AN ENSEMBLE KALMAN FILTER IMPLEMENTATION BASED ON MODIFIED CHOLESKY DECOMPOSITION FOR INVERSE COVARIANCE MATRIX ESTIMATION

ELIAS D. NINO-RUIZ†, ADRIAN SANDU‡, AND XINWEI DENG§

Abstract. This paper develops an efficient implementation of the ensemble Kalman filter based on a modified Cholesky decomposition for inverse covariance matrix estimation. This implementation is named EnKF-MC. Background errors corresponding to distant model components with respect to some radius of influence are assumed to be conditionally independent. This allows to obtain sparse estimators of the inverse background error covariance matrix. The computational effort of the proposed method is discussed and different formulations based on various matrix identities are provided. Furthermore, an asymptotic proof of convergence with regard to the ensemble size is presented. In order to assess the performance and the accuracy of the proposed method, experiments are performed making use of the Atmospheric General Circulation Model SPEEDY. The results are compared against those obtained using the local ensemble transform Kalman filter (LETKF). Tests are performed for dense observations (100% and 50% of the model components are observed) as well as for sparse observations (only 12%, 6%, and 4% of model components are observed). The results reveal that the use of modified Cholesky for inverse covariance matrix estimation can reduce the impact of spurious correlations during the assimilation cycle, i.e., the results of the proposed method are of better quality than those obtained via the LETKF in terms of root mean square error.

Key words. Modified Cholesky decomposition, background error covariance estimation, spurious correlations, ensemble Kalman filter

AMS subject classifications. 62L20, 62L12, 65C05, 65C60

1. Introduction. The goal of sequential data assimilation is to estimate the true state of a dynamical system \( x^{\text{true}} \in \mathbb{R}^{n \times 1} \) using information from numerical models, priors, and observations. A numerical model captures (with some approximation) the physical laws of the system and evolves its state forward in time [7]:

\[
    x_{k} = M_{t_{k-1} \rightarrow t_{k}} (x_{k-1}) \in \mathbb{R}^{n \times 1}, \quad \text{for } x \in \mathbb{R}^{n \times 1},
\]

where \( n \) is the dimension of the model state, \( k \) denotes time index, and \( M \) can represent, for example, the dynamics of the ocean and/or atmosphere. A prior estimation \( x_{k}^{b} \in \mathbb{R}^{n \times 1} \) of \( x_{k}^{\text{true}} \) is available:

\[
    x_{k}^{b} - x^{\text{true}} = \nu_{k} \sim \mathcal{N}(0, B_{k}) \in \mathbb{R}^{n \times 1},
\]

where the prior error \( \nu_{k} \) is assumed to have mean zero and a covariance matrix \( B_{k} \in \mathbb{R}^{n \times n} \). Noisy observations (measurements) of the true state \( y_{k} \in \mathbb{R}^{m \times 1} \) are taken, and the observation errors \( \epsilon \) are usually assumed to be normally distributed:

\[
    y_{k} - \mathcal{H}(x_{k}^{\text{true}}) = \epsilon_{k} \sim \mathcal{N}(0, R_{k}) \in \mathbb{R}^{m \times 1},
\]

\*Submitted to the editors DATE.

Funding: This work was supported in part by awards NSF CCF–1218454, AFOSR FA9550–12–1–0293–DEF, by the Computational Science Laboratory at Virginia Tech, and by the Applied Math and Computer Science Laboratory at Universidad del Norte.

†Department of Computer Science, Universidad del Norte, BAQ, Colombia (enino@uninorte.edu.co, https://sites.google.com/a/vt.edu/eliasnino/).

‡Computational Science Laboratory, Department of Computer Science, Virginia Tech, Blacksburg, VA 24060, USA (asandu7@vt.edu, http://people.cs.vt.edu/~asandu/).

§Department of Statistics, Virginia Tech, Blacksburg, VA 24060, USA (xdeng@vt.edu, http://www.apps.stat.vt.edu/deng/).
where \( m \) is the number of observed components, \( \mathcal{H} : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1} \) is the linear observation operator, and \( \mathbf{R}_k \in \mathbb{R}^{m \times m} \) is the data error covariance matrix.

Making use of Bayesian statistics and matrix identities, the assimilation of the observation (3) is performed as follows:

\[
\begin{align*}
\mathbf{x}^b_k & = \mathbf{x}^b_k + \mathbf{B}_k \cdot \mathbf{H}_k^T \cdot [\mathbf{H}_k \cdot \mathbf{B}_k \cdot \mathbf{H}_k^T + \mathbf{R}_k]^{-1} \cdot [\mathbf{y}_k - \mathcal{H} (\mathbf{x}^b_k)] \in \mathbb{R}^{n \times 1}, \\
\mathbf{A}_k & = [\mathbf{I} - \mathbf{B}_k \cdot \mathbf{H}_k^T \cdot [\mathbf{R}_k + \mathbf{H}_k \cdot \mathbf{B}_k \cdot \mathbf{H}_k^T]^{-1} \cdot \mathbf{H}_k] \cdot \mathbf{B}_k \in \mathbb{R}^{n \times n},
\end{align*}
\]

where \( \mathbf{H}_k \in \mathbb{R}^{m \times n} \) is a linear observation operator, \( \mathbf{A}_k \) is the analysis (posterior) covariance matrix, and \( \mathbf{x}^b_k \in \mathbb{R}^{n \times 1} \) is the analysis state. Typically, in the context of Kalman filtering, the observational operator is assumed linear [21]. Nonlinear observation operators are treated either by linearization, \( \mathbf{H}_k \approx \mathcal{H}'(\mathbf{x}^b_k) \), or by an ensemble approximation. For simplicity of presentation we will silently consider here a linear observation operator; this does not restrict the generality of the discussion below since the standard treatment of nonlinear operators can be directly applied.

According to equation (4) the elements of \( \mathbf{B}_k \) determine how the information about the observed model components contained in the innovations \( \mathbf{y}_k - \mathcal{H}(\mathbf{x}^b_k) \in \mathbb{R}^{m \times 1} \) is distributed to properly adjust all model components, including the unobserved ones. Thus, the successful assimilation of the observation (3) will rely, in part, on how well the background error statistics are approximated.

In the context of ensemble based methods, an ensemble of model realizations

\[
(5) \quad \mathbf{X}^b_k = \left[ \mathbf{x}^{b[1]}_k, \mathbf{x}^{b[2]}_k, \ldots, \mathbf{x}^{b[N_{ens}]}_k \right] \in \mathbb{R}^{n \times N_{ens}},
\]

is used in order to estimate the unknown moments of the background error distribution:

\[
(6a) \quad \bar{x}^b_k = \frac{1}{N_{ens}} \cdot \sum_{i=1}^{N_{ens}} \mathbf{x}^{b[i]}_k \in \mathbb{R}^{n \times 1}, \quad \mathbf{B}_k \approx \mathbf{P}^b = \frac{1}{N_{ens} - 1} \cdot \mathbf{U}^b_k \cdot (\mathbf{U}^b_k)^T \in \mathbb{R}^{n \times n},
\]

where \( N_{ens} \) is the number of ensemble members, \( \mathbf{x}^{b[i]}_k \in \mathbb{R}^{n \times 1} \) is the \( i \)-th ensemble member, \( \bar{x}^b_k \in \mathbb{R}^{n \times 1} \) is the background ensemble mean, \( \mathbf{P}^b \) is the background ensemble covariance matrix, and \( \mathbf{U}^b_k \in \mathbb{R}^{n \times N_{ens}} \) is the matrix of member deviations:

\[
(6b) \quad \mathbf{U}^b_k = \mathbf{X}^b_k - \bar{x}^b_k \cdot 1_{N_{ens}}^T \in \mathbb{R}^{n \times N_{ens}}.
\]

One attractive feature of \( \mathbf{P}^b_k \) is its flow-dependency which allows to approximate the background error correlations based on the dynamics of the numerical model (1).

However, in operational data assimilation, the number of model components is much larger than the number of model realizations \( n \gg N_{ens} \) and therefore \( \mathbf{P}^b_k \) is rank-deficient. Spurious correlations (e.g., correlations between distant model components in space) can degenerate the quality of the analysis corrections. One of the most successful EnKF formulations is the local ensemble transform Kalman filter (LETKF) in which the impact of spurious analysis correlations is avoided by making use of local domain analyses. In this context, every model component is surrounded by a box of a prescribed radius, and then the assimilation is performed within every local box. In this case the background error correlations are provided by the local ensemble covariance matrix. The local analyses are mapped back onto the global domain to obtain the global analysis state. Nevertheless, when sparse observational networks are
considered many boxes can contain no observations, in which case the local analyses coincide with the background. The local box sizes can be increased in order to include observations within the local domains, in which case local analysis corrections can be impacted by spurious correlations. Moreover, in practice, the size of local boxes can be still larger than the number of ensemble members and therefore, the local sample covariance matrix can be rank-deficient.

In order to address the above issues this paper proposes a better estimation of the inverse background error covariance matrix $B^{-1}$ obtained via a modified Cholesky decomposition. By imposing conditional independence between errors in remote model components we obtain sparse approximations of $B^{-1}$.

This paper is organized as follows. In Section 2 ensemble based methods and the modified Cholesky decomposition are introduced. Section 3 discusses the proposed ensemble Kalman filter based on a modified Cholesky decomposition for inverse covariance matrix estimation; a theoretical convergence of the estimator in the context of data assimilation as well as its computational effort are discussed. Section 4 presents numerical experiments using the Lorenz-96 model and the Atmospheric General Circulation Model SPEEDY; the results of the new filter are compared against those obtained by the local ensemble transform Kalman filter. Conclusions are drawn in Section 5.

2. Background. The ensemble Kalman filter is a sequential Monte Carlo method for state and parameter estimation of non-linear models such as those found in atmospheric and oceanic sciences [10, 30]. The EnKF popularity is due to its basic theoretical formulation and its relative ease of implementation [10]. Given the background ensemble (5) EnKF builds the analysis ensemble as follows:

\[
X^a = X^b + P^b \cdot H^T \cdot \left[ R + H \cdot P^b \cdot H^T \right] \cdot \Delta \in \mathbb{R}^{n \times N_{ens}},
\]

\[
\Delta = Y^s - H \cdot X^b \in \mathbb{R}^{m \times N_{ens}},
\]

and the matrix of perturbed observations $Y^s \in \mathbb{R}^{m \times N_{ens}}$ is:

\[
Y^s = \left[ y + \epsilon^1, y + \epsilon^2, \ldots, y + \epsilon^{N_{ens}} \right] \in \mathbb{R}^{m \times N_{ens}}, \quad \epsilon^i \sim \mathcal{N}(0, \Sigma),
\]

for $1 \leq i \leq N_{ens}$. For ease of notation we have omitted the time index superscripts.

The use of perturbed observations (7c) during the assimilation provides asymptotically correct analysis-error covariance estimates for large ensemble sizes and makes the formulation of the EnKF statistically consistent [32]. However, it also has been shown that the inclusion of perturbed observations introduces sampling errors in the assimilation [2, 16].

One of the important problems faced by current ensemble based methods is that spurious correlations between distant components in the physical space lead to spurious analysis corrections. Better approximations of the background error covariance matrix are proposed in the literature in order to alleviate this problem. A traditional approximation of $B$ is the Hollingworth and Lonnberg method [13] in which the difference between observations and background states are treated as a combination of background and observations errors. However, this method provides statistics of background errors in observation space, and requires dense observing networks (not the case in practice). EnKF formulations wherein inverse background error covariance
matrix has been proposed in order to exploit features of the precision matrices [31].
Another method has been proposed by Benedetti and Fisher [3] based on forecast
differences in which the spatial correlations of background errors are assumed to be
similar at 24 and 48 hours forecasts. This method can be efficiently implemented in
practice, however, it does not perform well in data-sparse regions, and the statistics
provided are a mixture of analysis and background errors. Another way to reduce the
impact of spurious correlations is based on adaptive modeling [19]. In this context,
the model learns and changes with regard to the data collected (i.e., parameters val-
ues and model structures). This allows to calibrate, in time, the error subspace rank
(i.e., number of empirical orthogonal functions used in the assimilation process), the
tapering parameter (i.e., local domain sizes), and the ensemble size, among others.
Yet another method based on error subspace statistical estimation is proposed in [20].
This approach develops an evolving error subspace, of variable size, that targets the
processes where the dominant errors occur. Then, the dominant errors are minimized
in order to estimate the best model state trajectory with regard to the observations.
Furthermore, spurious correlations can be damped out by using shrinkage covariance
matrix estimators which have been successfully implemented in the context of sequen-
tial data assimilation [25]. We proposed approximations based on autoregressive error
models [9] and using hybrid subspace techniques [8].
Covariance matrix localization artificially reduces correlations between distant
model components via a Schur product with a localization matrix \( \Pi \in \mathbb{R}^{n \times n} \) [29]:

\[
\hat{P}^b = \Pi \circ P^b \in \mathbb{R}^{n \times n},
\]
and then \( P^b \) is replaced by \( \hat{P}^b \in \mathbb{R}^{n \times n} \) in the EnKF analysis equation (7a). For
instance, the entries of \( \Pi \) can decrease with the distance between model components
depending on the radius of influence \( \zeta \):

\[
\{\Pi\}_{i,j} = \exp \left( -\frac{\pi(m_i, m_j)}{2 \cdot \zeta^2} \right), \quad \text{for } 1 \leq i \leq j \leq n,
\]
where \( \pi(m_i, m_j) \) represents the physical distance squared between the model compo-
nents \( m_i \) and \( m_j \). The exponential decay allows to reduce the impact of innovations
between distant model components. The use of covariance matrix localization allevi-
ates the impact of sampling errors. However, the explicit computation of \( \Pi \) (and even
\( P^b \)) is prohibitive owing to numerical model dimensions. Thus, domain localization
methods [6, 17] are commonly used in the context of operational data assimilation.
One of the best EnKF implementations based on domain localization is the local en-
semble transform Kalman filter (LETKF) [28]. In the LETKF the analysis increments
are computed in the space spanned by the ensemble perturbations \( U^b \) defined in (6b).
An approximation of the analysis covariance matrix in this space reads:

\[
\hat{P}^a = \left[(N_{\text{ens}} - 1) \cdot I + Q^T \cdot R^{-1} \cdot Q\right]^{-1} \in \mathbb{R}^{N_{\text{ens}} \times N_{\text{ens}}},
\]
where \( Q = H \cdot U^b \in \mathbb{R}^{m \times N_{\text{ens}}} \) and \( I \) is the identity matrix consistent with the
dimension. The analysis increments in the subspace are:

\[
\alpha^a = \hat{P}^a \cdot Q^T \cdot R^{-1} \cdot [y - H \cdot \bar{x}^b] \in \mathbb{R}^{N_{\text{ens}} \times 1},
\]
from which an estimation of the analysis mean in the model space can be obtained:

\[
\bar{x}^a = \bar{x}^b + U^b \cdot \alpha^a \in \mathbb{R}^{n \times 1}.
\]
Finally, the analysis ensemble reads:

\[
\mathbf{x}^a = \mathbf{x}^k \cdot 1_{N_{\text{ens}}}^T + \mathbf{U}^b \cdot \left[ (N_{\text{ens}} - 1) \cdot \mathbf{\hat{P}} \right]^{1/2} \in \mathbb{R}^{n \times N_{\text{ens}}}. \tag{10d}
\]

The domain localization in the LETKF is performed as follows: each model component is surrounded by a local box of radius $\zeta$. Within each local domain the analysis equations (10) are applied, and therefore a local analysis component is obtained. All local analysis components are mapped back onto the model space to obtain the global analysis state. The local sample covariance matrix (6) is utilized as the covariance estimator of the local $\mathbf{B}$. This can perform well when small radii $\zeta$ are considered during the assimilation step. However, for large values of $\zeta$, the analysis corrections can be impacted by spurious correlations since the local sample covariance matrix can be rank deficient. Consequently, the local analysis increments can perform poorly.

There is an opportunity to reduce the impact of sampling errors by improving the background error covariance estimation. We achieve this by making use of the modified Cholesky decomposition for inverse covariance matrix estimation [5]. Consider the sample of (approximately) $N_{\text{ens}}$ Gaussian random vectors (6b)

\[
\mathbf{U}^b = \begin{bmatrix} \mathbf{u}^{[1]} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{u}^{[N_{\text{ens}}]} \end{bmatrix} \in \mathbb{R}^{n \times N_{\text{ens}}},
\]

with statistical moments $\mathbf{u}^{[j]} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{B})$, for $1 \leq j \leq N_{\text{ens}}$. Denote by $\mathbf{x}^{[i]} \in \mathbb{R}^{N_{\text{ens}} \times 1}$ the vector holding the $i$-th component across all the samples (the $i$-th row of $\mathbf{U}^b$, transposed). The modified Cholesky decomposition arises from regressing each component on his predecessors:

\[
\mathbf{x}^{[i]} = \sum_{j=1}^{i-1} \mathbf{x}^{[j]} \cdot \beta_{i,j} + \mathbf{\varepsilon}^{[i]} \in \mathbb{R}^{N_{\text{ens}} \times 1}, \quad 2 \leq i \leq n, \tag{11}
\]

where $\mathbf{x}^{[j]}$ is the $j$-th model component which precedes $\mathbf{x}^{[i]}$ for $1 \leq j \leq i - 1$, $\mathbf{\varepsilon}^{[i]} = \mathbf{x}^{[1]}$, and $\mathbf{\varepsilon}^{[i]} \in \mathbb{R}^{N_{\text{ens}} \times 1}$ is the error in the $i$-th component regression for $i \geq 2$. Likewise, the coefficients $\beta_{i,j}$ in (11) can be computed by solving the optimization problem:

\[
\beta_{i,[i]} = \arg \min_{\beta} \| \mathbf{x}^{[i]} - \mathbf{Z}^{[i]} \cdot \beta \|_2^2, \tag{12}
\]

where

\[
\mathbf{Z}^{[i]} = \begin{bmatrix} \mathbf{x}^{[1]} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{x}^{[i-1]} \end{bmatrix} \end{bmatrix}^T \in \mathbb{R}^{(i-1) \times N_{\text{ens}}}, \quad 2 \leq i \leq n,
\]

\[
\beta_{i,[i]} = \begin{bmatrix} \beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,i-1} \end{bmatrix}^T \in \mathbb{R}^{(i-1) \times 1}.
\]

The simplest solution of (12) in which one can think reads,

\[
\beta_{i,[i]} = \left[ \mathbf{Z}^{[i]} \cdot \mathbf{Z}^{[i]} \right]^{-1} \cdot \mathbf{Z}^{[i]} \cdot \mathbf{x}^{[i]}.
\tag{13}
\]

The regression coefficients form the lower triangular matrix

\[
\{ \hat{T} \}_{i,j} = \begin{cases} -\beta_{i,j} & \text{for } 1 \leq j < i, \\ 1 & \text{for } j = i, \\ 0 & \text{for } j > i, \end{cases} \tag{14a}
\]
where \( \{ \hat{T} \}_{i,j} \) denotes the \((i,j)\)-th component of matrix \( \hat{T} \in \mathbb{R}^{n \times n} \). The empirical covariances \( \hat{\text{cov}} \) of the residuals \( \varepsilon_{[i]} \) form the diagonal matrix:

\[
\hat{\text{D}} = \text{diag} \left( \hat{\text{cov}}(\varepsilon_{[i]}) \right) = \text{diag} \left( \frac{1}{N_{\text{ens}} - 1} \sum_{j=1}^{N_{\text{ens}}} \{ \varepsilon_{[i]} \}_{j}^{2} \right) \in \mathbb{R}^{n \times n},
\]

where \( \{ \hat{\text{D}} \}_{1,1} = \hat{\text{cov}}(x_{[i]}) \). Then an estimate of \( B^{-1} \) can be computed as follows:

\[
\hat{B}^{-1} = \hat{T}^{T} \cdot \hat{\text{D}} \cdot \hat{T}^{-1} \in \mathbb{R}^{n \times n},
\]

or, by basic matrix algebra identities the estimate of \( B \) reads:

\[
\hat{B} = \hat{T}^{-1} \cdot \hat{\text{D}} \cdot \hat{T}^{-T} \in \mathbb{R}^{n \times n}.
\]

Note that the structure of \( \hat{B}^{-1} \) is strictly related to the structure of \( \hat{T} \). This can be exploited in order to obtain sparse estimators of \( B^{-1} \) by imposing that some entries of \( \hat{T} \) are zero. This is important for high dimensional probability distributions where the explicit computation of \( \hat{B} \) or \( \hat{B}^{-1} \) is prohibitive. The zero components in \( \hat{T} \) can be justified as follows: when two model components are conditionally independent for a given radius of influence \( \zeta \) their corresponding entry in \( \hat{B}^{-1} \) is zero. In the context of data assimilation, the conditional independence of background errors between different model components can be achieved by making use of local neighbourhoods [31]. For a given model component, its neighbourhood is formed by components within the scope of some radius of influence \( \zeta \). We can consider zero correlations between background errors corresponding to model components located at distances that exceed \( \zeta \). In the next section we present an ensemble Kalman filter implementation based on modified Cholesky decomposition for inverse covariance matrix estimation.

### 3. Ensemble Kalman Filter Based On Modified Cholesky Decomposition

In this section we discuss the new ensemble Kalman filter based on modified Cholesky decomposition for inverse covariance matrix estimation (EnKF-MC).

#### 3.1. Estimation of the inverse background covariance

As we mentioned before, the structure of \( \hat{B}^{-1} \) depends on that of \( \hat{T} \). If we assume that the correlations between model components are local, and there are no correlations outside a radius of influence \( \zeta \), we obtain lower-triangular sparse estimators of \( \hat{T} \). Consequently, the resulting \( \hat{B}^{-1} \) will also be sparse, and \( \hat{B} \) will be localized. Since the regression (11) can be performed only on the predecessors of each model component, an ordering (labeling) must be set on the model components prior the computation of \( \hat{T} \). Since we work with gridded models we consider column-major and row-major orders. Figure 1 shows the neighbourhood and the predecessors of the model component 6 when column-major order is utilized.
The estimation of $\hat{B}^{-1}$ proceeds as follows:

1. For the $i$-th model component, making use of the truncated singular value decomposition [15, 14, 12], estimate the regression coefficients $\beta_{i,j}$, for $1 \leq i \leq n$ and $j \in P(i, \zeta)$, that satisfies:

$$x_{[i]} = \sum_{j \in P(i, \zeta)} \beta_{i,j} \cdot x_{[j]} + \varepsilon_{[i]} \in \mathbb{R}^{N_{\text{ens}} \times 1},$$

where $P(i, \zeta)$ denotes the set of predecessors indexes for model component $i$ when the radius of influence is $\zeta$.

2. Build the matrices

$$\{\hat{T}\}_{i,j} = \begin{cases} -\beta_{i,j} & j \in P(i, \zeta), \\ 1 & i = j, \\ 0 & \text{otherwise} \end{cases}$$

and $\hat{D}$ according to equation (14b). Note that, the number of non-zero elements in the $i$-th row of $\hat{T}$ equals the number of predecessors for model component $i$. Note that, since residuals in (16) are never zero, they implicitly inflate the estimated variances, and therefore the use of covariance inflation may become unnecessary with the EnKF-MC assimilation.

3.2. Formulation of EnKF-MC. Once $\hat{B}^{-1}$ is estimated, the EnKF based on modified Cholesky decomposition (EnKF-MC) computes the analysis using Kalman’s formula:

$$(17a) \quad x^a = x^b + \hat{A} \cdot H^T \cdot R^{-1} \cdot \Delta \in \mathbb{R}^{n \times N_{\text{ens}}},$$

where $\hat{A} \in \mathbb{R}^{n \times n}$ is the estimated analysis covariance matrix

$$\hat{A} = \left(\hat{B}^{-1} + H^T \cdot R^{-1} \cdot H\right)^{-1},$$

and $\Delta \in \mathbb{R}^{m \times N_{\text{ens}}}$ is the innovation matrix on the perturbed observations given in (7b).

More efficient alternatives to (17a) can be obtained by making use of elementary matrix identities:

$$x^a = \hat{A} \cdot \left[\hat{B}^{-1} \cdot x^b + H^T \cdot \hat{B}^{-1} \cdot Y^x\right] \in \mathbb{R}^{n \times N_{\text{ens}}},$$

$$(17b) \quad x^a = x^b + \hat{T}^{-1} \cdot \hat{D}^{1/2} \cdot V_B^T \cdot \left[R + V_B \cdot \hat{B}^{-1} \cdot V_B^T\right]^{-1} \cdot \Delta,$$

$$(17c) \quad x^a = x^b + \hat{T}^{-1} \cdot \hat{D}^{1/2} \cdot V_B^T \cdot \left[R + V_B \cdot \hat{B}^{-1} \cdot V_B^T\right]^{-1} \cdot \Delta,$$
where \( \mathbf{V}_\mathbf{B} = \mathbf{H} \cdot \mathbf{\hat{T}}^{-1} \cdot \mathbf{\hat{D}}^{1/2} \in \mathbb{R}^{n \times m} \), and \( \mathbf{Y}^* \) are the perturbed observations. The formulation (17c) is well-known as the EnKF dual formulation, (17b) is known as the EnKF primal formulation, and the equation (17a) is the incremental form of the primal formulation. It can be easily shown that, by making use of the iterative Sherman Morrison formula [24], the computational effort of the method is bounded by,

\[
\mathcal{O} \left( m \cdot n + m \cdot n \cdot N_{\text{ens}} \right).
\]

### 3.3. Convergence of the covariance inverse estimator

In this section we prove the convergence of the \( \mathbf{\hat{B}}^{-1} \) estimator in the context of data assimilation.

**Comment 1 (Sparse Cholesky factors and localization).** The modified Cholesky decomposition for inverse covariance matrix estimation can be seen as a form of covariance matrix localization method in which the resulting matrix approximates the inverse of a localized ensemble covariance matrix. This process is implicit in the resulting estimator when only a local neighborhood for each model component is utilized in order to perform the local regression and to estimate \( \mathbf{\hat{T}} \) and \( \mathbf{\hat{D}} \).

**Comment 2 (Gaussian assumption relaxation).** The results proposed by Bickel and Levina in [4] are based on the assumption that samples follow a Normal distribution with covariance matrix \( \mathbf{B} \) whose component correlations decay exponentially. However, the Gaussian assumption can be relaxed to distributions whose second moment satisfies such condition and even more, high-order moments are bounded by second ones, for instance, probability distributions from the exponential family such as the Exponential, the Normal, and the Gamma.

To start our proof, the inverse of the (exact) background error covariance matrix \( \mathbf{B}^{-1} \) and of its estimator \( \mathbf{\hat{B}}^{-1} \) can be written as

\[
\begin{align*}
\mathbf{B}^{-1} &= \left[ \mathbf{I} - \mathbf{\hat{C}} \right]^T \cdot \mathbf{D}^{-1} \cdot \left[ \mathbf{I} - \mathbf{\hat{C}} \right] \in \mathbb{R}^{n \times n} \quad \text{(18a)} \\
\mathbf{\hat{B}}^{-1} &= \left[ \mathbf{I} - \mathbf{\hat{C}} \right]^T \cdot \mathbf{\hat{D}}^{-1} \cdot \left[ \mathbf{I} - \mathbf{\hat{C}} \right] \in \mathbb{R}^{n \times n},
\end{align*}
\]

respectively, where \( \mathbf{\hat{C}} = \mathbf{I} - \mathbf{\hat{T}} \in \mathbb{R}^{n \times n} \) and \( \mathbf{C} = \mathbf{I} - \mathbf{T} \in \mathbb{R}^{n \times n} \). Moreover, \( \mathbf{D} \) and \( \mathbf{\hat{D}} \) are diagonal matrices:

\[
\mathbf{D} = \text{diag} \left\{ d_1^2, d_2^2, \ldots, d_n^2 \right\}, \quad \mathbf{\hat{D}} = \text{diag} \left\{ \hat{d}_1^2, \hat{d}_2^2, \ldots, \hat{d}_n^2 \right\}
\]

where \( \{ \mathbf{D} \}_{i,i} = d_i^2 \) and \( \{ \mathbf{\hat{D}} \}_{i,i} = \hat{d}_i^2 \), for \( 1 \leq i \leq n \). In what follows we denote by \( \mathbf{\hat{c}}^{(j)} \in \mathbb{R}^{n \times 1} \) and \( \mathbf{c}^{(j)} \in \mathbb{R}^{n \times 1} \) the \( j \)-th columns of matrices \( \mathbf{\hat{C}} \) and \( \mathbf{C} \), respectively, for \( 1 \leq j \leq n \).

**Definition 1 (Class of matrices under consideration.).** We consider the class of covariance matrices with correlations decreasing quickly:

\[
\begin{align*}
\mathcal{U}^{-1} (\varepsilon_0, C, \alpha) &= \left\{ \mathbf{B} : 0 < \varepsilon_0 \leq \lambda_{\min} (\mathbf{B}) \leq \lambda_{\max} (\mathbf{B}) \leq \varepsilon_0^{-1}, \right. \\
&\left. \max_k \sum_{j \not\in P(k,\zeta)} \left| \{ \mathbf{T} \}_{k,j} \right| \leq C \cdot \zeta^{-\alpha} \right\}
\end{align*}
\]
where $B^{-1} = T^T D^{-1} T$, $\alpha$ is the decay rate (related to the dynamics of the numerical model), and $P(k, \zeta)$ for a given radius $\zeta$, denotes the row-indexes of predecessors for model component $k$.

**Theorem 1** (Error in the covariance inverse estimation). Uniformly for $B \in \mathcal{U}(\varepsilon_0, C, \alpha)$, if $\zeta \approx [N_{\text{ens}}^{-1} \cdot \log n]^{-1/2(\alpha+1)}$ and $N_{\text{ens}}^{-1} \cdot \log n = o(1)$,

\begin{equation}
(19b) \quad \|\hat{B}^{-1} - B^{-1}\|_\infty = O \left( \left( \frac{\log(n)}{N_{\text{ens}}} \right)^{\alpha(\alpha+1)/2} \right),
\end{equation}

where $\| \cdot \|_\infty$ denotes the infinity norm (matrix or vector)

In order to prove Theorem 1, we need the following result.

**Lemma 1.** Under the conditions of Theorem 1, uniformly on $\mathcal{U}^{-1}$

\begin{align}
(20a) & \quad \max \left\{ \|\hat{c}^{(j)} - c^{(j)}\|_\infty : 1 \leq j \leq n \right\} = O \left( N_{\text{ens}}^{-1/2} \log^{1/2} n \right), \\
(20b) & \quad \max \left\{ |d_j^2 - d_j^2| : 1 \leq j \leq n \right\} = O \left( [N_{\text{ens}}^{-1} \log n]^{(\alpha/2)(\alpha+1)} \right), \\
(20c) & \quad \|C\|_\infty = O(1) \text{ and } \|D^{-1}\|_\infty = O(1).
\end{align}

The proof of Lemma 1 is based on the following results of Bickel and Levina in [5].

**Lemma 2.** ([5, Lemma A.2]) Let $\nu^{[k]} \sim \mathcal{N}(\mathbf{0}, B)$ and $\lambda_{\max}(B) \leq \varepsilon_0^{-1} < \infty$, for $1 \leq k \leq N_{\text{ens}}$. Then, if $\{B\}_{i,j}$ denotes the $(i,j)$-th component of $B$, for $1 \leq i \leq j \leq n$,

\begin{equation}
(21) \quad \text{Prob} \left[ \sum_{k=1}^{N_{\text{ens}}} \left\{ \nu^{[k]} \right\}_k \cdot \left\{ \nu^{[k]} \right\}_j - \{B\}_{i,j} \geq N_{\text{ens}} \cdot \nu \right]
\leq C_1 \cdot \exp \left( -C_2 \cdot N_{\text{ens}} \cdot \nu^2 \right),
\end{equation}

for $|\nu| \leq \delta$, where $\{\nu^{[k]}\}_i$ is the $i$-th component of the sample $\nu^{[k]}$, for $1 \leq k \leq N_{\text{ens}}$, and $1 \leq i \leq n$. Likewise, $C_1$, $C_2$ and $\delta$ depend on $\varepsilon_0$ only.

**Proof of Lemma 1.** In what follows we denote by $\text{cov}$ and $\hat{\text{cov}}$ the true and the empirical covariances, respectively. In the context of EnKF we have that $\text{cov}(U^b) = B$.

Recall that

\begin{equation}
\hat{\text{cov}}(U^b) = P^b = \frac{1}{N_{\text{ens}} - 1} \cdot U_b \cdot U_b^T = \frac{1}{N_{\text{ens}} - 1} \sum_{k=1}^{N_{\text{ens}}} u_b^{[k]} \cdot u_b^{[k]T}
\end{equation}

$\iff \left\{ \hat{\text{cov}}(U^b) \right\}_{i,j} = \frac{1}{N_{\text{ens}} - 1} \sum_{k=1}^{N_{\text{ens}}} \left\{ u_b^{[k]} \right\}_i \cdot \left\{ u_b^{[k]} \right\}_j$.

For $\nu > 0$, $\left\{ \nu^{[k]} \right\}_i \cdot \left\{ \nu^{[k]} \right\}_j - \{B\}_{i,j} \geq N_{\text{ens}} \cdot \nu$ implies $\left\{ \nu^{[k]} \right\}_i \cdot \left\{ \nu^{[k]} \right\}_j - \{B\}_{i,j} \geq (N_{\text{ens}} - 1) \cdot \nu$, and therefore by Lemma 2 we have:

\begin{equation}
(22a) \quad \|\text{cov}(U^b) - \hat{\text{cov}}(U^b)\|_\infty = O \left( N_{\text{ens}}^{-1/2} \cdot \log^{1/2} n \right),
\end{equation}
since the entries of $\text{cov} (U^b) - \hat{\text{cov}} (U^b)$ can be bounded by:
\[
\left| \{ \text{cov} (U^b) - \hat{\text{cov}} (U^b) \} \right|_{ij} \leq N_{\text{ens}}^{-1} \sum_{k=1}^{N_{\text{ens}}} \left| \{ u_{b[k]}^i \} \cdot \{ u_{b[k]}^j \} \right| - \{ B \}_{i,j}.
\]

Lemma 2 ensures that:
\[
\operatorname{Prob} \left[ \max_{i,j} \left| \sum_{k=1}^{N_{\text{ens}}} \left| \{ u_{b[k]}^i \} \cdot \{ u_{b[k]}^j \} \right| - \{ B \}_{i,j} \right| \geq \nu \right] \leq C_1 \cdot n^2 \cdot \exp \left( -C_2 \cdot N_{\text{ens}} \cdot \nu^2 \right),
\]
for $|\nu| \leq \delta$. Let $\nu = \left( \frac{\log n^2}{N_{\text{ens}} \cdot C_2} \right)^{1/2} \cdot M$, for $M$ arbitrary.

Since $Z_{[i]}$ stores the columns of $U^b$ corresponding to the predecessors of model component $i$, an immediate consequence of (22a) is
\[
(22b) \quad \max_i \left\| \text{cov} (Z_{[i]}) - \hat{\text{cov}} (Z_{[i]}) \right\|_\infty = \mathcal{O} \left( N_{\text{ens}}^{-1/2} \cdot \log^{1/2} n \right).
\]

Also,
\[
\left\| B^{-1} \right\|_\infty = \left\| \text{cov} (U^b)^{-1} \right\|_\infty \leq \varepsilon_0^{-1}.
\]

According to (12), we have,
\[
\{ c_{[i]} \} = \{ \text{cov} (Z_{[i]})^{-1} Z_{[i]} \cdot x_{[i]} \} \quad \text{and} \quad \{ \hat{c}_{[i]} \} = \{ \hat{\text{cov}} (Z_{[i]})^{-1} Z_{[i]} \cdot x_{[i]} \},
\]
therefore:
\[
\max_k \left| \{ c_{[i]} \} - \{ \hat{c}_{[i]} \} \right| = \max_k \left| \{ \text{cov} (Z_{[i]})^{-1} Z_{[i]} \cdot x_{[i]} \} - \{ \hat{\text{cov}} (Z_{[i]})^{-1} Z_{[i]} \cdot x_{[i]} \} \right|
\]
\[
= \max_k \left| \{ \text{cov} (Z_{[i]})^{-1} Z_{[i]} \cdot x_{[i]} \} - \{ \text{cov} (Z_{[i]})^{-1} \cdot Z_{[i]} \cdot x_{[i]} \} \right|
\]
\[
= \max_k \left| \{ \text{cov} (Z_{[i]})^{-1} - \hat{\text{cov}} (Z_{[i]})^{-1} \cdot Z_{[i]} \cdot x_{[i]} \} \right|
\]
\[
= \mathcal{O} \left( N_{\text{ens}}^{-1/2} \cdot \log^{1/2} n \right)
\]
from which (20a) follows. Note that:
\[
x_{[i]} = \sum_{j \in P(i, \zeta)} \left( \{ \hat{c}_{[i]} \} \cdot x_{[j]} + \hat{e}_{[i]} \right) \Leftrightarrow \hat{\text{cov}} (x_{[i]}) = \text{cov} \left( \sum_{j \in P(i, \zeta)} \left( \{ \hat{c}_{[i]} \} \cdot x_{[j]} + \hat{e}_{[i]} \right) \right)
\]
\[
\Leftrightarrow \hat{\text{cov}} (x_{[i]}) = \text{cov} \left( \sum_{j \in P(i, \zeta)} \left( \{ \hat{c}_{[i]} \} \cdot x_{[j]} \right) \right) + \text{cov} (\hat{e}_{[i]})
\]
\[
\Leftrightarrow \hat{d}_{[i]}^2 = \text{cov} (x_{[i]}) - \text{cov} \left( \sum_{j \in P(i, \zeta)} \left( \{ \hat{c}_{[i]} \} \cdot x_{[j]} \right) \right),
\]
and similarly
\[
d_{[i]}^2 = \text{cov} (x_{[i]}) - \text{cov} \left( \sum_{j \in P(i, \zeta)} \{ c_{[i]} \} \cdot x_{[j]} \right).
\]
The claim \((20b)\) and the first part of \((20c)\) follow from \((22a)\), \((22b)\) and \((24)\).

Since

\[
\left| d_i^2 - d_i' \right| \leq \left| \text{cov} \left( \{x_i \} \right) - \text{cov} \left( \{x_i' \} \right) \right|
\]

\[
\text{(25)}
\]

\[
+ \left| \text{cov} \left( \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot x_{[j]} \right) \right|
\]

\[
+ \left| \text{cov} \left( \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot x_{[j]} \right) - \text{cov} \left( \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot x_{[j]} \right) \right|
\]

By Lemma 2 the maximum over \(i\) of the first term is:

\[
\max_i \left| \text{cov} \left( \{x_i \} \right) - \text{cov} \left( \{x_i' \} \right) \right| = \mathcal{O} \left( \frac{n_{\text{ens}}^{-1} \cdot \log^{\frac{1}{2}} n}{1} \right).
\]

The second term can be bounded as follows:

\[
\left| \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot \text{cov} \left( \{x_{[j]} \} \right) \right| \leq \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot \text{cov} \left( \{x_{[j]} \} \right)
\]

\[
\leq \max_k \left[ \left\{ \{c_i^{[k]} \}_k \right\}_k \cdot \text{cov} \left( \{x_{[k]} \} \right) \right] = \mathcal{O} \left( \frac{n_{\text{ens}}^{-1} \cdot \log n}{1} \right)
\]

by \((20a)\) and \(\|B\| \leq \varepsilon_0^{-1}\). The third term can be bounded similarly. Thus \((20b)\) follows. Furthermore,

\[
d_i^2 = \text{cov} \left( x_{[i]} - \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j \cdot x_{[j]} \right) \geq \varepsilon_0 \cdot \left( 1 + \sum_{j \in P(1, \zeta)} \left\{ \{c_i^{[j]} \}_j \right\}_j^2 \right) \geq \varepsilon_0,
\]

and the lemma follows. \(\square\)

We now are ready to prove Theorem 1.

**Proof of Theorem 1.** We need only check that:

\[
\text{(26a)} \quad \| \hat{B}^{-1} - B^{-1} \|_\infty = \mathcal{O} \left( \frac{n_{\text{ens}}^{-1/2} \cdot \log^{1/2} n}{1} \right)
\]

and

\[
\text{(26b)} \quad \| B^{-1} - \Phi_\zeta \left( B^{-1} \right) \|_\infty = \mathcal{O} \left( \frac{\zeta^{-\alpha}}{1} \right)
\]

where the entries of \(\Phi_\zeta \left( B^{-1} \right)\) are given by:

\[
\text{(26c)} \quad \left\{ \Phi_\zeta \left( B^{-1} \right) \right\}_{k,j} = \begin{cases} \left\{ B^{-1} \right\}_{k,j}, & \text{for } j \in P(k, \zeta) \\ 0, & \text{otherwise} \end{cases}, \text{ for } 1 \leq k \leq n
\]

We first prove \((26a)\). By definition,

\[
\text{(27)} \quad \hat{B}^{-1} - B^{-1} = \hat{T}^T \cdot \hat{D}^{-1} \cdot \hat{T} - T^T \cdot D^{-1} \cdot T.
\]
Applying the standard inequality:
\[
\|T^T \cdot D^{-1} \cdot T - \hat{T}^T \cdot \hat{D}^{-1} \cdot \hat{T}\| \leq \|T^T - \hat{T}^T\| \cdot \|D - \hat{D}\| \cdot \|\hat{T}\| \\
+ \|\hat{T}\| \cdot \|D - \hat{D}\| \cdot \|\hat{T}^T - T^T\| \\
+ \|\hat{D}\| \cdot \|T - \hat{T}\| \cdot \|\hat{T}^T - T^T\| \\
+ \|\hat{T}^T\| \cdot \|D - \hat{D}\| \cdot \|T - \hat{T}\| \\
+ \|D - \hat{D}\| \cdot \|T - \hat{T}\| \cdot \|T^T - \hat{T}^T\|
\]

all previous terms can be bounded making use of Lemma 1 and therefore, (26a) follows.
Likewise, for (26b), we need to note that for any matrix \(M\),
\[
\|M \cdot M^T - \Phi_\zeta(M) \cdot \Phi_\zeta(M)^T\|_\infty \leq 2\|M\|_\infty \|\Phi_\zeta(M) - M\|_\infty + \|\Phi_\zeta(M) - M\|_\infty^2
\]
and by letting \(M = T^T \cdot D^{-1/2}\), the theorem follows from Definition 1.

4. Numerical Experiments. We make use of two numerical models in order to assess the accuracy of the proposed EnKF implementation: the Lorenz-96 model and the Atmospheric General Circulation Model (AT-GCM) SPEEDY. Given its simple formulation and low computational cost, the Lorenz-96 model allows to run large numbers of experiments to determine the statistics of errors associated with ensemble forecasts. We use the Lorenz-96 model to investigate the effect of the ensemble inflation factor on the analysis quality. The SPEEDY model provides more realistic test scenarios, closer to those found in operational data assimilation.

The EnKF-MC analyses are compared against those obtained with the standard LETKF implementation of Hunt et al. [26, 27, 28]. As a measure of accuracy, the \(L_2\) norm of the analysis error is computed at each assimilation step:
\[
(28a) \quad \lambda_i = \|x_i^{ref} - x_i^0\|_2, \quad \text{for } 1 \leq i \leq M,
\]
where \(M\) denotes the number of observation times within the assimilation window, \(x_i^{ref}\) and \(x_i^0\) are the reference solution and the analysis states at time \(i\), respectively.
The root-mean-square-error across all times is used to assess the average accuracy of a filter solution throughout the entire assimilation window:
\[
(28b) \quad \text{RMSE} = \sqrt{\frac{1}{M} \sum_{k=1}^{M} \lambda_i^2}.
\]
The threshold used in the truncated singular value decomposition during the computation of \(\tilde{B}^{-1}\) is 0.10. During the assimilation steps, the data error covariance matrices \(R_k\) are used (no representativeness errors are involved during the assimilations). The different EnKF implementations are performed making use of FORTRAN and specialized libraries such as BLAS and LAPACK are used in order to perform the algebraic computations.

4.1. Lorenz-96 model. The Lorenz-96 model [22] is described by the following set of ordinary differential equations:
\[
(29) \quad \frac{dx_j}{dt} = \begin{cases} 
(x_2 - x_{n-1}) \cdot x_n - x_1 + F & \text{for } j = 1, \\
(x_{j+1} - x_{j-2}) \cdot x_{j-1} - x_j + F & \text{for } 2 \leq j \leq n - 1, \\
(x_1 - x_{n-2}) \cdot x_{n-1} - x_n + F & \text{for } j = n,
\end{cases}
\]
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.

The reference initial condition of the model is obtained after a long propagation of a random initial state. The background errors at the initial time are samples from a zero-mean normal distribution with covariance matrix $0.05 \cdot I \in \mathbb{R}^{n \times n}$ where $I$ is the identity matrix. Observations are taken every 0.5 time units (which corresponds to 3.5 days in the atmosphere). Observations errors are drawn from a zero-mean Gaussian distribution with covariance matrix $0.01 \cdot I \in \mathbb{R}^{m \times m}$. The assimilation window consists of 25 observations evenly distributed in time. Two ensemble sizes are utilized during the tests: $N_{\text{ens}} = 20$ and $N_{\text{ens}} = 60$. Two inflation factors are considered for the experiments, a smaller value $\alpha = 1.05$ and a larger value $\alpha = 1.09$.

For each configuration we perform 45 runs, each with a different reference initial condition, and with different realizations of the background and observation errors. At each time $m = 30$ components of the model space are observed; the observed components are randomly selected for each run, and are therefore different for each scenario.

The logarithm of mean errors (28a) at each assimilation time are reported in Figure 2 for all inflation factors and ensemble sizes. Both filters converge for all experimental configurations. The traditional LETKF analyses become more accurate when the inflation factor is increased; this behavior is expected and has been well-studied previously [1, 11]. Results reveal that the use of inflation can also improve the quality of EnKF-MC analyses. The quality of the results with both filters is also better for the larger ensemble size, as expected.

A comparison of EnKF-MC versus LETKF results shows that, for all ensemble sizes and inflation factors, the EnKF-MC converges faster than LETKF. Moreover, for each scenario, the logarithm of standard deviations of the EnKF-MC results across the 45 different runs are visibly smaller than the standard deviations of the LETKF results for most of the assimilation steps. This indicates that EnKF-MC provides the more robust implementation.

### 4.2. SPEEDY model.

In this section we study the performance of the proposed EnKF-MC implementation. The experiments are performed using the atmospheric general circulation model SPEEDY [18, 23]. SPEEDY is a hydrostatic, spectral coordinate, spectral transform model in the vorticity-divergence form, with semi-implicit treatment of gravity waves. The number of layers in the SPEEDY model is 8 and the T-63 model resolution ($192 \times 96$ grids) is used for the horizontal space discretization of each layer. Four model variables are part of the assimilation process: the temperature ($^\circ\text{C}$), the zonal and the meridional wind components ($\text{m/s}$), and the specific humidity ($\text{g/kg}$). The total number of model components is $n = 589,824$. The number of ensemble members is $N_{\text{ens}} = 94$ for all the scenarios. The model state space is approximately 6,274 times larger than the number of ensemble members ($n \gg N_{\text{ens}}$).

Starting with the state of the system $\mathbf{x}^{\text{ref}}_{-3}$ at time $t_{-3}$, the model solution $\mathbf{x}^{\text{ref}}_{-3}$ is propagated in time over one year:

$$\mathbf{x}^{\text{ref}}_{-2} = M_{t_{-3} \rightarrow t_{-2}} (\mathbf{x}^{\text{ref}}_{-3}).$$

The reference solution $\mathbf{x}^{\text{ref}}_{-2}$ is used to build a perturbed background solution:

$$\mathbf{x}^{b}_{-2} = \mathbf{x}^{\text{ref}}_{-2} + \epsilon^{b}_{-2}, \quad \epsilon^{b}_{-2} \sim \mathcal{N} \left(0, \text{diag} \left\{(0.05 \{x\}^2)\right\} \right).$$
Fig. 2: Experimental results with the Lorenz-96 model (29). The time evolution of mean analysis errors (first column) and their standard deviation (second column) across the 45 different experimental configurations are reported.
The perturbed background solution is propagated over another year to obtain the background solution at time $t_{-1}$:

$$\textbf{x}_b^{t_{-1}} = \mathcal{M}_{t_{-2} \rightarrow t_{-1}} (\hat{\textbf{x}}_b^{t_{-2}}).$$

(31)

This model propagation attenuates the random noise introduced in (30) and makes the background state (31) consistent with the physics of the SPEEDY model. Then, the background state (31) is utilized in order to build an ensemble of perturbed background states:

$$\hat{\textbf{x}}_b^{t_{-1}} = \hat{\textbf{x}}_b^{t_{-2}} + \textbf{e}_b^{t_{-1}}, \; \textbf{e}_b^{t_{-1}} \sim \mathcal{N} \left( 0_n, \text{diag} \left\{ (0.05 \{ \textbf{x}_b^{t_{-1}} \}_i)^2 \right\} \right), \; 1 \leq i \leq N_{\text{ens}},$$

(32)

from which, after three months of model propagation, the initial ensemble is obtained at time $t_0$:

$$\textbf{x}_{0}^{t_{-1} \rightarrow t_0} = \mathcal{M}_{t_{-1} \rightarrow t_0} (\hat{\textbf{x}}_b^{t_{-1}}).$$

Again, the model propagation of the perturbed ensemble ensures that the ensemble members are consistent with the physics of the numerical model.

The experiments are performed over a period of 24 days, where observations are taken every 2 days ($M = 12$). At time $k$ synthetic observations are built as follows:

$$\textbf{y}_k = \textbf{H}_k \cdot \textbf{x}_k^{\text{ref}} + \textbf{e}_k, \; \textbf{e}_k \sim \mathcal{N} \left( 0_m, \text{R}_k \right), \; \text{R}_k = \text{diag} \left\{ (0.01 \{ \textbf{H}_k \cdot \textbf{x}_k^{\text{ref}} \}_i)^2 \right\}.$$

The observation operators $\textbf{H}_k$ are fixed throughout the time interval. We perform experiments with several operators characterized by different proportions $p$ of observed components from the model state $\textbf{x}_k^{\text{ref}}$ ($m \approx p \cdot n$). We consider five different values for $p$: 1.00, 0.50, 0.12, 0.06 and 0.04 which represent 100%, 50%, 12 %, 6 % and 4 % of the total number of model components, respectively. All observation networks are uniformly distributed in space.

### 4.3. Results with dense observation networks.

We first consider dense observational networks in which 100% and 50% of the model components are observed. We vary the radius of influence $\zeta$ from 1 to 5 grid points. Figures 3a and 3b show the RMSE values for the LETKF and EnKF-MC analyses for different values of $\zeta$ for the specific humidity when 100% of model components are observed. When the radius of influence is increased the quality of the LETKF results degrades due to spurious correlations. This is expected since the local estimation of correlations in the context of LETKF is the sample covariance matrix. For instance, for a radius of influence of 1, the total number of local components for each local box is 36 which matches the dimension of the local background error distribution. Now, when we compare it against the ensemble size (96 ensemble members), sufficient degrees of freedom (95 degrees of freedom) are available in order to estimate the local background error distribution onto the ensemble space, and consequently all directions of the local probability error distribution are accounted during the estimation and posterior assimilation. On the other hand, when the radius of influence is 5, the local box sizes have dimension 484 (model components) which is approximately 5 times larger than the ensemble size. Thus, when the analysis increments are computed onto the ensemble space, just part of the local background error distribution is accounted during the assimilation. Consequently, the larger the local box, the more local background error information cannot be represented in the ensemble space.
E. D. NINO-RUIZ, A. SANDU, AND X. DENG

\[ \zeta = 2 \text{ and } p = 100\% . \]

\[ \zeta = 5 \text{ and } p = 100\% . \]

\[ \zeta = 2 \text{ and } p \sim 4\% . \]

\[ \zeta = 5 \text{ and } p \sim 4\% . \]

Fig. 3: Experimental results with the SPEEDY model. RMSE of specific humidity analyses with a dense observational network are reported. When the radius of influence \( \zeta \) is increased the performance of LETKF degrades.

Figures 4a and 4b show that EnKF-MC analyses improve with increasing radius of influence \( \zeta \). Since a dense observational network is considered during the assimilation, when the radius of influence is increased, a better estimation of the state of the system is obtained by the EnKF-MC. This can be seen clearly in Figure 4, where the RMSE values within the assimilation window are shown for the LETKF and the EnKF-MC solutions for the specific humidity variable and different values of \( \zeta \) and \( p \). The quality of the EnKF-MC analysis for \( \zeta = 5 \) is better than that of the LETKF with \( \zeta = 1 \). Likewise, when a full observational network is considered (\( p = 100\% \)), the proposed implementation outperforms the LETKF implementation. EnKF-MC is able to exploit the large amount of information contained in dense observational networks by properly estimating the local background error correlations.

4.4. Results with sparse observation networks. We vary the values of \( \zeta \) from 1 to 5. Three sparse observational networks with \( p = 12\% , 6\% , \) and \( 4\% , \) respectively are considered.

Figures 3c and 3d show the RMSE values of the specific humidity analyses for
different radii of influence and 4% of the model components being observed. The best performance of the LETKF analyses is obtained when the radius of influence is set to 2. Note that for $\zeta = 1$ the LETKF performs poorly, which is expected since during the assimilation most of model components will not have observations in their local boxes. For $\zeta \geq 3$ the effects of spurious correlations degrade the quality of the LETKF analysis. On the other hand, the background error correlations estimated by the modified Cholesky decomposition allows the EnKF-MC formulation to obtain good analyses even for largest radius of influence $\zeta = 5$.

Figures 4c and 4d show the RMSE values of the LETKF and the EnKF-MC implementations for different radii of influences and two sparse observational networks. Clearly, when the radius of influence is increased, in the LETKF context, the analysis corrections are impacted by spurious correlations. On the other hand, the quality of the results in the EnKF-MC case is considerably better. When data errors com-
ponents are uncorrelated $\zeta$ can be seen as a free parameter and the choice can be based on the “optimal performance of the filter”. For the largest radius of influence $\zeta = 5$ the RMSE values of the ENKF-MC and the LETKF implementations differ by one order of magnitude. Hence, the estimation of background errors via $\hat{\mathbf{B}}$ can reduce the impact of spurious correlations; the RMSE values of the EnKF-MC analyses remain small at all assimilation times, from which we infer that the background error correlations are properly estimated. On the other hand, the impact of spurious correlations is evident in the context of LETKF. Since most of the model components are unobserved, the background error correlations drive the quality of the analysis, and spurious correlations lead to a poor performance of the filter at many assimilation times. Figure 5 provides snapshots of the meridional and the zonal wind components, respectively, at the first assimilation time. For this particular case the percentage of observed model components is $p = 4\%$. At this step, only the initial observation has been assimilated in order to compute the analysis corrections by the EnKF-MC.

Fig. 5: Experimental results with the SPEEDY model. Snapshots of the reference solution, background state, and analysis fields from the EnKF-MC and LETKF for the second layer of the zonal wind component ($u$) are shown.
and the LETKF methods. The background solution contains erroneous waves for the zonal and the meridional wind components. For instance, for the \( u \) model variable, such waves are clearly present near the poles. After the first assimilation step, the LETKF analysis solution dissipates the erroneous waves but, the numerical values of the wind components are slightly greater than those of the reference solutions. This numerical difference increases at later times due to the highly-nonlinear dynamics of SPEEDY. On the other hand, the EnKF-MC implementation recovers the reference shape, and the analysis values of the numerical model components are close to that of the reference solution. This shows again that the use of the modified Cholesky decomposition as the estimator of the background error correlations can mitigate the impact of spurious error correlations.

5. Conclusions. This paper develops an efficient implementation of the ensemble Kalman filter, named EnKF-MC, that is based on a modified Cholesky decomposition to estimate the inverse background covariance matrix. This new approach has several advantages over classical formulations. First, a predefined sparsity structure can be built into the factors of the inverse covariance. This reflects the fact that if two distant model components are uncorrelated then the corresponding entry in the inverse covariance matrix is zero; the only nonzero entries in the Cholesky factors correspond to components of the model that are located in each other’s proximity. Therefore, imposing a sparsity structure on the inverse background covariance matrix is a form of covariance localization. Second, the formulation allows for a rigorous theoretical analysis; we prove the convergence of the covariance estimator for a number of ensemble members that is proportional to the logarithm of the number of states of the model therefore, when \( N_{\text{ens}} \approx \log n \), the background error correlations can be well-estimated making use of the modified Cholesky decomposition.

We discuss different implementations of the new EnKF-MC, and assess their computational effort. We show that domain decomposition can be used in order to decrease even more the computational effort of the proposed implementation. Numerical experiments are carried out using the Atmospheric General Circulation Model SPEEDY reveal that the analyses obtained by EnKF-MC are better than those of the LETKF in the root mean square sense when sparse observations are used in the analysis. For dense observation grids the EnKF-MC solutions are improved when the radius of influence increases, while the opposite holds true for LETKF analyses. (We stress the fact that these conclusions are true for our implementation of the basic LETKF; other implementations may incorporate advances that could make the filter perform considerably better). The use of modified Cholesky decomposition can mitigate the impact of spurious correlation during the assimilation of observations.

REFERENCES


