



# Technical Note: Comparison of ensemble Kalman filter and variational approaches for CO<sub>2</sub> data assimilation

A. Chatterjee<sup>1,2,\*</sup> and A. M. Michalak<sup>2</sup>

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

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

\* now at: Data Assimilation Research Section, The National Center for Atmospheric Research, Boulder, CO, USA

Correspondence to: A. Chatterjee (abhishek@ucar.edu)

Received: 16 April 2013 – Published in Atmos. Chem. Phys. Discuss.: 14 May 2013

Revised: 13 October 2013 – Accepted: 20 October 2013 – Published: 3 December 2013

**Abstract.** Data assimilation (DA) approaches, including variational and the ensemble Kalman filter methods, provide a computationally efficient framework for solving the CO<sub>2</sub> source–sink estimation problem. Unlike DA applications for weather prediction and constituent assimilation, however, the advantages and disadvantages of DA approaches for CO<sub>2</sub> flux estimation have not been extensively explored. In this study, we compare and assess estimates from two advanced DA approaches (an ensemble square root filter and a variational technique) using a batch inverse modeling setup as a benchmark, within the context of a simple one-dimensional advection–diffusion prototypical inverse problem that has been designed to capture the nuances of a real CO<sub>2</sub> flux estimation problem. Experiments are designed to identify the impact of the observational density, heterogeneity, and uncertainty, as well as operational constraints (i.e., ensemble size, number of descent iterations) on the DA estimates relative to the estimates from a batch inverse modeling scheme. No dynamical model is explicitly specified for the DA approaches to keep the problem setup analogous to a typical real CO<sub>2</sub> flux estimation problem. Results demonstrate that the performance of the DA approaches depends on a complex interplay between the measurement network and the operational constraints. Overall, the variational approach (contingent on the availability of an adjoint transport model) more reliably captures the large-scale source–sink patterns. Conversely, the ensemble square root filter provides more realistic uncertainty estimates. Selection of one approach over the other must therefore be guided by the carbon science questions being asked and the operational constraints under which the approaches are being applied.

## 1 Introduction

Data assimilation (DA) is best known as a tool in numerical weather prediction (NWP; e.g., Swinbank, 2010) and has been applied to analyze complex data sets and estimate parameters in a variety of fields, including atmospheric constituent (e.g., Lahoz and Errera, 2010; Elbern et al., 2010), oceanographic (e.g., Haines, 2010), and land surface (e.g., Reichle, 2008; Houser et al., 2010) assimilation problems. In all such applications, a DA system aims to optimally combine the information from available observations with a prior model estimate (or the background derived from a model forecast) based on their respective uncertainty estimates.

DA approaches for estimating CO<sub>2</sub> fluxes aim to constrain the spatial and temporal distributions of CO<sub>2</sub> sources and sinks by integrating atmospheric, terrestrial and oceanic data together into a common analysis framework. CO<sub>2</sub>-DA applications (e.g., ensemble-filter-based approaches; Peters et al., 2005; Feng et al., 2009; Miyazaki et al., 2011; Chatterjee et al., 2012; Kang et al., 2012); variational based approaches (Rayner et al., 2005; Chevallier et al., 2005; Rödenbeck, 2005; Baker et al., 2006), and hybrid ensemble–variational approaches such as the maximum likelihood ensemble filter (Zupanski et al., 2007; Lokupitiya et al., 2008) have been in vogue for nearly a decade and are viewed as an alternative to more traditional batch inverse modeling schemes. Unlike these newer DA applications, which use a combination of numerical approximations and time-stepping approaches, the batch schemes directly solve the linear system of equations relating the fluxes and the atmospheric CO<sub>2</sub> observations in a single step. The DA approaches are attractive because of

their computational efficiency (e.g., Rayner, 2010), but the impact of their underlying numerical approximations on the final estimates and their associated uncertainties is often unclear.

Chatterjee et al. (2012) pointed out fundamental differences between the carbon flux estimation (i.e., the inverse framework) and the NWP/constituent (i.e., assimilation framework) problems – namely that (a) the performance of the DA approaches are not necessarily equivalent for the two frameworks and (b) that only under specific inversion scenarios are the DA approaches able to perform optimally. Differences between the two frameworks are mainly driven by the ill-conditioned and highly diffusive nature of the flux estimation problem, as well as the absence of an explicit dynamical model that can evolve a set of estimated fluxes forward in time. By propagating the state vector between different assimilation time steps, a dynamical model directly contributes to the growth of the eigenvalue spectrum of the state covariance matrix in certain preferred directions and the decay in others (Bengtsson et al., 2003; Furrer and Bengtsson, 2007). For the CO<sub>2</sub> inverse problem, however, the dynamics are embedded within the atmospheric transport and are not used explicitly to inform the temporal evolution of the state vector. A few recent studies have attempted to use an explicit dynamical flux model (e.g., Kuppel et al., 2013), but they note that model errors significantly reduce the performance of the inversion in terms of the quality of the estimated fluxes. Currently, in most carbon flux estimation studies, dynamical models are not used, resulting in a loss of potentially valuable information to the DA system. The absence of this information, coupled with the availability of only sparse observational data sets, may result in the DA approaches performing suboptimally.

The authors are not aware of any study specifically related to the CO<sub>2</sub> flux estimation problem that attempts to evaluate the relative performance of DA techniques. This is unlike the weather forecasting community, where several studies have evaluated the strengths and weaknesses of ensemble and variational approaches for different weather-related applications ranging from simple to chaotic nonlinear systems (e.g., Lorenc, 2003; Caya et al., 2005; Fertig et al., 2007; Kalnay et al., 2007; Liu et al., 2008; Whitaker et al., 2009; Buehner et al., 2010a, b; Jardak et al., 2010; Zhang et al., 2011; also see the special collection of papers on intercomparison at [http://journals.ametsoc.org/page/Ensemble\\_Kalman\\_Filter](http://journals.ametsoc.org/page/Ensemble_Kalman_Filter)). Apart from NWP-related comparison studies, DA approaches have also been intercompared for chemical (e.g., Carmichael et al., 2008) and constituent (e.g., ozone – Wu et al., 2008) assimilation problems. These comparison studies cannot be used as a baseline, however, because of differences between the flux estimation and the NWP/constituent DA frameworks, as stated earlier.

The main purpose of this work is thus to fill this gap and build on the existing body of intercomparison studies from the perspective of the CO<sub>2</sub> flux estimation problem. Specifi-

cally, this study aims to answer the following two questions: (1) What is the relative performance of two state-of-the-art DA approaches (ensemble square root filter, EnSRF (e.g., Whitaker and Hamill, 2002), and 4-dimensional variational, 4D-VAR (e.g., Talagrand, 2010) for solving the CO<sub>2</sub> inverse problem, and (2) how well can the DA approaches reproduce the flux estimates and associated uncertainties from a batch inverse modeling (BIM) scheme?

To facilitate the intercomparison, we consider here a one-dimensional (1-D) passive tracer transport problem. Similar to previous studies (e.g., Liu and Rabier, 2002; Park and Kalnay, 2004), the 1-D framework allows us flexibility in setting up the problem because multiple experiments can be simulated in a computationally efficient manner. The low computational cost associated with the 1-D problem enables the implementation of a BIM approach in addition to the DA approaches. The DA estimates are thus compared both to the true signal and to the BIM estimates in order to isolate the degradation due to the underlying numerical approximations in the DA approaches. This study assesses whether these approximations limit the ability of the examined DA approaches to be used as suitable long-term replacements for the BIM approach under different inversion conditions.

When designing the 1-D problem, we focus on a framework that allows us to examine an *underdetermined and fine-scale* CO<sub>2</sub> flux estimation problem. This setup is necessary to mimic the challenges of a true CO<sub>2</sub> flux estimation problem in which atmospheric mixing coupled with the sparseness of available observations results in the inverse problem being highly underdetermined and ill posed. The underdetermined nature of the problem is accentuated by the need for estimating CO<sub>2</sub> fluxes at fine spatial and temporal scales, which is necessary to not only avoid spatiotemporal aggregation errors (e.g., Kaminski et al., 2001; Gourdji et al., 2012) but also to improve the understanding of the fine-scale processes driving the carbon cycle. This paradigm shift has brought about a computational bottleneck in solving the BIM problem, which requires the 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 in turn has prompted the use of computationally efficient alternatives, such as DA approaches, in which the number of atmospheric transport model runs is proportional to the number of ensemble members (in the ensemble approach) or the number of descent iterations (in the variational approach), both of which are typically set to be orders of magnitude lower than the number of estimated parameters or available observations. Analogous to a real CO<sub>2</sub> flux estimation problem, no dynamical model is explicitly specified for solving the 1-D problem.

The experiments are specifically targeted to evaluate the impact of three factors on the two DA approaches: (a) the impact of the observational density and homogeneity, (b) the impact of the model–data mismatch covariance, and (c) the impact of the operational parameters of the DA system (i.e.,

ensemble size, number of iterations). While examining the first two factors, issues of sampling and convergence error are minimized by specifying a large number of ensemble members and descent iterations for the EnSRF and the 4D-VAR<sup>1</sup> approaches, respectively. More realistic operational constraints are subsequently imposed in a latter set of experiments to not only evaluate the fundamental differences between the two DA approaches but also the effect of the compromises necessary to make the algorithms practically applicable. This study represents the first comparison of the EnSRF and the 4D-VAR approaches for a flux estimation problem, and is expected to guide the development of future intercomparison experiments with real data, including satellite observations of atmospheric CO<sub>2</sub>.

## 2 Experimental framework

### 2.1 Estimation methods

In a Bayesian framework, prior information and likelihood are expressed as probability density functions or pdfs (e.g., Enting, 2002). If the pdfs can be approximated as Gaussian, then maximizing the posterior probability of the state is equivalent to finding the minimum of a quadratic objective function:

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

where  $s$  is a  $m \times 1$  vector of the discretized state (e.g., CO<sub>2</sub> flux),  $z$  is the  $n \times 1$  vector of observations (e.g., CO<sub>2</sub> observations),  $h$  is a forward model that is often a combination of an observation operator and an atmospheric transport model with embedded dynamics,  $\mathbf{R}$  is the  $n \times n$  model–data mismatch covariance matrix,  $s^b$  is the  $m \times 1$  prior estimate of the state,  $\mathbf{Q}^b$  is the  $m \times m$  error covariance matrix of the prior estimate  $s^b$ , and the superscript  $T$  denotes the matrix transpose operation. Note that in the case of the atmospheric CO<sub>2</sub> inverse problem,  $h$  is linear and typically represented via its matrix form  $\mathbf{H}$  (i.e., sensitivity matrix with dimensions  $n \times m$ , or Jacobian matrix), which captures the sensitivity of the observations  $z$  to the fluxes  $s$  (i.e.,  $\mathbf{H}_{i,j} = \partial z_i / \partial s_j$ ). The inverse problem as formulated via Eq. (1) determines a least-squares fit of the state estimate to the observations and the

<sup>1</sup>Typically in the DA community the term 4D-VAR is used to represent the three-dimensional space plus time. In this study, the variational approach is applied to a one-dimensional space plus time, which may suggest that the term “1 + 1D-VAR” may be more appropriate. Within the geophysical community, however, the term 1-D in 1D-VAR specifically refers to the vertical column, and is quite popular for radio occultation data (e.g., Eyre et al., 1993; Poli et al., 2002), total column water vapor (e.g., Marécal and Mahfouf, 2002; Bauer et al., 2006), cloud (e.g., Janiskova et al., 2012) assimilation etc. Since in the current study 1-D refers to a single dimension along the horizontal space and not necessarily the vertical column, we persist with usage of the term 4D-VAR rather than the term 1D-VAR.

prior, with the ultimate aim of estimating the posterior pdf of the true state.

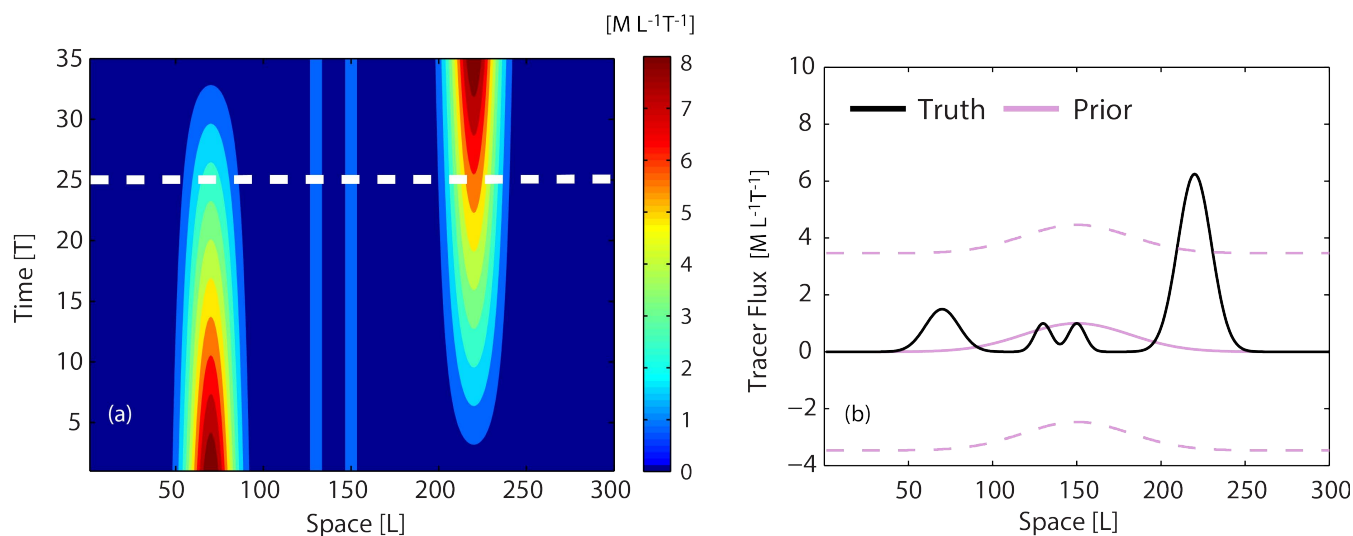
The estimate of the pdf of the true state is obtained here via three approaches – BIM, EnSRF and 4D-VAR. For Gaussian pdfs and a linear model, BIM analytically solves the linear system of equations resulting from the minimization outlined in Eq. (1) to obtain the parameters of the posterior pdf. 4D-VAR and EnSRF provide numerical approximations of this solution, and if perfectly implemented, they will yield the same solution, which in practice is never feasible due to operational constraints. The best estimate for EnSRF is defined as the mean estimate across an ensemble of source–sink realizations, and the posterior uncertainties are defined based on the ensemble spread at each location/time. The best estimate for 4D-VAR is defined as the maximum likelihood estimate that minimizes the misfit between a prior guess of the source–sink estimate and the observations that are available over a given assimilation window. The posterior analysis covariances for 4D-VAR need to be calculated indirectly, however, for example via a Monte Carlo algorithm (e.g., Chevallier et al., 2007). In the Monte Carlo framework, different 4D-VAR estimates based on perturbations of the observations (using the specified model–data mismatch statistics), perturbations of the prior (using the specified prior error statistics), and combinations thereof. In practice, for a real CO<sub>2</sub> flux inversion problem it may not be computationally feasible to run 4D-VAR with a large number of perturbations. The computational efficiency of the one-dimensional framework in this study allows us to be generous with the number of perturbations (25 total) to assess the quality of the Monte Carlo-algorithm-based error statistics relative to the uncertainty estimates obtained directly from the other two approaches.

A review of these approaches and their underlying mathematical framework is available in Appendix A, along with an exposition of the algorithmic choices necessary to adapt the DA approaches for solving the CO<sub>2</sub> flux estimation problem.

### 2.2 Problem description

A 1-D advection–diffusion problem of a passive tracer is used to emulate the CO<sub>2</sub> flux estimation problem. In the 1-D illustration, the passive tracer represents atmospheric CO<sub>2</sub>. Tracer fluxes get released from a series of locations over a finite duration and get transported by a tracer transport model that emulates the physics of advection and diffusion. No sink is specified, and there is therefore a gradual buildup of the passive tracer within the domain. Observations of the tracer are obtained at various locations and times within the domain. The locations and times of the observations as well as their precision can be regulated to simulate a variety of inversion scenarios. The inverse problem involves using the noisy tracer observations along with the transport information to infer the original tracer fluxes.

In the following description, the units of mass, length and time are reported as  $[M]$ ,  $[L]$ , and  $[T]$  to keep the problem



**Fig. 1.** (a) Spatiotemporal variability of the tracer flux. (b) True flux profile for a particular time period corresponding to the dashed white line in (a). Also shown in (b) is the prior estimate of the tracer flux profile with its  $\pm 2\sigma_Q$  prior uncertainty (dashed lines).

generic. Both the length of the 1-D domain and the time period of the experiment are arbitrarily discretized. The parameters for the experiment are: the grid size  $\Delta x = 1$  [L], the domain length  $x = 300$  [L], the time step of release  $\Delta t = 1$  [T], the total number of time periods over which the tracer flux is released  $t = 35$  [T], the longitudinal dispersion coefficient  $D_L = 2.0$  [ $L^2 T^{-1}$ ], and the advection velocity  $v = 50.0$  [ $L T^{-1}$ ].

The tracer flux  $s$  [ $M L^{-1} T^{-1}$ ] that is released (Fig. 1a) is modeled as

$$s(x_r, t_r) = 0.25(36 - t_r) \exp\left[-\frac{(x_r - 70)^2}{200}\right] + \exp\left[-\frac{(x_r - 130)^2}{50}\right] + \exp\left[-\frac{(x_r - 150)^2}{50}\right] + 0.25(t_r) \exp\left[-\frac{(x_r - 220)^2}{200}\right], \quad (2)$$

where  $x_r$  represents the locations along the 1-D domain ( $x_r = 1, 2, 3, \dots, 299, 300$  [L]) over which the tracer flux is released continuously over 35 fixed intervals ( $t_r = 1, 2, 3, \dots, 34, 35$  [T]), with each interval being a duration of 1 [T]. Equation 2 is designed to model two large peaks with fluctuating amplitudes (Fig. 1b) between 50 and 100 [L] and between 200 and 250 [L], as well as a smaller consistent double peak (Fig. 1b) between 100 and 200 [L]. Even though the spatial tracer flux profiles are different for each time period, the spatially averaged flux has a constant value of  $0.84$  [ $M L^{-1} T^{-1}$ ] across all time periods. Note that the true tracer flux  $s$  is used only for simulating the observations and is later assumed unknown throughout the analysis.

The tracer is sampled at locations  $x_o$  ( $x_o = 1, 2, \dots, 299, 300$  [L]) for 35 consecutive time periods ( $t_o = 1.5, 2.5, \dots, 34.5, 35.5$  [T]) to obtain the observational data set  $z$  [ $M L^{-1}$ ], such that the observation times are offset from the release times. The initial random error with a prespecified

variance ( $10$  [ $M^2 L^{-2}$ ]) is added to the tracer observations to simulate measurement, transport, aggregation, and representation errors. Later in the study, different configurations of  $x_o$  and error variances ( $\sigma_{R2}$ ) are prescribed to test the impact of factors.

The tracer observations ( $z$ ) and the tracer fluxes ( $s$ ) are related in the following fashion:

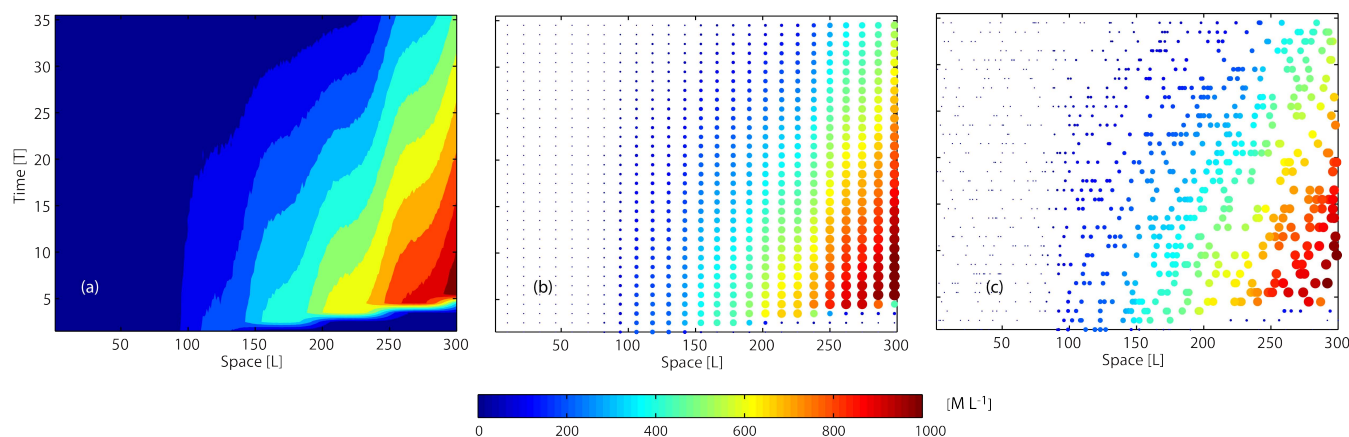
$$z = \mathbf{H}s + v, \quad \text{where } v \sim N(0, \mathbf{R}), \quad (3)$$

where  $\mathbf{H}$  is the sensitivity matrix that is generated using a 1-D tracer transport model as

$$\mathbf{H}_{o,r} = \frac{1}{2} \left[ \operatorname{erfc}\left(\frac{(x_o - x_r) - vt_o}{2\sqrt{D_L t_o}}\right) - \operatorname{erfc}\left(\frac{(x_o - x_r) - v(t_o - t_r)}{2\sqrt{D_L (t_o - t_r)}}\right) \right], \quad (4)$$

where  $x_r$  and  $t_r$  are the tracer flux release locations and times,  $x_o$  and  $t_o$  are the tracer observation locations and times,  $\operatorname{erfc}$  represents the complementary error function. The tracer transport model embedded in Eq. (4) assumes conservation of mass and is based on a well-known one-dimensional analytical solution for a conservative tracer with a continuous source under steady-state conditions (e.g., Ogata and Banks, 1961; Runkel, 1996).

The tracer observations obtained at a particular time step are sensitive to the tracer flux released at multiple previous time steps. Given that the total length of the domain is 300 [L] and the advection velocity is  $50$  [ $L T^{-1}$ ], the maximum residence time of the tracer within the domain is approximately 6 [T]. Based on the form of Eq. (2), however, the majority of fluxes occur at  $x_r > 50$  [L] (Fig. 1b), and the typical residence time is therefore  $< 5$  [T]. This means that an observation taken at time  $t_o$  [T] provides information about the tracer flux approximately up to time  $t_o - 5$  [T]. In all subsequent experiments, the lag window size for the DA approaches is thus set to 5 [T], such that this “long” lag window



**Fig. 2.** Observations of the tracer obtained from the three network configurations – REF (a), HM (b), and HT (c). Note that going from the REF to the HM and the HT networks, the total number of observations decreases by a factor of 12, whereas in going from the HM to the HT network, the observational network becomes more heterogeneous in space and time.

allows nearly all of the flux influence on observations to be represented within the DA approaches. The finite lag window also recognizes the operational limitations associated with the implementation of DA approaches in a real-world setting, and we foresee this as a realistic scenario for future global inversions (see Appendix A for a more detailed discussion).

The final piece of information necessary for setting up the inverse problem is the prior estimate ( $s^b$ ) of the tracer flux and its error characteristics (Fig. 1b).  $s^b$  is chosen here to be constant across all time periods:

$$s^b(\mathbf{x}_r, t_r) = \exp \left[ -\frac{(\mathbf{x}_r - 150)^2}{2000} \right]. \quad (5)$$

Its error covariance matrix  $\mathbf{Q}^b$  is based on an exponential decay model in space, with a correlation length ( $3l_Q$ ) of 90 [L] and variance ( $\sigma_Q^2$ ) of 3 [ $M^2 L^{-2} T^{-2}$ ].

The 1-D framework was designed to capture most of the characteristic features of the CO<sub>2</sub> flux estimation problem. For a real CO<sub>2</sub> inversion, units are most typically [ $\mu\text{mol} (\text{m}^2 \text{s})^{-1}$ ] for  $s$  and  $\sigma_Q$ , [ppm] for  $z$  and  $\sigma_R$ , [ $\text{ppm} \mu\text{mol}^{-1} (\text{m}^2 \text{s}^{-1})$ ] for  $\mathbf{H}$ , and [km] for  $l_Q$ .

### 2.3 Experiments

Experiments are designed to explore the impact of three factors on the ability of the DA approaches to solve the inverse problem: (a) the observational density and homogeneity, (b) the model–data mismatch covariance, and (c) the operational constraints of the DA system (i.e., ensemble size, number of descent iterations). In all the experiments, the size of the state vector or the total number of fluxes to be inferred is  $10\,500 \times 1$  (i.e., 300 locations  $\times$  35 times).

The first set of experiments (Table 1 – experiments A through C) aims to investigate the effect of the density and

spatiotemporal homogeneity of the observational network. Three different observational networks are designed (Fig. 2). In the first network configuration (denoted as REF – “reference observational set” as outlined in Sect. 2.2), observations are obtained throughout the domain ( $\mathbf{x}_0 = 1, 2, \dots, 299, 300$  [L]) and for all 35 measurement times ( $t_0 = 1.5, 2.5, \dots, 34.5, 35.5$  [T]) (Fig. 2a). The total number of observations available is thus 10 500 (i.e., 300 locations  $\times$  35 times). In the second network configuration (denoted as HM – “homogeneous”), observations are obtained at 25 equally spaced locations within the 1-D domain ( $\mathbf{x}_0 = 10, 22, 34, \dots, 298$  [L]) for all 35 time periods ( $t_0 = 1.5, 2.5, \dots, 34.5, 35.5$  [T]) (Fig. 2b). The total number of observations is thus reduced to 875 (i.e., 25 locations  $\times$  35 times). In the final configuration (denoted as HT – “heterogeneous”), observations are taken at 25 randomly selected locations for each measurement time ( $t_0 = 1.5, 2.5, \dots, 34.5, 35.5$  [T]) (Fig. 2c), and these locations vary from one time to the next. Similarly to HM, the total number of observations in HT is 875 (i.e., 25 locations  $\times$  35 times), but the observations are neither uniform in space nor consistent in time. Note that unlike REF, both HM and HT represent underdetermined inversion problems where the total number of observations is substantially lower than the number of unknowns in the state space to be estimated. In reality, the HT network configuration scheme is the closest to current CO<sub>2</sub> monitoring networks where different monitoring locations (ground based or remote sensing) can come online and go offline over different periods.

For all the three network configurations (experiments A through C), random errors with a variance of 10 [ $M^2 L^{-2}$ ] are added to the observations to represent measurement, transport, aggregation, and representation errors encountered in real applications. The model–data mismatch covariance matrix  $\mathbf{R}$  (Eq. 1) has this same variance as its diagonal elements. In contrast to the prior error covariance, the errors

**Table 1.** Summary of the experiments outlined in Sect. 2.3. The following parameters are held constant for all the experiments in this study:  $s^b$  (Eq. 5),  $3l_Q = 90$  [L] and  $\sigma_Q^2 = 3$  [ $M^2 L^{-2} T^{-2}$ ].

Impact considered	Experiment	Observation parameters		DA parameters	
		Network	Model–data mismatch Variance $\sigma_R^2$	Ensemble size	Descent iterations
Observational density and homogeneity	A	REF	10	1000	250
	B	HM	10	1000	250
	C	HT	10	1000	250
Model–data mismatch	AR	REF	400	1000	250
	BR	HM	400	1000	250
	CR	HT	400	1000	250
Operational limitations	AO	REF	10	100	25
	BO	HM	10	100	25
	CO	HT	10	100	25

have no spatial or temporal correlation. Finally, all three approaches use the same numerical realization of errors, thus ensuring that they are solving the same inverse problem.

The second set of experiments (Table 1 – experiments AR through CR) examines the effect of the model–data mismatch variance on the best estimates and their associated uncertainties. For all the network configurations, the variance of the random errors is increased to 400 [ $M^2 L^{-2}$ ] with the diagonal values of the model–data mismatch covariance matrix **R** increased accordingly. All other parameters are kept the same as in the first set of experiments.

The third set of experiments (Table 1 – experiments AO through CO) explores the impact of operational constraints, which are always an important consideration in implementing a DA system. To minimize numerical approximations and avoid sampling or convergence errors, the ensemble size (for EnSRF) and the number of descent iterations (for 4D-VAR) for the first two sets of experiments (Table 1 – experiments A–C and experiments AR–CR) are set to 1000 and 250, respectively. The number of descent iterations is prescribed to be lower than the number of ensemble members, keeping in mind that 4D-VAR typically requires more model integrations (i.e., both forward and adjoint model run) than EnSRF. Given that it is not feasible to either run a large number of ensemble members or specify a large number of descent iterations for real atmospheric applications, these numbers are reduced to 100 ensemble members for EnSRF and 25 descent iterations for 4D-VAR in the third set of experiments. The noise added to the observations is kept the same as in the first set of experiments, namely 10 [ $M^2 L^{-2}$ ], to allow for a direct comparison with experiments A–C.

### 3 Results

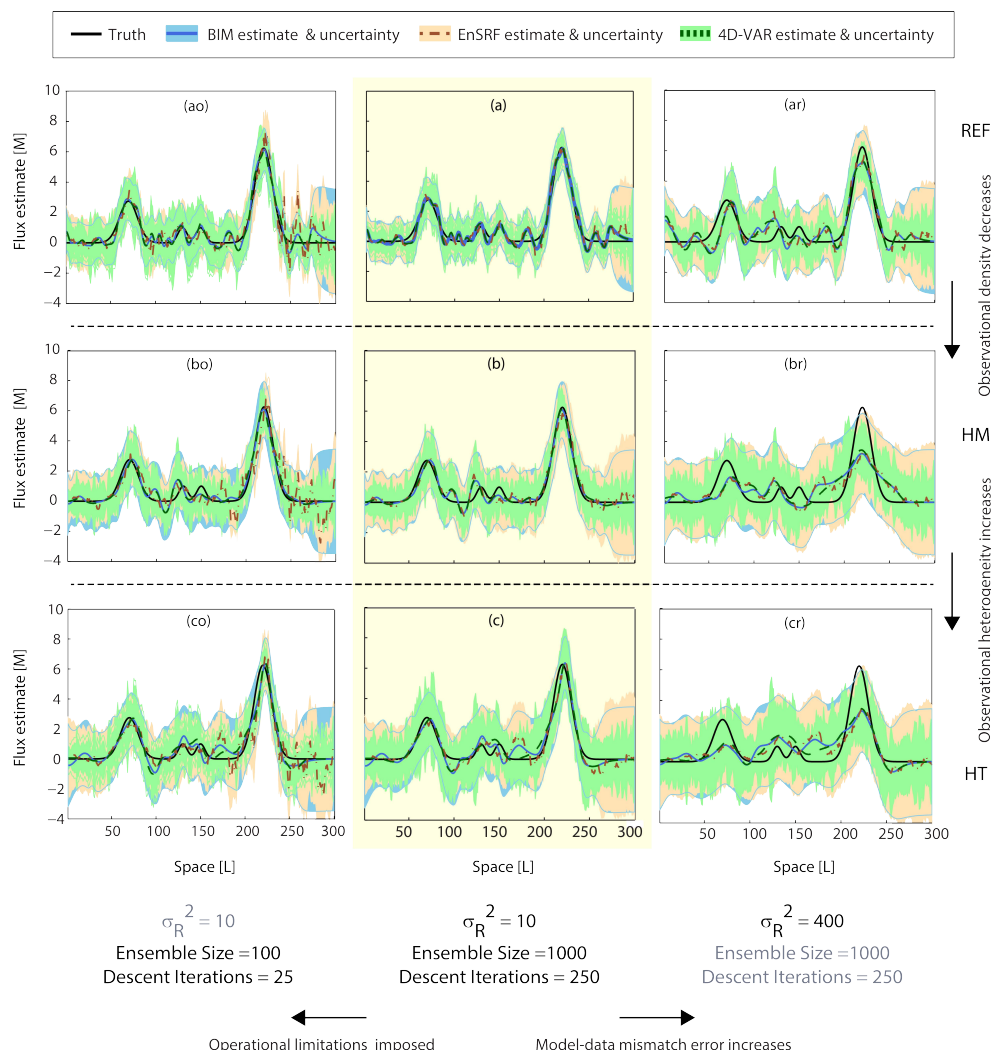
In the sections that follow, results from the nine experiments are interpreted at both the native and aggregated spatial scales, and estimates from the EnSRF and the 4D-VAR approaches are compared both to the truth and to the estimates from the BIM approach. Taylor diagrams (Taylor, 2001) are used to assess the root-mean-square difference (RMSD) and the correlation coefficient (CC) between the flux estimates and the truth, as well as the standard deviation (SD) of the flux estimates and the truth. These metrics are calculated across 30 time periods ( $t_r = 6, 7, \dots, 34, 35$  [T]) to be representative of the overall experiment after discarding the first five time periods as spin-up.

#### 3.1 Impact of observational density and homogeneity

For the REF network (experiment A), all three approaches perform well in recovering the true flux (e.g., Fig. 3a), and in fitting the observations within the specified model–data mismatch errors (results not shown). For the sample time period presented in Fig. 3a, both the 4D-VAR and the EnSRF estimates capture the flux profile, including its large and small peaks. These results are typical of the performance of the three approaches across other estimation times. The performance across the full examined time period is summarized in Fig. 4a, where all three approaches show a high CC ( $\sim 0.97$ ), low RMSD ( $\sim 0.3$  [ $M L^{-1} T^{-1}$ ]), and standard deviations ( $\sim 1.5$  [ $M L^{-1} T^{-1}$ ]) that are similar to that of the true fluxes.

The performance of all three approaches degrades as the observational density and homogeneity decrease in going from experiments A to C. This is evident by looking at Fig. 3b and c, where the estimates fail to capture the smaller





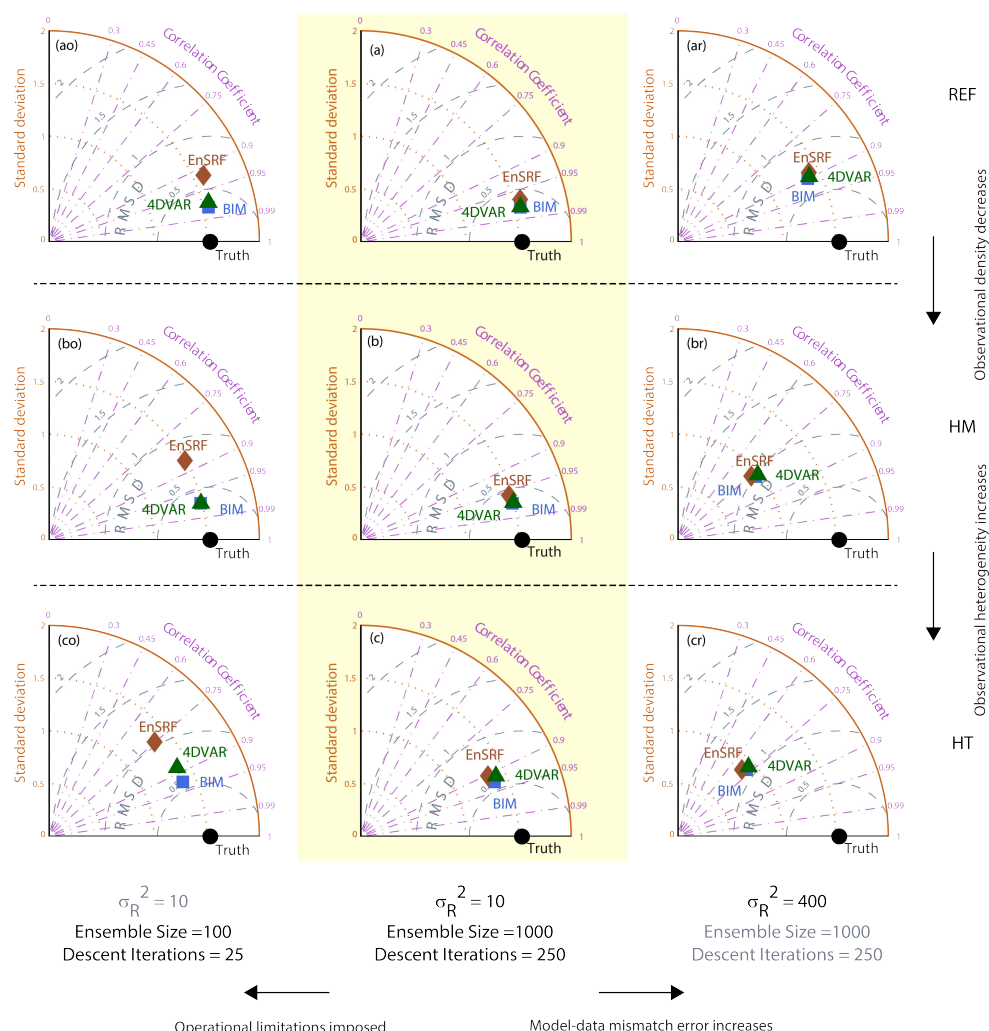
**Fig. 3.** Example of estimated tracer fluxes (lines) and associated  $\pm 2\sigma_{\hat{s}}$  uncertainties (shaded areas) for the different approaches assessed in this study. All values are shown for the 25th time period, which is representative of the observed performance over other time periods. The panel titles correspond to the different experiments outlined in Table 1.

double peak around 100–200 [L], and the Taylor diagrams in Fig. 4b and c show a corresponding drop in CC and an increase in RMSD. In general, for observations with spatially uncorrelated model–data mismatch errors such as those used here, decreasing the observational density is expected to decrease the analysis accuracy. The response of the two DA approaches mirrors the BIM approach in such cases, including the inference of an incorrect flux pattern for the HT network around 100 [L] in Fig. 3c. This result indicates that in the absence of operational constraints, best estimates from the DA approaches are consistent with the BIM estimate even for an underdetermined inverse problem.

In terms of the recovered posterior uncertainty estimates, the EnSRF uncertainty estimates are more consistent with the BIM uncertainty estimates relative to 4D-VAR (Fig. 3a–c). For the REF observational configuration, the average ra-

tios of the predicted posterior uncertainty of the individual flux estimates in EnSRF ( $\sigma_{\hat{s}_{\text{EnSRF}}}$ ) and 4D-VAR ( $\sigma_{\hat{s}_{\text{4D-VAR}}}$ ) to those from BIM ( $\sigma_{\hat{s}_{\text{BIM}}}$ ) are approximately 0.98 and 0.84, respectively; that is, on average, EnSRF and 4D-VAR underestimate the posterior uncertainties by 2 and 16 %, respectively, relative to BIM (Fig. 5a). As the observational density changes, EnSRF overestimates the uncertainty by 2 and 6 % for HM and HT, respectively, while 4D-VAR underestimates the uncertainty substantially, by 22 and 18 %, respectively (Fig. 5b, c). For all cases, the 4D-VAR uncertainties for individual locations/times over- or underestimates the BIM estimates even more substantially than evidenced by the average statistics, however, as seen by the spread in the histograms in Fig. 5.

In the EnSRF framework, the uncertainties are directly related to the ensemble spread. In the absence of a dynamical



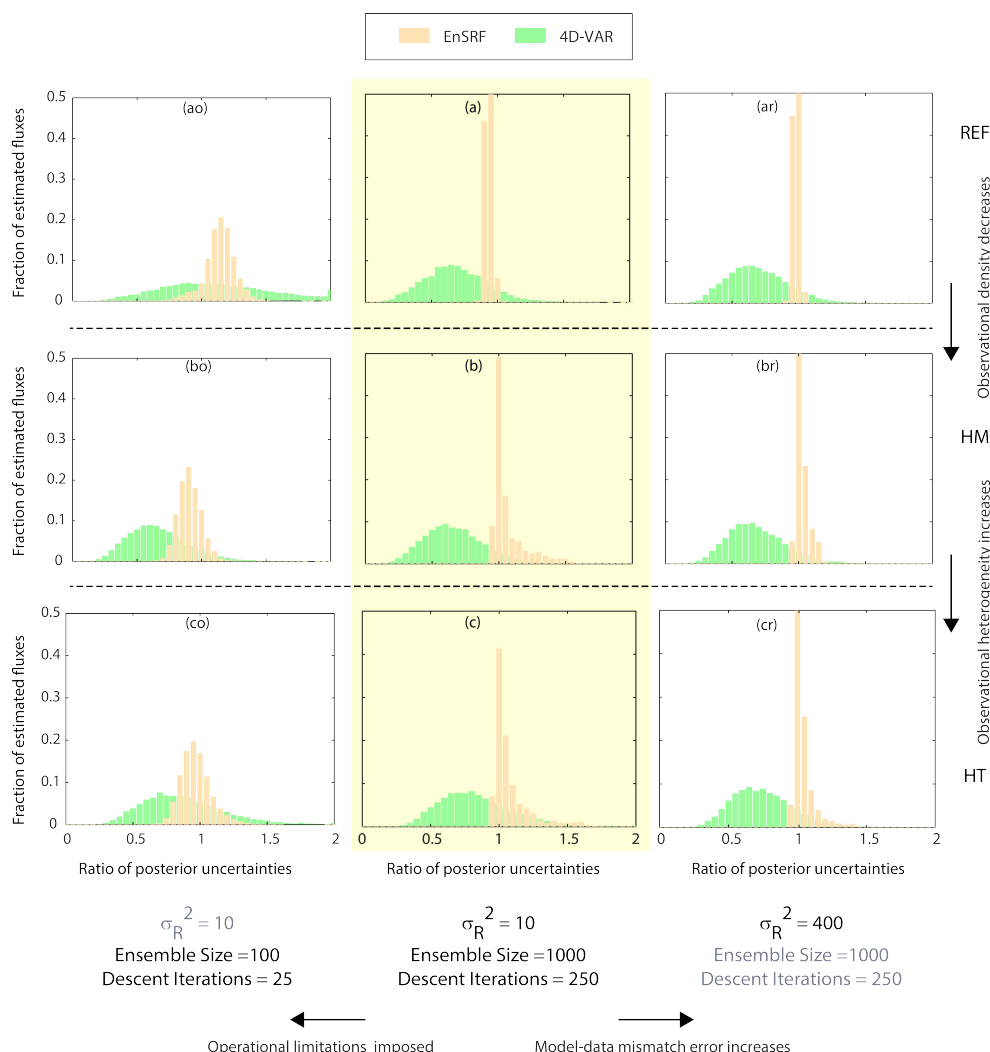
**Fig. 4.** Performance of the BIM, the EnSRF, and the 4D-VAR approaches for the different experiments outlined in Table 1. For each experiment, statistics are calculated between the estimates and the true fluxes across all locations and all 30 time periods and are represented on a Taylor diagram. For each Taylor diagram, the true flux is represented by a point along the abscissa corresponding to the standard deviation of the true fluxes (“Truth”). All other points (“BIM”, “EnSRF”, “4D-VAR”), which represent the estimated fluxes, are positioned such that their standard deviation is the radial distance from the origin, the correlation coefficient between the estimates and the truth is the cosine of the azimuthal angle, and the root-mean-square difference (RMSD) between the estimates and the truth is the distance to the observed point. In the limit of perfect agreement, these other points would coincide with “Truth” (i.e., RMSD = 0, CC = 1, and SD of the estimates would be the same as that of the truth).

model, there is little source of variability for the ensemble to maintain a consistent spread. As observations are assimilated, the ensemble members tend to collapse to the ensemble mean and the adaptive inflation (see Appendix A.3) has to compensate for this degeneracy by inflating the ensemble spread. In the HT case, however, the inflation technique has a delayed response in adjusting to the heterogeneity in the observational network, as different observation locations come into and out of the network. For the adaptive inflation component to function well, we find it beneficial to have a consistent set of observations to maintain a reasonable ensemble spread (e.g., Chatterjee et al., 2012). It is worthwhile

to mention here that the magnitudes of the inflation factors are very small in experiments A–C. This is not surprising given that a large number of ensemble members have been specified and the sampling error is hence quite low.

For 4D-VAR the posterior uncertainties are obtained via the Monte Carlo technique. For the time period shown in Fig. 3a–c, the 4D-VAR uncertainties not only underestimate the BIM uncertainties on average but they are also a lot more variable (i.e., too high or too low) for individual fluxes relative to BIM. We believe this heterogeneity in the posterior uncertainty estimates to be a result of the number of perturbations specified for the Monte Carlo technique. A sensitivity





**Fig. 5.** Histogram showing the ratio of the estimated posterior uncertainties from the EnSRF ( $\sigma_{\text{EnSRF}}$ ) and the 4D-VAR ( $\sigma_{\text{4D-VAR}}$ ) approaches to the posterior uncertainties from the BIM ( $\sigma_{\text{BIM}}$ ) approach. The ratios are calculated for each estimated flux time/location over all 30 time periods to be representative of the overall experiments outlined in Table 1. In the ideal case – that is, if the posterior uncertainty from a DA approach is equal to the posterior uncertainty from BIM – the histogram would be a single line centered around 1.

test in which the total number of perturbations is increased (or decreased) indicated that the heterogeneity of the uncertainty estimates obtained via the Monte Carlo technique decreased (or increased) correspondingly. When averaged over all the time periods (Fig. 5a–c), the posterior uncertainty estimates clearly underestimate the BIM uncertainty. Even then, the large spread in the histograms shown in Fig. 5a–c reinforces our earlier conclusions that the uncertainty estimates for individual fluxes over/underestimate the BIM uncertainty estimates substantially.

Overall, we find that both 4D-VAR and EnSRF can reproduce the performance of BIM in terms of the best estimates of the fluxes for all three observational network configurations. Even though small discrepancies are noticeable, the impact of a sparse and/or heterogeneous observational net-

work is similar for the DA approaches compared to the BIM approach. Both the DA approaches have some difficulty in reproducing the BIM posterior uncertainty estimates, albeit for different reasons. In the absence of any operational constraints, however, EnSRF provides more realistic and useful uncertainty bounds than 4D-VAR.

### 3.2 Impact of model–data mismatch covariance

For all the network configurations, the quality of the estimates degrades when a higher model–data mismatch error is prescribed (comparing Fig. 3a–c with Fig. 3a–c), although the heterogeneous network estimates show the most pronounced degradation. An increase in  $\sigma_{\text{R}}^2$  leads to higher uncertainties for the estimates, indicative of the decreased confidence in the analysis. Analogous to the first set of

experiments, EnSRF and 4D-VAR best estimates respond similarly to BIM when the model–data mismatch covariance changes, and both the approaches track the BIM best estimates quite well for all the three experiments. Figure 4ar–cr confirm that the best estimates from all the three approaches have a lower CC, higher RMSD, and lower SD when compared to Fig. 4a–c.

The standard deviation of the flux estimates change considerably between Fig. 4ar ( $\sim 1.45\text{--}1.5 [ML^{-1}T^{-1}]$ ) and 4cr ( $\sim 0.94\text{--}1.00 [ML^{-1}T^{-1}]$ ) for the three approaches. Increasing the  $\sigma_R^2$  to  $400 [M^2L^{-2}]$  results in the analysis rejecting the information from the observations and giving more weight to the prior, yielding overly smooth a posteriori estimates. A typical example of this is seen by comparing the estimated peak around  $50\text{--}100 [L]$  in Fig. 3c and cr. Estimates in both these panels are based on the same observational network but the estimates in Fig. 3cr do not capture the amplitude of the two large peaks in the true flux signal.

For experiments AR–CR, the posterior uncertainty estimates for all the three approaches are higher compared to experiments A–C, as expected due to the higher prescribed model–data mismatch error. Similarly to experiments A–C, the 4D-VAR uncertainty estimates for individual locations/times are too variable relative to BIM (Fig. 5). Averaged over time and space, the 4D-VAR uncertainty estimates underestimate the BIM uncertainty estimates by approximately 25 % (Fig. 5ar–cr). Thus, even though the 4D-VAR uncertainty estimates for experiments AR–CR are higher than the corresponding uncertainty estimates for experiments A–C, they fail to capture the full magnitude of the BIM uncertainty estimates. This makes intuitive sense due to the indirect approach adopted for generating the 4D-VAR uncertainty estimates. Conversely, as the observational network becomes sparser and more heterogeneous, the EnSRF slightly overestimates the BIM average uncertainties by 3 % (HM; Fig. 5br) and 5 % (HT; Fig. 5cr), while it underestimates the uncertainty by only 1 % for the reference network (Fig. 5ar). The EnSRF uncertainty estimates for individual locations/times are more closely distributed around the BIM estimates (Fig. 5). The better performance of EnSRF in terms of the uncertainty estimation can be directly related to the ensemble spread. Relative to experiments A–C, when the prescribed model–data mismatch error is high in experiments AR–CR, the initial ensemble spread is reduced by a lower amount as observations are now being given less weight and hence have lower impact on the ensemble spread. Consequently, the ensemble members maintain a large spread throughout the analysis and results in large posterior uncertainty estimates that are more realistic relative to 4D-VAR.

Experiments AR–CR reconfirm that in the absence of operational limitations, an increase (or decrease) in the model–data mismatch covariance does not affect the ability of the DA approaches to reproduce the BIM best estimates. Similar to experiments A–C, in terms of the posterior uncertainty estimates, however, (a) both the DA approaches are less skilled

at reproducing the uncertainty estimates from BIM, and (b) the uncertainty estimates from 4D-VAR severely underestimate the BIM uncertainty estimates and are less realistic than the EnSRF uncertainty estimates.

### 3.3 Impact of operational constraints

Operational constraints hinder the performance of the DA approaches, and the impact is further intensified as the observational network becomes more heterogeneous.

For 4D-VAR, an inadequate number of iterations may lead to a failure to find the minimum of the quadratic objective function (convergence results not shown here). When the observational network is heterogeneous, the minimization has even more difficulty in finding the path towards the minimum. Thus, comparing Fig. 4ao–co, the 4D-VAR estimates diverge from the BIM estimates for the HT network configuration. In general, we find that for the HT network, 4D-VAR needs approximately 50 iterations to converge completely for the case studies presented here. Conversely, for the REF and the HM network, 4D-VAR requires only approximately 20 to 30 iterations to reach full convergence. For all the three experiments, however, the value of the objective function is reduced relative to that for the prior fluxes, indicating an improvement over the prior estimates. As pointed out by Rödenbeck (2005), the minimization determines the large-scale gradient in the initial iterations, while in subsequent iterations fine-scale tuning is performed to capture the optimum. By artificially limiting the number of iterations in experiments AO–CO, the ability of 4D-VAR to make small-scale adjustments is hindered, which manifests itself clearly in experiment CO (panel co in Figs. 3, 4 and 5).

For EnSRF, the degradation is attributable to sampling error caused by the limited ensemble size. This reduces the estimation accuracy (both flux estimates and their uncertainties) and makes the filter sensitive to the observational density. Note that a Schur-based localization scheme was implemented for EnSRF (see Appendix A3). Since the localization length scale is dependent on the ensemble size, when an ensemble size of 1000 was used (experiments A–C, or AR–CR), a long localization length scale of  $90\text{--}120 [L]$  can be used. The localization length scale is determined subjectively based on sensitivity tests, and hence a range of values (i.e.,  $90\text{--}120 [L]$ ) is acceptable within which the EnSRF estimates are not contaminated by spurious noise. Reducing the ensemble size to 100 requires the use of a shorter localization length scale. It was found beneficial to have different length scales for the different observational networks, namely  $10\text{--}30 [L]$  for the REF network and  $45\text{--}60 [L]$  for the sparser networks. Specifying a longer localization length scale than  $30 [L]$  for the REF network led to a divergence of the EnSRF system. In this case, the spurious noise in the ensemble outweighs the positive impact of the observations. The complex interplay between the ensemble size and the observational density makes it difficult to identify a mathematical or

physical basis for selecting an appropriate localization length scale. We refer the reader to Chatterjee et al. (2012), as well as the sensitivity tests presented therein, for a more detailed discussion of the role of localization for the CO<sub>2</sub> source–sink estimation problem.

Furthermore, when operational constraints are imposed, the posterior flux uncertainty estimates obtained using both EnSRF and 4D-VAR are unable to reproduce those from BIM. The degree to which posterior uncertainties from both DA approaches over/underestimate the BIM uncertainties for individual locations and times increases. This is clearly evident in Figs. 3 and 5, panels ao–co. Averaged over all time periods and locations, the posterior uncertainty estimates from EnSRF are still closer to the BIM uncertainty estimates, being within 10 % of the averaged BIM uncertainties for experiments AO, BO, and CO, whereas the average 4D-VAR uncertainties underestimate the BIM uncertainties by up to 25 %. The degradation in the uncertainty estimates provided by 4D-VAR and EnSRF is due to different reasons. While for EnSRF the large sampling error plays a dominant role, for 4D-VAR the perturbations in the Monte Carlo technique are unable to capture the true range of the posterior uncertainties. The impact for 4D-VAR is accentuated for experiments BO and CO, where the sparse network exacerbates the need for more iterations to reach convergence.

An important caveat here is that the results for both the DA approaches could potentially be improved through further tuning of each algorithm. For example, the implementation of more sophisticated algorithms to precondition and obtain faster convergence, or stronger localization schemes to dampen the spurious noise in the ensemble members, might provide slightly different responses and reduce the error incurred due to the numerical approximations. In spite of having state-of-the-art algorithms, however, once the underlying numerical approximations come into play, (a) EnSRF fails to reproduce the BIM best estimates, with the EnSRF performance decreasing as the observational network becomes sparser and more heterogeneous, and (b) 4D-VAR also fails to reproduce the BIM best estimates when the observational network is heterogeneous but still performs better than EnSRF. The better performance of 4D-VAR in terms of the flux estimates is offset by the fact that EnSRF provides more realistic uncertainty bounds on the recovered flux estimates.

The DA approaches are particularly sensitive to the information flow from the observations because of the lack of an explicit dynamical model. As discussed in Sect. 1, a dynamical model adds to the information content of the system, albeit at the cost of additional model errors that must be taken into account. Identifying/developing an appropriate dynamical model relevant to the CO<sub>2</sub> flux estimation problem, and subsequently repeating experiments such as those presented here, may further inform the assessment of the interplay between the operational constraints and the observational network.

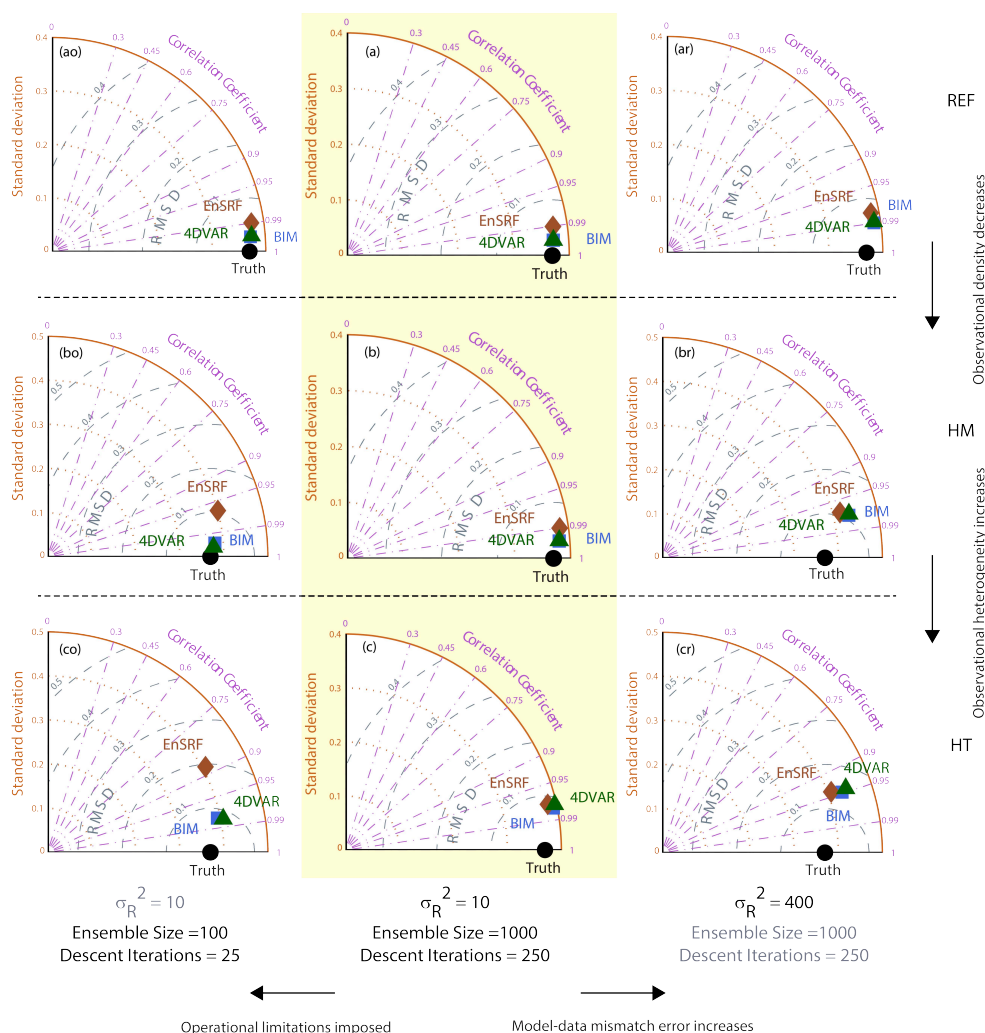
### 3.4 Examining results at aggregated spatial and temporal scales

In the previous sections, the performance of the DA and the BIM approaches were analyzed and reported at the native estimation scales (both in space and time). In contemporary CO<sub>2</sub> inversion studies, the fine-scale estimates and their uncertainties are typically averaged a posteriori in space and/or in time to coarser scales (e.g., daily grid-scale fluxes averaged to monthly continental scales) for additional interpretation.

To represent this process, the domain used here is divided into two subregions, 1–150 [L] and 151–300 [L], and the true flux and the flux estimates are aggregated over each of these areas and examined across time. This is qualitatively analogous to aggregating fluxes a posteriori to “large regions” (e.g., biomes, continents) within the inversion domain. Because the true flux differs between these two subregions, and these differences themselves vary in time, it is possible to examine the ability of the various approaches to capture these spatiotemporal variations. Figure 6 presents the comparison at the aggregated scales in the form of a Taylor diagram. For all the 9 experiments, 4D-VAR is able to match the temporal variation of the spatially aggregated BIM estimates better than EnSRF. Even when the number of descent iterations is reduced (Fig. 6ao–co), the differences between the BIM and the 4D-VAR best estimates are negligible at the examined spatially aggregated scales. Comparing Figs. 4co and 6co demonstrates that the differences observed at fine scales are substantially reduced when aggregating the estimates to a coarser resolution.

Because there is no explicit dynamical model to evolve the information between assimilation time periods, the EnSRF estimates are always contaminated with small-scale sampling errors, but these errors partially cancel out when the estimates are spatially aggregated. Overall, the EnSRF estimates still, however, exhibit more spurious variability relative to the 4D-VAR estimates. Especially when the number of ensemble members is reduced (Fig. 6ao–co), both the sampling error as well as the observational density and heterogeneity start to play a role, leading to less reliable estimates at aggregated scales. For example, in Fig. 6ao and 6co, the CC between the spatially aggregated EnSRF and the true flux estimates drops from 0.99 to 0.90, and the RMSD increases from 0.03 to 0.21 [ $ML^{-1}T^{-1}$ ]. The standard deviation of these spatially aggregated EnSRF estimates also increases from 0.37 to 0.40 [ $ML^{-1}T^{-1}$ ], leading to an overestimation relative to the true flux, which has a standard deviation of 0.36 [ $ML^{-1}T^{-1}$ ].

Analysis at aggregated scales demonstrates that when operational constraints are not imposed, both the DA approaches provide aggregated estimates that are close to the aggregated BIM estimates. Even when operational constraints are imposed, the aggregated flux estimates for EnSRF and 4D-VAR (Fig. 6co) have substantially higher CC



**Fig. 6.** Performance of BIM, EnSRF, and 4D-VAR at spatially aggregated scales for the different experiments outlined in Table 1. The details of the plot are as described in the caption of Fig. 4.

and lower RMSD than the corresponding estimates at the fine scale (Fig. 4co). This is encouraging from the perspective of a real CO<sub>2</sub> flux estimation problem, as it implies that the DA approaches may provide reliable flux estimates at aggregated scales, even under circumstances when their performance at fine scales is compromised.

#### 4 Discussion

Overall, the choice between 4D-VAR and EnSRF approaches for the CO<sub>2</sub> flux estimation problem should be based on the carbon science questions being targeted, as well as the tradeoff between the impact of incomplete convergence of the minimization algorithm (for 4D-VAR) and the impact of sampling error (for EnSRF) on the estimated fluxes and their uncertainties.

When an adjoint for the atmospheric transport model is available, and if the best estimates are the primary target, 4D-VAR outperforms EnSRF, especially when the measurement network is heterogeneous. With a small number of iterations, 4D-VAR may not converge to the minimum, but it still reliably captures the majority of the large-scale features, with the final estimates close to those expected from a BIM solution. The main disadvantages of the 4D-VAR approach, on the other hand, are its more cumbersome implementation and its inability to provide explicit estimates of the analysis error, which can instead be obtained indirectly via the Monte Carlo technique. In addition, the 4D-VAR approach may potentially be more advantageous due to its ability to more easily account for correlated model–data mismatch errors (O. Talagrand, personal communication, 2012). Although this specific issue has not been examined in this study, it is worth keeping in mind with the increasing use of satellite-based CO<sub>2</sub> measurements. Recent work by Brankart et al. (2009)

has demonstrated techniques to cope with such correlations in an ensemble filter setting as well. Such techniques, however, are numerically efficient only for certain types of error correlation structures.

In the absence of correlated model–data mismatch errors, however, the serial EnSRF is easier to implement than a 4D-VAR system, and does not require the development and maintenance of an adjoint model. Due to restrictions on the size of the ensemble, however, it is necessary to adapt and tune ancillary algorithms such as localization and inflation, which improve the ensemble approximations to the full-rank Kalman filter. Unlike adaptive inflation, the choices to be made in implementing a localization scheme remain highly subjective. Experiments in this study show, for example, that the localization length scale is dependent on both the ensemble size and the observational density. Will increasing volumes of observations push us towards specifying shorter localization length scales? If so, what is the limit beyond which decreasing the localization length scale may actually degrade the analysis? It is necessary to identify more rigorously a basis for selecting the localization parameters (e.g., Anderson, 2012) or adaptive approaches that may be less sensitive to variations in the observational network.

One critical requirement of an inverse problem is to obtain reliable second-order statistical moments for the estimated system states. Even though we have demonstrated that posterior error statistics can be obtained for both EnSRF (directly) and 4D-VAR (indirectly via a Monte Carlo technique), the EnSRF uncertainty estimates reproduce the BIM uncertainty estimates consistently better. The indirect techniques necessary for obtaining posterior error statistics for 4D-VAR has important caveats, as demonstrated by the large over/underestimation of uncertainties for individual flux locations and times, as well as the underestimation of the posterior uncertainties relative to BIM when averaged across all locations and times. Therefore, EnSRF is more desirable for attribution purposes, wherein source–sink estimates with realistic confidence bounds can be used to gain a better understanding of the mechanistic processes driving the carbon cycle or to reconcile estimates from top-down and bottom-up approaches.

With both 4D-VAR and EnSRF, there is a direct trade-off between computational savings and estimation accuracy, which is intensified when solving an underdetermined problem with a heterogeneous observational network. For large-scale flux estimation problems, operational constraints will always exist, as will scarce and inconsistent observations, transport model biases and uncertainties (thus further limiting the use of available observations), etc. The HT scheme with a limited number of ensemble members/descent iterations (panel c in Figs. 3, 4 and 5) serves as the closest analogue to a real inversion problem. Even if we account for the increase in remote-sensing measurements of CO<sub>2</sub>, the observational network is going to be a complex hybrid between the REF and the HT scheme. In this scenario, the accuracy

and precision of either of the DA approaches will be compromised relative to BIM, as demonstrated in this study.

## 5 Conclusions

We present a comparative assessment of two advanced DA approaches with the BIM approach for the atmospheric CO<sub>2</sub> inversion problem. The performance of the DA approaches is found to depend on a complex interplay between the underlying numerical approximations and the information available from the observations. Overall, the 4D-VAR scheme is found to be more robust for obtaining the best estimates, while the EnSRF scheme has the advantage of providing direct estimates of analysis error that are more realistic than those operationally feasible for 4D-VAR. The relative performance of 4D-VAR and EnSRF best estimates, when a large and homogeneous set of observations is available, is consistent with the conclusions obtained from intercomparison studies carried out by other DA communities. The sensitivity of the approaches to the observational scheme in the absence of an explicit dynamical model, and specifically for solving an underdetermined inverse problem, however, had not previously been thoroughly explored. Beyond CO<sub>2</sub> source–sink estimation problems, the conclusions of this study are therefore also relevant to other DA problems where a dynamical model is lacking.

The sensitivity experiments demonstrate that when a large number of ensemble members or descent iterations is specified, the best estimates obtained from state-of-the-art implementations of the 4D-VAR and the EnSRF approaches are similar to those from BIM, irrespective of the observational characteristics. Even under these optimal conditions, however, the uncertainty estimates from 4D-VAR are unable to reproduce those from BIM. When operational constraints are imposed, both the characteristics of the observational network and the numerical approximations play a greater role in differentiating the performance of the two DA approaches. Because such operational constraints are always present for real CO<sub>2</sub> source–sink estimation problems, the choice of an approach should be based on (a) the carbon science questions being targeted (i.e., whether the science calls only for best estimates of fluxes, or for flux estimates with realistic uncertainties) and (b) the inversion conditions under which they are being applied (i.e., characteristics of the observational network, such as data density and heterogeneity).

Recently, 4D-VAR and EnSRF have begun to influence each other's development, and EnVar DA approaches (Lorenc, 2013) may take center stage for solving the CO<sub>2</sub> flux problem in the future. In addition, the emerging focus on the assimilation of satellite CO<sub>2</sub> observations will require additional intercomparisons that go beyond the scope of the work presented here, including the consideration of temporal error correlations in the atmospheric CO<sub>2</sub> observations.



## Appendix A

### Estimation methods for solving the inverse problem

#### A1 Batch inverse modeling (BIM)

In the BIM approach (e.g., Enting, 2002), the analytical solution for the a posteriori estimate and the associated covariances of the objective function (Eq. 1) are given by

$$\hat{s}^a = s^b + \mathbf{K} \left( z - \mathbf{H} s^b \right), \quad (\text{A1})$$

$$\mathbf{Q}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{Q}^b, \quad (\text{A2})$$

$$\mathbf{K} = \mathbf{Q}^b \mathbf{H}^T \left( \mathbf{H} \mathbf{Q}^b \mathbf{H}^T + \mathbf{R} \right)^{-1}. \quad (\text{A3})$$

where  $\hat{s}^a$  is the posterior best estimate of the state and  $\mathbf{Q}^a$  is the a posteriori covariance of the recovered best estimate. The diagonal elements of  $\mathbf{Q}^a$  represent the predicted error variance ( $\sigma_s^2$ ) of individual elements in  $\hat{s}^a$ . As stated in Sect. 1, for CO<sub>2</sub> inversion studies, the generation of the matrix  $\mathbf{H}$  requires an atmospheric transport model to be run either once per estimated element of the state vector, or once per observation. The large number of model runs ultimately makes the BIM approach computationally intractable for solving very large-scale problems.

#### A2 Four-dimensional variational (4D-VAR)

In 4D-VAR, the fluxes  $\hat{s}^a$  that minimize the objective function in Eq. (1) are sought iteratively by minimizing the misfit between a feasible state trajectory and the observations that are available over a given assimilation window. The overall approximation lies in the fact that the minimization can be stopped by artificially limiting the number of iterations or by requiring that the norm of the gradient decreases by a predefined amount during each iteration. Most minimization schemes rely on the availability of the gradient of the objective function with respect to the state (or control vector in 4D-VAR terms),

$$\nabla J(s) = \left( \mathbf{Q}^b \right)^{-1} \left[ s - s^b \right] + \mathbf{H}^T \mathbf{R}^{-1} [z - h(s)]. \quad (\text{A4})$$

Instead of analytically calculating the gradient, the adjoint of the forward transport model is used to compute the term  $\sigma_s^2$  directly, which is then added to  $\hat{s}$  in Eq. (A4) above. While a variety of minimization schemes may be used (e.g., conjugate gradient or BFGS, e.g., Nocedal and Wright, 2006), in this study, we use a Lanczos minimizer algorithm (e.g., Nocedal and Wright, 2006), which produces similar results as the conjugate-gradient technique in terms of the reduction of the cost function and gradient norm but converges substantially faster. Additionally, a new variable ( $\Xi$ ) is defined to precondition the minimization as

$$\Xi = \left( \mathbf{Q}^b \right)^{-\frac{1}{2}} s, \quad (\text{A5})$$

where  $\Xi$  now becomes the control variable with respect to which the objective function is minimized instead of  $s$  directly. The optimal preconditioning matrix to reduce the number of iterations required to solve the minimization problem is the inverse Hessian (Axelsson and Barker, 2001). For a real high-dimensional CO<sub>2</sub> inversion problem, however, the size and structure of the prior covariance matrix may constitute a significant impediment for calculating the inverse Hessian. Any algebraic manipulations, such as taking inverses or calculating the square roots for preconditioning purposes, become operationally cumbersome and thus matrix manipulation techniques (e.g., Yadav and Michalak, 2013) are necessary to sidestep these operational challenges.

The main caveat with the variational approaches, such as 4D-VAR, is that a direct estimate of the analysis error is not available (no clear analogue of Eq. A2). Mathematically this can be obtained from the inverse of the Hessian (e.g., Le Dimet et al., 2002; Rödenbeck, 2005; Meirink et al., 2008) but operational challenges restrict the calculation and storage of the Hessian for high-dimensional problems. Recent applications of 4D-VAR for NWP problems have shown that computationally efficient alternatives do exist (e.g., Cheng et al., 2010; Gejadze et al., 2013). Although some of these techniques are more suited for nonlinear problems (e.g., Gejadze et al., 2013), they retain potential applicability to the CO<sub>2</sub> flux estimation problem as well. In this study, we use a Monte Carlo technique (e.g., Chevallier et al., 2007) where both the observations and the prior are perturbed multiple times with the prespecified model–data mismatch error statistics and prior error statistics, respectively.

4D-VAR is implemented with successive overlapping windows to (a) account for the long residence times of CO<sub>2</sub> in the atmosphere and (b) to avoid the operational cost associated with calculating the inverse of the full prior covariance matrix  $\mathbf{Q}^b$  in Eqs. (A4) and (A5). In order to determine the length of the moving window, however, one needs to keep in mind two primary factors: (a) observations typically inform fluxes over a finite period of time (e.g., 4- to 6-month time frame based on the analysis reported in Bruhwiler et al., 2005), and (b) temporal correlation in the prior flux errors (e.g., for the prior land flux errors Chevallier et al. (2012) has suggested the temporal correlation to be strongly positive for lags <85 days and mildly positive for lags >275 days). Given a sufficiently long window the 4D-VAR approach using a moving window will emulate the 4D-VAR approach using a single window, while at the same time providing substantial computational savings. In the absence of a dynamical model, this implementation of 4D-VAR becomes similar to the FGAT-3DVAR (Massart et al., 2010) variant occasionally used within the NWP community.

#### A3 Ensemble square root filter (EnSRF)

In the ensemble filter approach, the key innovation is to work in a reduced subspace of the error covariance matrices.

Observations are assimilated to update the ensemble representation of the error covariance matrices. The optimal analysis states and an estimate of the analyses error are determined in a similar fashion to Eqs. (A1) and (A2), but the calculation of  $\mathbf{Q}^b \mathbf{H}^T$  and  $\mathbf{H} \mathbf{Q}^b \mathbf{H}^T$  is approximated by running transport model using the ensemble members directly:

$$\mathbf{H} \mathbf{Q}^b \mathbf{H}^T \approx \frac{1}{N-1} [h(s')][h(s')]^T, \quad (\text{A6})$$

$$\mathbf{Q}^b \mathbf{H}^T \approx \frac{1}{N-1} (s') [h(s')]^T. \quad (\text{A7})$$

While several variants of the ensemble filter approach exist, here we have used a serial ensemble square root filter (Whitaker and Hamill, 2002) implemented in a fixed-lag smoother form (e.g., Whitaker and Compo, 2002; Chatterjee et al., 2012). Similarly to an ensemble square root filter, the ensemble smoother uses Monte Carlo estimates of the error covariances to compute a Kalman smoother gain matrix. This is applied iteratively to a time series of observations, where the analysis at the first time step is equivalent to an ensemble square root filter analysis as it only utilizes observations taken up to and including the analysis time. All subsequent time steps utilize observations taken a number of observing times past the analysis time. The localization scheme (i.e., to cut down spurious noise in the ensemble members) is based on Houtekamer and Mitchell (2001) using a fifth-order Gaspari–Cohn function (Gaspari and Cohn, 1999), while the adaptive inflation algorithm (i.e., to counter spurious variance deficiency among the ensemble members) is based on Anderson (2009). Implementation of these algorithms within an ensemble smoother framework is described in further detail in Chatterjee et al. (2012). Despite these ancillary algorithms, the overall implementation of EnSRF is quite simple and computationally efficient.

#### A4 Sensitivity tests and additional approaches considered

The setup of 4D-VAR (with overlapping time windows) and EnSRF (expressed as a fixed-lag smoother) reflect state-of-the-art implementations of these two DA approaches, keeping in mind the nature of the atmospheric CO<sub>2</sub> process, the fact that CO<sub>2</sub> observations only provide significant information about fluxes over a finite preceding time window, and the operational constraints for estimating CO<sub>2</sub> fluxes at high spatiotemporal scales (e.g., spatial  $\sim 1^\circ$ , temporal  $\sim$ daily). Note that we have implemented both approaches with a single long window as a sensitivity test. Other than an increase in the computational cost, the conclusions regarding the performance of the approaches relative to the BIM approach are the same as those reported in the manuscript. We have also tested other varieties of the ensemble filter (EnKF with perturbed observations (Evensen, 2003) and the variational approach (PSAS; Courtier, 1997) and found that the overall conclusions from the presented experiments remain con-

sistent across these algorithmic choices. We encourage the reader to look at Lorenc (2003) and Nichols (2010) (and references therein) for a more detailed discussion on these DA approaches.

**Acknowledgements.** The authors would like to thank Jeffrey Anderson, Derek Posselt, Peter Rayner, and Chris Snyder for providing valuable suggestions and feedback on the intercomparison analyses as well as the manuscript. Comments from the two anonymous reviewers also helped improve the quality of the manuscript. This work was supported through a (a) National Aeronautics and Space Administration Earth System Science Fellowship for Abhishek Chatterjee, grant no. NNX09AO10H, and the (b) NOAA Climate and Global Change Postdoctoral Fellowship Program administered by the University Corporation for Atmospheric Research.

Edited by: M. Heimann

#### References

- Anderson, J. L.: Spatially and temporally varying adaptive covariance inflation for ensemble filters, *Tellus A – Dyn. Meteorol. Oceanogr.*, 61, 72–83, doi:10.1111/j.1600-0870.2008.00361.x, 2009.
- Anderson, J. L.: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation, *Mon. Weather Rev.*, 140, 2359–2371, doi:10.1175/mwr-d-11-00013.1, 2012.
- Axelsson, O. and Barker, V. A.: Finite-Element Solution of Boundary-value Problems. Theory and Computation, vol. 35 of *Classics in Applied Mathematics*, SIAM, Philadelphia, PA, 432 pp. (Reprint of the 1984 original), 2001.
- Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO<sub>2</sub>, *Tellus Series B-Chemical and Physical Meteorology*, 58, 359–365, doi:10.1111/j.1600-0889.2006.00218.x, 2006.
- Bauer, P., Lopez, P., Benedetti, A., Salmond, D., and Moreau, E.: Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1 D-Var, *Q. J. Roy. Meteorol. Soc.*, 132, 2307–2332, doi:10.1256/qj.05.189, 2006.
- Bengtsson, T., Snyder, C., and Nychka, D.: Toward a nonlinear ensemble filter for high-dimensional systems, *J. Geophys. Res.*, 108, 8775–8785, doi:10.1029/2002JD002900, 2003.
- Brankart, J.-M., Ubelmann, C., Testut, C.-E., Cosme, E., Brasseur, P., and Verron, J.: Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry, *Mon. Weather Rev.*, 137, 1908–1927, doi:10.1175/2008MWR2693.1, 2009.
- Bruhwyler, L. M. P., Michalak, A. M., Peters, W., Baker, D. F., and Tans, P.: An improved Kalman Smoother for atmospheric inversions, *Atmos. Chem. Phys.*, 5, 2691–2702, doi:10.5194/acp-5-2691-2005, 2005.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments, *Mon. Weather Rev.*, 138, 1550–1566, 2010a.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the

- Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations, *Mon. Weather Rev.*, 138, 1567–1586, 2010b.
- Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., Tang, Y.: Predicting Air Quality: Improvements through advanced methods to integrate models and measurements, *J. Comput. Phys.*, 227, 3540–3571, doi:10.1016/j.jcp.2007.02.024, 2008.
- Caya, A., Sun, J., and Snyder, C.: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation, *Mon. Weather Rev.*, 133, 3081–3094, 2005.
- Chatterjee, A., Michalak, A. M., Anderson, J. L., Mueller, K. L., and Yadav, V.: Towards reliable ensemble Kalman filter estimates of CO<sub>2</sub> fluxes, *J. Geophys. Res.-Atmos.*, 117, D22306, doi:10.1029/2012JD018176, 2012.
- Cheng, H. Y., Jardak, M., Alexe, M., and Sandu, A.: A hybrid approach to estimating error covariances in variational data assimilation, *Tellus A – Dyn. Meteorol. Oceanogr.*, 62A, 288–297, doi:10.1111/j.1600-0870.2010.00442.x, 2010.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Breon, F.-M., Chedin, A., and Ciais, P.: Inferring CO<sub>2</sub> sources and sinks from satellite observations: Method and application to TOVS data, *J. Geophys. Res.-Atmos.*, 110, D24309, doi:10.1029/2005JD006390, 2005.
- Chevallier, F., Breon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO<sub>2</sub> sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.-Atmos.*, 112, D09307, doi:10.1029/2006JD007375, 2007.
- Chevallier, F., Wang, T., Ciais, P., Maignan, F., Bocquet, M., Altaf Arain, M., Cescatti, A., Chen, J., Dolman, A. J., Law, B. E., Margolis, H. A., Montagnani, L., and Moors, E. J.: What eddy-covariance measurements tell us about prior land flux errors in CO<sub>2</sub>-flux inversion schemes, *Global Biogeochem. Cy.*, 26, GB1021, doi:10.1029/2010GB003974, 2012.
- Courtier, P.: Dual formulation of four-dimensional variational assimilation, *Q. J. Roy. Meteorol. Soc.*, 123, 2449–2461, 1997.
- Elbern, H., Strunk, A., and Nieradzik, L.: Inverse Modeling and Combined State-Source Estimation for Chemical Weather, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., Menard, R., 491–515, Springer-Verlag Berlin, 2010.
- Enting, I. G.: *Inverse Problems in Atmospheric Constituent Transport*. Atmospheric and Space Science Series, 392 pp., Cambridge University Press, Cambridge, 2002.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, 2003.
- Eyre, J. R., Kelly, G. A., McNally, A. P., Andersson, E., and Persson, A.: Assimilation of TOVS radiance information through one-dimensional variational analysis, *Q. J. Roy. Meteorol. Soc.*, 119, 1427–1463, doi:10.1002/qj.49711951411, 1993.
- Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: 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, 2619–2633, doi:10.5194/acp-9-2619-2009, 2009.
- Fertig, E. J., Harlim, J., and Hunt, B. R.: A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: perfect model simulations with Lorenz-96, *Tellus A – Dyn. Meteorol. Oceanogr.*, 59, 96–100, 2007.
- Furrer, R. and Bengtsson, T.: Estimation of highdimensional prior and posterior covariance matrices in Kalman filter variants, *J. Multivar. Anal.*, 98, 227–255, doi:10.1016/j.jmva.2006.08.003, 2007.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Q. J. Roy. Meteorol. Soc.*, 125, 723–757, doi:10.1256/smsqj.55416, 1999.
- Gejadze, I. Yu., Shutyaev, V., and Dimet, F.-X. L.: Analysis error covariance versus posterior covariance in variational data assimilation, *Q. J. R. Meteorol. Soc.*, 139, 1826–1841, doi:10.1002/qj.2070, 2013.
- Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: 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, 2012.
- Haines, K.: Ocean Data Assimilation, in *Data Assimilation, Making Sense of Observations*, edited by: W. Lahoz, B. Khattatov, R. Menard, 517–547, Springer-Verlag Berlin, 2010.
- Houser, P. R., De Lannoy, G. J. M., and Walker, J. P.: Land Surface Data Assimilation, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., and Menard, R., 549–597, Springer-Verlag Berlin, 2010.
- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2, 2001.
- Janiskova, M., Lopez, P., and Bauer, P.: Experimental 1D+4D-Var assimilation of CloudSat observations, *Q. J. Roy. Meteorol. Soc.*, 138, 1196–1220, doi:10.1002/qj.988, 2012.
- Jardak, M., Navon, I. M., and Zupanski, M.: Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation, *Int. J. Num. Meth. Fluid.*, 62, 374–402, doi:10.1002/fld.2020, 2010.
- Kalnay, E., H. Li, Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, *Tellus A – Dyn. Meteorol. Oceanogr.*, 59, 758–773, 2007.
- Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.-Atmos.*, 106, 4703–4715, doi:10.1029/2000JD900581, 2001.
- Kang, J.-S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, *J. Geophys. Res.-Atmos.*, 117, D24101, doi:10.1029/2012JD018259, 2012.
- Kuppel, S., Chevallier, F., and Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems, *Geosci. Model Dev.*, 6, 45–55, doi:10.5194/gmd-6-45-2013, 2013.
- Lahoz, W. and Errera, Q.: Constituent Assimilation, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., and Menard, R., 449–489, Springer-Verlag Berlin, 2010.
- Le Dimet, F.-X., Navon, I. M., and Daescu, D. N.: Second-Order Information in Data Assimilation,

- Mon. Weather Rev., 130, 629–648. doi:10.1175/1520-0493(2002)130<0629:SOIIDA>2.0.CO;2, 2002.
- Liu, J., Fertig, E. J., Li, H., Kalnay, E., Hunt, B. R., Kostelich, E. J., Szunyogh, I., and Todling, R.: Comparison between Local Ensemble Transform Kalman Filter and PSAS in the NASA finite volume GCM – perfect model experiments, *Nonlin. Processes Geophys.*, 15, 645–659, doi:10.5194/npg-15-645-2008, 2008.
- Liu, Z.-Q. and Rabier, F.: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study, *Q. J. Roy. Meteorol. Soc.*, 128, 1367–1386, 2002.
- Lokupitiya, R. S., Zupanski, D., Denning, A. S., Kawa, S. R., Gurney, K. R., and Zupanski, M.: Estimation of global CO<sub>2</sub> fluxes at regional scale using the maximum likelihood ensemble filter, *J. Geophys. Res.-Atmos.*, 113, D20110, doi:10.1029/2007JD009679, 2008.
- Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var, *Q. J. Roy. Meteorol. Soc.*, 129, 3183–3203, 2003.
- Lorenc, A. C.: Recommended Nomenclature for EnVar Data Assimilation Methods, available online at [http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01\\_Lorenc\\_Andrew\\_EnVar\\_nomenclature.pdf](http://www.wcrp-climate.org/WGNE/BlueBook/2013/individual-articles/01_Lorenc_Andrew_EnVar_nomenclature.pdf) (last access: 10 October 2013), 2013.
- Marécal, V. and Mahfouf, J. F.: Four-dimensional variational assimilation of total column water vapor in rainy areas, *Mon. Weather Rev.*, 130, 43–58, doi:10.1175/1520-0493(2002)130<0043:fdvaot>2.0.CO;2, 2002.
- Massart, S., Pajot, B., Piacentini, A., and Pannekoucke, O.: On the Merits of Using a 3D-FGAT Assimilation Scheme with an Outer Loop for Atmospheric Situations Governed by Transport, *Mon. Wea. Rev.*, 138, 4509–4522. doi:10.1175/2010MWR3237.1, 2010.
- Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, *Atmos. Chem. Phys.*, 8, 6341–6353, doi:10.5194/acp-8-6341-2008, 2008.
- Miyazaki, K., Maki, T., Patra, P., and Nakazawa, T.: 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.-Atmos.*, 116, D16306, doi:10.1029/2010JD015366, 2011.
- Nichols, D.: Mathematical Concepts of Data Assimilation, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., Menard, R., 13–40, Springer-Verlag Berlin, 2010.
- Nocedal, J. and Wright, S. J.: *Numerical Optimization*, 224–229, Springer Ser. Oper. Res., Springer-Verlag, Berlin, 2006.
- Ogata, A. and Banks, R. B.: A Solution of the Differential Equation of Longitudinal in Porous Media, US Geological Survey Professional Paper 411-A, available at: <http://pubs.usgs.gov/pp/0411a/report.pdf>, 1961.
- Park, S. K. and Kalnay, E.: Inverse three-dimensional variational data assimilation for an advection-diffusion problem: Impact of diffusion and hybrid application, *Geophys. Res. Lett.*, 31, L04102, doi:10.1029/2003GL018830, 2004.
- Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO<sub>2</sub> surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.-Atmos.*, 110, D24304, doi:10.1029/2005JD006157, 2005.
- Poli, P., Joiner, J., and Kursinski, E. R.: 1DVAR analysis of temperature and humidity using GPS radio occultation refractivity data, *J. Geophys. Res.-Atmos.*, 107, 4448, doi:10.1029/2001jd000935, 2002.
- Rayner, P. J.: The current state of carbon-cycle data assimilation, *Current Opinion in Environmental Sustainability*, 2, 289–296, doi:10.1016/j.cosust.2010.05.005, 2010.
- Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cy.*, 19, GB2026, doi:10.1029/2004GB002254, 2005.
- Reichle, R. H.: Data Assimilation Methods in the Earth Sciences, *Adv. Water Resour.*, 31, 1411–1418, doi:10.1016/j.advwatres.2008.01.001, 2008.
- Rödenbeck, C.: Estimating CO<sub>2</sub> sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport. Technical Report 6, Max Planck Institute for Biogeochemistry, Jena, 2005.
- Runkel, R. L.: Solution of the advection-dispersion equation: continuous load of finite duration, *J. Environ. Eng.*, 122, 830–832, 1996.
- Swinbank, R.: Numerical Weather Prediction, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., Menard, R., 381–407, Springer-Verlag Berlin, 2010.
- Talagrand, O.: Variational Assimilation, in *Data Assimilation, Making Sense of Observations*, edited by: Lahoz, W., Khattatov, B., Menard, R., 41–67, Springer-Verlag Berlin, 2010.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, 2001.
- Whitaker, J. S. and Compo, G. P.: An ensemble Kalman smoother for reanalysis. *Proc. Symp. on Observations, Data Assimilation and Probabilistic Prediction*, Orlando, FL, Amer. Meteor. Soc., 144–147, 2002.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, 130, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002.
- Whitaker, J. S., Compo, G. P., and Thepaut, J. N.: A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations, *Mon. Weather Rev.*, 137, 1991–1999, 2009.
- Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, *J. Geophys. Res.-Atmos.*, 113, D20310, doi:10.1029/2008JD009991, 2008.
- Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation, *Geosci. Model Dev.*, 6, 583–590, doi:10.5194/gmd-6-583-2013, 2013.
- Zhang, M., Zhang, F. Q., Huang, X. Y., and Zhang, X.: Inter-comparison of an Ensemble Kalman Filter with Three- and Four-Dimensional Variational Data Assimilation Methods in a Limited-Area Model over the Month of June 2003, *Mon. Weather Rev.*, 139, 566–572, 2011.

Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W., and Corbin, K. D.: Carbon flux bias estimation employing maximum likelihood ensemble filter (MLEF), *J. Geophys. Res.-Atmos.*, 112, D17107, doi:10.1029/2006JD008371, 2007.