

## Toward reliable ensemble Kalman filter estimates of CO<sub>2</sub> fluxes

Abhishek Chatterjee,<sup>1,2</sup> Anna M. Michalak,<sup>2</sup> Jeffrey L. Anderson,<sup>3</sup> Kim L. Mueller,<sup>1,4</sup> and Vineet Yadav<sup>2</sup>

Received 25 May 2012; revised 29 September 2012; accepted 11 October 2012; published 28 November 2012.

[1] The use of ensemble filters for estimating sources and sinks of carbon dioxide (CO<sub>2</sub>) is becoming increasingly common, because they provide a relatively computationally efficient framework for assimilating high-density observations of CO<sub>2</sub>. Their applicability for estimating fluxes at high-resolutions and the equivalence of their estimates to those from more traditional “batch” inversion methods have not been demonstrated, however. In this study, we introduce a Geostatistical Ensemble Square Root Filter (GENSRF) as a prototypical filter and examine its performance using a synthetic data study over North America at a high spatial (1° × 1°) and temporal (3-hourly) resolution. The ensemble performance, both in terms of estimates and associated uncertainties, is benchmarked against a batch inverse modeling setup in order to isolate and quantify the degradation in the estimates due to the numerical approximations and parameter choices in the ensemble filter. The examined case studies demonstrate that adopting state-of-the-art covariance inflation and localization schemes is a necessary but not sufficient condition for ensuring good filter performance, as defined by its ability to yield reliable flux estimates and uncertainties across a range of resolutions. Observational density is found to be another critical factor for stabilizing the ensemble performance, which is attributed to the lack of a dynamical model for evolving the ensemble between assimilation times. This and other results point to key differences between the applicability of ensemble approaches to carbon cycle science relative to its use in meteorological applications where these tools were originally developed.

**Citation:** Chatterjee, A., A. M. Michalak, J. L. Anderson, K. L. Mueller, and V. Yadav (2012), Toward reliable ensemble Kalman filter estimates of CO<sub>2</sub> fluxes, *J. Geophys. Res.*, 117, D22306, doi:10.1029/2012JD018176.

### 1. Introduction

[2] Over the last decade it has become increasingly apparent that quantification of global carbon sources and sinks with sufficient accuracy and precision is critical to balancing the global carbon budget and monitoring of carbon-management activities [Schimel, 2007]. It has also become clear that our understanding of, and ability to accurately model, the carbon-cycle is severely constrained by the sparse distribution of the present atmospheric CO<sub>2</sub> measurement network [Scholes et al., 2009]. The sparse and spatially non-uniform network is neither sufficient to constrain regional budgets with the needed certainty, nor understand the nature, geographic distribution and temporal variability of CO<sub>2</sub> sources and sinks. This absence of

spatially and temporally dense measurements of atmospheric CO<sub>2</sub> has spurred the development of space-based measurement sensors. Measurements from passive sensors such as the Atmospheric Infrared Sounder (AIRS) on Aqua [Chahine et al., 2006], the Tropospheric Emissions Spectrometer (TES) on Aura [Kulawik et al., 2010], the Infrared Atmospheric Sounding Interferometer (IASI) on Met-Op-1 [Crevoisier et al., 2009], the Scanning Imaging Absorption spectrometer for Atmospheric Cartography (SCIAMACHY) on EnviSAT [Buchwitz et al., 2005], the Greenhouse gases Observing SATellite (GOSAT) [Hamazaki et al., 2004], as well as planned future sensors such as the Orbiting Carbon Observatory-2 (OCO-2) [Crisp et al., 2008], and the Active Sensing of CO<sub>2</sub> Emissions over Nights, Days, and Seasons (ASCENDS) satellite [National Research Council, 2007] are expected to improve our scientific understanding of regional carbon cycle processes and budgets. Although remote-sensing measurements of CO<sub>2</sub> do not achieve the precision possible from in situ measurements [Rayner and O'Brien, 2001], they provide a large number of observations with near-global coverage, beyond what is possible from a surface network alone [e.g., Buchwitz et al., 2007].

[3] The global coverage provided by these space-based measurements has demonstrated promise in improving the accuracy and precision of regionally resolved flux estimates [e.g., Baker et al., 2010; Ciais et al., 2010], but the solution

<sup>1</sup>Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA.

<sup>2</sup>Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA.

<sup>3</sup>National Centre for Atmospheric Research, Boulder, Colorado, USA.

<sup>4</sup>Now at Science and Technology Policy Institute, Washington, D. C., USA.

Corresponding author: A. Chatterjee, Department of Global Ecology, 260 Panama St., Stanford, CA 94305, USA. (abhishch@umich.edu)

©2012. American Geophysical Union. All Rights Reserved.  
0148-0227/12/2012JD018176

of the associated inverse problem has also resulted in a substantial increase in computational cost. The computational challenge results from the fact that inverse modeling techniques (a.k.a. top-down approaches) have historically been solved in “batch” mode, where the inversion is performed by solving a system of linear equations relating the CO<sub>2</sub> fluxes and the atmospheric CO<sub>2</sub> observations [e.g., Enting, 2002]. Solving the batch problem requires an atmospheric transport model to be run either once per estimated flux region/period combination, or once per observation if an adjoint to the transport model is available. This becomes computationally infeasible given the increasing spatial and temporal resolution at which CO<sub>2</sub> fluxes are being estimated, and the increasing number of concentration measurements available from remote-sensing observations.

[4] To address the increasing computational challenge of atmospheric inversions, data assimilation (DA) techniques (e.g., ensemble Kalman filter methods [Peters et al., 2005; Feng et al., 2009; Miyazaki et al., 2011]; variational methods [Rayner et al., 2005; Chevallier et al., 2005; Rödenbeck, 2005; Baker et al., 2006], or hybrid approaches such as the Maximum Likelihood Ensemble Filter [Zupanski et al., 2007; Lokupitiya et al., 2008]) have recently been employed for estimating CO<sub>2</sub> fluxes, in some cases as part of advanced systems where meteorological and carbon variables are simultaneously assimilated [e.g., Kang et al., 2011]. Application of data assimilation techniques to the CO<sub>2</sub> problem is, however, much less mature [Rayner, 2010] than its use in numerical weather prediction [Swinbank, 2010, and references therein] or for assimilating other atmospheric constituents such as humidity and ozone [e.g., Rood, 2005; Lahoz and Errera, 2010]. An important question for carbon-cycle research that has hitherto remained unanswered concerns the impact of the numerical data assimilation framework on the precision and accuracy of fine-scale flux estimates and their associated uncertainties. Previous CO<sub>2</sub> DA studies have evaluated flux estimates by comparing them to biospheric model and inventory estimates and/or by assessing how well they reproduce available atmospheric CO<sub>2</sub> observations [e.g., Peters et al., 2005; Chevallier et al., 2007; Lokupitiya et al., 2008; Feng et al., 2009; Baker et al., 2010; Kang et al., 2011; Miyazaki et al., 2011]. Given the host of error sources (e.g., transport error, aggregation error, etc.) that impact inversions, these diagnostics provide an assessment of the overall inversion framework, but do not isolate any errors incurred due to the numerical approximations in the implemented DA approach.

[5] This study is primarily motivated by an attempt to isolate and quantify such errors, specifically from the perspective of an ensemble Kalman filter applied to the estimation of CO<sub>2</sub> fluxes at fine spatial and temporal scales. Ensemble filters and their variants [e.g., Peters et al., 2005, 2007, 2010; Zupanski et al., 2007; Lokupitiya et al., 2008; Feng et al., 2009, 2011; Kang et al., 2011; Miyazaki et al., 2011] have gained popularity within the carbon-cycle community due to their simple conceptual formulation and relative ease of implementation. So far, the examination of the use of ensemble Kalman filters for estimating fluxes at fine spatial and temporal scales has been limited, however. Except for Kang et al. [2011] and Miyazaki et al. [2011], where fluxes were estimated at the grid resolution of the atmospheric transport model used in the studies (~2.8°), almost all other studies have estimated fluxes

at large spatial scales (e.g., continental or ecoregion). The temporal scales at which fluxes have been estimated range from several days to weeks.

[6] The work presented here estimates fluxes at substantially finer scales (1° by 1° and 3-hourly) relative to previous application of ensemble filters. In general, high resolution estimates of carbon fluxes are advantageous for 1) improving budgeting and mechanistic understanding of the carbon cycle at local to regional scales, and 2) minimizing spatial and temporal aggregation errors [e.g., Kaminski et al., 2001; Peters et al., 2010; Gourdji et al., 2010] that may otherwise bias the final flux estimates. The impact of spatial aggregation errors has long been discussed and documented in the inverse modeling literature [e.g., Kaminski et al., 2001; Engelen et al., 2002; Peters et al., 2010], and recent studies have shown that a priori temporal aggregation has similar impacts. Gourdji et al. [2010, 2012] found that biases occurred when fluxes were estimated at multiday or even daily timescales, and recommended a 3-hourly temporal resolution to allow the inversion to resolve the diurnal cycle. Huntzinger et al. [2011] further found that differences between the diurnal representations among a suite of terrestrial ecosystem models yielded significant difference at CO<sub>2</sub> monitoring locations, suggesting that adopting a fixed diurnal cycle from one particular model a priori could bias flux estimates at larger scales. Although desirable from a scientific perspective, applying an ensemble approach to a fine-scale flux estimation problem is challenging due to two issues associated with the ensemble filter.

[7] The first challenge common to all applications of ensemble filters is the error due to representing the probability density function of the fluxes by a finite number of randomly generated flux realizations or system states. Experience in the NWP area has suggested that because of the finite number of ensemble members, the ensemble filter can suffer from variance underestimation, rank deficiency and sampling error [e.g., Houtekamer and Mitchell, 2005; Anderson, 2007a, 2007b; Ehrendorfer, 2007; Meng and Zhang, 2011], all of which impact both the final estimates and their uncertainty. Anderson [2007a] notes that, even in low-order perfect-model applications for NWP, mitigating the impacts of the limited ensemble size requires the introduction and tuning of several additional algorithms. Considerable expertise exists in these algorithms for NWP [e.g., Hamill and Whitaker, 2005; Anderson, 2007a, 2007b; Uzunoglu et al., 2007; Sacher and Bartello, 2008; Anderson, 2009; Bergemann and Reich, 2010; Bishop and Hodyss, 2011] and for DA of other constituents [e.g., Schutgens et al., 2010]. However, these algorithms and their impact on flux estimates and uncertainties are less well-understood for the carbon flux estimation problem. Applications to CO<sub>2</sub> have investigated the impact of ensemble size [e.g., Peters et al., 2005; Zupanski et al., 2007; Feng et al., 2009; Miyazaki et al., 2011] and different localization/inflation parameters (see Sections 2.3 and 2.4) but have refrained from drawing conclusions as to the optimal values of parameters that may aid future filter designs. Because previous studies have not compared ensemble filter estimates to those from batch inversions, it is difficult to isolate the impact of the parameter and algorithm choices from other errors present within any inversion framework.

[8] The second challenge that differentiates the CO<sub>2</sub> flux estimation problem from NWP-related applications is that there is currently no dynamical model to directly evolve the

carbon flux state vector forward in time [Peters et al., 2005; Lokupitiya et al., 2008; Miyazaki et al., 2011]. In other words, given the estimated flux at one time, there is no model to predict the flux at the following assimilation time. The lack of such a dynamical model represents a loss of valuable information to the ensemble, as along with the transport model, a dynamical model would capture the flow-dependent error covariance patterns. In NWP-related applications, several studies have been carried out to test the impact of dynamical model errors [e.g., Houtekamer et al., 2005; Szunyogh et al., 2005; Houtekamer et al., 2009; Hamill and Whitaker, 2011], but no study has evaluated the impact of a complete absence of a dynamical model. The absence of the dynamical model may make the ensemble filter extremely sensitive to the observation network and coverage. Given the spatial and temporal variability of atmospheric CO<sub>2</sub> measurements (whether in situ or satellite-based), this raises questions about the applicability of ensemble filters for leveraging the information content of available CO<sub>2</sub> observations.

[9] In order to understand these issues, we introduce a geostatistical variant of the Bayesian ensemble square root filter (EnSRF) [Whitaker and Hamill, 2002]. The geostatistical ensemble square root filter (GEnSRF) is based on a geostatistical inverse modeling (GIM) formulation of the flux estimation problem [Michalak et al., 2004]. The GIM formulation is not limited to the use of prior CO<sub>2</sub> flux information from biospheric models and/or inventories, and has been applied for inversions conducted at very high spatiotemporal resolutions [e.g., Gourdji et al., 2012].

[10] The GEnSRF is used as a prototype filter in exploring the impacts of the challenges outlined above for CO<sub>2</sub> flux estimates at fine spatial and temporal scales. The sensitivity of the ensemble filter to different scenarios is judged by comparing the GEnSRF estimates to the estimates from an equivalent batch GIM setup. This comparison is carried out using synthetic data from the growing season (June 2008) over North America. Both GEnSRF and GIM are used to estimate fluxes and their associated uncertainties at a 1° × 1° (spatial) and 3-hourly (temporal) resolution. Test cases are designed to gauge the performance of the ensemble system and to evaluate whether the numerically approximate ensemble scheme can accurately capture the characteristic features of the CO<sub>2</sub> cycle, such as the spatial location of sources and sinks and the amplitude and phase of the diurnal flux cycle. The test cases are used to assess the baseline performance of the ensemble system, as well as to explore the impact of the measurement network, ensemble size, and the implementation of covariance inflation and localization algorithms designed to improve ensemble performance.

[11] Overall, this work provides (1) an assessment of the relative performance of the ensemble filter in comparison to the batch approach and of the conditions necessary for the ensemble approach to be a suitable replacement for batch inversions, and (2) an investigation of the error sources in the ensemble system and their implications for adjustments to ensemble systems that need to be made relative to NWP applications. The remainder of this paper is organized as follows. Section 2 provides the rationale for the proposed filter followed by an overview of GEnSRF. Section 3 provides a description of the examined synthetic data test cases. Results are presented and discussed in Section 4. Finally, we

conclude in Section 5 with a summary of the findings of this study and recommendations for future research.

## 2. Methodology

### 2.1. Choosing a Filter Formulation

[12] The underlying framework in all ensemble filters is a low-rank ensemble representation of the error covariance matrices. The ensembles themselves are scaled matrix square-roots of the covariance matrices, and are updated during the assimilation of observations either stochastically [e.g., Houtekamer and Mitchell, 1998; Burgers et al., 1998; Pham, 2001] or deterministically [e.g., Bishop et al., 2001; Anderson, 2001; Whitaker and Hamill, 2002; Ott et al., 2004]. The details of this update step distinguish most ensemble variants. Based on existing studies [e.g., Tippett et al., 2003; Lawson and Hansen, 2004; Nerger et al., 2005; O’Kane and Frederiksen, 2008], it can be concluded that for a linear problem – (1) deterministic filters are more accurate than their stochastic counterparts, and (2) although all the deterministic filters will produce analysis ensembles that span the same state subspace and have the same covariance, implementation of a serial EnSRF has the lowest computational cost, if the observation errors are assumed to be independent.

[13] The simplest serial EnSRF that can be implemented for inferring CO<sub>2</sub> surface fluxes is one using a Bayesian formulation (e.g., CarbonTracker [Peters et al., 2010]), which uses prior information about the CO<sub>2</sub> fluxes from bottom-up models and/or inventories. Because of the highly ill-posed nature of the CO<sub>2</sub> flux estimation problem and the sparseness of the current observing network, the posterior flux estimates and uncertainties are quite sensitive to the a priori prescribed flux patterns and their associated error covariance parameters [Peters et al., 2010]. By adapting the ensemble system to the geostatistical approach, we avoid some of the reliance on prior/model assumptions, albeit at the cost of an increase in complexity. Therefore, it can be argued that the niche filled by a geostatistical ensemble square root filter lies in more directly isolating the information content of the available atmospheric measurements.

### 2.2. Geostatistical Ensemble Square Root Filter

[14] The GEnSRF, like the EnSRF, is a Monte Carlo technique based on a state-space formulation of the Kalman Filter using an ensemble of model states to represent, propagate and update the estimates of the state and state error covariance. The aim is to minimize a cost function of the form:

$$L(\mathbf{s}, \beta) = \frac{1}{2} [\mathbf{z} - h(\mathbf{s})]^T \mathbf{R}^{-1} [\mathbf{z} - h(\mathbf{s})] + \frac{1}{2} [\mathbf{s} - \mathbf{X}\beta]^T \mathbf{Q}^{-1} [\mathbf{s} - \mathbf{X}\beta] \quad (1)$$

where  $\mathbf{z}$  is a  $n \times 1$  vector of observations,  $h$  represents the atmospheric transport model,  $\mathbf{s}$  is a  $m \times 1$  state vector composed of the discretized *unknown* surface flux distribution,  $\mathbf{R}$  is the  $n \times n$  model-data mismatch covariance,  $\mathbf{X}$  is an  $m \times 1$  vectors of ones in the test cases presented in the work, but could also include auxiliary variables related to carbon flux [see Gourdji et al., 2008, 2012] for further details on the selection of auxiliary variables),  $\beta$  is an unknown constant

here, but could also include unknown drift coefficients that scale the auxiliary variables in  $\mathbf{X}$ , the prior covariance matrix  $\mathbf{Q}$  describes the expected variability in flux departures from  $\mathbf{X}\beta$  as a function of the separation distance in space and time between fluxes (see Section 3.1.3 for further details on the structure of  $\mathbf{Q}$ ), and  $T$  represents the transpose operator. In a batch setup, instead of running the transport model  $h$  directly as part of the inversion, an  $n \times m$  sensitivity matrix  $\mathbf{H}$  (a.k.a. Jacobian matrix) is generated that represents the sensitivity of the observations  $\mathbf{z}$  to the fluxes  $\mathbf{s}$  (i.e.,  $H_{ij} = \partial z_i / \partial s_j$ ).

[15] Equation (1) represents a compromise between reproducing the atmospheric measurements ( $\mathbf{z}$ ) and staying close to the statistical model of the trend ( $\mathbf{X}\beta$ ), where the covariance matrices determine the relative weight of these competing objectives. Although some implementations of ensemble approaches include more variable in the state vector  $\mathbf{s}$ , including atmospheric concentrations of CO<sub>2</sub> themselves, the focus here is on constraining only the underlying fluxes. Correspondingly, any updates in the atmospheric CO<sub>2</sub> distribution must therefore be attributable to a change in the underlying fluxes.

[16] The GENSRF is implemented as a smoother [e.g., Bruhwiler *et al.*, 2005; Michalak, 2008], such that individual time steps through the smoother include (a) fluxes that are no longer being estimated, (b) fluxes that are being updated at least for the second time (i.e., that have been previously estimated), and (c) fluxes being estimated for the first time (i.e., for which no prior information is available). In the following discussion, the  $m \times 1$  vector of the *estimated* surface flux distribution is represented as  $\hat{\mathbf{s}}$ .  $\hat{\mathbf{s}}_j$  denotes estimates of fluxes that are being updated at least for the second time,  $\hat{\mathbf{s}}_k$  denotes fluxes that are being obtained for the first time, and  $\hat{\mathbf{s}}_{j+k}$  denotes both sets of fluxes being estimated. Finally, the superscripts  $a$  and  $b$  represents the analyzed (or updated) estimate and the previous (or background) estimate.

[17] Given an initial prior covariance  $\mathbf{Q}^b$ , GENSRF starts by creating an ensemble of  $N$  state fields (where  $N \ll m$ ). These are created as unconditional realizations of the matrix  $\mathbf{Q}^b$  through Cholesky decomposition.

$$\mathbf{Q}^b \approx \frac{1}{N-1} (\mathbf{s}^b) (\mathbf{s}^b)^T \quad (2)$$

where  $\mathbf{s}^b$  represents the estimated error statistics of CO<sub>2</sub> flux deviations from the trend. In the limit of  $N \rightarrow \infty$  this representation of  $\mathbf{Q}^b$  is exact. In GENSRF, observations are assimilated serially. When the  $i^{\text{th}}$  observation is being assimilated, the estimates of fluxes are given by:

$$\begin{bmatrix} \hat{\mathbf{s}}_j^a \\ \hat{\mathbf{s}}_k^a \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{s}}_j^b \\ \mathbf{0} \end{bmatrix} + \Lambda \left[ z_i - h_i(\hat{\mathbf{s}}_j^b) \right] \quad (3)$$

where  $\Lambda$  is calculated by solving the following system of equations,

$$\begin{bmatrix} \mathbf{H}_{i,j+k} \mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T + \mathbf{R}_{i,i} & h_i(\mathbf{X}_k) \\ (h_i(\mathbf{X}_k))^T & 0 \end{bmatrix} \begin{bmatrix} \Lambda^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} (\mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T)^T \\ [\mathbf{0} \quad \mathbf{X}_k^T] \end{bmatrix} \quad (4)$$

[18] Consistent with a GIM framework, fluxes being estimated for the first time ( $\hat{\mathbf{s}}_k$ ) need not be initialized with a prior value, and the zero in equation (3) is not equivalent to a

prior in the classical Bayesian setup. Since direct matrix computation of  $\mathbf{H}_{i,j+k} \mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T$  and  $\mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T$  can be expensive, these are approximated by running the transport model directly with the ensemble of state deviations.

$$\mathbf{H}_{i,j+k} \mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T \approx \frac{1}{N-1} \left( h_i(\mathbf{s}_{j+k}^b) \right) \left( h_i(\mathbf{s}_{j+k}^b) \right)^T \quad (5)$$

$$\mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T \approx \frac{1}{N-1} \left( \mathbf{s}_{j+k}^b \right) \left( h_i(\mathbf{s}_{j+k}^b) \right)^T \quad (6)$$

[19] Once  $\Lambda$  is obtained, it is used in equation (3) to estimate the fluxes. If the same  $\Lambda$  is used to update the ensemble of state deviations from the mean  $\mathbf{s}_{j+k}^b$ , it would result in an underestimation of the analysis error covariance [Whitaker and Hamill, 2002]. Instead  $\Lambda$  is reduced in magnitude ( $\tilde{\Lambda}$ ; equation (7)) such that the spread of the ensemble is reduced less by the analysis (equation (8)), in order to maintain an error covariance consistent with the full-rank Kalman filter.

$$\tilde{\Lambda} = \left( 1 + \sqrt{\frac{\mathbf{R}_{i,i}}{\mathbf{H}_{i,j+k} \mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T + \mathbf{R}_{i,i}}} \right)^{-1} \Lambda \quad (7)$$

$$\mathbf{s}_{j+k}^a = \mathbf{s}_{j+k}^b - \tilde{\Lambda} \left( h_i(\mathbf{s}_{j+k}^b) \right) \quad (8)$$

[20] When observations are serially processed, equation (7) reduces to the computation of a scalar factor. Notice that the piece  $h_i(\mathbf{s}_{j+k}^b)$  is already available, and hence updating the ensemble via equation (8) is no more computationally expensive than equation (3).

[21] Finally, before assimilating the next observation, we update the sampled observational ensemble and the sampled observation state corresponding to all future observations that are yet to be assimilated:

$$h_{i+1}(\mathbf{s}^a) = h_{i+1}(\mathbf{s}^b) - h_{i+1} \left( \tilde{\Lambda} h_i(\mathbf{s}_{j+k}^b) \right) \quad (9)$$

$$h_{i+1}(\hat{\mathbf{s}}^a) = h_{i+1}(\hat{\mathbf{s}}^b) + h_{i+1} \left( \Lambda \left[ z_i - h_i(\hat{\mathbf{s}}_j^b) \right] \right) \quad (10)$$

[22] Equations (9) and (10) require two additional transport model runs that could be avoided as described in Peters *et al.* [2005] by approximating these equations in a manner similar to equation (5). When the ensemble size is much smaller than the size of the state vector (as will be the case when fluxes are estimated at a high spatial and temporal resolution) this results in a poor approximation, however. The additional cost of running the transport model might well offset the errors incurred due to the approximation. However, with satellite measurements, running the transport model each time an observation is assimilated makes the direct implementation of equations (9) and (10) impractical. Hence, work is underway to find suitable alternatives to both these equations without incurring large errors in the analysis. In the work presented here, these additional runs are performed.

[23] Using equations (3) to (10), a best estimate of CO<sub>2</sub> fluxes is obtained. Once all observations have been

assimilated, the a posteriori covariance  $\mathbf{Q}^a$  for the flux estimates is reconstructed from the ensemble (equation (11)). The diagonal values of this posterior covariance matrix correspond to the uncertainty (expressed as a variance) of each estimated flux component in  $\hat{\mathbf{s}}$ .

$$\mathbf{Q}^a = \frac{1}{N-1} (\mathbf{s}^a) (\mathbf{s}^a)^T. \quad (11)$$

[24] Finally, in regular NWP applications a dynamical model (nonlinear forecast operator) would have been used to propagate the state vector between the two observational time periods. For the CO<sub>2</sub> problem, no suitable deterministic model is available to directly propagate fluxes from one time step to the next. This differentiates the CO<sub>2</sub> flux estimation problem from the NWP and other trace gas assimilation problems, and may have critical implications for good filter performance. Note that this drawback is not specific to GEnSRF but to all variants of the ensemble filter that have been employed for CO<sub>2</sub> flux estimation.

### 2.3. Covariance Localization

[25] Covariance localization aims to heuristically improve the error covariance estimates in the case of small ensemble sizes. In all ensemble filters (including GEnSRF), the number  $N$  of ensemble members is small relative to the size  $m$  of the state-space, hence the representation of the prior covariance matrix in  $N$ -dimensional space is not perfect. This results in a number of erroneous flux correlations as a consequence of which a state variable may be incorrectly impacted by an observation that is physically remote.

[26] Several covariance localization techniques have been proposed for the NWP problem [Houtekamer and Mitchell, 2001; Hamill et al., 2001; Ott et al., 2004; Anderson, 2007b] to account for the statistical noise of the ensemble. For CO<sub>2</sub> applications, implemented localization schemes have varied depending on the particular ensemble filter variant being used. For example, Peters et al. [2005] chose a simple exponential decay function, while Miyazaki et al. [2011] subjectively specify different cutoff radii based on the type and location of observation data used in their analysis. Zupanski et al. [2007] and Lokupitiya et al. [2008] chose a more dynamic scheme based on information theory, where the localization length scale is a function of the information content in the assimilated observations.

[27] Similarly to Peters et al. [2005], we implement a simple covariance localization scheme in GEnSRF. This is achieved by performing a Schur (or Hadamard [Horn and Mathias, 1990]) product, or element-wise multiplication (denoted  $\bullet$  in equation (12)) of a correlation matrix  $\rho$  with the covariance model generated by the ensemble as shown in equation (12).

$$\mathbf{Q}_{j+k,j+k} \mathbf{H}_{i,j+k}^T \approx \rho \bullet \left[ \frac{1}{N-1} (\mathbf{s}_{j+k}^{rb}) (h_i (\mathbf{s}_{j+k}^{rb}))^T \right]. \quad (12)$$

[28] Here,  $\rho$  is defined using a standard fifth-order Gaspari-Cohn function [Gaspari and Cohn, 1999] with a finite length scale. Note that Peters et al. [2005] used an exponential decay function to define their  $\rho$ . Both the Gaspari-Cohn function and the exponential function are compactly supported [Gneiting, 2002; Bergemann and Reich, 2010], which means that the function is nonzero in only a small (local)

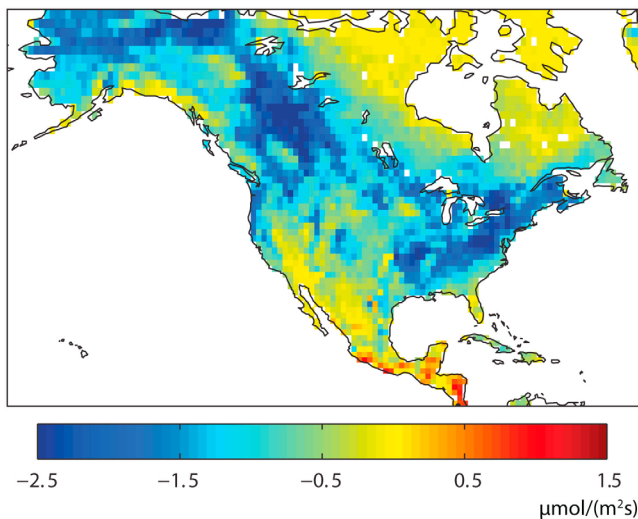
region specified by a length scale. We find that the overall conclusions presented in Section 4.2 are valid for a variety of compactly supported functions. The key ingredient in all compactly supported functions is the length scale, which ensures that spurious correlations are removed, but correctly specified physical correlations are maintained and not excessively damped.

[29] Covariance localization using the Schur product might be a simple approach to increase the effective rank of the covariance matrix, but there are several important caveats for CO<sub>2</sub> applications. Previous studies [e.g., Lokupitiya et al., 2008] have raised questions regarding selection of an appropriate localization length scale, and whether the atmospheric advection of CO<sub>2</sub> is consistent with the use of a compactly supported correlation function such as the Gaspari-Cohn. Since the prior covariance matrix holds information on the spatial and temporal autocorrelation of flux deviations from the trend (Section 3.1.3), by including the Schur product and thereby modifying this matrix, covariance localization may disrupt the autocorrelation structure (see Karspeck and Anderson [2007], Oke et al. [2007], and Kepert [2009] for a similar discussion related to NWP problems). In spite of these concerns, in this study, we have persisted with the Gaspari-Cohn function because we want to assess the applicability of this simple scheme for atmospheric CO<sub>2</sub> inversions. Future work could explore the applicability of more sophisticated dynamic localization schemes [e.g., Zupanski et al., 2007] or balance-aware localization schemes proposed for NWP [e.g., Bishop and Hodyss, 2011; Kepert, 2011; Jun et al., 2011] or more adaptive techniques based on the prior ensemble [e.g., Anderson, 2012].

### 2.4. Adaptive Covariance Inflation

[30] A second algorithm for combatting insufficient variance in ensemble filters is covariance inflation. Insufficient variance (or undersampling) is primarily caused by sampling error resulting from the use of small ensembles [Furrer and Bengtsson, 2007]. Over successive assimilation periods, undersampling can become more severe, and in the worst case can lead to filter divergence, where the filter effectively rejects the observations and the assimilation reduces to the prior. Also, because the posterior analysis error covariance is generated from the ensemble at the end of the assimilation (equation (11)), insufficient variance leads to an under-estimation of the analysis error covariance (i.e., the flux uncertainties).

[31] Several ad hoc and adaptive techniques have been proposed in NWP applications to counter this loss of variance [e.g., Anderson and Anderson, 1999; Zhang et al., 2004; Hamill and Whitaker, 2005; Anderson, 2007a; Li et al., 2009; Anderson, 2009; Peña et al., 2010]. Various inflation schemes have also been employed with CO<sub>2</sub> applications, depending on the variant of ensemble Kalman filter used in a particular study. Some have avoided using the ensemble spread as a measure of uncertainty altogether by instead deriving final uncertainties from a set of sensitivity experiments [e.g., Peters et al., 2010]. Feng et al. [2009] chose to use an ensemble of the same size as the state vector, thereby minimizing undersampling directly. For the maximum likelihood ensemble filter [e.g., Zupanski et al., 2007; Lokupitiya et al., 2008] a multiplicative inflation scheme was used for covariance inflation, where the ensemble is inflated by a



**Figure 1.** “True” CASA-GFEDv2 fluxes aggregated to the monthly scale.

constant factor that is homogeneous in space (although different inflation factors are used for land and ocean regions) and time. This approach has some limitations because neither the observation network nor the CO<sub>2</sub> dynamics are homogeneous in space and time, and the cost of tuning experiments to find an appropriate inflation factor that is applicable everywhere is prohibitive. Recognizing these drawbacks, more recent studies have employed either conditional covariance inflation [Miyazaki *et al.*, 2011] or a mix of adaptive and covariance relaxation techniques [Kang *et al.*, 2011].

[32] In GEnSRF, we adopt the more generalized version of the adaptive technique used by Kang *et al.* [2011] (as originally proposed by Anderson [2009]) to calculate spatially and temporally varying inflation factors for each state component (i.e., flux at each time and grid point). This adaptive algorithm applies Bayesian estimation theory to the probability density function of the inflation factors. First, a normally distributed inflation random variable is associated with each element of the state vector. Then, via Bayes theorem, these inflation factors are incrementally updated during serial assimilation of the observations. Note that the atmospheric CO<sub>2</sub> observations can be used to optimize the inflation factors for the CO<sub>2</sub> fluxes due to the link between these quantities provided by the atmospheric transport model.

[33] In order to calculate the spatially and temporally varying inflation factors, however, it is necessary to implement covariance localization first. The adaptive technique uses sample correlations of the ensemble between observation space and the model space to convert the inflation estimates in the observation space to those in the model space. Covariance localization plays an important role in reducing the sampling noise in the sample correlations. If no covariance localization is pursued, then the sampling error manifests itself in the adaptive inflation step resulting in spurious inflation factors. Thus, using the adaptive technique, we have specifically adjusted the covariance inflation strategy to take into account the information provided by the atmospheric CO<sub>2</sub> measurements.

[34] We refer the reader to Anderson [2009] (and the subsequent review by Miyoshi [2011]) for the mathematical

underpinnings of the adaptive approach. It is worthwhile to reiterate that this particular adaptive technique has not previously been applied to any CO<sub>2</sub> inversion study. Hence, as part of the sensitivity tests described later, we will examine both the advantages and disadvantages of this adaptive technique for the CO<sub>2</sub> source-sink estimation problem.

### 3. Sample Application

[35] The GEnSRF approach is applied to a synthetic data study over the North American continent (Section 3.1). A series of analyses are designed (Section 3.2) to compare the estimates from GEnSRF with the estimates from GIM. These comparisons are done by aggregating the posterior estimates to a range of spatial and temporal scales (Section 3.3) to assess the accuracy and precision of the ensemble approach relative to a batch inversion.

[36] In the context of this study, the main advantage of the GIM approach relative to the GEnSRF technique is that it solves the entire system of equations analytically (without any approximations) and hence provides a “gold” standard for evaluating the ensemble results. By keeping the atmospheric data sets consistent for GEnSRF and GIM, it is possible to isolate the degradation due to the numerical approximations in the ensemble filter framework. The drawback of this setup is that the effects of the transport model errors have been removed by using the same transport model to both create the synthetic measurements as well as estimate the fluxes in the inversion.

#### 3.1. Experimental Design

##### 3.1.1. Flux Data and Basis Functions

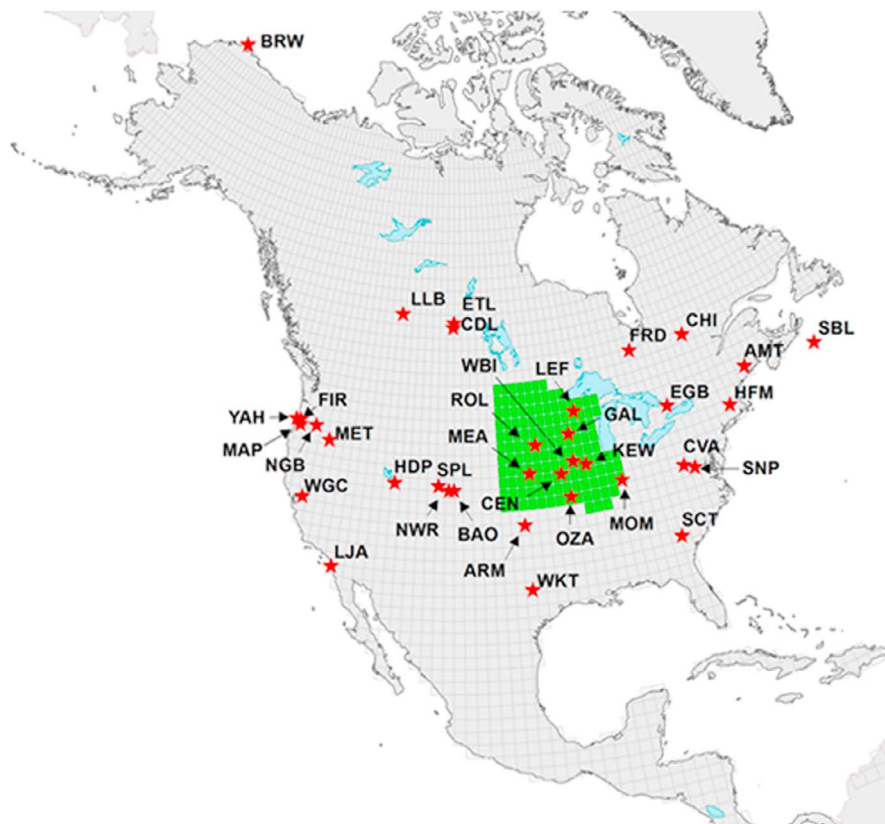
[37] Biospheric fluxes from the Carnegie Ames Stanford Approach terrestrial carbon cycle model, as configured for the Global Fire Emissions Database v2 project (henceforth referred to as CASA-GFEDv2) [Randerson *et al.*, 1997; van der Werf *et al.*, 2006] are used as the true fluxes for generating the synthetic atmospheric data. The monthly averaged 1° × 1° CASA-GFEDv2 Net Ecosystem Exchange (NEE) for June 2008 (Figure 1) is temporally downsampled to 3-hourly resolution using the method of Olsen and Randerson [2004].

[38] The sensitivity matrix  $\mathbf{H}$  is obtained by coupling the Weather Research Forecasting (WRF) model [Skamarock *et al.*, 2005] to the Stochastic Time-Inverted Lagrangian Transport Model (STILT) [Lin *et al.*, 2003], as outlined in Gourdjji *et al.* [2010]. Calculating and pre-storing the sensitivity matrix  $\mathbf{H}$  is necessary for performing the batch GIM analysis but not for GEnSRF, where the transport model can be run directly as part of the DA system. Given that  $\mathbf{H}$  was available in this case, it is also used as the transport model for the ensemble implementation.

##### 3.1.2. Synthetic Observation Data

[39] The basis functions generated via WRF-STILT are used with the CASA-GFEDv2 fluxes to generate the synthetic observations  $\mathbf{z}$  (i.e.,  $h(\mathbf{s})$ ) for the 35 continuous observation towers (see Table S1 in auxiliary material Text S1) that were operational in June 2008 (Figure 2).<sup>1</sup> First, a full set of synthetic data is generated for all the towers at the 3-hourly scale, and small random errors (standard deviation of 0.1 ppm)

<sup>1</sup>Auxiliary materials are available in the HTML. doi:10.1029/2012JD018176.



**Figure 2.** Location of the 35 tower network (stars), and the regions used for interpreting the flux estimates, i.e., North America and the MCI region (green shaded area). The background grid represents the flux estimation resolution of  $1^\circ \times 1^\circ$ . The three-letter codes for the towers are defined in Table S1 in auxiliary material Text S1.

are then added to the synthetic data. Such small errors were used to represent an, albeit somewhat unrealistic, best case scenario for the performance of the ensemble approach. Next only afternoon measurements are retained for the shorter towers (height  $\leq 150$  m) consistent with typical data choices in inversion studies [e.g., Göckede *et al.*, 2010; Gourdji *et al.*, 2010], motivated by lower transport model errors for afternoon conditions [e.g., Geels *et al.*, 2007; Gerbig *et al.*, 2008]. Finally, data gaps are simulated in the synthetic observations consistent with missing data from the actual June 2008 observations (due to either instrument down time or calibration needs). By mimicking the random outage in the collection/storage of the data, the synthetic data set is highly variable (in both space and time), but realistic. The ratio of the number of fluxes ( $m$ ) to the number of observations ( $n$ ) is on the order of  $\sim 250:1$ . Conversely, if the full set (i.e., eight 3-hourly averaged observations per day) of synthetic observations without data gaps were retained, then the ratio would be on the order of  $\sim 75:1$ .

### 3.1.3. Error Covariance Matrices

[40] The model-data mismatch covariance matrix  $\mathbf{R}$  is a diagonal matrix, with values of  $0.01 \text{ ppm}^2$  along the diagonal (i.e., all towers are assumed to have the same model-data mismatch error), corresponding to the variance of the errors introduced into the synthetic observations.

[41] The prior covariance matrix  $\mathbf{Q}^b$  captures the spatio-temporal autocorrelation of the flux deviations from the

model of the trend  $\mathbf{X}\beta$ . In this study, only spatial correlation is assumed a priori in order to keep the structure of  $\mathbf{Q}^b$  simple, although accounting for both spatial and temporal correlation could further improve estimates [e.g., Gourdji *et al.*, 2010; Chevallier *et al.*, 2012].

[42]  $\mathbf{Q}^b$  is prescribed as a block diagonal matrix, with each block describing the correlation between grid-scale fluxes for each time period of the inversion. Based on previous work [e.g., Michalak *et al.*, 2004; Gourdji *et al.*, 2010], each block is modeled by an exponential covariance function:

$$\mathbf{Q}^b(d|\sigma^2, l) = \sigma^2 \exp\left(-\frac{d}{l}\right) \quad (13)$$

where  $d$  is the spatial separation distance between the grid points where fluxes are to be estimated,  $\sigma^2$  represents the variance of the flux residuals at large separation distances, and  $l$  is the range parameter. The correlation length beyond which correlation between the flux residuals becomes negligible is approximately  $3l$  [Chilès and Delfiner, 1999] for an exponential model.

[43] The covariance parameters in  $\mathbf{Q}^b$  (i.e.,  $\sigma^2$  and  $l$ ) can be obtained via different methods [e.g., Michalak *et al.*, 2005; Rödenbeck, 2005; Chevallier *et al.*, 2010] ranging from analyzing the variability in biospheric model outputs to statistically inferring these parameters directly from the atmospheric measurements. In this study, we follow the

latter approach, and optimize for the covariance parameters using the Restricted Maximum Likelihood [e.g., *Kitanidis*, 1995; *Michalak et al.*, 2004] approach.

### 3.2. Test Cases

[44] Two primary inversion setups (TC1 and TC2) are outlined, both of which estimate 3-hourly fluxes at a  $1^\circ \times 1^\circ$  scale over North America for the month of June 2008. However, TC1 uses a sparse measurement data set (as described in Section 3.1.2), while TC2 uses all 24 h of measurements for all towers, yielding a temporally denser and homogeneous data set. Covariance parameters were estimated separately for the two test cases. GIM is run once for each test case to obtain the batch estimates, while the GEnSRF is run multiple times for both TC1 and TC2 with different configurations of ensemble size, localization and inflation parameters. The details of these runs are expanded upon in the following paragraphs.

[45] A control run of GEnSRF is defined based on TC1 with a 500-member ensemble and without any covariance localization or adaptive inflation. This run, designated as TC1E500, is used to gauge the incremental benefits of subsequent modifications.

[46] Given the absence of a dynamical model to propagate the state vector forward in time, our hypothesis is that the inversion conditions (at least in terms of measurement quantity and density) may play a significant role in the performance of the ensemble filter. *Miyazaki et al.* [2011] concluded that the absence of a dynamical model resulted in the posterior analysis being sensitive to the initial error covariance, but this earlier study did not test the influence of the measurement network. TC2 is designed to explore the impact of the measurement network sparseness as it represents the best possible scenario that one can attain with the existing ground-based continuous measurement network. If this were a real-data study, there would be several caveats regarding using all 3-hourly measurements, especially from shorter towers [*Gourdji et al.*, 2012]. Hence other inversion scenarios in which measurements are progressively reduced in space and time were also evaluated. The conclusions from these additional test cases mirrored those from TC1 and TC2, and hence these have been omitted here for the sake of brevity.

[47] In order to provide insight into an optimal and practical setup of the ensemble filter that can provide accurate flux estimates of CO<sub>2</sub> with reliable uncertainties, the parameters of the ensemble system were varied for both TC1 and TC2. The GEnSRF is run with three ensemble sizes – 100, 500, and 2500, denoted as E100, E500, and E2500, respectively. In addition, three different localization length scales were prescribed – 500, 1500 and 3000 km, denoted as L500, L1500 and L3000, respectively. Finally, the adaptive inflation algorithm requires a priori estimates of inflation factors and their associated variance. Again three different specifications of the prior inflation variance were provided – a prior inflation factor of 1 with a standard deviation of 0.01, a prior inflation factor of 1 with a standard deviation of 0.05, and a prior inflation factor of 1 with a standard deviation of 0.25. These runs are denoted as I001, I005 and I025, respectively. These parameters were chosen based on a combination of extensive literature review of ensemble filter applications, subjective knowledge of CO<sub>2</sub> transport and its correlation scales, and some preliminary testing with a 1D advection-diffusion problem.

[48] Overall, a total of 2 GIM and 56 GEnSRF runs were carried out for this study. The 2 GIM runs represent the batch estimates for each of the test cases, and are simply denoted as GIM TC1 and GIM TC2. For GEnSRF, the first run for each setup is with an ensemble size of 500 and without any localization and inflation applied (i.e., the control run GEnSRF TC1E500 and GEnSRF TC2E500). GEnSRF is then run with varying ensemble sizes, localization length scales, and different a priori inflation values as described above. As an example, the GEnSRF run for setup TC1 using a 500 member ensemble, a localization length scale of 1500 km and a prior inflation factor of 1 with a standard deviation of 0.05 is denoted GEnSRF TC1E500\_L1500I005.

[49] Finally, as mentioned previously, the covariance parameters in  $\mathbf{Q}^b$  (equation (13)) are optimized separately for the two inversion setups. The flux standard deviation ( $\sigma$ ) was  $6.7 \mu\text{mol}/(\text{m}^2 \text{ s})$  for TC1 and  $6.4 \mu\text{mol}/(\text{m}^2 \text{ s})$  for TC2, while the correlation length ( $3l$ ) was 1630 km for TC1 and 1590 km for TC2. For all the runs, GEnSRF was spun up for 8 days prior to June 1, 2008. The lag window for the smoother was set to 10 days to take into account that CO<sub>2</sub> information is preserved over the continent for a maximum of 10 days. Note that a much longer lag window would have been required for global applications where there is no finite residence time for an air mass in the domain, or if substantial flux temporal correlation had been assumed a priori. Even with a 10-day window longer integrations of the transport model are required as well as more parameters need to be estimated. Thus to represent the covariance matrix properly, it becomes necessary to have a large number of ensemble members.

### 3.3. Evaluating the Analysis

[50] The posterior flux estimates from the different GEnSRF runs and GIM are compared using both quantitative and qualitative metrics at different spatial and temporal scales. Results are presented for a subset of GEnSRF runs (Table 1) that answer the specific questions posed in the study, and other setups are discussed where appropriate.

[51] In terms of time-averaged diagnostics, the two quantitative metrics used are the root mean square difference (RMSD) and the correlation coefficients (CC). The GEnSRF and the GIM flux estimates are aggregated to a monthly timescale and the RMSD and the CC calculated at the native  $1^\circ \times 1^\circ$  spatial resolution for all grid-cells across the continent. Both these quantities are reported aggregated over North America (NA) and the Mid-Continent Intensive (MCI) region. The MCI region [e.g., *Lauvaux et al.*, 2012] that is shown as the green shaded area in Figure 2, was not only well constrained by a dense measurement network in 2008, but also lies in the interior of the study domain and hence is immune to biases that may arise along the boundaries of the study domain [*Dirren et al.*, 2007]. The monthly fluxes and uncertainties are also aggregated to seven ecoregions (Figure 3) that are loosely defined based on the work of *Olson et al.* [2001] and demarcate large (mostly contiguous) regions with similar climate, land cover and land use.

[52] In terms of diagnostics at fine time scales, the GEnSRF performance is evaluated at 3-hourly and daily time scales, aggregated spatially to the full North American domain. By domain-averaging the recovered 3-hourly fluxes, we assess the ability of GEnSRF to accurately recover the



**Table 1.** Summary of the GEnSRF Configurations Reported in Section 4

Test Case Name	Inversion Setup	Number of Observations	Parameters		
			Ensemble Size	Localization Length	Prior Inflation Standard Deviation
GEnSRF TC1E500	TC1	3-hourly with data gaps <sup>a,b</sup>	500	n/a	n/a
GEnSRF TC1E2500	TC1	3-hourly with data gaps <sup>a,b</sup>	2500	n/a	n/a
GEnSRF TC1E500_L1500I005	TC1	3-hourly with data gaps <sup>a,b</sup>	500	1500 km	0.05
GEnSRF TC2E500	TC2	3-hourly <sup>a</sup>	500	n/a	n/a
GEnSRF TC2E500_L1500I005	TC2	3-hourly <sup>a</sup>	500	1500 km	0.05

<sup>a</sup>Three-hourly implies observations are available at 8 time periods during the day.

<sup>b</sup>For the shorter towers (height  $\leq 150$  m) only afternoon measurements are used; for the very short towers (height  $\leq 30$  m) only those measurements recommended by the data providers are used.

diurnal cycle of the CO<sub>2</sub> fluxes. Daily RMSD between the GEnSRF and the GIM grid-scale flux estimates are also examined as a function of time to evaluate the filter stability.

[53] The degree to which the GEnSRF fluxes reproduced the atmospheric CO<sub>2</sub> observations was also evaluated (results not shown), but a direct comparison of GEnSRF fluxes rather than atmospheric concentrations provides a more direct measure of the impact of the numerical DA scheme.

## 4. Results

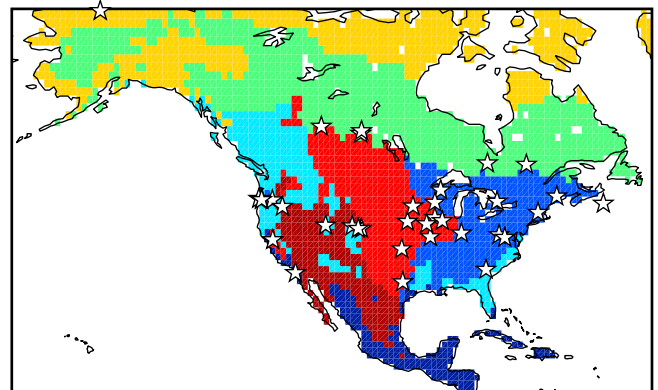
### 4.1. Multiscale Evaluation of the Ensemble Estimates for the Control Run (TC1E500)

[54] Monthly averaged grid-scale flux estimates and uncertainties for TC1 are presented in Figure 4. Qualitatively, it is clear that the control run (GEnSRF TC1E500, Figures 4c and 4d) is not capable of reproducing the monthly averaged GIM estimates or their associated uncertainties (Figures 4a and 4b). The under-estimated uncertainties should not be interpreted as more confident estimates, but rather point to the problem of insufficient variance in the ensemble. While the ensemble approach correctly captures the flux estimates over the Eastern corridor and the Southern parts of the continent, its performance degrades over the Northwestern region in Alaska and Canada, where scattered sources are incorrectly inferred throughout. In real-data application, this could be caused by two reasons: (1) the use of a limited number of ensemble members resulting in large sampling error, and (2) in general, this area has a sparse network with several of the available towers located in complex terrains where the transport is difficult to model. Given that this synthetic data study does not include transport model errors, the erroneous fluxes suggested by GEnSRF are a product of spurious ensemble noise. As indicated in Table 2, the difference in the spatial patterns between the two sets of estimates manifests itself in low CC and high RMSD between GIM and GEnSRF over North America.

[55] Monthly averaged ecoregion-scale flux estimates and associated uncertainties are presented in Figure 5. GEnSRF TC1E500 estimates suggest a smaller sink throughout all ecoregions relative to the GIM TC1 estimates, and the 95% uncertainty bounds based on the ensemble estimate only capture the true fluxes in 4 of the 7 ecoregions. At the continental scale, the GEnSRF TC1E500 estimate ( $-23.8 (\pm 3.4)$  gC/(m<sup>2</sup>month)) is significantly higher than the GIM TC1 estimate ( $-32.8 (\pm 2.7)$  gC/(m<sup>2</sup>month)), and

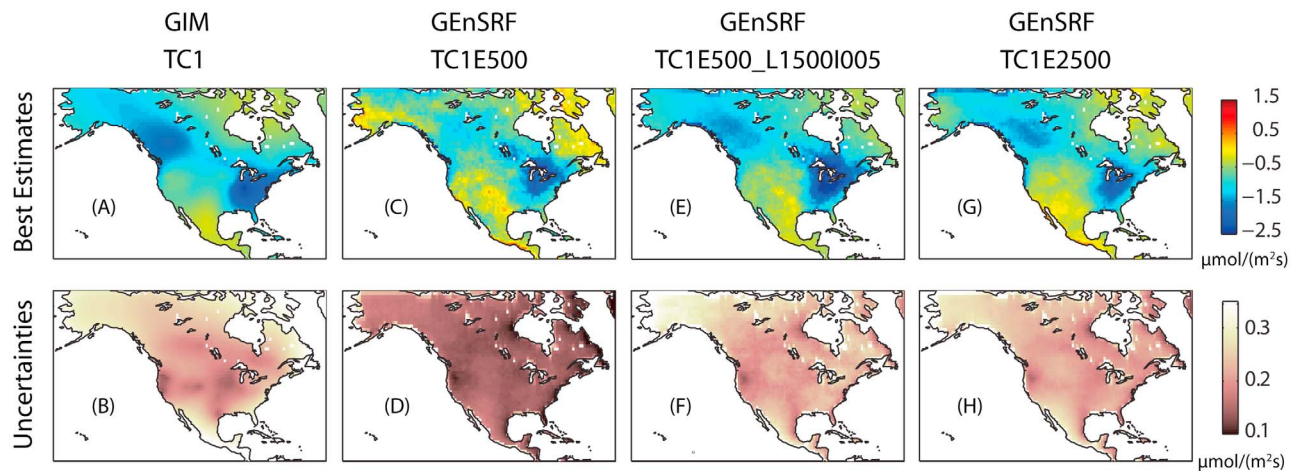
unlike the GIM estimate, does not capture the true flux of  $-30.46$  gC/(m<sup>2</sup>month).

[56] The inferred monthly averaged diurnal cycle at the continental scale is shown in Figure 6. GEnSRF TC1E500 does not reproduce the GIM TC1 diurnal pattern, with the difference between the two estimates spiking around 0400 h and 1900 h UTC. The largest differences between the estimated diurnal cycles coincide with times with the greatest temporal gradient in the true underlying fluxes, as well as times when observation locations are coming into/out-of the TC1 network. Another mechanism that could cause these observed errors is the sampling error due to a small ensemble size which could result in spurious temporal correlations in the estimates, leading to a dampened diurnal cycle relative to GIM. Further analysis (Sections 4.2 and 4.3) suggests that the gradient in the true diurnal cycle is the better explanatory factor. Conclusions based on the inferred diurnal cycle for the MCI region, which spans a much narrower longitudinal range and therefore exhibits less smearing of the diurnal



- Tropical & Subtropical (Trop)
- Temperate Broadleaf & Mixed Forests (TBMF)
- Temperate Coniferous Forests (TCof)
- Boreal Forests/Taiga (Bore)
- Tundra (Tund)
- Temperate Grasslands, Savannas & Shrublands (TGSS)
- Desert & Xeric Shrublands (DeXS)

**Figure 3.** Ecoregion map, modified from *Olson et al.* [2001], which is used for analyzing inversion results at spatially aggregated scales. Stars represent the location of the 35 tower network.



**Figure 4.** TC1 (top) flux estimates and (bottom) associated uncertainties aggregated to the monthly scale for (a and b) GIM and (c–h) three different GEnSRF runs.

cycle, are consistent with those for the full continent (Figure S1 in auxiliary material).

[57] Overall, the conclusion from the control run is that the small ensemble size and limited observational information in TC1E500 hinder the ensemble filter’s ability to reproduce GIM estimates across spatial and temporal scales. Sampling errors and sparse measurements may both result in a dramatic failure of the ensemble filter to infer fluxes.

#### 4.2. Sensitivity to Ensemble Size, and Covariance Localization and Inflation Algorithms

[58] A straightforward solution to reducing the sampling error is to increase the ensemble size, which in effect increases the rank of the ensemble estimate of the prior error covariance matrix. In the absence of a dynamical model and at the limit of a large ensemble, the ensemble filter asymptotically approaches the Kalman filter (assuming the error characteristics remain Gaussian) at a convergence rate of  $1/\sqrt{N}$ . A large ensemble (GEnSRF TC1E2500, Figures 4g and 4h) indeed appropriately reduces the spurious noise in the best estimates at fine spatial scales, and yields uncertainty estimates close to those from GIM TC1, albeit at the expense of an increase in computational cost compared to GEnSRF TC1E500 proportional to the increase in the size of the ensemble.

[59] An alternate approach that does not carry substantial additional computational cost is to implement covariance localization and inflation, which dampen the sampling error and

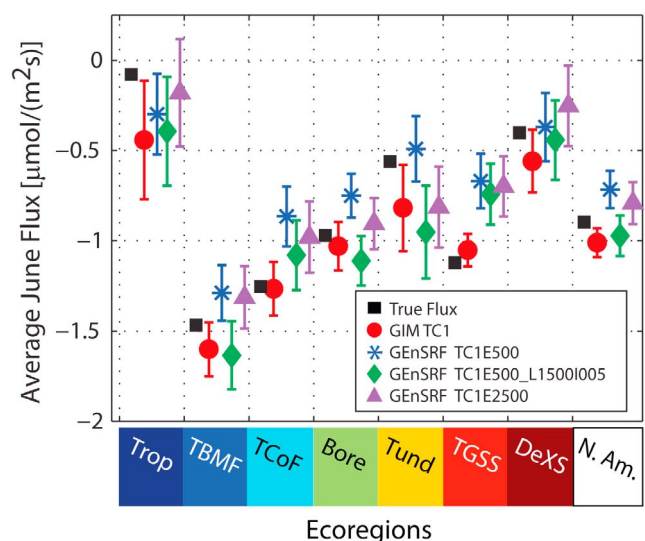
improve estimates of the flux uncertainties, as seen in GEnSRF TC1E500\_L1500I005 (Figures 4e and 4f). The improved performance resulting from increasing the ensemble size and implementing localization and inflation is confirmed in Table 2, where GEnSRF TC1E2500 and GEnSRF TC1E500\_L1500I005 both show higher CC and lower RMSD values relative to the control run. Both approaches also improve the ecoregion and continental scale results (Figure 5). The continental scale flux estimate for both GEnSRF TC1E2500 ( $-28.6 (\pm 3.9)$  gC/(m<sup>2</sup>month)) and GEnSRF TC1E500\_L1500I005 ( $-33.3 (\pm 3.7)$  gC/(m<sup>2</sup>month)) capture the true CASA fluxes within their 95% confidence intervals.

[60] The impact of increasing the ensemble size or of implementing inflation and localization is less conclusive for the estimation of the diurnal cycle either over the full continent (Figure 6) or over the MCI region (Figure S1 in auxiliary material). From Figure 6, GEnSRF TC1E2500

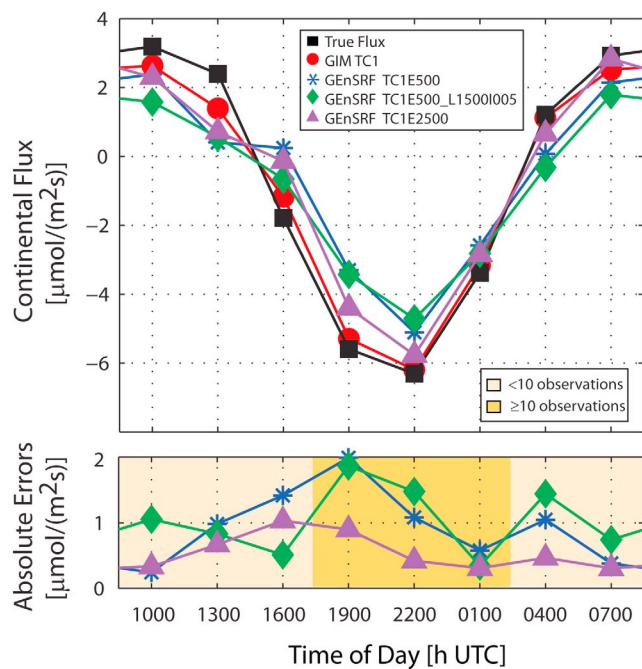
**Table 2.** Correlation Coefficients (CC) and Root Mean Square Difference (RMSD;  $\mu\text{mol}/(\text{m}^2 \text{s})$ ), Calculated Based on Grid-Scale, Monthly Averaged Flux Estimates Between the Various Runs of GEnSRF and GIM, for TC1 and TC2<sup>a</sup>

Test Case	North America (NA)		Mid-Continent Intensive (MCI)	
	CC	RMSD	CC	RMSD
GEnSRF TC1E500 <sup>a</sup>	0.64	0.52	0.77	0.32
GEnSRF TC1E2500	0.81	0.35	0.91	0.29
GEnSRF TC1E500_L1500I005	0.75	0.37	0.83	0.35
GEnSRF TC2E500	0.68	0.48	0.77	0.30
GEnSRF TC2E500_L1500I005	0.76	0.39	0.85	0.32

<sup>a</sup>Control runs for TC1 and TC2.



**Figure 5.** Estimated monthly averaged flux estimates and the associated uncertainties aggregated to ecoregions (Figure 3) and continental scales. The error bars represent 95% uncertainty bounds.



**Figure 6.** (top) Estimated flux diurnal cycle, and (bottom) absolute errors of the individual GEnSRF estimates with respect to the GIM estimates, aggregated to the continental scale. Also highlighted in Figure 6 (bottom) is the average observation density (light yellow denotes <10 observations, medium yellow denotes  $\geq 10$  observations) used in TC1 over the day.

captures the diurnal cycle very well initially but the error peaks around 1600 h UTC. The implementation of inflation and localization in GEnSRF TC1E500\_L1500I005 does not yield a clear reduction in errors especially at 1900 h UTC, although the discontinuity observed in GEnSRF TC1E500 between 1300 h UTC and 1600 h UTC is eliminated. The overall diurnal cycle, however, becomes even more washed out and fails to capture the true amplitude of the fluxes. The lack of error reduction resulting from the implementation of inflation and localization highlights the fact that although sampling error does contribute somewhat to the errors in the estimated diurnal cycle, the dominant cause is either the variable measurement network prescribed in TC1 or the inability of a small ensemble to capture sharp gradients in the flux diurnal cycle, or a combination of both. This is problematic, especially if we were to use these estimates either for mechanistic understanding of the carbon cycle at sub-diurnal scales, or for reconciling with estimates from biospheric models.

[61] Changing the localization length scale to either 500 km or 3000 km for the 500-member ensemble negatively impacts the estimates (Table 3). A tight isotropic localization scale (GEnSRF TC1E500\_L500) imposes high locality, as a consequence of which the autocorrelation information modeled in the prior covariance is completely lost. Measurements impact fluxes in their immediate vicinity, while areas in which no local observations are available are not constrained at all. It is possible though that strong localization could be imposed if a wide network of measurements were available to compensate for the loss of

remote influence. On the other hand, a large localization scale (GEnSRF TC1E500\_L3000) cannot significantly reduce the spurious correlations among distant flux locations. This suggests that the optimal value of the localization length scale (1500 km) may be linked with the correlation length scale of the fluxes themselves ( $\sim 1600$  km, see Section 3.2). However, tests also revealed that the optimal filter length scale is a function of the size of the ensemble, with a smaller ensemble size requiring a shorter optimal length scale. This is due to the fact that if the number of ensemble members is large, the noise in the covariance estimates does not overwhelm the signal until much farther from the observations. This makes it harder to identify a universal mathematical or physical basis for selecting these length scales. Nevertheless, the correlation length scale of the fluxes can be used as a starting point for the localization length scale in future filter designs.

[62] Likewise, the estimates are found to be sensitive to the parameters of the adaptive inflation technique, especially in terms of the recovered uncertainties over data-sparse regions and periods. As evident in Table 3, the change in CC and RMSD is small for the different starting parameters of the adaptive inflation, but the impact is more visible when the uncertainties associated with the GEnSRF are compared to the uncertainties from the GIM. For example, with very tightly constrained inflation factors (I001), GEnSRF underestimates the standard deviation of the individual flux estimates by an average of 13% relative to GIM. Conversely, with very loose prior inflation factors (I025) the initial inflation in the ensemble is large. During assimilation of subsequent observations, the ensemble should be deflated gradually. Yet for TC1, even after the full analysis, the ensemble remains over-inflated, resulting in an overestimate of the posterior standard deviations by GEnSRF by 31%. A prior inflation factor standard deviation of 0.05 provides a good balance, with uncertainties being underestimated by GEnSRF by only 4% (Figure 4f).

[63] In understanding the response of the adaptive inflation technique, two factors need to be considered: 1) the specification of a large and spatially uniform prior inflation factor uncertainty, i.e., one that does not vary between data sparse and data dense regions, and 2) a delayed response on the part of the adaptive inflation technique in adjusting to the changes in the measurement network as specific

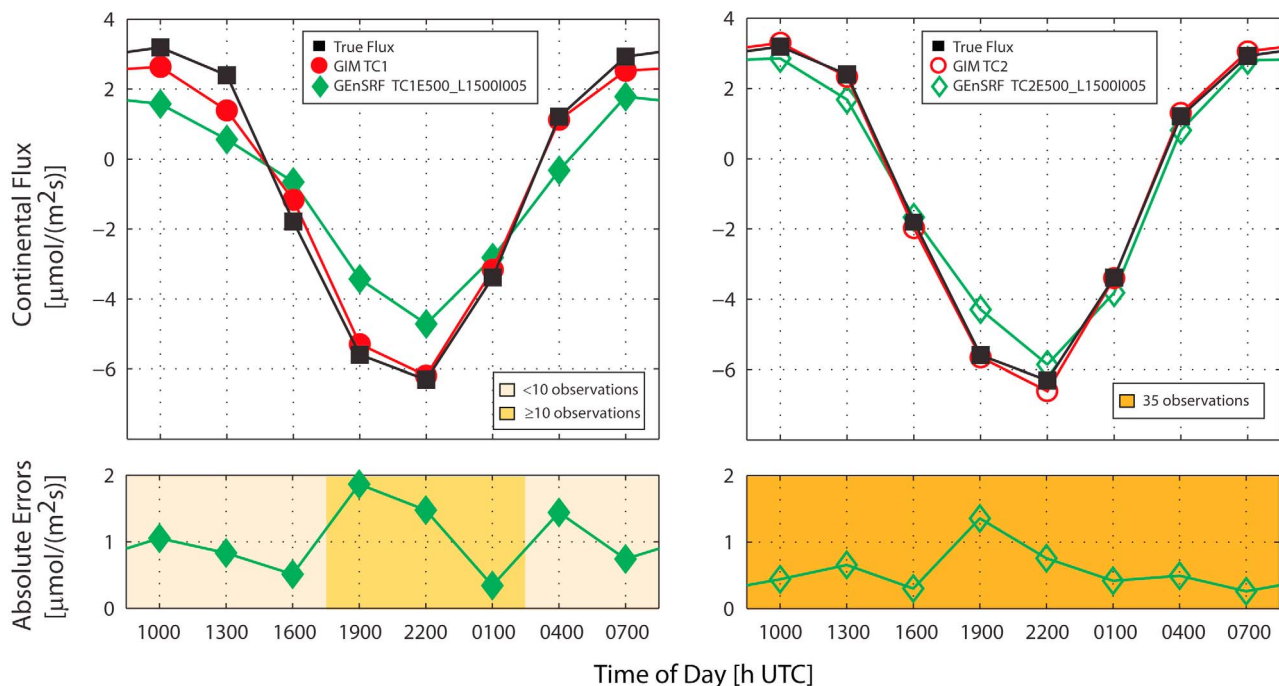
**Table 3.** Correlation Coefficients (CC) and Root Mean Square Difference (RMSD;  $\mu\text{mol}/(\text{m}^2 \text{ s})$ ), Calculated Based on Grid-Scale, Monthly Averaged Flux Estimates Between the Different Runs of GEnSRF and GIM for TC1

Test Case	North America (NA)		Mid-Continent Intensive (MCI)	
	CC	RMSD	CC	RMSD
GEnSRF TC1E500 <sup>a</sup>	0.64	0.52	0.77	0.32
GEnSRF TC1E500_L500 <sup>b</sup>	0.62	0.42	0.90	0.31
GEnSRF TC1E500_L1500 <sup>b</sup>	0.75	0.37	0.84	0.32
GEnSRF TC1E500_L3000 <sup>b</sup>	0.69	0.44	0.82	0.28
GEnSRF TC1E500_L1500I001 <sup>c</sup>	0.75	0.37	0.84	0.33
GEnSRF TC1E500_L1500I005 <sup>c</sup>	0.75	0.37	0.83	0.35
GEnSRF TC1E500_L1500I025 <sup>c</sup>	0.75	0.38	0.79	0.44

<sup>a</sup>Control run.

<sup>b</sup>Cases that specifically show the impact of localization on the final estimate.

<sup>c</sup>Cases that specifically show the impact of adaptive inflation on the final estimate.



**Figure 7.** (top) Estimated flux diurnal cycle, and (bottom) absolute errors of GEnSRF TC1E500\_L1500I005 and GEnSRF TC2E500\_L1500I005 with respect to the corresponding GIM estimates, aggregated to the continental scale. For TC2 the average observation density is 35 (dark yellow) but for TC1 it varies (light yellow denotes  $<10$  observations, medium yellow denotes  $\geq 10$  observations) over the day.

measurement location come into and out of the network throughout the day. Recall that the adaptive inflation technique is based on a Bayesian inverse modeling framework; hence, its dependency on the measurement network is not surprising. Significant improvement in the performance is obtained if the inflation is damped toward 1 as a function of time. Damping the inflation value over time makes the technique less dependent on the measurement coverage, and has been successfully implemented in other operational tests of the adaptive technique [e.g., Torn, 2010].

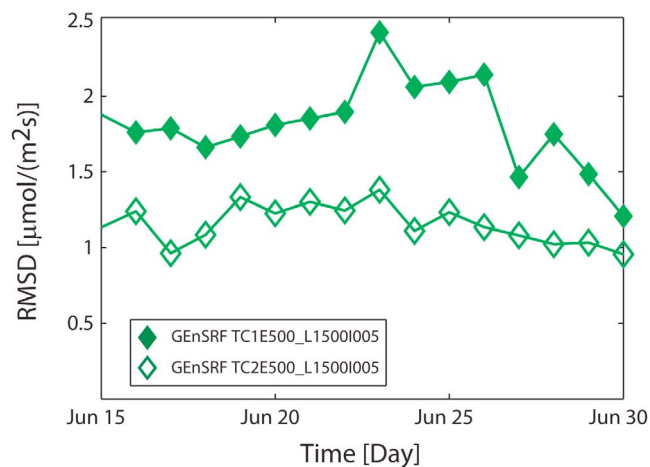
#### 4.3. Sensitivity to the Measurement Network

[64] For any inversion framework based on Bayesian estimation theory, the addition of measurements in space and time will improve both the estimation accuracy and the uncertainty. As expected by increasing the temporal density of measurements in TC2, the performance of both GIM and GEnSRF estimates at the grid and ecoregion scale improve significantly (results not shown). The continental scale flux estimates for both GIM TC2 ( $-31.7 (\pm 1.9) \text{ gC}/(\text{m}^2 \text{ month})$ ) and GEnSRF TC2E500\_L1500I005 ( $-32.1 (\pm 2.7) \text{ gC}/(\text{m}^2 \text{ month})$ ) improve substantially, allowing them to capture the true CASA flux estimate ( $-30.5 \text{ gC}/(\text{m}^2 \text{ month})$ ) within their 95% confidence intervals.

[65] Of greater interest is that the GEnSRF estimates now capture the amplitude and phase of the diurnal cycle better than in the case of TC1 (Figure 7 over the full continent and Figure S2 in auxiliary material over the MCI region), even without needing to increase the ensemble size. From Figure 7, GEnSRF TC2E500\_L1500I005 estimates mirror

the GIM TC2 3-hourly estimates, indicating the positive impact that the additional measurements have had on the ensemble filter, especially between 0100 h to 1600 h UTC. Comparing the error in Figure 7 (bottom), one can see that the denser homogeneous network in TC2 plays a significant role in aiding the ensemble filter to correctly capture the diurnal cycle. However, the higher errors at 1900 h UTC still persist, showing that these errors are more likely to be attributable to the sharp gradient in the true diurnal cycle at this time, rather than due to temporal heterogeneity in the measurement network. Either hypothesis could have been supported by results from TC1, because transition times in network size coincide with times with sharp gradients in the diurnal cycle. Allowing the GEnSRF to directly estimate sub-continental spatial and sub-daily temporal patterns therefore also made it possible to identify the filter sensitivity to the measurement network prescribed in TC1 and TC2.

[66] Results from TC2 also confirm that ensemble filter performance improves with a denser measurement network. This follows from the hypothesis stated earlier in Section 3.2, that for an under-determined inversion problem, the ensemble system is sensitive to the spatiotemporal density of the measurements. Additional runs with a temporally homogeneous 10 tower network confirmed that the total number of observations is a better determinant of ensemble performance at fine temporal scales relative to their temporal heterogeneity/homogeneity. Without the guidance of a dynamical model and in the absence of a rich observational constraint, the ensemble deviates from the truth, resulting in increased ensemble degeneracy and inaccurate estimates. In fact, as



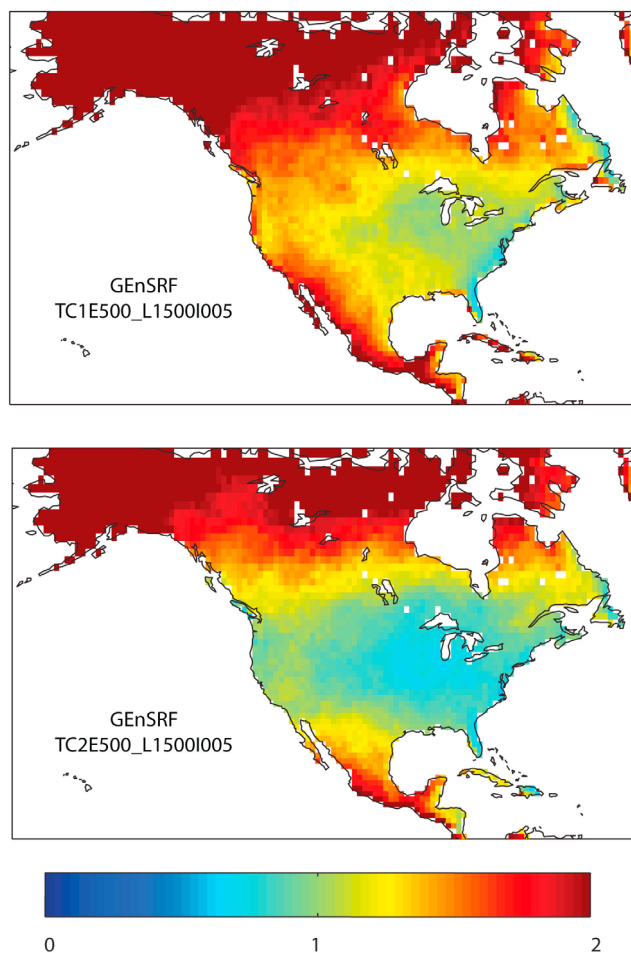
**Figure 8.** Time series of the Root Mean Square Difference (RMSD) between grid-scale daily averaged estimates from GEnSRF and GIM over North America for TC1 and TC2. The time series shown here is for the latter half of the assimilation cycle to emphasize that with the measurement network in TC1, the ensemble filter does not stabilize and suffers from divergence.

shown in Figure 8, it is only in the case of TC2 that the filter is stable and reaches an asymptotic level of accuracy.

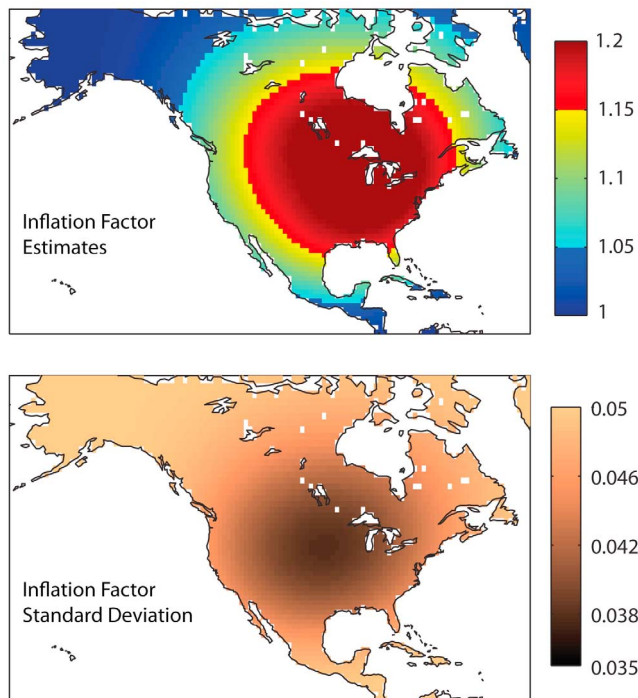
[67] The influence of the measurement density on the ensemble behavior can be examined using several diagnostics that are commonly available in the NWP literature. All of these diagnostics, however, require knowledge of the true state against which the ensemble mean is evaluated. In this study, the true state is available from the CASA-GFED v2 fluxes, but in real application this would be unknown and hence these diagnostics could not be calculated. The diagnostic selected here examines the ratio of the time-averaged ensemble spread to the error in the ensemble mean [Liu *et al.*, 2008] at every estimation grid point, and highlights how measurement availability controls ensemble behavior. In this case, the ensemble spread is obtained as the difference between individual ensemble members and the ensemble mean, while the error in the ensemble mean is calculated as the mean squared difference from the true state. This ratio is an indication of the optimality of the DA system, and illustrates the impact of the measurements in adjusting this ratio. Figure 9 shows this ratio for GEnSRF TC1E500\_L1500I005 and GEnSRF TC2E500\_L1500I005. In the case of GEnSRF TC2E500\_L1500I005, the ratio is close to 1.0 over most of the continent, such that on an average the analysis spread among the ensemble members is consistent with the true errors, i.e., mean squared difference between the ensemble mean and the truth. In the case of GEnSRF TC1E500\_L1500I005, however, the ratio is close to 1.0 only over a small portion of the continent that is both removed from domain boundaries and that is relatively well constrained by observations, while in other areas the value of the ratio is near 2.0. This indicates that the ensemble overestimates the uncertainties in these regions by a factor of 2.

[68] This result is indicative of the better performance of the adaptive inflation technique with a richer observational constraint, as seen by the fact that TC2E500\_L1500I005 has a reduced mismatch between the ensemble mean error and

the ensemble spread, except over very sparsely observed areas like the Tundra. The dependency of the adaptive inflation technique on the spatial heterogeneity of the measurement network might seem a disadvantage at first. However, we argue that the adaptive algorithm provides inflation values that are preferable than having to manually tune the system with a single inflation value that would be applied everywhere. Figure 10 shows the spatially dependent monthly averaged inflation factors and their uncertainties, as determined by the adaptive inflation algorithm for GEnSRF TC2E500\_L1500I005. Although the time-averaged values in Figure 10 mask the significant temporal variations of the inflation factors, they do highlight the spatial structure that is clearly consistent with the spatial density of the measurement network. This spatial variability in the inflation factors underscores the need for adopting an inflation strategy that can be adjusted recursively. Conversely, if a single inflation value had been used over the entire continent, for example a value of 1.1, then this would have under-inflated the ensemble in the data-dense regions, but over-inflated it over the data-sparse regions. This would have led to additional errors in the final estimated flux estimates and their uncertainties. Overall, results indicate that the density of the



**Figure 9.** The ratio of grid-scale time-averaged ensemble spread and ensemble mean error for TC1E500\_L1500I005 and TC2E500\_L1500I005. A ratio of 1.0 (or green color) indicates optimal data assimilation.



**Figure 10.** Monthly averaged a posteriori inflation factor estimates and associated standard deviations for the case GEnSRF TC2E500\_L1500I005. Note that the largest change in the inflation factors and the largest reduction in the prior inflation standard deviations are over areas with more measurements.

measurement network not only controls the estimation accuracy, but also ensures that the entire ensemble system and its associated algorithms function well. In the absence of a dynamical model, the measurements play an even more integral role in the assimilation process, as they drive both the ensemble mean and the ensemble spread. In practice, obtaining such a network not only requires additions to the existing monitoring network, but also improvements to atmospheric transport model that would enable the use of observations collected throughout the day, as was done in TC2.

[69] An important question that has not been addressed here is the response of the ensemble to either transport error or measurements of varying quality. Given the sensitivity of the ensemble to the density of measurements, we expect differential and/or correlated model-data mismatch errors to play a key role as well. Tests with higher model-data mismatch covariance errors demonstrated that no additional degradation in the performance of GEnSRF was observed relative to that for GIM (results not shown). These tests did not include correlated observational errors, however.

## 5. Summary and Conclusions

[70] Application of data assimilation techniques for estimating sources and sinks of CO<sub>2</sub> provides unique opportunities to better understand the mechanistic processes governing the carbon cycle. In this work, we examined the parameter space of the ensemble filter in terms of estimating CO<sub>2</sub> fluxes at high spatial and temporal resolutions. A new ensemble square root filter (GEnSRF) based on the geostatistical

inverse modeling technique was presented and applied to a synthetic data study over North America.

[71] The application of GEnSRF to different inversion regimes illustrates a dynamic interplay between three factors: (1) the spatial and temporal density of the measurements available to inform the filter, (2) the ensemble size, and the resultant sampling error, and (3) the implementation of covariance inflation and localization algorithms to ameliorate the latter. Together, these factors determine not only the relative precision and accuracy of the best estimates but also their associated uncertainties. For the ensemble filter to serve as an appropriate replacement for batch estimation of fine-scale fluxes, experiments in this study demonstrated that it may be necessary to have a dense network of measurements in space and time. To some extent, this bodes well for future applications with high-density remote sensing measurements of CO<sub>2</sub>. Additional studies will be necessary, however, to quantify the impact of biases, correlated errors, temporal heterogeneity etc. in the remote sensing measurements.

[72] It can be argued that the requirement for more measurements may be relaxed if a dynamical model is developed to propagate the CO<sub>2</sub> fluxes in time. Alternately, if the inversion can be formulated as an over-determined problem it will be better constrained by the measurements, which would also lead to better ensemble behavior. This may be problematic, however, since by solving at large spatial and temporal scales existing deficiencies in the ensemble filter are masked, and aggregation errors grow. In the long run, solving at large spatial and temporal scales may limit methodological advancements in the design of future filters for the CO<sub>2</sub> source-sink estimation problem.

[73] As the popularity of the ensemble filter within the carbon science community rises, future developments will most likely revolve around reducing the impact of sampling error. In this study, this was the largest source of error resulting from the use of a limited ensemble size. Sensitivity tests with different ensemble sizes established that approximately 500 ensemble members, used in combination with covariance inflation and localization, may be used for estimating 3-hourly fluxes over North America at a  $1^\circ \times 1^\circ$  scale. Estimates at both native and aggregated spatial scales were reliable, as were estimates at aggregated temporal scales. Capturing the diurnal cycle of the underlying fluxes proved most difficult, even when covariance inflation and localization were used. By designing inflation and localization techniques that are more tailored or customized to the CO<sub>2</sub> flux estimation problem, the requisite number of ensemble members may be reduced further to increase the computational efficiency. The two algorithms implemented here are drawn from NWP-related problems. Although they perform reasonably well, questions remain over the behavior of the adaptive inflation technique in data-sparse regions, the appropriateness of existing localization techniques, etc. In spite of these shortcomings, these algorithms can be implemented with a limited ensemble size to obtain reliable posterior CO<sub>2</sub> flux estimates but with over-inflated uncertainties over data-sparse regions, as was done here.

[74] **Acknowledgments.** This work was supported by the National Aeronautics and Space Administration (NASA) through an Earth System Science Fellowship for Abhishek Chatterjee, under grant NNX09AO10H. Additional support was provided through NASA grant NNX12AB90G and a contract from Sandia National Laboratories, Albuquerque, NM,

funded under a Laboratory Directed Research and Development project. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The authors acknowledge Ave Arellano, Andy Jacobson and Derek Posselt for fruitful comments and discussions regarding this work. We specially thank Sharon Gourdji for her suggestions and help on the GIM analyses. Finally, we would like to thank the following scientists for sharing their atmospheric measurements that were used in simulating the data-gaps in the observation scenario used in TC1: Arlyn Andrews, Ken Davis, Danilo Dragoni, Marc Fischer, Mathias Goeckede, Ralph Keeling, Bev Law, Natasha Miles, Bill Munger, Matt Parker, Scott Richardson, Britt Stephens, Colm Sweeney, Steven Wofsy and Douglas Worthy.

## References

- Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*(12), 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Anderson, J. L. (2007a), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus, Ser. A*, *59*(2), 210–224, doi:10.1111/j.1600-0870.2006.00216.x
- Anderson, J. L. (2007b), Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D*, *230*, 99–111, doi:10.1016/j.physd.2006.02.011.
- Anderson, J. L. (2009), Spatially and temporally varying adaptive covariance inflation for ensemble filters, *Tellus, Ser. A*, *61*(1), 72–83, doi:10.1111/j.1600-0870.2008.00361.x
- Anderson, J. L. (2012), Localization and sampling error correction in ensemble Kalman filter data assimilation, *Mon. Weather Rev.*, *140*(7), 2359–2371, doi:10.1175/MWR-D-11-00013.1.
- Anderson, J. L., and S. L. Anderson (1999), A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, *127*(12), 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
- Baker, D. F., S. C. Doney, and D. S. Schimel (2006), Variational data assimilation for atmospheric CO<sub>2</sub>, *Tellus, Ser. B*, *58*(5), 359–365, doi:10.1111/j.1600-0889.2006.00218.x.
- Baker, D. F., H. Boesch, S. C. Doney, D. O'Brien, and D. S. Schimel (2010), Carbon source/sink information provided by column CO<sub>2</sub> measurements from the Orbiting Carbon Observatory, *Atmos. Chem. Phys.*, *10*(9), 4145–4165, doi:10.5194/acp-10-4145-2010.
- Bergemann, K., and S. Reich (2010), A localization technique for ensemble Kalman filters, *Q. J. R. Meteorol. Soc.*, *136*(648), 701–707, doi:10.1002/qj.591.
- Bishop, C. H., and D. Hodyss (2011), Adaptive ensemble covariance localization in Ensemble 4D-VAR state estimation, *Mon. Weather Rev.*, *139*(4), 1241–1255, doi:10.1175/2010MWR3403.1.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar (2001), Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, *Mon. Weather Rev.*, *129*(3), 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.
- Bruhwyler, L. M. P., et al. (2005), An improved Kalman smoother for atmospheric inversions, *Atmos. Chem. Phys.*, *5*, 2691–2702, doi:10.5194/acp-5-2691-2005.
- Buchwitz, M., et al. (2005), Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: Initial comparison with chemistry and transport models, *Atmos. Chem. Phys.*, *5*, 941–962, doi:10.5194/acp-5-941-2005.
- Buchwitz, M., O. Schneising, J. P. Burrows, H. Bovensmann, M. Reuter, and J. Notholt (2007), First direct observation of the atmospheric CO<sub>2</sub> year-to-year increase from space, *Atmos. Chem. Phys.*, *7*(16), 4249–4256, doi:10.5194/acp-7-4249-2007.
- Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*(6), 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.
- Chahine, M. T., et al. (2006), Improving weather forecasting and providing new data on greenhouse gases, *Bull. Am. Meteorol. Soc.*, *87*(7), 911–926, doi:10.1175/BAMS-87-7-911.
- Chevallier, F., M. Fisher, P. Peylin, S. Serrar, P. Bousquet, F.-M. Bréon, A. Chédin, and P. Ciais (2005), 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.
- Chevallier, F., F.-M. Bréon, and P. J. Rayner (2007), 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.
- Chevallier, F., et al. (2010), CO<sub>2</sub> surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, *115*, D21307, doi:10.1029/2010JD013887.
- Chevallier, F., et al. (2012), What eddy-covariance measurements tell us about prior land flux errors in CO<sub>2</sub>-flux inversion schemes, *Global Biogeochem. Cycles*, *26*, GB1021, doi:10.1029/2010GB003974.
- Chilès, J.-P., and P. Delfiner (Eds.) (1999), *Geostatistics: Modeling Spatial Uncertainty*, 1st ed., 720 pp., Wiley-Intersci., Hoboken, N. J., doi:10.1002/9780470316993.
- Ciais, P., P. Rayner, F. Chevallier, P. Bousquet, M. Logan, P. Peylin, and M. Ramonet (2010), Atmospheric inversions for estimating CO<sub>2</sub> fluxes: Methods and perspectives, *Clim. Change*, *103*(1–2), 69–92, doi:10.1007/s10584-010-9909-3.
- Crevoisier, C., A. Chedin, H. Matsueda, T. Machida, R. Armante, and N. A. Scott (2009), First year of upper tropospheric integrated content of CO<sub>2</sub> from IASI hyperspectral infrared observations, *Atmos. Chem. Phys.*, *9*(14), 4797–4810, doi:10.5194/acp-9-4797-2009.
- Crisp, D., C. E. Miller, and P. L. DeCola (2008), NASA Orbiting Carbon Observatory: Measuring the column averaged carbon dioxide mole fraction from space, *J. Appl. Remote Sens.*, *2*, 023508, doi:10.1117/1.2898457.
- Dirren, S., R. D. Tom, and G. J. Hakim (2007), A data assimilation case study using a limited-area ensemble Kalman filter, *Mon. Weather Rev.*, *135*, 1455–1473, doi:10.1175/MWR3358.1.
- Ehrendorfer, M. (2007), A review of issues in ensemble-based Kalman filtering, *Meteorol. Z.*, *16*(6), 795–818, doi:10.1127/0941-2948/2007/0256.
- Engelen, R. J., A. S. Denning, and K. R. Gurney (2002), On error estimation in atmospheric CO<sub>2</sub> inversions, *J. Geophys. Res.*, *107*(D22), 4635, doi:10.1029/2002JD002195.
- Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport. Atmospheric and Space Science Series*, 392 pp., Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9780511535741.
- Feng, L., P. I. Palmer, H. Boesch, and S. Dance (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, *Atmos. Chem. Phys.*, *9*(8), 2619–2633, doi:10.5194/acp-9-2619-2009.
- Feng, L., P. I. Palmer, Y. Yang, R. M. Yantosca, S. R. Kawa, J.-D. Paris, H. Matsueda, and T. Machida (2011), Evaluating a 3-D transport model of atmospheric CO<sub>2</sub> using ground-based, aircraft and space-borne data, *Atmos. Chem. Phys.*, *11*, 2789–2803, doi:10.5194/acp-11-2789-2011.
- Furrer, R., and T. Bengtsson (2007), Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *J. Multivariate Anal.*, *98*(2), 227–255, doi:10.1016/j.jmva.2006.08.003.
- Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, *125*(554), 723–757, doi:10.1002/qj.49712555417.
- Geels, C., et al. (2007), Comparing atmospheric transport models for future regional inversions over Europe - Part 1: Mapping the atmospheric CO<sub>2</sub> signals, *Atmos. Chem. Phys.*, *7*(13), 3461–3479, doi:10.5194/acp-7-3461-2007.
- Gerbig, C., S. Korner, and J. C. Lin (2008), Vertical mixing in atmospheric tracer transport models: Error characterization and propagation, *Atmos. Chem. Phys.*, *8*, 591–602, doi:10.5194/acp-8-591-2008.
- Gneiting, T. (2002), Compactly supported correlation functions, *J. Multivariate Anal.*, *83*(2), 493–508, doi:10.1006/jmva.2001.2056.
- Göckede, M., A. M. Michalak, D. Vickers, D. P. Turner, and B. E. Law (2010), Atmospheric inverse modeling to constrain regional scale CO<sub>2</sub> budgets at high spatial and temporal resolution, *J. Geophys. Res.*, *115*, D15113, doi:10.1029/2009JD012257.
- Gourdji, S., K. Mueller, K. Schaefer, and A. M. Michalak (2008), Global monthly averaged CO<sub>2</sub> fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, *J. Geophys. Res.*, *113*, D21115, doi:10.1029/2007JD009733.
- Gourdji, S. M., A. I. Hirsch, K. L. Mueller, V. Yadav, A. E. Andrews, and A. M. Michalak (2010), Regional-scale geostatistical inverse modeling of North American CO<sub>2</sub> fluxes: A synthetic data study, *Atmos. Chem. Phys.*, *10*(13), 6151–6167, doi:10.5194/acp-10-6151-2010.
- Gourdji, S. M., et al. (2012), North American CO<sub>2</sub> Exchange: Inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, *Biogeosciences*, *9*, 457–475, doi:10.5194/bg-9-457-2012.
- Hamazaki, T., A. Kuze, and K. Kondo (2004), Sensor system for greenhouse gas observing satellite (GOSAT), *Proc. SPIE Int. Soc. Opt. Eng.*, *5543*(1), 275–282, doi:10.1117/12.560589.
- Hamill, T. M., and J. S. Whitaker (2005), Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches, *Mon. Weather Rev.*, *133*(11), 3132–3147, doi:10.1175/MWR3020.1.
- Hamill, T. M., and J. S. Whitaker (2011), What constrains spread growth in forecasts initialized from ensemble Kalman filters?, *Mon. Weather Rev.*, *139*(1), 117–131, doi:10.1175/2010MWR3246.1.
- Hamill, T. M., J. S. Whitaker, and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, *129*(11), 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

- Horn, R. A., and R. Mathias (1990), An analog of the Cauchy-Schwarz inequality for Hadamard-products and unitarily invariant norms, *SIAM J. Matrix Anal. Appl.*, *11*(4), 481–498, doi:10.1137/0611034.
- Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*(3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Houtekamer, P. L., and H. L. Mitchell (2001), A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, *129*(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- Houtekamer, P. L., and H. L. Mitchell (2005), Ensemble Kalman filtering, *Q. J. R. Meteorol. Soc.*, *131*(613), 3269–3289, doi:10.1256/qj.05.135.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and M. Hansen (2005), Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Mon. Weather Rev.*, *133*(3), 604–620, doi:10.1175/MWR-2864.1.
- Houtekamer, P. L., H. L. Mitchell, and X. Deng (2009), Model error representation in an operational ensemble Kalman filter, *Mon. Weather Rev.*, *137*(7), 2126–2143, doi:10.1175/2008MWR2737.1.
- Huntzinger, D. N., S. M. Gourdji, K. L. Mueller, and A. M. Michalak (2011), The utility of continuous atmospheric measurements for identifying biospheric CO<sub>2</sub> flux variability, *J. Geophys. Res.*, *116*, D06110, doi:10.1029/2010JD015048.
- Jun, M., I. Szunyogh, M. G. Genton, F. Zhang, and C. H. Bishop (2011), A statistical investigation of the sensitivity of ensemble-based Kalman filters to covariance filtering, *Mon. Weather Rev.*, *139*(9), 3036–3051, doi:10.1175/2011MWR3577.1.
- Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*(D5), 4703–4715, doi:10.1029/2000JD900581.
- Kang, J. S., E. Kalnay, J. J. Liu, I. Fung, T. Miyoshi, and K. Ide (2011), “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation, *J. Geophys. Res.*, *116*, D09110, doi:10.1029/2010JD014673.
- Karspeck, A. R., and J. L. Anderson (2007), Experimental implementation of an ensemble adjustment filter for an intermediate ENSO model, *J. Clim.*, *20*(18), 4638–4658, doi:10.1175/JCLI4245.1.
- Keper, J. D. (2009), Covariance localisation and balance in an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, *135*(642), 1157–1176, doi:10.1002/qj.443.
- Keper, J. D. (2011), Balance-aware covariance localisation for atmospheric and oceanic ensemble Kalman filters, *Comput. Geosci.*, *15*(2), 239–250, doi:10.1007/s10596-010-9188-0.
- Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inverting, *Water Resour. Res.*, *31*(10), 2411–2419, doi:10.1029/95WR01945.
- Kulawik, S. S., et al. (2010), Characterization of Tropospheric Emission Spectrometer (TES) CO<sub>2</sub> for carbon cycle science, *Atmos. Chem. Phys.*, *10*(12), 5601–5623, doi:10.5194/acp-10-5601-2010.
- Lahoz, W., and Q. Errera (2010), Constituent assimilation, in *Data Assimilation, Making Sense of Observations*, edited by W. Lahoz, B. Khattatov, and R. Menard, pp. 449–490, Springer, Berlin.
- Lauvaux, T., et al. (2012), Constraining the CO<sub>2</sub> budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmos. Chem. Phys.*, *12*, 337–354, doi:10.5194/acp-12-337-2012.
- Lawson, W. G., and J. A. Hansen (2004), Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth, *Mon. Weather Rev.*, *132*(8), 1966–1981, doi:10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2.
- Li, H., E. Kalnay, and T. Miyoshi (2009), Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, *135*(639), 523–533, doi:10.1002/qj.371.
- Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger (2003), A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J. Geophys. Res.*, *108*(D16), 4493, doi:10.1029/2002JD003161.
- Liu, J., E. J. Fertig, H. Li, E. Kalnay, B. R. Hunt, E. J. Kostelich, I. Szunyogh, and R. Todling (2008), Comparison between local ensemble transform Kalman filter and PSAS in the NASA finite volume GCM—Perfect model experiments, *Nonlinear Process. Geophys.*, *15*(4), 645–659, doi:10.5194/npg-15-645-2008.
- Lokupitiya, R. S., D. Zupanski, A. S. Denning, S. R. Kawa, K. R. Gurney, and M. Zupanski (2008), Estimation of global CO<sub>2</sub> fluxes at regional scale using the maximum likelihood ensemble filter, *J. Geophys. Res.*, *113*, D20110, doi:10.1029/2007JD009679.
- Meng, Z., and F. Zhang (2011), Limited-area ensemble-based data assimilation, *Mon. Weather Rev.*, *139*(7), 2025–2045, doi:10.1175/2011MWR3418.1.
- Michalak, A. M. (2008), Technical Note: Adapting a fixed-lag Kalman smoother to a geostatistical atmospheric inversion framework, *Atmos. Chem. Phys.*, *8*(22), 6789–6799, doi:10.5194/acp-8-6789-2008.
- Michalak, A. M., L. Bruhwiler, and P. P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *J. Geophys. Res.*, *109*, D14109, doi:10.1029/2003JD004422.
- Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *J. Geophys. Res.*, *110*, D24107, doi:10.1029/2005JD005970.
- Miyazaki, K., T. Maki, P. Patra, and T. Nakazawa (2011), Assessing the impact of satellite, aircraft, and surface observations on CO<sub>2</sub> flux estimation using an ensemble-based 4-D data assimilation system, *J. Geophys. Res.*, *116*, D16306, doi:10.1029/2010JD015366.
- Miyoshi, T. (2011), The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter, *Mon. Weather Rev.*, *139*(5), 1519–1535, doi:10.1175/2010MWR3570.1.
- National Research Council (2007), *Earth Science and Applications From Space: National Imperatives for the Next Decade and Beyond*, Natl. Acad. Press, Washington, D. C.
- Nerger, L., W. Hiller, and J. Schröter (2005), A comparison of error sub-space Kalman filters, *Tellus, Ser. A*, *57*(5), 715–735, doi:10.1111/j.1600-0870.2005.00141.x.
- O’Kane, T. J., and J. S. Frederiksen (2008), Comparison of statistical dynamical, square root and ensemble Kalman filters, *Entropy*, *10*(4), 684–721, doi:10.3390/e10040684.
- Oke, P. R., P. Sakov, and S. P. Corney (2007), Impacts of localisation in the EnKF and EnOI: Experiments with a small model, *Ocean Dyn.*, *57*(1), 32–45, doi:10.1007/s10236-006-0088-8.
- Olsen, S. C., and J. T. Randerson (2004), Differences between surface and column atmospheric CO<sub>2</sub> and implications for carbon cycle research, *J. Geophys. Res.*, *109*, D02301, doi:10.1029/2003JD003968.
- Olson, D. M., et al. (2001), Terrestrial ecoregions of the world: A new map of life on Earth, *BioScience*, *51*(11), 933–938, doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.
- Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke (2004), A local ensemble Kalman filter for atmospheric data assimilation, *Tellus, Ser. A*, *56*(5), 415–428, doi:10.1111/j.1600-0870.2004.00076.x.
- Peña, M., Z. Toth, and M. Z. Wei (2010), Controlling noise in ensemble data assimilation schemes, *Mon. Weather Rev.*, *138*(5), 1502–1512, doi:10.1175/2009MWR2854.1.
- Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P. P. Tans (2005), 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.
- Peters, W., et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 18,925–18,930, doi:10.1073/pnas.0708986104.
- Peters, W., et al. (2010), Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Global Change Biol.*, *16*(4), 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x.
- Pham, D. T. (2001), Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Mon. Weather Rev.*, *129*(5), 1194–1207, doi:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.
- Randerson, J. T., M. V. Thompson, T. J. Conway, I. Y. Fung, and C. B. Field (1997), The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cycles*, *11*(4), 535–560, doi:10.1029/97GB02268.
- Rayner, P. J. (2010), The current state of carbon-cycle data assimilation, *Curr. Opin. Environ. Sustain.*, *2*(4), 289–296, doi:10.1016/j.cosust.2010.05.005.
- Rayner, P. J., and D. M. O’Brien (2001), The utility of remotely sensed CO<sub>2</sub> concentration data in surface source inversions, *Geophys. Res. Lett.*, *28*(1), 175–178, doi:10.1029/2000GL011912.
- Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cycles*, *19*, GB2026, doi:10.1029/2004GB002254.
- Rödenbeck, C. (2005), Estimating CO<sub>2</sub> sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, *Tech. Rep. 6*, Max Planck Inst. for Biogeochem., Jena, Germany.
- Rood, R. B. (2005), Assimilation of stratospheric meteorological and constituent observations: A review, *SPARC Newsl.*, *25*, 31–37.
- Sacher, W., and P. Bartello (2008), Sampling errors in ensemble Kalman filtering. Part I: Theory, *Mon. Weather Rev.*, *136*(8), 3035–3049, doi:10.1175/2007MWR2323.1.
- Schmiedl, D. (2007), Carbon cycle conundrums, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(47), 18,353–18,354, doi:10.1073/pnas.0709331104.
- Scholes, R. J., P. M. S. Monteiro, C. L. Sabine, and J. G. Canadell (2009), Systematic long-term observations of the global carbon cycle, *Trends Ecol. Evol.*, *24*(8), 427–430, doi:10.1016/j.tree.2009.03.006.



- Schutgens, N. A. J., T. Miyoshi, T. Takemura, and T. Nakajima (2010), Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, *Atmos. Chem. Phys.*, *10*, 6583–6600, doi:10.5194/acp-10-6583-2010.
- Skamarock, W. C., et al. (2005) A description of the advanced research WRF version 2, *NCAR Tech. Note 468+STR*, NCAR, Boulder, Colo. [Available at [http://www.mmm.ucar.edu/wrf/users/docs/arw\\_v2.pdf](http://www.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf).]
- Swinbank, R. (2010), Numerical weather prediction, in *Data Assimilation, Making Sense of Observations*, edited by W. Lahoz, B. Khatatov, and R. Menard, pp. 381–406, Springer, Berlin.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke (2005), Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model, *Tellus, Ser. A*, *57*(4), 528–545, doi:10.1111/j.1600-0870.2005.00136.x
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker (2003), Ensemble square root filters, *Mon. Weather Rev.*, *131*(7), 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.
- Torn, R. D. (2010), Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test, *Mon. Weather Rev.*, *138*, 4375–4392, doi:10.1175/2010MWR3361.1.
- Uzunoglu, B., S. J. Fletcher, M. Zupanski, and I. M. Navon (2007), Adaptive ensemble reduction and inflation, *Q. J. R. Meteorol. Soc.*, *133*(626), 1281–1294, doi:10.1002/qj.96.
- van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and A. F. Arellano (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, *6*, 3423–3441, doi:10.5194/acp-6-3423-2006.
- Whitaker, J. S., and T. M. Hamill (2002), Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, *130*(7), 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- Zhang, F., C. Snyder, and J. Sun (2004), Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter, *Mon. Weather Rev.*, *132*(5), 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.
- Zupanski, D., A. S. Denning, M. Uliasz, M. Zupanski, A. E. Schuh, P. J. Rayner, W. Peters, and K. D. Corbin (2007), Carbon flux bias estimation employing maximum likelihood ensemble filter (MLEF), *J. Geophys. Res.*, *112*, D17107, doi:10.1029/2006JD008371.