A Matrix-Free Posterior Ensemble Kalman Filter Implementation Based on a Modified Cholesky Decomposition

Elias D. Nino-Ruiz

Applied Math and Computational Science Laboratory, Department of Computer Science, Universidad del Norte, Barranquilla 080001, Colombia; enino@uninorte.edu.co

Abstract: In this paper, a matrix-free posterior ensemble Kalman filter implementation based on a modified Cholesky decomposition is proposed. The method works as follows: the precision matrix of the background error distribution is estimated based on a modified Cholesky decomposition. The resulting estimator can be expressed in terms of Cholesky factors which can be updated based on a series of rank-one matrices in order to approximate the precision matrix of the analysis distribution. By using this matrix, the posterior ensemble can be built by either sampling from the posterior distribution or using synthetic observations. Furthermore, the computational effort of the proposed method is linear with regard to the model dimension and the number of observed components from the model domain. Experimental tests are performed making use of the Lorenz-96 model. The results reveal that, the accuracy of the proposed implementation in terms of root-mean-square-error is similar, and in some cases better, to that of a well-known ensemble Kalman filter (EnKF) implementation: the local ensemble transform Kalman filter. In addition, the results are comparable to those obtained by the EnKF with large ensemble sizes.

Keywords: ensemble kalman filter; posterior ensemble; modified cholesky decomposition

1. Introduction

Data Assimilation is the process by which imperfect numerical forecasts and sparse observational networks are fused in order to estimate the state $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$ of a system [1] which (approximately) evolves according to some model operator,

$$\mathbf{x}_p^* = \mathcal{M}_{t_{p-1} \to t_p}\left(\mathbf{x}_{p-1}^*\right), \text{ for } 1 \le p \le M, \tag{1}$$

where, for instance, $\mathcal{M} : \mathbb{R}^{n \times 1} \to \mathbb{R}^{n \times 1}$ is a numerical model which mimics the ocean and/or the atmosphere dynamics, n is the number of model components, M is the number of observations (which form the assimilation window), and p denotes time index at time t_p. Sequential and smoothing methods are commonly utilized in order to perform the estimation process [2–4]. In the context of sequential data assimilation, a forecast state $\mathbf{x}^b \in \mathbb{R}^{n \times 1}$ is adjusted according to a real observation $\mathbf{y} \in \mathbb{R}^{m \times 1}$, where m is the number of observed components from the numerical grid. When Gaussian assumptions are done over background and observational errors, background (forecast) errors can be characterized by,

$$\mathbf{x} \sim \mathcal{N}\left(\mathbf{x}^b, \mathbf{B}\right), \tag{2}$$

while observational errors are described by,

$$\mathbf{y} \sim \mathcal{N}\left(\mathcal{H}\left(\mathbf{x}\right), \mathbf{R}\right), \tag{3}$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is the unknown background error covariance matrix, $\mathcal{H} : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times 1}$ is the observation operator, and $\mathbf{R} \in \mathbb{R}^{m \times m}$ stands for the data error covariance matrix. By using Bayes rule, the prior (2) and the likelihood (3) distributions can be harnessed in order to estimate the posterior error distribution,

$$\mathcal{P}\left(\mathbf{x}|\mathbf{y}\right) \propto \exp\left(-\mathcal{J}\left(\mathbf{x}\right)\right), \tag{4}$$

where

$$\mathcal{J}\left(\mathbf{x}\right) = \frac{1}{2} \cdot \left\|\mathbf{x} - \mathbf{x}^b\right\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \cdot \left\|\mathbf{y} - \mathcal{H}\left(\mathbf{x}\right)\right\|_{\mathbf{R}^{-1}}^2, \tag{5}$$

is well-known as the three dimensional variational cost function. The posterior mode of the distribution (4) can be computed via the minimizer of the cost function (5),

$$\mathbf{x}^a = \arg\min_{\mathbf{x}} \mathcal{J}\left(\mathbf{x}\right), \tag{6}$$

where $\mathbf{x}^a \in \mathbb{R}^{n \times 1}$ is well-known as the analysis state. The solution of the optimization problem (6) can be computed as follows,

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathcal{H}\left(\mathbf{x}^b\right)\right] \in \mathbb{R}^{n \times 1}, \tag{7a}$$

where $\mathcal{H}'(\mathbf{x}) \approx \mathbf{H}^T \in \mathbb{R}^{n \times m}$ is a linearized observation operator (with the linearization performed about the background state) and the analysis covariance matrix reads,

$$\mathbf{A} = \left[\mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \in \mathbb{R}^{n \times n}. \tag{7b}$$

Typically, the moments of the prior distribution can be estimated based on an ensemble of model realizations [5]. However, since ensemble members come at high computational costs owing to current operational data assimilation settings (i.e., numerical grid resolutions), ensemble sizes are bounded by the hundreds while their underlying error distributions range in the order of billions [6]. Consequently, sampling errors can impact the quality of the analysis state [7]. In practice, localization methods can be utilized in order to mitigate the impact of sampling errors during the assimilation steps [8]. For instance, in the EnKF implementation based on a modified Cholesky decomposition (EnKF-MC) [9,10], the covariance matrix estimator proposed by Bickel and Levina in [11] and the conditional independence of model components regarding their spatial distances are exploited in order to estimate a sparse precision matrix of the background distribution, to reduce the computational cost of the analysis step, and to mitigate the impact of spurious correlations during the assimilation of observations. Given the relation between $\mathbf{A}$ and $\mathbf{B}^{-1}$ in (7b) and by using the Bickel and Levina estimator, we think that a sparse precision matrix of the analysis distribution can be estimated without the needing of actually computing (7b), and therefore, the assimilation of observations can be efficiently performed.

This paper is organized as follows: in Section 2, efficient implementations of the EnKF which account for localization are discussed, Section 3 presents a matrix-free posterior ensemble Kalman filter implementation based on a modified Cholesky decomposition; in Section 4, experimental tests are performed making use of the Lorenz-96 model in order to assess the accuracy of the proposed method, and finally, Section 5 states the conclusions of this research.

2. Preliminaries

2.1. The Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) is a sequential Monte-Carlo method for parameter and state estimation in highly non-linear models [12]. The popularity of the EnKF owes to his simple formulation and relatively ease implementation [13]. In the EnKF context, an ensemble of model realizations is utilized,

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \ldots, \mathbf{x}^{b[N]} \right] \in \mathbb{R}^{n \times N},$$ (8a)

in order to estimate the moments of the background error distribution (2) via the empirical moments of the ensemble,

$$\mathbf{x}^b \approx \bar{\mathbf{x}}^b = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^{b[e]} \in \mathbb{R}^{n \times 1},$$ (8b)

and,

$$\mathbf{B} \approx \mathbf{P}^b = \frac{1}{N-1} \cdot \Delta\mathbf{X} \cdot \Delta\mathbf{X}^T \in \mathbb{R}^{n \times n},$$ (8c)

where N is the number of ensemble members, $\mathbf{x}^{b[e]} \in \mathbb{R}^{n \times 1}$ denotes the e-th ensemble member, for $1 \leq e \leq N$, $\bar{\mathbf{x}}^b$ is the ensemble mean, $\mathbf{P}^b$ is the ensemble covariance matrix, and

$$\Delta\mathbf{X} = \mathbf{X}^b - \bar{\mathbf{x}}^b \cdot \mathbf{1}^T \in \mathbb{R}^{n \times N},$$ (8d)

is the matrix of member deviations with **1** being a vector of consistent dimension whose components are all ones. When an observation **y** becomes available, the analysis ensemble can be computed as follows,

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^a \cdot \Delta\mathbf{Y} \in \mathbb{R}^{n \times N},$$ (9a)

where the scaled matrix of innovations on the perturbed observations $\Delta\mathbf{Y}$ reads,

$$\Delta\mathbf{Y} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} \cdot \mathbf{1}^T + \mathbf{R}^{1/2} \cdot \mathbf{E} - \mathbf{H} \cdot \mathbf{X}^b \right] \in \mathbb{R}^{n \times N},$$ (9b)

the columns of matrix $\mathbf{E} \in \mathbb{R}^{m \times N}$ are samples from a multivariate standard Normal distribution, and

$$\mathbf{P}^a = \left[\left[\mathbf{P}^b \right]^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \in \mathbb{R}^{n \times n},$$ (9c)

is the analysis ensemble covariance matrix.

2.2. Localization Methods

As we mentioned before, the ensemble size is much lower than the model resolution ($N \ll n$) and as a consequence, $\mathbf{P}^b$ is rank-deficient which implies that, the precision matrix (9c) can not be computed. In practice, localization methods are commonly utilized in order to artificially increase the rank of $\mathbf{P}^b$ and to mitigate the impact of spurious correlations during the analysis steps [14]. In general, we can think in two different flavours of localization: covariance matrix localization [15,16], and spatial localization [17].

In the context of covariance matrix localization, a decorrelation matrix $\widehat{\mathbf{\Lambda}} \in \mathbb{R}^{n \times n}$ is typically utilized in order to dissipate spurious correlations between distant model components,

$$\widehat{\mathbf{P}}^b = \widehat{\mathbf{\Lambda}} \circ \mathbf{P}^b \in \mathbb{R}^{n \times n}, \tag{10}$$

where $\circ$ denotes component-wise multiplication, $\widehat{\mathbf{P}}^b$ is a localized covariance matrix, and the components of the localization matrix $\widehat{\mathbf{\Lambda}}$, for instance, reads,

$$\left\{ \widehat{\mathbf{\Lambda}} \right\}_{i,j} = \exp\left(-\frac{1}{2} \cdot \frac{d(i,j)^2}{r^2} \right), \text{ for } 1 \leq i, j \leq n,$$

where r is the localization radius, and $d(i,j)$ stands for the spatial distance between components i and j.

In spatial localization schemes, each model component is surrounded by a local box of radius r and all the information contained within such box is utilized in order to perform a local assimilation step [18]. All local analysis components are then mapped back to the global domain from where a global analysis state is estimated. Some examples of spatial localization are shown in Figure 1 for a two dimensional domain. In general, radius sizes can vary among different model components.
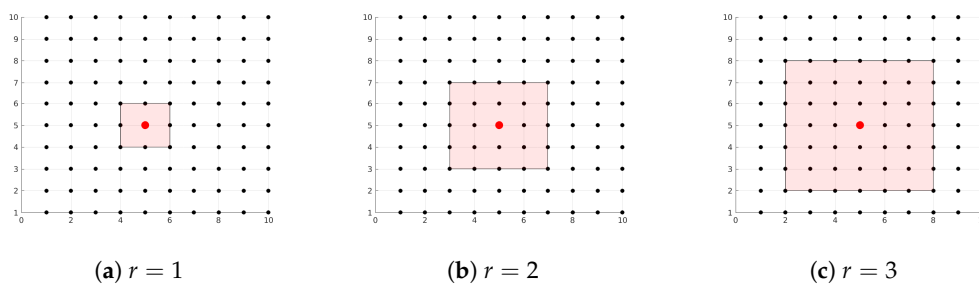

(**a**) $r = 1$　　　　　　(**b**) $r = 2$　　　　　　(**c**) $r = 3$

Figure 1. Local domains for different radii of influence r. The red dot is the model component to be assimilated, the red square denotes components within the scope of r, and components outside the region are unused during the local assimilation process.

It is important to note that, the use of covariance matrix localization or spatial localization during the assimilation of observations should rely, mainly, on computational aspects [19].

2.3. Efficient EnKF Implementations: Accounting for Localization

An efficient EnKF implementation in the context of spatial localization is the local ensemble transform Kalman filter (LETKF) which belongs to the family of deterministic EnKF formulations [20–23]. This method has been sucesfully tested in operational data assimilation centres for weather forecast [24]. In the LETKF, the mode of the analysis distribution is estimated in the ensemble space as follows,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{\Delta X} \cdot \mathbf{w}^a \in \mathbb{R}^{n \times 1}, \tag{11a}$$

where $\mathbf{w}^a = \widetilde{\mathbf{P}^a} \cdot [\mathbf{H} \cdot \mathbf{\Delta X}]^T \cdot \left[\mathbf{y} - \mathcal{H}\left(\overline{\mathbf{x}}^b \right) \right] \in \mathbb{R}^{N \times 1}$, and $\widetilde{\mathbf{P}^a} \in \mathbb{R}^{N \times N}$ is a projection of the analysis covariance matrix (9c) onto such space wherein the ensemble covariance matrix (8c) is well-conditioned. The analysis ensemble is then built about the estimated state (11a) as follows,

$$\mathbf{X}^a = \overline{\mathbf{x}}^a \cdot \mathbf{1}^T + \sqrt{N-1} \cdot \mathbf{\Delta X} \cdot \widetilde{\mathbf{P}^a}^{1/2} \in \mathbb{R}^{n \times N}. \tag{11b}$$

The assimilation step (11) is applied to each model component for a given radius of influence r from where the global analysis state is obtained. Hence, since the most expensive computation during the assimilation step of LETKF is the inversion of $\widetilde{\mathbf{P}^a}$, the computational effort of the LETKF reads,

$$\mathcal{O}\left(\varphi \cdot n \cdot N^3\right),$$

where φ denotes the local box sizes.

Another efficient EnKF implementation which makes use of covariance matrix localization is the EnKF based on a modified Cholesky decomposition (EnKF-MC) [25]. In this context, given a radius of influence r, uncorrelated errors are assumed between the component i and the model components $P(i, r)$, for $1 \leq i \leq n$, where $P(i, r)$ stands for the predecessors of component i whose spatial distances do not exceed r. For a component i, the predecessors $P(i, r)$ depend on some ordering of grid components, for instance, Figure 2 shows an example for a two dimensional domain when $r = 1$, $i = 6$, and grid components are labelled by using column-major format. In the EnKF-MC, the resulting estimator $\widehat{\mathbf{B}}^{-1}$ can be obtained in terms of Cholesky factors by using the Bickel and Levina estimator,

$$\mathbf{B}^{-1} \approx \widehat{\mathbf{B}}^{-1} = \mathbf{L}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{L} \in \mathbb{R}^{n \times n}, \tag{12a}$$

where $\mathbf{L} \in \mathbb{R}^{n \times n}$ is a lower triangular matrix,

$$\{\mathbf{L}\}_{i,j} = \begin{cases} -\beta_{i,j} & , j \in P(i,r) \\ 1 & , i = j \\ 0 & , otherwise \end{cases}, \tag{12b}$$

whose non-zero sub-diagonal elements are obtained by fitting models of the form,

$$\mathbf{x}_{[i]}^T = \sum_{j \in P(i,r)} \beta_{i,j} \cdot \mathbf{x}_{[j]}^T + \gamma_i \in \mathbb{R}^{N \times 1}, \ 1 \leq i \leq n, \tag{12c}$$

$\mathbf{x}_{[i]}^T \in \mathbb{R}^{N \times 1}$ denotes the *i*-th row (model component) of the ensemble (8a), and the components of vector $\gamma_i \in \mathbb{R}^{N \times 1}$ are samples from a zero-mean Normal distribution with unknown variance σ^2. Likewise, $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements read,

$$\{\mathbf{D}\}_{i,i} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[i]}^T - \sum_{j \in P(i,r)} \beta_{i,j} \cdot \mathbf{x}_{[j]}^T\right)^{-1} \approx \mathbf{var}\left(\gamma_i\right)^{-1} = \frac{1}{\sigma^2} > 0, \ \text{with} \ \{\mathbf{D}\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[1]}^T\right)^{-1}, \tag{12d}$$

where $\mathbf{var}(\bullet)$ and $\widehat{\mathbf{var}}(\bullet)$ denote the actual and the empirical variances, respectively.

Figure 2. Local model components (local box) and local predecessors for the model component 6 when $r = 1$. Column-major ordering is utilized to label the model components. (a) In blue, local box for the model component 6 when $r = 1$. (b) In blue, predecessors of the model component 6 for $r = 1$.

By using the precision matrix (12a), the analysis covariance matrix (7b) can be estimated as follows,

$$\widehat{\mathbf{A}} = \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \in \mathbb{R}^{n \times n}. \tag{12e}$$

Thusly, the posterior ensemble is given by,

$$\mathbf{X}^a = \mathbf{X}^b + \widehat{\mathbf{A}} \cdot \mathbf{\Delta Y} \in \mathbb{R}^{n \times N}. \tag{12f}$$

Notice, since the number of predecessors for each model component depends on r, the resulting factor (12b) can be sparse, this implies huge savings in terms of memory usage under current operational data assimilation settings wherein n can be very large. Besides, $\widehat{\mathbf{B}}^{-1}$ can be represented in terms of his Cholesky factors and therefore, efficient manners to compute the ensemble (12f) can be derived. Since the number of predecessors is some multiple of the local box sizes, the computational effort of the analysis step in the EnKF-MC can be expressed as follows,

$$\mathcal{O}\left(\varphi^2 \cdot m \cdot n + \varphi^2 \cdot n \right),$$

which is linear with regard to n and m. Note that, the EnKF-MC and the LETKF provide similar computational costs.

We think that, by updating the Cholesky factors of (12e) it is possible to obtain an estimate of the precision matrix $\mathbf{A}^{-1}$ in terms of sparse Cholesky factors which can be exploited during the assimilation step (12f) in order to avoid the explicit solution of the linear system,

$$\left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right] \cdot \mathbf{Z} = \mathbf{\Delta Y},$$

and yet, to derive another efficient implementation of the assimilation cycle. The next section presents an EnKF implementation based on this general idea.

3. A Posterior Ensemble Kalman Filter Based On Modified Cholesky Decomposition

Before we start the derivation of our filter implementation, we assume that, the observational operator is nearly linear and/or it can be easily applied [26], the data error covariance matrix possesses a simple structure and/or it can be easily decomposed [27], and that observational networks are sparse [28].

3.1. General Formulation of the Filter

To be concise, we want to obtain an estimate of the precision matrix for (12e) of the form,

$$\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} = \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} \in \mathbb{R}^{n \times n}, \tag{13}$$

where $\widehat{\mathbf{L}} \in \mathbb{R}^{n \times n}$ is a sparse lower triangular matrix whose diagonal elements are all ones and $\widehat{\mathbf{D}} \in \mathbb{R}^{n \times n}$ is a diagonal matrix. In this manner, we can approximate the posterior mode of the error distribution,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \cdot \mathbf{\Delta y} = \overline{\mathbf{x}}^b + \mathbf{\Delta x}^a, \tag{14}$$

via the solution of a lower and an upper triangular system of equations,

$$\widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} \cdot \mathbf{\Delta x}^a = \mathbf{\Delta y} \in \mathbb{R}^{n \times 1},$$

which can be solved by using forward and backward substitutions, where $\mathbf{\Delta y} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathcal{H}\left(\overline{\mathbf{x}}^b\right) \right] \in \mathbb{R}^{n \times 1}$ is the scaled innovation vector on the observations. Note that, by using $\overline{\mathbf{x}}^a$ and $\widehat{\mathbf{A}}$, an estimate of the posterior error distribution can be proposed as follows:

$$\mathbf{x} \sim \mathcal{N}\left(\overline{\mathbf{x}}^a, \widehat{\mathbf{A}}\right), \tag{15}$$

and therefore, the posterior ensemble (the analysis ensemble members) can be approximated by either

1.　drawing samples from the Normal distribution (15),

$$\mathbf{X}^a = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \widehat{\mathbf{A}}^{1/2} \cdot \mathbf{E} = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \mathbf{V}^a, \tag{16}$$

where $\mathbf{V}^a \in \mathbb{R}^{n \times N}$ is given by the solution of an upper triangular system of equations,

$$\widehat{\mathbf{A}}^{-1/2} \cdot \mathbf{V}^a = \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}}^{1/2} \cdot \mathbf{V}^a = \mathbf{E} \in \mathbb{R}^{n \times N},$$

and the columns of $\mathbf{E} \in \mathbb{R}^{n \times N}$ are samples from a multivariate standard Normal distribution, or

2.　using the synthetic data (9b),

$$\mathbf{X}^a = \overline{\mathbf{x}}^b \cdot \mathbf{1}_N^T + \widehat{\mathbf{A}} \cdot \mathbf{\Delta Y} = \overline{\mathbf{x}}^b \cdot \mathbf{1}_N^T + \widehat{\mathbf{V}}^a, \tag{17}$$

where $\widehat{\mathbf{V}}^a \in \mathbb{R}^{n \times N}$ is given by the solution of the next linear system of equations,

$$\widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} \cdot \widehat{\mathbf{V}}^a = \mathbf{\Delta Y}.$$

Note that, the use of $\overline{\mathbf{x}}^b$ instead of $\mathbf{x}^{b[i]}$ in Equation (17) does not change the statistical moments of the posterior ensemble, consider, for instance, the e-th column $\mathbf{x}^{a[e]}$ of matrix $\mathbf{X}^a$, for $1 \leq e \leq N$,

$$
\begin{aligned}
\mathbb{E}\left(\mathbf{x}^{a[e]}\right) &= \mathbb{E}\left(\overline{\mathbf{x}}^b + \widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) = \mathbb{E}\left(\overline{\mathbf{x}}^b\right) + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) \\
&= \mathbb{E}\left(\frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^{b[e]}\right) + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbb{E}\left(\mathbf{x}^{b[e]}\right) + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) \\
&= \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^b + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) = \mathbf{x}^b + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) = \mathbb{E}\left(\mathbf{x}^{b[e]}\right) + \mathbb{E}\left(\widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right) \\
&= \mathbb{E}\left(\mathbf{x}^{b[e]} + \widehat{\mathbf{A}} \cdot \mathbf{\Delta y}^{[e]}\right),
\end{aligned}
$$

where $\mathbf{\Delta y}^{[e]} \in \mathbb{R}^{n \times 1}$ is the e-th column of matrix $\mathbf{\Delta Y}$.

The approximation (16) is named *the posterior ensemble Kalman filter* (P-EnKF) since the analysis ensemble is built based on samples from the posterior distribution (15) while the approximation (17) is called *the synthetic posterior ensemble Kalman filter* (P-EnKF-S) since synthetic data is utilized in order to compute the analysis increments as is commonly done in stochastic EnKF formulations. Details about the computations of $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{D}}$ are discussed next.

3.2. Computing the Cholesky Factors of the Precision Analysis Covariance

By letting $\mathbf{Z} = \mathbf{H}^T \cdot \mathbf{R}^{-1/2} \in \mathbb{R}^{n \times m}$ in (13), the precision analysis covariance matrix can be written as follows,

$$\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{B}}^{-1} + \mathbf{Z} \cdot \mathbf{Z}^T = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L} + \mathbf{Z} \cdot \mathbf{Z}^T = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L} + \sum_{j=1}^{m} \cdot \mathbf{z}^{[j]} \cdot \left[\mathbf{z}^{[j]}\right]^T, \tag{18}$$

where $\mathbf{z}^{[j]} \in \mathbb{R}^{n \times 1}$ is the j-th column of matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$, for $1 \leq j \leq m$. Consider the sequence of matrices,

$$
\begin{aligned}
\widehat{\mathbf{A}}^{(0)} &= \left[\mathbf{L}^{(0)}\right]^T \cdot \mathbf{D}^{(0)} \cdot \left[\mathbf{L}^{(0)}\right] = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L} = \widehat{\mathbf{B}}^{-1}, \\
\widehat{\mathbf{A}}^{(1)} &= \widehat{\mathbf{A}}^{(0)} + \mathbf{z}^{[1]} \cdot \left[\mathbf{z}^{[1]}\right]^T = \left[\mathbf{L}^{(1)}\right]^T \cdot \mathbf{D}^{(1)} \cdot \left[\mathbf{L}^{(1)}\right], \\
\widehat{\mathbf{A}}^{(2)} &= \widehat{\mathbf{A}}^{(1)} + \mathbf{z}^{[2]} \cdot \left[\mathbf{z}^{[2]}\right]^T = \left[\mathbf{L}^{(2)}\right]^T \cdot \mathbf{D}^{(2)} \cdot \left[\mathbf{L}^{(2)}\right], \\
&\vdots \\
\widehat{\mathbf{A}}^{(m)} &= \widehat{\mathbf{A}}^{(m-1)} + \mathbf{z}^{[m]} \cdot \left[\mathbf{z}^{[m]}\right]^T = \left[\mathbf{L}^{(m)}\right]^T \cdot \mathbf{D}^{(m)} \cdot \left[\mathbf{L}^{(m)}\right] = \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} = \widehat{\mathbf{A}}^{-1},
\end{aligned}
$$

where $\mathbf{L}^{(0)} \in \mathbb{R}^{n \times n}$ and $\mathbf{D}^{(0)} \in \mathbb{R}^{n \times n}$ are the Cholesky factors of $\widehat{\mathbf{B}}^{-1}$, note that, at any intermediate step j, for $1 \leq j \leq m$, we have,

$$
\begin{aligned}
\widehat{\mathbf{A}}^{(j)} &= \left[\mathbf{L}^{(j-1)}\right]^T \cdot \mathbf{D}^{(j-1)} \cdot \left[\mathbf{L}^{(j-1)}\right] + \mathbf{z}^{[j]} \cdot \left[\mathbf{z}^{[j]}\right]^T \\
&= \left[\mathbf{L}^{(j-1)}\right]^T \cdot \left[\mathbf{D}^{(j-1)} + \mathbf{p}^{(j)} \cdot \left[\mathbf{p}^{(j)}\right]^T\right] \cdot \left[\mathbf{L}^{(j-1)}\right],
\end{aligned} \tag{19}
$$

where $\left[\mathbf{L}^{(j-1)}\right]^T \cdot \mathbf{p}^{(j)} = \mathbf{z}^{[j]} \in \mathbb{R}^{n \times 1}$. By computing the Cholesky decomposition of,

$$
\mathbf{D}^{(j-1)} + \mathbf{p}^{(j)} \cdot \left[\mathbf{p}^{(j)}\right]^T = \left[\widetilde{\mathbf{L}}^{(j-1)}\right]^T \cdot \mathbf{D}^{(j)} \cdot \left[\widetilde{\mathbf{L}}^{(j-1)}\right], \tag{20}
$$

from Equation (19), $\widehat{\mathbf{A}}^{(j)}$ can be written as follows,

$$
\widehat{\mathbf{A}}^{(j)} = \left[\widetilde{\mathbf{L}}^{(j-1)} \cdot \mathbf{L}^{(j-1)}\right]^T \cdot \mathbf{D}^{(j)} \cdot \left[\widetilde{\mathbf{L}}^{(j-1)} \cdot \mathbf{L}^{(j-1)}\right] = \left[\mathbf{L}^{(j)}\right]^T \cdot \mathbf{D}^{(j)} \cdot \left[\mathbf{L}^{(j)}\right],
$$

where $\mathbf{L}^{(j)} = \widetilde{\mathbf{L}}^{(j-1)} \cdot \mathbf{L}^{(j-1)} \in \mathbb{R}^{n \times n}$. In Equation (20), the components of the factors $\widetilde{\mathbf{L}}^{(j-1)}$ and $\widetilde{\mathbf{D}}^{(j)}$ can be easily computed from the elements of $\mathbf{D}^{(j-1)}$ and $\mathbf{p}^{(j)}$ based on the Dolittle's method for matrix factorization, it is enough to note that,

$$
\left\{\left[\widetilde{\mathbf{L}}^{(j-1)}\right]^T \cdot \mathbf{D}^{(j)} \cdot \left[\widetilde{\mathbf{L}}^{(j-1)}\right]\right\}_{i,k} = \delta_{i,k} \cdot \left\{\mathbf{D}^{(j-1)}\right\}_{i,i} + \left\{\mathbf{p}^{(j)}\right\}_i \cdot \left\{\mathbf{p}^{(j)}\right\}_k,
$$

from which, the next equations are then obtained,

$$
\left\{\mathbf{D}^{(j)}\right\}_{n,n} = \left[\left\{\mathbf{p}^{(j)}\right\}_n\right]^2 + \left\{\mathbf{D}^{(j-1)}\right\}_{n,n}, \tag{21a}
$$

$$
\left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{i,k} = \frac{1}{\left\{\mathbf{D}^{(j)}\right\}_{i,i}} \cdot \left[\left\{\mathbf{p}^{(j)}\right\}_i \cdot \left\{\mathbf{p}^{(j)}\right\}_k - \sum_{q \in P(i,r)} \left\{\mathbf{D}^{(j)}\right\}_{q,q} \cdot \left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{q,i} \cdot \left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{q,k}\right], \tag{21b}
$$

and

$$
\left\{\mathbf{D}^{(j)}\right\}_{i,i} = \left[\left\{\mathbf{p}^{(j)}\right\}_i\right]^2 + \left\{\mathbf{D}^{(j-1)}\right\}_{i,i} - \sum_{q \in P(i,r)} \left\{\mathbf{D}^{(j)}\right\}_{q,q} \cdot \left[\left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{q,i}\right]^2, \tag{21c}
$$

for $n - 1 \geq i \geq 1$ and $k \in P(i, r)$, where $\delta_{i,j}$ is the Kronecker delta function, and the diagonal elements of $\widetilde{\mathbf{L}}^{(j-1)}$ are all ones. In Algorithm 1 the updating process for a column $\mathbf{z}^{[j]}$ is detailed while the

computations of factors $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{D}}$ for $\widehat{\mathbf{A}}^{-1}$ are shown in Algorithm 2. We also report an estimate of the number of multiplications performed at each step of the proposed algorithms. φ denotes the maximum number of predecessors across all model components. Typically, it will be a function of the radius of influence r with $\varphi \ll n$. Note that, since the index k is constrained to the predecessors of component i, the structure of $\mathbf{L}^{(j-1)}$ is replicated in $\widetilde{\mathbf{L}}^{(j-1)}$ and therefore, the structure of $\mathbf{L}^{(j-1)}$ is preserved in $\mathbf{L}^{(j)}$. Consequently, the sparsity pattern of $\widehat{\mathbf{L}}$ equals that of $\mathbf{L}$.

Algorithm 1 Rank-one update for the factors $\mathbf{L}^{(j-1)}$ and $\mathbf{D}^{(j-1)}$.

1: **function** UPD_CHOLESKY_FACTORS($\mathbf{L}^{(j-1)}, \mathbf{D}^{(j-1)}, \mathbf{z}^{[j]}$) ▷ **COST**

2:     Compute $\mathbf{p}^{(j)}$ from $\left[\mathbf{L}^{(j-1)}\right]^T \cdot \mathbf{p}^{(j)} = \mathbf{z}^{[j]}$. ▷ $\mathcal{O}(\varphi \cdot n)$

3:     Compute $\left\{\mathbf{D}^{(j)}\right\}_{n,n}$ via Equation (21a). ▷ $\mathcal{O}(1)$

4:     **for** $i = n-1 \to 1$ **do** ▷ $n-1$ times lines 5–9, $\mathcal{O}(\varphi^2 \cdot n)$

5:        Set $\left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{i,i} \leftarrow 1$. ▷ $\mathcal{O}(1)$

6:        **for** $k \in P(i, r)$ **do** ▷ φ times line 7, $\mathcal{O}(\varphi^2)$

7:           Compute $\left\{\widetilde{\mathbf{L}}^{(j-1)}\right\}_{i,k}$ according to (21b). ▷ $\mathcal{O}(\varphi)$

8:        **end for**

9:        Compute $\left\{\mathbf{D}^{(j)}\right\}_{i,i}$ via Equation (21c). ▷ $\mathcal{O}(\varphi)$

10:     **end for**

11:     Set $\mathbf{L}^{(j)} \leftarrow \widetilde{\mathbf{L}}^{(j-1)} \cdot \mathbf{L}^{(j-1)}$. ▷ $\mathcal{O}(\varphi^2 \cdot n)$

12:     **return** $\mathbf{L}^{(j)}, \mathbf{D}^{(j)}$

13: **end function**

Algorithm 2 Computing the factors $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{D}}$ of $\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}}$.

1: **function** COMPUTE_ANALYSIS_FACTORS($\mathbf{L}^{(0)}, \mathbf{D}^{(0)}, \mathbf{H}, \mathbf{R}$) ▷ **COST**

2:     Set $\mathbf{Z} \leftarrow \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$. ▷ $\mathcal{O}(m \cdot n)$

3:     **for** $j = 1 \to m$ **do** ▷ m times line 4, $\mathcal{O}(\varphi^2 \cdot m \cdot n)$

4:        Set $[\mathbf{L}^{(j)}, \mathbf{D}^{(j)}] \leftarrow$ UPD_CHOLESKY_FACTORS $\left(\mathbf{L}^{(j-1)}, \mathbf{D}^{(j-1)}, \mathbf{z}^{[j]}\right)$ ▷ $\mathcal{O}(\varphi^2 \cdot n)$

5:     **end for**

6:     **return** $\mathbf{L}^{(m)}$ as $\widehat{\mathbf{L}}$, $\mathbf{D}^{(m)}$ as $\widehat{\mathbf{D}}$.

7: **end function**

3.3. Computational Cost of the Analysis Step

Once the Cholesky factors of the matrix (13) are estimated, the posterior state (14) can be computed and the analysis ensemble (16) can be built. Putting it all together, the assimilation step of the P-EnKF is detailed in Algorithm 3. Based on the Algorithms 1 and 2, it is clear that the computational effort of the P-EnKF is,

$$\mathcal{O}\left(\varphi^2 \cdot m \cdot n + \varphi^2 \cdot n\right),$$

which is linear with regard to the number of model components n and the number of observed components from the model domain m. This computational effort obeys, mainly, to the special structure of the Cholesky factors $\mathbf{L}^{(j)}$ and $\widetilde{\mathbf{L}}^{(j)}$, for $1 \leq j \leq m$, and how these structures can be exploited in practice in order to reduce the overall computational effort of the analysis step, for instance, the backward substitution in line 2 of Algorithm 1 can be efficiently performed as follows,

$$\left\{ \mathbf{p}^{(j)} \right\}_{n-i+1} = \left\{ \mathbf{z}^{[j]} \right\}_{n-i+1} - \sum_{k \in P(i,r)} \left\{ \mathbf{p}^{(j)} \right\}_k \cdot \left\{ \left[\mathbf{L}^{(j-1)} \right] \right\}_{i,k} , \text{ for } 1 \leq i \leq n \,,$$

and therefore, the number of multiplications in this computation is bounded by $\mathcal{O}\left(\varphi \cdot n \right)$ while the matrix multiplication in line 11 can be performed as follows,

$$\left\{ \mathbf{L}^{(j)} \right\}_{i,k} = \sum_{q \in P(i,r) \cap P(k,r)} \left\{ \widetilde{\mathbf{L}}^{(j-1)} \right\}_{i,q} \cdot \left\{ \mathbf{L}^{(j-1)} \right\}_{q,k} , \text{ for } 1 \leq i \leq n, \text{ and } k \in P(i,r) \,,$$

where, evidently, the maximum number of elements in $P(i,r) \cap P(k,r)$ is bounded by φ and therefore, the total number of multiplications is bounded by $\mathcal{O}\left(\varphi^2 \cdot n \right)$. Notice, since the matrices $\mathbf{L}^{(j)}$ and $\widetilde{\mathbf{L}}^{(j)}$ are sparse and lower triangular, their non-zero bands can be represented by one-dimensional arrays. In this manner, for instance, the computations derived for the elements of $\mathbf{L}^{(j)}$ and $\widetilde{\mathbf{L}}^{(j)}$ can be performed on the elements of such vectors. Efficient matrix storage schemes are now proposed by the literature [29,30] as well as scientific computational languages which exploit such structures in order to speed-up matrix computations and to save memory usage [31]. From here, matrix-free implementations of the P-EnKF can be easily derived in order to make it practical under operational data assimilation settings.

Algorithm 3 Assimilation of observations via the posterior ensemble Kalman filter (16).

1: **function** ANALYSIS_P-ENKF($\mathbf{X}^b$, **L**, **D** , **y**, **H**, **R**)　　　　　　　▷ Having, $\widehat{\mathbf{B}}^{-1} = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L}$.

2:　　　Set $[\widehat{\mathbf{L}}, \widehat{\mathbf{D}}] \leftarrow$ COMPUTE_ANALYSIS_FACTORS (**L**, **D**, **H**, **R**).

3:　　　Compute $\overline{\mathbf{x}}^a$ according to Equation (14).

4:　　　Compute the posterior ensemble $\mathbf{X}^a$ based on Equation (16).

5:　　　**return** $\mathbf{X}^a$.

6: **end function**

Readily, the computational effort of the P-EnKF-S is similar to that of the P-EnKF. We detail the assimilation step of the P-EnKF-S in Algorithm 4.

Algorithm 4 Assimilation of observations via the posterior ensemble Kalman filter (17) .

1: **function** ANALYSIS_P-ENKF($\mathbf{X}^b$, **L**, **D** , **y**, **H**, **R**)　　　　　　　▷ Having, $\widehat{\mathbf{B}}^{-1} = \mathbf{L}^T \cdot \mathbf{D} \cdot \mathbf{L}$.

2:　　　Set $[\widehat{\mathbf{L}}, \widehat{\mathbf{D}}] \leftarrow$ COMPUTE_ANALYSIS_FACTORS (**L**, **D**, **H**, **R**).

3:　　　Compute the posterior ensemble $\mathbf{X}^a$ based on Equation (17).

4:　　　**return** $\mathbf{X}^a$.

5: **end function**

3.4. Inflation Aspects

While localization methods reduce the impact of spurious correlations, covariance inflation mitigates the impact of under-estimation of sample variances [32–35]. Typically, ensemble members

are inflated prior the forecast step in order to enrich the background error information for the next assimilation cycle and to reduce the odds of ensemble collapsing. For instance, after the assimilation step of EnKF, the ensemble members (9a) are inflated by a factor of $\rho > 1$ about the analysis mean,

$$\mathbf{X}^{a,\rho} = \overline{\mathbf{x}}^a \cdot \mathbf{1}^T + \rho \cdot \mathbf{\Delta X}^a \, ,$$

where $\mathbf{\Delta X}^a = \mathbf{X}^a - \overline{\mathbf{x}}^a \cdot \mathbf{1}^T \in \mathbb{R}^{n \times N}$ are the innovations about the analysis mean. Thus, the (co) variances in $\mathbf{P}^a$ are inflated by a factor of ρ^2. This idea can be incorporated in the P-EnKF by noting that, the estimated precision analysis covariance matrix can be inflated as follows,

$$\rho^2 \cdot \widehat{\mathbf{A}} \Leftrightarrow \frac{1}{\rho^2} \cdot \widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{L}}^T \cdot \left[\frac{1}{\rho^2} \cdot \widehat{\mathbf{D}} \right] \cdot \widehat{\mathbf{L}} \, ,$$

and therefore, the inflation can be performed before the analysis members are drawn from the distribution (15),

$$\mathbf{X}^{a,\rho} = \overline{\mathbf{x}}^a \cdot \mathbf{1}^T + \widehat{\mathbf{Z}} \, ,$$

where $\widehat{\mathbf{Z}} \in \mathbb{R}^{n \times N}$ is given by the solution of the sparse upper triangular system of equations,

$$\widehat{\mathbf{L}}^T \cdot \left[\frac{1}{\rho} \cdot \widehat{\mathbf{D}}^{1/2} \right] \cdot \widehat{\mathbf{Z}} = \mathbf{E} \, .$$

Recall that, the columns of matrix $\mathbf{E}$ are samples from a multivariate standard Normal distribution. Note that, the use of covariance inflation does not increase the computational effort of the P-EnKF. Similarly, inflation can be applied to the assimilation step in the P-EnKF-S formulation.

3.5. Main Differences Between the EnKF-MC, the P-EnKF, and the P-EnKF-S

In essence, the EnKF-MC, the P-EnKF, and the P-EnKF-S are stochastic filters based on a modified Cholesky decomposition [11]. Nevertheless, the manner how they operate is quite different. Consider again the analysis covariance matrix (13),

$$\widehat{\mathbf{A}} = \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \, ,$$

technically speaking, in the context of the EnKF-MC, the posterior ensemble is built by solving the linear system of equations,

$$\widehat{\mathbf{A}} \cdot \widehat{\mathbf{Z}} = \mathbf{\Delta Y}, \text{ with } \mathbf{\Delta Y} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} \cdot \mathbf{1}^T + \mathbf{R}^{1/2} \cdot \mathbf{E} - \mathbf{H} \cdot \mathbf{X}^b \right] \, ,$$

where the columns of $\mathbf{E} \in \mathbb{R}^{n \times N}$ follows a standard Normal distribution. $\widehat{\mathbf{Z}}$ can be computed by using the iterative Sherman Morrison formula [36] in order to avoid the direct inversion of $\widehat{\mathbf{A}}$ while the trivial linear system,

$$\widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}}^{1/2} \cdot \mathbf{V}^a = \mathbf{E} \, ,$$

is the one to solve in order to compute the analysis increments for the background members in the P-EnKF formulation. Likewise, the linear system to solve in the P-EnKF-S case depends on the Cholesky factors and the synthetic data (9b),

$$\widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} \cdot \widehat{\mathbf{V}^a} = \mathbf{\Delta Y} \, ,$$

whose solution can be obtained by using backward and forward substitutions. Yet another important difference is that, in the P-EnKF, the computation of the posterior mode is similar to that of square root filter formulations,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \Delta \mathbf{x}^a \text{, with } \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{L}} \cdot \Delta \mathbf{x}^a = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathcal{H}\left(\overline{\mathbf{x}}^b \right) \right] ,$$

while the analysis mean,

$$\overline{\mathbf{x}}^a = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^{a[e]} \in \mathbb{R}^{n \times 1} ,$$

provides the estimator of $\mathbf{x}^a$ in the EnKF-MC and the P-EnKF-S methods.

4. Experimental Results

In this section, we tests the proposed EnKF implementations and compare our results with those obtained by the EnKF-MC and the LETKF implementations discussed in Section 2. We make use of the Lorenz-96 model [37] as our surrogate model during the experiments. The Lorenz-96 model is described by the following set of ordinary differential equations [38]:

$$\frac{dx_j}{dt} = \begin{cases} (x_2 - x_{n-1}) \cdot x_n - x_1 + F & \text{for } j = 1, \\ (x_{j+1} - x_{j-2}) \cdot x_{j-1} - x_j + F & \text{for } 2 \leq j \leq n-1, \\ (x_1 - x_{n-2}) \cdot x_{n-1} - x_n + F & \text{for } j = n, \end{cases} \tag{22}$$

where F is external force and $n = 40$ is the number of model components. Periodic boundary conditions are assumed. When $F = 8$ units the model exhibits chaotic behavior, which makes it a relevant surrogate problem for atmospheric dynamics [39,40]. A time unit in the Lorenz-96 represents 7 days in the atmosphere. The experimental settings are described below:

- An initial random solution is integrated over a long time period in order to obtain an initial condition $\mathbf{x}^*_{-2} \in \mathbb{R}^{n \times 1}$ dynamically consistent with the model (22).
- A perturbed background solution $\tilde{\mathbf{x}}^b_{-2}$ is obtained at time t_{-2} by drawing a sample from the Normal distribution,

$$\tilde{\mathbf{x}}^b_{-2} \sim \mathcal{N}\left(\mathbf{x}^*_{-2}, 0.05^2 \cdot \mathbf{I} \right) ,$$

  this solution is then integrated for 10 time units (equivalent to 70 days in the atmosphere) in order to obtain a background solution $\mathbf{x}^b_{-1}$ consistent with the numerical model.
- An initial perturbed ensemble is built about the background state by taking samples from the distribution,

$$\tilde{\mathbf{x}}^{b[\hat{e}]}_{-1} \sim \mathcal{N}\left(\mathbf{x}^b_{-1}, 0.05^2 \cdot \mathbf{I} \right) \text{, for } 1 \leq \hat{e} \leq \widehat{N} ,$$

  and in order to make them consistent with the model dynamics, the ensemble members are propagated for 10 time units, from which the initial ensemble members $\mathbf{x}^{b[\hat{e}]}_0 \in \mathbb{R}^{n \times 1}$ are obtained. We create the initial pool $\widehat{\mathbf{X}^b}_0$ of $\widehat{N} = 10^6$ members.
- The assimilation window consists of $M = 25$ observations. These are taken every 3.5 days and their error statistics are associated with the Gaussian distribution,

$$\mathbf{y}_k \sim \mathcal{N}\left(\mathbf{H}_k \cdot \mathbf{x}^*_k, 0.01^2 \cdot \mathbf{I} \right) \text{, for } 1 \leq k \leq M .$$

where the number of observed components from the model space is $m = 30$. The components are randomly chosen at the different assimilation steps. Thus, $\mathbf{H}_k$ is randomly formed at the different assimilation cycles.

- Values of the inflation factor ρ are ranged in $1 \leq \rho \leq 1.1$.
- We try different values for the radii of influence r, these range in $1 \leq r \leq 7$.
- The ensemble size for the benchmarks is $N = 20$. These members are randomly chosen from the pool $\widehat{\mathbf{X}^b}_0$ for the different pairs (r, ρ) in order to form the initial ensemble $\mathbf{X}_0^b$ for the assimilation window. Evidently, $\mathbf{X}_0^b \subset \widehat{\mathbf{X}^b}_0$.
- The assimilation steps are also performed by the EnKF with full size of $\widehat{\mathbf{X}^b}_0$ in order to obtain a reference solution regarding what to expect from the EnKF formulations. Note that, the ensemble size is large enough in order to dissipate the impact of sampling errors. No inflation is needed as well.
- The L-2 norm of the error is utilized as a measure of accuracy at the assimilation step p,

$$\lambda_p = \sqrt{\left[\mathbf{x}_p^* - \mathbf{x}_p^a\right]^T \cdot \left[\mathbf{x}_p^* - \mathbf{x}_p^a\right]}, \tag{23}$$

where $\mathbf{x}_p^*$ and $\mathbf{x}_p^a$ are the reference and the analysis solutions, respectively.

- The Root-Mean-Square-Error (RMSE) is utilized as a measure of performance, in average, on a given assimilation window,

$$\epsilon = \sqrt{\frac{1}{M} \cdot \sum_{p=1}^{M} \lambda_p^2}.$$

In Figure 3, the results are shown in terms of $L - 2$ error norms and assimilation steps (p) for the compared methods, we group the results by radius sizes. As can be seen, there are parameter values for r and ρ in which the results proposed by the filters are comparable with those obtained by the EnKF with a large ensemble size of $N = 10^6$. This supports the theory that, all filters can provide good approximations of posterior states with just a few ensemble members. However, the performance of the LETKF degrades as the radius size is increased, this can be explained as follows: in the LETKF, local background error correlations are estimated based on a local ensemble covariance matrix, therefore, sampling noise can easily impact the sample covariance matrix when the ensemble size is not much larger than the number of components within the local boxes. Such cases are not evident in the EnKF implementations based on a modified Cholesky decomposition. All these filters perform remarkably well and in all cases, the initial error (uncertainty) is decreased. The importance of having filters which are not so sensible to the parameters r and ρ is that, in practice, such parameters can be hard to tune. Thus, one can prefer filter formulations whose accuracy is not highly impacted by changes in those parameters.

We can analyze the results in terms of RMSE values as well. Those are reported in Figure 4. Note that, as we mentioned before, the accuracy of the LETKF can be sensible to the radius size regardless the inflation factor. After $r = 3$, the quality of the analysis obtained by this filter is clearly affected by sampling errors. On the other hand, the proposed filters behave similarly to the EnKF-MC since their background error correlations are estimated making use of the same estimator (12a).
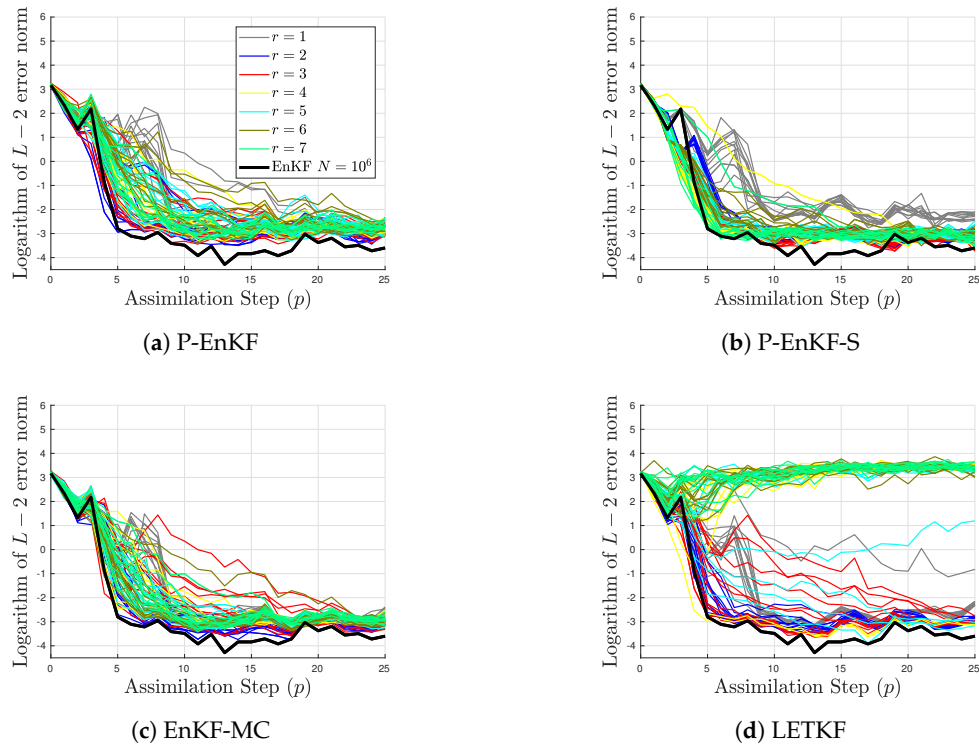
Figure 3. Experimental results with the Lorenz-96 model (22). The results are grouped by values of *r*. inflation factors are ranged in $1 \leq \rho \leq 1.1$ for each group.
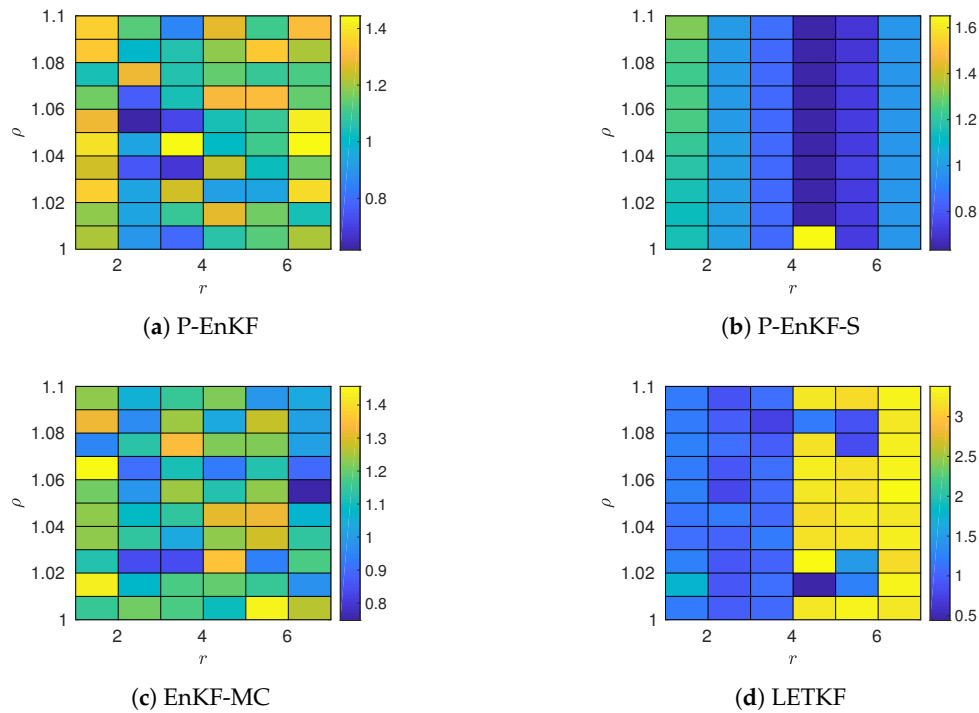

Figure 4. Experimental results with the Lorenz-96 model (22). The The Root-Mean-Square-Error (RMSE) values are shown for the compared filter implementations for different values of *r* and *ρ*.

5. Conclusions

This paper proposes a posterior ensemble Kalman filter based on a modified Cholesky decomposition which works as follows: the precision matrix of the background error distribution is estimated in terms of Cholesky factors via a modified Cholesky decomposition, these factors are updated making use of a series of rank-one updates in order to estimate a precision matrix of the posterior distribution. By using this matrix, the posterior mode of the error distribution can be estimated. The posterior ensemble can be built by either sampling from the posterior distribution or making use of synthetic data. Experimental tests are performed in order to assess the accuracy of the proposed method as the inflation factor and the localization radius are varied. The numerical model operator utilized during the experiments is the Lorenz-96 model. The results reveal that the accuracy in terms of root-mean-square-error of the proposed methods are similar to that of the local ensemble transform Kalman filter and even more, our results are comparable to those obtained by the EnKF with large ensemble sizes.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Lorenc, A.C. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **1986**, *112*, 1177–1194.
2. Buehner, M.; Houtekamer, P.; Charette, C.; Mitchell, H.L.; He, B. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Weather Rev.* **2010**, *138*, 1550–1566.
3. Buehner, M.; Houtekamer, P.; Charette, C.; Mitchell, H.L.; He, B. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Weather Rev.* **2010**, *138*, 1567–1586.
4. Caya, A.; Sun, J.; Snyder, C. A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Weather Rev.* **2005**, *133*, 3081–3094.
5. Nino-Ruiz, E.D.; Sandu, A. Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation. *Ocean Dyn.* **2015**, *65*, 1423–1439.
6. Anderson, J.L. Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Mon. Weather Rev.* **2012**, *140*, 2359–2371.
7. Jonathan, P.; Fuqing, Z.; Weng, Y. The Effects of Sampling Errors on the EnKF Assimilation of Inner-Core Hurricane Observations. *Mon. Weather Rev.* **2014**, *142*, 1609–1630.
8. Buehner, M. Ensemble-derived Stationary and Flow-dependent Background-error Covariances: Evaluation in a Quasi-operational NWP Setting. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 1013–1043.
9. Nino-Ruiz, E.D.; Sandu, A.; Deng, X. A parallel ensemble Kalman filter implementation based on modified Cholesky decomposition. In Proceedings of the 6th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems, Austin, TX, USA, 15–20 November 2015; p. 4.
10. Nino-Ruiz, E.D.; Sandu, A.; Deng, X. A parallel implementation of the ensemble Kalman filter based on modified Cholesky decomposition. *J. Comput. Sci.* **2017**, https://doi.org/10.1016/j.jocs.2017.04.005.
11. Bickel, P.J.; Levina, E. Regularized estimation of large covariance matrices. *Ann. Stat.* **2008**, *36*, 199–227.
12. Evensen, G. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dyn.* **2003**, *53*, 343–367.
13. Lorenc, A.C. The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3183–3203.
14. Hamill, T.M.; Whitaker, J.S.; Snyder, C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **2001**, *129*, 2776–2790.
15. Cheng, H.; Jardak, M.; Alexe, M.; Sandu, A. A Hybrid Approach to Estimating Error Covariances in Variational Data Assimilation. *Tellus A* **2010**, *62*, 288–297.

16. Chatterjee, A.; Engelen, R.J.; Kawa, S.R.; Sweeney, C.; Michalak, A.M. Background Error Covariance Estimation for Atmospheric CO_2 Data Assimilation. *J. Geophys. Res. Atmos.* **2013**, *118*, doi:10.1002/jgrd.50654.

17. Keppenne, C.L. Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. *Mon. Weather Rev.* **2000**, *128*, 1971–1981.

18. Buehner, M. Evaluation of a Spatial/Spectral Covariance Localization Approach for Atmospheric Data Assimilation. *Mon. Weather Rev.* **2011**, *140*, 617–636.

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

20. Tippett, M.K.; Anderson, J.L.; Bishop, C.H.; Hamill, T.M.; Whitaker, J.S. Ensemble Square Root Filters. *Mon. Weather Rev.* **2003**, *131*, 1485–1490.

21. Bishop, C.H.; Toth, Z. Ensemble Transformation and Adaptive Observations. *J. Atmos. Sci.* **1999**, *56*, 1748–1765.

22. Ott, E.; Hunt, B.R.; Szunyogh, I.; Zimin, A.V.; Kostelich, E.J.; Corazza, M.; Kalnay, E.; Patil, D.J.; Yorke, J.A. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* **2004**, *56*, 415–428.

23. Hunt, B.R.; Kostelich, E.J.; Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Phys. D: Nonlinear Phenom.* **2007**, *230*, 112–126.

24. Ott, E.; Hunt, B.; Szunyogh, I.; Zimin, A.V.; Kostelich, E.J.; Corazza, M.; Kalnay, E.; Patil, D.J.; Yorke, J.A. A Local Ensemble Transform Kalman Filter Data Assimilation System for the NCEP Global Model. *Tellus A* **2008**, *60*, 113–130.

25. Nino, E.D.; Sandu, A.; Deng, X. An Ensemble Kalman Filter Implementation Based on Modified Cholesky Decomposition for Inverse Covariance Matrix Estimation. *arXiv* **2016**, arXiv:1605.08875.

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

27. Evensen, G. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst.* **2009**, *29*, doi:10.1109/MCS.2009.932223.

28. Anderson, J.L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **2001**, *129*, 2884–2903.

29. Smailbegovic, F.; Gaydadjiev, G.N.; Vassiliadis, S. Sparse matrix storage format. In Proceedings of the 16th Annual Workshop on Circuits, Systems and Signal Processing, Veldhoven, The Netherlands, 17–18 November 2005; pp. 445–448.

30. Langr, D.; Tvrdik, P. Evaluation criteria for sparse matrix storage formats. *IEEE Trans. Parallel Distrib. Syst.* **2016**, *27*, 428–440.

31. Dongarra, J.J.; Demmel, J.W.; Ostrouchov, S. LAPACK: A Linear Algebra Library for High-Performance Computers. In *Computational Statistics: Volume 1, Proceedings of the 10th Symposium on Computational Statistics*; Springer Science & Business Media: Berlin, Germany, 2013; p. 23.

32. Westgate, P.M. A covariance correction that accounts for correlation estimation to improve finite-sample inference with generalized estimating equations: A study on its applicability with structured correlation matrices. *J. Stat. Comput. Simul.* **2016**, *86*, 1891–1900.

33. Lei, L.; Whitaker, J.S. Evaluating the tradeoffs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *J. Adv. Model. Earth Syst.* **2017**, doi:10.1002/2016MS000864.

34. Lee, Y.; Majda, A.J.; Qi, D. Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence. *Mon. Weather Rev.* **2017**, *145*, 669–682.

35. Putnam, B.J.; Xue, M.; Jung, Y.; Snook, N.A.; Zhang, G. Ensemble Probabilistic Prediction of a Mesoscale Convective System and Associated Polarimetric Radar Variables using Single-Moment and Double-Moment Microphysics Schemes and EnKF Radar Data Assimilation. *Mon. Weather Rev.* **2017**, doi:10.1175/MWR-D-16-0162.1.

36. Ruiz, E.D.N.; Sandu, A.; Anderson, J. An efficient implementation of the ensemble Kalman filter based on an iterative Sherman—Morrison formula. *Stat. Comput.* **2015**, *25*, 561–577.

37. Lorenz, E.N. Designing Chaotic Models. *J. Atmos. Sci.* **2005**, *62*, 1574–1587.

38. Fertig, E.J.; Harlim, J.; Hunt, B.R. A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus A* **2007**, *59*, 96–100.

39. Karimi, A.; Paul, M.R. Extensive chaos in the Lorenz-96 model. *Chaos: An Interdiscip. J. Nonlinear Sci.* **2010**, *20*, 043105.

40. Gottwald, G.A.; Melbourne, I. Testing for chaos in deterministic systems with noise. *Phys. D: Nonlinear Phenom.* **2005**, *212*, 100–110.