Quarterly Journal of the Royal Meteorological Society *Q. J. R. Meteorol. Soc.* 142: 1075–1089, January 2016 B DOI:10.1002/qj.2711

RMetS
Royal Meteorological Society

Localization and the iterative ensemble Kalman smoother

M. Bocquet*

CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

*Correspondence to: M. Bocquet, CEREA, École des Ponts ParisTech, 6–8 Avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne 77455, Marne-la-Vallée Cedex, France. E-mail: bocquet@cerea.enpc.fr

The iterative ensemble Kalman smoother (IEnKS) is a data assimilation method meant for tracking the state of nonlinear geophysical models efficiently. It combines an ensemble of model states to estimate the errors similarly to the ensemble square-root Kalman filter, with a four-dimensional variational analysis performed within the ensemble space. As such, it belongs to the class of four-dimensional ensemble variational methods. It could require the use of localization of the analysis when the state-space dimension is high. However, its localization needs to be defined across time and to be as consistent as possible with the dynamical flow within the data assimilation window where the four-dimensional variational analysis is performed. We show that a Liouville equation governs the time evolution of the localization operator, which is linked to the evolution of the error correlations. It is argued that the time evolution of the localization operator depends strongly on the forecast dynamics. Using either covariance localization or domain localization, we propose and test several localization strategies meant to address the issue: (i) a static and uniform localization, (ii) propagation through the window of a restricted set of dominant modes of the error covariance matrix and (iii) the approximate propagation of the localization operator using covariant local domains that are moved in accordance with the dynamical flow. These schemes are illustrated with the one-dimensional Lorenz 40 variable model and with a two-dimensional barotropic vorticity model. In both cases, local analysis based on the covariant local domains leads to a systematic improvement of the data assimilation performance.

Key Words: iterative ensemble Kalman smoother; localization; ensemble variational method; ensemble Kalman filter; 4DEnVar; ensemble Kalman smoother

Received 27 May 2015; Revised 9 November 2015; Accepted 17 November 2015; Published online in Wiley Online Library 5 February 2016

1. Introduction

1.1. Localization in ensemble-based data assimilation methods

The key idea of ensemble-based data assimilation methods is to represent the analysis and forecast errors using a limited number of state vectors, which can be stored and propagated with limited computational resources. Following this approach, the ensemble Kalman filter (EnKF) and many variants thereof were proposed as *reduced-rank Monte Carlo* Kalman filters (Evensen, 2009, and references therein). Their efficiency in geophysical systems is due to the limited number of influential degrees of freedom within these systems compared with the dimension of the state space, so that error statistics can be represented with a restricted number of ensemble members. Nonetheless, the dynamics of high-dimensional chaotic models and the size of their unstable subspace may still require an ensemble of size much higher than can be afforded numerically.

The ensemble members are viewed as samples that can be used to estimate the state vector and the error covariance matrix, which are unknown. The limited size of the ensemble and the related rank-deficiency of the sampled error covariance matrix generate spurious correlations at long distances that affect the filter performance strongly. For this reason, the technique of *localization* has been introduced to regularize the analysis of the errors (Hamill *et al.*, 2001; Houtekamer and Mitchell, 2001; Evensen, 2003; Ott *et al.*, 2004). Localization can be performed by the Schur (Hadamard or point-wise) product of the empirical error covariance matrix with a limited-range regularizing correlation matrix. This technique is also called *tapering* or *covariance localization* (CL). Alternatively, one can perform local analyses using observations located near to the state variable to be updated. This technique is also called *local analysis* or *domain localization* (DL).

These localization techniques rely fundamentally on the vanishing correlations exhibited by geophysical systems at long distances. The implementations of the two classes of techniques differ significantly, from a mathematical as well as a numerical complexity standpoint. Shared properties and differences of the two classes of scheme are discussed by Sakov and Bertino (2011) and Nerger *et al.* (2012), for instance.

Many variants of these localization techniques have been proposed and validated for ensemble Kalman filters, depending

on their specific implementation. For instance, covariance localization has been applied to the serial ensemble square-root filter by tapering each scalar covariance individually (Anderson, 2012). Covariance localization has also been applied in hybrid systems using the so-called α control variable (Lorenc, 2003; Buehner, 2005). Domain localization can be implemented either by tapering of the ensemble anomalies or via inverse squared tapering of the observation-error variances (Evensen, 2003; Ott *et al.*, 2004). It has been chosen for the ensemble transform Kalman filter (ETKF: Hunt *et al.*, 2007) because the analysis is performed in ensemble space rather than state space. The stochastic ensemble Kalman filter can be localized either by covariance or domain localization. More recently, several schemes of adaptive localization, where an adequate localization length-scale is estimated, have been proposed (Anderson and Lei, 2013; Ménétrier *et al.*, 2015).

1.2. Localization in an ensemble variational context

Throughout this study, a perfect forecast model will be assumed. Nevertheless, this hypothesis does not prevent us from using the methods in a realistic imperfect model context. This assumption is often used in the four-dimensional variational (4D-Var) method (Le Dimet and Talagrand, 1986), as well as in the iterative ensemble Kalman smoother (IEnKS), which is the focus of this article (Bocquet and Sakov, 2014).

This IEnKS aims at a variational analysis within a data assimilation time window (DAW) using an ensemble of state trajectories within this window. The conceptual framework is that of ensemble variational (EnVar) methods (Liu *et al.*, 2008; Buehner *et al.*, 2010; Bocquet and Sakov, 2013; Desroziers *et al.*, 2014; Fairbairn *et al.*, 2014). This analysis requires the estimation of spatial and temporal covariances from the ensemble of trajectories within the DAW. The regularization of these space and time covariances requires localization in not only space but also time. The techniques developed for space localization in the context of the EnKF or of a hybrid EnKF–3D-Var are likely to be inadequate. Indeed, four-dimensional (4D) EnVar methods may require the estimation of covariances with significant time separation within the DAW. A straightforward reuse of a two- or three-dimensional localization scheme could become inadequate for long DAWs.

A proper 4D localization operator (Bishop and Hodyss, 2007; Bishop *et al.*, 2011) would ideally account for the dynamical flow. If covariances are estimated with an ensemble at the beginning of the DAW, t_0, they can be estimated throughout the DAW by the propagation of the ensemble within the DAW, leading to the computation of new sampled covariances at later times t_k. Assuming the ensemble of trajectories is rank-deficient, localization should be used at t_0 to regularize the sample background-error covariance matrix $\mathbf{P}_0$. Using covariance localization, i.e. applying a Schur product with a space localization matrix ρ estimated from climatological considerations, yields $\rho \circ \mathbf{P}_0$. However, localization should also be used at later times. Using the same climatological localization matrix, one obtains $\rho \circ \mathbf{P}_k$ at time t_k. Even though it can be considered physically sound, this regularization at later times is mathematically inconsistent with the regularization at t_0. Denoting by $\mathbf{M}_{k:0}$ the tangent linear model of the forecast model from t_0 to t_k, we have in general

$$\mathbf{M}_{k:0}\left(\rho \circ \mathbf{P}_0\right)\mathbf{M}_{k:0}^{\mathrm{T}} \neq \rho \circ \left(\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}\right). \tag{1}$$

In mathematical terms, the localization operator does not commute with the forecast model resolvent. This inconsistency was first discussed by Fairbairn *et al.* (2014) and pointed out in Bocquet and Sakov (2014) as an obstacle to the definition of a local IEnKS. It can be problematic if the variational problem is significantly nonlinear and several minimization iterations (*outer loop*) are required, a regime the IEnKS is meant to deal with.

1.3. Objective and outline

The goal of this methodological article is to investigate the principles, use and efficiency of localization in the 4D ensemble variational context of the iterative ensemble Kalman smoother, with relevance to any other four-dimensional EnVar method.

The principle and the algorithm of the global IEnKS, i.e. without localization, will first be recalled in section 2. The definition of a 4D localization operator and its consistency will be studied in section 3. The link between this operator and the system correlations and the impact of the system dynamics on the time integration of the operator will be discussed. Building on the outcome of this study, several strategies to implement 4D localization will be proposed. Some of them will lead to simple implementations described in section 4, which will be illustrated with the Lorenz 40 variables low-order model in section 5 and a two-dimensional barotropic vorticity model in section 6. Conclusions will be drawn in section 7.

2. The iterative ensemble Kalman smoother

The iterative ensemble Kalman smoother (IEnKS) has been introduced and justified through a Bayesian derivation that can be found in Bocquet and Sakov (2014). The IEnKS extends the so-called iterative ensemble Kalman filter of Sakov *et al.* (2012) to a 4D-Var analysis. The IEnKS has been named after the article by Bell (1994), who first described such an algorithm but in the Kalman filter context. The method is meant for not only smoothing but also filtering as well as forecasting. In the following, we describe the simplest variant of the algorithm, known as the single data assimilation scheme (SDA IEnKS), and we emphasize aspects of the scheme that will be relevant for localization. A more complete description of the variants of the IEnKS can be found in Bocquet and Sakov (2013, 2014).

2.1. The algorithm

Assume that batches of observations $\mathbf{y}_k \in \mathbb{R}^d$ are collected every Δt, at times t_k. 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, uncorrelated in time, with observation-error covariance matrices $\mathbf{R}_k$.

Let us first consider the analysis step of the scheme. It is performed over a window of L intervals of length Δt ($L\Delta t$ in time units). By convention, the time at the end of the DAW will be called t_L, while the time at the beginning of the DAW will be called t_0. At t_0 (i.e. $L\Delta t$ in the past), the prior distribution is estimated from an ensemble of N state vectors of $\mathbb{R}^M$: $\mathbf{x}_{0,[1]}, \ldots, \mathbf{x}_{0,[n]}, \ldots, \mathbf{x}_{0,[N]}$. Index 0 refers to time, while $[n]$ refers to the ensemble member index. They can be gathered into the ensemble matrix $\mathbf{E}_0 = [\mathbf{x}_{0,[1]}, \ldots, \mathbf{x}_{0,[N]}] \in \mathbb{R}^{M\times N}$. The ensemble can equivalently be given in terms of its mean $\overline{\mathbf{x}}_0 = (1/N)\sum_{n=1}^{N}\mathbf{x}_{0,[n]}$ and its (normalized) anomaly matrix $\mathbf{X}_0 = (1/\sqrt{N-1})[\mathbf{x}_{0,[1]} - \overline{\mathbf{x}}_0, \ldots, \mathbf{x}_{0,[N]} - \overline{\mathbf{x}}_0]$.

As in the ensemble Kalman filter, this prior is modelled as a Gaussian distribution of mean $\overline{\mathbf{x}}_0$ and covariance matrix $\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$, the first- and second-order sampled moments of the ensemble. The background-error covariance matrix 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 $N \ll M$. Hence, in the absence of localization, one seeks an analysis-state vector $\mathbf{x}_0$ in the ensemble space $\overline{\mathbf{x}}_0 + \mathrm{Vec}\left\{\mathbf{x}_{[1]} - \overline{\mathbf{x}}_0, \ldots, \mathbf{x}_{[N]} - \overline{\mathbf{x}}_0\right\}$, which can be written as $\mathbf{x}_0 = \overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^N$ is a vector of coefficients in ensemble space. The analysis will be defined as the most likely deterministic state of a cost function in the reduced space of the ensemble and will later be identified as the mean of the updated ensemble.

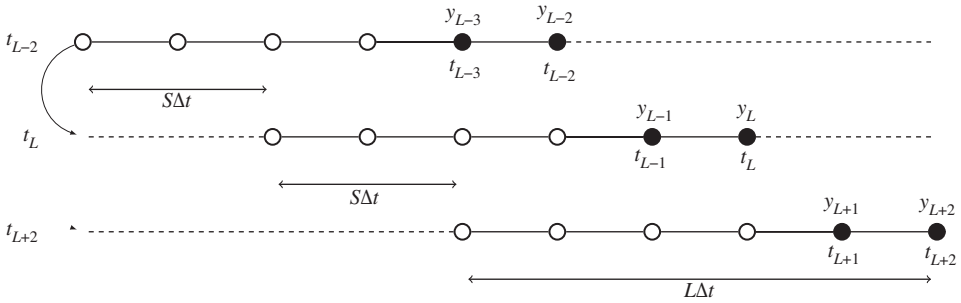

Figure 1. Cycling of the SDA IEnKS where $L = 5$ and $S = 2$, in units of Δt, the time interval between two updates. The method performs a smoothing update throughout the window but only assimilates the newest observation vectors (that have not been already assimilated), marked by black dots. Note that the time index of the dates and the observations is absolute for this schematic, not relative.

The forward model from t_k to t_{k+1} is given by the possibly nonlinear resolvent $\mathcal{M}_{k+1:k}$ defined by $x_{k+1} = \mathcal{M}_{k+1:k}(x_k)$. The variational analysis of the IEnKS over $[t_0, t_L]$ stems from a cost function. The restriction of this cost function to the ensemble space reads

$$\widetilde{\mathcal{J}}(\mathbf{w}) = \sum_{k=L-S+1}^{L} \frac{1}{2} \left\| \mathbf{y}_k - \mathcal{F}_{k:0}\left(\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}\right)\right\|^2_{\mathbf{R}_k} + \frac{1}{2}\|\mathbf{w}\|^2 , \quad (2)$$

where $\mathcal{F}_{k:0} = H_k \odot \mathcal{M}_{k:0}$, $\|\mathbf{z}\|^2_{\mathbf{R}_k} = \mathbf{z}^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{z}$, $\|\mathbf{w}\|^2 = \mathbf{w}^{\mathrm{T}}\mathbf{w}$ and $\mathcal{M}_{k:0}$ is the model resolvent from t_0 to t_k, with the $\odot$ symbol representing the composition of operators. The second term represents the prior in ensemble space (Hunt *et al.*, 2007). S batches of observations are assimilated in this analysis. The case $L = 0$, $S = 1$, where $\mathcal{M}_{0:0}$ is defined as the identity, corresponds to an ensemble transform variant of the maximum-likelihood filter of Zupanski (2005). Otherwise, one has $1 \leq S \leq L$ and a 4D variational analysis.

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

$$\mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} - \widetilde{\mathcal{H}}^{-1}_{(j)} \nabla \widetilde{\mathcal{J}}_{(j)}(\mathbf{w}^{(j)}) , \quad (3)$$

using the gradient $\nabla\widetilde{\mathcal{J}}$ and an approximation $\widetilde{\mathcal{H}}$ of the Hessian in ensemble space:

$$\mathbf{x}_0^{(j)} = \overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}^{(j)} , \quad (4)$$

$$\nabla\widetilde{\mathcal{J}}_{(j)} = \mathbf{w}^{(j)} - \sum_{k=L-S+1}^{L} \mathbf{Y}^{\mathrm{T}}_{k,(j)}\mathbf{R}_k^{-1}\left[\mathbf{y}_k - \mathcal{F}_{k:0}(\mathbf{x}_0^{(j)})\right] , \quad (5)$$

$$\widetilde{\mathcal{H}}_{(j)} = \mathbf{I}_N + \sum_{k=L-S+1}^{L} \mathbf{Y}^{\mathrm{T}}_{k,(j)}\mathbf{R}_k^{-1}\mathbf{Y}_{k,(j)} . \quad (6)$$

The notation (j) refers to the iteration index of the minimization. At the first iteration, one sets $\mathbf{w}^{(0)} = \mathbf{0}$ so that $\mathbf{x}_0^{(0)} = \overline{\mathbf{x}}_0$. $\mathbf{I}_N$ is the identity matrix in ensemble space. $\mathbf{Y}_{k,(j)} = [\mathcal{F}_{k:0}]'_{|\mathbf{x}_0^{(j)}}\mathbf{X}_0$ is the tangent linear of the operator from ensemble space to the observation space. The estimation of these sensitivities using the ensemble avoids the use of the model adjoint. Note that other iterative minimization schemes could be used in place of Gauss–Newton, such as quasi-Newton methods or Levenberg–Marquardt (Bocquet and Sakov, 2012) with distinct convergence properties.

These sensitivities can be computed using finite differences (the so-called *bundle* IEnKS variant). The ensemble is rescaled closer to the mean trajectory by a factor ε. It is then propagated through the model and the observation operator ($\mathcal{F}_{k:0}$), after which it is rescaled back by the inverse factor ε^{-1}. The operation reads

$$\mathbf{Y}_{k,(j)} \approx \frac{1}{\varepsilon}\mathcal{F}_{k:0}\left(\mathbf{x}_0^{(j)}\mathbf{1}^{\mathrm{T}} + \varepsilon\mathbf{X}_0\right)\left(\mathbf{I}_N - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}}{N}\right) , \quad (7)$$

where $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$. Note that the $\sqrt{N-1}$ factor needed to normalize the anomalies has been incorporated in ε. It is also possible to compute these sensitivities without rescaling the anomalies (Sakov *et al.*, 2012; Bocquet and Sakov, 2014) with very similar IEnKS performance (the so-called *transform variant*). However, we shall use the bundle variant in this study because it refers explicitly to the tangent linear of the model, which is instrumental to our study.

The iterative minimization is stopped when $\|\mathbf{w}^{(j)} - \mathbf{w}^{(j-1)}\|$ reaches a predetermined threshold e. Let us denote by $\mathbf{w}_0^\star$ the solution of the cost function minimization. The '$\star$' symbol will be used with any quantity obtained at the minimum. Subsequently, a posterior ensemble can be generated at t_0:

$$\mathbf{E}_0^\star = \mathbf{x}_0^\star\mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{X}_0\mathbf{W}\mathbf{U} , \quad (8)$$

where

$$\mathbf{W} = \left[\mathbf{I}_N + \sum_{k=L-S+1}^{L} \left(\mathbf{Y}_k^\star\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k^\star\right]^{-1/2} , \quad (9)$$

$\mathbf{U}$ is an arbitrary orthogonal matrix satisfying $\mathbf{U}\mathbf{1} = \mathbf{1}$, meant to keep the posterior ensemble centred on the analysis, and $\mathbf{x}_0^\star = \overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}_0^\star$. The update of the anomalies $\mathbf{X}_0^\star = \mathbf{X}_0\mathbf{W}$ is a right transform of the initial anomalies $\mathbf{X}_0$. It can be used with domain localization but is inadequate for covariance localization, since it operates in ensemble space. The analysis description is complete for fixed-lag smoothing. If filtering is considered, it additionally requires an independent model forecast throughout the DAW and possibly beyond the present time for forecasting.

During the forecast step of the scheme cycle, the ensemble is propagated for $S\Delta t$ time units, with S an integer that has been introduced via Eq. (2):

$$\mathbf{E}_S^\star = \mathcal{M}_{S:0}(\mathbf{E}_0^\star) . \quad (10)$$

This forecast ensemble at t_S will form the prior for the next analysis. A typical cycling of the analysis and forecast steps is displayed schematically in 1.

In complex data assimilation systems, asynchronous observations can be collected over a DAW and assimilated at fixed analysis times, using covariances across space and time and relying on the mechanisms of the standard ensemble Kalman smoother (Sakov *et al.*, 2010). The IEnKS can be understood as carrying out this task over the DAW but with fully nonlinear updates. The DAW of the asynchronous observation would then be of length $L\Delta t$. Nonetheless, the IEnKS should mainly be understood as using the DAW akin to a 4D-Var. In that latter case, asynchronous observations are collected over Δt, while the IEnKS DAW is of length $L\Delta t$.

A pseudo-code of the IEnKS is proposed in Algorithm 1, given in Table 1. It describes a full cycle of the (SDA) IEnKS that includes the analysis and forecast steps, with the exception of the model forecast within the DAW required to complete the analysis

Table 1. A cycle of the lag-L/shift-S/SDA/bundle/Gauss–Newton IEnKS.

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

$j = 0$, $\mathbf{w} = \mathbf{0}$
$\mathbf{x}_0^{(0)} = \mathbf{E}_0\mathbf{1}/N$
$\mathbf{X}_0 = \left(\mathbf{E}_0 - \mathbf{x}_0^{(0)}\mathbf{1}^{\mathrm{T}}\right)/\sqrt{N-1}$
repeat
  $\mathbf{x}_0 = \mathbf{x}_0^{(0)} + \mathbf{X}_0\mathbf{w}$
  $\mathbf{E}_0 = \mathbf{x}_0\mathbf{1}^{\mathrm{T}} + \epsilon\mathbf{X}_0$
  for $k = 1, \ldots, L$ **do**
    $\mathbf{E}_k = \mathcal{M}_{k:k-1}(\mathbf{E}_{k-1})$
    if $k \in [L - S + 1, L - 1]$ **then**
      $\overline{\mathbf{y}}_k = H_k(\mathbf{E}_k)\mathbf{1}/N$
      $\mathbf{Y}_k = (H_k(\mathbf{E}_k) - \overline{\mathbf{y}}_k\mathbf{1}^{\mathrm{T}})/\epsilon$
    end if
  end for
  $\overline{\mathbf{y}}_L = H_L(\mathbf{E}_L)\mathbf{1}/N$
  $\mathbf{Y}_L = (H_L(\mathbf{E}_L) - \overline{\mathbf{y}}_L\mathbf{1}^{\mathrm{T}})/\epsilon$
  $\nabla\widetilde{\mathcal{J}} = \mathbf{w} - \sum_{k=L-S+1}^{L}\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}_k - \overline{\mathbf{y}}_k)$
  $\widetilde{\mathcal{H}} = \mathbf{I}_N + \sum_{k=L-S+1}^{L}\mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k$
  Solve $\widetilde{\mathcal{H}}\Delta\mathbf{w} = \nabla\widetilde{\mathcal{J}}$
  $\mathbf{w} := \mathbf{w} - \Delta\mathbf{w}$
  $j := j + 1$
until $||\Delta\mathbf{w}|| \leq e$ **or** $j \geq j_{\max}$
$\mathbf{E}_0 = \mathbf{x}_0\mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{X}_0\mathcal{H}^{-1/2}\mathbf{U}$
$\mathbf{E}_S = \mathcal{M}_{S:0}(\mathbf{E}_0)$

for all times in the DAW, which is independent of the main IEnKS cycle.

Because it combines the propagation of the errors via the ensemble with a nonlinear 4D variational analysis, the IEnKS has been shown to outperform 4D-Var, the ensemble Kalman filter and smoother on several low-order models (Lorenz-63, Lorenz-95, Lorenz-95 coupled to a tracer model or a chemistry model and a 2D turbulence model). This good performance persisted in all tested regimes of nonlinearity, applied to either smoothing, filtering or forecasting. It was also shown to be useful for parameter estimation, because it combines the straightforward state augmentation technique of the EnKF with a variational analysis, where the adjoint sensitivity to the parameters is estimated within the scheme in ensemble space. The only fundamental approximation of the scheme is the Gaussian assumption made in the generation of the posterior ensemble and the second-order statistics assumption made when estimating the prior from the ensemble. When $S = L$ and using only the first iteration of the minimization, the transform variant SDA IEnKS becomes formally equivalent to the so-called 4D-ETKF (Hunt *et al.*, 2004), with an analysis formally similar to that of 4DEnVar (Liu *et al.*, 2008).

2.2. Obstacles to localization

However, as concluded by Bocquet and Sakov (2014), the main obstacle to the application of the IEnKS in high-dimensional systems is the implementation of a satisfying localization of the IEnKS analysis. Algorithm 1 could be written in state space rather than ensemble space. In such an algorithm, not only $\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$ would appear but also covariances coming as products of sensitivities, typically $\mathbf{X}_0\mathbf{Y}_k^{\mathrm{T}}$ between a state variable at t_0 and observations at t_k within the DAW, as well as $\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}$ between observations at t_k within the DAW. As with the EnKF, it is sound to assume that one can localize the background-error covariance matrix at t_0, $\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}} \mapsto \rho \circ \left(\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\right)$, using localization matrix ρ defined at the beginning of the IEnKS window. However, the regularization of covariances such as $\mathbf{X}_0\mathbf{Y}_k^{\mathrm{T}}$, $\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}$ is not straightforward. To gain some insight into the regularization of such covariances, we will study in the next section the definition of an adequate 4D localization operator in general and heuristic terms. We will come back to the IEnKS later and propose solutions inspired by these considerations.

3. Constraints on a consistent 4D localization operator

3.1. Time evolution of the localization operator

As explained in the Introduction and incarnated by Eq. (1), the localization operator is not necessarily consistent with the dynamical flow of the model. Let us consider a *linear* localization operator that applies to covariances: $\mathcal{L} : \mathbf{P} \mapsto \mathcal{L} \cdot \mathbf{P}$. Schur localization is an example of such a generic linear operator. The transport of the covariances by the tangent linear model from t to $t + \delta t$ is formalized with the following linear operator:

$$\mathcal{T}_{t+\delta t:t} : \mathbf{P}_t \mapsto \mathcal{T}_{t+\delta t:t} \cdot \mathbf{P}_t = \mathbf{M}_{t+\delta t:t}\mathbf{P}_t\mathbf{M}_{t+\delta t:t}^{\mathrm{T}}\,. \tag{11}$$

Assuming that the localization operator is not static but depends on time, we wish to impose that localization commutes with the tangent linear model, which can be written as

$$\mathcal{T}_{t+\delta t:t} \cdot \mathcal{L}_t \cdot \mathbf{P}_t = \mathcal{L}_{t+\delta t} \cdot \mathcal{T}_{t+\delta t:t} \cdot \mathbf{P}_t\,. \tag{12}$$

Ideally, this dynamical constraint should apply to any $t, \delta t$ and $\mathbf{P}_t$. In the limit where δt goes to zero, the transport operator can be expanded as

$$\mathcal{T}_{t+\delta t:t} = \mathbf{I} \otimes \mathbf{I} + \mathcal{K}_t\delta t + o(\delta t)\,, \tag{13}$$

where $\mathcal{K}_t$ is a linear operator that can be written $\mathcal{K}_t = \mathbf{M}_t \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{M}_t$ and $\mathbf{I} \otimes \mathbf{I}$ is the identity operator in the space of the covariance matrices. As a consequence, Eq. (12) can in turn be expanded in the $\delta t \to 0$ limit to yield

$$\frac{\partial\mathcal{L}_t}{\partial t} = [\mathcal{K}_t, \mathcal{L}_t]\,, \tag{14}$$

where $[A, B] = AB - BA$ is the commutator of linear operators. This *Liouville equation* prescribes the dynamics of a consistent localization operator for synchronous covariances of the form $\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}$. As explained in the previous section, a consistent asynchronous localization operator that can be used to regularize covariances such as $\mathbf{X}_0\mathbf{Y}_k^{\mathrm{T}}$ is also needed. In that case, the same Liouville equation still applies but with an asymmetrical transport $\mathcal{K}_t = \mathbf{I} \otimes \mathbf{M}_t$ (or $\mathcal{K}_t = \mathbf{M}_t \otimes \mathbf{I}$ for $\mathbf{Y}_k\mathbf{X}_0^{\mathrm{T}}$).

A solution to this equation is desirable but seems out of reach without further specification of the dynamics. In Appendix A, assuming the model is univariate, continuous and one-dimensional, we explore solutions of this equation under restrictive conditions. Looking for a solution is greatly simplified if the dynamics are assumed to be hyperbolic. In practice, it means that the model partial differential equation depends only on the first-order partial derivatives and the model can theoretically be integrated using the method of characteristics (see for instance Evans *et al.*, 1999). In that case, the evolution of the covariances, variances and correlations is simple (Cohn, 1993; Ménard *et al.*, 2000), but we observe that the general Liouville equation remains complex if it is not simplified further. However, if, in addition to hyperbolic dynamics, the localization is assumed to be a Schur product, i.e. $\mathcal{L} : \mathbf{P} \mapsto \boldsymbol{\rho} \circ \mathbf{P}$, where $\boldsymbol{\rho}$ is a positive-definite correlation function, then we show that this localization function $\boldsymbol{\rho}$ must have the same evolution as any correlation function. In that case, $\boldsymbol{\rho}$ can be propagated using characteristics. Nonetheless, this is likely to be incorrect if the dynamics are not hyperbolic.

If the dynamics incorporate diffusion and hence are not hyperbolic but parabolic, the equation for $\boldsymbol{\rho}$ cannot be closed in the space of the correlation functions, which demonstrates how difficult finding a consistent localization can be in the presence of diffusion.

Finally, we also show in Appendix A under the restrictive conditions of hyperbolic dynamics and Schur localization that an evolution equation can be derived for the correlation length of $\boldsymbol{\rho}$, i.e. for the localization length.

3.2. Impact of the dynamics

If these considerations shed some light on 4D localization and are helpful in that respect, they do not necessarily lead to practical algorithms for realistic geophysical flows. Here, we discuss why.

We have seen that if the system is hyperbolic then one can use the characteristics method to evolve either the covariances or a Schur localization function. The inviscid Burgers equation is the archetype of such nonlinear model. However, we have also observed that it is difficult to compute the evolution of the covariances (Cohn, 1993) or of the localization operator (this study) when the system dynamics are parabolic. If the system is taken as approximately hyperbolic, focusing for instance only on the advection and the wind variables of a geophysical fluid, one can use a semi-Lagrangian method using the wind field as a surrogate model to evolve either the covariances or a Schur localization function. In that case and in the rest of this study, by characteristics we shall mean the characteristic curves of the advection surrogate model, usually implemented by Lagrangian integrations.

This view is especially relevant for chemistry and transport models with their advection–diffusion–reaction equations (Ménard *et al.*, 2000). Characteristic curves can be used to transport the covariances or the localization operator, even if approximately. This is supported by the traditional interpretation of the advection–diffusion equation as a Fokker–Planck equation that emerges from the stochastic integration of particles with a drift equal to the wind (Risken, 1989, and references therein).

Still, we cannot expect these (approximate or not) characteristic curves to always hold for the range of dynamical models to be used in the geosciences. Most turbulent geophysical fluids can hardly be seen as hyperbolic. At small scales, two nearby perturbations in the fluid diverge from each other rapidly, at variance with the picture of integration by characteristics. Hence, some fluid velocity average over small scales will be necessary in turbulent geophysical fluids. This may coincide with the dissipation range. Other transient phenomena such as waves, which could carry information too, may blur the line and prevent a clear definition of how the information attached to a perturbation is carried away by the dynamical flow.

At the end of this range of dynamics lie strongly chaotic models, such as the Lorenz-95 model (Lorenz and Emanuel, 1998), which will be used as illustration in section 5, where no obvious surrogate advection model emerges. Information propagates in these systems, but it is not necessarily simply advected following the fluid parcels. On the other hand, some climatology-based average response to a perturbation (related to the sensitivity that we must compute in the IEnKS) can be estimated. This will be carried out for the Lorenz-95 model in section 5.

If the dynamical system is of hyperbolic type, we have seen in the specific context of Appendix A that the evolution of the localization function follows that of the error correlation closely. At the other end of the dynamical range, for significantly chaotic models, invariant quantities such as energy or enstrophy may lead to an invariant statistical state distribution. For instance, using the existence of an invariant distribution in the Lorenz-95 system due to the conservation of energy by the convective term and its Liouville property, Majda *et al.* (2005) showed that a fluctuation dissipation theorem precisely relates the average response to a perturbation and the climatological correlations in the system. In this case, a localization operator, meant to be efficient on average, could be estimated from the climatological correlations in the system. Hence, we remark that the time localization function could be related to the time evolution of the error correlations or to the climatological model correlations, depending on the nature of the model dynamics.

Finally, for long DAWs, it is very likely that the picture of a consistent 4D localization operator will fade away as (effective) diffusion starts impacting the dynamics significantly and spoils the formal framework discussed earlier. Hence our effort should be targeted at defining and improving 4D localization for reasonably short DAWs.

3.3. Strategies for 4D localization

From sections 3.1 and 3.2, two general strategies emerge. One path consists of applying spatial localization to the initial sampled covariances at t_0, followed by the propagation of these covariances using characteristics to t_k. It is designated by the directed path A in Figure 2. Another route, path B, consists of propagating the original ensemble of perturbations defined at t_0 to t_k without approximation, propagating the initial localization function at t_0 using characteristics and, finally, regularizing the covariances of the propagated perturbations using the propagated localization function.

Paths A and B are theoretically equivalent if the dynamics is truly hyperbolic. However, it is likely that, with realistic models, only path B is potentially viable. Interest in the direct path A can nonetheless emerge from the use of promising methods such as the propagation of the local metric tensor of the covariance matrices (Pannekoucke *et al.*, 2014). It may also come from specific but physically relevant dynamics which are close to hyperbolic, such as those of atmospheric constituents (Ménard *et al.*, 2000), where the method of characteristics is relevant. In contrast, path B has the major advantage that it relies only partially on the characteristics method, which will often be an approximation in practice.

A third route (path C) is largely independent of the previous considerations. It consists of applying covariance localization at t_0 first. The regularized covariance is then decomposed into modes $\mathbf{Z}_0 = [\mathbf{z}_0^{[1]}, \mathbf{z}_0^{[2]}, \cdots, \mathbf{z}_0^{[M]}]$. Those modes can be obtained using the normalized eigenvectors $\mathbf{v}^{[m]}$ and corresponding eigenvalues $\lambda^{[m]}$ of the regularized covariance matrix and setting $\mathbf{z}^{[m]} = \sqrt{\lambda^{[m]}}\mathbf{v}^{[m]}$. The eigenpairs are ordered by their decreasing eigenvalue. The dominant modes are then selected: $\mathbf{Z}_0^r = [\mathbf{z}_0^{[1]}, \mathbf{z}_0^{[2]}, \cdots, \mathbf{z}_0^{[r]}]$, with $r \ll M$ chosen in order to keep a sufficiently large ratio $\sum_{m=1}^{r} \lambda^{[m]} / \sum_{m=1}^{M} \lambda^{[m]}$. We obtain

$$\rho \circ (\mathbf{X}_0\mathbf{X}_0) = \mathbf{Z}_0(\mathbf{Z}_0)^{\mathrm{T}} \simeq \mathbf{Z}_0^r(\mathbf{Z}_0^r)^{\mathrm{T}}. \tag{15}$$

Depending on the minimal r required by the data assimilation system, the method may or may not be numerically affordable.

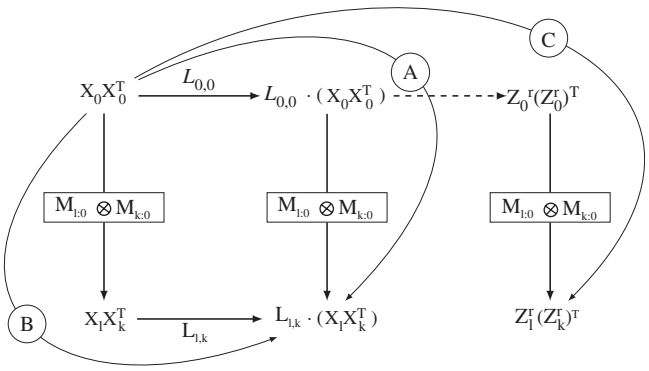

Figure 2. Schematic of several strategies for localization in the IEnKS. The left square diagram is commutative if the dynamics are of hyperbolic type and if the localization is a Schur product (sufficient conditions). The diagram accounts for both synchronous ($l, k := k, k$) and asynchronous ($l, k := 0, k$) covariances.

The r dominant perturbations are then propagated to t_k, where covariances can be estimated from this forecast ensemble of perturbations. This strategy has been used successfully in petroleum reservoir modelling data assimilation, where the unstable subspace is not as large as for geophysical flows (Chen and Oliver, 2013), hence requiring fewer modes than in chaotic systems, but the model nonlinearity is strong.

These considerations were focused on covariance localization. However, we believe that they are relevant for domain localization too. Still assuming some underlying hyperbolic dynamics, if a local domain D_η is centred at a location η at t_0, then this domain will be transported by the flow into $\mathcal{S}_{k:0}(D_\eta)$ at t_k. The model $\mathcal{S}$ is the surrogate model of hyperbolic dynamics meant to approximate the complete model, which can be integrated forward or backward in time using the characteristic curves. The three-dimensional domain D_η is seen as *covariant* with the dynamical flow. To update the model state at η, one will be able to use observations at t_k that are contained within the image $\mathcal{S}_{k:0}(D_\eta)$ of the initial domain D_η. To determine those observations, it is possible to pull the locations of the potentially useful observations at t_k backward to t_0 using the characteristic curves and select only those contained in D_η.

If the flow cannot be seen as hyperbolic, $\mathcal{S}_{k:0}(D_\eta)$ should at least encompass the evolved perturbation initially centred at η.

4. Possible implementations

In the following, we explore several ways to implement localization in the IEnKS using the guidelines from section 3, with relevance to other four-dimensional EnVar systems.

4.1. Covariance localization with the adjoint model

Algorithm 1 is not suitable for covariance localization because it mostly operates in ensemble space. For covariance localization, an IEnKS implementation should operate in state and observation spaces.

That is why we consider the cost function of the IEnKS, as in Eq. (2), but written in state space rather than in ensemble space. Moreover, we choose $S = 1$ for the sake of clarity, i.e. one observation vector is assimilated at the end of the DAW. Note that with such minimal S the algorithm inherits the properties of a quasi-static variational algorithm (Pires *et al.*, 1996; Bocquet and Sakov, 2014). The cost function reads

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x})\right\|_{\mathbf{R}}^2 + \frac{1}{2}\left\|\mathbf{x} - \overline{\mathbf{x}}_0\right\|_{\mathbf{B}_0}^2. \tag{16}$$

The background covariance matrix $\mathbf{B}_0$ is chosen to be the Schur product of the sampled covariance matrix with the localization matrix ρ: $\mathbf{B}_0 = \rho \circ \left(\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\right)$. We further assume that the chosen minimization scheme is gradient-based, a requirement for high-dimensional applications. The gradient of the cost function is

$$\nabla_{\mathbf{x}}\mathcal{J} = -\mathbf{F}_{L:0}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x})\right\} + \mathbf{B}_0^{-1}(\mathbf{x} - \overline{\mathbf{x}}_0), \tag{17}$$

where $\mathbf{F}_{L:0} = \mathbf{H}_L\mathbf{M}_{L:0}$ is the tangent linear of the observation and forecast model operator applied to the model resolvent $H_L \odot \mathcal{M}_{L:0}$, where $\odot$ is the operator composition. Depending on the iterative minimizer that is used or on a change of variable for better conditioning, one may alternatively use

$$\mathbf{B}_0\nabla_{\mathbf{x}}\mathcal{J} = -\mathbf{B}_0\mathbf{F}_{L:0}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x})\right\} + \mathbf{x} - \overline{\mathbf{x}}_0. \tag{18}$$

This gradient is computed without approximation in the global IEnKS. With an approximate localized IEnKS, it is desirable to keep sufficient precision in the gradient, especially for nonlinear systems that would require several iterations (outer loop) before convergence.

As in any 4D variational system, the computation of the gradient requires the adjoint model $\mathbf{F}^{\mathrm{T}}$, as seen from Eq. (17). In a global IEnKS, this is addressed without the explicit adjoint model. Indeed, in that case $\mathbf{B}_0 = \mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$ and, as seen from Eq. (17), the adjoint model output is projected on to the ensemble space. Unfortunately, within a CL IEnKS, $\mathbf{B}_0$ becomes full-rank and projection on a reduced space is ruled out, as seen in Eq. (18). If one has at their disposal the tangent linear and adjoint of model $H \odot \mathcal{M}$, then it is possible from Eq. (17) or (18) to compute the gradient within a CL IEnKS and to find the state analysis via an efficient minimization of the cost function. For the same reason, Buehner *et al.* (2010) with the 4D-Var-Ben and more specifically Fairbairn *et al.* (2014) pointed out the theoretical advantage of 4D-Var and its adjoint model over 4DEnVar systems when it comes to localizing. The advantage of using the adjoint in the four-dimensional localization was also demonstrated numerically by Poterjoy and Zhang (2015) in a comparison between 4DEnVar, which does not use the explicit tangent and adjoint models, and E4DVar, which uses them. Focusing on the perturbations update of the IEnKS, this argument is pushed even further in the following.

The updated anomalies are, in this state-space context, obtained from a left transform of the prior anomalies $\mathbf{X}_0^\star = \mathbf{T}\mathbf{X}_0$, rather than from the right transform Eq. (9). Following Sakov and Bertino (2011), the left transform matrix is

$$\mathbf{T} = \left[\mathbf{I}_N + \mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\left(\mathbf{F}_{L:0}^\star\right)^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{L:0}^\star\right]^{-1/2}. \tag{19}$$

This expression theoretically enables covariance localization of $\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}$ through

$$\mathbf{T} = \left[\mathbf{I}_N + \mathbf{B}_0\left(\mathbf{F}_{L:0}^\star\right)^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{L:0}^\star\right]^{-1/2}, \tag{20}$$

where $\mathbf{B}_0 = \boldsymbol{\rho} \circ \left(\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\right)$. If the adjoint is available, the action of matrix $\mathbf{B}_0\left(\mathbf{F}_{L:0}^\star\right)^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{L:0}^\star$ on any perturbation is numerically affordable in principle. Since one only needs to apply $\mathbf{T}$ on the prior anomalies, the use of a Krylov subspace method is theoretically possible to compute $\mathbf{T}\mathbf{X}_0$. In conclusion, if the adjoint of $\mathcal{F}_{L:0}$ is available, it is theoretically possible to implement covariance localization rigorously in the IEnKS at a scalable numerical cost.

Even though avoiding the development and maintenance of the adjoint was the prime reason for the development of 4DEnVar (Liu *et al.*, 2008), it is only one motivation for the IEnKS, the precision of the algorithm being another. That point clarified, it is of course desirable in practice to perform localization of the IEnKS without the adjoint model. So, from now on, let us assume that we do not have access to the explicit adjoint model of $H \odot \mathcal{M}$.

4.2. Covariance localization without the adjoint model

Let us start from cost function Eq. (16) and let us use the Gauss–Newton minimizer as in Algorithm 1. The iteration has the explicit state-space form (an inner loop problem):

$$\begin{aligned}\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} &- \left[\mathbf{I}_M + \mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{(j)}\right]^{-1} \\ &\times\left[-\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}\mathbf{R}_L^{-1}\left\{\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x}^{(j)})\right\} + \mathbf{x}^{(j)} - \overline{\mathbf{x}}_0\right],\end{aligned} \tag{21}$$

where $\mathbf{F}_{(j)}$ is a shortcut for $\mathbf{F}_{L:0}^{(j)}$. An equivalent update formula is obtained using the Sherman–Morrison–Woodbury lemma:

$$\begin{aligned}\mathbf{x}^{(j+1)} = \overline{\mathbf{x}}_0 &+ \mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}\left[\mathbf{R}_L + \mathbf{F}_{(j)}\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}\right]^{-1} \\ &\times\left[\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x}^{(j)}) + \mathbf{F}_{(j)}\left\{\mathbf{x}^{(j)} - \overline{\mathbf{x}}_0\right\}\right].\end{aligned} \tag{22}$$

Approximations of the gradient due to localization put the convergence of the variational minimization at risk. The

advantage of Eq. (21) over Eq. (22) is that it maintains the explicit call to the gradient Eq. (18). Hence, the impact on the iterative minimization of an approximation carried out in Eq. (21) due to localization can be scrutinized better than an approximation made in Eq. (22).

Equation (22) resembles a standard state estimate of the EnKF, up to a relaxation term to the prior (Sakov *et al.*, 2012). Localization could be applied through

$$\begin{aligned}\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}} &\simeq \rho \circ \left\{\mathbf{X}_0 \left(\mathbf{F}_{(j)}\mathbf{X}_0\right)^{\mathrm{T}}\right\} ,\\ \mathbf{F}_{(j)}\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}} &\simeq \rho \circ \left\{\mathbf{F}_{(j)}\mathbf{X}_0 \left(\mathbf{F}_{(j)}\mathbf{X}_0\right)^{\mathrm{T}}\right\} .\end{aligned} \tag{23}$$

Unfortunately, $\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}$ are asynchronous covariances, which cannot easily be regularized by a climatological localization matrix ρ. Moreover, as explained in section 3, we would like the localization operator to be consistent with the evolution of the covariances. Such approximations are not fundamentally detrimental to the EnKF, since such error in the localization scheme will be accounted for by inflation later in the cycle. They could, however, be detrimental to the IEnKS if the variational analysis requires several iterations in significantly nonlinear systems.

Following path A of section 3, we would propagate $\mathbf{B}_0$ using characteristics either on the right (asynchronous covariances) or on the left and the right (synchronous covariances) to obtain estimates of $\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}$ and $\mathbf{F}_{(j)}\mathbf{B}_0\mathbf{F}_{(j)}^{\mathrm{T}}$. Following path B of section 3, we would propagate the ensemble as in the right-hand side of Eq. (23), followed by localization using the *covariant* localization operator at t_L that is obtained by the characteristic method. Both methods are rigorous if the flow is hyperbolic, and approximate otherwise. In the latter case, path B is probably sounder and safer since, the Lagrangian transport would only apply to the localization operator, whereas path A is likely to be unviable.

Provided localization can be implemented with a limited number of modes, then path C has the advantage that it can maintain consistency and convergence within the IEnKS minimization. Assume $\mathbf{B}_0$ has been regularized and decomposed into r dominant modes $\mathbf{B}_0 = \boldsymbol{\rho} \circ \left(\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\right) \simeq \mathbf{Z}_0^r \left(\mathbf{Z}_0^r\right)^{\mathrm{T}}$. The state analysis is carried out rigorously as the global IEnKS, following Algorithm 1 but using these r perturbations rather than the initial perturbations. The ensemble update is more problematic. We would like to apply the left transform Eq. (20):

$$\mathbf{T} = \left[\mathbf{I}_N + \mathbf{Z}_0^r \left(\mathbf{Z}_0^r\right)^{\mathrm{T}} \left(\mathbf{F}_{L:0}^{\star}\right)^{\mathrm{T}} \mathbf{R}_L^{-1}\mathbf{F}_{L:0}^{\star}\right]^{-1/2} , \tag{24}$$

to $\mathbf{X}_0$ but cannot do so in the present form. Another equivalent expression for $\mathbf{T}$ leads to the update formula:

$$\begin{aligned}\mathbf{X}_0^{\star} = &\Big[\mathbf{I}_M - \mathbf{Z}_0^r \left(\mathbf{I}_r + \mathbf{Y}_r^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{Y}_r + \sqrt{\mathbf{I}_r + \mathbf{Y}_r^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{Y}_r}\right)^{-1}\\ &\times \mathbf{Y}_r^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{L:0}^{\star}\Big]\mathbf{X}_0,\end{aligned} \tag{25}$$

with $\mathbf{Y}_r = \mathbf{F}_{L:0}^{\star}\mathbf{Z}_0^r$. This update is practical only if the number of required modes r remains small. A proof is given in Appendix B. It can even be approximated to avoid the square-root matrix, following the idea of Sakov and Oke (2008), if the analysis remains close to the prior:

$$\mathbf{X}_0^{\star} \simeq \mathbf{X}_0 - \frac{1}{2}\mathbf{Z}_0^r \left(\mathbf{I}_r + \mathbf{Y}_r^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{Y}_r\right)^{-1} \mathbf{Y}_r^{\mathrm{T}}\mathbf{R}_L^{-1}\mathbf{F}_{L:0}^{\star}\mathbf{X}_0 . \tag{26}$$

This consistent implementation of the IEnKS has an interesting theoretical advantage. When the number of modes r increases, the solution converges to a fully consistent localization scheme of the IEnKS based on Eq. (16). By increasing the number of modes, one should obtain some baseline performance for a local IEnKS based on a dynamically consistent covariance localization, a performance that can be achieved in high dimension if the model adjoint is available. Even though the method cannot be used with high-dimensional systems, we shall use it as a baseline performance method in the first numerical illustration of this study.

It is also possible to perform a similar analysis for covariance localization in the context of the serial ensemble square-root Kalman filter (Whitaker and Hamill, 2002; Tippett *et al.*, 2003). The implementation of paths A and B would be analogous to what has been described for the matrix ensemble square-root filter and the adaptation is straightforward. When following path B, one noticeable simplification is that only the asynchronous localization operator is needed. Indeed, the serial equivalent to the covariances $\mathbf{X}_0\mathbf{Y}_L^{\mathrm{T}}$ and $\mathbf{Y}_L\mathbf{Y}_L^{\mathrm{T}}$ is scalar regression coefficients; $\mathbf{X}_0\mathbf{Y}_L^{\mathrm{T}}$ becomes an asynchronous regression coefficient that needs tapering with the proper localization coefficient estimated with a characteristic curve from t_0 to t_L and $\mathbf{Y}_L\mathbf{Y}_L^{\mathrm{T}}$ becomes a variance without tapering (unlike the batch IEnKS).

4.3. *Domain localization*

In the following, domain localization is implemented using analyses in ensemble space, similarly to the local ETKF. Forgetting temporarily about the issues raised and discussed in section 3, let us implement domain localization building on Algorithm 1 for the global IEnKS and on Hunt *et al.* (2007) for the local ETKF. For each iteration of the IEnKS outer loop, a local analysis is performed on the ensemble coefficients vector $\mathbf{w}^m$ for each grid-space point η_m indexed by m. Each local analysis encompasses local observations, which can be formalized with the observation-error covariance matrix. The observation-error covariance matrix $\mathbf{R}$ becomes $\mathbf{R}^m$, where all observation errors outside the domain centred on η_m are strongly inflated, using either a boxcar scheme or a smooth function that inflates the variances progressively away from η_m to enable smoother transitions in between nearby local analyses. The vectors $\mathbf{w}^m$ are updated locally, for $m = 1, \ldots, M$, which defines a global updated state vector at t_0. A new set of local perturbations is also generated. The computations of the required sensitivities are obtained by forecasting these perturbations within the DAW using the global model $\mathcal{F}$. Hence, in practice, this forecast step within the DAW is not more numerically expensive than in the global IEnKS. The final ensemble update is the same as in the local ETKF. Pseudo-code for this algorithm is given by Algorithm 2, shown in Table 2, and can be compared with the global Algorithm 1.

This algorithm has one theoretical caveat. The algorithm actually minimizes a collection of local cost functions, which may not be glued to form a global cost function without approximations. Hence, convergence of the global minimization is not guaranteed if the transition between local domain analyses is not smooth enough. In practical terms, the precision of the algorithm for the local IEnKS depends on the threshold e in Algorithm 1 more significantly than the global IEnKS. In the absence of guaranteed convergence, the threshold cannot be set to too small a precision target. The physical reason for this caveat is that the locality that is usually assumed implicitly in the local ETKF between the state variable being updated and the observations is broken for long enough DAWs. A solution to this issue is to follow the suggestion of covariant localization at the end of section 3, the implementation of which is discussed in the following. Hence, we now wish to implement in this domain localization framework a localization where the domains are covariant, i.e. where they move in accordance with the flow within the DAW.

The inner loop problem solved by Algorithm 2 in the case $S = 1$ at iteration j in the local domain D_m around

Table 2. A cycle of the local/lag-L/shift-S/SDA/bundle/Gauss–Newton IEnKS.

Require: t_L is present time. Transition model $\mathcal{M}_{k+1:k}$, observation operators H_k at t_k. Algorithm parameters: ϵ, e, $j_{\max}$. $\mathbf{E}_0$, the ensemble at t_0, $\mathbf{y}_k$ the observation at t_k. $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$. $\mathbf{U}_m$ is an orthogonal matrix in $\mathbb{R}^{N\times N}$ satisfying $\mathbf{U}_m\mathbf{1} = \mathbf{1}$.

$j = 0$
for $m = 1, \ldots, M$ **do**
  $\mathbf{w}^m = \mathbf{0}$
end for
$\mathbf{x}_0^{(0)} = \mathbf{E}_0\mathbf{1}/N$
$\mathbf{X}_0 = \left(\mathbf{E}_0 - \mathbf{x}_0^{(0)}\mathbf{1}^{\mathrm{T}}\right)/\sqrt{N-1}$
repeat
  for $m = 1, \ldots, M$ **do**
    $[\mathbf{x}_0]_m = [\mathbf{x}_0^{(0)}]_m + [\mathbf{X}_0\mathbf{w}^m]_m$
  end for
  $\mathbf{E}_0 = \mathbf{x}_0\mathbf{1}^{\mathrm{T}} + \epsilon\mathbf{X}_0$
  for $k = 1, \ldots, L$ **do**
    $\mathbf{E}_k = \mathcal{M}_{k:k-1}(\mathbf{E}_{k-1})$
    if $k \in [L - S + 1, L - 1]$ **then**
      $\overline{\mathbf{y}}_k = H_k(\mathbf{E}_k)\mathbf{1}/N$
      $\mathbf{Y}_k = (H_k(\mathbf{E}_k) - \overline{\mathbf{y}}_k\mathbf{1}^{\mathrm{T}})/\epsilon$
    end if
  end for
  $\overline{\mathbf{y}}_L = H_L(\mathbf{E}_L)\mathbf{1}/N$
  $\mathbf{Y}_L = (H_L(\mathbf{E}_L) - \overline{\mathbf{y}}_L\mathbf{1}^{\mathrm{T}})/\epsilon$
  for $m = 1, \ldots, M$ **do**
    $\nabla\widetilde{\mathcal{J}}_m = \mathbf{w}^m - \sum_{k=L-S+1}^{L}\mathbf{Y}_k^{\mathrm{T}}(\mathbf{R}_k^m)^{-1}(\mathbf{y}_k - \overline{\mathbf{y}}_k)$
    $\widetilde{\mathcal{H}}_m = \mathbf{I}_N + \sum_{k=L-S+1}^{L}\mathbf{Y}_k^{\mathrm{T}}(\mathbf{R}_k^m)^{-1}\mathbf{Y}_k$
    Solve $\widetilde{\mathcal{H}}_m\Delta\mathbf{w}^m = \nabla\widetilde{\mathcal{J}}_m$
    $\mathbf{w}^m := \mathbf{w}^m - \Delta\mathbf{w}^m$
  end for
  $j := j + 1$
until $\sum_{m=1}^{M} ||\Delta\mathbf{w}^m|| \leq Me$ **or** $j \geq j_{\max}$
for $m = 1, \ldots, M$ **do**
  $\widetilde{\mathcal{H}}_m = \mathbf{I}_N + \sum_{k=L-S+1}^{L}\mathbf{Y}_k^{\mathrm{T}}(\mathbf{R}_k^m)^{-1}\mathbf{Y}_k$
  $[\mathbf{E}_0]_m = [\mathbf{x}_0\mathbf{1}^{\mathrm{T}}]_m + \sqrt{N-1}\mathbf{X}_0\widetilde{\mathcal{H}}_m^{-1/2}\mathbf{U}_m$
end for
$\mathbf{E}_S = \mathcal{M}_{S:0}(\mathbf{E}_0)$

η_m minimizes the *local* cost function

$$\widetilde{\mathcal{J}}_{(j)}^m(\mathbf{w}) = \left\|\mathbf{y}_L - \mathcal{F}_{L:0}(\mathbf{x}_0^{(j)}) - \mathbf{Y}_{L,(j)}(\mathbf{w} - \mathbf{w}_{(j)}^m)\right\|_{\mathbf{R}_L^{D_m}}^2 + \frac{1}{2}\left\|\mathbf{w} - \mathbf{w}_{(j)}^m\right\|^2, \tag{27}$$

where $\mathbf{R}_L^{D_m} \equiv \mathbf{R}_L^m$. The first term of $\widetilde{\mathcal{J}}_{(j)}^m$ uses observations at t_L contained within D_m. However, we argued in section 3 that it should instead use observations within the image of D_m by a surrogate model: $\mathcal{S}_{L:0}(D_m)$, as schematized in Figure 3.

The inner loop problem is only affected via the specification of a new local observation-error covariance matrix: $\mathbf{R}_L^{D_m} \longrightarrow \mathbf{R}_L^{\mathcal{S}_{L:0}(D_m)}$, which only makes use of the observations contained in $\mathcal{S}_{L:0}(D_m)$. Let us assume in the following that the surrogate model is an advection model and let us see how to define $\mathbf{R}_L^{\mathcal{S}_{L:0}(D_m)}$. Let us first suppose that the local analysis relies on a boxcar scheme. In order to determine the observations contained in the subdomains $\mathcal{S}_{L:0}(D_m)$, one can pull back all observation sites at t_0 using the backward surrogate model. Since the surrogate model is an advection, this pull back is simple and numerically efficient to implement. Moreover, it can easily be parallelized. The observations with pulled-back sites contained in D_m should be selected for this local analysis. These have the strongest influence on D_m and obviously the variable at its centre, η_m. Secondly, this scheme is easily adapted if a smooth tapering function in place of a boxcar function is used to render the precision matrix $\mathbf{R}_L^{-1}$ local. In that case, the precision of an observation at t_L is tapered depending on the distance between the centre of the domain η_m and the pulled-back site of this observation at t_0. Through this precision tapering scheme, the observations are selected and rated according to their influence on the variables defined in D_m.

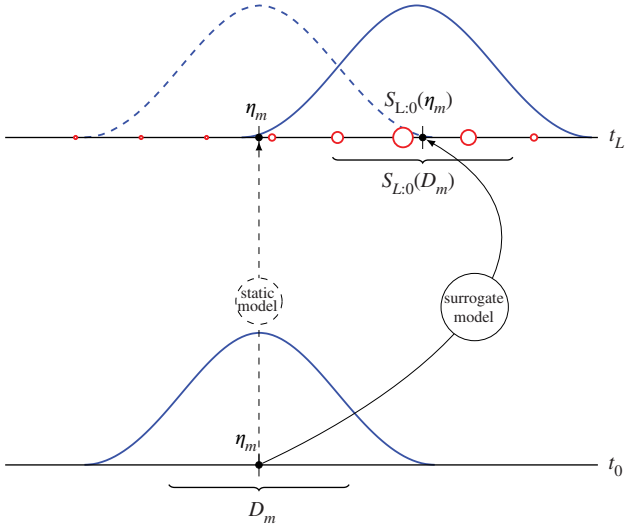

Figure 3. Schematic of the covariant DL IEnKS analysis with comoving domains: state variables indexed by m at η_m, t_0 are updated using observations at t_L, the precisions of which are tapered depending on the distance between the pulled-back locations of these observations and the centre of the initial domain D_m. Here the surrogate model used to pull back the observations is a simple uniform advection model. The curves show the GC function used to taper the precision of the observations and smooth their contribution to the analysis of state variables at η_m, t_0 and transported, or not, at t_L. The static DL IEnKS does not move the domains. The observations used in the covariant DL analysis are represented by red circles, the diameters of which depend on the GC tapering.

Algorithm 2 is only modified through the change: $\mathbf{R}^m \rightarrow \mathbf{R}_L^{\mathcal{S}_{L:0}(D_m)}$. In practice, this requires one to perform the pull back of the observations using the mean trajectory within the DAW. It could be performed once for the whole variational analysis or once per inner loop problem, depending on the required precision.

If the surrogate model is a uniform advection, the pull back of the observation sites can equivalently be replaced with a simple forecast $\mathcal{S}_{L:0}(\eta_m)$ of the centre η_m of the domain within the DAW, in order to select the relevant observations centred on the image $\mathcal{S}_{L:0}(\eta_m)$. This latter scheme will be used in section 5, whereas the former scheme will be used in section 6.

5. Numerical illustration with a one-dimensional model

In the following, we illustrate localization of the IEnKS, using several implementations proposed in section 4 and the Lorenz-95 model (Lorenz and Emanuel, 1998).

The Lorenz-95 model is a low-order one-dimensional discrete model of $M = 40$ variables that obeys the following ordinary differential equations for $m = 1, \ldots, M$:

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F, \tag{28}$$

where $F = 8$ is a forcing parameter. Periodic boundary conditions are assumed ($x_{M+1} = x_1$, $x_0 = x_M$ and $x_{-1} = x_{M-1}$). A time-step of 0.05 in this model represents 6 h of real time. The model mimics a midlatitude circle and its Rossby waves. It is chaotic; it has 13 positive Lyapunov exponents, plus one equal to 0. In particular, Lorenz-95 is useful in that it has been used extensively for the development of data assimilation methods. It is well known

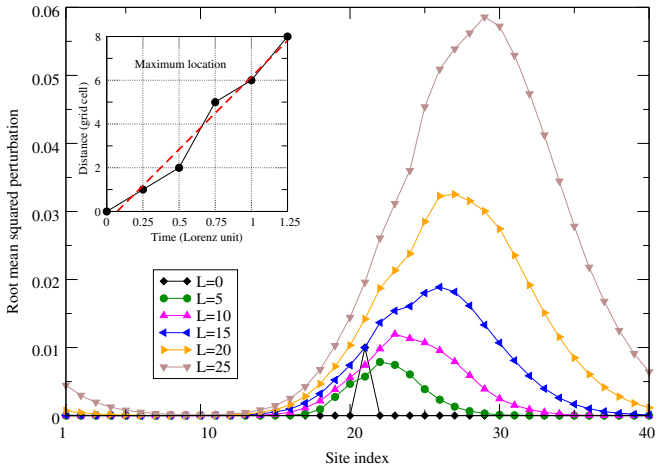

Figure 4. Root-mean-squared perturbations averaged over many simulations. The initial perturbation ($L = 0$) is an increment of x_{20} by 0.01. This perturbation is evolved to times $\Delta t = 0.25, 0.50, 0.75, 1.00, 1.25$, or $L = 5, 10, 15, 20, 25$. A tentative velocity of the perturbation propagation of $v_p = 6.6 \pm 0.5$ is deduced from the location of the maximum of the perturbation by linear regression (in the insert).

and it has documented expected performances for state-of-the-art methods such as the IEnKS. Localization techniques can be easily tested with the model. The short-range correlations in the model are consistent with the physics it is roughly meant to represent. Some important features of localization in ensemble methods, such as imbalance and variable sparsity of the observation network, can be less conveniently experimented upon with this model, but those aspects are not the focus of the present article.

The undamped unforced Lorenz-95 dynamical system has no known Hamiltonian structure and is quite far from an ideal hyperbolic dynamical system, even though it conserves energy. As intended, the dynamics are dominated by Rossby-like waves. Linear stability theory shows that the unstable model modes have westward (negative) phase velocity and eastward (positive) group velocity (Lorenz and Emanuel, 1998). To the author's knowledge, there is no obvious advection surrogate model that would approximate the Lorenz-95 system and help define some approximate characteristic curves.

Hence, it is more difficult to implement the methods of section 4 on Lorenz-95 than with an (inviscid or not) one-dimensional Burgers model, because a perturbation does not remain neatly contained within a progressively growing domain, under the model's action. To go further, we wish to estimate the average response in the domain to a localized impulse from t_0 to t_L. This defines an average sensitivity. This can be estimated by numerics, or in a less intensive way computed from the climatological correlations in the system and using a fluctuation dissipation theorem to relate the two quantities (Majda *et al.*, 2005). Figure 4 shows the average standard deviation of the response to a small perturbation at a point in the domain, over a time range relevant to our use of the IEnKS. The location of the maximum of the root-mean-squared perturbation as a function of time leads to the definition of a mean velocity for the propagation of the perturbation. It could be related to the group velocity of the most unstable mode in the system (which was also checked visually on graphical outputs of the model).

We conduct twin experiments with the goal of tracking the truth. Firstly, we illustrate the use of localization in the IEnKF, i.e. the IEnKS with $L = S = 1$ in regimes that are significantly nonlinear. Unlike other EnVar methods, the IEnKS can handle these regimes with a finer variational minimization based on an outer loop. Secondly, we will illustrate the use of localization in the IEnKS, with a varying DAW length. In this regime, the IEnKS relies significantly on the prior and strong transmission of information in the forecast, so that most of the time only one iteration of the IEnKS variational analysis is required.

In the following, the system will be observed fully. Observations are generated from the truth using the observation operator $\mathbf{H} = \mathbf{I}_M$. They are perturbed independently with a white unbiased Gaussian noise of variance 1. Hence, the observation-error covariance matrix, which is assumed to be known, is $\mathbf{R} = \mathbf{I}_M$. The interval between observation acquisitions is Δt, which will be varied in the first experiment. The metric we use to measure the distance of the analysis from the truth is the average root-mean-square error (RMSE). The runs are $2 \times 10^5 \Delta t$ long with an additional spin-up of $5 \times 10^3 \Delta t$. These lengths ensure that the statistical indicators have converged.

Here, covariance localization schemes rely on the Schur product of the sampled covariance matrix with the fifth-order piecewise rational correlation function $r \mapsto \mathrm{GC}(r/c)$ of Gaspari and Cohn (1999), with a localization length of $c = 12$ grid size units. Smooth domain localization is implemented around η by dividing the observation variance at η' by $\mathrm{GC}(\|\eta - \eta'\|/c)$, using the same $c = 12$.

Inflation is addressed systematically with the finite-size scheme (Bocquet, 2011; Bocquet and Sakov, 2012), an adaptive scheme that avoids the burden of tuning inflation, accounting for sampling errors in the variances, and yields performances similar to those obtained by optimally tuned inflation.

5.1. *Illustration of the use of localization in the IEnKF*

The first experiment illustrates local variants of the IEnKF, when the time interval between updates is increased progressively from $\Delta t = 0.05$ to $\Delta t = 0.60$. With increasing Δt, the model nonlinearity acts over longer periods and the EnKF becomes less efficient. This deficiency is addressed by the IEnKF (Sakov *et al.*, 2012) using several iterations in the variational analysis.

The average analysis RMSEs of several approaches are reported in Figure 5. We compare the following.

- For reference: the global ensemble square root Kalman filter with an ensemble of size $N = 20$ [R_1].
- For reference: the global IEnKF with an ensemble of size $N = 20$, larger than the unstable and neutral subspace size (14), so that localization is unnecessary in this regime [R_2].
- The IEnKF with an ensemble of size $N = 10$, with covariance localization using the propagation of $N_e = 40$ modes throughout the DAW [R_3]. Following path C of section 3, the method is described at the end of section 4. This approach is not scalable and cannot be used with higher dimensional systems. It serves here as a baseline where the localized covariance matrix at t_0 is propagated fully and correctly within the DAW. It is theoretically equivalent to the performance of the IEnKS where an adjoint model is used. Hence, one cannot hope to achieve a better precision than this method using covariance localization at t_0 and perfect propagation of the regularized errors throughout the DAW.
- The static domain localization of the IEnKF (Algorithm 2) with an ensemble of size $N = 10$ [DL_1].
- The domain localization of the IEnKF with an ensemble of size $N = 10$ with covariant domains [DL_2] (Algorithm 2 with covariant domains). Covariant domains means that the analysis at η_0 is performed using observations in the domain centred at $\eta_L = \mathcal{S}_{L:0}(\eta_0)$, where η_L is the end point of the characteristic curve. Since there is no such curve in the Lorenz-95 model, here $\mathcal{S}$ is the advection with the constant velocity $v_p = 6.6$ grid cells per time unit, as determined in Figure 4. Note, however, that the motion of this domain is only correct on average.

The static domain localization implementation performs rather well. As explained in section 4, it does require adjusting the threshold to $e = 10^{-2}$ for optimal performance, an issue that has never been a concern with the global IEnKF. With a velocity 6.6, the shift between the centres of the initial domain and of the image

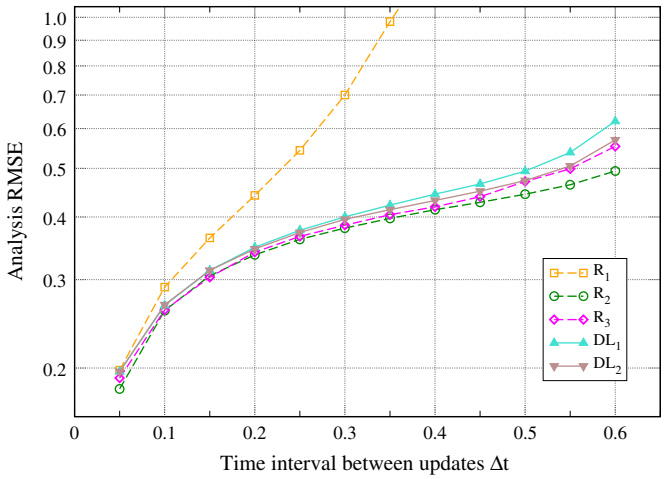

Figure 5. Average analysis RMSEs as a function of the time step between updates for several systems based on the IEnKF that are described in the text. Three systems are references and baseline performances, two are domain localization IEnKF.

domain is about four grid cells for $\Delta t = 0.60$. The local domain is large enough to accommodate this shift approximately. That said, domain localization with covariant domains performs better and for long Δt, almost as well as the baseline covariance localization scheme with unapproximated propagation of the covariance. This idea of covariant domains for localization with a predetermined velocity and applied to the Lorenz-95 model has also been put forward by Kalnay *et al.* (2012), but in the context of sensitivity analysis and ensemble forecast based on a local EnKF.

5.2. *Illustration of the use of localization in the IEnKS*

The second experiment illustrates localization in the IEnKS, when the DAW length L is increased progressively from $L = 0$ (the EnKF) to $L = 30$ (7.5 days), while the observations are collected every $\Delta t = 0.05$. The filtering performance (analysis at present time t_L) and also the smoothing performance (analysis at the beginning of the DAW t_0, i.e. $L\Delta t$ in the past) are reported in Figure 6. The efficiency of the IEnKS compared with the EnKF, the ensemble Kalman smoother (EnKS: Evensen, 2009; Cosme *et al.*, 2012) and 4D-Var has been discussed in Bocquet and Sakov (2013). Here, we compare the following.

- For reference: the global EnKS with an ensemble of size $N = 20$ where localization is unnecessary [R_1].
- For reference: the global IEnKS with an ensemble of size $N = 20$ where localization is unnecessary [R_2].
- Similarly to the IEnKF experiment, the IEnKS with an ensemble of size $N = 10$, with a covariance localization using the propagation of $N_e = 40$ modes throughout the DAW [R_3]. This serves as a baseline performance for covariant CL localization.
- Like the IEnKF experiment, static domain localization of the IEnKF with an ensemble of size $N = 10$ [DL_1].
- Like the IEnKF experiment, the domain localization of IEnKF with an ensemble of size $N = 10$ with covariant domains [DL_2].
- Static (but standard) domain localization of the EnKS with an ensemble of size $N = 10$ [DL_3].
- The IEnKS with an ensemble of size $N = 10$, with a covariance localization using the propagation of $N_e = 15$ modes throughout the DAW, just above the size of the unstable and neutral subspace size [CL_1].

We observe that the fully consistent propagation of the regularized covariance in the covariance localized IEnKS [R_3] is not as successful as the global IEnKS with a sufficiently large ensemble [R_2]. However, it does outperform the EnKS. The propagation of a restricted number of dominant modes [CL_1] has disappointing performance, casting some doubt on the interest in using this scheme in such a chaotic system. As for the IEnKF, the static domain localization approach [DL_1] performs satisfyingly but degrades beyond $L = 6$ for filtering and beyond $L = 13$ for smoothing. However, the domain localization approach with covariant domains [DL_2] performs very well up to $L = 15$ for filtering and $L = 20$ for smoothing, almost as good as the baseline performance covariance localization [R_3].

With DAW lengths of up to $L = 30$, the dynamical shifts in the domain are possibly larger than in the IEnKF experiment (of about 10 grid cells). The gain in covariant domain localization using characteristics or an approximation thereof is very significant. More surprisingly, it is almost as good as the dynamically consistent but non-scalable CL IEnKS, which amounts to using the adjoint model in order to propagate the covariances within the DAW implicitly.

6. Numerical illustration with a two-dimensional model

The localization of the IEnKS is also tested with a two-dimensional vorticity field q, which satisfies the following partial differential equations:

$$\frac{\partial q}{\partial t} + J(\psi, q) + \beta \frac{\partial \psi}{\partial x} = -\xi q - \nu \Delta^4 q + F, \qquad q = \Delta \psi, \tag{29}$$

where $J(\psi, q) = \partial_x \psi \partial_y q - \partial_y \psi \partial_x q$ and ψ is the two-dimensional stream function. The domain of simulation is delimited by $0 \leq x < 1, 0 \leq y < 1$, with doubly periodic boundary conditions.

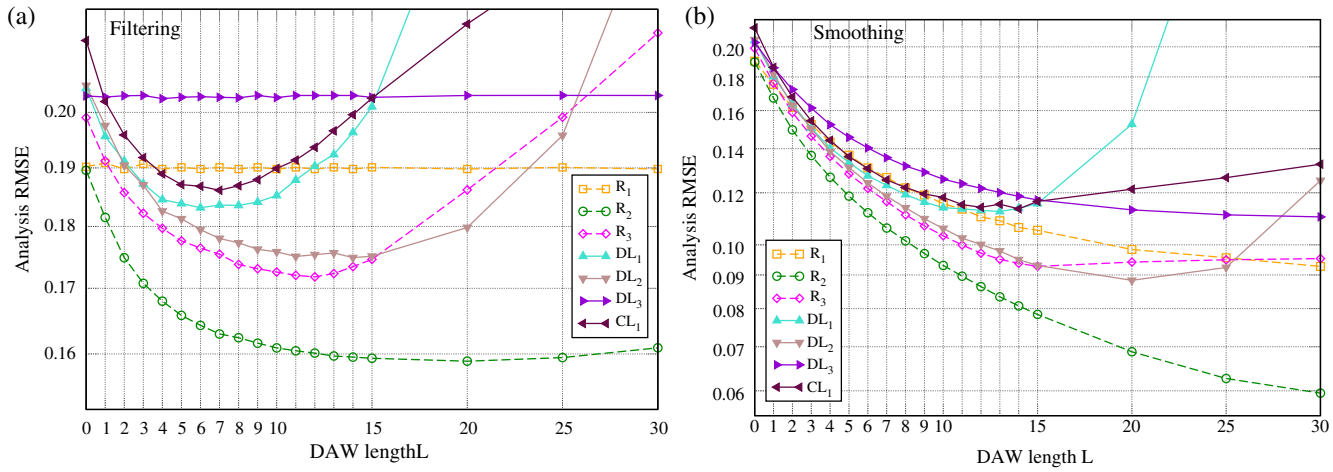

Figure 6. (a) Average filtering and (b) smoothing RMSEs as a function of the DAW length for several systems based on the IEnKS or EnKS that are described in the text. Three systems offer references and baseline performances, three are domain localization IEnKS and one is a covariance localization IEnKS.

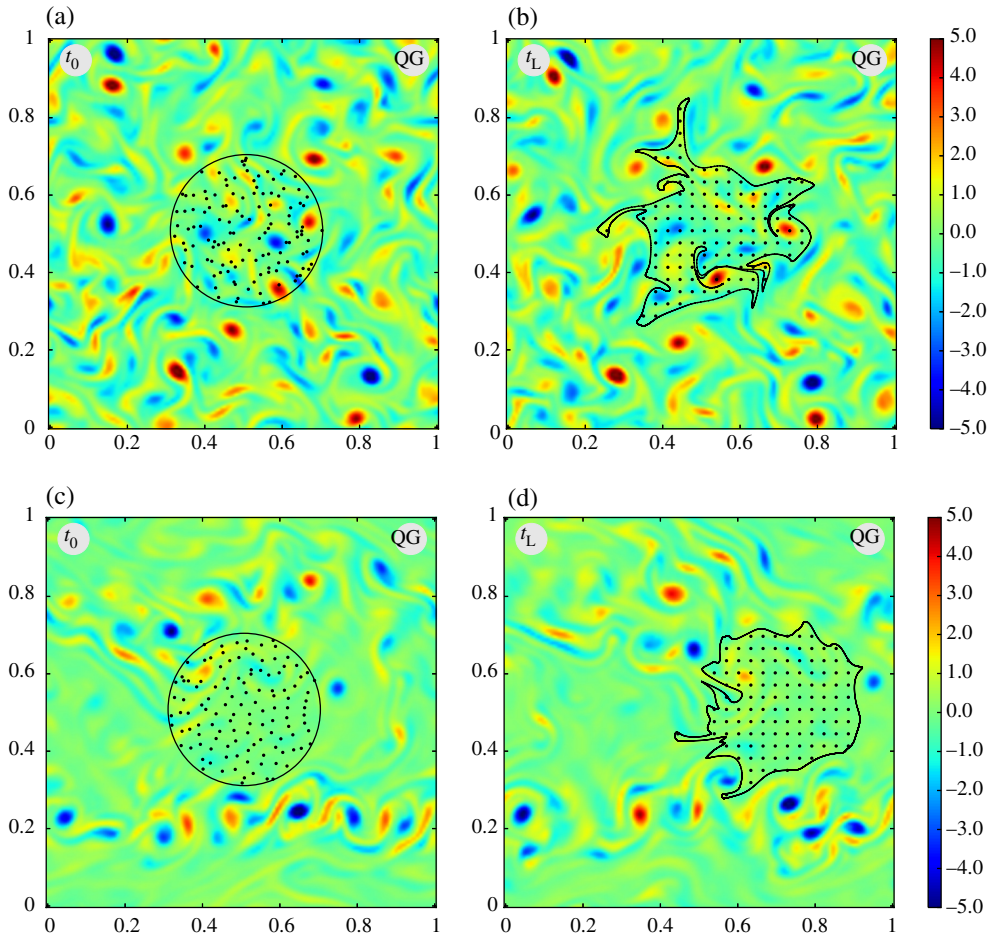

Figure 7. Representation of the vorticity field at the beginning of the DAW (t_0, a,c) and at the end of the DAW (t_L, b,d). The circle at t_0 delimits the domain with centre hosting the vorticity variable to be updated locally. The deformed domain at t_L, which is the result of the surrogate advection from t_0 up to t_L, encompasses the most influential observations to be used in the local update, the locations of which are marked with crosses. Their pulled back locations are also represented at t_0 (left panels). The TB model is used in the top panels, while the QG model is used in the bottom panels.

Here F is a deterministic forcing given by

$$F(x, y) = \Phi \sin(16\pi x) \sin(16\pi y)\,, \tag{30}$$

ξ sets the amplitude of the friction that is meant to extract energy at the largest scales and ν is the amplitude of the hyperdiffusion, which is meant to dissipate vorticity at the smallest scales. We have chosen $\xi = 10^{-2}$ s^{-1} and $\nu = 10^{-19}$ m^8 s^{-1}. Finally, the β-plane term on the left-hand side of the vorticity equation can be switched on ($\beta = 1$) or off ($\beta = 0$).

The two-dimensional grid is of size 128×128. The equations are integrated in space using a fully dealiased pseudo-spectral method (Patterson Jr and Orszag, 1971) and integrated in time using an exponential time-differencing fourth-order Runge–Kutta method (Kassam and Trefethen, 2005), with a time step of 0.05. This allows us to use a large range of diffusion parameters and avoids the numerical annoyance of their stiffness in spectral space.

Without the β-plane term ($\beta = 0$), the model simulates forced two-dimensional turbulence (denoted TB), which is statistically homogeneous. In this case, we choose the forcing magnitude to be $\Phi = 0.5$. In the presence of the β-plane term ($\beta = 1$), the model simulates a single-layer quasi-geostrophic flow (denoted QG). An eastward-bound zonal jet is created in the region $y \in [0.25, 0.75]$, while a jet moving in the opposite direction forms in the region $y \in [0, 0.25] \cup [0.75, 1]$. The flow is less turbulent in the central region of the jets, whereas vortices are mainly created in the storm tracks $y \sim 0.25$ and $y \sim 0.75$. Hence, the flow is quite inhomogeneous. In this case, we chose $\Phi = 0.6$, higher than in the TB case because part of the energy used to generate turbulence is diverted into the jet kinetic energy.

The system is partially observed on a regular square mesh, one observation for 16 grid points, i.e. $d = 1024$ observations for 16 384 vorticity variables. A frequency of observation of $\Delta t = 0.5$ corresponds to a model autocorrelation of 0.96, with a similar structure function to the Lorenz-95 model over short time intervals. The error covariance matrix is chosen to be $\mathbf{R} = 0.09\mathbf{I}_d$, which corresponds to about one tenth of the vorticity climatological variability in the following experiments.

The mechanism of covariant localization in the interval $[t_0, t_L]$ is illustrated in Figure 7 for both models TB (top) and QG (bottom). Circular domains of radius 25 grid cells are depicted in the left panels. They are centred around the grid cell for which the variables should be updated. The surrogate model is defined to be advection by the wind defined by the stream function ψ. The length of the DAW is set to $L = 10$, i.e. a time interval of $L\Delta t = 5$ time units. The images at t_L of the initial domains are shown in the right panels. They encompass the observations that must be selected for local analysis. The pulled back sites of these observations are shown in the left panels.

In all data assimilation experiments, the ensemble size of the IEnKS is set to $N = 20$ and the localization radius of the GC function used in the local analysis scheme is set to $c = 25$ grid spacings (which is near to optimal in both cases).

Similarly to the Lorenz-95 experiments, we first consider a synthetic experiment with the IEnKF ($S = 1$, $L = 1$) where we choose $\Delta t = 3$, greater that the mildly nonlinear $\Delta t = 0.5$. This significantly nonlinear regime requires a multiplicative inflation on the prior ensemble of about 1.5. Static and covariant domain localizations are used. The probability density function (pdf) of the filtering analysis errors, accumulated over a run of 10^3 times units after a spin-up of 50 time units, is plotted in Figure 8. Each scalar analysis error can be seen as the outcome of a local analysis. Similarly to the Lorenz-95 experiments, the second experiment uses the IEnKS ($S = 1, L = 10$) based on a longer DAW, but with

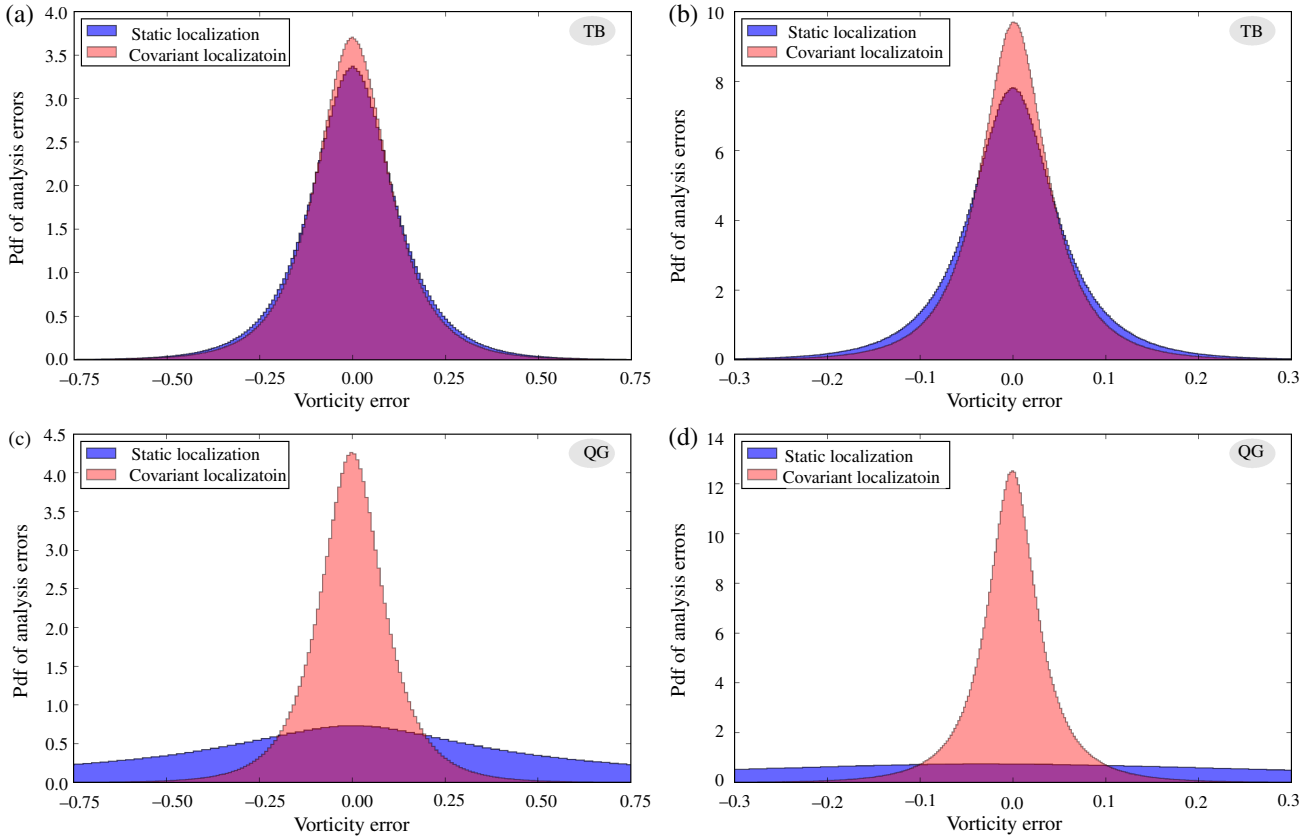

Figure 8. Probability density functions of the pointwise filtering analysis errors for (a,c) the IEnKF with $\Delta t = 3$ and (b,d) the IEnKS with $L = 10$ and $\Delta t = 0.5$, applied to (a,b) the TB model and (c,d) the QG model.

the mildly nonlinear $\Delta t = 0.5$, in a regime where 4D-LETKF and 4DEnVar could be used too, since, on average, one outer loop is sufficient. Hence these data assimilation runs are more Gaussian, requiring a multiplicative inflation on the prior ensemble of about 1.05. The pdf of the errors is plotted in Figure 8 for the static and covariant domain localization IEnKS. Since the DAW is longer, the perturbations are displaced further out and the benefit of covariant localization should be stronger. The two types of experiment are carried out for both models TB and QG.

Covariant localization leads to a systematic improvement, which can be significant when the DAW gets longer. Moreover, the static localization implementation actually leads to divergence from the truth when applied to the QG model (yielding an almost flat pdf of the errors). Indeed, the advection is strong in the zonal jets, so that the observations within a local domain are not relevant to updating the local vorticity. Because the TB model is homogeneous with a null mean flow, the observations at t_L remain relevant to the local vorticity at t_0 over a longer period. Still, even with the TB model, covariant localization systematically improves the stability and the performance of the IEnKS.

7. Conclusion

Localization of four-dimensional ensemble variational data assimilation methods requires special attention because it is necessary to define tapering of covariances consistently across the full 4D data assimilation window. We have studied such 4D localization in the framework of the iterative ensemble Kalman smoother, using a perfect model assumption.

We have shown that a localization operator consistent with the dynamics has a time evolution that satisfies a Liouville equation. This equation simplifies if the localization operator consists of a Schur product with correlation function, but this restricted equation is not closed for all types of dynamics. However, it closes if the dynamics are of hyperbolic type. In that case, the localization function has the same time evolution as the correlation function. If the dynamics are close to hyperbolic, the covariance function as well as the localization function can be evolved using integrations along characteristic curves. Despite this, the range of dynamics of geophysical systems is vast, so that this picture is not always expected to hold, especially in turbulent geo-fluids. A local average of the fluid velocity fields may be necessary. For strongly chaotic dynamics, the fluid velocity, or a local average of it, could be inappropriate. In this case, climatology could be used to infer the evolution of the localization function instead.

Using covariance localization, we found that the most appealing route consists of (i) propagating the ensemble within the DAW, (ii) propagating the localization operator within the DAW using the characteristic curves of a surrogate model and (iii) regularizing covariances applying this localization operator to the 4D ensemble. Another possible method is to keep the dominant modes of the regularized covariance matrix at the beginning of the DAW and propagate these modes within the DAW, from which sampled covariances are computed.

If an adjoint model is available and used in the IEnKS to compute the 4D sensitivities, there is no need to emulate a covariant localization operator, since localization is propagated implicitly as in 4D-Var-Ben or E4DVar. Moreover, it was explained that the adjoint is also very useful in updating the perturbations of the IEnKS, going further in that respect than 4D-Var-Ben or E4DVar.

Domain localization was also discussed in this 4D context. A static domain localization for the IEnKS has fixed domains: a point is influenced by past, present and future observations spatially confined within a static domain around this point. One can borrow the previous findings on 4D covariance localization to improve on this naive approach. A point at the beginning of the DAW is influenced by observations contained within the image of the domain centred on this point by the hyperbolic flow or other surrogate advection model. This results in a simple modification of the static domain localization IEnKS by pulling back the locations of these observations and considering those falling back into the original domain.

These ideas and algorithms have been illustrated with the Lorenz-95 low-order model and a 2D barotropic vorticity equation, which can represent forced 2D turbulence or a single-layer QG model. Among the main results, it was observed that the static domain localization approach, although not optimal, is satisfying in practice for a moderately long data assimilation window. Moreover, the improved covariant domain localization approach turned out to be efficient. In the case of the Lorenz-95 model, it was almost as efficient (without the adjoint model) as the baseline performance obtained by a full propagation of the regularized error covariance matrix at the beginning of the DAW, a performance that can only be obtained in high-dimensional systems with the help of the adjoint model. With the 2D vorticity model, covariant localization led to a systematic stability and performance improvement. It was obvious with the QG model, which has zonal jets with a significant mean flow, but it was also observed to a lesser extent with the TB model, which has a null mean flow.

These considerations are directly relevant to 4DEnVar systems, which require the localization of 4D ensembles. The covariant domain localization obtained by pulling back observations within the DAW is applicable to any 4DEnVar system that would be based on domain localization. Moreover, these considerations point to the fact that 4D localization will be even more critical if outer loop iterations are considered in 4DEnVar systems and critical in an update of the perturbations that is fully consistent with the analysis, two key aspects constitutive of the IEnKS. These considerations could be relevant to the control of any chaotic geophysical flow, such as those met in meteorology and oceanography.

As is much emphasized here, improved 4D localization schemes with the IEnKS and other 4DEnVar systems depend strongly on the dynamics. Hence, further experiments with other types of dynamics are still required to validate these ideas fully; these will be the focus of subsequent studies. In particular, the addition of some diffusion in the propagation of covariant localization meant to account for model diffusion, but which does not fit well into the hyperbolic view of the issue, needs to be investigated.

Acknowledgements

The author is grateful to Pavel Sakov for co-initiating this work and for enlightening discussions. He also thanks Jean-Matthieu Haussaire for a careful reading of the manuscript. The author thanks the Associate Editor, Peter Houtekamer, and two anonymous reviewers for their useful comments, suggestions and thorough reading of the manuscript. The author also acknowledges illuminating exchanges with Gérald Desroziers, Loïk Berre and Yann Michel on improving covariance localization in 4DEnVar. This article is a contribution to the INSU LEFE/MANU DAVE project.

Appendix A: Exploring the consequences of the Liouville equation

In this Appendix, several consequences of the Liouville equation Eq. (14) for the localization operator are studied in the restricted framework of *univariate* and *one-dimensional* systems. Generalizations to two- or three-dimensional systems do not present any new mathematical difficulty. We will use a *continuous* formalism in space and time. We will be guided by ideas from the inspiring articles by Cohn (1993) and Ménard *et al.* (2000).

A1. Covariances

Let us consider a one-dimensional model of state $\eta \in \mathbb{R} \mapsto x(\eta) \in \mathbb{R}$, i.e. a function in $\mathbb{R}$. The dynamics of the trajectory $(\eta, t) \mapsto x(\eta, t)$ are governed by

$$\frac{\partial x}{\partial t} = \mathcal{M}(x)\,. \tag{A1}$$

We are interested in the evolution in time of covariance functions $(\eta_1, \eta_2, t) \mapsto p(\eta_1, \eta_2, t)$, which are continuum limits of error covariance matrices. The related tangent linear equation that describes the evolution of a perturbation e that depends on (η, t) is

$$\frac{\partial e}{\partial t} = \mathbf{M}e\,, \tag{A2}$$

where $\mathbf{M}$ is the tangent linear model derived from $\mathcal{M}$. One can either think of the model dynamics as stochastic and see covariance functions $p(\eta_1, \eta_2, t)$ as an expectation over perturbations defined at time t $\overline{\{e(\eta_1) - \overline{e}(\eta_1)\}\{e(\eta_2) - \overline{e}(\eta_2)\}}$ (Cohn, 1993) or one can see the covariance functions as sampled from an ensemble of perturbations defined at time t:

$$p(\eta_1, \eta_2) = \frac{1}{N-1}\sum_{n=1}^{N}\left\{e_{[n]}(\eta_1) - \overline{e}(\eta_1)\right\}\left\{e_{[n]}(\eta_2) - \overline{e}(\eta_2)\right\}\,, \tag{A3}$$

where $\overline{e}$ is the ensemble mean. Either way, the time evolution of p can be computed using the Leibniz rule and one obtains

$$\frac{\partial p}{\partial t} = \{\mathbf{M}(\eta_1) + \mathbf{M}(\eta_2)\}\,p\,. \tag{A4}$$

This relates to the computation of $\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}$, but not to that of $\mathbf{X}_0\mathbf{Y}_k^{\mathrm{T}}$, which connects two distinct times within the DAW. Hence, we are also interested in asynchronous covariances such as $p_1(\eta_1, t, \eta_2)$, which is defined as the covariance between a point η_2 at some *fixed time*, say t_0, and a point η_1 at time t. The time evolution of such an asynchronous covariance is

$$\frac{\partial p_1}{\partial t} = \mathbf{M}(\eta_1)p_1\,. \tag{A5}$$

In the following we shall mostly work with synchronous covariance function Eq. (A4), but it should be kept in mind that asynchronous covariance functions can be dealt with as well, following Eq. (A5).

A2. Liouville equation for the localization operator

Let us denote by $\mathcal{L}$ the functional operator for localization: $p \mapsto \mathcal{L}\cdot p$. It acts on covariance functions $p(\eta_1, \eta_2)$, but it does not have to be a pointwise product. It is assumed to be linear and may depend on time. If p is a sampled covariance, we know that p follows Eq. (A4). To make the regularized covariance $\mathcal{L}\cdot p$ consistent throughout the DAW, we would like to impose this constraint on $\mathcal{L}\cdot p$ as well, i.e.

$$\frac{\partial \mathcal{L}\cdot p}{\partial t} = \{\mathbf{M}(\eta_1) + \mathbf{M}(\eta_2)\}\,\mathcal{L}\cdot p\,. \tag{A6}$$

Since $\mathcal{L}$ is linear, we can apply the Leibniz rule to the time derivation of $\mathcal{L}\cdot p$ in the left-hand side of Eq. (A6), for any p, and use Eq. (A4) to infer

$$\frac{\partial \mathcal{L}}{\partial t} = [\mathbf{M}(\eta_1) + \mathbf{M}(\eta_2), \mathcal{L}]\,, \tag{A7}$$

which is one instance of the Liouville equation Eq. (14). A similar equation can be obtained for the asynchronous localization operator.

A3. Closure of the Liouville equation for hyperbolic dynamics

As explained by Cohn (1993), it is difficult to explore such results further without making assumptions regarding the dynamics.

In particular, one cannot write a general equation for the time evolution of variances or correlations in a closed form.

Hence, it is assumed further that the dynamics are hyperbolic: $\mathcal{M}(x, \partial_\eta x)$ depends only on x or on the first-order partial derivative of x. Consequently, the tangent linear model equation is of the form

$$\frac{\partial e}{\partial t} = \frac{\partial \mathcal{M}}{\partial x} e + \frac{\partial \mathcal{M}}{\partial(\partial_\eta x)} \frac{\partial e}{\partial \eta}, \tag{A8}$$

so that the tangent linear operator can be written in the generic form $\mathbf{M} = a(\eta, t)\partial_\eta + b(\eta, t)$ following the notations of Cohn (1993).

With this assumption on the dynamics, Cohn (1993) has shown that the covariance p and the standard deviation σ satisfy

$$\frac{\partial p}{\partial t} = \left\{a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2} + b(\eta_2) + b(\eta_1)\right\} p\,, \tag{A9}$$

$$\frac{\partial \sigma}{\partial t} = \left\{a(\eta)\partial_\eta + b(\eta)\right\} \sigma\,, \tag{A10}$$

from which the equation for the correlation function $c(\eta_1, \eta_2) = p(\eta_1, \eta_2)/(\sigma(\eta_1)\sigma(\eta_2))$ can be obtained as

$$\frac{\partial c}{\partial t} = \left\{a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2}\right\} c\,. \tag{A11}$$

Partial differential equations Eqs (A9)–(A11) can in principle be solved using integration via the characteristics method (see for instance Evans *et al.*, 1999).

With the same dynamics and Eq. (A7), the localization operator obeys

$$\frac{\partial \mathcal{L}}{\partial t} = [a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2} + b(\eta_1) + b(\eta_2), \mathcal{L}]\,. \tag{A12}$$

The computation of the commutator leads to

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} = &\left\{a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2}\right\} \mathcal{L} \\ &+ [a(\eta_1), \mathcal{L}]\partial_{\eta_1} + [a(\eta_2), \mathcal{L}]\partial_{\eta_2} \\ &+ [b(\eta_1), \mathcal{L}] + [b(\eta_2), \mathcal{L}]\,.\end{aligned} \tag{A13}$$

In particular, we observe that $\mathcal{L}$ generally depends on partial derivative operators. One sufficient condition to close this equation within a restricted operator space for $\mathcal{L}$ is to look for solutions that are Schur-product operators: $\mathcal{L} : p \mapsto \rho \circ p = \rho\, p$, where ρ is a positive-definite correlation function. In that case, $\mathcal{L}$ will be identified as ρ. All four commutators in Eq. (A13) vanish and the Liouville equation becomes

$$\frac{\partial \rho}{\partial t} = \left\{a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2}\right\} \rho\,, \tag{A14}$$

which coincides with the equation for the correlations Eq. (A11). This makes a connection between the localization function (for a localization operator of a Schur form) and the correlation function when the dynamics is hyperbolic. This equation can also be solved using characteristics and the localization operator at the initial time. Moreover, it shows that the propagation of the localization function is simpler than that of the covariances. For instance, *in situ* constituent chemical reactions, which would appear in $b(\eta, t)$, need to be modelled only in the propagation of perturbations, not in the Schur localization function propagation. Even for the hyperbolic one-dimensional inviscid Burgers equation of state vector $x(\eta, t)$, where $a(\eta, t) = -x$, $b(\eta, t) = -\partial x/\partial\eta$, the integration along characteristics would be simplified into mere advection.

It is also possible to investigate more general dynamics. If the dynamics are parabolic with a tangent linear operator of the form $\mathbf{M} = a(\eta)\partial_\eta + b(\eta) + \nu(\eta)\partial_\eta^2$, then the Liouville equation cannot be simply closed even if $\mathcal{L}$ is a Schur product. Indeed, we obtain in this case

$$\begin{aligned}\frac{\partial \rho}{\partial t} = &\left\{a(\eta_1)\partial_{\eta_1} + a(\eta_2)\partial_{\eta_2} + \nu(\eta_1)\partial_{\eta_1}^2 + \nu(\eta_2)\partial_{\eta_2}^2\right\} \rho \\ &+ 2\nu(\eta_1)\frac{\partial \rho}{\partial \eta_1}\partial_{\eta_1} + 2\nu(\eta_2)\frac{\partial \rho}{\partial \eta_2}\partial_{\eta_2}\,.\end{aligned} \tag{A15}$$

An approximate closure of this equation amounts to ignoring the operators in the second line.

A4. Evolution of the localization length

With the localization function having the same dynamics as the correlation function, we use the results of Cohn (1993) to our advantage in order to determine the time evolution of a typical localization length. The localization function, as a correlation function, is symmetric in η_1 and η_2; it equals 1 when $\eta_1 = \eta_2$, so that $\eta = (\eta_1 + \eta_2)/2$ and $\xi = (\eta_1 - \eta_2)/2$ are adequate variables for a finer description. Moreover, homogeneity and isotropy of this localization function are naturally discussed in terms of η and ξ. In one dimension, the local curvature of the localization function at η will be given by

$$\rho_2(\eta) = \frac{\partial^2 \rho}{\partial \xi^2}(\eta, \xi = 0)\,. \tag{A16}$$

It is related to some localization length-scale Λ by $\Lambda = -1/\rho_2^2$. Piggybacking on the findings of Cohn (1993), an equation for $\rho_2(\eta)$ can be found by expanding Eq. (A14) in the separation variable ξ. At second order in ξ, one finds

$$\frac{\partial \rho_2}{\partial t} = \left\{a(\eta)\partial_\eta + 2\frac{\partial a(\eta)}{\partial \eta}\right\} \rho_2\,, \tag{A17}$$

which can still be solved using time integration by the characteristics method. It also shows that the localization length is governed by the same equation as that of the error correlation length.

Finally, let us mention that the method developed by Pannekoucke *et al.* (2014) could be a promising approach to propagate ρ using the local metric tensor of ρ and its propagation. The time evolution of ρ_2 offers a direct connection to this earlier work, since it relates directly to the metric tensor.

Appendix B: Proof of equation (25)

Let f be an analytic function, such that $f(0) = 1$. Its formal power series is $f(x) = \sum_{n=0}^{\infty} \alpha_n x^n$, with $\alpha_0 = 1$. Define $g(x) = (f(x) - 1)/x$. The formal power series of g is $g(x) = \sum_{n=0}^{\infty} \alpha_{n+1} x^n$. If $\mathbf{A}$ is a matrix of size $M \times N$ and if $\mathbf{B}$ is a matrix of size $N \times M$, then

$$\begin{aligned}f(\mathbf{AB}) &= \sum_{n=0}^{\infty} \alpha_n (\mathbf{AB})^n \\ &= \mathbf{I}_M + \mathbf{A}\left(\sum_{n=0}^{\infty} \alpha_{n+1} (\mathbf{BA})^n\right)\mathbf{B} \\ &= \mathbf{I}_M + \mathbf{A}g(\mathbf{BA})\mathbf{B}\,.\end{aligned} \tag{B1}$$

In particular, if $f(x) = 1/\sqrt{1+x}$, then $g(x) = -1/(1 + x + \sqrt{1+x})$. When applied to $\mathbf{AB}$, one obtains

$$(\mathbf{I}_M + \mathbf{AB})^{-1/2} = \mathbf{I}_M - \mathbf{A}\left(\mathbf{I}_N + \mathbf{BA} + \sqrt{\mathbf{I}_N + \mathbf{BA}}\right)^{-1}\mathbf{B}\,. \tag{B2}$$

Choosing $\mathbf{A} = \mathbf{X}_0^r$ and $\mathbf{B} = \left(\mathbf{X}_0^r\right)^{\mathrm{T}} \left(\mathbf{F}_{L:0}^\star\right)^{\mathrm{T}} \mathbf{R}_L^{-1} \mathbf{F}_{L:0}^\star$ and applying Eq. (B2), Eq. (24) turns out equivalent to Eq. (25).

References

Anderson JL. 2012. Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Weather Rev.* **140**: 2359–2371.

Anderson JL, Lei L. 2013. Empirical localization of observation impact in ensemble Kalman filters. *Mon. Weather Rev.* **141**: 4140–4153.

Bell BM. 1994. The iterated Kalman smoother as a Gauss–Newton method. *SIAM J Optim.* **4**: 626–636.

Bishop CH, Hodyss D. 2007. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 2029–2044.

Bishop CH, Hodyss D, Steinle P, Sims H, Clayton AM, Lorenc AC, Barker DM, Buehner M. 2011. Efficient ensemble covariance localization in variational data assimilation. *Mon. Weather Rev.* **139**: 573–580.

Bocquet M. 2011. Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Proc. Geophys.* **18**: 735–750, doi: 10.5194/ npg-18-735-2011/.

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

Bocquet M, Sakov P. 2013. Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinar Proc. Geophys.* **20**: 818, doi: 10.5194/npg-20-803-2013.

Bocquet M, Sakov P. 2014. An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.* **140**: 1521–1535, doi: 10.1002/qj.2236.

Buehner M. 2005. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**: 1013–1043.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010. 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.

Chen Y, Oliver DS. 2013. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**: 689–703.

Cohn SE. 1993. Dynamics of short-term univariate forecast error covariances. *Mon. Weather Rev.* **121**: 3123–3149.

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

Desroziers G, Camino JT, Berre L. 2014. 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Q. J. R. Meteorol. Soc.* **140**: 2097–2110, doi: 10.1002.2325/qj.

Evans G, Blackledge J, Yardley P. 1999. *Analytic Methods for Partial Differential Equations*. Springer-Verlag: Berlin.

Evensen G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53**: 343–367.

Evensen G. 2009. *Data Assimilation: The Ensemble Kalman Filter* (4th edn). Springer-Verlag: Berlin Heidelberg.

Fairbairn D, Pring SR, Lorenc AC, Roulstone I. 2014. A comparison of 4DVar with ensemble data assimilation methods. *Q. J. R. Meteor. Soc.* **140**: 281–294.

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757.

Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790.

Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.

Hunt BR, Kalnay E, Kostelich EJ, Ott E, Patil DJ, Sauer T, Szunyogh I, Yorke JA, Zimin AV. 2004. Four-dimensional ensemble Kalman filtering. *Tellus A* **56**: 273–277.

Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230**: 112–126.

Kalnay E, Ota Y, Miyoshi T, Liu J. 2012. A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus A* **64**: 18 462.

Kassam AK, Trefethen LN. 2005. Fourth-order time-stepping for stiff PDEs. *SIAM J. Sci. Comput.* **26**: 1214–1233.

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

Liu C, Xiao Q, Wang B. 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Weather Rev.* **136**: 3363–3373.

Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**: 3183–3203, doi:10.1256/qj.02.132.

Lorenz EN, Emanuel KE. 1998. Optimal sites for supplementary weather observations: Simulation with a small model. *J Atmos. Sci.* **55**: 399–414.

Majda A, Abramov RV, Grote MJ. 2005. *Information Theory and Stochastics for Multiscale Nonlinear Systems*, Vol. 25. American Mathematical Society: Providence, Rhode Island USA.

Menard R, Cohn SE, Chang LP, Lyster PM. 2000. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation. *Mon. Weather Rev.* **128**: 2654–2671.

Ménétrier B, Montmerle T, Michel Y, Berre L. 2015. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Weather Rev.* **143**: 1622–1643.

Nerger L, Janjić T, Schröter J, Hiller W. 2012. A regulated localization scheme for ensemble-based Kalman filters. *Q. J. R. Meteorol. Soc.* **138**: 802–812, doi: 10.1002/qj.945.

Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke A. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* **56**: 415–428.

Pannekoucke O, Emili E, Thual O. 2014. Modelling of local length-scale dynamics and isotropizing deformations. *Q. J. R. Meteorol. Soc.* **140**: 1387–1398.

Patterson GS, Jr, Orszag SA. 1971. Spectral calculations of isotropic turbulence: Efficient removal of aliasing interactions. *Phys. Fluids* **14**: 2538–2541.

Pires C, Vautard R. Talagrand 0.1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A* **48**: 96–121.

Poterjoy J, Zhang F. 2015. Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background-error covariance: E4DVar versus 4DEnVar. *Mon. Weather Rev.* **143**: 1601–1621.

Risken H. 1989. *The Fokker-Planck Equation* (2nd edn). Springer: Berlin Heidelberg.

Sakov P, Bertino L. 2011. Relation between two common localisation methods for the EnKF. *Comput. Geosci.* **15**: 225–237.

Sakov P, Oke PR. 2008. A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus A* **60**: 361–371.

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

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

Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Ensemble square root filters. *Mon. Weather Rev.* **131**: 1485–1490.

Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**: 1913–1924.

Zupanski M. 2005. Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Weather Rev.* **133**: 1710–1726.

