SIAM J. SCI. COMPUT.
Vol. 40, No. 2, pp. A867–A886

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 one 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 EnKF-MC 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

DOI. 10.1137/16M1097031

1. Introduction.

The goal of sequential data assimilation is to estimate the true state of a dynamical system $\mathbf{x}^{\text{true}} \in \mathbb{R}^{n\times 1}$ 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]:

$$\mathbf{x}_k = \mathcal{M}_{t_{k-1}\to t_k}\left(\mathbf{x}_{k-1}\right) \in \mathbb{R}^{n\times 1} \text{ for } \mathbf{x} \in \mathbb{R}^{n\times 1}, \tag{1}$$

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

$$\mathbf{x}_k^{\text{b}} - \mathbf{x}^{\text{true}} = \boldsymbol{\nu}_k \sim \mathcal{N}\left(\mathbf{0},\, \mathbf{B}_k\right) \in \mathbb{R}^{n\times 1}, \tag{2}$$

where the prior error $\boldsymbol{\nu}_k$ is assumed to have mean zero and a covariance matrix $\mathbf{B}_k \in \mathbb{R}^{n\times n}$. Noisy observations (measurements) of the true state $\mathbf{y}_k \in \mathbb{R}^{m\times 1}$ are

*Submitted to the journal's Methods and Algorithms for Scientific Computing section October 3, 2016; accepted for publication (in revised form) November 8, 2017; published electronically March 20, 2018.

http://www.siam.org/journals/sisc/40-2/M109703.html

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 (asandu7@vt.edu, http://people.cs.vt.edu/~asandu/).

§Department of Statistics, Virginia Tech, Blacksburg, VA 24060 (xdeng@vt.edu, http://www.apps.stat.vt.edu/deng/).

taken, and the observation errors $\boldsymbol{\epsilon}$ are usually assumed to be normally distributed:

$$\mathbf{y}_k - \mathcal{H}\left(\mathbf{x}_k^{\text{true}}\right) = \boldsymbol{\epsilon}_k \sim \mathcal{N}\left(\mathbf{0},\, \mathbf{R}_k\right) \in \mathbb{R}^{m\times 1}, \tag{3}$$

where m is the number of observed components, $\mathcal{H} : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ 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:

$$\begin{aligned} \mathbf{x}_k^{\text{a}} &= \mathbf{x}_k^{\text{b}} + \mathbf{B}_k \cdot \mathbf{H}_k^T \cdot \left[\mathbf{H}_k \cdot \mathbf{B}_k \cdot \mathbf{H}_k^T + \mathbf{R}_k\right]^{-1} \cdot \left[\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^{\text{b}})\right] \in \mathbb{R}^{n\times 1}, \\ \mathbf{A}_k &= \left[\mathbf{I} - \mathbf{B}_k \cdot \mathbf{H}_k^T \cdot \left[\mathbf{R}_k + \mathbf{H}_k \cdot \mathbf{B}_k \cdot \mathbf{H}_k^T\right]^{-1} \cdot \mathbf{H}_k\right] \cdot \mathbf{B}_k \in \mathbb{R}^{n\times n}, \end{aligned} \tag{4}$$

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 $\mathbf{x}_k^{\text{a}} \in \mathbb{R}^{n\times 1}$ is the analysis state. Typically, in the context of Kalman filtering, the observational operator is assumed linear [20]. Nonlinear observation operators are treated either by linearization, $\mathbf{H}_k \approx \mathcal{H}'(\mathbf{x}_k^{\text{b}})$, 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 (4) the elements of $\mathbf{B}_k$ determine how the information about the observed model components contained in the innovations $\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^{\text{b}}) \in \mathbb{R}^{m\times 1}$ 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,

$$\mathbf{X}_k^{\text{b}} = \left[\mathbf{x}_k^{\text{b}[1]}, \mathbf{x}_k^{\text{b}[2]}, \ldots, \mathbf{x}_k^{\text{b}[\text{N}_{\text{ens}}]}\right] \in \mathbb{R}^{n\times \text{N}_{\text{ens}}}, \tag{5}$$

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

$$\overline{\mathbf{x}}_k^{\text{b}} = \frac{1}{\text{N}_{\text{ens}}} \cdot \sum_{i=1}^{\text{N}_{\text{ens}}} \mathbf{x}_k^{\text{b}[i]} \in \mathbb{R}^{n\times 1}, \quad \mathbf{B}_k \approx \mathbf{P}^{\text{b}} = \frac{1}{\text{N}_{\text{ens}} - 1} \cdot \mathbf{U}_k^{\text{b}} \cdot \left(\mathbf{U}_k^{\text{b}}\right)^T \in \mathbb{R}^{n\times n}, \tag{6a}$$

where N_{ens} is the number of ensemble members, $\mathbf{x}_k^{\text{b}[i]} \in \mathbb{R}^{n\times 1}$ is the ith ensemble member, $\overline{\mathbf{x}}_k^{\text{b}} \in \mathbb{R}^{n\times 1}$ is the background ensemble mean, $\mathbf{P}_k^{\text{b}}$ is the background ensemble covariance matrix, and $\mathbf{U}_k \in \mathbb{R}^{n\times \text{N}_{\text{ens}}}$ is the matrix of member deviations:

$$\mathbf{U}_k^{\text{b}} = \mathbf{X}_k^{\text{b}} - \overline{\mathbf{x}}_k^{\text{b}} \cdot \mathbf{1}_{\text{N}_{\text{ens}}}^T \in \mathbb{R}^{n\times \text{N}_{\text{ens}}}. \tag{6b}$$

One attractive feature of $\mathbf{P}_k^{\text{b}}$ is its flow dependency which allows one 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, $n \gg \text{N}_{\text{ens}}$, and, therefore, $\mathbf{P}_k^{\text{b}}$ 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 ensemble Kalman filter (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 $\mathbf{B}^{-1}$ obtained via a modified Cholesky decomposition. By imposing conditional independence between errors in remote model components we obtain sparse approximations of $\mathbf{B}^{-1}$.

This paper is organized as follows. In section 2 ensemble-based methods and the modified Cholesky decomposition are introduced. Section 3 discusses the proposed EnKF 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 LETKF. Conclusions are drawn in section 5.

2. Background. The EnKF is a sequential Monte Carlo method for state and parameter estimation of nonlinear models such as those found in atmospheric and oceanic sciences [9, 29]. The EnKF popularity is due to its basic theoretical formulation and its relative ease of implementation [9]. Given the *background* ensemble (5) EnKF builds the *analysis* ensemble as follows:

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{b}} + \mathbf{P}^{\mathrm{b}} \cdot \mathbf{H}^{T} \cdot \left[\mathbf{R} + \mathbf{H} \cdot \mathbf{P}^{\mathrm{b}} \cdot \mathbf{H}^{T}\right] \cdot \boldsymbol{\Delta} \in \mathbb{R}^{n \times \mathrm{N_{ens}}}, \tag{7a}$$

where

$$\boldsymbol{\Delta} = \mathbf{Y}^{\mathrm{s}} - \mathbf{H} \cdot \mathbf{X}^{\mathrm{b}} \in \mathbb{R}^{m \times \mathrm{N_{ens}}}, \tag{7b}$$

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

$$\mathbf{Y}^{\mathrm{s}} = \left[\mathbf{y} + \boldsymbol{\epsilon}^{[1]}, \mathbf{y} + \boldsymbol{\epsilon}^{[2]}, \ldots, \mathbf{y} + \boldsymbol{\epsilon}^{[\mathrm{N_{ens}}]}\right] \in \mathbb{R}^{m \times \mathrm{N_{ens}}}, \quad \boldsymbol{\epsilon}^{[i]} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right), \tag{7c}$$

for $1 \leq i \leq \mathrm{N_{ens}}$. 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 [31]. However, it also has been shown that the inclusion of perturbed observations introduces sampling errors in the assimilation [2, 15].

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 $\mathbf{B}$ is the Hollingworth and Lonnberg method [12] 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, with an inverse background error covariance matrix, have been proposed in order to exploit features of the precision matrices [30]. 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 hour 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 [18]. In this context, the model learns and changes with regard to the data collected (i.e., parameter values and model structures). This allows one 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 [19]. 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 sequential data assimilation [24]. We proposed approximations based on autoregressive error models [8] and using hybrid subspace techniques.[7].

Covariance matrix localization artificially reduces correlations between distant model components via a Schur product with a localization matrix $\mathbf{\Pi} \in \mathbb{R}^{n\times n}$ [28]:

$$\widehat{\mathbf{P}}^{\mathrm{b}} = \mathbf{\Pi} \circ \mathbf{P}^{\mathrm{b}} \in \mathbb{R}^{n\times n}\,, \tag{8}$$

and then $\mathbf{P}^{\mathrm{b}}$ is replaced by $\widehat{\mathbf{P}}^{\mathrm{b}} \in \mathbb{R}^{n\times n}$ in the EnKF analysis equation (7a). For instance, the entries of $\mathbf{\Pi}$ can decrease with the distance between model components depending on the radius of influence ζ:

$$\{\mathbf{\Pi}\}_{i,j} = \exp\left(-\frac{\pi\,(m_i,\, m_j)}{2\cdot\zeta^2}\right) \text{ for } 1 \le i \le j \le n\,, \tag{9}$$

where $\pi\,(m_i,\, m_j)$ represents the physical distance squared between the model components m_i and m_j. The exponential decay allows one to reduce the impact of innovations between distant model components. The use of covariance matrix localization alleviates the impact of sampling errors. However, the explicit computation of $\mathbf{\Pi}$ (and even $\mathbf{P}^{\mathrm{b}}$) is prohibitive owing to numerical model dimensions. Thus, domain localization methods [6, 16] are commonly used in the context of operational data assimilation. One of the best EnKF implementations based on domain localization is the LETKF [27]. In the LETKF the analysis increments are computed in the space spanned by the ensemble perturbations $\mathbf{U}^{\mathrm{b}}$ defined in (6b). An approximation of the analysis covariance matrix in this space reads

$$\widehat{\mathbf{P}}^{\mathrm{a}} = \left[(\mathrm{N_{ens}}-1)\cdot\mathbf{I} + \mathbf{Q}^T\cdot\mathbf{R}^{-1}\cdot\mathbf{Q}\right]^{-1} \in \mathbb{R}^{\mathrm{N_{ens}}\times\mathrm{N_{ens}}}, \tag{10a}$$

where $\mathbf{Q} = \mathbf{H}\cdot\mathbf{U}^{\mathrm{b}} \in \mathbb{R}^{m\times\mathrm{N_{ens}}}$ and $\mathbf{I}$ is the identity matrix consistent with the dimension. The analysis increments in the subspace are:

$$\boldsymbol{\alpha}^{\mathrm{a}} = \widehat{\mathbf{P}}^{\mathrm{a}}\cdot\mathbf{Q}^T\cdot\mathbf{R}^{-1}\cdot\left[\mathbf{y} - \mathbf{H}\cdot\overline{\mathbf{x}}^{\mathrm{b}}\right] \in \mathbb{R}^{\mathrm{N_{ens}}\times 1} \tag{10b}$$

from which an estimation of the analysis mean in the model space can be obtained:

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{b}} + \mathbf{U}^{\mathrm{b}}\cdot\boldsymbol{\alpha}^{\mathrm{a}} \in \mathbb{R}^{n\times 1}. \tag{10c}$$

Finally, the analysis ensemble reads

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

The domain localization in the LETKF is performed as follows: each model component is surrounded by a local box of radius ζ. 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 ζ are considered during the assimilation step. However, for large values of ζ, 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) $\mathrm{N_{ens}}$ Gaussian random vectors (6b)

$$\mathbf{U}^{b} = \left[\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[\mathrm{N_{ens}}]}\right] \in \mathbb{R}^{n \times \mathrm{N_{ens}}}$$

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

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

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

$$\boldsymbol{\beta}_{[i]} = \arg\min_{\boldsymbol{\beta}} \left\| \mathbf{x}_{[i]} - \mathbf{Z}_{[i]} \cdot \boldsymbol{\beta} \right\|_2^2, \tag{12}$$

where

$$\mathbf{Z}_{[i]} = \left[\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \ldots, \mathbf{x}^{[i-1]}\right]^{T} \in \mathbb{R}^{(i-1) \times \mathrm{N_{ens}}}, \quad 2 \leq i \leq n,$$
$$\boldsymbol{\beta}_{[i]} = \left[\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,i-1}\right]^{T} \in \mathbb{R}^{(i-1) \times 1}.$$

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

$$\boldsymbol{\beta}^{[i]} = \left[\mathbf{Z}^{[i]} \cdot \mathbf{Z}^{[i]\,T}\right]^{-1} \cdot \mathbf{Z}^{[i]} \cdot \mathbf{x}_{[i]}\,. \tag{13}$$

The regression coefficients form the lower triangular matrix

$$\left\{\widehat{\mathbf{T}}\right\}_{i,j} = \begin{cases} -\beta_{i,j} \text{ for } 1 \leq j < i, \\ \quad 1 \text{ for } j = i, \\ \quad 0 \text{ for } j > i, \end{cases} \qquad 1 \leq i \leq n, \tag{14a}$$

where $\{\widehat{\mathbf{T}}\}_{i,j}$ denotes the (i,j)th component of matrix $\widehat{\mathbf{T}} \in \mathbb{R}^{n\times n}$. The empirical variances $\widehat{\mathbf{cov}}$ of the residuals $\boldsymbol{\varepsilon}_{[i]}$ form the diagonal matrix

$$\widehat{\mathbf{D}} = \underset{1\le i\le n}{\operatorname{diag}} \left(\widehat{\mathbf{cov}}(\boldsymbol{\varepsilon}_{[i]})\right) = \underset{1\le i\le n}{\operatorname{diag}} \left(\frac{1}{\mathrm{N_{ens}}-1} \sum_{j=1}^{\mathrm{N_{ens}}} \left\{\boldsymbol{\varepsilon}_{[i]}\right\}_j^2 \right) \in \mathbb{R}^{n\times n}, \tag{14b}$$

where $\{\widehat{\mathbf{D}}\}_{1,1} = \widehat{\mathbf{cov}}\left(\mathbf{x}_{[1]}\right)$. Then an estimate of $\mathbf{B}^{-1}$ can be computed as follows:

$$\widehat{\mathbf{B}}^{-1} = \widehat{\mathbf{T}}^T \cdot \widehat{\mathbf{D}}^{-1} \cdot \widehat{\mathbf{T}} \in \mathbb{R}^{n\times n} \tag{15a}$$

or, by basic matrix algebra identities, the estimate of $\mathbf{B}$ reads

$$\widehat{\mathbf{B}} = \widehat{\mathbf{T}}^{-1} \cdot \widehat{\mathbf{D}} \cdot \widehat{\mathbf{T}}^{-T} \in \mathbb{R}^{n\times n}. \tag{15b}$$

Note that the structure of $\widehat{\mathbf{B}}^{-1}$ is strictly related to the structure of $\widehat{\mathbf{T}}$. This can be exploited in order to obtain sparse estimators of $\mathbf{B}^{-1}$ by imposing that some entries of $\widehat{\mathbf{T}}$ are zero. This is important for high dimensional probability distributions where the explicit computation of $\widehat{\mathbf{B}}$ or $\widehat{\mathbf{B}}^{-1}$ is prohibitive. The zero components in $\widehat{\mathbf{T}}$ can be justified as follows: when two model components are *conditionally independent* for a given radius of influence ζ their corresponding entry in $\widehat{\mathbf{B}}^{-1}$ 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 neighborhoods [30]. For a given model component, its neighborhood is formed by components within the scope of some radius of influence ζ. We can consider zero correlations between background errors corresponding to model components located at distances that exceed ζ. In the next section we present an EnKF implementation based on modified Cholesky decomposition for inverse covariance matrix estimation.

3. EnKF based on modified Cholesky decomposition. In this section we discuss the new EnKF 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 $\widehat{\mathbf{B}}^{-1}$ depends on that of $\widehat{\mathbf{T}}$. If we assume that the correlations between model components are local, and there are no correlations outside a radius of influence ζ, we obtain lower-triangular sparse estimators of $\widehat{\mathbf{T}}$. Consequently, the resulting $\widehat{\mathbf{B}}^{-1}$ will also be sparse, and $\widehat{\mathbf{B}}$ 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 $\widehat{\mathbf{T}}$. Since we work with gridded models we consider column-major and row-major orders. Figure 1 shows the neighborhood and the predecessors of the model component 6 when column-major order is utilized.

The estimation of $\widehat{\mathbf{B}}^{-1}$ proceeds as follows:

1. For the ith model component, making use of the truncated singular value decomposition [14, 13, 11], estimate the regression coefficients $\beta_{i,j}$ for $1 \le i \le n$ and $j \in P(i,\zeta)$, that satisfies

$$\mathbf{x}_{[i]} = \sum_{j\in P(i,\zeta)} \beta_{i,j} \cdot \mathbf{x}^{[j]} + \boldsymbol{\varepsilon}_{[i]} \in \mathbb{R}^{\mathrm{N_{ens}}\times 1}, \tag{16}$$

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

(a) In blue, neihgborhood for the model component 6 when $\zeta = 1$.

(b) In blue, predecessors of the model component 6 for $\zeta = 1$.

FIG. 1. *Neighborhood and local predecessors within the scope of* $\zeta = 1$ *for the model component* 6. *Column-major ordering is utilized to label the model components.*

2. Build the matrices

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

and $\widehat{\mathbf{D}}$ according to (14b). Note that the number of nonzero elements in the ith row of $\widehat{\mathbf{T}}$ 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 $\widehat{\mathbf{B}}^{-1}$ is estimated, EnKF-MC computes the analysis using Kalman's formula

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{A}} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{\Delta} \in \mathbb{R}^{n \times \mathrm{N_{ens}}}, \tag{17a}$$

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

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

and $\mathbf{\Delta} \in \mathbb{R}^{m \times \mathrm{N_{ens}}}$ 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:

$$\mathbf{x}^{\mathrm{a}} = \widehat{\mathbf{A}} \cdot \left[\widehat{\mathbf{B}}^{-1} \cdot \mathbf{x}^{\mathrm{b}} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Y}^{\mathrm{s}}\right] \in \mathbb{R}^{n \times \mathrm{N_{ens}}} \tag{17b}$$

$$= \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{T}}^{-1} \cdot \widehat{\mathbf{D}}^{1/2} \cdot \mathbf{V}_{\widehat{\mathbf{B}}}^T \cdot \left[\mathbf{R} + \mathbf{V}_{\widehat{\mathbf{B}}} \cdot \mathbf{V}_{\widehat{\mathbf{B}}}^T\right]^{-1} \cdot \mathbf{\Delta}\,, \tag{17c}$$

where $\mathbf{V}_{\widehat{\mathbf{B}}} = \mathbf{H} \cdot \widehat{\mathbf{T}}^{-1} \cdot \widehat{\mathbf{D}}^{1/2} \in \mathbb{R}^{n \times m}$, and $\mathbf{Y}^{\mathrm{s}}$ are the perturbed observations. The formulation (17c) is well known as the EnKF dual formulation (17b) is known as the EnKF primal formulation, and (17a) is the incremental form of the primal formulation. It can be easily shown that, by making use of the iterative Sherman–Morrison formula [23], the computational effort of the method is bounded by

$$\mathcal{O}\left(m \cdot n + m \cdot n \cdot \mathrm{N_{ens}}\right).$$

3.3. Convergence of the covariance inverse estimator. In this section we prove the convergence of the $\widehat{\mathbf{B}}^{-1}$ 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* $\widehat{\mathbf{T}}$ *and* $\widehat{\mathbf{D}}$.

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

$$\widehat{\mathbf{B}}^{-1} = \left[\mathbf{I} - \widehat{\mathbf{C}}\right]^T \cdot \widehat{\mathbf{D}}^{-1} \cdot \left[\mathbf{I} - \widehat{\mathbf{C}}\right] \in \mathbb{R}^{n\times n} \tag{18a}$$

and

$$\mathbf{B}^{-1} = [\mathbf{I} - \mathbf{C}]^T \cdot \mathbf{D}^{-1} \cdot [\mathbf{I} - \mathbf{C}] \in \mathbb{R}^{n\times n}, \tag{18b}$$

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

$$\mathbf{D} = \mathbf{diag}\left\{d_1^2, d_2^2, \ldots, d_n^2\right\}, \quad \widehat{\mathbf{D}} = \mathbf{diag}\left\{\widehat{d}_1^2, \widehat{d}_2^2, \ldots, \widehat{d}_n^2\right\},$$

where $\{\mathbf{D}\}_{i,i} = d_i^2$ and $\{\widehat{\mathbf{D}}\}_{i,i} = \widehat{d}_i^2$ for $1 \le i \le n$. In what follows we denote by $\widehat{\mathbf{c}}^{\{j\}} \in \mathbb{R}^{n\times 1}$ and $\mathbf{c}^{\{j\}} \in \mathbb{R}^{n\times 1}$ the jth columns of matrices $\widehat{\mathbf{C}}$ and $\mathbf{C}$, respectively, for $1 \le j \le n$.

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

$$\mathcal{U}^{-1}(\varepsilon_0, C, \alpha) = \Bigg\{ \mathbf{B} : 0 < \varepsilon_0 \le \lambda_{min}(\mathbf{B}) \le \lambda_{max}(\mathbf{B}) \le \varepsilon_0^{-1}, \max_k \sum_{j\notin P(k,\zeta)} \left|\{\mathbf{T}\}_{k,j}\right| \le C\cdot\zeta^{-\alpha} \Bigg\}, \tag{19a}$$

where $\mathbf{B}^{-1} = \mathbf{T}^T\mathbf{D}^{-1}\mathbf{T}$, α *is the decay rate (related to the dynamics of the numerical model), and* $P(k,\zeta)$, *for a given radius* ζ, *denotes the row indexes of predecessors for model component* k.

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

$$\left\|\widehat{\mathbf{B}}^{-1} - \mathbf{B}^{-1}\right\|_\infty = \mathcal{O}\left(\left[\frac{\log(n)}{\mathrm{N_{ens}}}\right]^{\alpha(\alpha+1)/2}\right), \tag{19b}$$

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* $\mathcal{U}^{-1}$

$$\max\left\{\left\|\widehat{\mathbf{c}}^{\{j\}} - \mathbf{c}^{\{j\}}\right\|_\infty : 1 \le j \le n\right\} = \mathcal{O}\left({\mathrm{N_{ens}}}^{-1/2}\log^{1/2} n\right), \tag{20a}$$

$$\max\left\{\left|\widehat{d}_j^2 - d_j^2\right| : 1 \le j \le n\right\} = \mathcal{O}\left(\left[{\mathrm{N_{ens}}}^{-1}\log n\right]^{\alpha/(2(\alpha+1))}\right), \tag{20b}$$

$$\left\|\mathbf{C}\right\|_\infty = \mathcal{O}(1), \text{ and } \left\|\mathbf{D}^{-1}\right\|_\infty = \mathcal{O}(1). \tag{20c}$$

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

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

$$\begin{aligned} \text{(21)} \qquad & \mathrm{Prob}\left[\sum_{k=1}^{\mathrm{N_{ens}}}\left[\left\{\boldsymbol{\nu}^{[k]}\right\}_i \cdot \left\{\boldsymbol{\nu}^{[k]}\right\}_j - \{\mathbf{B}\}_{i,j}\right] \geq \mathrm{N_{ens}} \cdot \nu\right] \\ & \leq C_1 \cdot \exp\left(-C_2 \cdot \mathrm{N_{ens}} \cdot \nu^2\right) \end{aligned}$$

for $|\nu| \leq \delta$, *where* $\left\{\boldsymbol{\nu}^{[k]}\right\}_i$ *is the* i*th component of the sample* $\boldsymbol{\nu}^{[k]}$ *for* $1 \leq k \leq \mathrm{N_{ens}}$ *and* $1 \leq i \leq n$. *Likewise,* C_1, C_2, *and* δ *depend on* ε_0 *only.*

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

Recall that

$$\widehat{\mathbf{cov}}\left(\mathbf{U}^{\mathrm{b}}\right) = \mathbf{P}^{\mathrm{b}} = \frac{1}{\mathrm{N_{ens}}-1} \cdot \mathbf{U}^{\mathrm{b}} \cdot \mathbf{U}^{\mathrm{b}T} = \frac{1}{\mathrm{N_{ens}}-1} \cdot \sum_{k=1}^{\mathrm{N_{ens}}} \mathbf{u}^{\mathrm{b}[k]} \cdot \mathbf{u}^{\mathrm{b}[k]T}$$

$$\Leftrightarrow \left\{\widehat{\mathbf{cov}}\left(\mathbf{U}^{\mathrm{b}}\right)\right\}_{i,j} = \frac{1}{\mathrm{N_{ens}}-1} \cdot \sum_{k=1}^{\mathrm{N_{ens}}} \left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_i \cdot \left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_j.$$

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

$$\text{(22a)} \qquad \left\|\mathbf{cov}\left(\mathbf{U}^{\mathrm{b}}\right) - \widehat{\mathbf{cov}}\left(\mathbf{U}^{\mathrm{b}}\right)\right\|_\infty = \mathcal{O}\left(\mathrm{N_{ens}}^{-1/2} \cdot \log^{1/2} n\right)$$

since the entries of $\mathbf{cov}\left(\mathbf{U}^{\mathrm{b}}\right) - \widehat{\mathbf{cov}}\left(\mathbf{U}^{\mathrm{b}}\right)$ can be bounded by

$$\left|\left\{\mathbf{cov}\left(\mathbf{U}^{\mathrm{b}}\right) - \widehat{\mathbf{cov}}\left(\mathbf{U}^{\mathrm{b}}\right)\right\}_{i,j}\right| \leq \mathrm{N_{ens}}^{-1} \cdot \sum_{k=1}^{\mathrm{N_{ens}}} \left|\left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_i \cdot \left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_j - \{\mathbf{B}\}_{i,j}\right|.$$

Lemma 2 ensures that

$$\begin{aligned} & \mathrm{Prob}\left[\max_{i,j}\left|\mathrm{N_{ens}}^{-1} \cdot \sum_{k=1}^{\mathrm{N_{ens}}} \left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_i \cdot \left\{\mathbf{u}^{\mathrm{b}[k]}\right\}_j - \{\mathbf{B}\}_{i,j}\right| \geq \nu\right] \\ & \leq C_1 \cdot n^2 \cdot \exp\left(-C_2 \cdot \mathrm{N_{ens}} \cdot \nu^2\right) \end{aligned}$$

for $|\nu| \leq \delta$. Let $\nu = \left(\frac{\log n^2}{\mathrm{N_{ens}} \cdot C_2}\right)^{1/2} \cdot M$ for M arbitrary.

Since $\mathbf{Z}_{[i]}$ stores the columns of $\mathbf{U}^{\mathrm{b}}$ corresponding to the predecessors of model component i, an immediate consequence of (22a) is

$$\text{(22b)} \qquad \max_i \left\|\mathbf{cov}\left(\mathbf{Z}_{[i]}\right) - \widehat{\mathbf{cov}}\left(\mathbf{Z}_{[i]}\right)\right\|_\infty = \mathcal{O}\left(\mathrm{N_{ens}}^{-1/2} \cdot \log^{1/2} n\right).$$

Also,

$$\left\|\mathbf{B}^{-1}\right\|_\infty = \left\|\mathbf{cov}\left(\mathbf{U}^{\mathrm{b}}\right)^{-1}\right\|_\infty \leq \varepsilon_0^{-1}.$$

According to (12), we have,

$$\left\{\mathbf{c}^{[i]}\right\}_j = \left\{\mathbf{cov}\left(\mathbf{Z}_{[i]}\right)^{-1}\mathbf{Z}_{[i]}\cdot\mathbf{x}_{[i]}\right\}_j, \quad \left\{\widehat{\mathbf{c}}^{[i]}\right\}_j = \left\{\widehat{\mathbf{cov}}\left(\mathbf{Z}_{[i]}\right)^{-1}\mathbf{Z}_{[i]}\cdot\mathbf{x}_{[i]}\right\}_j;$$

therefore,

$$\begin{aligned}
&\max_k \left|\left\{\mathbf{c}^{[i]}\right\}_k - \left\{\widehat{\mathbf{c}}^{[i]}\right\}_k\right| \\
&= \max_k \left|\left\{\mathbf{cov}\left(\mathbf{Z}_{[i]}\right)^{-1}\mathbf{Z}_{[i]}\mathbf{x}_{[i]}\right\}_k - \left\{\widehat{\mathbf{cov}}\left(\mathbf{Z}_{[i]}\right)^{-1}\cdot\mathbf{Z}_{[i]}\mathbf{x}_{[i]}\right\}_k\right| \qquad (23)\\
&= \max_k \left|\left\{\left[\mathbf{cov}\left(\mathbf{Z}_{[i]}\right)^{-1} - \widehat{\mathbf{cov}}\left(\mathbf{Z}_{[i]}\right)^{-1}\right]\cdot\mathbf{Z}_{[i]}\mathbf{x}_{[i]}\right\}_k\right| \\
&= \mathcal{O}\left(\mathrm{N_{ens}}^{-1/2}\cdot\log^{1/2} n\right), \qquad (24)
\end{aligned}$$

from which (20a) follows. Note that

$$\begin{aligned}
&\mathbf{x}_{[i]} = \sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \mathbf{x}_{[j]} + \widehat{\boldsymbol{\varepsilon}}^{[i]} \Leftrightarrow \widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right) = \widehat{\mathbf{cov}}\left(\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \mathbf{x}_{[j]} + \widehat{\boldsymbol{\varepsilon}}^{[i]}\right) \\
&\Leftrightarrow \widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right) = \widehat{\mathbf{cov}}\left(\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \cdot\mathbf{x}_{[j]}\right) + \widehat{\mathbf{cov}}\left(\widehat{\boldsymbol{\varepsilon}}^{[i]}\right) \\
&\Leftrightarrow \widehat{d}_i^2 = \widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right) - \widehat{\mathbf{cov}}\left(\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \cdot\mathbf{x}_{[j]}\right)
\end{aligned}$$

and, similarly,

$$d_i^2 = \mathbf{cov}\left(\mathbf{x}_{[i]}\right) - \mathbf{cov}\left(\sum_{j\in P(i,\zeta)}\left\{\mathbf{c}^{[i]}\right\}_j \cdot\mathbf{x}_{[j]}\right).$$

The claim (20b) and the first part of (20c) follow from (22a), (22b), and (24). We have

$$\begin{aligned}
\left|\widehat{d}_i^2 - d_i^2\right| &\le \left|\mathbf{cov}\left(\mathbf{x}_{[i]}\right) - \widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right)\right| \\
&+ \left|\widehat{\mathbf{cov}}\left(\sum_{j\in P(i,\zeta)}\left[\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j - \left\{\mathbf{c}^{[i]}\right\}_j\right]\cdot\mathbf{x}_{[j]}\right)\right| \qquad (25)\\
&+ \left|\widehat{\mathbf{cov}}\left(\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \cdot\mathbf{x}_{[j]}\right) - \mathbf{cov}\left(\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j \cdot\mathbf{x}_{[j]}\right)\right|.
\end{aligned}$$

By Lemma 2 the maximum over i of the first term is

$$\max_i \left|\mathbf{cov}\left(\mathbf{x}_{[i]}\right) - \widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right)\right| = \mathcal{O}\left(\mathrm{N_{ens}}^{-1/2}\cdot\log^{1/2} n\right).$$

The second term can be bounded as follows:

$$\left|\sum_{j\in P(i,\zeta)}\left[\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j-\left\{\mathbf{c}^{[i]}\right\}_j\right]^2\widehat{\mathbf{cov}}\left(\mathbf{x}_{[j]}\right)\right|\le\sum_{j\in P(i,\zeta)}\left[\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j-\left\{\mathbf{c}^{[i]}\right\}_j\right]^2\left|\widehat{\mathbf{cov}}\left(\mathbf{x}_{[j]}\right)\right|$$
$$\le\max_k\left[\left\{\widehat{\mathbf{c}}^{[i]}\right\}_k-\left\{\mathbf{c}^{[i]}\right\}_k\right]^2\cdot\max_i\left|\widehat{\mathbf{cov}}\left(\mathbf{x}_{[i]}\right)\right|=\mathcal{O}\left(\zeta^2\cdot{\mathrm{N_{ens}}}^{-1}\cdot\log n\right)$$
$$=\mathcal{O}\left(\left[{\mathrm{N_{ens}}}^{-1}\cdot\log n\right]^{\alpha/2\cdot(\alpha+1)}\right)$$

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

$$d_i^2=\mathbf{cov}\left(\mathbf{x}_{[i]}-\sum_{j\in P(i,\zeta)}\left\{\widehat{\mathbf{c}}^{[i]}\right\}_j\cdot\mathbf{x}_{[j]}\right)\ge\varepsilon_0\cdot\left(1+\sum_{j\in P(i,\zeta)}\left[\widehat{\mathbf{c}}_j^{[i]}\right]^2\right)\ge\varepsilon_0\,,$$

and the lemma follows. ∎

Comment 2 (Gaussian assumption relaxation). *The Lemma* 2 *here, originally from Bickel and Levina in* [5]*, is based on the assumption that samples follow a Gaussian distribution with covariance matrix* $\mathbf{B}$ *such that the sample covariance component has the exponential decay property* (21)*. However, the Gaussian assumption can be relaxed to other distributions whose second moment satisfies such a decay condition (even may be slower than exponential decay) 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.*

We now are ready to prove Theorem 1.

Proof of Theorem 1. We need only check that

$$\left\|\widehat{\mathbf{B}}^{-1}-\mathbf{B}^{-1}\right\|_\infty=\mathcal{O}\left({\mathrm{N_{ens}}}^{-1/2}\cdot\log^{1/2}(n)\right)\tag{26a}$$

and

$$\left\|\mathbf{B}^{-1}-\Phi_\zeta\left(\mathbf{B}^{-1}\right)\right\|_\infty=\mathcal{O}\left(\zeta^{-\alpha}\right)\,,\tag{26b}$$

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

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

We first prove (26a). By definition,

$$\widehat{\mathbf{B}}^{-1}-\mathbf{B}^{-1}=\widehat{\mathbf{T}}^T\cdot\widehat{\mathbf{D}}^{-1}\cdot\widehat{\mathbf{T}}-\mathbf{T}^T\cdot\mathbf{D}^{-1}\cdot\mathbf{T}.\tag{27}$$

Applying the standard inequality

$$\begin{aligned}\left\|\mathbf{T}^T\cdot\mathbf{D}^{-1}\cdot\mathbf{T}-\widehat{\mathbf{T}}^T\cdot\widehat{\mathbf{D}}^{-1}\cdot\widehat{\mathbf{T}}^T\right\|\le{}&\left\|\mathbf{T}^T-\widehat{\mathbf{T}}^T\right\|\cdot\left\|\widehat{\mathbf{D}}\right\|\cdot\left\|\widehat{\mathbf{T}}\right\|\\&+\left\|\mathbf{D}-\widehat{\mathbf{D}}\right\|\cdot\left\|\widehat{\mathbf{T}}^T\right\|\cdot\left\|\widehat{\mathbf{T}}\right\|\\&+\left\|\mathbf{T}-\widehat{\mathbf{T}}\right\|\cdot\left\|\widehat{\mathbf{T}}\right\|\cdot\left\|\widehat{\mathbf{D}}\right\|\\&+\left\|\widehat{\mathbf{T}}\right\|\cdot\left\|\mathbf{D}-\widehat{\mathbf{D}}\right\|\cdot\left\|\widehat{\mathbf{T}}^T-\mathbf{T}^T\right\|\\&+\left\|\widehat{\mathbf{D}}\right\|\cdot\left\|\mathbf{T}-\widehat{\mathbf{T}}\right\|\cdot\left\|\widehat{\mathbf{T}}^T-\mathbf{T}^T\right\|\\&+\left\|\widehat{\mathbf{T}}^T\right\|\cdot\left\|\mathbf{D}-\widehat{\mathbf{D}}\right\|\cdot\left\|\widehat{\mathbf{T}}-\mathbf{T}\right\|\\&+\left\|\mathbf{D}-\widehat{\mathbf{D}}\right\|\cdot\left\|\mathbf{T}-\widehat{\mathbf{T}}\right\|\cdot\left\|\mathbf{T}^T-\widehat{\mathbf{T}}^T\right\|,\end{aligned}$$

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 $\mathbf{M}$,

$$\left\|\mathbf{M}\cdot\mathbf{M}^T-\Phi_\zeta\left(\mathbf{M}\right)\cdot\Phi_\zeta\left(\mathbf{M}\right)^T\right\|_\infty\leq 2\left\|\mathbf{M}\right\|_\infty\left\|\Phi_\zeta\left(\mathbf{M}\right)-\mathbf{M}\right\|_\infty+\left\|\Phi_\zeta\left(\mathbf{M}\right)-\mathbf{M}\right\|_\infty^2,$$

and by letting $\mathbf{M}=\mathbf{T}^T\cdot\mathbf{D}^{-1/2}$, 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 SPEEDY. Given its simple formulation and low computational cost, the Lorenz-96 model allows one 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 Ott et al. [25, 26, 27]. As a measure of accuracy, the L_2 norm of the analysis error is computed at each assimilation step:

$$\lambda_i=\left\|\mathbf{x}_i^{\text{ref}}-\mathbf{x}_i^a\right\|_2\quad\text{for } 1\leq i\leq M\,, \tag{28a}$$

where M denotes the number of observation times within the assimilation window, $\mathbf{x}_i^{\text{ref}}$ and $\mathbf{x}_i^a$ are the reference solution and the analysis states at time i, respectively. The root mean square error (RMSE) across all times is used to assess the average accuracy of a filter solution throughout the entire assimilation window:

$$\text{RMSE}=\sqrt{\frac{1}{M}\cdot\sum_{k=1}^{M}\lambda_i^2}\,. \tag{28b}$$

The threshold used in the truncated singular value decomposition during the computation of $\widehat{\mathbf{B}}^{-1}$ is 0.10. During the assimilation steps, the data error covariance matrices $\mathbf{R}_k$ are used (no representative 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 [21] is described by the following set of ordinary differential equations:

$$\frac{dx_j}{dt}=\begin{cases}(x_2-x_{n-1})\cdot x_n-x_1+F & \text{for } j=1,\\ (x_{j+1}-x_{j-2})\cdot x_{j-1}-x_j+F & \text{for } 2\leq j\leq n-1,\\ (x_1-x_{n-2})\cdot x_{n-1}-x_n+F & \text{for } j=n,\end{cases} \tag{29}$$

where F is the 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 $0.05\cdot\mathbf{I}\in\mathbb{R}^{n\times n}$, where $\mathbf{I}$ 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 $0.01 \cdot \mathbf{I} \in \mathbb{R}^{m \times m}$. The assimilation window consists of 25 observations evenly distributed in time. Two ensemble sizes are utilized during the tests: $\mathrm{N_{ens}} = 20$ and $\mathrm{N_{ens}} = 60$. 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, 10]. 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 [17, 22]. 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 (192×96 grids) is used for the horizontal space discretization of each layer. Four model variables are part of the assimilation process: the temperature (K), the zonal and the meridional wind components (m/s), and the specific humidity (g/kg). The total number of model components is $n = 589,824$. The number of ensemble members is $\mathrm{N_{ens}} = 94$ for all the scenarios. The model state space is approximately 6,274 times larger than the number of ensemble members ($n \gg \mathrm{N_{ens}}$).

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

$$\mathbf{x}^{\mathrm{ref}}_{-2} = \mathcal{M}_{t_{-3} \to t_{-2}} \left(\mathbf{x}^{\mathrm{ref}}_{-3} \right).$$

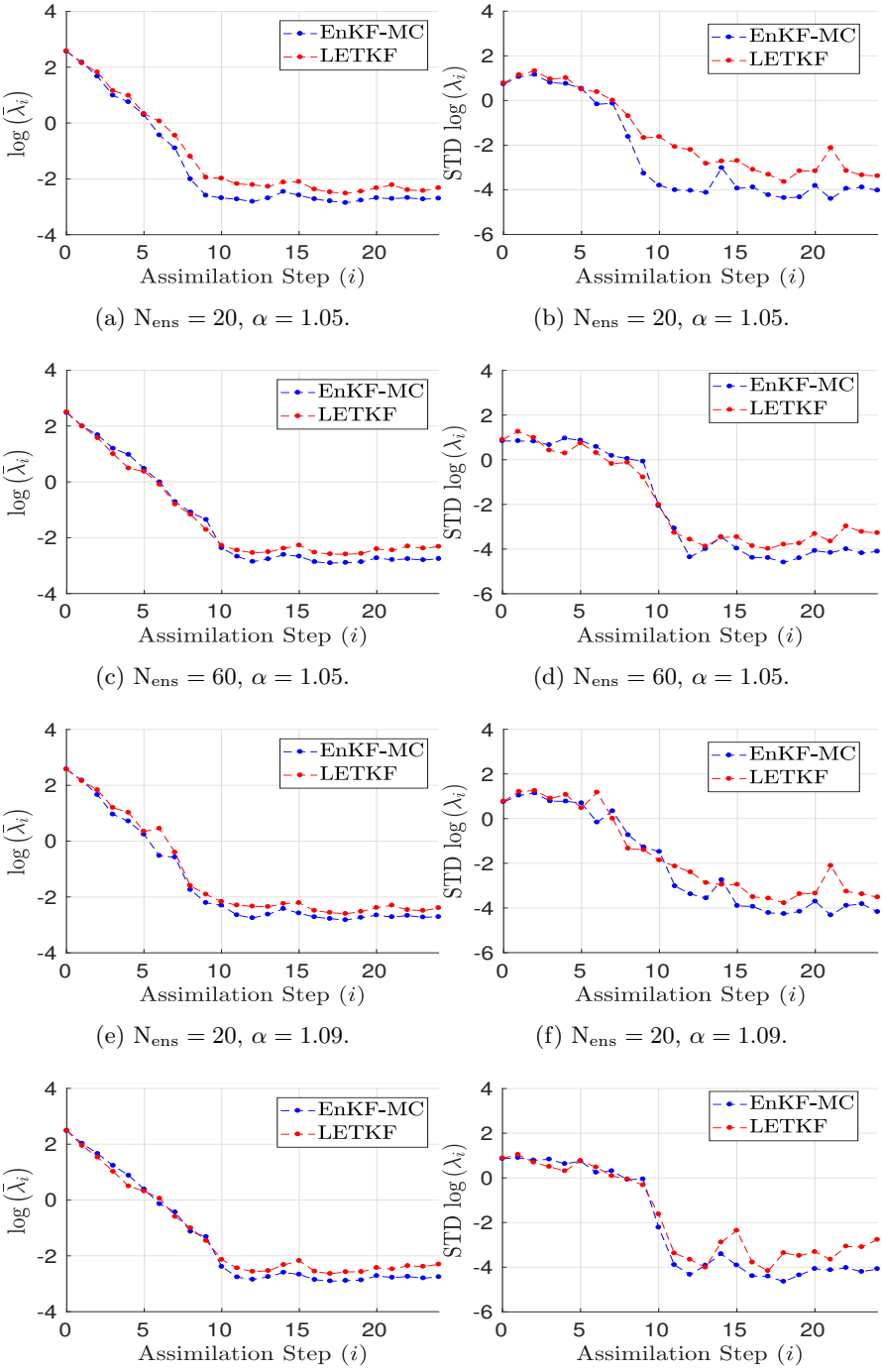
The reference solution $\mathbf{x}^{\mathrm{ref}}_{-2}$ is used to build a perturbed background solution:

$$\widehat{\mathbf{x}}^{\mathrm{b}}_{-2} = \mathbf{x}^{\mathrm{ref}}_{-2} + \boldsymbol{\epsilon}^{\mathrm{b}}_{-2}, \quad \boldsymbol{\epsilon}^{\mathrm{b}}_{-2} \sim \mathcal{N}\left(\mathbf{0}_n, \, \mathrm{diag}\left\{ (0.05 \, \{\mathbf{x}^{\mathrm{ref}}_{-2}\}_i)^2 \right\} \right). \tag{30}$$

The perturbed background solution is propagated over another year to obtain the background solution at time t_{-1}:

$$\mathbf{x}^{\mathrm{b}}_{-1} = \mathcal{M}_{t_{-2} \to t_{-1}} \left(\widehat{\mathbf{x}}^{\mathrm{b}}_{-2} \right). \tag{31}$$

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,



(a) $N_{ens} = 20$, $\alpha = 1.05$. (b) $N_{ens} = 20$, $\alpha = 1.05$.

(c) $N_{ens} = 60$, $\alpha = 1.05$. (d) $N_{ens} = 60$, $\alpha = 1.05$.

(e) $N_{ens} = 20$, $\alpha = 1.09$. (f) $N_{ens} = 20$, $\alpha = 1.09$.

(g) $N_{ens} = 60$, $\alpha = 1.09$. (h) $N_{ens} = 60$, $\alpha = 1.09$.

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 background state (31) is utilized in order to build an ensemble of perturbed background states:

$$\widehat{\mathbf{x}}_{-1}^{\mathrm{b}[i]} = \mathbf{x}_{-1}^{\mathrm{b}} + \boldsymbol{\epsilon}_{-1}^{\mathrm{b}}, \quad \boldsymbol{\epsilon}_{-1}^{\mathrm{b}} \sim \mathcal{N}\left(\mathbf{0}_n,\, \underset{i}{\mathrm{diag}}\left\{(0.05\,\{\mathbf{x}_{-1}^{\mathrm{b}}\}_i)^2\right\}\right), \quad 1 \le i \le \mathrm{N_{ens}}, \tag{32}$$

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

$$\mathbf{x}_0^{\mathrm{b}[i]} = \mathcal{M}_{t_{-1}\to t_0}\left(\widehat{\mathbf{x}}_{-1}^{\mathrm{b}[i]}\right).$$

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

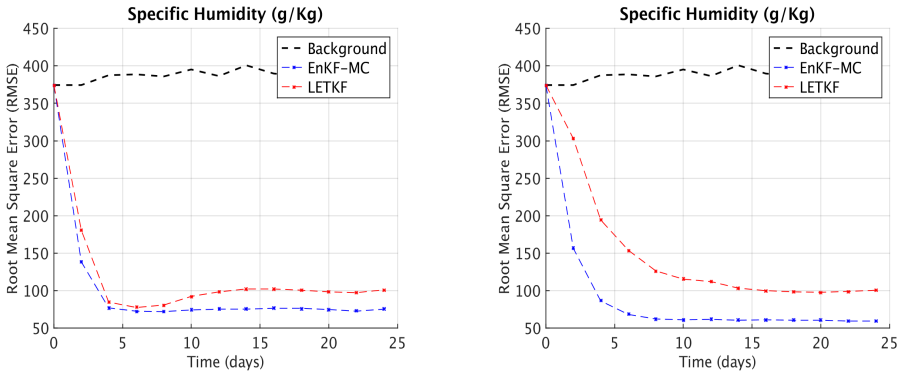
The experiments are performed over a period of 24 days, where observations are taken every 2 days ($M = 12$). At time k synthetic observations are built as follows:

$$\mathbf{y}_k = \mathbf{H}_k \cdot \mathbf{x}_k^{\mathrm{ref}} + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}\left(\mathbf{0}_m,\, \mathbf{R}_k\right), \quad \mathbf{R}_k = \mathrm{diag}_i\left\{(0.01\,\{\mathbf{H}_k\,\mathbf{x}_k^{\mathrm{ref}}\}_i)^2\right\}.$$

The observation operators $\mathbf{H}_k$ are fixed throughout the time interval. We perform experiments with several operators characterized by different proportions p of observed components from the model state $\mathbf{x}_k^{\mathrm{ref}}$ ($m \approx p \cdot n$). 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.

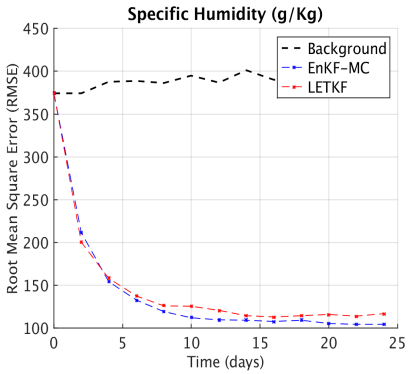
4.3. Results with dense observation networks. We first consider dense observational networks in which 100% and 50% of the model components are observed. We vary the radius of influence ζ from 1 to 5 grid points. Figures 3(a) and 3(b) show the RMSE values for the LETKF and EnKF-MC analyses for different values of ζ 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 for 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 for during the assimilation. Consequently, the larger the local box, the more local background error information cannot be represented in the ensemble space.

Figures 4(a) and 4(b) show that EnKF-MC analyses improve with increasing radius of influence ζ. 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 ζ and p. The quality of the EnKF-MC analysis for $\zeta = 5$ is better than that of the LETKF

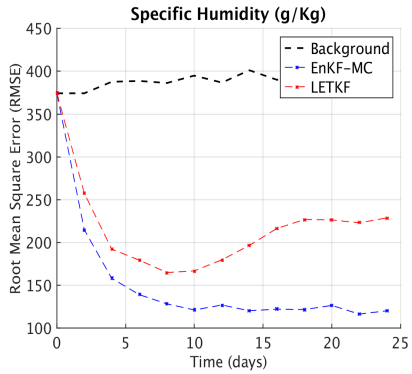


(a) $\zeta = 2$ and $p = 100\%$.

(b) $\zeta = 5$ and $p = 100\%$.



(c) $\zeta = 2$ and $p \sim 4\%$.



(d) $\zeta = 5$ and $p \sim 4\%$.

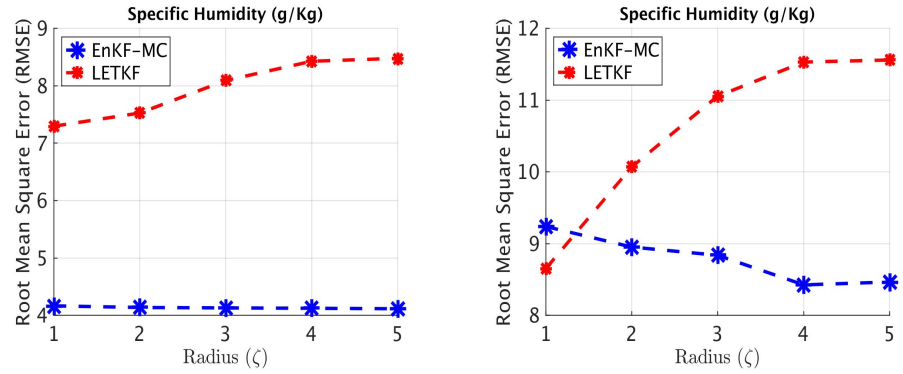
FIG. 3. *Experimental results with the SPEEDY model. RMSE of specific humidity analyses with a dense observational network are reported. When the radius of influence ζ is increased the performance of LETKF degrades.*

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 ζ from 1 to 5. Three sparse observational networks with $p = 12\%$, 6%, and 4%, respectively, are considered.

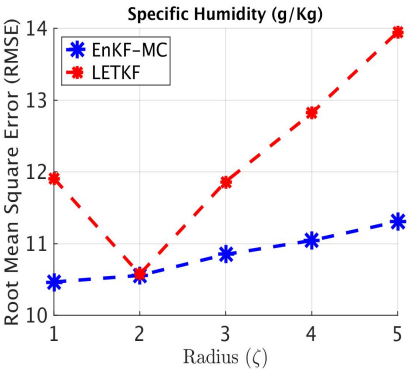
Figures 3(c) and 3(d) show the RMSE values of the specific humidity analyses for different radii of influence and 4% of the model components being observed. The best performance of the LETKF analyses is obtained when the radius of influence is set to 2. Note that for $\zeta = 1$ the LETKF performs poorly, which is expected since during the assimilation most of the 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 the largest radius of influence $\zeta = 5$.

Figures 4(c) and 4(d) show the RMSE values of the LETKF and the EnKF-MC implementations for different radii of influences and two sparse observational

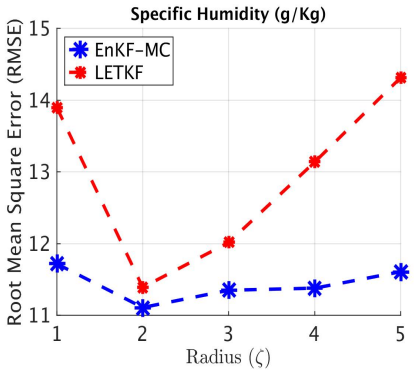


(a) $p = 100\%$.

(b) $p = 50\%$.



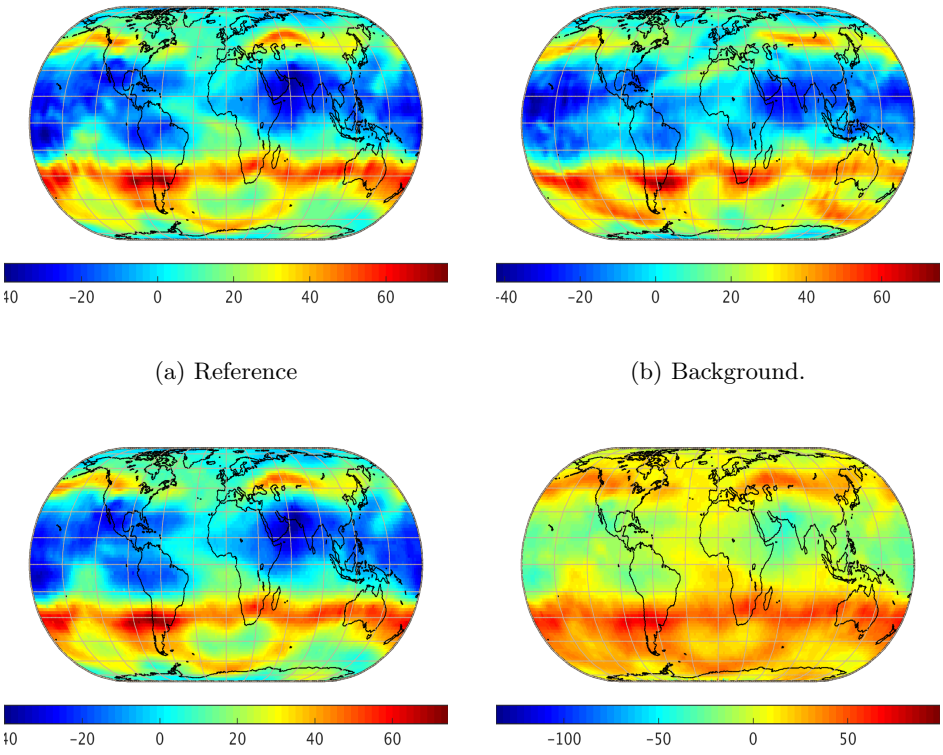
(c) $p = 6\%$.



(d) $p = 4\%$.

FIG. 4. *Experimental results with the SPEEDY model. Analysis RMSEs for the specific humidity variable are reported. The RMSE values of the assimilation window are shown for different values of ζ and percentage of observed components p. When the local domain sizes are increased the accuracy of the LETKF analysis degrades, while the accuracy of EnKF-MC analysis improves.*

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 error components are uncorrelated ζ 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 $\widehat{\mathbf{B}}$ 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 and the LETKF methods. The background solution contains erroneous



(a) Reference (b) Background.

(c) EnKF-MC (d) LETKF

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.*

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 EnKF, 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 $N_{ens} \approx \log n$, 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 asses 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

[1] J. L. Anderson, *Spatially and temporally varying adaptive covariance inflation for ensemble filters*, Tellus A, 61 (2009), pp. 72–83.

[2] J. L. Anderson, *Localization and sampling error correction in ensemble Kalman filter data assimilation*, Mon. Weather Rev., 140 (2012), pp. 2359–2371, https://journals.ametsoc.org/doi/pdf/10.1175/MWR-D-11-00013.1.

[3] A. Benedetti and M. Fisher, *Background error statistics for aerosols*, Quart. J. Roy. Meteorol. Soc., 133 (2007), pp. 391–405, https://doi.org/10.1002/qj.37.

[4] D. R. Bickel and M. Padilla, *A prior-free framework of coherent inference and its derivation of simple shrinkage estimators*, J. Statist. Plann. Inference, 145 (2014), pp. 204–221, https://doi.org/10.1016/j.jspi.2013.08.011.

[5] P. J. Bickel and E. Levina, *Regularized estimation of large covariance matrices*, Ann. Statist., 36 (2008), pp. 199–227, https://doi.org/10.1214/009053607000000758.

[6] M. Buehner, *Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation*, Mon. Weather Rev., 140 (2011), pp. 617–636, https://doi.org/10.1175/MWR-D-10-05052.1.

[7] H. Cheng, M. Jardak, M. Alexe, and A. Sandu, *A hybrid approach to estimating error covariances in variational data assimilation*, Tellus A, 62 (2010), pp. 288–297, https://doi.org/10.1111/j.1600-0870.2010.00442.x

[8] E. Constantinescu, A. SSandu, T. Chai, and G. Carmichael, *Autoregressive models of background errors for chemical data assimilation*, J. Geophys. Res., 112 (2007), D12309, https://doi.org/10.1029/2006JD008103.

[9] G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dyn., 53 (2003), pp. 343–367, https://doi.org/10.1007/s10236-003-0036-9.

[10] T. M. Hamill, J. S. Whitaker, and C. Snyder, *Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter*, Mon. Weather Rev., 129 (2001), pp. 2776–2790.

[11] P. C. Hansen, *Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank*, SIAM J. Sci. Stat. Comput., 11 (1990), pp. 503–518, https://doi.org/10.1137/0911028.

[12] A. Hollingsworth and P. Lonnberg, *The statistical structure of short-range forecast errors as determined from radiosonde data. Part* I: *The wind field*, Tellus A, 38 (1986), pp. 111–136, https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0870.1986.tb00460.x/full.

[13] S. V. Huffel, *Iterative algorithms for computing the singular subspace of a matrix associated with its smallest singular values*, Linear Algebra Appl., 154 (1991), pp. 675–709, https://www.sciencedirect.com/science/article/pii/002437959190399H.

[14] E. P. Jiang and M. W. Berry, *Solving total least-squares problems in information retrieval*, Linear Algebra Appl., 316 (2000), pp. 137–156, https://doi.org/10.1016/S0024-3795(00)00030-6.

[15] P. Jonathan, Z. Fuqing, and Y. Weng, *The effects of sampling errors on the EnKF assimilation of inner-core hurricane observations*, Mon. Weather Rev., 142 (2014), pp. 1609–1630, http://journals.ametsoc.org/doi/pdf/10.1175/MWR-D-13-00305.1.

[16] C. L. Keppenne, *Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter*, Mon. Weather Rev., 128 (2000), pp. 1971–1981, https://doi.org/10.1175/1520-0493(2000)128%3C1971:DAIAPE%3E2.0.CO;2.

[17] F. Kucharski, F. Molteni, and A. Bracco, *Decadal interactions between the Western Tropical Pacific and the North Atlantic oscillation*, Clim. Dyn., 26 (2006), pp. 79–91, https://doi.org/10.1007/s00382-005-0085-5.

[18] P. F. J. Lermusiaux, *Adaptive modeling, adaptive data assimilation and adaptive sampling*, Phys. D, 230 (2007), pp. 172–196, https://doi.org/10.1016/j.physd.2007.02.014.

[19] P. F. J. Lermusiaux and A. R. Robinson, *Data assimilation via error subspace statistical estimation. Part* I: *Theory and schemes*, 127 (1999), pp. 1385–1407, https://doi.org/10.1175/1520-0493(1999)127%3C1385:DAVESS%3E2.0.CO;2

[20] A. C. Lorenc, *The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var*, Quart. J. Roy. Meteorol. Soc., 129 (2003), pp. 3183–3203.

[21] E. N. Lorenz, *Designing chaotic models*, J. Atmos. Sci., 62 (2005), pp. 1574–1587, https://doi.org/10.1175/JAS3430.1.

[22] F. Molteni, *Atmospheric simulations using a GCM with simplified physical parametrizations.* I: *Model climatology and variability in multi-decadal experiments*, Clim. Dyn., 20 (2003), pp. 175–191, https://doi.org/10.1007/s00382-002-0268-2.

[23] E. Nino Ruiz, A. Sandu, and J. Anderson, *An efficient implementation of the ensemble Kalman filter based on an iterative Sherman–Morrison formula*, Stat. Comput., 25 (2014), pp. 561–577, https://doi.org/10.1007/s11222-014-9454-4.

[24] E. D. Nino-Ruiz and A. Sandu, *Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation*, Ocean Dyn., 65 (2015), pp. 1423–1439.

[25] E. Ott, B. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, *A local ensemble transform Kalman filter data assimilation system for the NCEP global model*, Tellus A, 60 (2008), pp. 113–130, https://doi.org/10.1111/j.1600-0870.2007.00274.x.

[26] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, *A local ensemble Kalman filter for atmospheric data assimilation*, Tellus A, 56 (2004), pp. 415–428, https://doi.org/10.3402/tellusa.v56i5.14462.

[27] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, *A local ensemble Kalman filter for atmospheric data assimilation*, Tellus A, 56 (2004), pp. 415–428, https://doi.org/10.3402/tellusa.v56i5.14462.

[28] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, Cambridge, 2015.

[29] P. Sakov and P. R. Oke, *A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters*, Tellus A, 60 (2008), pp. 361–371, https://doi.org/10.1111/j.1600-0870.2007.00299.x.

[30] G. Ueno and T. Tsuchiya, *Covariance regularization in inverse space*, Quart. J. Roy. Meteorol. Soc., 135 (2009), pp. 1133–1156.

[31] J. S. Whitaker and M. H. Thomas, *Ensemble data assimilation without perturbed observations*, Mon. Weather Rev., 16 (2002), pp. 1913–1924, https://doi.org/10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2.