Joint state and parameter estimation with an iterative ensemble Kalman smoother

Marc Bocquet, Pavel Sakov

Nonlin. Processes Geophys., 20, 803–818, 2013
www.nonlin-processes-geophys.net/20/803/2013/
doi:10.5194/npg-20-803-2013

Nonlinear Processes in Geophysics
Open Access

Joint state and parameter estimation with an iterative ensemble Kalman smoother

M. Bocquet[1,2] **and P. Sakov**[3]

[1]Université Paris-Est, CEREA joint laboratory École des Ponts ParisTech and EDF R&D, France
[2]INRIA, Paris Rocquencourt research centre, France
[3]Bureau of Meteorology, Melbourne, Australia

Correspondence to: M. Bocquet (bocquet@cerea.enpc.fr)

Received: 4 June 2013 – Revised: 2 September 2013 – Accepted: 4 September 2013 – Published: 23 October 2013

Abstract. Both ensemble filtering and variational data assimilation methods have proven useful in the joint estimation of state variables and parameters of geophysical models. Yet, their respective benefits and drawbacks in this task are distinct. An ensemble variational method, known as the iterative ensemble Kalman smoother (IEnKS) has recently been introduced. It is based on an adjoint model-free variational, but flow-dependent, scheme. As such, the IEnKS is a candidate tool for joint state and parameter estimation that may inherit the benefits from both the ensemble filtering and variational approaches.

In this study, an augmented state IEnKS is tested on its estimation of the forcing parameter of the Lorenz-95 model. Since joint state and parameter estimation is especially useful in applications where the forcings are uncertain but nevertheless determining, typically in atmospheric chemistry, the augmented state IEnKS is tested on a new low-order model that takes its meteorological part from the Lorenz-95 model, and its chemical part from the advection diffusion of a tracer. In these experiments, the IEnKS is compared to the ensemble Kalman filter, the ensemble Kalman smoother, and a 4D-Var, which are considered the methods of choice to solve these joint estimation problems. In this low-order model context, the IEnKS is shown to significantly outperform the other methods regardless of the length of the data assimilation window, and for present time analysis as well as retrospective analysis. Besides which, the performance of the IEnKS is even more striking on parameter estimation; getting close to the same performance with 4D-Var is likely to require both a long data assimilation window and a complex modeling of the background statistics.

1 Introduction

Data assimilation in geophysics is often concerned with the estimation of the state of the system (e.g. atmosphere, ocean). Yet, non-observed parameters of the model can also be seen as control variables. They can indirectly be estimated through the assimilation of observations. In such context, data assimilation can be a powerful inverse modeling tool.

With the progress in techniques as well as the rise in popularity of data assimilation in geosciences, this topic has become of increasing interest. Parameter estimation is useful because it can account for model error through a parametric representation of the uncertain processes, and could serve as a tool to enhance the system state estimation. For instance, it is now accepted that air quality forecasting can benefit considerably from the online estimation of forcing parameters. Parameter estimation is also a fundamental tool *per se* in the estimation of the parameters which are often of physical or societal interests. For instance, again regarding air quality, data assimilation can help assess effective kinetic rates of interest to chemists, or it can help assess regulated pollutant emissions of interest to policy makers.

1.1 Data assimilation techniques for parameter estimation

As is the case with data assimilation for state estimation, two types of approach have been used for parameter estimation: filtering methods and variational methods.

The estimation of parameters by the filtering approaches is based on the augmentation of the state vector with the parameter variables. If the state space has dimension M and if the number of parameters is P, then the augmented control

vector has dimension $M + P$. Through the assimilation of observations, the joint analysis of the state variables and the parameters aims at building covariances (or higher-order dependencies for non-Gaussian filters) between them; these are crucially needed because of the non-observability of most parameters. The augmented state principle is likely to be used with any type of filter: extended Kalman filters (e.g., Kondrashov et al., 2008), ensemble Kalman filters (e.g., Aksoy et al., 2006; Wirth and Verron, 2008; Barbu et al., 2009), particle filters (e.g., Vossepoel and van Leeuwen, 2007; Weir et al., 2013), and stochastic sampling and genetic algorithms (e.g., Jackson et al., 2004; Liu et al., 2005; Bocquet, 2012; Posselt and Bishop, 2012). In an enlightening review, Ruiz et al. (2013) have discussed the use of ensemble Kalman filters (EnKFs) for parameter estimation. When the filtering method accounts for asynchronous observations by building covariances between parameter errors defined at distinct times, the method is usually referred to as a smoother (Evensen, 2003; Hunt et al., 2004; Sakov et al., 2010; Cosme et al., 2010).

The estimation of parameters with the variational approach is based on the explicit dependence of the cost function in not only the state variables, but also the parameters. If the dependence is not explicit, one should at least be able to compute the gradient of the cost function with respect to the parameters. The four-dimensional variational method, or 4D-Var (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Rabier et al., 2000), has the distinct advantage of being a natural smoother since it works within a temporal window to assimilate asynchronous observations. However, it requires the use of the adjoint evolution model to compute gradients of the cost function. Computing the gradient with respect to the model parameters requires the same adjoint model, and also the extra effort of computing the explicit derivative of the cost function with respect to the parameters, in terms of the adjoint variables. It has been used for parameter estimation by, e.g., Pulido and Thuburn (2006), Bocquet (2012), and Kazantsev (2012).

This list of contributions to the field is far from being exhaustive and merely illustrates some of the methodologies used in atmospheric, ocean and climate sciences. In particular, there is a vast literature in atmospheric chemistry dedicated to the inversion of sources of pollutants and tracers. The extended and ensemble Kalman filters and variational methods (3D-Var and 4D-Var) have been employed in this field for over two decades (Zhang et al., 2012, and references within). Owing to the (quasi-)linearity of some chemical species, simpler four-dimensional smoothing analysis merely using a Best Linear Unbiased Estimator (BLUE) have also been extensively used to estimate sources.

1.2 The iterative ensemble Kalman smoother

The iterative ensemble Kalman smoother (IEnKS) has been recently proposed (Bocquet and Sakov, 2013) as an extension of the iterative ensemble Kalman filter (Sakov et al., 2012; Bocquet and Sakov, 2012). It is meant to solve the variational problem of 4D-Var with the help of a 4D ensemble. As such, it is a 4D ensemble variational method of the type used in the work by Buehner et al. (2010), Chen and Oliver (2012) and Fairbairn et al. (2013), and has (more remote) connections with ensemble of variational methods (Raynaud et al., 2009; Bowler et al., 2013). It does not require the use of the adjoint observation and evolution models since the sensitivities are estimated with the ensemble (Gu and Oliver, 2007; Liu et al., 2008). Moreover, the IEnKS generates the posterior ensemble using Gaussian assumptions and forecasts the ensemble to the next update step in the same way an ensemble square root Kalman filter does. Note that the scheme is not a hybrid method since it does not combine two distinct methods.

Because the IEnKS fundamentally solves a variational problem, it may require iterations for the cost function minimization. The number of iterations depends on the nonlinearity of the system. This number is expected to be small (1 or 2) for weak nonlinearity (typical of synoptic scale meteorology).

Using perfect model assumptions, Bocquet and Sakov (2013) have tested the IEnKS on two low-order models in different regimes representing different nonlinearities and lengths of the data assimilation window (DAW). The IEnKS (often significantly) outperforms EnKF and the standard ensemble Kalman smoother (EnKS) in all these regimes, not only regarding the smoothing performance (retrospective state estimation) but also regarding the filtering performance (state estimation at present and future time). Here, we will also show that the IEnKS also outperforms 4D-Var in this context.

In addition, the IEnKS has been shown on these models to be able to handle long DAWs, especially when assimilating observations several times (in a mathematically consistent manner).

Because the IEnKS offers the advantages of both filtering and variational methods, and because it is capable of operating on long DAWs, it has considerable potential as an efficient parameter estimation method.

1.3 Objective and outline

The objective of this article is to introduce a straightforward extension of the IEnKS to joint state and parameter estimation, and to test the potential of the approach on low-order models. The physical context is that of chaotic geophysical models, and of atmospheric chemical/tracer models, in which a joint state and parameter estimation is, in our opinion, a key to successful forecasts.

The algorithm of the IEnKS will be described in Sect. 2, in a compact but comprehensive manner. The method will then be generalized to joint state and parameter estimation. In Sect. 3, the capabilities of the IEnKS on the Lorenz-95 model (Lorenz and Emmanuel, 1998) will be reported. Additional

tests will be performed: a comparison with the state-of-the-art EnKF and standard EnKS, as well as with a 4D-Var, and with a new cycling of the IEnKS DAWs. Then the IEnKS will be tested for joint state and parameter estimation on the Lorenz-95 model (Lorenz and Emmanuel, 1998). In Sect. 4, an original extension of the Lorenz-95 with the advection of a tracer will be introduced. It is meant to represent the dynamics of an online atmospheric chemistry model, or meteorological models with a constituent such as moisture, with two unobserved parameters: the Lorenz-95 forcing parameter, and the emission flux. The IEnKS, the EnKF/EnKS, and a 4D-Var will be tested and compared in this context. The results will be discussed in Sect. 5. Conclusions will be drawn in Sect. 6.

2 The iterative ensemble Kalman smoother for joint state and parameter estimation

2.1 The algorithm

A Bayesian derivation of the IEnKS can be found in Bocquet and Sakov (2013). However, we would like to introduce the IEnKS comprehensively in this article: reference to Bocquet and Sakov (2013) will only be made regarding details that are not directly relevant to this study. Here, we describe the algorithm with its main justifications, and then provide its pseudo-code.

2.1.1 The core algorithm

Observation vectors $\boldsymbol{y} \in \mathbb{R}^d$ are assumed to be collected every time step Δt. Time is discretized into the times t_k when the observations are collected. The number d of scalar observations within $\boldsymbol{y}$ can be time-dependent. The observations are related to the state vector through a possibly nonlinear, possibly time-dependent observation operator H_k. The observation errors are assumed to be Gaussian-distributed, unbiased, and uncorrelated in time, and to have an observation error covariance matrix $\mathbf{R}_k$.

The analysis step of the assimilation scheme is performed over a window of length $L\Delta t$ in time units. Unless otherwise stated, time index k is relative to present time. With this convention, present time is set to be always t_L, so that the initial condition of the DAW is conveniently always t_0.

Let us first describe the update step. At t_0 (i.e. $L\Delta t$ in the past), the background is obtained from an ensemble of N state vectors of $\mathbb{R}^M$: $\boldsymbol{x}_{0,[1]}, \ldots, \boldsymbol{x}_{0,[n]}, \ldots, \boldsymbol{x}_{0,[N]}$. Index 0 refers to time while $[n]$ refers to the ensemble member index. They can be stored in a matrix $\mathbf{E}_0 = [\boldsymbol{x}_{0,[1]}, \ldots, \boldsymbol{x}_{0,[N]}] \in \mathbb{R}^{M\times N}$. One can equivalently represent the ensemble with its mean $\overline{\boldsymbol{x}}_0 = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_{0,[n]}$ and its anomaly matrix $\mathbf{A}_0 = [\boldsymbol{x}_{0,[1]} - \overline{\boldsymbol{x}}_0, \ldots, \boldsymbol{x}_{0,[N]} - \overline{\boldsymbol{x}}_0]$.

As in the ensemble Kalman filter, this background is approximated as a Gaussian distribution of mean $\overline{\boldsymbol{x}}_0$, and covariance matrix $\mathbf{A}_0\mathbf{A}_0^{\mathrm{T}}/(N-1)$, the first- and second-order empirical moments of the ensemble. The background is rarely full rank since the anomalies of the ensemble span a vector space of dimension smaller than or equal to $N-1$ and in a realistic context $N \ll M$. Therefore, one solves for the analysis state vector $\boldsymbol{x}_0$ in the ensemble space $\overline{\boldsymbol{x}}_0 + \mathrm{Vec}\left\{\boldsymbol{x}_{[1]} - \overline{\boldsymbol{x}}_0, \ldots, \boldsymbol{x}_{[N]} - \overline{\boldsymbol{x}}_0\right\}$, which can be written $\boldsymbol{x}_0 = \overline{\boldsymbol{x}}_0 + \mathbf{A}_0\boldsymbol{w}$, where $\boldsymbol{w} \in \mathbb{R}^N$ is a vector of coefficients in ensemble space.

The analysis of IEnKS over $[t_0, t_L]$ is obtained from a cost function. The restriction of this cost function in state space to the ensemble space yields:

$$\widetilde{\mathcal{J}}(\boldsymbol{w}) = \frac{1}{2}(N-1)\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \frac{1}{2}\sum_{k=1}^{L}\beta_k \boldsymbol{\delta}_k^{\mathrm{T}}(\boldsymbol{w})\mathbf{R}_k^{-1}\boldsymbol{\delta}_k(\boldsymbol{w}),$$
$$\boldsymbol{\delta}_k(\boldsymbol{w}) = \boldsymbol{y}_k - H_k \circ \mathcal{M}_{k\leftarrow 0}\left(\boldsymbol{x}_0^{(0)} + \mathbf{A}_0\boldsymbol{w}\right). \quad (1)$$

The tilde symbol signifies that $\widetilde{\mathcal{J}}$ is a mathematical object defined in ensemble space. $\mathcal{M}_{k\leftarrow 0}$ is the possibly nonlinear transition operator from t_0 to t_k. $\{\beta_k\}_{1\le k\le L}$ are scalars in $[0,1]$ that weight the observations within the DAW. The choice of the β_k can be made mathematically consistent and can have dramatic consequences on the performance of the data assimilation system. We refer to Bocquet and Sakov (2013) for a justification and numerical tests. Nonetheless, the rational for the choice of the $\{\beta_k\}_{1\le k\le L}$ will be discussed later.

This cost function is iteratively minimized in the ensemble space following the Gauss–Newton algorithm:

$$\boldsymbol{w}^{(j+1)} = \boldsymbol{w}^{(j)} - \widetilde{\mathcal{H}}_{(j)}^{-1}\nabla\widetilde{\mathcal{J}}_{(j)}(\boldsymbol{w}^{(j)}), \quad (2)$$

using the gradient $\nabla\widetilde{\mathcal{J}}_{(j)}$ and an approximate Hessian $\widetilde{\mathcal{H}}_{(j)}$ of the cost function:

$$\nabla\widetilde{\mathcal{J}}_{(j)} = -\sum_{k=1}^{L}\beta_k \mathbf{Y}_{k,(j)}^{\mathrm{T}}\mathbf{R}_k^{-1}\left[\boldsymbol{y}_k - H_k\circ\mathcal{M}_{k\leftarrow 0}(\boldsymbol{x}_0^{(j)})\right] + (N-1)\boldsymbol{w}^{(j)}, \quad (3)$$

$$\widetilde{\mathcal{H}}_{(j)} = (N-1)\mathbf{I}_N + \sum_{k=1}^{L}\beta_k \mathbf{Y}_{k,(j)}^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_{k,(j)}, \quad (4)$$

$$\boldsymbol{x}_0^{(j)} = \boldsymbol{x}_0^{(0)} + \mathbf{A}_0\boldsymbol{w}^{(j)}. \quad (5)$$

$\widetilde{\mathcal{H}}_{(j)}$ is an approximation of the full Hessian because it disregards the contribution of the second-order derivatives of the innovation vectors $\boldsymbol{\delta}_k(\boldsymbol{w})$ in the cost function. The notation (j) refers to the iteration index of the minimization. At the first iteration one sets $\boldsymbol{w}^{(0)} = \mathbf{0}$. $\mathbf{I}_N$ is the identity matrix in ensemble space. $\mathbf{Y}_{k,(j)} = [H_k\circ\mathcal{M}_{k\leftarrow 0}]'_{|\boldsymbol{x}_0^{(j)}}\mathbf{A}_0$ is the tangent linear of the operator from ensemble space to the observation space. The estimation of this sensitivity using the ensemble is what allows one to avoid the use of the model adjoint. Two implementations, referred to as the transform and the bundle variants, have been put forward (Sakov et al., 2012; Bocquet and Sakov, 2012). With the bundle scheme, for instance, the

ensemble is rescaled closer to the mean trajectory by a factor ε. It is then propagated through the model and the observation operators, after which it is rescaled back by the inverse factor ε^{-1}. The operation reads:

$$\mathbf{Y}_{k,(j)} \approx \frac{1}{\varepsilon} H_k \circ \mathcal{M}_{k\leftarrow 0}\left(\boldsymbol{x}_0^{(j)}\mathbf{1}^{\mathrm{T}} + \varepsilon \mathbf{A}_0\right)\left(\mathbf{I}_N - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}}{N}\right), \quad (6)$$

where $\mathbf{1} = (1,\ldots,1)^{\mathrm{T}} \in \mathbb{R}^N$.

Note that each iterative update Eq. (2) solves the inner quadratic variational problem:

$$\begin{aligned}\widetilde{\mathcal{J}}^{(j)}(\boldsymbol{w}) = {} & \frac{1}{2}(N-1)\left\|\boldsymbol{w} - \boldsymbol{w}^{(j)}\right\|^2 \\ & + \frac{1}{2}\sum_{k=1}^{L}\beta_k \Big\| \boldsymbol{y}_k - H_k \circ \mathcal{M}_{k\leftarrow 0}(\boldsymbol{x}_0^{(j)}) \\ & - \mathbf{Y}_{k,(j)}(\boldsymbol{w} - \boldsymbol{w}^{(j)})\Big\|^2_{\mathbf{R}_k}, \end{aligned} \quad (7)$$

where $\|\mathbf{z}\|^2_{\mathbf{G}} = \mathbf{z}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{z}$.

The iteration is stopped when $\|\boldsymbol{w}^{(j)} - \boldsymbol{w}^{(j-1)}\|$ becomes smaller than a predetermined threshold e. Let us denote $\boldsymbol{w}^\star$ the solution of the cost function minimization. The symbol $\star$ will be used with any quantity obtained at the minimum. Subsequently, a posterior ensemble can be generated at t_0:

$$\mathbf{E}_0^\star = \boldsymbol{x}_0^\star \mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{A}_0 \widetilde{\mathcal{H}}_\star^{-1/2}\mathbf{U}, \quad (8)$$

where $\mathbf{U}$ is an orthogonal matrix that is arbitrary but satisfies $\mathbf{U}\mathbf{1} = \mathbf{1}$ – meant to keep the posterior ensemble centered on the analysis – and $\boldsymbol{x}_0^\star = \overline{\boldsymbol{x}}_0 + \mathbf{A}_0\boldsymbol{w}^\star$.

The Gauss–Newton minimization scheme shown in Eq. (2) can easily be replaced by a quasi-Newton scheme that avoids the computation of the Hessian, or by a Levenberg–Marquardt algorithm that guarantees convergence of the minimization. These alternatives have been suggested and successfully tested in Bocquet and Sakov (2012). In the context of the standard models tested in Sects. 3 and 4, the nonlinearity is mild enough that a Levenberg–Marquardt scheme is unnecessary, and the Gauss-Newton scheme is very efficient.

This ends the part of the analysis step that is required to cycle the data assimilation scheme. An optional analysis step is required when a state estimation is desired at times $t_1,\ldots,t_L$, i.e. up to present time, or when a forecast to future times is desired. This additional step depends on the choice of the β_k and whether the DAWs are overlapping. In the simplest case, when observations are assimilated once and only once, this subsequent analysis takes the form of a forecast of the mean state $\boldsymbol{x}_k^\star = \mathcal{M}_{k\leftarrow 0}(\boldsymbol{x}_0^\star)$, or a forecast of the full ensemble if one is additionally interested in estimating forecast uncertainty $\mathbf{E}_k^\star = \mathcal{M}_{k\leftarrow 0}(\mathbf{E}_0^\star)$.

During the forecast step of the scheme cycle, not to be confused with the forecast of the analysis step we just mentioned, the ensemble is propagated for $S\Delta t$, with S an integer:

$$\mathbf{E}_S^\star = \mathcal{M}_{S\leftarrow 0}(\mathbf{E}_0^\star). \quad (9)$$

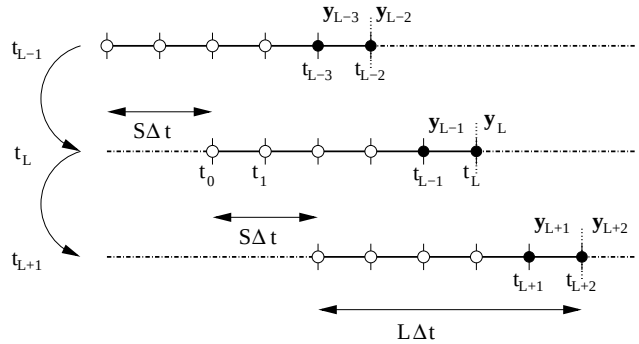


Fig. 1. Chaining of the SDA IEnKS cycles. The schematic illustrates the case $L = 5$ and a shift of $S = 2$ time intervals Δt is applied between two updates. The method performs a smoothing update throughout the window but only assimilates the newest observations vectors (that have not been already assimilated) marked by black dots. Note that the time index of the dates and the observations are absolute, not relative, for this schematic.

If the optional analysis step implied forecasting the ensemble to or beyond t_S, then there is no need to forecast it again. This ensemble at t_S will form the background for the next analysis.

A typical chaining of the analysis and forecast steps is schematically displayed in Fig. 1.

A pseudo-code of the IEnKS is displayed in Algorithm 1. It does not show the optional analysis step, since the cycling of data assimilation does not depend on it. It is the same as the one presented in Bocquet and Sakov (2013), except that here it is given in the general case, $1 \le S \le L$, rather than the specific case $S = 1$. The pseudo-code accounts for the possible use of inflation (lines 20, 21).

In summary, the IEnKS solves the variational problem of 4D-Var in the ensemble range. Because the variational problem is solved in a reduced space, there is no need for the adjoint evolution and observation models. The IEnKS generates and propagates the posterior perturbations following the scheme of the ensemble Kalman filter. As such, it uses sampled *errors of the day*.

2.1.2 Single and multiple assimilation of observations

There are some degrees of freedom in the choice of L, S and the $\{\beta_k\}_{1\le k\le L}$. Let us just mention a few legitimate choices.

Firstly, for any L and S, such that $1 \le S \le L$, the most natural choice for the $\{\beta_k\}_{1\le k\le L}$ is $\beta_k = 1$ for $k = L - S + 1,\ldots L$, and $\beta_k = 0$ otherwise. That way, the observations are assimilated once and only once. We call this the single data assimilation scheme (SDA IEnKS). It is simple, and the optional analysis of the update step is merely a forecast of the analyzed state at t_0, or possibly a forecast of the full ensemble from t_0. When $S = L$, the DAWs do not overlap, but they do so when $S < L$. The chaining of the data assimilation cycles in the SDA case is displayed in Fig. 1.

Algorithm 1 A cycle of the lag-L / shift-S / MDA / bundle / Gauss-Newton IEnKS.

Require: t_L is present time. Transition model $\mathcal{M}_{k+1\leftarrow k}$, observation operators H_k at t_k. Algorithm parameters: ϵ, e, $j_{\max}$. $\mathbf{E}_0$, the ensemble at t_0, $\boldsymbol{y}_k$ the observation at t_k. λ is the inflation factor. $\mathbf{U}$ is an orthogonal matrix in $\mathbb{R}^{N\times N}$ satisfying $\mathbf{U}\mathbf{1}=\mathbf{1}$. β_k, $1\le k\le L$, are the observation weights within the DAW.

$j=0$, $\boldsymbol{w}=\mathbf{0}$
$\boldsymbol{x}_0^{(0)}=\mathbf{E}_0\mathbf{1}/N$
$\mathbf{A}_0=\mathbf{E}_0-\boldsymbol{x}_0^{(0)}\mathbf{1}^{\mathrm{T}}$
repeat
  $\boldsymbol{x}_0=\boldsymbol{x}_0^{(0)}+\mathbf{A}_0\boldsymbol{w}$
  $\mathbf{E}_0=\boldsymbol{x}_0\mathbf{1}^{\mathrm{T}}+\epsilon\mathbf{A}_0$
  for $k=1,\dots,L$ **do**
    $\mathbf{E}_k=\mathcal{M}_{k\leftarrow k-1}(\mathbf{E}_{k-1})$
    $\overline{\boldsymbol{y}}_k=H_k(\mathbf{E}_k)\mathbf{1}/N$
    $\mathbf{Y}_k=(H_k(\mathbf{E}_k)-\overline{\boldsymbol{y}}_k)/\epsilon$
  end for
  $\nabla\widetilde{\mathcal{J}}=(N-1)\boldsymbol{w}-\sum_{k=1}^{L}\beta_k\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}(\boldsymbol{y}_k-\overline{\boldsymbol{y}}_k)$
  $\widetilde{\mathcal{H}}=(N-1)\mathbf{I}_N+\sum_{k=1}^{L}\beta_k\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k$
  Solve $\widetilde{\mathcal{H}}\Delta\boldsymbol{w}=\nabla\widetilde{\mathcal{J}}$
  $\boldsymbol{w}:=\boldsymbol{w}-\Delta\boldsymbol{w}$
  $j:=j+1$
until $||\Delta\boldsymbol{w}||\le e$ **or** $j\ge j_{\max}$
$\mathbf{E}_0=\boldsymbol{x}_0\mathbf{1}^{\mathrm{T}}+\sqrt{N-1}\mathbf{A}_0\widetilde{\mathcal{H}}^{-\frac{1}{2}}\mathbf{U}$
$\mathbf{E}_S=\mathcal{M}_{S\leftarrow 0}(\mathbf{E}_0)$
$\overline{\boldsymbol{x}}_s=\mathbf{E}_S\mathbf{1}/N$
$\mathbf{E}_S:=\overline{\boldsymbol{x}}_S\mathbf{1}^{\mathrm{T}}+\lambda\left(\mathbf{E}_S-\overline{\boldsymbol{x}}_S\mathbf{1}^{\mathrm{T}}\right)$

For very long data assimilation windows, the use of *multiple assimilation* (or *splitting*) of observations, denoted MDA in the following, can prove numerically efficient (Bocquet and Sakov, 2013). An observation vector $\boldsymbol{y}$ is said to be assimilated with weight β ($0\le\beta\le 1$) if the following Gaussian observation likelihood is used in the analysis:

$$p(\boldsymbol{y}^{\beta}|\boldsymbol{x})=\frac{e^{-\frac{\beta}{2}(\boldsymbol{y}-H(\boldsymbol{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{y}-H(\boldsymbol{x}))}}{\sqrt{(2\pi/\beta)^d|\mathbf{R}|}}, \tag{10}$$

where $|\mathbf{R}|$ is the determinant of $\mathbf{R}$. The upper index of $\boldsymbol{y}^{\beta}$ refers to the partial assimilation of $\boldsymbol{y}$ with weight β. The prior errors attached to the several occurrences of one observation are *chosen* to be independent. In that light, the $\{\beta_k\}_{1\le k\le L}$ are merely the weights of the observation vectors $\{\boldsymbol{y}_k\}_{1\le k\le L}$ within the DAW. Statistical consistency necessitates that a unique observation vector is assimilated in such a way that the sum of all its weights in the data assimilation experiment is 1. For instance, if $1=S\le L$, consistency requires that $\sum_{k=1}^{L}\beta_k=1$. In a more general case in which the observation vectors have the same number of non-zero weights, L is a multiple of $S: L=QS$, where Q is an integer. As a result, consistency requires $\sum_{q=0}^{Q-1}\beta_{Sq+l}=1$ with $l=1,\dots,S$.

In the MDA case (except the SDA subcase) the optional analysis step is more complex since it requires re-weighting the observations within the DAW to obtain the correct analyses for states t_1 to t_L and beyond. More details that are not directly relevant to this study can be found in Bocquet and Sakov (2013).

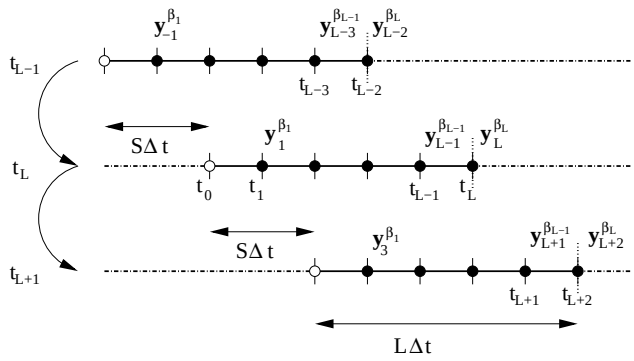


Fig. 2. Chaining of the MDA IEnKS cycles. The schematic illustrates the case $L=5$, and $S=2$. The method performs a smoothing update throughout the window potentially using all observations within the window (marked by black dots), except for the first observation vector assumed to be already entirely assimilated. Note that the time index for the dates and the observations are absolute for this schematic, not relative.

Note that when the constraint $\sum_{k=1}^{L}\beta_k=1$ is not satisfied, the underlying smoothing probability density function (pdf) will not be the one targeted, but, with well chosen $\{\beta_k\}_{1\le k\le L}$, could be a power of it (Bocquet and Sakov, 2013).

These MDA approaches are mathematically consistent in the sense that they are demonstrated to be correct in the linear model, Gaussian statistics case. An heuristic argument based on Bayesian ideas justifies the use of the method in the nonlinear case (Bocquet and Sakov, 2013).

The chaining of the data assimilation cycles in the MDA case is displayed in Fig. 2.

In the experimental Sects. 3 and 4, both SDA and MDA schemes will be used.

2.2 Augmented state formalism

We wish to estimate a set of model parameters $\boldsymbol{\theta}\in\mathbb{R}^P$ along with the state variables. To do so, the state space is augmented from $\boldsymbol{x}\in\mathbb{R}^M$ to a vector

$$\mathbf{z}=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{\theta}\end{pmatrix}\in\mathbb{R}^{M+P}, \tag{11}$$

of the joint state and parameter space. From the mathematical point of view, the analysis step of the IEnKS is unchanged.

As is usual in a parameter estimation context, a forward model needs to be introduced for the parameters. This model could be, for instance, the persistence model ($\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k$), or some jittering such as a Brownian motion, could be assumed ($\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+\boldsymbol{\epsilon}_k$). Depending on the constraints on the parameters, this jittering could also be constrained.

Technically, there is nothing more in the joint state and parameter IEnKS than in the state IEnKS. As opposed to the EnKF and EnKS, the objective of the joint state and parameter IEnKS is not to build covariances to help estimate hidden parameters, but instead to minimize a cost function that depends on the full augmented state. In a strongly nonlinear context, this approach could prove superior to the standard EnKF and EnKS.

As mentioned in the introduction, the estimation of model parameters within 4D-Var requires the adjoint model. Besides, the computation of the derivative of the cost function with respect to the parameters in terms of the adjoint field can be tedious. Parameter estimation with the IEnKS avoids this time-consuming task.

A potential advantage of the IEnKS over 4D-Var is that the errors of the day are by construction estimated within the IEnKS for all types of variables or parameters, whereas the 4D-Var modeling of background statistics of heterogeneous variables and parameters can be complex (see, for instance, Elbern et al. (2007), relating the modeling of inter-species correlation in a 4D-Var applied to air quality, or Montmerle and Berre (2010) in a meteorological convective scale context).

Similarly to state estimation, joint state and parameter estimation with the IEnKS in theory combines appealing features of both variational and ensemble Kalman filtering techniques. The purpose of the following numerical exploration is to investigate whether this holds true in experiments with low-order models.

3 Numerical experiments with the Lorenz-95 model

The Lorenz-95 one-dimensional model (Lorenz and Emmanuel, 1998) represents a mid-latitude zonal circle of the global atmosphere. It has $M = 40$ variables $\{x_m\}_{m=1,\dots,M}$. Its dynamics is given by the following set of ordinary differential equations:

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F\,, \tag{12}$$

for $m = 1,\dots,M$, and the domain is periodic (circle-like). F is chosen to be 8 so that the dynamics is chaotic and has 13 positive Lyapunov exponents. A time step of $\Delta t = 0.05$ is meant to represent a time interval of 6 h in the real atmosphere. Unless otherwise stated, the time interval between each observational update will be $\Delta t = 0.05$, meant to be representative of a data assimilation cycle of global meteorological models. With such a value for Δt, the data assimilation system is considered weakly nonlinear, leading to statistics of errors weakly diverging from Gaussianity. This model is integrated using the fourth-order Runge–Kutta scheme with a time step of 0.05.

3.1 Setup

Twin experiments are conducted. The truth is represented by a free model run (nature run), meant to be tracked by the data assimilation system. The system is assumed to be fully observed ($d = 40$) every Δt, so that $\mathbf{H}_k = \mathbf{I}_d$, with the observation error covariance matrix $\mathbf{R}_k = \mathbf{I}_d$. The related synthetic observations are generated from the truth, and perturbed according to the same observation error prior. The performance of a scheme is measured by the temporal mean of a root mean square difference between a state estimate ($\boldsymbol{x}^a$) and the truth ($\boldsymbol{x}^t$). Typically, one averages the following analysis root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(x_m^a - x_m^t\right)^2} \tag{13}$$

over the data assimilation cycles. When this RMSE concerns the system state at present time, i.e., the state at the end of the DAW, we call it the *filtering RMSE*. When this RMSE concerns the state defined $L\Delta t$ in the past, i.e., at the beginning of the DAW, we call it the *smoothing RMSE*. All data assimilation runs will extend over 10^5 cycles after a burn-in period of 5×10^3 cycles. This guarantees a sufficient convergence of the error statistics.

Unless otherwise stated, the size of the ensemble used with the ensemble methods will be $N = 20$, which is greater than the size of the unstable subspace, and, in the case of this model, makes localization unnecessary.

In this context, we have chosen to implement the inflation using the finite-size counterparts of the filters/smoothers (Bocquet et al., 2011). For this model, except in quasi-linear conditions ($\Delta t \sim 0.01$), this inflation leads to performances that are quantitatively very close to the same filter/smoother with optimally tuned uniform inflation (Bocquet et al., 2011; Bocquet and Sakov, 2012). In the following methods like EnKF/IEnKS/EnKS should be understood as EnKF/IEnKS/EnKS with optimally tuned uniform inflation, and will actually be implemented with a single run of the finite-size variants, i.e. EnKF-N/IEnKS-N/EnKS-N, which is much more economical. Any reader not interested in implementing the finite-size IEnKS (whose pseudo-code is presented in Algorithm 2), or IEnKS-N, can alternatively optimally tune the uniform inflation of an EnKF/IEnKS/EnKS to attain very similar results.

3.2 New experiments with the IEnKS

This section is meant to recall and extend to 4D-Var and the case $S = L$ several numerical tests of Bocquet and Sakov (2013), before considering joint state and parameter estimation. The following five systems are compared:

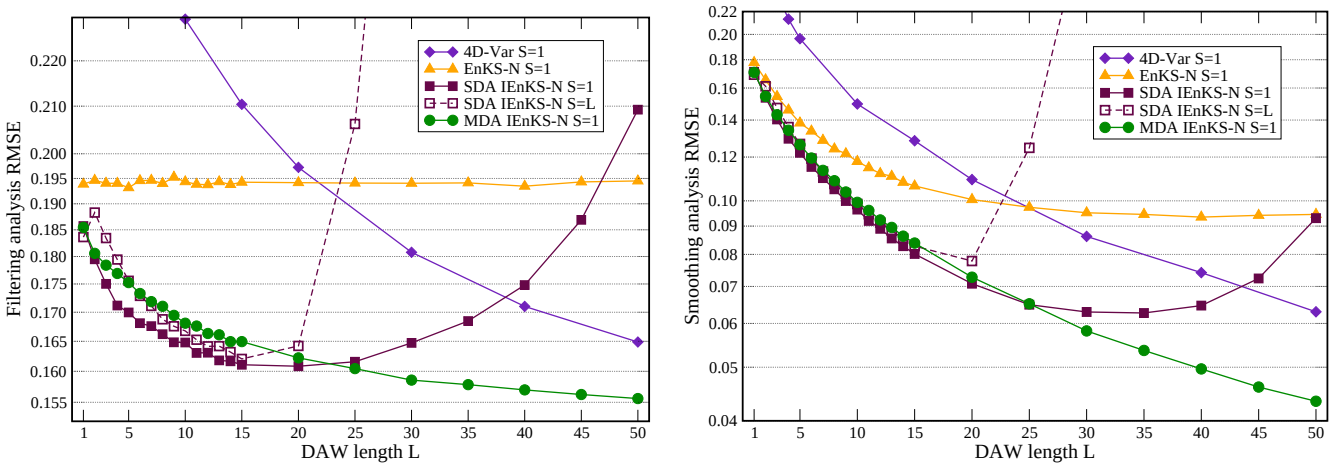


Fig. 3. Comparison of the filtering (left) and smoothing (right) performance of the SDA IEnKS, MDA IEnKS, the EnKS and 4D-Var, in weakly nonlinear conditions corresponding to $\Delta t = 0.05$.

Algorithm 2 A cycle of the lag-L / shift-S / MDA / bundle / Gauss-Newton IEnKS-N. Same as algorithm 1 with the exception of the following lines:

Require: Same requirements as algorithm 1. $\varepsilon_N = 1$.
12: $\nabla\widetilde{\mathcal{J}} = N\frac{\boldsymbol{w}}{\varepsilon_N+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}} - \sum_{k=1}^{L}\beta_k\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}(\boldsymbol{y}_k-\overline{\boldsymbol{y}}_k)$
18: $\widetilde{\mathcal{H}} = N\frac{(\varepsilon_N+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w})\mathbf{I}_N-2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}}{(\varepsilon_N+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w})^2} + \sum_{k=1}^{L}\beta_k\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k$
19: $\mathbf{E}_0 = \boldsymbol{x}_0\mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{A}_0\widetilde{\mathcal{H}}^{-\frac{1}{2}}\mathbf{U}$
20: $\mathbf{E}_S = \mathcal{M}_{S\leftarrow 0}(\mathbf{E}_0)$
21:

- The SDA IEnKS, $S = 1$.
- The MDA IEnKS, $S = 1$. The $\{\beta_k\}_{1\le k\le L}$ are chosen to be uniform in the DAW and constant in time.
- The SDA IEnKS, with S equal to the length of the DAW $S = L$, so that the DAWs do not overlap. This approach is meant to be computationally economical, and is much more economical than the quasi-static case $S = 1$, since there is no overlapping of DAWs.
- The standard ensemble Kalman smoother (EnKS), with $S = 1$. The standard ensemble Kalman smoother has been defined in Evensen and van Leeuwen (2000); Evensen (2003, 2009); Cosme et al. (2012).
- 4D-Var with a shift $S = 1$, corresponding to overlapping DAWs and quasi-static conditions. The gradient is obtained by finite differences, which is affordable and precise enough in this small dimensional context. The performance of 4D-Var strongly depends on the background statistics. Since the correlations in the Lorenz-95 system are rather short-ranged, the **B**-matrix is chosen diagonal. The performance of 4D-Var does not vary much if we introduce some correlation and off-diagonal terms. However, the scaling of the **B**-matrix is crucial in this context (Kalnay et al., 2007). The longer the DAW is, the smaller the scaling factor should be, since the first guess becomes more accurate. For each experiment, we tuned this scaling so as to obtain the best filtering analysis RMSE.

To avoid tuning inflation, the finite-size variants of the filters and smoothers are employed (SDA IEnKS-N, MDA IEnKS-N, EnKS-N). All EnKF and EnKS, and their finite-size variants in this article are based on the ensemble transform square root Kalman filter (Bishop et al., 2001; Hunt et al., 2007; Bocquet et al., 2011). These five data assimilation systems are compared in weakly nonlinear conditions ($\Delta t = 0.05$) chosen to roughly represent synoptic scale meteorology dynamics (Lorenz and Emmanuel, 1998), and more nonlinear conditions ($\Delta t = 0.20$ between updates). The time-averaged analysis RMSE is plotted in Fig. 3 for the former case, and in Fig. 4 for the latter case, as a function of the length of the DAW.

Let us first notice that the filtering performance of the EnKS is, by construction, given by that of the EnKF, whatever the length of the DAWs. This explains why the filtering RMSE of EnKS is constant, modulo statistical noise. When comparing the filtering performances of the EnKF/EnKS and 4D-Var, the conclusions of Kalnay et al. (2007) are reinforced. 4D-Var does not perform as well for short DAWs and performs better for long DAWs. In addition, we note that the same conclusion applies to the smoothing performance, even though the crossover point might be different.

Considering filtering as well as smoothing, the MDA IEnKS $S = 1$ significantly outperforms 4D-Var and the EnKF/EnKS in all regimes. The SDA IEnKS $S = 1$, also performs very well, but its performance wanes with longer DAWs, which is why the MDA IEnKS was introduced by

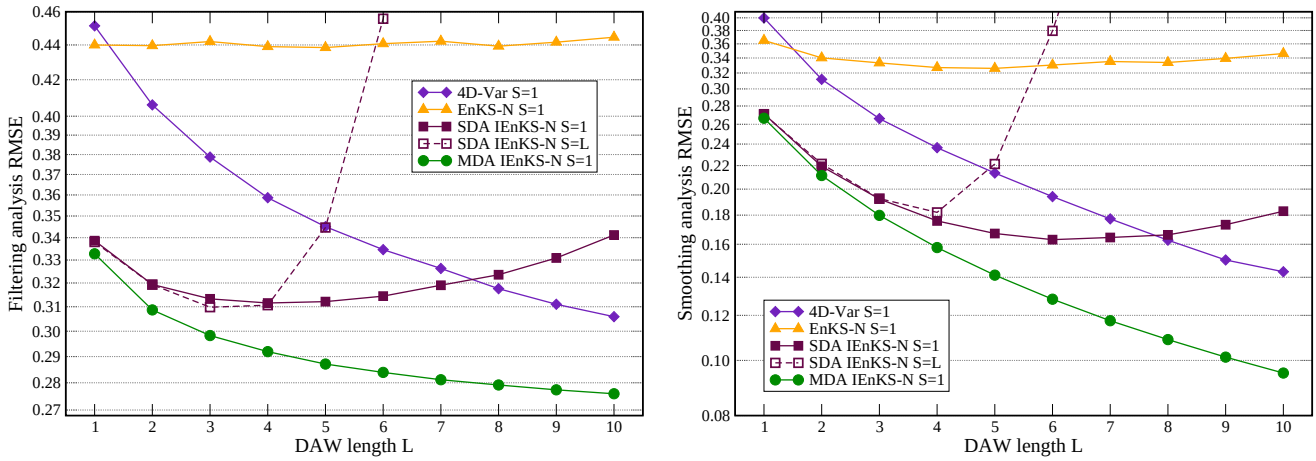


Fig. 4. Comparison of the filtering (left) and smoothing (right) performance of the SDA IEnKS, MDA IEnKS, the EnKS and 4D-Var, in nonlinear conditions corresponding to $\Delta t = 0.20$.

Bocquet and Sakov (2013). For very short DAWs ($L = 1, 2$ in the case $\Delta t = 0.05$), the performances of the SDA IEnKS $S = 1$ and MDA IEnKS $S = 1$ are equal ($L = 1$) or very close ($L = 2$). For intermediate DAW lengths, the SDA IEnKS $S = 1$ can slightly outperform MDA IEnKS $S = 1$. This is not surprising, since the SDA IEnKS algorithm is meant to be optimal for sufficiently short DAWs, whereas the MDA IEnKS algorithm is only guaranteed to be optimal in linear/Gaussian conditions.

Practically, in weakly nonlinear conditions ($\Delta t = 0.05$), the IEnKS $S = 1$ only requires one to two propagations of the ensemble within the DAW. Consistently, it was shown in Bocquet and Sakov (2013) that a linearized variant of the algorithm, requiring one propagation of the ensemble within the DAW to compute the sensitivity, performed just as well in these conditions. It is nevertheless tempting to check whether this cost can be reduced by using non-overlapping windows $S = L$, and performing the analysis every $L\Delta t$. This would divide the cost of model runs by L, but this effect might nevertheless be offset by an higher number of iterations required for the analysis.

Quite surprisingly, the SDA IEnKS $S = L$ performs very well for DAWs of length smaller than 0.80 (about twice the doubling time of the Lorenz-95 model). It is useless beyond that length, which was to be expected since the background at the beginning of the DAW results from a long forecast within the DAW, as opposed to a forecast of only Δt in the quasi-static $S = 1$ case.

In stronger nonlinear conditions, the variational methods (4D-Var and IEnKS) easily outperform the EnKF/EnKS. In particular, 4D-Var outperforms the EnKF/EnKS as soon as the the DAW reaches $L = 2$.

3.3 Joint state and forcing F estimation

A twin experiment is conducted in a situation where F is unknown. The true model (nature run) has forcing $F = 8$. The model used for assimilation and forecast has the initial value $F = 7$.

In addition to the state variables, the forcing parameter F will be estimated as well. Hence, the state vector $\boldsymbol{x} \in \mathbb{R}^M$ with $M = 40$ will be extended to the joint vector of size $M + P = 41$, with its 41st entry being the forcing parameter. The persistence model will be assumed for the evolution of the model parameter.

Because the filters and smoothers used here are all deterministic, the only source of stochasticity to generate the variability in F comes from the initialization of the ensemble. The forcing parameter of a member is initialized to $7 + \varepsilon$, where ε is independently drawn from a normal distribution of standard deviation 0.1. The augmented state IEnKS will be compared to several augmented state alternatives. Specifically, we shall consider in this experiment:

- The ensemble Kalman filter (EnKF).
- The ensemble Kalman smoother (EnKS) $S = 1$.
- The MDA IEnKS $S = 1$. The $\{\beta_k\}_{1 \le k \le L}$ are chosen to be uniform in the DAW and constant in time.
- 4D-Var with $S = 1$. The background's magnitude is tuned so as to minimize the global (on all 41 extended variables) RMSE.

To avoid tuning inflation, the finite-size variants – EnKF-N, EnKS-N, MDA IEnKS-N – are employed. The DAW length is varied in the MDA IEnKS and 4D-Var cases up to $L = 50$, before a degeneration of the performance sets in. In the EnKS case, the DAW length is varied up to $L = 100$, which

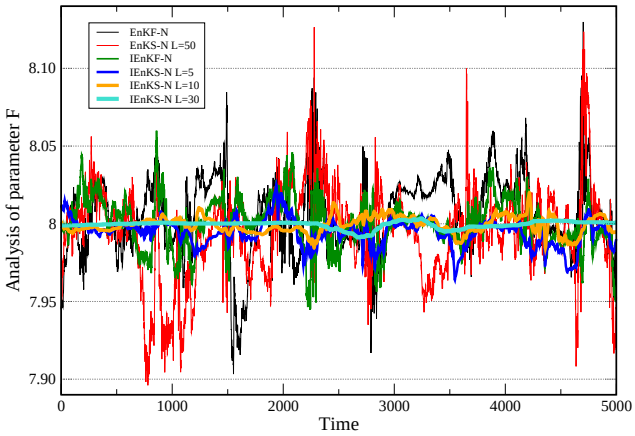


Fig. 5. Plot of the Lorenz-95 forcing parameter F as a function of the cycle index of the data assimilation experiment. F is estimated by several filters and smoothers with an ensemble of size $N = 20$. The forcing of the true model is $F = 8$. The MDA IEnKS for $L = 1, 5, 10$ and 30 is compared to the EnKF and the EnKS ($L = 50$). The finite-size variants of these methods are used: they do not require inflation and perform in this context as well as with optimally tuned inflation.

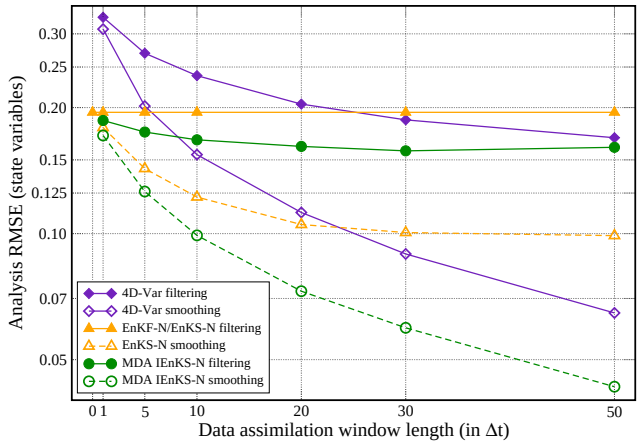


Fig. 6. Root mean square errors for the analysis of the state vector at present time (filtering) or the retrospective analysis of the state vector (smoothing) for the EnKF, EnKS and the IEnKS, in the case of the Lorenz-95 model, with $\Delta t = 0.05$.

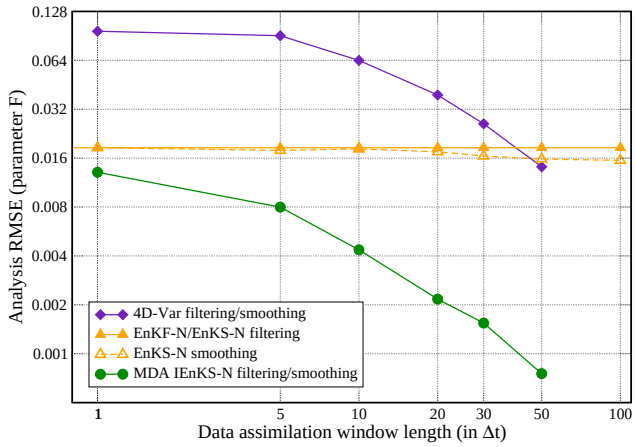


Fig. 7. Root mean square errors for the analysis of F at present time (filtering) or the retrospective analysis of F (smoothing) for the EnKF, EnKS and the IEnKS, in the case of the Lorenz-95 model, with $\Delta t = 0.05$.

corresponds to the optimal performance for the smoothing estimation of F by the EnKS.

The forcing parameter is plotted in Fig. 5, over a 5×10^3-cycle-long segment of the experiment. In the case of the EnKS, the smoothing estimator for F (at the beginning of the DAW) is plotted because it is better than the filtering estimate of F (at the end of the DAW). Because the persistence model is assumed for F, the smoothing and the filtering estimates of F are the same for the IEnKS and 4D-Var. In addition, because the true F is static, the smoothing and filtering RMSEs should coincide. From Fig. 5, it is clear that the IEnKS significantly outperforms the EnKF and the EnKS.

The time-averaged analysis root mean square errors (RMSEs) are computed over a much longer run of 10^5 cycles. The scores for the state variables are reported in Fig. 6. The filtering RMSEs (i.e., the RMSEs at present time) of the EnKF or of the EnKS for any L are, by construction, the same. The estimation of the forcing F is good enough that the performance is indistinguishable from the EnKF performance when $F = 8$ is known. Nevertheless, even in this weakly nonlinear regime, the IEnKS with $L \geq 1$ outperforms the EnKF and EnKS. Confirming the results of Bocquet and Sakov (2013), the gap in the smoothing performance between the EnKS and the IEnKS significantly increases as L increases. In this weakly nonlinear regime, the number of iterations required by the IEnKS is close to one, and its performance equals that of the linearized IEnKS (Bocquet and Sakov, 2013).

The scores for the estimation of the forcing parameter are reported in Fig. 7. By construction, the filtering performance of the EnKF and the EnKS at any L is the same, approximately 0.018. The parameter smoothing RMSE for the EnKS is approximately 0.015 and is optimal for $L \sim 100$. By construction, the analysis at present time and retrospective analysis of F by the IEnKS is the same. Even in the case $L = 1$, the so-called iterative ensemble Kalman filter (IEnKF) outperforms the EnKS with an RMSE of 0.013. With increasing L, this performance improves more and more, reaching the RMSE of 7.5×10^{-4} for $L = 50$.

The estimation of 4D-Var only becomes better than that of the EnKF for DAWs of length $L = 50$. This counter-performance can only be explained by a poor specification of the error covariance matrix. Indeed, the scaling of the background error statistics for the state variables should be

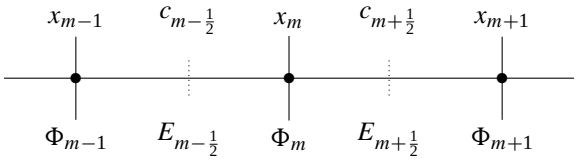
different from the scaling of the background error statistics for the parameter. However, the separate tuning of scalings requires additional work that the IEnKS does not require. This hypothesis will be checked in Sect. 5.

4 Numerical experiments with a coupled Lorenz-95 – tracer model

In this section we introduce a simple extension of the Lorenz-95 model with a tracer field advected by the Lorenz-95 field to represent an advective wind. This is meant to test the ability of the IEnKS to carry out joint state and parameter estimation in the dynamical context of an online atmospheric chemical model, with heterogeneous variables.

4.1 Extending the Lorenz-95 model

We shall think of the variables x_m of the Lorenz-95 as wind speed and direction variables defined on the circle. A tracer field $c_{m+\frac{1}{2}}$, $m = 1, \ldots, M = 40$ will be added to the model variables, for a total of 80 variables. These variables are defined on the circle using a C-grid. A schematic of the grid is shown below:



The tracer is advected by the *wind* field of the Lorenz-95 model. We have chosen to use the simple Godunov upwind scheme, which is positive and conservative. It is quite diffusive but this diffusion could be seen as a feature of the modeled physics. The equations read:

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F\,, \tag{14}$$

$$\frac{\mathrm{d}c_{m+\frac{1}{2}}}{\mathrm{d}t} = \Phi_m - \Phi_{m+1} - \lambda c_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}\,, \tag{15}$$

where
$$\Phi_m = x_m c_{m-\frac{1}{2}} \quad \text{if} \quad x_m \geq 0\,, \tag{16}$$

$$= x_m c_{m+\frac{1}{2}} \quad \text{if} \quad x_m < 0\,. \tag{17}$$

The tracer is emitted on the whole domain, and the emission fluxes are denoted $E_{m+\frac{1}{2}}$. It is deposited on the whole domain, using a simple scavenging scheme parameterized by a scavenging ratio λ. A stationary point of the dynamics is $x_m = F$ and $c_{m+\frac{1}{2}} = E_{m+\frac{1}{2}}/\lambda$. This provides orders of magnitude for the wind and concentration variables.

For simplicity, the emission flux will be made constant and uniform: $E_{m+\frac{1}{2}} \equiv E$. Obviously, however, a more complex setting with urban/rural/sea emission type and diurnal/nocturnal cycle could be chosen. The values of our reference simulation's parameters are $\lambda = 0.1$, and $E = 1$, so that the typical concentration value is 10.

The Courant–Friedrichs–Lewy (CFL) condition is almost always satisfied: the Lorenz-95 model variables $|x_m|$ very rarely exceed 15; by construction, one has $\Delta t = 0.05$ and $\Delta x = 1$, so that CFL $\leq 0.75 < 1$.

Free run simulations help one to understand some dynamical characteristics of the model. The model exhibits features of a realistic tracer model. For instance, consider two of the model's distinct trajectories in which the wind fields model trajectories are the same. It turns out that the concentrations $c_{m+\frac{1}{2}}$ of the two trajectories converge with each other. The positive part of the Lyapunov spectrum of the model is consistently very close to that of the Lorenz-95 model, with a number of positive Lyapunov exponents equal to 13, and a much broader negative part of the Lyapunov spectrum. However, we observed that the relaxation time of such two trajectories is quite long (typically $\tau = 10$), so that it seems difficult to break down the system into fast and slow dynamics.

A free run (after spin-up) is displayed in Fig. 8. The peaks of the tracer are correlated with the waves of the Lorenz-95, though not in an obvious way (see Sect. 5).

The causality and propagation of information in this model is special, and presumably similar to much more complex online atmospheric chemistry models. This impacts the effectiveness of data assimilation. For instance, measuring a tracer plume at t_0 (actually a peak in this one-dimensional context) does not enable one to detect a swift change in the local wind at t_0. Only future observations of the tracer concentrations will enable a diagnosis of this change in the local wind. As a consequence, variational schemes such as 4D-Var and the IEnKS that work over larger DAWs appear to be ideal tools in this context.

4.2 Numerical tests

We have performed data assimilation tests of the IEnKS using this model in order to estimate winds and concentrations, and unknown parameters F and E. Initially, we had carried out the same test but estimating E and λ instead of F and E. Parameters E and λ are typical of the kind one would like to control in an atmospheric chemistry model to improve forecast and re-analysis, when they are not themselves the focus of interest (Bocquet, 2012). The results were quite similar to those presented here. Yet, because the deposition and emission are antagonistic processes, the inverse problem of estimating them is very ill posed, requiring a specific prior distribution for those two parameters. In the absence of such strongly constraining prior, 4D-Var's performance would be hampered. That is why we choose to estimate F and E instead.

One of the potential difficulties in data assimilation with this model is the positivity of the concentrations $c_{m+\frac{1}{2}} \geq 0$ and of the parameters $F \geq 0$ and $E \geq 0$. This problem can be dealt with straightforwardly with 4D-Var, since the positivity of the variables can be enforced by the minimizer, or by a change of variables that is easy to implement in this context.

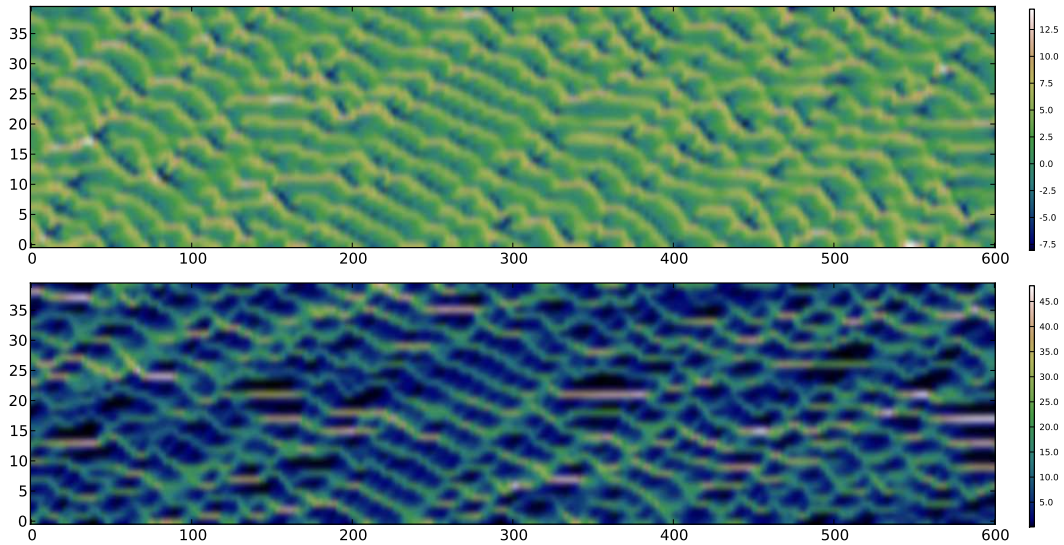


Fig. 8. Time evolution of the wind (top) and concentration (bottom) fields of the coupled Lorenz-95 – tracer model.

The problem is more severe with the EnKF, since the Best Linear Unbiased Estimator (BLUE) analysis and the ensemble generation that are at the heart of the methods will generate negative concentrations or parameters. A simple but fairly effective trick is to perform clipping by setting all negative variables of the analysis to zero; this, however, is suboptimal and could also induce imbalances and harming positive biases. A more elegant solution is to perform an analytical anamorphosis (Cohn, 1997; Bocquet et al., 2010; Simon and Bertino, 2012), so that the BLUE analysis and the ensemble generation are carried out in a space where the variables are defined on $\mathbb{R}$ and their statistics are closer to a Gaussian. For instance, in our case one could perform the state augmentation using the extended state vector:

$$\mathbf{z} = [x_1, \ldots x_M, \ln c_{\frac{1}{2}}, \ldots, \ln c_{M-\frac{1}{2}}, \ln F, \ln E]^{\mathrm{T}}. \tag{18}$$

Note that in practice this problem does not apply to F, which is well estimated and close to $F = 8$ because of a strong sensitivity of the model to F. The choice of $\ln(F)$ as the parameter to be estimated is only justified by the need of an homogeneous error metric for the two parameters.

Our numerical tests (EnKF as well as IEnKS) showed that the anamorphosis on the concentration variables is useless for improving precision, and can even lead to instability. This is at variance with the findings of Simon and Bertino (2012) who applied anamorphosed analysis on a 1D ocean ecosystem model, and who also found benefit in using anamorphosis on the state variables. Choosing a more complex gamma or lognormal distribution for anamorphosis function would avoid favoring large concentration values, as does the instability-prone logarithm anamorphosis (L. Bertino, personal communication, 2013). Aside from the choice of the anamorphosis function, this difference can also be explained by the fact that static anamorphosis is more efficient on distributions that are not too dynamical.

In addition, we found that occurrences of negative concentrations in the analysis and the posterior ensemble are extremely rare in the present case. By contrast, we found that the anamorphosis on E is useful and avoids instabilities. Because parameters F and E are not observed, their anamorphosis is a mere change of variables (as in 4D-Var) that does not require much work. Therefore, in the following the extended state vector will be

$$\mathbf{z} = [x_1, \ldots, x_M, c_{\frac{1}{2}}, \ldots, c_{M-\frac{1}{2}}, \ln F, \ln E]^{\mathrm{T}}. \tag{19}$$

A twin experiment, similar to that described in Sect. 3, is performed with $\Delta t = 0.05$. The winds and the concentrations are fully observed, with $\mathbf{R}_d = \mathbf{I}_d$, $d = M = 40$, in the wind observation space as well as in the tracer concentration space. The observations are generated from the truth and perturbed according to these error statistics. All runs are performed over 10^5 cycles after a burn-in period of 5×10^3 cycles. The following methods are compared:

- The SDA IEnKS $S = 1$.
- The MDA IEnKS $S = 1$. The $\{\beta_k\}_{1 \le k \le L}$ are chosen to be uniform in the DAW and constant in time.
- The EnKS with $S = 1$.
- 4D-Var with $S = 1$, corresponding to overlapping windows, quasi-static conditions. The scaling of the background is tuned so as to minimize the global (on all 82 extended variables) RMSE.

To avoid tuning inflation, the finite-size variants are employed: SDA/MDA IEnKS-N and EnKS-N.

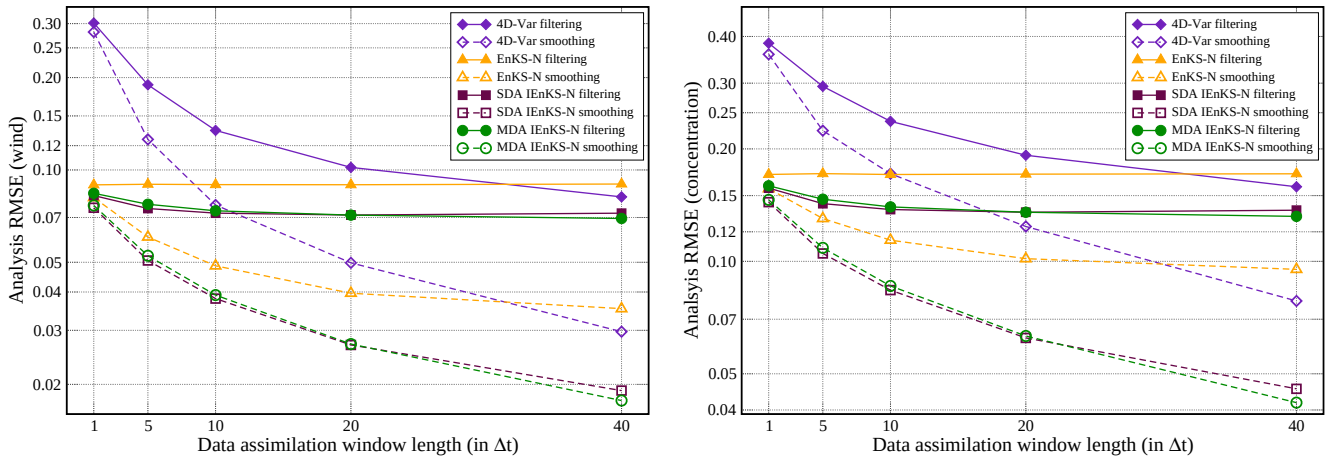


Fig. 9. Mean filtering and smoothing analysis root mean square errors of the wind variables (left) and concentration variables (right) of the online tracer model, as a function of the DAW length for the IEnKS (finite-size variant), the EnKF/EnKS, and 4D-Var (with optimal inflation of the prior).

The time-averaged analysis RMSEs on the wind and concentration variables are plotted in Fig. 9 as a function of the DAW length. Both the mean filtering and smoothing RMSEs are reported. Again, the results are consistent with those of Kalnay et al. (2007). 4D-Var is not as precise as the EnKF/EnKS for short DAWs ($L \leq 20$), but it outperforms the EnKF/IEnKF for large DAWs, in both filtering and smoothing. Moreover, the IEnKS significantly outperforms the EnKS/EnKF in all regimes and for both filtering and smoothing. In terms of performance, the difference between the SDA IEnKS and the MDA IEnKS is very similar to that reported in Sect. 3. However, the RMSE differences are much weaker, which may be explained by the doubled number of observations.

The RMSEs of the logarithm of the two parameters, i.e.,

$$\mathrm{RMSE} = \sqrt{\frac{1}{2}(\ln F^a - \ln F^t)^2 + \frac{1}{2}(\ln E^a - \ln E^t)^2}, \qquad (20)$$

where $F^t = 8$ and $E^t = 1$ are plotted in Fig. 10 as a function of the DAW length. The filters and smoothers perform significantly better than 4D-Var. The EnKF/EnKS and 4D-Var remain quite far from the performance of the SDA and MDA IEnKS. This shows that smoothing over a large window and flow-dependent error statistics are both crucial for the parameter estimation.

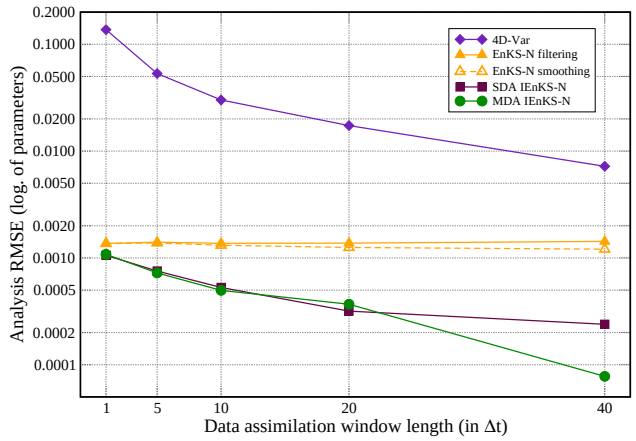


Fig. 10. Mean filtering and smoothing analysis root mean square errors of the two parameters of the online tracer model, as a function of the DAW length for the IEnKS (finite-size variant), the EnKF/EnKS, and 4D-Var (with optimal inflation of the prior).

5 Discussion

One of the possible limitations of the IEnKS is the potentially large average number of iterations. The number of required ensemble propagations could be the most costly part of the algorithm for complex high-dimensional models. We have seen in Sect. 3.2 in the context of the Lorenz-95 model that with non-overlapping windows ($S = L$), at moderate S ($1 \leq S \leq 15$), the IEnKS is performing well, and better that the EnKF/EnKS and 4D-Var. In the case of weak nonlinearity $\Delta t = 0.05$, only one iteration of the minimization is required on average for the computation of the sensitivities $\mathbf{Y}_{k,(j)}$. Additionally, accounting for the propagation of the ensemble of the (ensemble) forecast step, an average of two propagations of the ensemble through the DAW is required. A further exploration of the computational performance of the IEnKS is out of scope of this article, but it seems quite promising for the success of the IEnKS with complex models.

In the numerical experiments, parameters F and E were chosen to be static. This type of parameters is frequently modelled in geophysical systems. Furthermore, they make 4D-Var and the IEnKS with large DAW ideal tools. When the

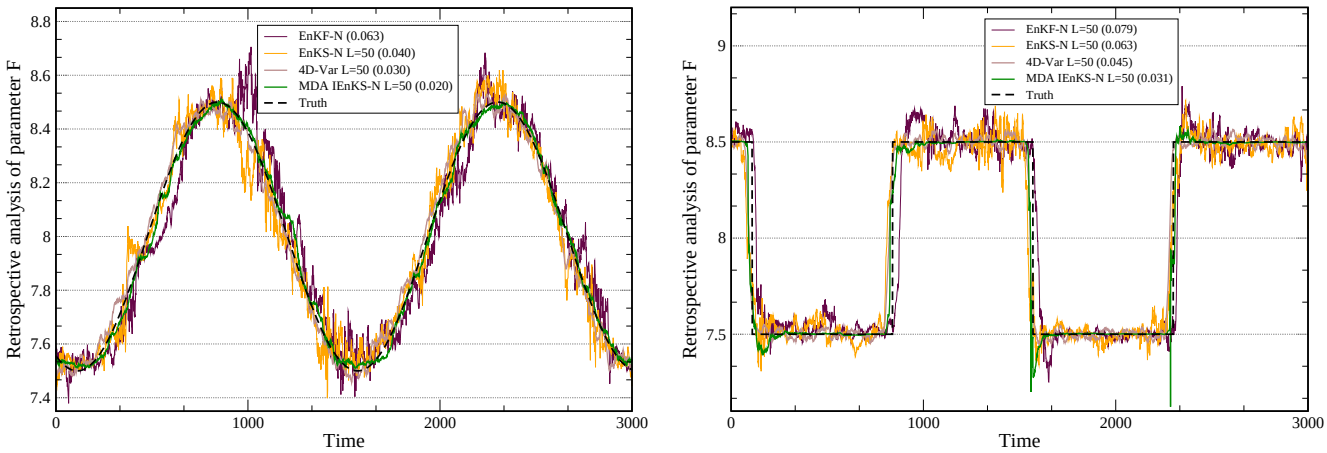


Fig. 11. Comparison of the retrospective analysis of F using the EnKF, the EnKS $L = 50$, 4D-Var $L = 50$ and the MDA IEnKS $L = 50$, when the truth (black dashed line) evolves as a sinusoid (left), and a step-wise function (right).

parameters evolve in time, the variational methods may not perform as well as the EnKF, EnKS and IEnKS with smaller DAWs. In particular, the persistence model for the parameters becomes imperfect. We have repeated the same experiment as in Sect. 3.3, but with F varying in time according to a sinusoid and a step-wise function, within the interval $[7.5; 8.5]$, with a period of one year (1456 time units of the Lorenz-95 model). Not only is model error made intrinsic by incorporating parameter F in the control variables as has been done so far, but model error also becomes extrinsic because the assumed persistence model for F is wrong (permanently in the sinusoid case and intermittently in the step-wise case).

Some results are displayed in Fig. 11. The evolution of the retrospective analysis of F is shown for the EnKF-N, the EnKS-N $L = 50$, the MDA IEnKS-N $L = 50$ $S = 1$, and 4D-Var $L = 50$ $S = 1$. The RMSEs are indicated in parenthesis in the legends. Although the IEnKS-N $L = 50$ remains the best performer in both cases, the gap in performance is narrower, because of the incorrect persistence assumption within the DAW. Let us remark that, in these cases, the RMSE of the retrospective analysis of the IEnKS is different from the RMSE of the filtering analysis because the truth that serves as a point of comparison changes within the DAW. Note also that because of the imperfection of the persistence model, a multiplicative inflation of 1.01 of the ensemble anomalies has been applied to the finite-size methods since they are not meant to intrinsically account for extrinsic model error (Bocquet et al., 2011), whereas the EnKF requires an inflation of 1.05 here to account for both model and sampling errors.

The last and main point of the discussion is dedicated to the improvement of the 4D-Var background and its comparison to the IEnKS. The background error statistics that determine the prior of variational methods, such as 4D-Var and the IEnKS, have less impact with longer data assimilation

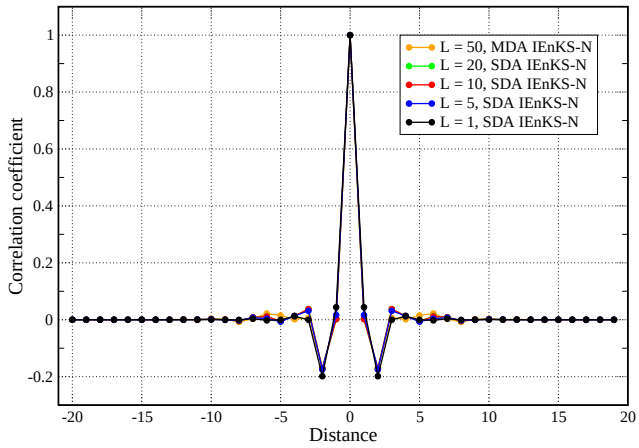


Fig. 12. Structure function of the correlation of the errors of the initial condition in the IEnKS, for several DAW lengths.

windows. These methods estimate the background state by a forecast of the analysis or of the posterior ensemble. Nevertheless, in the context of our low-order models, the IEnKS outperforms 4D-Var, especially in the joint state and parameter case. Therefore the difference should lie in the specification of the background error covariance matrix. It could be that the time-dependence of the background error statistics remains essential in the long DAWs length limit. Or it could be that the climatological statistics of the background in our implementation of 4D-Var is poorly specified.

To explore those hypotheses, we derived climatological statistics of the errors on the initial state of the DAW inferred from the SDA IEnKS-N. We first considered the Lorenz-95 model without parameter estimation. Since the system is statistically homogeneous, the error covariance matrix is circulant, so that is can be represented by a one-dimensional

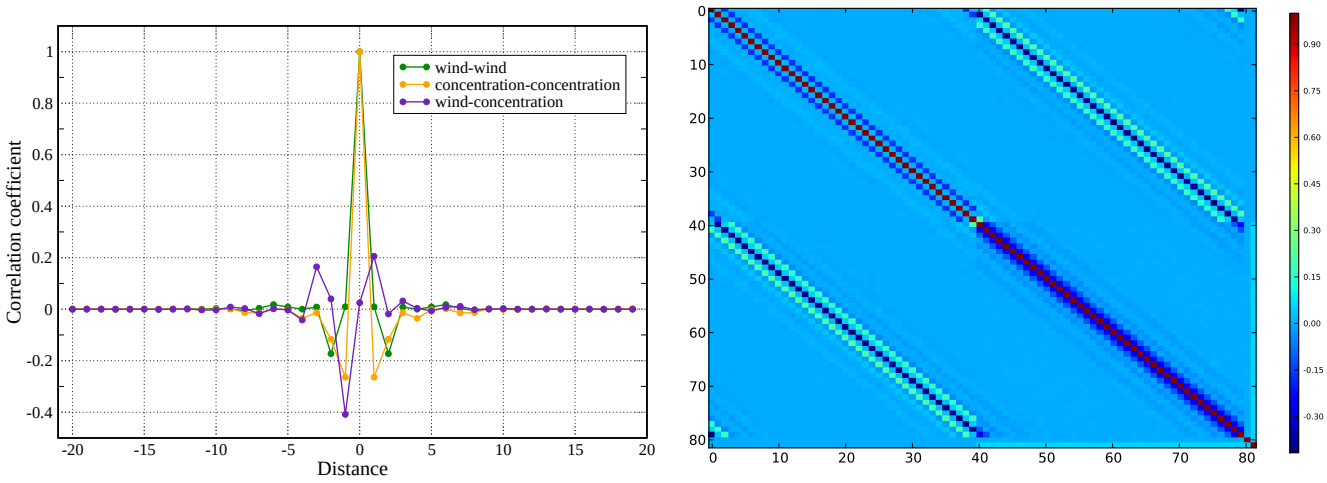


Fig. 13. Structure functions of the correlation of the errors of the initial condition from the IEnKS applied to the online tracer model. The structure functions for the wind correlation, for the concentration correlation, and for the cross-correlation between winds and concentrations are shown in the left panel. The full correlation matrix is displayed in the right panel.

structure function that depends on the distance between sites on the circle. The correlation structure function is plotted in Fig. 12 instead of the full correlation matrix. When L is varied, the differences are small, except in the case $L = 1$ that shows slightly modified next-to-nearest correlations.

The related covariance matrix, defined up to an optimally tuned scaling parameter, is used in 4D-Var as a new prior in place of the identity matrix. The new 4D-Var scores barely change from when using a covariance matrix proportional to the identity. This is consistent with the findings of Kalnay et al. (2007) who also tried, in a similar experimental context, to improve the performance of 4D-Var with a finer error covariance structure.

In a second experiment, we derived the climatological statistics of the errors on the initial extended state vector, in the Lorenz-95 case when F is unknown and estimated. The error covariance matrix turns out to be almost identical to the Lorenz-95 case with a fixed $F = 8$. The only difference is in the covariances that involve F. As expected, the correlation between the error on F and the error on any state variable is uniform. The covariance of the identity that was used in Sect. 3.3 is clearly not a good model for this case. Furthermore, the errors on F and the state variables are not homogeneous, so, again, choosing the error covariance matrix proportional to the identity matrix is not ideal. The climatological statistics of the errors have been inferred from the SDA IEnKS-N, when the state vector is augmented to incorporate F. It was done for each L because the ratio of the variances of the error on a typical state variable to the error on F is a non-uniform but increasing function of L. Using this procedure, we did not obtain real improvement in the state variable RMSE. However, the precision of the parameter estimation was remarkably improved to the level of the IEnKS-N. This shows that a fine specification of the background statistics is very helpful in the estimation of the static parameter F. Nevertheless, these statistics are static, and they do not significantly aid the estimation of the rapidly changing state variables.

In the last experiment, we applied the same procedure to the online tracer model based on the Lorenz-95 model. The error covariance matrix derived from the SDA IEnKS, for several L, entails significant covariances between the wind field and the concentration field. Again, because of the statistical homogeneity of the subsystems, one can represent the correlations of the winds and of the concentrations, and the cross-correlations between the winds and the concentrations, by using structure functions. These structure functions are displayed in Fig. 13, obtained from the IEnKS-N, $L = 20$.

We believe that the structure function of the cross-correlations is non-symmetric because of the preferred orientation of the winds. The waves in the Lorenz-95 preferentially travel westward and create fronts of tracer on one preferred side of the wave, yielding a non-trivial cross-correlation structure function. For 4D-Var, results similar to those from the previous experiment were obtained. Using the climatological priors, the errors of the state variables barely reduce. The fine correlations that build between the errors of the wind and concentration variables are dynamical and seem to be of little use when averaged in the climatological background error covariance matrix. However, the parameters are much better estimated. Nevertheless, unlike in the previous experiment, they do not quite match the precision of the IEnKS. For instance, in the $L = 1$ case, the 4D-Var RMSE of the logarithm of the parameters is reduced from 1.5×10^{-1} to 1.3×10^{-2}, but is still far from 1.0×10^{-3} of IEnKS.

6 Conclusion

In this article, the iterative ensemble Kalman smoother (IEnKS) has been explored numerically. Using the Lorenz-95 low-order model, it has been compared to the ensemble Kalman filter (EnKF) and the standard ensemble Kalman smoother (EnKS). It has also been compared to a 4D-Var, for a wide range of data assimilation window (DAW) lengths. The IEnKS systematically outperformed the EnKF, EnKS and 4D-Var. This conclusion holds true even when the background error covariance matrix of 4D-Var is better specified and tuned. The IEnKS has been extended to joint state and parameter estimation, using the augmented state formalism. Endowed with assets of 4D-Var and ensemble Kalman methods, IEnKS appeared to us as an ideal candidate method to tackle such problems.

It was applied to the joint estimation of the Lorenz-95 state vector and its forcing parameter F. The IEnKS outperformed the EnKF, EnKS and 4D-Var for a wide range of lengths of the DAW. In addition, the estimation of F was shown to be even more precise compared to the standard methods.

Motivated by future applications of the IEnKS to atmospheric chemistry models where the estimation of the forcings is crucial, we introduced an extension of the Lorenz-95 model, adding a tracer field advected by the Lorenz-95 field. Key parameters of the tracer emission and deposition were also meant to be estimated. Again, the IEnKS managed to finely estimate the parameters without any tuning.

A better specification of the error covariance matrix of 4D-Var (obtained from the IEnKS) led to a spectacular improvement in the estimation of the static parameters. Yet, it did not help in the improvement of joint estimation of the state variables. This stresses the importance of the time-dependence of the error covariance matrix for rapidly varying variables. By contrast, the IEnKS completely avoids the need to build any background error covariance matrix.

One way to account for model error is to parameterize this error and estimate the related parameters, as was suggested in this study. Another way is to implement a weak constraint formulation of the underlying variational problem. However, this formulation remains to be defined for the IEnKS, whereas it is already implemented in the standard EnKS.

Following this study, we are planning to test the IEnKS on a more complex low-order model with several reactive species and test the estimation of the concentration variables as well as some parameters such as kinetic constants. If this is successful, our eventual plan is to implement the method on a high-dimensional air quality model. However, the definition of a satisfying implementation of localization in the IEnKS context will first be needed (work in progress).

Acknowledgements. The authors thank the editor, Olivier Talagrand, the reviewers, Laurent Bertino and Takemasa Miyoshi, for their useful comments and suggestions. The authors also thank the organizers of the International Conference on Ensemble Methods in Geophysical Sciences held in Toulouse, France, from the 12th to the 16th of November 2012.

Edited by: O. Talagrand
Reviewed by: L. Bertino and T. Miyoshi

References

Aksoy, A., Zhang, F., and Nielsen-Gammon, J.: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, Mon. Weather Rev., 134, 2951–2969, 2006.

Barbu, A. L., Segers, A. J., Schaap, M., Heemink, A. W., and Builtjes, P. J. H.: A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe, Atmos. Environ., 43, 1622–1631, 2009.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Weather Rev., 129, 420–436, 2001.

Bocquet, M.: Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var, Q. J. R. Meteorol. Soc., 138, 664–681, doi:10.1002/qj.961, 2012.

Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, Nonlin. Processes Geophys., 19, 383–399, doi:10.5194/npg-19-383-2012, 2012.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Q. J. R. Meteorol. Soc., doi:10.1002/qj.2236, in press, 2013.

Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, 2010.

Bocquet, M., Wu, L., and Chevallier, F.: Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism, Q. J. R. Meteorol. Soc., 137, 1340–1356, doi:10.1002/qj.837, 2011.

Bowler, N., Flowerdew, J., and Pring, S.: Tests of different flavours of EnKF on a simple model, Q. J. R. Meteorol. Soc., 139, 1505–1519, doi:10.1002/qj.2055, 2013.

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, 2010.

Chen, Y. and Oliver, D. S.: Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother, Math. Geosci., 44, 1–26, 2012.

Cohn, S. E.: An Introduction to Estimation Theory, J. Meteor. Soc. Japan, 75, 257–288, 1997.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a Reduced-rank, square-root smoother for ocean data assimilation, Ocean Model., 33, 87–100, 2010.

Cosme, E., Verron, J., Brasseur, P., Blum, J., and Auroux, D.: Smoothing problems in a Bayesian framework and their linear Gaussian solutions, Mon. Weather Rev., 140, 683–695, 2012.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational

inversion, Atmos. Chem. Phys., 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007.

Evensen, G.: The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dynam., 53, 343–367, 2003.

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer-Verlag, 2nd Edn., 2009.

Evensen, G. and van Leeuwen, P. J.: An Ensemble Kalman Smoother for Nonlinear Dynamics, Mon. Weather Rev., 128, 1852–1867, 2000.

Fairbairn, D., Pring, S. R., Lorenc, A. C., and Roulstone, I.: A comparison of 4DVar with ensemble data assimilation methods, Q. J. R. Meteorol. Soc., 0, 0–0, in press, 2013.

Gu, Y. and Oliver, D. S.: An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation, SPE Journal, 12, 438–446, 2007.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., D. J. Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, Tellus A, 56, 273–277, 2004.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, 2007.

Jackson, C., Sen, M. K., and Stoffa, P. L.: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictsions, J. Climate, 17, 2828–2841, 2004.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4D-Var or ensemble Kalman filter?, Tellus A, 59A, 758–773, 2007.

Kazantsev, E.: Sensitivity of a shallow-water model to parameters, Nonlinear Anal. R. World Appl., 13, 1416–1428, doi:10.1016/j.nonrwa.2011.11.006, 2012.

Kondrashov, D., Sun, C., and Ghil, M.: Data assimilation for a coupled ocean atmosphere model. Part II: Parameter estimation, Mon. Weather Rev., 5, 5062–5076, 2008.

Le Dimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, Tellus A, 38, 97–110, 1986.

Liu, C., Xiao, Q., and Wang, B.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test, Mon. Weather Rev., 132, 3363–3373, 2008.

Liu, Y., Gupta, H. V., Sorooshian, S., Bastidas, L. A., and Shuttlewort, W. J.: Constraining land surface and atmospheric parameters of a locally coupled model using observational data, J. Hydrometeorol., 6, 156–172, 2005.

Lorenz, E. N. and Emmanuel, K. E.: Optimal sites for supplementary weather observations: simulation with a small model, J. Atmos. Sci., 55, 399–414, 1998.

Montmerle, T. and Berre, L.: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale, Q. J. R. Meteorol. Soc., 136, 1408–1420, 2010.

Posselt, D. J. and Bishop, C. H.: Nonlinear Parameter Estimation: Comparison of an Ensemble Kalman Smoother with a Markov Chain Monte Carlo Algorithm, Mon. Weather Rev., 140, 1957–1974, 2012.

Pulido, M. and Thuburn, J.: Gravity wave drag estimation from global analyses using variational data assimilation principles. Part II: A case study, Q. J. R. Meteorol. Soc., 132, 1527–1543, 2006.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, Q. J. R. Meteorol. Soc., 126, 1143–1170, 2000.

Raynaud, L., Berre, L., and Desroziers, G.: Objective filtering of ensemble-based background-error variances, Q. J. R. Meteorol. Soc., 135, 1177–1199, 2009.

Ruiz, J. J., Pulido, M., and Miyoshi, T.: Estimating model parameters with ensemble-based data assimilation: A Review, J. Meteorol. Soc. Japan, 91, 79–99, doi:10.2151/jmsj.2013-201, 2013.

Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, Tellus A, 62, 24–29, 2010.

Sakov, P., Oliver, D., and Bertino, L.: An iterative EnKF for strongly nonlinear systems, Mon. Weather Rev., 140, 1988–2004, 2012.

Simon, E. and Bertino, L.: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: application to a 1D ocean ecosystem model, J. Mar. Syst., 89, 1–18, 2012.

Talagrand, O. and Courtier, P.: Variational Assimilation of Meteorological Observation with the Adjoint Vorticity Equation. I: Theory, Q. J. R. Meteorol. Soc., 113, 1311–1328, 1987.

Vossepoel, F. C. and van Leeuwen, P. J.: Parameter Estimation Using a Particle Method: Inferring Mixing Coefficients from Sea Level Observations, Mon. Weather Rev., 135, 1006–1020, 2007.

Weir, B., Miller, R. N., and Spitz, Y. H.: Implicit Estimation of Ecological Model Parameters, Bull. Math. Biol., 75, 223–257, doi:10.1007/s11538-012-9801-6, 2013.

Wirth, A. and Verron, J.: Estimation of friction parameters and laws in 1.5D shallow-water gravity currents on the f-plane, by data assimilation, Ocean Dynam., 58, 247–257, 2008.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-Time Air Quality Forecasting, Part II: State of the Science, Current Research Needs, and Future Prospects, Atmos. Environ., 60, 656–676, doi:10.1016/j.atmosenv.2012.02.041, 2012.