

Technical Note: Propagating correlations in atmospheric inversions using different Kalman update smoothers

J. Tang^{1,2} and Q. Zhuang^{1,2,3}

¹Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, IN, USA

²Purdue Climate Change Research Center, West Lafayette, IN, USA

³Department of Agronomy, Purdue University, West Lafayette, IN, USA

Received: 27 July 2010 – Published in Atmos. Chem. Phys. Discuss.: 16 August 2010

Revised: 20 December 2010 – Accepted: 24 January 2011 – Published: 2 February 2011

Abstract. The scheme to propagate correlations between on-line and off-line state variables in atmospheric inversions using the fixed-lag Kalman smoother proposed in Bruhwiler et al. (2005) is explained as a process to impose a balanced constraint on the on-line state variables. It is then extended to the fixed-lag ensemble square root Kalman smoother and fixed-lag square root sigma-point Kalman smoother, allowing us to treat nonlinear observation operators easily. Further, to constrain the posterior fluxes within their feasible ranges, the constrained fixed-lag Kalman smoother is presented and the variable transform technique is proposed for the other two smoothers. Comparisons between various methods and observational data are conducted using a synthetic inversion of atmospheric CH₄ fluxes. The results indicate that our developed methods are good alternatives to existing methods for conducting sequential inversion of atmospheric trace gases. It is also shown that the benefit to include the correlations between on-line and off-line state variables is case dependent.

1 Introduction

Closing the budget of various greenhouse gases, such as CO₂, CH₄ and N₂O, has been an important task in our understanding of the human-induced climate change. A good knowledge of the different sources and sinks of these greenhouse gases is invaluable to mitigate or avoid the environmental risk due to the increasing atmospheric content of those trace gases. Atmospheric inversion modeling plays an

important role in quantifying the sources and sinks of various trace gases (Enting, 2002). It involves the comparison of forward model simulations from atmospheric transport models using prior sources and sinks with the spatiotemporally discrete observations. The prior sources and sinks are optimized by minimizing a cost function defined by the distances between the forward model simulation and observations (e.g., Gurney et al., 2002).

The inversion problem is usually formulated in the form of Bayesian inference (Tarantola, 2005). In the Bayesian theorem, the fluxes and their associated error characteristics that are known as the prior ($\Pr(s)$) and the observations and their error characteristics that define the likelihood function ($\Pr(o | s)$) are used to obtain the posterior fluxes ($\Pr(s | o)$) as

$$\Pr(s | o) = C^{-1} \Pr(s) \Pr(o | s) \quad (1)$$

where C is some constant to normalize the posterior distributions. The Eq. (1) can be solved in the “batch” mode (Gelb, 1974) that treats all observations simultaneously and infer all the sources and sinks at the same time. This works efficiently when the number of observations and the number of fluxes involved are small. However, as it is often the case, the “batch” mode is cumbersome to implement when more observations become available and higher spatiotemporal resolution fluxes are to be inverted. Other methods, e.g. the fixed-lag Kalman smoother (KS), adjoint-based nonlinear optimization method, and the fixed-lag ensemble Kalman smoother, have been used to overcome the difficulties found in the “batch” inversion technique (e.g., Hartley and Prinn, 1993; Houweling et al., 1999; Bruhwiler et al., 2005; Peters et al., 2005). These various methods have proven to be efficient in solving the well-configured problems in their studies.



Correspondence to: J. Tang
(tang16@purdue.edu)

Among these methods, the fixed-lag KS has been used in a number of studies (e.g., Hartley and Prinn, 1993; HaasLaursen et al., 1996; Bruhwiler et al., 2005; Michalak, 2008), because of its easiness to implement and its efficiency to assimilate observations sequentially. The fixed-lag KS was developed based on the fact that the signal of a certain flux at a given time period in a given region will be fully blended into the background field, such that no useful information will be extracted from the observations after a sufficient time of transport of the flux (Hartley and Prinn, 1993; Bruhwiler et al., 2005).

Bruhwiler et al. (2005) noted that the conventional implementation of fixed-lag KS treats the fluxes that are no longer estimated as if they are exact, even with known uncertainties. They recommended to use correlations between the on-line state variables (fluxes that are still under optimization) and the off-line state variables (fluxes that are no longer updated) to improve the posterior fluxes. They showed by propagating such correlations with the fixed-lag KS, superior results to those from inversions without considering such correlations can be obtained in their CO₂ inversion experiments. Bruhwiler et al. (2005) derived the correlation propagating scheme as a correction to the prior covariance of the on-line state variables before assimilating the atmospheric observations.

In this note, we put their derivation of the correlation propagating scheme into a more general context, by showing that the correlations between the on-line and off-line state variables effectively act as a special type of observational constraint without measurement error. This makes it possible to extend the correlation propagation scheme to the ensemble methods directly. We then extend the fixed-lag KS in Bruhwiler et al. (2005) to two ensemble-based methods, the fixed-lag ensemble square root Kalman smoother (ESRKS) and the fixed-lag square root central difference Kalman smoother (SRCDKS). We apply the new developments to an atmospheric inversion problem of CH₄ fluxes, and show the new developments are good alternatives to the fixed-lag KS in solving the atmospheric inversion problem. In addition, in our use of these three methods, we find that some of the inverted fluxes could have non-realistic values, i.e., either too large or even negative because of the ill-posedness of the inversion problem or insufficient constraints on the fluxes to be inverted compared to that in the linear batch inversion. Solutions to avoid such spurious inversions are provided with the three methods.

In Sect. 2.1, we introduce the inversion problem and its lagged-form. In Sect. 2.2, we show that the correlation propagation scheme between the off-line and online state variables is an effective way to impose a balanced constraint on the on-line state variables. We extend the development to the fixed-lag ESRKS in Sect. 2.3 and to the fixed-lag SRCDKS in Sect. 2.4. Techniques to impose interval constraint are presented in Sect. 2.5. The designation of the synthetic inversion experiment is described in Sects. 2.6 and 2.7. Results and

discussions are made in Sect. 3, followed by a conclusion of findings in Sect. 4.

2 Methods

2.1 The inversion problem and its lagged-form

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{v} \quad (2)$$

where \mathbf{z} is the vector of observations, \mathbf{s} is the vector of sinks and sources, \mathbf{H} is the sensitivity matrix that maps the fluxes into the measurement space, and \mathbf{v} is the uncertainty of the approximated observations $\mathbf{H}\mathbf{s}$ with respect to the real observation \mathbf{z} .

The aim of Bayesian inversion is to solve for \mathbf{s} in Eq. (2) using the Bayes theorem Eq. (1), by assuming variables \mathbf{z} , \mathbf{s} and \mathbf{v} as random variables with certain probability distributions.

In the lagged form, the forward equation Eq. (2) is

$$\mathbf{z}_J = [\mathbf{H}_{J,J} \mathbf{H}_{J,J-1} \cdots \mathbf{H}_{J,1}] \left[\mathbf{s}_J^T \mathbf{s}_{J-1}^T \cdots \mathbf{s}_1^T \right]^T + \mathbf{v} \quad (3)$$

$$= [\mathbf{H}_u \mathbf{H}_v] \left[\mathbf{s}_u^T \mathbf{s}_v^T \right]^T + \mathbf{v} \quad (4)$$

where \mathbf{s}_u is the vector of on-line state variables defined by fluxes that are still in estimation, from time J back to time $J-L+1$, and \mathbf{s}_v is the vector of off-line state variables defined by fluxes that are no longer updated, from time $J-L$ back to time 1. The observation operators \mathbf{H}_u and \mathbf{H}_v are defined accordingly for \mathbf{s}_u and \mathbf{s}_v .

2.2 A revisit of the Kalman smoother

The posterior distribution of the fluxes \mathbf{s}_u^+ is

$$p(\mathbf{s}_u^+ | \mathbf{s}_v^+, \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{s}_u^-, \mathbf{s}_v^+) p(\mathbf{s}_u^- | \mathbf{s}_v^+) p(\mathbf{s}_v^+)}{\int p(\mathbf{z}) d\mathbf{z}} \quad (5)$$

where superscript $-$ means the prior forecast, and superscript $+$ indicates the posterior inference.

Since \mathbf{s}_v^+ is no longer estimated, it holds that $p(\mathbf{s}_v^+)$ is constant, which leads to

$$p(\mathbf{s}_u^+ | \mathbf{s}_v^+, \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{s}_u^-, \mathbf{s}_v^+) p(\mathbf{s}_u^- | \mathbf{s}_v^+)}{\int p(\mathbf{z}) d\mathbf{z}} \quad (6)$$

For the special case when normal distributions are assumed for the state variables and observations, the posterior distribution of \mathbf{s}_u is

$$\begin{aligned} p(\mathbf{s}_u^+ | \mathbf{s}_v^+, \mathbf{z}) &= CN(\mathbf{z} - \mathbf{H}_u \mathbf{s}_u^- - \mathbf{H}_v \mathbf{s}_v^+, \mathbf{R}) \\ &\quad N\left(\left[\mathbf{s}_u^{+T} - \mathbf{s}_u^{-T}, \mathbf{s}_v^{+T} - \mathbf{s}_v^{+T}\right]^T, \mathbf{Q}\right) \\ &= CN(\mathbf{z} - \mathbf{H}_u \mathbf{s}_u^- - \mathbf{H}_v \mathbf{s}_v^+, \mathbf{R}) \\ &\quad \times N(\mathbf{s}_u^+ - \mathbf{s}_u^-, \mathbf{Q}_{aa}) \end{aligned} \quad (7)$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{uu}^- & \mathbf{Q}_{uv}^- \\ \mathbf{Q}_{vu}^- & \mathbf{Q}_{vv}^- \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{Q}_{aa}^{-1} & \mathbf{Q}_{ab}^{-1} \\ \mathbf{Q}_{ba}^{-1} & \mathbf{Q}_{bb}^{-1} \end{bmatrix} \quad (8)$$

$$\mathbf{Q}_{aa} = \mathbf{Q}_{uu}^- - \mathbf{Q}_{uv}^- \mathbf{Q}_{vv}^{-1} \mathbf{Q}_{uv}^{-T} \quad (9)$$

and $N(\mathbf{a}, \mathbf{B})$ means a multi-dimensional normal distribution with mean \mathbf{a} and covariance \mathbf{B} . C is a normalization constant. \mathbf{Q}_{uu} is the covariance of the on-line state variables, \mathbf{Q}_{vv} is the covariance of the off-line state variables, and \mathbf{Q}_{uv} is the correlation matrix between the on-line and off-line state variables. \mathbf{Q}_{bb} , and \mathbf{Q}_{ab} are defined exactly as the Eqs. (22–23) in Bruhwiler et al. (2005).

The Kalman update of s_u^+ is then

$$s_u^+ = s_u^- + \mathbf{K}_u (z - \mathbf{H}_u s_u^- - \mathbf{H}_v s_v^+) \quad (10)$$

where the Kalman gain is

$$\begin{aligned} \mathbf{K}_u &= \mathbf{P}_{uz} \mathbf{Q}_{zz}^{-1} \\ &= \mathbf{Q}_{aa} \mathbf{H}_u^T (\mathbf{R} + \mathbf{H}_u \mathbf{Q}_{aa} \mathbf{H}_u^T)^{-1} \end{aligned} \quad (11)$$

where \mathbf{P}_{uz} is the correlation matrix between the state variables s_u and the measurement z defined as the product $\mathbf{Q}_{aa} \mathbf{H}_u^T$, and \mathbf{Q}_{zz} is the total error covariance of the measurement with respect to the forecast defined as $(\mathbf{R} + \mathbf{H}_u \mathbf{Q}_{aa} \mathbf{H}_u^T)$.

For the posterior covariance \mathbf{Q}_{uu}^+ of s_u^+ , we require it as close as possible to that derived when both s_u and s_v are updated with z . Therefore, using the Kalman smoother update of the posterior covariance of the full state variables,

$$\begin{aligned} \mathbf{Q}_a &= \mathbf{Q} - \mathbf{K} \mathbf{H} \mathbf{Q} \\ &= \begin{bmatrix} \mathbf{Q}_{uu}^- & \mathbf{Q}_{uv}^- \\ \mathbf{Q}_{vu}^- & \mathbf{Q}_{vv}^- \end{bmatrix} - \mathbf{K} \mathbf{H} \begin{bmatrix} \mathbf{Q}_{uu}^- & \mathbf{Q}_{uv}^- \\ \mathbf{Q}_{vu}^- & \mathbf{Q}_{vv}^- \end{bmatrix} \end{aligned} \quad (12)$$

we obtain

$$\mathbf{Q}_{uu}^+ = \mathbf{Q}_{uu}^- - \mathbf{K}_1 \mathbf{H} \begin{bmatrix} \mathbf{Q}_{uu}^- \\ \mathbf{Q}_{vu}^- \end{bmatrix} = \mathbf{Q}_{uu}^- - \mathbf{K}_1 (\mathbf{P}_{uz}^s)^T \quad (13)$$

where

$$\mathbf{K}_1 = [\mathbf{Q}_{uu}^-, \mathbf{Q}_{uv}^-] \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{Q} \mathbf{H}^T)^{-1} \quad (14)$$

$\mathbf{H} \equiv [\mathbf{H}_u, \mathbf{H}_v]$, and \mathbf{P}_{uz}^s is the correlation between the full state variables $[s_u^T, s_v^T]^T$ and the measurement z .

It can be proved the above formula for \mathbf{Q}_{uu}^+ is equivalent to the formula given in Bruhwiler et al. (2005) (their Eq. 25).

An alternative derivation for the above formulae is available when viewing s_v as a special type of measurement. Define a measurement operator \mathbf{H}_{vu} , such that

$$s_v = \mathbf{H}_{vu} s_u + \epsilon \quad (15)$$

$$\mathbf{Q}_{vv} = \mathbf{H}_{vu} \mathbf{Q}_{uu} \mathbf{H}_{vu}^T \quad (16)$$

$$\mathbf{Q}_{uv} = \mathbf{Q}_{uu} \mathbf{H}_{vu}^T \quad (17)$$

where ϵ is a random variable with distribution $N(\mathbf{o}, \sqrt{\gamma} \mathbf{I})$, with $\gamma \approx 0$ (a small positive number). Therefore, using the Kalman update we have

Therefore, using the Kalman update we have

$$s_u^{-'} = s_u^- + \mathbf{Q}_{uv} (\mathbf{Q}_{vv} + \gamma \mathbf{I})^{-1} (s_v - \mathbf{H}_{vu} s_u^-) \quad (18)$$

Thence, in the limit $\gamma \approx 0$, $E(s_u^{-'}) = E(s_u^-)$. Further invoking the Kalman update of covariance, it can be shown that the covariance of $s_u^{-'}$ is defined by Eq. (9). Now, it is clear that the correlation between on-line and off-line state variables effectively acts as a balanced constraint on the on-line state variables. It indeed helps to reduce the background covariance before the measurement data are used to constrain the on-line state variables, but makes no update of the prior mean $E(s_u^-)$. The assimilation equation is still Eq. (10), and the final update of covariance matrix is Eq. (13). This alternative derivation more clearly shows that the covariance correction Eq. (9) is necessary in order to sufficiently constrain the on-line state variables. In addition, the new derivation provides a way to assimilate the unusual correlation information (which corrects the covariance of the on-line state variables) to improve the inversion.

2.3 Further development of the ensemble square root Kalman smoother

A detailed description of the ensemble square root Kalman filter can be found in Tippett et al. (2003). We here give only the steps needed in our study. According to Eq. (18), before assimilating the observations, an adjustment to the scaled ensemble perturbations from the evolution of the forecast model should first be carried out to assure that the correlation is properly accounted for. One possible implementation of such adjustment is

$$\mathbf{S}_{u,1}^- = \mathbf{S}_u^- - \mathbf{Q}_{uv} \mathbf{Q}_{vv}^{-1} \mathbf{S}_v \quad (19)$$

Then the ensemble mean is updated with Eq. (10), while the Kalman gain is computed as

$$\mathbf{K}_u = \mathbf{S}_{u,1}^- \mathbf{V}_1 (\mathbf{R} + \mathbf{V}_1^T \mathbf{V}_1)^{-1} \quad (20)$$

where $\mathbf{V}_1 \equiv (\mathbf{H}_u \mathbf{S}_{u,1}^-)^T$.

The scaled ensemble perturbations for next evolution cycle are

$$\mathbf{S}_u^+ = \mathbf{S}_u^- \mathbf{X}_2 \mathbf{U}_2 \quad (21)$$

where

$$\mathbf{X}_2 \mathbf{X}_2^T = \left[\mathbf{I} - \mathbf{V}_2 (\mathbf{V}_2^T \mathbf{V}_2 + \mathbf{R})^{-1} \mathbf{V}_2^T \right] \quad (22)$$

where $\mathbf{V}_2 \equiv (\mathbf{H} \mathbf{S}^-)^T$, with $\mathbf{H} = [\mathbf{H}_u, \mathbf{H}_v]$, $\mathbf{S}^- = [\mathbf{S}_u^{-T}, \mathbf{S}_v^{+T}]^T$ and \mathbf{U}_2 is an arbitrary $m \times m$ orthogonal matrix. \mathbf{X}_2 is the square root matrix of the above equation. The posterior covariance is

$$\mathbf{Q}_{uu}^+ = \mathbf{S}_u^+ \mathbf{S}_u^{+T} \quad (23)$$

2.4 Further development of the sigma-point square root central difference Kalman smoother

The fixed-lag sigma-point Kalman smoother deterministically samples a group of points in the state variable space for the ensemble simulations (Ambadan and Tang, 2009; Van der Merwe, 2004). It is derivative free in assimilating the measurements (Nørgaard et al., 1998). The sigma-point Kalman smoother is accurate up to second order in estimating mean and covariance (Julier and Uhlmann, 2004; Nørgaard et al., 1998). Among the several different choices, we use the square root form of the central difference Kalman smoother for this study. Also, we assume the measurement and model errors are additive, which greatly reduces the computation requirements, while resulting in little loss of accuracy compared to the complete forms presented elsewhere (e.g., Nørgaard et al., 1998). The SRCDKS uses a set of sigma points of size $2L_1 + 1$ derived from the distribution defined by Eq. (7) to approximate the dynamic system,

$$\mathbf{S}_u^- = \begin{cases} \bar{s}_u^-, & i = 0 \\ \bar{s}_u^- + h\Theta_{a,i}, & i = 1, \dots, L_1 \\ \bar{s}_u^- - h\Theta_{a,i}, & i = L_1 + 1, \dots, 2L_1 \end{cases} \quad (24)$$

where h is a scaling parameter of value $\sqrt{3}$ to make the approximation with second order accuracy, and

$$\Theta_a = \left[\sqrt{\mathbf{Q}_{aa} q(L_1)}, \sqrt{\mathbf{Q}_{aa} q(L_1)} \right] \quad (25)$$

where $q(L_1)$ means taking L_1 columns of $\sqrt{\mathbf{Q}_{aa}}$ with some specific criterion, which is detailed in the end of this section.

The predicted mean of the measurement is then

$$\bar{z} = \sum_{i=2L_1}^{i=0} w_i^{(m)} z_{u,i} \quad (26)$$

$$w_0^{(m)} = \frac{h^2 - L_1}{h^2}, w_i^{(m)} = \frac{1}{2h^2}, i = 1, \dots, 2L_1 \quad (27)$$

where $z_{u,i}$ means the measurement variable calculated with sigma point $s_{u,i}^-$.

The correlation matrix between state variables and measurement is

$$\mathbf{P}_{uz}^u = \sqrt{\mathbf{Q}_{aa} q(L_1)} \mathbf{C}_{1,1:n}^T \quad (28)$$

and the Kalman gain is

$$\mathbf{K}_u = \mathbf{P}_{uz}^u \mathbf{Q}_{zz}^{-1} \quad (29)$$

with

$$\mathbf{Q}_{zz} = \mathbf{C}_1 \mathbf{C}_1^T + \mathbf{C}_2 \mathbf{C}_2^T + \mathbf{R} \quad (30)$$

and

$$\mathbf{C}_{1,i} = w_1^{(c)} (z_{u,i} - z_{u,L_1+i}), i = 1, \dots, L_1 \quad (31)$$

$$\mathbf{C}_{2,i} = w_2^{(c)} (z_{u,i} + z_{u,L_1+i} - 2z_{u,0}), i = 1, \dots, L_1 \quad (32)$$

$$w_1^{(c)} = \frac{1}{2h}, w_2^{(c)} = \frac{\sqrt{2h^2 - 1}}{2h^2} \quad (33)$$

The mean state is updated as

$$s_u^+ = s_u^- + \mathbf{K}_u (z - \bar{z}) \quad (34)$$

The posterior covariance is updated by making use of the full state variables, whose sigma points are represented as

$$\mathbf{S}_{u,v} = \begin{cases} \bar{s}_{u,v}^-, & i = 0 \\ \bar{s}_{u,v}^- + h\Theta_i, & i = 1, \dots, L_2 \\ \bar{s}_{u,v}^- - h\Theta_i, & i = L_2 + 1, \dots, 2L_2 \end{cases} \quad (35)$$

where $\bar{s}_{u,v}^- = [\bar{s}_u^{-T}, \bar{s}_v^{-T}]^T$, and

$$\Theta = \left[\sqrt{\mathbf{Q}_{ss} q(L_2)}, \sqrt{\mathbf{Q}_{ss} q(L_2)} \right] \quad (36)$$

where $q(L_2)$ means taking L_2 columns of $\sqrt{\mathbf{Q}_{ss}}$ with some specific criterion (see the last paragraph of this section).

$$\mathbf{Q}_{ss}^- = \begin{bmatrix} \mathbf{Q}_{uu}^- & \mathbf{Q}_{uv}^- \\ \mathbf{Q}_{vu}^- & \mathbf{Q}_{vv}^- \end{bmatrix} \quad (37)$$

The correlation matrix between the full state variables and observations, and the Kalman gain are computed in a similar way as for the mean update step, except that only the u components that are related to the on-line state variables are updated. The posterior covariance is computed using Eq. (13).

We now give the method to specify $q(L_1)$ and $q(L_2)$. In the original formulation of the sigma-point Kalman filter (see e.g., Van der Merwe, 2004), the total number of sigma points is chosen equal to $2L + 1$, where L is the total dimension of the problem, including the dimension of the state variables, the dimension of measurements and the dimension of process noise of the forecast model. For high dimensional systems, as often found in inversion, such formulation would require tremendous computation. For instance, for a lag length of 6 and a total number of 62 measurement site in this study, the total dimension L would be $195 (= 19 \times 6 + 19 + 62)$, without model noise being accounted for. Even by assuming additive error between the model and measurement, L is 133, which means 267 sigma-points are needed to propagate the state variables. So to reduce the computation burden, only a subset of the total sigma points is chosen to propagate the system. Those sigma points are chosen such that they are closest to the error sub-space that is spanned by the dominant eigen vectors of the error space. In implementation, the error sub-space is formed using the principal component analysis (Smith, 2002).

2.5 Dealing with interval constraint

When implementing our developed methods in inversion, we find some of the inverted fluxes is negative or of extremely large values. The following techniques are used to impose the interval constraint to overcome this problem. Since the KS can only deal with linear operators, we use the projection operator method (Simon and Chia, 2002) to impose the

constraint. After inversion using the unconstrained formula in Sect. 2.2, the following minimization problem is solved

$$J_2 = (s_u^{++} - s_u^+)^T (\mathbf{Q}_{uu}^+)^{-1} (s_u^{++} - s_u^+) \quad (38)$$

$$lb_u \leq s_u^{++} \leq ub_u$$

We use the method associated with the concept of active set (Murty, 1988) to solve the above optimization problem iteratively. Specifically, before each iteration, we first identify the set of variables (say $w(s)$) that violate the interval constraint. This would give equality constraint of those variables

$$\mathbf{c}(s_u^{++}) = \mathbf{o} \quad (39)$$

where \mathbf{c} is a linear operator to impose the proper constraints.

Then the problem of Eqs. (38) and (39) is solved using the Kalman update

$$s_u^{++} = s_u^+ - \mathbf{Q}_{uu}^+ \mathbf{c}^T (\mathbf{c} \mathbf{Q}_{uu}^+ \mathbf{c}^T)^{-1} (\mathbf{c}(s_u^+)) \quad (40)$$

$$\mathbf{Q}_{uu}^{++} = \mathbf{Q}_{uu}^+ - \mathbf{Q}_{uu}^+ \mathbf{c}^T (\mathbf{c} \mathbf{Q}_{uu}^+ \mathbf{c}^T)^{-1} \mathbf{c} \mathbf{Q}_{uu}^+ \quad (41)$$

After one iteration, if there are still (usually fewer) variables violating the constraint, the above iteration is repeated until all constraints are satisfied.

For the ESRKS and the SRCDKS, the following variable transform is used to impose the constraint

$$y = lb + \left(0.5 + \frac{\text{atan}(x)}{\pi} \right) (ub - lb) \quad (42)$$

The above transform ensures the variable y will always fall within the range (lb, ub) . In the inversion, the state variable is defined as a vector containing the scaling factors of the flux adjustments defined with respect to the prior fluxes. The posterior fluxes are thus the sum of prior fluxes and their adjustments. The constraints are imposed by letting the scaling factors fall in the range $[-0.95, 2.0]$ for all the flux adjustments, except for that of the stratospheric destruction, which is set to $[-0.2, 0.2]$.

2.6 GEOS-Chem model and sensitivity matrix

To test our developed methods, we use the CH_4 forward transport simulations from an atmospheric transport model GEOS-Chem (Bey et al., 2001; Wang et al., 2004) to derive the elements of the sensitivity matrix \mathbf{H} . GEOS-Chem is a global 3-D offline transport model developed in the atmospheric chemistry group at Harvard University. The version used in this study is based on v8-01-03 (<http://www.as.harvard.edu/chemistry/trop/geos>), with proper modifications to transport CH_4 . The model is driven by meteorological data assimilated by the Goddard Earth Observing System (GEOS) at the NASA Global Modeling and Assimilation Office (GMAO) (Bey et al., 2001). An evolution of different versions of GEOS assimilated meteorology data from GEOS-1 to most recent GEOS-5 are used in GEOS-Chem simulations. In this study, we use the GEOS-5 meteorology

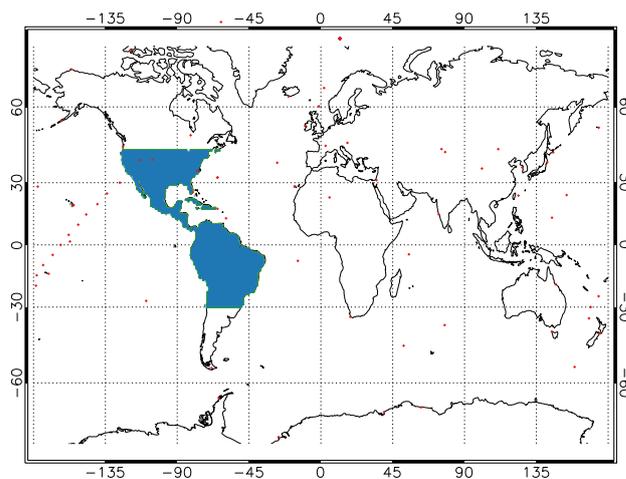


Fig. 1. Map of measurement sites involved in the inversion experiments. The blue shaded region is used to compare the inverted wetland CH_4 flux by the different methods.

data. The horizontal resolution of the simulation is 4° longitude by 5° latitude, with 47 hybrid eta levels in the vertical direction that extends from 1000 hPa at surface to 0.01 hPa at top.

The surface fluxes are organized into two categories. The seasonal fluxes, i.e. the first category fluxes, vary month by month and the aseasonal fluxes, i.e. the second category fluxes, keep constant throughout the inversion period. For seasonal fluxes, 1 Tg CH_4 is emitted in one month and is then shut down and allowed to decay until the end of forward simulation. For aseasonal fluxes, 1 Tg $\text{CH}_4 \text{ month}^{-1}$ is emitted in a year and is then shut down and allowed to decay until the end of forward simulation. We sample the responses to the emissions at the site-locations involved in the globalview- CH_4 -2009 product (GLOBALVIEW- CH_4 , 2009). The forward simulation is done from 1 January 2004 to 1 January 2010, totally 6 yr. We choose a subset of 62 measurement sites (Fig. 1) to form the observational network. Therefore, for every month, the sensitivity matrix is of maximum size 38×62 in our study. The maximum overall size of \mathbf{H} for a “batch” inversion is thence $(19 \times 72 + 19) \times (62 \times 72)$, which is much larger than that involved in the fixed-lag inversions.

2.7 Implementation and comparison experiments

The different methods are coded with Fortran 95. The linear algebra is done with publicly available packages of BLAS and LAPACK from the Intel Math Kernel Library version 10.3.

Comparisons are made for an ideal inversion problem with known true values of the fluxes. The synthetic observations are sampled at 62 sites (GLOBALVIEW- CH_4 , 2009) (Fig. 1). To quantify the uncertainty in model simulations when compared to globalview measurements, the relative residual error

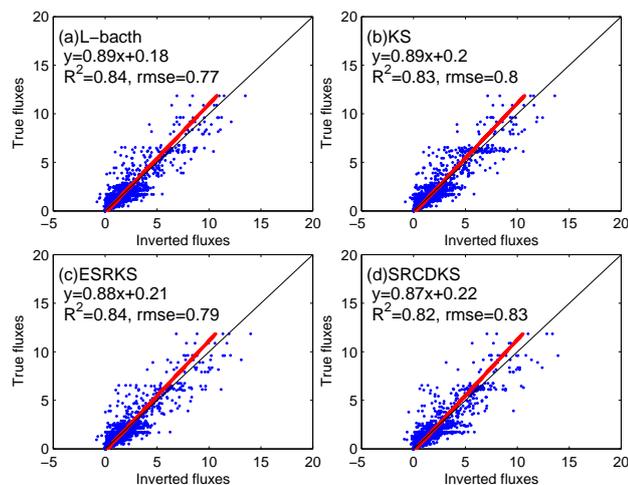


Fig. 2. Inversion results from the interval unconstrained inversion using different methods. A lag length of 7 and a correlation length of 0 are used for all Kalman update smoothers. The ensemble size for ESRKS is 500, and that for SRCDKS is 101. The red lines are the linear regressions of the inversions against the true fluxes.

(RRE) are derived for time series at each sample location by comparing a reference run with the globalview CH₄ data. The RREs are different from January to December and are assumed of no interannual variability. These RREs are then used to derive the uncertainty (one- σ error) at different sampling locations for the given temporal domain. The observations are finally obtained by perturbing the sampled time series with the deduced one- σ error. The prior fluxes for the inversion are perturbed randomly from the true fluxes, and are used to run a prior simulation to sample the prior CH₄ concentrations at the measurement sites.

3 Results and discussions

Inversions using different lag lengths for the different Kalman update smoothers are shown in Table 1 and Fig. 2. It is found a lag length 6 is enough to obtain stable inversions, and inversions with a lag length 7 are conducted for the comparisons. Results from a linear batch inversion and those from the KS, ESRKS and SRCDKS with a lag length 7 are compared in Fig. 2. Due to the ill-posedness of the inversion problem, we find that all the inversions are not able to fully reveal the true fluxes. Some unreasonable negative fluxes are found in all inversions, due to the insufficient observational constraint. This justifies the necessity of using some method to impose the interval constraint on the posterior fluxes.

Posterior fluxes and their uncertainties obtained from the interval unconstrained inversions using different months of correlations are compared at the selected region for a two-year period in Fig. 3. For the KS, differing from the results in Bruhwiler et al. (2005) and Michalak (2008), we find

Table 1. Inversion results using different lag lengths without taking correlations into account. All statistics are tested for statistical significance with $p < 0.001$. The first value in the brackets is the root mean square error of the inversion; the second value is the R^2 value of the linear regression. The regressions are the posterior fluxes against the true fluxes. The SRCDKS is implemented with the reduced set of sigma points chosen using the PCA method. No interval constraint is imposed for the inversions.

Methods	Lag=6	Lag=8
KS	$0.88x+0.20$ (0.80, 0.83)	$0.89x+0.20$ (0.80, 0.83)
ESRKS-500	$0.88x+0.21$ (0.79, 0.84)	$0.88x+0.22$ (0.80, 0.83)
SRCDKS	$0.87x+0.22$ (0.83, 0.82)	$0.87x+0.22$ (0.82, 0.82)

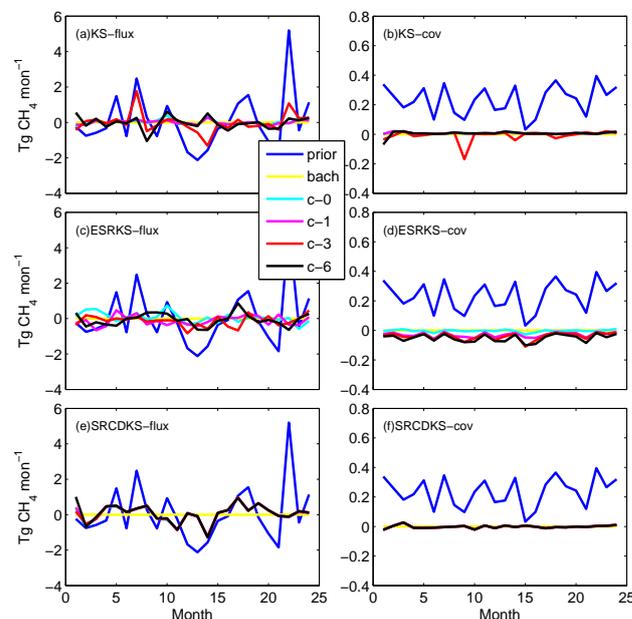


Fig. 3. Two-year comparison of the posterior fluxes and posterior uncertainties between inversions using the different Kalman update smoothers and those from the linear batch inversion. Results from the linear batch inversion have been subtracted from the time series for a better illustration. The ensemble size for ESRKS is 500, and that for SRCDKS is 101. None of the Kalman type smoothers uses interval constraints.

that the incorporation of correlations between on-line and off-line state variables does not improve the inverted fluxes and their posterior uncertainties appeared even smaller when compared to the batch inversion. Including more months of correlations further reduce the posterior uncertainty. Such results are explained by the ill-posedness of the problem. In the linear batch inversion, we find the inversion failed when the state variable is updated using Eq. (11), and the covariance is updated using Eq. (13). So their equivalent forms (e.g. Eqs. (12–13) in Bruhwiler et al., 2005) derived using the Sherman-Morrison-Woodbury identity are used in the

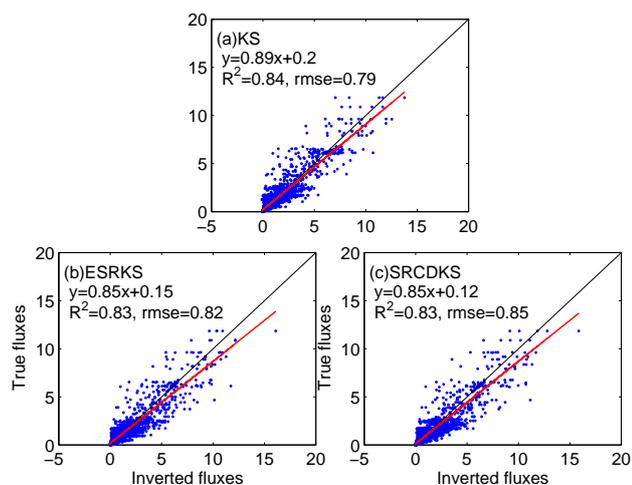


Fig. 4. Inversion results from the interval constrained inversion using different methods. A lag length of 7 and a correlation length of 0 are used for all Kalman update smoothers. The ensemble size for ESRKS is 500, and that for SRCDKS is 101. The red lines are the linear regressions of the inversions against the true fluxes.

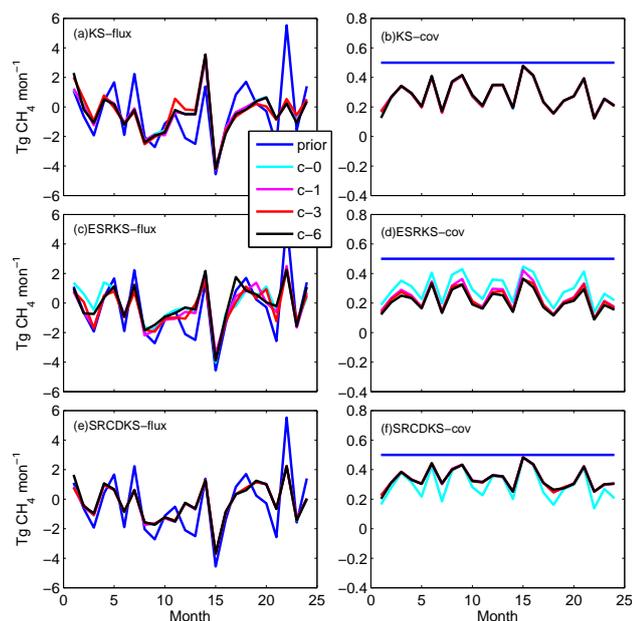


Fig. 5. Two-year comparison of the posterior fluxes and posterior uncertainties between inversions using the different Kalman update smoothers. The true fluxes have been subtracted from the time series for a better illustration. The ensemble size for ESRKS is 500, and that for SRCDKS is 101. Interval constraints are applied for all Kalman update smoothers using the different algorithms described in the text.

computation to ensure the numerical stability. For ESRKS, the inclusion of one month of correlations slightly degrades the inversion. Including more months of correlations does improve the inversion when compared to the true fluxes and

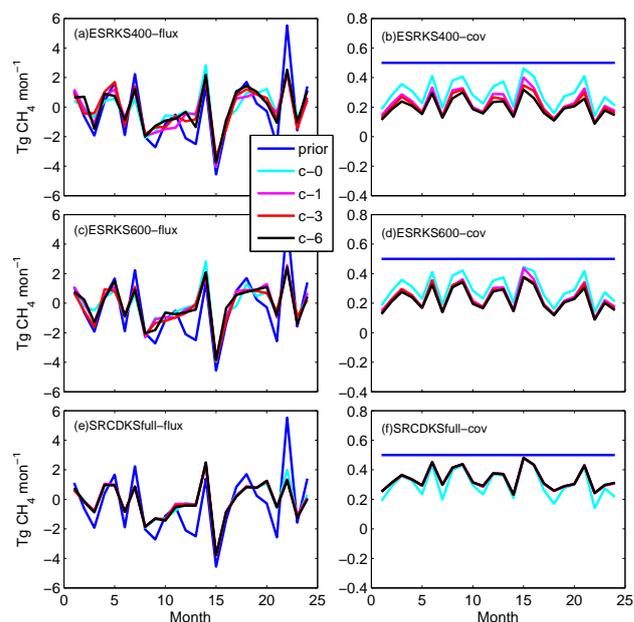


Fig. 6. Two-year comparison of the posterior fluxes and posterior uncertainties between inversions using the different ensemble Kalman update smoothers. The true fluxes have been subtracted from the time series for a better illustration. The ESRKS in panels (a) and (b) uses an ensemble size 400, and uses an ensemble size 600 in (c) and (d). The SRCDKS in panels (e) and (f) uses a ensemble size 305. Interval constraints are applied for all inversions using the variable transform algorithm described in the text.

reduces the posterior uncertainty. The results from including six months of correlations are even better than that from the linear batch inversion. The inclusion of correlations does not affect the inversion result much when the SRCDKS is used. This is explained by the fact that better approximations of the covariance matrices are obtained by the SRCDKS because of its second order accuracy when the Gaussian error is assumed. Also the information that can be extracted from the approximated correlation matrix provides much less constraint on the on-line state variables than the observations. Therefore, we conclude that the benefit from including correlations between the on-line and off-line state variables depend on the property of the problem to be solved. While in the carefully designed problems in Bruhwiler et al. (2005) and Michalak (2008), the correlation propagation is useful to further constrain the inversion. For the problems in this study, little or no benefit is obtained from the inclusion of the correlations between the on-line and off-line state variables.

Since the interval unconstrained inversions could produce some unphysical results, the interval constrained inversions are conducted using the three Kalman update smoothers. The results are compared with the true fluxes in Fig. 4, Fig. 5 and Table 2. After applying the interval constraint, we find all negative fluxes are corrected into the reasonable range. The impact of inclusion of the correlations between the on-line

Table 2. Inversion results using different correlation lengths (CL) with the lag length equals to 7. All statistics are tested for statistical significance with $p < 0.001$. The first value in the brackets is the root mean square error of the inversion; the second value is the R^2 value of the linear regression. The regressions are the posterior fluxes against the true fluxes. Interval constraints are applied using the methods described in the text. The SRCDKS is implemented with the reduced set of sigma points chosen using the PCA method.

Methods	CL=0	CL=1	CL=3
KS	$0.89x+0.16$ (0.79, 0.83)	$0.89x+0.16$ (0.80, 0.83)	$0.88x+0.17$ (0.82, 0.83)
ESRKS-500	$0.85x+0.15$ (0.82, 0.83)	$0.85x+0.17$ (0.80, 0.84)	$0.85x+0.17$ (0.81, 0.84)
SRCDKS	$0.85x+0.12$ (0.85, 0.83)	$0.84x+0.13$ (0.86, 0.82)	$0.84x+0.13$ (0.85, 0.83)

Table 3. Inversion results using different correlation lengths (CL) with the lag length equals to 7. All statistics are tested for statistical significance with $p < 0.001$. The first value in the brackets is the root mean square error of the inversion; the second value is the R^2 value of the linear regression. The regressions are the posterior fluxes against the true fluxes. Interval constraints are applied using the variable transform method described in the text. The SRCDKS is implemented with the full set of sigma points determined by the error covariance matrix of the state variables.

Methods	CL=0	CL=1	CL=3
ESRKS-400	$0.86x+0.14$ (0.79, 0.84)	$0.85x+0.17$ (0.81, 0.84)	$0.85x+0.17$ (0.80, 0.84)
ESRKS-600	$0.86x+0.15$ (0.80, 0.84)	$0.85x+0.17$ (0.80, 0.84)	$0.85x+0.17$ (0.81, 0.84)
SRCDKS-full	$0.86x+0.17$ (0.78, 0.85)	$0.87x+0.16$ (0.77, 0.85)	$0.87x+0.15$ (0.77, 0.85)

and off-line state variables again differs among the inversions using different methods. For the interval constrained KS, the inclusion of correlations degrades the inverted posterior fluxes. However, for ESRKS and SRCDKS, the inclusion of correlations does not affect the posterior fluxes significantly (Table 2). Including more months of correlations in KS does not change the posterior uncertainty of the fluxes significantly because of the numerical noise in the implementation. With the ESRKS, more months of correlations reduces the posterior uncertainty for the selected fluxes at the specific time period. In contrast, with the SRCDKS, the inclusion of correlations increases the posterior uncertainty and the inversion is not significantly dependent on the number of months that are used to propagate the correlations (Fig. 5).

For the ensemble filters ESRKS and SRCDKS, we test the impact of ensemble size on the inversions (Fig. 6 and Table 3). When the ESRKS is implemented with an ensemble size 400, the inverted posterior fluxes have a slightly worse linear fitting against the true fluxes when compared to that from the inversion using an ensemble size 500 (Table 2). Increasing the ensemble size to 600, the inversion changes slightly. Therefore 500 is a proper ensemble size to obtain a good inversion with ESRKS. The inclusion of correlations between on-line and off-line state variables impacts the results more when the ensemble size is smaller. This is because of the insufficient approximation of the correlations for the inversion using a smaller ensemble size and it tends to degrade the inverted fluxes. The posterior uncertainties appear smaller as the correlation length increases for the inversion. The uncertainty reduction is most significant by the inclu-

sion of the first month of correlations. For the SRCDKS with the full set of sigma points determined by the covariance matrix, the inversion provides superior results to that uses sigma points chosen with the PCA method. In addition, we notice that the inclusion of correlations in such case increases the posterior uncertainty for the specific fluxes used in comparison. The inclusion of the first month of correlations again improves the inversion most (Table 3), similar to the finding in Bruhwiler et al. (2005).

We also repeat the above comparisons using observations from 151 globalview sites. The results (not shown) are similar to the inversions using 62 globalview sites. The linear batch inversion still results in some negative fluxes and the inclusion of correlations between on-line and off-line state variables only affects the posterior fluxes from the Kalman update smoothers slightly. We conclude that our developed methods are robust in atmospheric CH_4 inversions, irrespective of the number of sites being used.

4 Conclusions

Propagating correlations between on-line and off-line state variables is necessary to improve the sequential atmospheric inversion with the KS proposed in (Bruhwiler et al., 2005). We show that such correlations act as a balanced constraint on the on-line state variables. We then extend the correlation propagating scheme to two different ensemble smoothers. The extensions are able to account for nonlinearity in the inversions and impose interval constraints on the inverted

fluxes easily. The comparison studies indicate the new methods are good alternatives to existing methods in inverting fluxes of trace gases (e.g., CO₂ and CH₄) using atmospheric measurements. The methods are potentially useful to solve very high dimensional inversion problems.

Acknowledgements. J. Tang is supported with graduate fellowships from the Purdue Climate Change Research Center, the NASA Earth System Science program and the Bilsland Dissertation Fellowship. The research is also supported by the National Science Foundation (projects DEB-0919331, ARC-0554811 and EAR-0630319). We thank the GEOS-Chem group at Harvard University for providing the original model code and meteorology data for the simulations. Comments and suggestions from two anonymous reviewers and the editor also helped improve greatly the earlier version of this work.

Edited by: M. Heimann

References

- Ambadan, J. T. and Tang, Y.: Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems, *J. Atmos. Sci.*, 66, 261–285, 2009.
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B., Fiore, A. M., Li, Q., Liu, H., Mickley, L. J., and Schultz, M.: Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation, *J. Geophys. Res.*, 106, 23073–23096, 2001.
- Bruhwiller, 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.
- Enting, I. G.: *Inverse Problems in Atmospheric Constituent Transport*, Cambridge University Press, UK, 2002.
- Gelb, A.: *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.
- GLOBALVIEW-CH4: Cooperative Atmospheric Data Integration Project – Methane, CD-ROM, NOAA ESRL, Boulder, Colorado, also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ceg/ch4/GLOBALVIEW], 2009.
- Gurney, K. R., Law, R. W., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiller, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, 2002.
- HaasLaursen, D. E., Hartley, D. E., and Prinn, R. G.: Optimizing an inverse method to deduce time-varying emissions of trace gases, *J. Geophys. Res.*, 101, 22823–22831, 1996.
- Hartley, D. and Prinn, R.: Feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model, *J. Geophys. Res.*, 98, 5183–5197, doi:10.1029/92JD02594, 1993.
- Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M.: Inverse modeling of methane sources and sinks using the adjoint of a global transport model, *J. Geophys. Res.*, 104, 26137–26160, doi:10.1029/1999JD900428, 1999.
- Julier, S. J. and Uhlmann, J. K.: Unscented filtering and nonlinear estimation, *Proc. IEEE*, 92, 401–422, 2004.
- Michalak, A. M.: Technical Note: Adapting a fixed-lag Kalman smoother to a geostatistical atmospheric inversion framework, *Atmos. Chem. Phys.*, 8, 6789–6799, doi:10.5194/acp-8-6789-2008, 2008.
- Murty, K. G.: *Linear complementarity, linear and nonlinear programming*, Heldermann Verlag, Berlin, 1988.
- Nørgaard, M., Poulsen, N. K., and Ravn, O.: *Advances in Derivative-Free State Estimation for Nonlinear Systems*, Tech. rep., 1998.
- 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 CO₂ surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110, D24304, doi:10.1029/2005JD006157, 2005.
- Simon, D. and Chia, T. L.: Kalman filtering with state equality constraints, *IEEE T. Aero. Elec. Syst.*, 38, 128–136, 2002.
- Smith, L. I.: A tutorial on Principal Components Analysis, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf (last access: July 2010), 2002.
- Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, PA, USA, 2005.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square-root filters, *Mon. Weather Rev.*, 131, 1485–1490, 2003.
- Van der Merwe, R.: *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*, Ph.D. thesis, Oregon Health and Science University, 2004.
- Wang, J. S., Logan, J., McElroy, M., Duncan, B., Megretskaia, I., and Yantosca, R.: A 3-D model analysis of the slowdown and interannual variability in the methane growth rate from 1988 to 1997, *Global Biogeochem. Cy.*, 18, GB3011, doi:10.1029/2003GB002180, 2004.