

An Ensemble Kalman Smoother Based On Modified Cholesky Decomposition

Elias D. Nino-Ruiz · Sebastian Garrido-Cepeda

Received: date / Accepted: date

Abstract In this paper, we propose an ensemble Kalman smoother based on a modified Cholesky decomposition. The proposed method works as follows: based on an ensemble of model realizations, moments of the background error distribution are estimated at times wherein observations are available, the ensembles are utilized for creating sub-spaces onto which observations are projected and assimilation steps are performed. The sub-spaces are built by means of a modified Cholesky decomposition of the precision background covariance. Experimental tests are performed by using the Lorenz 96 model. The results reveal that, in terms of root-mean-square-error values, the accuracy of the proposed filter is much better than that of well-known smoothing methods from the specialized literature.

Keywords First keyword · Second keyword · More

1 Introduction

To be concise, Data Assimilation (DA) is the process by which an imperfect numerical forecast and a set of noisy observations $\{\mathbf{y}_k\}_{k=0}^{M-1}$ are fused in order

Elias D. Nino-Ruiz
Applied Math and Computer Science Laboratory
Department of Computer Science
Universidad del Norte, BAQ 0180001, Colombia
Tel.: +64-5-3509268
E-mail: enino@uninorte.edu.co

Sebastian Garrido-Cepeda
Applied Math and Computer Science Laboratory
Department of Computer Science
Universidad del Norte, BAQ 0180001, Colombia
Tel.: +64-5-3509268
E-mail: sebastiangarrido@uninorte.edu.co

to estimate the state of a dynamical system $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$ which (approximately) evolves according to some numerical model

$$\mathbf{x}_{k+1} = \mathcal{M}_{t_k \rightarrow t_{k+1}}(\mathbf{x}_k), \text{ for } \mathbf{x}_k \in \mathbb{R}^{n \times 1}, \text{ and } 0 \leq k \leq M-1, \quad (1)$$

where n is the number of model components, $\mathbf{x}_k \in \mathbb{R}^{n \times 1}$ is a vector state, M is the number of observations which form the assimilation window, $\mathbf{y}_k \in \mathbb{R}^{m \times 1}$ is an observation, and m is the number of observed components from the vector state. For ease in the notation, we assume that observed components across the assimilation window remain unchanged. In the context of the ensemble Kalman smoother (EnKFS), at assimilation steps, an ensemble of model realizations

$$\mathbf{X}_k^b = [\mathbf{x}_k^{b[1]}, \mathbf{x}_k^{b[2]}, \dots, \mathbf{x}_k^{b[N]}] \in \mathbb{R}^{n \times N}, \quad (2)$$

is commonly utilized in order to approximate the moments of background error distributions

$$\mathbf{x}_k \sim \mathcal{N}(\mathbf{x}_k^b, \mathbf{B}_k), \quad (3)$$

where N is the ensemble size, $\mathbf{x}_k^{b[e]}$ is the e -th ensemble member, for $1 \leq e \leq N$. Based on these samples, cross-correlations of model components in time are computed in order to exploit all information contained within the assimilation window. This provides, for instance, an initial ensemble \mathbf{X}_0^b whose model trajectory fits the observations. Due to computational costs, model dimensions are several times ensemble sizes and therefore, spurious correlations can impact the quality of analysis corrections. Typically, these problems can be lessened by using localization methods, for instance, covariance matrix localization. In this context, ensemble covariances are component-wise multiplied by a localization matrix in order to reduce far-field sampling errors. In practice, compact representations of localization matrices are available but, most of them depend on square-root approximations of such matrix which can still be a huge matrix. We think that, such matrices are not necessary and better manners to perform the smoothing process are possible via a modified Cholesky decomposition. By using this method, for instance, the precision covariances at assimilation times can be estimated in terms of Cholesky factors and besides, such factors can be sparse by exploiting conditional independences of model components in time and in space. This implies huge savings in terms of memory usage.

This paper is organized as follows: in section 2 ensemble based methods are discussed as well as some variational approaches, section 3 presents an ensemble Kalman smoother based on a modified Cholesky decomposition which exploits the conditional independence of model components in time and in space regarding their physical locations, in section 4 experimental tests are performed by using the Lorenz 96 model and a quasi-geostrophic model, and lastly, section 5 states the conclusions of this research.

2 Preliminaries

2.1 The Ensemble Kalman Filter

In the ensemble Kalman filter (EnKF) [2, 4, 13, 5], the moments of the background error distribution

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}_k^b, \mathbf{B}_k), \text{ for } 0 \leq k \leq M-1, \quad (4)$$

are estimated via an ensemble of model realizations (2), therefore,

$$\mathbf{x}_k^b \approx \bar{\mathbf{x}}_k^b = \frac{1}{N} \cdot \sum_{e=1}^N \mathbf{x}_k^{b[e]} \in \mathbb{R}^{n \times 1}, \quad (5a)$$

$$\mathbf{P}_k^b = \frac{1}{N-1} \cdot \Delta \mathbf{X}_k \cdot \Delta \mathbf{X}_k^T \in \mathbb{R}^{n \times n}, \quad (5b)$$

and $\Delta \mathbf{X}_k \in \mathbb{R}^{n \times N}$ is the matrix of member deviations:

$$\Delta \mathbf{X}_k = \mathbf{X}_k^b - \bar{\mathbf{x}}_k^b \cdot \mathbf{1}^T, \quad (5c)$$

where $\mathbf{1}$ is a vector of consistent dimension whose components are all ones. In practice, model dimensions are several times ensemble sizes and therefore, the covariance matrix (5b) is typically rank deficient. By using covariance localization, the effect of sampling errors over the estimated background error correlations (5b) can be mitigated. In this context, a recent filter is the ensemble Kalman filter based on a modified Cholesky decomposition (EnKF-MC) [8, 9]. In the EnKF-MC, the ensemble covariance (5b) is replaced by a regularized estimation $\hat{\mathbf{B}}_k^{-1}$ of the precision covariance \mathbf{B}_k^{-1} , this estimator can be obtained in terms of Cholesky factors:

$$\hat{\mathbf{B}}_k^{-1} = \mathbf{L}_k \cdot \mathbf{D}_k \cdot \mathbf{L}_k^T, \quad (6)$$

where $\mathbf{L}_k \in \mathbb{R}^{n \times n}$ is a lower triangular matrix whose non-zero sub-diagonal elements are obtained by fitting models of the form:

$$\left[\mathbf{x}^{[i]} \right]^T = \sum_{j=1}^{i-1} -\{\mathbf{L}_k\}_{i,j} \cdot \left[\mathbf{x}^{[j]} \right]^T + \gamma_i \in \mathbb{R}^{N \times 1}, \text{ for } 1 \leq i \leq n, \quad (7)$$

where $\mathbf{x}^{[i]} \in \mathbb{R}^{1 \times N}$ stands for the i -th row of the ensemble (2). Likewise, $\mathbf{D}_k \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements are the reciprocal variances of residuals in (7),

$$\{\mathbf{D}_k\}_{i,i} = \text{var} \left(\left[\mathbf{x}^{[i]} \right]^T + \sum_{j=1}^{i-1} \{\mathbf{L}_k\}_{i,j} \cdot \left[\mathbf{x}^{[j]} \right]^T \right)^{-1}. \quad (8)$$

The analysis ensemble can then be estimated as follows:

$$\mathbf{X}_k^a = \mathbf{X}_k^b + \hat{\mathbf{A}}_k \cdot \mathbf{H}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Y}_k^s, \quad (9)$$

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

$$\hat{\mathbf{A}} = \left[\hat{\mathbf{B}}^{-1} + \mathbf{H}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{H}_k \right]^{-1}, \quad (10)$$

and columns of matrix $\mathbf{Y}_k^s \in \mathbb{R}^{m \times N}$ are drawn from a Normal distribution with parameters $\mathcal{N}(\mathbf{y}_k, \mathbf{R}_k)$. After the assimilation step, ensemble members are propagated in time until the next assimilation cycle,

$$\mathbf{x}_{k+1}^{b[e]} = \mathcal{M}_{t_k \rightarrow t_{k+1}} \left(\mathbf{x}_k^{a[e]} \right), \text{ for } 1 \leq e \leq N,$$

from which a new background ensemble is obtained. The direct computation of matrix (10) is not actually needed and some matrix-free methods for the EnKF-MC are now proposed in the literature [7, 6].

2.2 Hybrid Ensemble Four Dimensional Methods

In the context of four-dimensional data assimilation (4D-Var) cost functions of the form,

$$\mathcal{J}_{4D}(\mathbf{x}) = \frac{1}{2} \cdot \|\mathbf{x}_0 - \mathbf{x}_0^b\|_{[\mathbf{P}_0^b]^{-1}}^2 + \frac{1}{2} \cdot \sum_{k=0}^{M-1} \|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2, \quad (11)$$

are considered in order to find the initial condition $\mathbf{x}_0 \in \mathbb{R}^{n \times 1}$, for an imperfect numerical model $\mathcal{M} : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{n \times 1}$, whose model trajectory

$$\mathbf{x}_k = \mathcal{M}_{t_{k-1} \rightarrow t_k}(\mathbf{x}_{k-1}), \text{ for } 0 \leq k \leq M-1, \quad (12)$$

best fit the given observations $\{\mathbf{y}_k\}_{k=0}^{M-1}$, where M is the number of observation which form the assimilation window, n denotes the number of model components, $\mathbf{x}_0^b \in \mathbb{R}^{n \times 1}$ is the initial background state, $\mathbf{P}_0^b \in \mathbb{R}^{n \times n}$ is the background covariance matrix at time 0, and at time k , for $0 \leq k \leq M-1$, $\mathbf{y}_k \in \mathbb{R}^{m \times 1}$ is the observation, m is the number of observed components from the model state, $\mathbf{R}_k \in \mathbb{R}^{m \times m}$ is the data error covariance matrix, $\mathcal{H}_k : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$ is the observation operator which maps states from the model space to the observation space. In strong constraint 4D-Var, typically, the best initial condition can be found via the optimization problem:

$$\mathbf{x}_0^* = \arg \min_{\mathbf{x}_0} \mathcal{J}_{4D}(\mathbf{x}_0), \text{ subject to (12)}. \quad (13)$$

Before 4D-Var, variational methods were constrained in their capacity to quantify appropriately the flow-dependent background and model error statistics. However, commonly, only the first moment, the initial analysis is recovered from the optimization process. 4D-Var methods rely on linearized and adjoints

versions of the numerical model (1). In practice, linearizing highly non-linear processes can demand significant cost in terms of time and resources (human capital) [15, 12, 11, 10]. Ensemble based methods are normally utilized in order to avoid such computation. For instance, in the subject of four dimensional ensemble Kalman filter methods (4DEnVar) [1, 3, 14], the model trajectory is restricted to the range of matrices (5c),

$$\mathbf{x}_k = \bar{\mathbf{x}}_k^b + \Delta \mathbf{X}_k \cdot \mathbf{w} \in \mathbb{R}^{n \times 1}, \quad (14)$$

where $\mathbf{w} \in \mathbb{R}^{N \times 1}$ is a vector in redundant coordinates to be determined. By replacing (14) in (11) we obtain

$$\mathcal{J}_{4E}(\mathbf{w}) = \frac{N-1}{2} \cdot \|\mathbf{w}\|^2 + \frac{1}{2} \cdot \sum_{k=0}^{M-1} \|\mathbf{d}_k - \mathbf{Z}_k \cdot \mathbf{w}\|_{\mathbf{R}_k^{-1}}^2. \quad (15)$$

where $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \cdot \bar{\mathbf{x}}_k^b \in \mathbb{R}^{m \times 1}$ and $\mathbf{Z}_k = \mathbf{H}_k \cdot \Delta \mathbf{X}_k \in \mathbb{R}^{m \times N}$. The initial analysis (13) can be approximated by solving the N -dimensional optimization problem:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathcal{J}_{4E}(\mathbf{w}), \text{ for } \mathbf{w} \in \mathbb{R}^{N \times 1}, \quad (16)$$

whose solution, for linear observation operators, reads

$$\mathbf{w}^* = \left[(N-1) \cdot \mathbf{I} + \sum_{k=1}^{M-1} \mathbf{Z}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Z}_k \right]^{-1} \cdot \left[\sum_{k=1}^{M-1} \mathbf{Z}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{d}_k \right],$$

and therefore $\mathbf{x}_0^* \approx \bar{\mathbf{x}}_0^a = \bar{\mathbf{x}}_0^b + \Delta \mathbf{X} \cdot \mathbf{w}^*$. Note that, the control vector \mathbf{w} has N components while the vector \mathbf{x}_0 in the strong constraint 4D-Var optimization problem has n elements. Hence, since $N \ll n$, the computational benefits of using 4DEnVar methods are just evident. However, the low-rank property of covariance matrices \mathbf{P}_k implies that sampling errors will impact the resulting state $\bar{\mathbf{x}}_0^a$. Consequently, some sort of localization must be used in order to mitigate such effect.

2.3 The Ensemble Kalman Smoother

Roughly speaking, in the ensemble Kalman smoother (EnKFS), an ensemble of snapshots can be used in order to perform a smoothing process of observations over the assimilation windows:

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \left[\left[\hat{\mathbf{P}}^b \right]^{-1} + \hat{\mathbf{H}}^T \cdot \hat{\mathbf{R}}^{-1} \cdot \hat{\mathbf{H}} \right]^{-1} \cdot \left[\hat{\mathbf{y}} - \hat{\mathbf{H}} \cdot \hat{\mathbf{x}}^b \right] \in \mathbb{R}^{\hat{n} \times N}, \quad (17)$$

where $\left\{ \mathbf{x}_k^{b[e]} \right\}_{k=0}^{N-1}$ is the e -th ensemble member at time k , for $1 \leq e \leq N$, $\hat{n} = n \cdot M$, $\hat{\mathbf{P}}^b \in \mathbb{R}^{\hat{n} \times \hat{n}}$, $\hat{\mathbf{R}} \in \mathbb{R}^{\hat{m} \times \hat{m}}$ and $\hat{\mathbf{H}} \in \mathbb{R}^{\hat{m} \times \hat{n}}$ are block-diagonal matrices

with diagonal blocks $\{\mathbf{P}_k^b\}_{k=0}^{M-1}$, $\{\mathbf{R}_k\}_{k=0}^{M-1}$, and $\{\mathbf{H}_k\}_{k=0}^{M-1}$, respectively, and $\mathbf{P}_k^b \in \mathbb{R}^{n \times n}$ is the ensemble covariance. Likewise, $\hat{\mathbf{x}}^a \in \mathbb{R}^{\hat{n} \times 1}$, $\hat{\mathbf{x}}^b \in \mathbb{R}^{\hat{n} \times 1}$, and $\hat{\mathbf{y}} \in \mathbb{R}^{\hat{m} \times 1}$ are augmented vectors with elements,

$$\hat{\mathbf{x}}^a = \begin{bmatrix} \bar{\mathbf{x}}_0^a \\ \bar{\mathbf{x}}_1^a \\ \vdots \\ \bar{\mathbf{x}}_{M-1}^a \end{bmatrix} \in \mathbb{R}^{\hat{n} \times 1}, \hat{\mathbf{x}}^b = \begin{bmatrix} \bar{\mathbf{x}}_0^b \\ \bar{\mathbf{x}}_1^b \\ \vdots \\ \bar{\mathbf{x}}_{M-1}^b \end{bmatrix} \in \mathbb{R}^{\hat{n} \times 1}, \text{ and } \hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{M-1} \end{bmatrix} \in \mathbb{R}^{\hat{m} \times 1} \quad (18)$$

where $\{\bar{\mathbf{x}}_k^b\}_{k=0}^{M-1} \in \mathbb{R}^{n \times 1}$ denotes ensemble means at observation times. Similarly, $\{\bar{\mathbf{x}}_k^a\}_{k=0}^{M-1} \in \mathbb{R}^{n \times 1}$ stands for analysis means. Thus, $\bar{\mathbf{x}}_0^a \approx \mathbf{x}_0^*$ in (13). A main issue with this approximation is that, sampling errors degenerate the quality of background error correlations in $\hat{\mathbf{P}}^b$ and therefore, the analysis corrections are highly impacted by such estimation. Localization methods are commonly utilized in order to overcome this situation but, the direct localization of covariance matrices such as $\hat{\mathbf{P}}^b$ is prohibitive owing to his dimension, indeed, it is impossible to explicitly store such matrix in memory. This paper is highly motivated by such constraint and therefore, we seek an efficient implementation of the ensemble Kalman smoother by using conditional independence of model components in time and space via a modified Cholesky decomposition. This decomposition allow us to write $\hat{\mathbf{P}}^b$ in terms of Cholesky factors and even more, such factors can be highly sparse which implies huge saving in terms of memory usage during the assimilation of observations.

3 Proposed Method

4 Experimental Results

5 Conclusions

References

1. Gérald Desroziers, Jean-Thomas Camino, and Loïk Berre. 4denvar: link with 4d state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2097–2110, 2014.
2. Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
3. D Fairbairn, SR Pring, AC Lorenc, and I Roulstone. A comparison of 4dvar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140(678):281–294, 2014.
4. Thomas M Hamill. Ensemble-based atmospheric data assimilation. *Predictability of weather and climate*, pages 124–156, 2006.
5. PL Houtekamer and Fuqing Zhang. Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12):4489–4532, 2016.
6. Elias D Nino-Ruiz. A matrix-free posterior ensemble kalman filter implementation based on a modified cholesky decomposition. *Atmosphere*, 8(7):125, 2017.

7. Elias D Nino-Ruiz, Alfonso Mancilla, and Juan C Calabria. A posterior ensemble kalman filter based on a modified cholesky decomposition. *Procedia Computer Science*, 108:2049–2058, 2017.
8. Elias D Nino-Ruiz, Adrian Sandu, and Xinwei Deng. A parallel ensemble kalman filter implementation based on modified cholesky decomposition. In *Proceedings of the 6th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems*, page 4. ACM, 2015.
9. Elias D Nino-Ruiz, Adrian Sandu, and Xinwei Deng. A parallel implementation of the ensemble kalman filter based on modified cholesky decomposition. *Journal of Computational Science*, 2017.
10. Elias D Nino Ruiz and Adrian Sandu. A derivative-free trust region framework for variational data assimilation. *Journal of Computational and Applied Mathematics*, 293:164–179, 2016.
11. O Stiller. Efficient moist physics schemes for data assimilation. ii: Deep convection. *Quarterly Journal of the Royal Meteorological Society*, 135(640):721–738, 2009.
12. O Stiller and SP Ballard. Efficient moist physics schemes for data assimilation. i: Large-scale clouds and condensation. *Quarterly Journal of the Royal Meteorological Society*, 135(640):707–720, 2009.
13. Christian Sutton, Thomas M Hamill, and Thomas T Warner. Will perturbing soil moisture improve warm-season ensemble forecasts? a proof of concept. *Monthly weather review*, 134(11):3174–3189, 2006.
14. Xiangjun Tian, Zhenghui Xie, and Aiguo Dai. An ensemble-based explicit four-dimensional variational assimilation method. *Journal of Geophysical Research: Atmospheres*, 113(D21), 2008.
15. Shuowen Yang and Qin Xu. Statistical errors in variational data assimilation—a theoretical one-dimensional analysis applied to doppler wind retrieval. *Journal of the atmospheric sciences*, 53(18):2563–2577, 1996.