Wind Energy Potential Estimation via a Hybrid Data Assimilation Method

Ph.D. Thesis of

Juan Carlos Calabria Sarmiento

at the Department of Computer Science
Universidad del Norte, Barranquilla, Colombia

Reviewer 1: Daladier Jabba Molinares, Ph.D.
Reviewer 2: Miguel Angel Jimeno Paba, Ph.D.
Reviewer 3: Alvin Arturo Henao Perez, Ph.D.
Reviewer 4: Ivan Saavedra Antolinez, Ph.D.
Advisor: Elias D. Nino-Ruiz, Ph.D.
Co-Advisor: Jose R. Capacho-Portilla, Ph.D.

Fall 2016 – Spring 2020

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

PLACE, DATE

(Juan Carlos Calabria Sarmiento)

Acknowledgments

I immensely thank Jehovah, Jesus Christ and the wonderful Holy Spirit for giving me the wisdom and the right path to carry out this life project.
To my sentimental partner Margarita Rosa, to my children María Carolina, Carlos Miguel and Carlos Alfonso for their support and understanding every time I have to be away.
To my nephew-son Mateo Rafael Calabria Stand because he is always there supporting me in moments of fatigue with his thanks and smile.
To my parents Carlos Alfonso Calabria Niebles and Adela Sarmiento Navarro for the education they gave me and instilled in me the love for Jesus Christ, my idol and respect for others.
To my great director, Engineer Elias David Niño Ruiz, a male servant of Jehovah, for his unconditional support and patience for the development and completion of this thesis.

Abstract

This research proposes a 4D-Var ensemble-based data assimilation framework for wind energy potential estimation. In this formulation, in the 4D-Var context, the intrinsic need of adjoint models is avoided via the use of an ensemble of model realizations. These ensembles are employed to build control spaces onto which analysis increments are estimated. Control spaces are built via a modified Cholesky decomposition. The particular structure of this estimator allows for a matrix-free implementation of the proposed filter formulation. Experimental tests are performed, making use of wind turbines catalogs and the Atmospheric General Circulation Model Speedy. The results reveal that our proposed framework can properly estimate wind energy potential capacities within reasonable accuracies in terms of Root-Mean-Square-Error, and even more, these estimations are better than those of traditional 4D-Var ensemble-based methods. Besides, Wind Turbine Generators (WTGs) with low rate-capacity are the ones which provide homogeneous behavior of error estimations around the globe. As the rate-capacity increases, the potential energy increases as well, but the error dispersion of ensemble members grow, which can difficult decision-making processes. Of course, rate-capacity is just a single parameter of many in the WTG context, and we do not consider, for instance, economic aspects in our study, which can be crucial for deciding whether or not to employ green sources of energy.

Contents

List of Figures

List of Tables

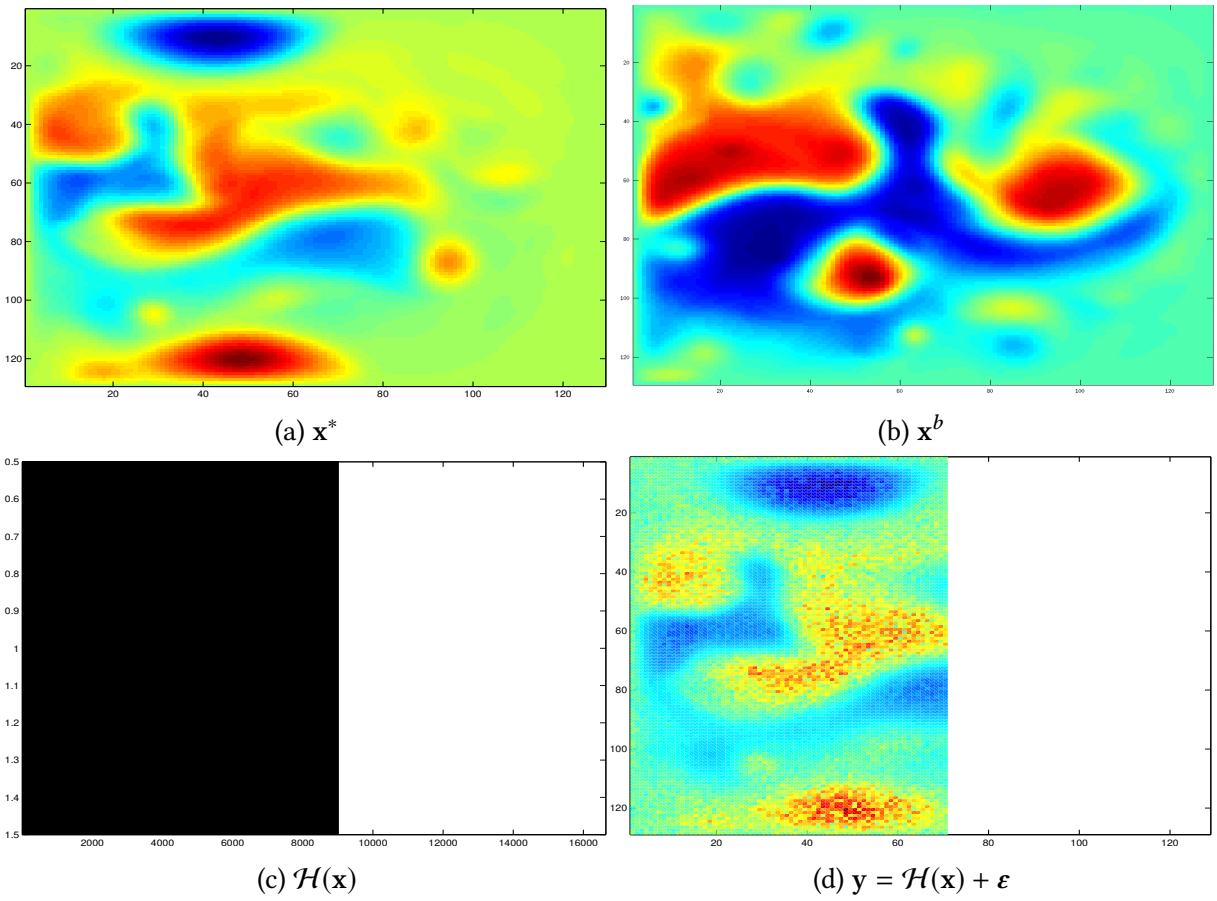
1 Introduction

The Nature Conservancy (TNC) [TNC19], an international NGO responsible for promoting trends that mitigate the negative effect on the environment, states that Colombia at 2050 will, at least, duplicate its energy consumption. Thus, it requires to develop a smart strategy to deploy a Data Assimilation method to produce wind power. Latin American and Caribbean (LAC) countries are widely known for their massive power generation capacity using renewable energies, which makes them highly attractive for clean energy investment [FAF+17]. Recent studies indicate that the full deployment of this capacity can be almost seven times larger than the current world installed one, and even more, it can constitute a near-zero carbon emissions option for countries in development path [DRB08]. As [VIR+14] states, "this could provide substantial societal benefits, including energy security, local and global environmental benefits, domestic job creation, and improved balance of payments, amongst others". Given Colombia's geography, wind turbines (wind farms) can be exploited for having clean energy sources. Economic benefits of wind farms are better than those of traditional sources such as solar parks, which make them the first option desirable for long-term plans at national levels. The proper planning and scheduling of wind power systems can lead to almost no impact on Colombian ecosystems, nor visual or audible. Besides, this can serve as a complement of the hydro-dominated electricity grids [MWOH17], as the winds are stronger during the dry season when hydroelectric generation is most limited. Moreover, wind farms provide a full or complementary source of energy in some areas of difficult access; the application of wind turbines is primarily in windmills that are used to generate electricity [GJSAGJ17]. These wind turbines can be used to avail off-grid electricity in remote regions (i.e., some islands). Even more, in some cases, they are the only potential source of energy. Nowadays, some government agencies provide decades of useful information about global weather variables such as wind components, temperatures, and humidity. For instance, the amount of data on the US National Oceanic and Atmospheric Administration (NOAA) website range in the order of petabytes. This can be exploited by using Data Assimilation methods to estimate the economic benefits of wind farm placements in Colombia, for instance. Hence, a data assimilation method is required to forecast wind components and wind velocity to estimate energetic power in Colombia. With these forecasts, we can identify highly windy areas to get the most of power availability.

This document is organized as follows: Section 1.1 discusses topics related to Data Assimilation in ensemble-based and variational contexts. In Section 1.2, we present some wind turbines catalogs to be employed in this research for wind-potential-energy estimation. Section 1.3 presents some relevant issues found during this exploratory step. In section 1.4, the expected outcomes of this research are detailed. Section 1.6 states the objective of this research. In Section 1.6, the steps to accomplish the objectives are denoted. Lastly, Section 1.7 show the advances of our research so far as well as the numerical model to be employed during the experimental part.

1.1 Data Assimilation

Data Assimilation (DA) is the process by which an imperfect numerical forecast $\{\mathbf{x}_k^b\}_{k=0}^G$ is adjusted according to real noisy observations $\{\mathbf{y}_k\}_{k=0}^G$ [NR18, NRCB18], where $\mathbf{x}_k^b \in \mathbb{R}^{n\times 1}$ and $\mathbf{y}_k \in \mathbb{R}^{m\times 1}$ are the background state and the observations at step k, for $0 \leq k \leq G$, respectively, n is the model size (model resolution), m denotes the number of observations per assimilation step, and G is the size of the assimilation window (the number of times wherein observations are available). The elements of DA are shown in figure 1.1.



(a) $\mathbf{x}^*$ (b) $\mathbf{x}^b$

(c) $\mathcal{H}(\mathbf{x})$ (d) $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \boldsymbol{\varepsilon}$

Figure 1.1: Data Assimilation Components

In strong constraint Four-Dimensional Variational (4D-Var) methods, cost function of the form:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2} \cdot \left\|\mathbf{x}_0 - \mathbf{x}_0^b\right\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \cdot \sum_{k=0}^{G} \left\|\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k)\right\|_{\mathbf{R}^{-1}}^2 , \tag{1.1}$$

where:

$$\begin{aligned} \mathbf{x}_0 - \mathbf{x}_0^b &\sim \mathcal{N}(\mathbf{0}, \mathbf{B}) , \text{ and} \\ \mathbf{y}_k - \mathcal{H}(\mathbf{x}_k) &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}) , \end{aligned}$$

are employed to perform the assimilation process, where $\mathbf{B} \in \mathbb{R}^{n\times n}$ and $\mathbf{R} \in \mathbb{R}^{m\times m}$ are the background error covariance matrix and the estimated data error covariance matrix, respectively. **$\mathbf{x}^b$ and B must be estimated.** Likewise,

$$\mathbf{x}_f = \mathcal{M}_{t_{f-1}\to t_f}\left(\mathbf{x}_{f-1}\right)\ ,\ \text{for } 1 \le f \le G\,, \tag{1.2}$$

where $\mathcal{M} : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ is a numerical model which, for instance, mimics the behavior of the ocean and/or the atmosphere and even more, we assume that the model can predict wind components (or wind-speed). We then seek the initial condition which best fit the data:

$$\mathbf{x}_0^a = \arg\min_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0)\,,\ \text{subject to (1.2)}\,. \tag{1.3}$$

To solve (1.3), for instance, we can make use of adjoint models or an ensemble of model realizations. Regardless of which one is chosen, we think that 4D-Var DA can be exploited in the context of Wind Turbine Generators (WTG). For instance, we can make use of numerical models to estimate wind components in areas (regions) of interests, then by using variables from wind turbines such as cut-in speed, cut-out speed, and rated capacity, we can forecast their WTG capacities. This can be exploited in places such as the Latin American and Caribbean (LAC) countries, specifically, Colombia.

1.1.1 Data Assimilation: Sequential vs Variational

Consider G observations, ie, $\{\mathbf{y}_k\}_{k=0}^G$, and $\mathbf{x}_k = \mathcal{M}_{x_{k-1}\to x_k}\left(\mathbf{x}_{k-1}\right)$.

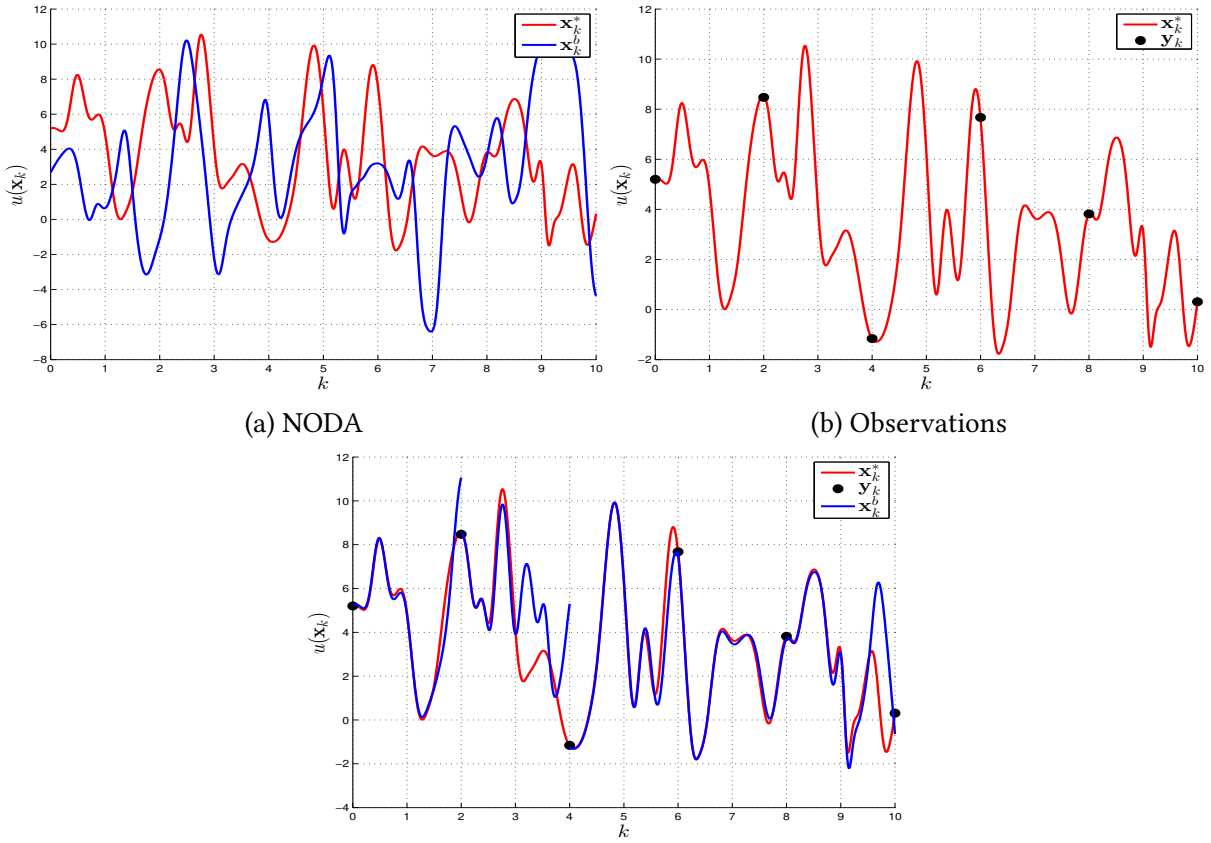
- Sequential method: it corrects the model estimate as observations are available, and then, propagate the estimate. Mathematically,

  $$\mathcal{P}\left(\mathbf{x}_k|\mathbf{y}_k\right) \propto \mathcal{P}\left(\mathbf{x}_k\right)\cdot\mathcal{L}\left(\mathbf{x}_k|\mathbf{y}_k\right)$$

  The **Analysis** at time k is given by:

  $$\mathbf{x}_k^a = \arg\max_{\mathbf{x}} \mathcal{P}\left(\mathbf{x}_k|\mathbf{y}_k\right)\ .$$

  This process can be seen in figure 1.2.



(a) NODA

(b) Observations

(c) Sequential Assimilation

Figure 1.2: Sequential Data Assimilation for an Assimilation Window.

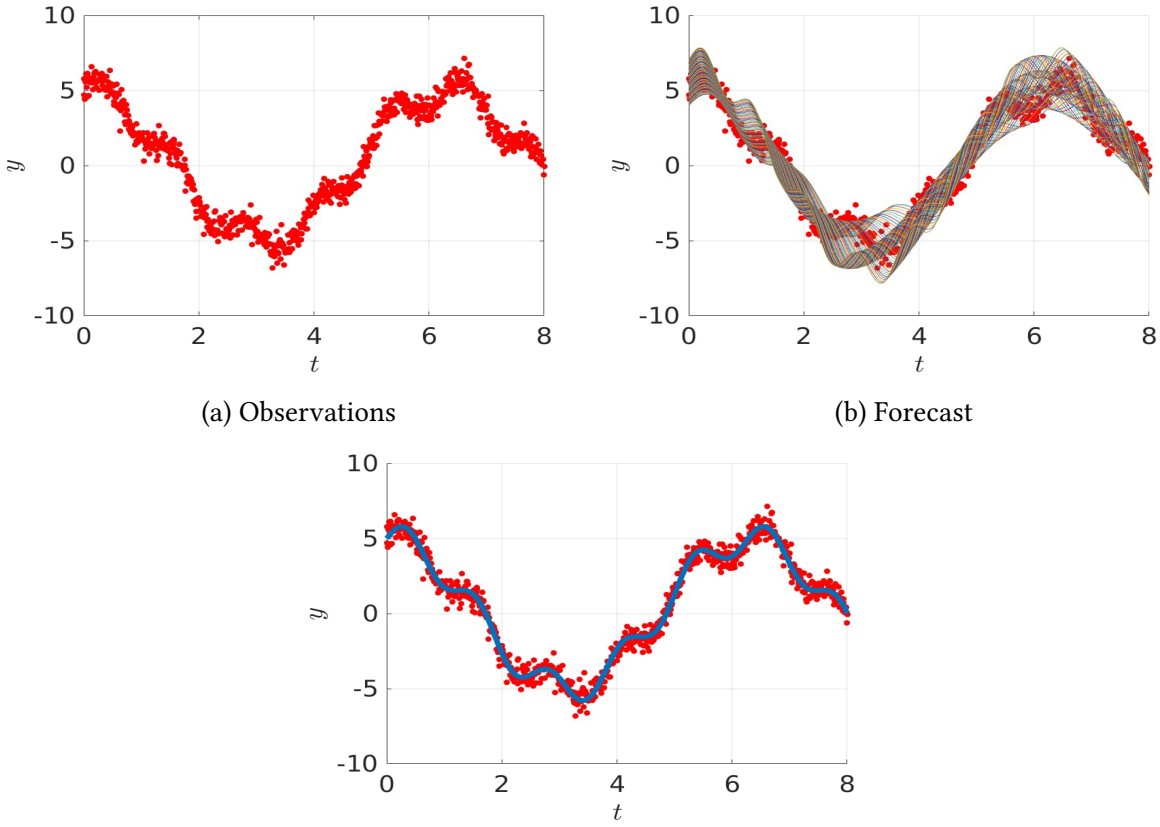
- Variational method: performs then optimization process in a single shot, it means using all observations at once.

$$\mathcal{P}\left(\mathbf{x}_0 | \{\mathbf{y}_k\}_{k=0}^G\right) \propto \mathcal{P}(\mathbf{x}_0) \cdot \prod_{k=0}^{G} \mathcal{L}(\mathbf{x}_k | \mathbf{y}_k)$$

The **Analysis** at time 0:

$$\mathbf{x}_0^a = \arg\max_{\mathbf{x}_0} \mathcal{P}\left(\mathbf{x}_0 | \{\mathbf{y}_k\}_{k=0}^G\right) .$$

The smoothing process can be seen in figure 1.3.



(a) Observations

(b) Forecast

(c) Initial Analysis

Figure 1.3: Smoothing Data Assimilation for an Assimilation Window.

1.1.2 Ensemble-based methods

- **Ensemble Kalman Filter (EnKF):**

  The ensemble Kalman filter (ENKF) is a sequential Monte Carlo method for parameter and state estimation of highly non-linear models [Eve06]. The popularity of the EnKF obeys to his simple formulation and relatively ease implementation [GMC+06]. In the EnKF, an ensemble of N model realizations,

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \ldots, \mathbf{x}^{b[N]}\right] \in \mathbb{R}^{n \times N}, \tag{1.4a}$$

  where $x^{b[e]} \in \mathbb{R}^{n \times 1}$ denotes the e-th ensemble member, for $1 \leq e \leq N$, is utilized in order to estimate the moments of the prior error distribution,

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

via the empirical moments of the ensemble,

$$\mathbf{x}^b \approx \overline{\mathbf{x}}^b = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^{b[e]} \in \mathbb{R}^{n\times 1}\,, \tag{1.4b}$$

and

$$\mathbf{B} \approx \mathbf{P}^b = \frac{1}{N-1} \cdot \mathbf{\Delta X} \cdot \mathbf{\Delta X}^T \in \mathbb{R}^{n\times n}\,, \tag{1.4c}$$

where $\overline{\mathbf{x}}^b$ and $\mathbf{P}^b$ are the ensemble mean and the ensemble covariance matrix, respectively. Likewise, $\mathbf{\Delta X} \in \mathbb{R}^{n\times N}$ is the matrix of member deviations,

$$\mathbf{\Delta X} = \mathbf{X}^b - \overline{\mathbf{x}}^b \cdot \mathbf{1}^T\,, \tag{1.4d}$$

where $\mathbf{1}$ is a vector of a consistent dimension whose components are all ones. The assimilation process, for instance, can be stochastically performed as follows,

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{R} + \mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T\right]^{-1} \cdot \mathbf{\Gamma} \in \mathbb{R}^{n\times N}$$

Or, equivalently

$$\mathbf{X}^a = \mathbf{X}^b + \left[\left[\mathbf{P}^b\right]^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \mathbf{\Delta Y} \in \mathbb{R}^{n\times N}\,, \tag{1.5}$$

where $\mathcal{H}'(\mathbf{x}) \approx \mathbf{H}^T \in \mathbb{R}^{n\times m}$ is a linearized observation operator (with the linearization performed about the background state $\mathcal{H}(\mathbf{x}) \approx \mathcal{H}(\overline{\mathbf{x}}^b) + \mathbf{H} \cdot \left[\mathbf{x} - \overline{\mathbf{x}}^b\right]$), $\mathbf{\Delta Y} \in \mathbb{R}^{n\times m}$ is the scaled matrix of innovations on the synthetic observations,

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

where the e-th column of $\mathbf{\Gamma} \in \mathbb{R}^{m\times 1}$ is:

$$\mathbf{d}^{[e]} = \mathbf{y} + \boldsymbol{\varepsilon}^{[e]} - \mathcal{H}\left(\mathbf{x}^{b[e]}\right) \in \mathbb{R}^{m\times 1},\ \text{with } \boldsymbol{\varepsilon}^{[e]} \sim \mathcal{N}(\mathbf{0},\, \mathbf{R})\ .$$

and the columns of matrix $\mathbf{E} \in \mathbb{R}^{m\times N}$ are samples from a multivariate standard Normal distribution. Since current model resolutions range in the order of the millions while ensemble sizes does it in the hundreds, the ensemble covariance (1.4c) is typically rank-deficient. To counteract the effects of sampling noise, localization methods have been proposed [Kep00, NRMR18]. These have triggered the formulation and the implementation of efficient EnKF based methods [BT99, OHS+04, NRSA14]. A recent EnKF implementation relies on the Bickel and Levina estimator [BP14] in order to estimate background error correlations: the EnKF based on a modified Cholesky decomposition

(EnKF-MC) [NRSD15, NSD16, NRSD17, NRMC17]. In this method, for each model component $1 \leq i \leq n$, a neighbourhood $P(i\,, \delta)$ is defined based on its predecessors according to some labelling of model components and a radius of influence $\delta \in \mathbb{R}^+$, therefore [NRSD18],

$$j \in P(i,\, \delta) \Leftrightarrow d\left(x_i,\, x_j\right)^2 \leq \delta^2, \text{and } j < i\,,$$

where $d(\bullet,\, \bullet)$ denotes a consistent distance function. Notice, each model component is conditionally correlated only with its predecessors given all components, from here, a sparse precision matrix for the background distribution can be estimated as follows,

$$\widehat{\mathbf{B}}^{-1} = \widehat{\mathbf{V}}^T \cdot \widehat{\mathbf{\Gamma}}^{-1} \cdot \widehat{\mathbf{V}} \in \mathbb{R}^{n \times n}\,, \tag{1.6}$$

where the diagonal entries of the factor $\widehat{\mathbf{V}} \in \mathbb{R}^{n \times n}$ are all ones while its non-zero elements of row i are given by fitting models of the form,

$$\left[\mathbf{x}_{[i]}\right]^T = \sum_{j \in P(i,\, \delta)} -\left\{\widehat{\mathbf{V}}\right\}_{i,j} \cdot \left[\mathbf{x}_{[j]}\right]^T + \boldsymbol{\xi}^{[i]}\,, \text{ for } 1 \leq i \leq n\,,$$

where $\mathbf{x}_{[i]} \in \mathbb{R}^{N \times 1}$ denotes the i-th row of matrix, $\{\widehat{\mathbf{V}}\}_{i,j} \in \mathbb{R}$ is the $(i,\, j)$-th element of matrix $\widehat{\mathbf{V}}$, and the components of $\boldsymbol{\xi}^{[i]} \in \mathbb{R}^{N \times 1}$ are described by a zero-mean Normal distribution with unknown variance σ^2. Likewise, $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal entries are the empirical variances $\widehat{\mathbf{var}}(\bullet)$ of the residuals,

$$\left\{\widehat{\mathbf{\Gamma}}\right\}_{i,i} = \widehat{\mathbf{var}}\left(\left[\mathbf{x}_{[i]}\right]^T - \sum_{j \in P(i,\, \delta)} \left\{-\widehat{\mathbf{V}}\right\}_{i,j} \cdot \left[\mathbf{x}_{[j]}\right]^T\right) \approx \sigma^2\,, \tag{1.7}$$

for $2 \leq i \leq n$, with $\left\{\widehat{\mathbf{\Gamma}}\right\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[1]}\right)$. By replacing (1.6) in (1.5) the EnKF-MC is obtained. Another efficient EnKF implementation which exploits the structure of $\widehat{\mathbf{B}}^{-1}$ is the Posterior EnKF (P-EnKF) [NR17]. On its square root formulation, the P-EnKF approximates the posterior covariance matrix by a sequence of rank-one updates over the prior factors in (1.6),

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

where $\mathbf{V}^{(0)} \in \mathbb{R}^{n \times n}$ and $\mathbf{\Gamma}^{(0)} \in \mathbb{R}^{n \times n}$ are the Cholesky factors of $\widehat{\mathbf{B}}^{-1}$, and $\boldsymbol{\omega}^{[j]}$ is the j-th column of matrix $\mathbf{\Omega} = \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$, for $1 \leq j \leq m$. The posterior ensemble can then be simply built by drawing samples from the distribution,

$$
\mathbf{X}^a = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \left[\widehat{\mathbf{V}}^T \cdot \widehat{\mathbf{\Gamma}}^{1/2}\right]^{-1} \cdot \mathbf{E} \in \mathbb{R}^{n \times N}, \tag{1.8}
$$

where the columns of $\mathbf{E} \in \mathbb{R}^{n \times N}$ are formed by samples from a multivariate standard Normal distribution, note that, the inversion of matrix $\widehat{\mathbf{V}}^T \cdot \widehat{\mathbf{\Gamma}}^{1/2} \in \mathbb{R}^{n \times n}$ is not actually performed since $\widehat{\mathbf{V}}^T$ is an upper triangular matrix while $\widehat{\mathbf{\Gamma}}^{1/2}$ is diagonal and therefore, forward substitutions are sufficient in order to solve the subjacent linear system.

There are many other methods in the context of ensemble-based formulations for (nearly) linear observation operators, which we do not discuss further here. Recently, a complete survey of those is detailed by Bannister in [Ban16].

- **Ensemble Square-Root Filters (EnSRF):** The use of perturbed observations $\mathbf{y}^s$ makes the filter statistically consistent, but it induces sampling errors. The EnSRF are ensemble methods that do not make use of synthetic data. We match the moments as follow:

$$
\begin{aligned}
\overline{\mathbf{x}}^a &= \overline{\mathbf{x}}^b + \mathbf{\Delta X}^b \cdot \mathbf{V}^T \cdot \left[\mathbf{R} + \mathbf{V} \cdot \mathbf{V}^T\right]^{-1} \cdot \mathbf{d} \in \mathbb{R}^{n \times 1} \\
\mathbf{P}^a &= \left[\mathbf{I} - \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{R} + \mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T\right]^{-1} \cdot \mathbf{H}\right] \cdot \mathbf{P}^b \in \mathbb{R}^{n \times n} \\
\mathbf{X}^a &= \overline{\mathbf{x}}^a \otimes \mathbf{1}_N^T + \mathbf{\Delta X}^a \in \mathbb{R}^{n \times N}, \mathbf{P}^a = \frac{1}{N-1} \mathbf{\Delta X}^a \cdot \left[\mathbf{\Delta X}^a\right]^T.
\end{aligned}
$$

Where $\mathbf{\Delta y} = \mathbf{y} - \mathbf{H} \cdot \overline{\mathbf{x}}^b \in \mathbb{R}^{m \times 1}$. $\mathbf{V}$ is a matrix square root of the covariance matrix, and offer us a way to perform calculations without using the full precision matrix, preserving all its properties.

- **Issues of the ensemble-based methods**

  1. The ensemble size is constrained by the computational cost ($N \ll n$).

2. The sampling error impacts the estimation of the prior moments. Localization techniques can be exploited to reduce those effects, though those techniques are a matter of study yet.
3. Synthetic observations make the filter statistically consistent, but induces sampling errors.
4. Localization and inflation of the covariance matrix are mandatory steps. In practice, it is not clear how to set these parameters.

1.1.3 Variational Methods

Consider the 4D-Var cost function:

$$\mathcal{J}(\mathbf{x}_0) = \left\|\mathbf{x}_0 - \mathbf{x}_0^b\right\|^2_{\mathbf{B}_0^{-1}} + \sum_{k=0}^{G} \left\|\mathbf{y}_k - \mathcal{H}_k\left(\mathbf{x}_k\right)\right\|^2_{\mathbf{R}_k^{-1}} ,$$

then:

$$\begin{aligned} \mathbf{x}_0^a &= \arg\min_{\mathbf{x}_0} \mathcal{J}\left(\mathbf{x}_0\right) \equiv \mathbf{x}_0^a = \arg\max_{\mathbf{x}_0} \mathcal{P}\left(\mathbf{x}_0 | \left\{\mathbf{y}_k\right\}_{k=0}^{G}\right) , \\ \mathbf{x}_k &= \mathcal{M}_{k,k-1}\left(\mathbf{x}_{k-1}\right) \ \textbf{strong constrain} . \end{aligned}$$

The solution of this problem requires the use of adjoints: $\mathcal{M}_{k,k-1}(\mathbf{x}) \approx \mathcal{M}\left(\mathbf{x}_{k-1}\right) + \mathbf{M}_{k,k-1} \cdot [\mathbf{x} - \mathbf{x}_{k-1}]$. Adjoints are computationally expensive (at run) and sensitive to human errors. The validation of $\mathbf{M}_{k,k-1}$ can take several years.

1.1.4 Hybrid methods

To solve the optimization problem (1.3), we can employ adjoint models for gradient approximations. However, these models can be labor-intensive to develop and computationally expensive to run. For instance, the adjoint model of the High-Resolution Limited Area Modelling (HIRLAM) 4D-Var [GB14, SUL+09] was developed in 10 years in which most of the time was spent to detect and to fix errors in the tangent and the adjoint models [Gus07]. To avoid the use of adjoint models, we can employ an ensemble of model realizations [NRS15] as follows [HM98, SKW18]:

$$\mathbf{X}_k^b = \left[\mathbf{x}_k^{b[1]}, \mathbf{x}_k^{b[2]}, \ldots, \mathbf{x}_k^{b[N]}\right] \in \mathbb{R}^{n \times N} \tag{1.9}$$

where $\mathbf{x}_k^{b[e]} \in \mathbb{R}^{n \times 1}$ stands for the e-th ensemble member, for $1 \leq e \leq N$, at time k, for $0 \leq k \leq G$. Then, the ensemble mean:

$$\overline{\mathbf{x}}_k^b = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}_k^{b[e]} \in \mathbb{R}^{n \times 1} , \tag{1.10}$$

and the ensemble covariance matrix:

$$\mathbf{P}_k^b = \frac{1}{N-1} \cdot \Delta\mathbf{X}_k^b \cdot \left[\Delta\mathbf{X}_k^b\right]^T \in \mathbb{R}^{n \times n} , \tag{1.11}$$

act as estimates of the forecast state $\mathbf{x}_k^b$ and the forecast error covariance matrix $\mathbf{B}_k$, respectively, where the matrix of member deviations reads:

$$\Delta\mathbf{X}_k^b = \mathbf{X}_k^b - \overline{\mathbf{x}}_k^b \cdot \mathbf{1}^T \in \mathbb{R}^{n\times N}\,. \tag{1.12}$$

The model trajectory in (1.1) can be constrained to the space spanned by the background ensemble members (1.9), this is:

$$\mathbf{x}_k = \overline{\mathbf{x}}_k^b + \Delta\mathbf{X}_k \cdot \mathbf{w}\,, \tag{1.13}$$

where $\mathbf{w} \in \mathbb{R}^{N\times 1}$ is a vector in redundant coordinates to be determined later. This is equivalent to:

$$\mathbf{x}_k - \overline{\mathbf{x}}_k^b \in \mathbf{range}\left\{\Delta\mathbf{X}_k\right\} \approx \mathbf{range}\left\{\mathbf{B}_k^{1/2}\right\}\,,$$

therefore, the estimation of analysis increments is performed onto the sub-space given by a low-rank square root approximation of the background error covariance matrices (1.11) at observation times. By replacing (1.13) into the equation (1.1) one obtains:

$$\begin{aligned}\mathcal{J}(\mathbf{x}_0) &= \mathcal{J}\left(\overline{\mathbf{x}}_0^b + \Delta\mathbf{X}_0 \cdot \mathbf{w}\right)\\ &= \widehat{\mathcal{J}}(\mathbf{w}) = \frac{(N-1)}{2}\cdot\|\mathbf{w}\|^2 + \frac{1}{2}\cdot\sum_{k=0}^{G}\|\mathbf{d}_k - \mathbf{Q}_k\cdot\mathbf{w}\|^2_{\mathbf{R}_k^{-1}}\,,\end{aligned} \tag{1.14}$$

where $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\cdot\overline{\mathbf{x}}_k^b \in \mathbb{R}^{m\times 1}$ is the innovation vector and $\mathbf{Q}_k = \mathbf{H}_k\cdot\Delta\mathbf{X}_k^b \in \mathbb{R}^{m\times N}$. Note that, this cost function does not rely in the numerical model (1.2) anymore. The optimal value of the control variable $\mathbf{w}$ is then seek:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}}\widehat{\mathcal{J}}(\mathbf{w})\,. \tag{1.15}$$

The gradient of (1.14) equals:

$$\begin{aligned}\nabla_{\mathbf{w}}\widehat{\mathcal{J}}(\mathbf{w}) &= (N-1)\cdot\mathbf{w} - \sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot[\mathbf{d}_k - \mathbf{Q}_k\cdot\mathbf{w}]\\ &= \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{Q}_k\right]\cdot\mathbf{w} - \sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{d}_k \in \mathbb{R}^{N\times 1},\end{aligned} \tag{1.16}$$

and from here, the optimal weight (1.15) can be approximated as follows:

$$\mathbf{w}^* = \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{Q}_k\right]^{-1}\cdot\sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{d}_k \in \mathbb{R}^{N\times 1}, \tag{1.17}$$

from which the initial analysis state can be estimated:

$$\overline{\mathbf{x}}_0^a = \overline{\mathbf{x}}_0^b + \mathbf{\Delta X}_0^b \cdot \mathbf{w}^* \,. \tag{1.18}$$

Since in (1.13), $\mathbf{x}_k^a$ represents an approximation rather than an exact analysis trajectory, the initial analysis is recovered and then, it is evolved in time by using the numerical model (1.2) from which we obtain an estimate of the optimal trajectory of (1.3). Note that, all computations are performed onto the ensemble space (1.13) and therefore, the computational cost of estimating (1.18) is linearly bounded regadring n and m [NRS16]:

$$O\left(N \cdot n \cdot m + N^2 \cdot m\right) \,.$$

Readily, posterior (initial) members can be estimated via the implicit covariance matrix in (1.17):

$$\overline{\mathbf{x}}_0^{a(e)} = \overline{\mathbf{x}}_0^b + \mathbf{\Delta X}_0^b \cdot \mathbf{w}^{(e)}, \text{ for } 1 \leq e \leq N \,,$$

where:

$$\mathbf{w}^{(e)} \sim \mathcal{N}\left(\mathbf{w}^*, \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Q}_k\right]^{-1}\right) .$$

In practice, model dimensions range in the order of millions while ensemble sizes are constrained by the hundreds and as a direct consequence, undersampling degrades the quality of analysis corrections onto the space spanned by (1.12). To counteract the effects of sampling noise, localizations methods are commonly employed [GKM+11, CO10], in practice. For instance, methods such as covariance matrix localization (**B**-localization) [LWB18], domain localization, and observation localization (**R**-localization) [And01, HZS18, And19] are employed under operational DA scenarios. Yet another possible choice is to make use of precision covariance matrix estimation. In this context, for instance, the use of the *spatial-predecessors* concept can be employed to obtain sparse estimators of precision matrices [LRZ+08]. The predecessors of model component i, from now on $\Pi(i,\ \delta)$, for $1 \leq i \leq n$ and a radius of influence $\delta \in \mathbb{Z}^+$, are given by the set of components whose labels are lesser than that of the i-th one. Of course, this will depend on the format employed to label components on a numerical grid. In practice, column major and row major format are commonly employed. This idea is exploited in the EnKF formulation proposed in [NRSD17, NRSD18] wherein the following estimator is employed to approximate precision covariances [BL+08c]:

$$\widehat{\mathbf{B}}_k^{-1} = \widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1} \cdot \widehat{\mathbf{V}}_k \in \mathbb{R}^{n\times n}, \tag{1.19}$$

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

$$\left\{\widehat{\mathbf{V}}_k\right\}_{i,g} = \begin{cases} -\beta_{i,g,k} & , g \in P(i,\delta) \\ 1 & , i = g \\ 0 & , otherwise \end{cases}, \tag{1.20}$$

whose (sub-diagonal) elements $\beta_{i,g,k}$ are estimated by fitting linear models:

$$\mathbf{x}^T_{[i]_k} = \sum_{g \in \Pi(i,\,\delta)} \beta_{i,g,k} \cdot \mathbf{x}^T_{[g]_k} + \boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N\times 1}\,,\ 1 \le i \le n\,, \tag{1.21}$$

where $\mathbf{x}^T_{[i]_k} \in \mathbb{R}^{N\times 1}$ denotes the model component i from the ensemble (1.9). Likewise, $\boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N\times 1} \sim \mathcal{N}\left(\mathbf{0},\ \sigma^2 \cdot \mathbf{I}\right)$, where the variance σ^2_k is unknown, and the diagonal matrix $\boldsymbol{\Gamma}_k \in \mathbb{R}^{n\times n}$ holds the variance of residuals:

$$\{\boldsymbol{\Gamma}_k\}_{i,i} = \widehat{\mathbf{var}}\left(\mathbf{x}^T_{[i]_k} - \sum_{g \in \Pi(i,\,\delta)} \beta_{i,g,k} \cdot \mathbf{x}^T_{[j]_k}\right)^{-1} \tag{1.22}$$

$$\approx \mathbf{var}\left(\boldsymbol{\gamma}_{ik}\right)^{-1} = \frac{1}{\sigma^2_k} > 0\,,\ \text{with } \{\boldsymbol{\Gamma}_k\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}^T_{[1]_k}\right)^{-1}, \tag{1.23}$$

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

To summarize:

- Exploits the best of both ensemble-based methods, as it posses a flux dependent covariance (i.e., it mimics the model dynamic) and the robustness of the assimilation of many observations at once.
- States are limited to the ensemble space:

  $$\mathbf{x}_k = \overline{\mathbf{x}}_k + \boldsymbol{\Delta}\mathbf{X}^b_k \cdot \mathbf{w}\,,\ \text{where } \mathbf{P}^b_k = \frac{1}{N-1}\boldsymbol{\Delta}\mathbf{X}^b_k \cdot \left[\boldsymbol{\Delta}\mathbf{X}^b_k\right]^T,$$

  $\mathbf{w} \in \mathbb{R}^{N\times 1}$ belongs to the control space (ensemble space).
- The 4D-Var function in the ensemble espace, with $\mathbf{Q}_k = \mathbf{H}_k \cdot \boldsymbol{\Delta}\mathbf{X}^b_k$:

  $$\begin{aligned} \mathcal{J}(\mathbf{x}_0) &= \mathcal{J}\left(\overline{\mathbf{x}}^b_0 + \boldsymbol{\Delta}\mathbf{X}^b_0 \cdot \mathbf{w}\right) = \frac{N-1}{2} \cdot \|\mathbf{w}\|^2 \\ &+ \sum_{k=0}^{G} \|\mathbf{d}_k - \mathbf{Q}_k \cdot \mathbf{w}\|^2_{\mathbf{R}^{-1}_k}\,,\ \text{with } \boldsymbol{\Delta}\mathbf{y}_k = \mathbf{y}_k - \mathcal{H}\left(\overline{\mathbf{x}}^b_k\right). \end{aligned}$$

- The optimization problem reads:

  $$\mathbf{w}^a = \arg\min_{\mathbf{w}} \mathcal{J}\left(\overline{\mathbf{x}}^b_0 + \boldsymbol{\Delta}\mathbf{X}^b_0 \cdot \mathbf{w}\right).$$

which **do not require adjoints!**, so, the solution is:

$$\left[(N-1)\cdot\mathbf{I}+\sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{Q}_k\right]\cdot\mathbf{w}^a=\left[\sum_{k=0}^{G}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{d}_k\right],$$

estimation: $\mathbf{x}_0^a=\mathbf{x}_0^b+\mathbf{\Delta X}_0^b\cdot\mathbf{w}^a$.

- This methods are sensitive to sampling errors ($n \gg N$). Localization of $\mathbf{\Delta X}_k^b$ is not trivial.

1.2 Wind Potential Estimation

The effects of climate change have triggered alarms to employ alternatives and to reduce Carbon Dioxide ($CO2$) emissions around the world. In many countries, regulation and $CO2$ reduction goals promote the substitution of fossil energy sources with Renewable Energy Sources (RES) [KST17]. For instance, China, the largest energy consumer worldwide, has an economic motivation to execute such substitution [Liu17]: traditional power systems (mainly composed of nuclear, hydro, and thermal generators) are drastically decreasing, and now, they are trying to integrate RES as a shock absorber of this situation. However, RES integration is not straightforward since it brings new issues and challenges that need to be analyzed and addressed. One of the main challenges comes from the intermittency of RES [VDVDH17]. Intermittency combines variability and uncertainty. The former is produced by the movement of large cloud systems owing to high and low-pressure areas. Uncertainty, also known as unpredictability, comes from the forecast error, which in turn depends on the numerical model (1.2). Thus, uncertainty amplification relies on model errors (i.e., physics simplifications to make it computationally feasible). For instance, if the accuracy of the numerical model is poor, and no Data Assimilation is performed, the bias on the resulting estimate will be large concerning the actual wind speed. Thus, wind speeds can be poorly estimated, and as a direct consequence, wind energy potentials can be underestimated. Hence, Data Assimilation can be employed in this context to mitigate the impact of poor potential energy estimations via real noisy observations of wind speeds. The potential energy $p(v)$ in MegaWatts (MW) of a wind turbine given a wind speed v (km/h) can be estimated as follows:

$$p(v)=\begin{cases}P_{nom}\cdot\left(\frac{v^3-v_c^3}{v_r^3-v_c^3}\right) & v_c\le v\le v_r\\ P_{nom} & v_r\le v\le v_f\\ 0 & \text{otherwise}\end{cases},\tag{1.24}$$

where v_c, v_r, v_f, and v_p are the cut-in wind speed, the rated wind speed, the cut-out wind speed, and the rated power of wind turbine, respectively. Table 1.1 shows the 12 wind turbine generators types assumed and utilized in many case studies [XB10]. The outage rate of each wind turbine reads 0.04. Commonly, the useful life of a wind turbine is about 25 years; this does not depend on its size. It is ubiquitous to assume an interest rate of 0.08 for turbines. We also report the capital cost, and the maintenance and operating cost for each turbine, these are taken from [W$^+$08, Mas13].

Type	R.C. (MW)	v_c (km/h)	v_r (km/h)	v_f (km/h)	C.C.	$M\&O$
1	0.5	10	40	80	1350	36
2	0.5	10	45	70	1350	36
3	1	12	40	80	1250	35
4	2	12	30	55	1120	30
5	1	13	33	60	1220	33
6	1	14	40	90	1250	32
7	2	15	33	50	1100	35
8	2	15	33	60	1100	30.5
9	1	15	37	70	1200	32
10	1	18	48	70	1250	32
11	2	18	45	70	1100	30
12	2	18	35	75	1100	30

Table 1.1: WTG Unit Parameters

1.3 Problem Statement

The previous cutting-edge research throw the following issues into the Data Assimilation community:

1.3.1 Current Limitations

1. Conceptual/Theoretical:
    - Meteorological Simulations demand high computational efforts.
    - Ensemble-based methods are highly sensitive to sampling noise.
    - Variational methods demand the use of adjoint models, which are expensive to run and hard to develop. They are very sensitive to human errors. Years can be taken for their proper validation.
    - Hybrid methods can be sensitive to sampling noise since ensemble sub-space is employed to estimate analysis increments.
    - High model resolutions can difficult the analysis of wind components in the Colombian national territory.
2. Operational:
    - No real-time is accessible in Colombia. Data is commonly requested by governmental agencies (INVEMAR e IDEAM), and then it is available weeks later.
    - There are no numerical models to estimate wind components in Colombia (and to exploit our knowledge about our ecosystems).

1.4 Expected Main Contributions

To be concise, we expect with our research:

- to develop an Adjoint-Free 4D-Var method which mitigates the impact of sampling noise (advance the cutting-edge of data assimilation),
- to reduce the uncertainty in wind-potential-energy forecasts, and
- to develop Colombian maps with wind-potential-energy potential estimations per regions constrained by numerical model resolutions.

1.5 Objectives

1.5.1 Main Objective

To design and implement a Four-Dimensional Variational Data Assimilation method for wind-energy-potential estimation.

1.5.2 Specific Objectives

1. To design and implement an Adjoint-Free Four-Dimensional Variational Data Assimilation method for wind component estimations.
2. To map wind-fields to wind-speeds and wind-energy potentials.
3. To validate the proposed method by employing real and synthetic observations.
4. To analyze, statistically, the wind-energy potentials for the employed wind turbines.

1.6 Methodology

In this section, we briefly describe the steps to accomplish the objectives in Section .

1. To define a control space via a modified Cholesky decomposition.
   - To build an ensemble of model realizations via an Atmospheric General Circulation Model (AT-GCM).
   - To estimate square root approximation of background error covariances via a modified Cholesky decomposition.
2. To implement the proposed method via a scientific computational language.
3. To employ synthetic data and real data (from the National Oceanic and Atmospheric Administration - NOAA [CTW$^+$94]) to validate the proposed variational method.
   - Data can be downloaded from the NOAA website, for free, via FTP.
   - Space interpolation is needed to adjust the NOAA data to the SPEEDY model.

4. To map wind-fields to wind-potential estimations via non-linear mappings.
5. To design Colombian maps for wind-potential estimations and different wind turbines.

1.7 Advances - Proposed Method: Hybrid 4D-Var DA Method for Wind Energy Potential Estimation in Colombia

In this section, we show our advances in the proposed adjoint-free 4D-Var method for potential energy estimation. The variational filter is divided into four stages. First, we build an ensemble of snapshots at assimilation times. Second, these snapshots are employed to build control spaces via a modified Cholesky decomposition. Third, analysis increments are computed onto control spaces. Lastly, wind components are employed to estimate wind speeds, which allow us to approximate potential energies of wind turbines. We also develop a matrix-free analysis formulation to avoid the direct inversion of linear systems during assimilation steps. All these stages are clearly detailed next.

1.7.1 Building an Ensemble of Snapshots

Snapshots of an ensemble of model realizations (1.9) are taken at $G+1$ observation times. At step k, for $0 \leq k \leq G$. The background ensemble $\mathbf{X}_k^b$ is then employed to estimate a full-rank square-root approximation of the precision covariance matrix $\mathbf{B}_k^{-1}$ via a modified Cholesky decomposition (1.19):

$$\widehat{\mathbf{B}}_k^{-1/2} = \widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1/2} \in \mathbb{R}^{n \times n} . \tag{1.25}$$

This square-root estimation serves as a control space onto which analysis increments can be estimated:

$$\mathbf{x}_k - \overline{\mathbf{x}}_k^b \in \mathbf{range} \left\{ \widehat{\mathbf{B}}_k^{1/2} \right\} ,$$

or equivalently:

$$\mathbf{x}_k = \overline{\mathbf{x}}_k^b + \widehat{\mathbf{B}}_k^{1/2} \cdot \boldsymbol{\alpha} \in \mathbb{R}^{n \times 1} , \tag{1.26}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{n \times 1}$ is a vector in redundant coordinates to be determined later. We assume that:

$$\mathbf{range} \left\{ \left[\widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1/2} \right]^{-1} \right\} \approx \mathbf{range} \left\{ \mathbf{B}_k^{1/2} \right\} .$$

Note that, since the square root approximations (1.25) are full-rank, the dimension of the spaces (1.26) equal those of the range of $\mathbf{B}^{1/2}$. We then expect to capture all error dynamics onto the spaces (1.25). Since the initial background error covariance matrix $\mathbf{B}_0$ onto the control space (1.26) is nothing but the identity matrix, the following error statistics hold for the prior weights $\boldsymbol{\alpha}^{b(e)}$:

$$\boldsymbol{\alpha}^{b(e)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) , \text{ for } 1 \leq e \leq N .$$

Due to this, the 4D-Var cost function (1.1) onto the space (1.26) can be written as follows:

$$\mathcal{J}(\mathbf{x}_0) \quad = \quad \mathcal{J}\left(\overline{\mathbf{x}}_0^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}\right) = \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) = \frac{1}{2} \cdot \|\boldsymbol{\alpha}\|^2 + \frac{1}{2} \cdot \sum_{k=0}^{G} \left\| \widetilde{\mathbf{d}}_k - \widetilde{\mathbf{Q}}_k \cdot \boldsymbol{\alpha} \right\|_{\mathbf{R}_k^{-1}}^2 , \quad (1.27)$$

where $\widetilde{\mathbf{d}}_k = \mathbf{y}_k - \mathbf{H} \cdot \overline{\mathbf{x}}_k^b \in \mathbb{R}^{m \times 1}$, and $\widetilde{\mathbf{Q}}_k = \mathbf{H} \cdot \widehat{\mathbf{B}}_k^{1/2} \in \mathbb{R}^{m \times n}$. Again, this cost function does not rely on the numerical model (1.2).

1.7.2 Adjoint-Free 4D-Var Optimization

Consider the cost function (1.27), the adjoint-free optimization problem to solve reads:

$$\boldsymbol{\alpha}^a = \arg\min_{\boldsymbol{\alpha}} \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) \, . \quad (1.28)$$

The gradient of the cost function (1.27) can be written as follows:

$$\nabla_{\boldsymbol{\alpha}} \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) = \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right] \cdot \boldsymbol{\alpha} - \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{d}}_k \, ,$$

whose root reads:

$$\boldsymbol{\alpha}^a = \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1} \cdot \left[\sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{d}}_k \right] , \quad (1.29)$$

and therefore an estimate of the initial analysis state (1.3) can be computed as follows:

$$\overline{\mathbf{x}}_0^a = \overline{\mathbf{x}}_0^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}^a \, ,$$

whose model trajectory provides a forecast which accounts for the given data into the assimilation window. The posterior ensemble onto the control space can then be built by using a square root approximation of the information matrix, its can be easily shown that the posterior error statistics read:

$$\boldsymbol{\alpha}^{a[e]} \sim \mathcal{N}\left(\boldsymbol{\alpha}^a, \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1} \right) , \text{ for } 1 \leq e \leq N \, , \quad (1.30)$$

with corresponding analysis members in the model space:

$$\mathbf{x}_0^{a[e]} = \overline{\mathbf{x}}^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}^{a[e]} \, .$$

Given the special structure of our estimator $\mathbf{B}_k^{-1}$ woodbury matrix identities can be exploited to avoid direct inversions [RSA15]. We denote this filter implementation *Four Dimensional Variational Data Assimilation via a Modified Cholesky Decomposition* (4D-Var-MC).

1.7.3 Matrix-Free Formulation of the 4D-Var-MC

Following the ideas discussed in [NR17], we can develop a matrix-free equation for the analysis step of the 4D-Var-MC implementation. We can proceed as follows, consider:

$$\Omega = [\Omega_0,\ \Omega_1,\ \ldots,\ \Omega_G] \in \mathbb{R}^{n\times O}$$

where $\Omega_k = \widetilde{\mathbf{Q}}_k \cdot \mathbf{R}_k^{-1/2} \in \mathbb{R}^{n\times m}$, and $O = m \cdot G$, the precision covariance matrix in (1.30) can be written as follows,

$$\widehat{\mathbf{A}}^{-1} = \mathbf{I} + \Omega \cdot \Omega^T = \mathbf{T}^T \cdot \mathbf{C} \cdot \mathbf{T} + \Omega \cdot \Omega^T = \mathbf{T}^T \cdot \mathbf{C} \cdot \mathbf{T} + \sum_{o=1}^{O} \cdot \boldsymbol{\omega}^{[o]} \cdot \left[\boldsymbol{\omega}^{[o]}\right]^T , \tag{1.31}$$

where $\boldsymbol{\omega}^{[o]} \in \mathbb{R}^{n\times 1}$ is the o-th column of matrix Ω, for $0 \le o \le O$, and $\mathbf{I} = \mathbf{T}^T \cdot \mathbf{C} \cdot \mathbf{T}$ is the Cholesky decomposition of $\mathbf{I}$ (all factors equal the identity matrix). Consider the sequence of matrices,

$$\begin{aligned}
\widehat{\mathbf{A}}^{(0)} &= \left[\mathbf{V}^{(0)}\right]^T \cdot \boldsymbol{\Gamma}^{(0)} \cdot \left[\mathbf{V}^{(0)}\right] = \mathbf{T}^T \cdot \mathbf{C} \cdot \mathbf{T} = \mathbf{I}, \\
\widehat{\mathbf{A}}^{(1)} &= \widehat{\mathbf{A}}^{(0)} + \boldsymbol{\omega}^{[1]} \cdot \left[\boldsymbol{\omega}^{[1]}\right]^T = \left[\mathbf{V}^{(1)}\right]^T \cdot \boldsymbol{\Gamma}^{(1)} \cdot \left[\mathbf{V}^{(1)}\right], \\
\widehat{\mathbf{A}}^{(2)} &= \widehat{\mathbf{A}}^{(1)} + \boldsymbol{\omega}^{[2]} \cdot \left[\boldsymbol{\omega}^{[2]}\right]^T = \left[\mathbf{V}^{(2)}\right]^T \cdot \boldsymbol{\Gamma}^{(2)} \cdot \left[\mathbf{V}^{(2)}\right], \\
&\vdots \\
\widehat{\mathbf{A}}^{(O)} &= \widehat{\mathbf{A}}^{(O-1)} + \boldsymbol{\omega}^{[O]} \cdot \left[\boldsymbol{\omega}^{[O]}\right]^T = \left[\mathbf{V}^{(O)}\right]^T \cdot \boldsymbol{\Gamma}^{(O)} \cdot \left[\mathbf{V}^{(O)}\right] = \widehat{\mathbf{V}}^T \cdot \widehat{\boldsymbol{\Gamma}} \cdot \widehat{\mathbf{V}} = \widehat{\mathbf{A}}^{-1},
\end{aligned}$$

where $\mathbf{V}^{(0)} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Gamma}^{(0)} \in \mathbb{R}^{n\times n}$ are the factors of the Cholesky decomposition of the identity matrix $\mathbf{I}$. Among iteration steps o, for $0 \le o \le O$, one can see that:

$$\begin{aligned}
\widehat{\mathbf{A}}^{(o)} &= \left[\mathbf{V}^{(o-1)}\right]^T \cdot \boldsymbol{\Gamma}^{(o-1)} \cdot \left[\mathbf{V}^{(o-1)}\right] + \boldsymbol{\omega}^{[o]} \cdot \left[\boldsymbol{\omega}^{[o]}\right]^T \\
&= \left[\mathbf{V}^{(o-1)}\right]^T \cdot \left[\boldsymbol{\Gamma}^{(o-1)} + \boldsymbol{\gamma}^{(o)} \cdot \left[\boldsymbol{\gamma}^{(o)}\right]^T\right] \cdot \left[\mathbf{V}^{(o-1)}\right],
\end{aligned} \tag{1.32}$$

where $\left[\mathbf{V}^{(j-1)}\right]^T \cdot \boldsymbol{\gamma}^{(j)} = \boldsymbol{\omega}^{[j]} \in \mathbb{R}^{n\times 1}$. Via the Cholesky factors of,

$$\boldsymbol{\Gamma}^{(o-1)} + \boldsymbol{\gamma}^{(o)} \cdot \left[\boldsymbol{\gamma}^{(o)}\right]^T = \left[\widetilde{\mathbf{V}}^{(o-1)}\right]^T \cdot \boldsymbol{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)}\right], \tag{1.33}$$

and, by considering equation (1.32), the matrix $\widehat{\mathbf{A}}^{(j)}$ can be decomposed as follows,

$$\widehat{\mathbf{A}}^{(o)} = \left[\widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}\right]^T \cdot \boldsymbol{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}\right] = \left[\mathbf{V}^{(o)}\right]^T \cdot \boldsymbol{\Gamma}^{(o)} \cdot \left[\mathbf{V}^{(o)}\right],$$

where $\mathbf{V}^{(o)} = \widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)} \in \mathbb{R}^{n\times n}$. By taking a close look at equation (1.33), the elements of factors $\widetilde{\mathbf{V}}^{(o-1)}$ and $\widetilde{\boldsymbol{\Gamma}}^{(o)}$ can be easily related to those of $\boldsymbol{\Gamma}^{(o-1)}$ and $\boldsymbol{\gamma}^{(o)}$ via the Dolittle's method for matrix factorization, for instance, we can note that:

$$\left[\widetilde{\mathbf{V}}^{(o-1)}\right]^T \cdot \boldsymbol{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)}\right]_{i,b} = \delta_{i,b} \cdot \boldsymbol{\Gamma}_{i,i}^{(o-1)} + \boldsymbol{\gamma}_i^{(o)} \cdot \boldsymbol{\gamma}_b^{(o)},$$

and therefore, the next relations hold:

$$\mathbf{\Gamma}_{n,n}^{(o)} = \left[\boldsymbol{\gamma}_n^{(o)}\right]^2 + \mathbf{\Gamma}_{n,n}^{(o-1)}, \tag{1.34a}$$

$$\widetilde{\mathbf{V}}_{i,b}^{(o-1)} = \frac{1}{\mathbf{\Gamma}_{i,i}^{(o)}} \cdot \left[\boldsymbol{\gamma}_i^{(o)} \cdot \boldsymbol{\gamma}_b^{(o)} - \sum_{q \in \Pi(i,\,\delta)} \mathbf{\Gamma}_{q,q}^{(o)} \cdot \widetilde{\mathbf{V}}_{q,i}^{(o-1)} \cdot \widetilde{\mathbf{V}}_{q,b}^{(o-1)}\right], \tag{1.34b}$$

and

$$\mathbf{\Gamma}_{i,i}^{(o)} = \left[\boldsymbol{\gamma}_i^{(o)}\right]^2 + \mathbf{\Gamma}_{i,i}^{(o-1)} - \sum_{q \in \Pi(i,\,\delta)} \mathbf{\Gamma}_{q,q}^{(o)} \cdot \left[\widetilde{\mathbf{V}}_{q,i}^{(o-1)}\right]^2, \tag{1.34c}$$

for $1 \leq i \leq n-1$ and $b \in \Pi(i,\,\delta)$, where the Kronecker delta function $\delta_{i,j}$ equals 1 for $i = j$ and 0 otherwise, and the diagonal entries of matrix $\widetilde{\mathbf{V}}^{(o-1)}$ are all equal to one. In algorithm 1, we show how the Cholesky factors $\mathbf{V}^{(o)}$ and $\mathbf{\Gamma}^{(o)}$ can be updated with the information brought by the rank-one matrix $\boldsymbol{\omega}^{[o]} \cdot \left[\boldsymbol{\omega}^{[o]}\right]^T$, the general updating process of factors $\widehat{\mathbf{V}}$ and $\mathbf{\Gamma}$ for the estimation of $\widehat{\mathbf{A}}^{-1}$ are detailed in algorithm 2. The number of long computations is reported as well for each step of our proposed updating process. We let by φ the largest number of non-zero elements per row in the $\mathbf{V}^{(o)}$ factor. This value will depend on chosen radius of influence δ during assimilation steps, and intuitively $\varphi \ll n$. Note that, $\mathbf{V}^{(o-1)}$ and $\widetilde{\mathbf{V}}^{(o-1)}$ hold the same structure since this is given by the predecessors of k. Thus, the structure (form) of $\mathbf{V}^{(o-1)}$ is preserved in $\mathbf{V}^{(o)}$. Consequently, we can hold a desired structure in the resulting estimator $\widehat{\mathbf{V}}$, for instance, this can be equal to that of $\mathbf{V}$. Note that, the number of long computations in the 4D-Var-MC reads:

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

Note that, this bound increases linearly regarding the number of model components.

Algorithm 1 Rank-one update of factors $\mathbf{V}^{(o-1)}$ and $\mathbf{\Gamma}^{(o-1)}$ via Doolittle's method.

function UPDATE_RANK_ONE($\mathbf{V}^{(o-1)}$, $\mathbf{\Gamma}^{(o-1)}$, $\boldsymbol{\omega}^{[o]}$) ▷ **COST**
  Solve $\left[\mathbf{V}^{(o-1)}\right]^T \cdot \mathbf{p}^{(o)} = \boldsymbol{\omega}^{[o]}$. ▷ $\mathcal{O}(\varphi \cdot n)$
  Compute $\mathbf{\Gamma}_{n,n}^{(o)}$ via equation (1.34a). ▷ $\mathcal{O}(1)$
  for $i = n-1 \rightarrow 1$ **do** ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
    Let $\widetilde{\mathbf{V}}_{i,i}^{(o-1)} \leftarrow 1$. ▷ $\mathcal{O}(1)$
    for $k \in \Pi(i,\,\delta)$ **do** ▷ $\mathcal{O}\left(\varphi^2\right)$
      Compute $\widetilde{\mathbf{V}}_{i,k}^{(o-1)}$ according to (1.34b). ▷ $\mathcal{O}(\varphi)$
    end for
    Compute $\mathbf{\Gamma}_{i,i}^{(o)}$ via equation (1.34c). ▷ $\mathcal{O}(\varphi)$
  end for
  Let $\mathbf{V}^{(o)} \leftarrow \widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}$. ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
  return $\mathbf{V}^{(o)}$, $\mathbf{\Gamma}^{(o)}$
end function

Algorithm 2 Computing the posterior factors $\widehat{\mathbf{V}}$ and $\boldsymbol{\Gamma}$ of $\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{V}}^T \cdot \boldsymbol{\Gamma} \cdot \widehat{\mathbf{V}} = \mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k$.

1: **function** COMPUTE_POSTERIOR_CHOLESKY_FACTORS($\mathbf{V}^{(0)}$, $\boldsymbol{\Gamma}^{(0)}$, $\mathbf{H}$, $\mathbf{R}$) ▷ **COST**
2: Let $\boldsymbol{\Omega} \leftarrow \left[\mathbf{H}_0^T \cdot \mathbf{R}_0^{-1/2}, \mathbf{H}_1^T \cdot \mathbf{R}_1^{-1/2}, \ldots, \mathbf{H}_G^T \cdot \mathbf{R}_G^{-1/2}\right]$. ▷ $\mathcal{O}\,(O \cdot n)$
3: Let $O \leftarrow m \cdot G$
4: **for** $j = 1 \rightarrow O$ **do** ▷ O times line 4, $\mathcal{O}\left(\varphi^2 \cdot O \cdot n\right)$
5: Let $[\mathbf{V}^{(j)}, \boldsymbol{\Gamma}^{(j)}] \leftarrow$ UPDATE_RANK_ONE $\left(\mathbf{V}^{(j-1)}, \boldsymbol{\Gamma}^{(j-1)}, \boldsymbol{\omega}^{[j]}\right)$ ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
6: **end for**
7: **return** $\mathbf{V}^{(O)}$ as $\widehat{\mathbf{V}}$, $\boldsymbol{\Gamma}^{(O)}$ as $\boldsymbol{\Gamma}$.
8: **end function**

1.7.4 Post-Processing of Data, Potential Energy Estimation

Once the model trajectory is computed for each ensemble member, we proceed to map wind fields to wind energy potentials in two steps:

1. whenever is necessary, the wind components of ensemble members are mapped to wind-speeds,

2. this subset of information is exploited to estimate the wind energy potential of each analysis ensemble member:

$$\widehat{\mathbf{x}}_k^{a(e)} = f\left(\mathbf{x}_k^{a(e)}\right), \text{ for } 1 \leq e \leq N, \text{ and } 0 \leq k \leq G, \tag{1.35}$$

where $f : \mathbb{R}^{n\times 1} \rightarrow \mathbb{R}^{h\times 1}$ is a function that maps model states to potential energy states (this is, for each ensemble member, its wind-speed components are mapped to wind energy potentials), where h is the number of wind-speed components (with $h \leq n$), and $\widehat{\mathbf{x}}_k^{a(e)} \in \mathbb{R}^{h\times 1}$ is the k-th transformed member. The mapping process depends on the wind turbine employed.

Once ensemble members are mapped to wind energy potentials, empirical moments of these samples can be exploited to estimate mean and standard deviations of wind powers. Besides, covariances of such samples can be estimated via a modified Cholesky decomposition to understand better (and to estimate) their uncertainties.

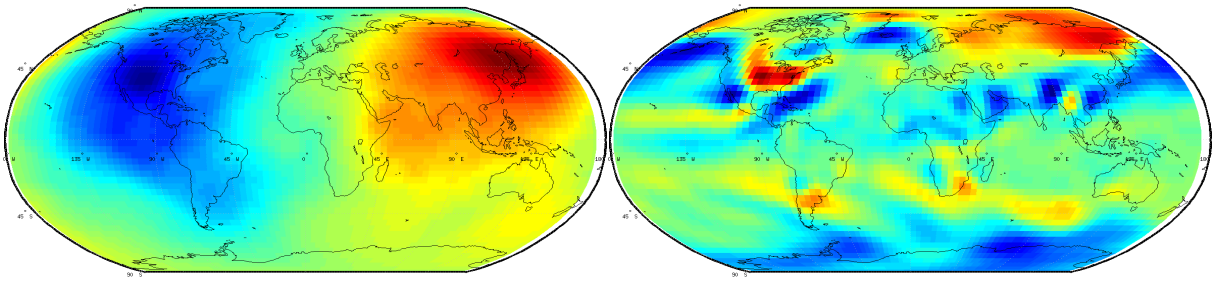
1.7.5 Surrogate Model to be Employed for Wind Speed Estimation

The SPEEDY model is a general circulation model that mimics the behavior of the atmosphere across different pressure levels [BKKM04, Miy11]. The number of numerical layers in this model is 7, and we employ a T-30 spectral model resolution (96×48 grid components) for the space discretization of each model layer [Mol03, KMB06]. The number of physical variables is 5. These are detailed in the Table 1.2 with their corresponding units and number of layers.

Name	Notation	Units	Number of Layers
Temperature	T	K	7
Zonal Wind Component	u	m/s	7
Meridional Wind Component	v	m/s	7
Specific Humidity	Q	g/kg	7
Pressure	T	K	1

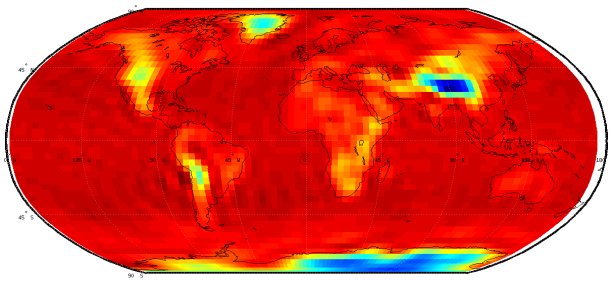
Table 1.2: Physical variables of the SPEEDY model.

Note that the total number of model components to be estimated reads $n = 133,632$. We let the number of model realizations (ensemble size) as $N = 30$ for all experimental scenarios. Besides, the SPEEDY model has more than 15 post-processing physical variables that can be employed for further research. Some model snapshots are shown in figure 1.4.



(a) Chi

(b) Omega



(c) Pressure

Figure 1.4: Some post-processing model variables

1.8 Organization of this document

This thesis is organized as follows: in Chapter 2, we discuss an efficient ensemble Kalman filter implementation based on a modified Cholesky decomposition. Chapter 3 discusses an efficient

and matrix-free implementation of the ensemble Kalman filter based on a posterior sampling method. In Chapter 4, the space spanned by the proposed ensemble based formulations is exploited to obtain matrix-free Four-Dimensional-Variational optimization problems and to estimate wind potential at different regions of the world. Conclusions of this research are stated in Chapter 5.

2 A Posterior EnKF Based On A Modified Cholesky Decomposition

In this chapter, we propose a posterior ensemble Kalman filter (EnKF) based on a modified Cholesky decomposition. The main idea behind our approach is to estimate the moments of the analysis distribution based on an ensemble of model realizations. The method proceeds as follows: initially, an estimate of the precision background error covariance matrix is computed via a modified Cholesky decomposition and then, based on rank-one updates, the Cholesky factors of the inverse background error covariance matrix are updated in order to obtain an estimate of the inverse analysis covariance matrix. The special structure of the Cholesky factors can be exploited in order to obtain a matrix-free implementation of the EnKF. Once the analysis covariance matrix is estimated, the posterior mode of the distribution can be approximated and samples about it are taken in order to build the posterior ensemble. Experimental tests are performed making use of the Lorenz 96 model in order to assess the accuracy of the proposed implementation. The results reveal that, the accuracy of the proposed implementation is similar to that of the well-known local ensemble transform Kalman filter and even more, the use of our estimator reduces the impact of sampling errors during the assimilation of observations.

To be concise, Data Assimilation is the process by which imperfect numerical forecasts are adjusted according real noisy observations [Eve03]. In practice, Gaussian errors are commonly assumed for background and observational errors during the assimilation of observations [Eve06]. While observational errors can be well-estimated in the context of operational data assimilation, background error correlations can be hard to approximate, mainly, owing to the size of the vector state which, typically, ranges in the order of millions [Zup09]. In the ensemble Kalman filter (EnKF), an ensemble of model realizations is utilized in order to estimate the moments of the underlying background error distribution [GMC$^+$06]. Since the ensemble size is constrained by computational aspects, localization methods can be utilized in order to mitigate the impact of sampling errors [JFW14, And12]. Efficient formulations of the EnKF account for some sort of implicit localization during the analysis step in order to damp out spurious correlations [Kep00, TAB$^+$03, AA99, NRSA14]. For instance, in the EnKF based on a modified Cholesky decomposition [BL08b], a band estimate of the inverse background error covariance matrix can be obtained, on the fly, during the assimilation of observations. Even more, this precision matrix can be expressed in terms of Cholesky factors which are composed by a diagonal matrix and a band lower triangular matrix. We think that, these factors can be updated in order to estimate the Cholesky factors of the inverse analysis covariance matrix. This covariance matrix can be estimated by applying a series of rank-one updates on the Cholesky factors of the inverse background error covariance matrix. With such covariance matrix, samples from the posterior error distribution can be approximately taken with low-computational efforts.

This paper is organized as follows: in section 2.1 efficient EnKF formulation are discussed, section 2.2 presents the proposed EnkF implementation, in section 2.3 the accuracy of the proposed EnKF is assessed and compared with well-known EnKF formulations and finally, conclusions are stated in section 2.4.

2.1 Preliminaries

In this section, efficient EnKF implementations in order to avoid the impact of sampling errors on the analysis innovations are discussed. Inflation aspects and other sources of misestimation of model states and ensemble collapsing are well-studied in [LKM09, And07].

In the ensemble Kalman filter, an ensemble of model realizations,

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \ldots, \mathbf{x}^{b[N]}\right] \in \mathbb{R}^{n\times N}, \tag{2.1}$$

is utilized in order to estimate, the moments of the background error distribution,

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

via the empirical moments of the ensemble (2.1), therefore,

$$\mathbf{x}^b \approx \overline{\mathbf{x}}^b = \frac{1}{N}\cdot\sum_{i=1}^{N}\mathbf{x}^{b[i]} \in \mathbb{R}^{n\times 1}, \tag{2.2a}$$

and

$$\mathbf{B} \approx \mathbf{P}^b = \frac{1}{N-1}\cdot\Delta\mathbf{X}\cdot\Delta\mathbf{X}^T \in \mathbb{R}^{n\times n}, \tag{2.2b}$$

where n is the model dimension, N is the ensemble size, $\mathbf{x}^{b[i]} \in \mathbb{R}^{n\times 1}$ is the i-th ensemble member, for $1 \leq i \leq N$, $\mathbf{x}^b \in \mathbb{R}^{n\times 1}$ is well-known as the background state while $\mathbf{B} \in \mathbb{R}^{n\times n}$ stands for background error covariance matrix, $\overline{\mathbf{x}}^b$ is the ensemble mean, and $\mathbf{P}^b$ is the ensemble covariance matrix. Likewise, the matrix of member deviations $\Delta\mathbf{X} \in \mathbb{R}^{n\times N}$ reads,

$$\Delta\mathbf{X} = \mathbf{X}^b - \overline{\mathbf{x}}^b\cdot\mathbf{1}_N^T.\Delta\mathbf{X} \in \mathbb{R}^{n\times N} \tag{2.3}$$

When an observation $\mathbf{y} \in \mathbb{R}^{m\times 1}$ is available, the analysis ensemble can be computed as follows,

$$\mathbf{X}^a = \mathbf{X}^b + \Omega \in \mathbb{R}^{n\times N}, \tag{2.4}$$

where $\Omega \in \mathbb{R}^{n\times N}$ can be obtained by the solution of the linear system of equations,

$$\left[\left[\mathbf{P}^b\right]^{-1} + \mathbf{H}^T\cdot\mathbf{R}^{-1}\cdot\mathbf{H}\right]\cdot\Omega = \mathbf{H}^T\cdot\mathbf{R}^{-1}\cdot\Delta\mathbf{Y} \tag{2.5}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ is a linearized observational operator, $\mathbf{R} \in \mathbb{R}^{m\times m}$ is the estimated data-error covariance matrix, the matrix of innovations on the observations $\Delta\mathbf{Y} \in \mathbb{R}^{m\times N}$ reads,

$$\Delta\mathbf{Y} = \mathbf{y}\cdot\mathbf{1}_N^T + \mathbf{E} - \mathbf{H}\cdot\mathbf{X}^b, \tag{2.6}$$

and the columns of $\mathbf{E} \in \mathbb{R}^{m \times N}$ are samples from a zero-mean Normal distribution with covariance matrix $\mathbf{R}$. The forecast is the approximated by propagating the ensemble (2.4) until new observations are available,

$$\mathbf{x}_{\text{next}}^{b[i]} = \mathcal{M}_{t_{\text{current}} \to t_{\text{next}}} \left(\mathbf{x}_{\text{current}}^{a[i]} \right), \text{ for } 1 \leq i \leq N, \tag{2.7}$$

where $\mathbf{x}^{a[i]} \in \mathbb{R}^{n \times 1}$ denotes the i-th analysis member, and $\mathcal{M}_{\text{current} \to \text{next}} : \mathbb{R}^{n \times 1} \to \mathbb{R}^{n \times 1}$ is an imperfect numerical model (i.e., a model which mimics the behaviour of the ocean and/or atmosphere).

In operational data assimilation, ensemble members comes at high computational efforts [Ler07, LR99] and therefore, the ensemble moments (2.2) are corrupted by sampling noise [EWH13]. Hence, localization methods are commonly utilized in the EnKF context in order to mitigate the impact of sampling errors. One of the best EnKF implementations is the local ensemble transform Kalman filter (LETKF) [BT99, OHS+04]. In the LETKF, the analysis is approximated in the ensemble space,

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

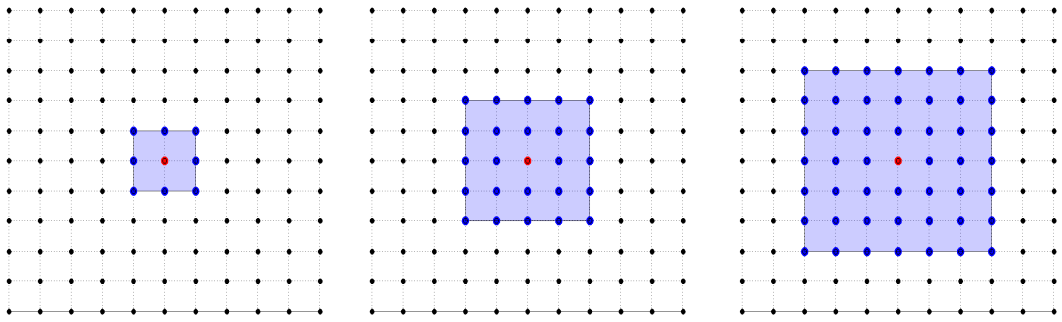
where,

$$\mathbf{w}^a = \mathbf{Q} \cdot \mathbf{V}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathbf{H} \cdot \overline{\mathbf{x}}^b \right] \in \mathbb{R}^{N \times 1}, \tag{2.8}$$

$\mathbf{V} = \mathbf{H} \cdot \Delta\mathbf{X} \in \mathbb{R}^{m \times N}$, and an estimate of the analysis covariance matrix in such space reads,

$$\mathbf{Q} = \left[(N-1) \cdot \mathbf{I} + \mathbf{V}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{V} \right]^{-1} \in \mathbb{R}^{N \times N},$$

with $\mathbf{I} \in \mathbb{R}^{N \times N}$ being the identity matrix in the ensemble space. This covariance matrix can then be utilized in order to build an ensemble about the posterior model of the distribution. In this context, localization methods are performed by using domain decomposition [OHS+08]: each model component is surrounded by a local box of radius δ and only local information (i.e., observed components) are utilized during the assimilation step. Examples of local boxes for different radii of influences are shown in figure 2.1. The global analysis is obtained by assembling all local analysis states onto the global domain.



(a) $\delta = 1$ (b) $\delta = 2$ (c) $\delta = 3$

Figure 2.1: Local domains for different radii of influence δ. The red dot is the model component to be assimilated, blue components are within the scope of δ, and black model components are unused during the local assimilation process.

In the ensemble Kalman filter based on a modified Cholesky decomposition (EnKF-MC) [NRSD15] background error correlations are estimated via the Cholesky decomposition proposed by Bicket and Levina in [BL08a]. This provides an estimate of the inverse background error covariance matrix of the form,

$$\widehat{\mathbf{B}}^{-1} = \mathbf{V}^T \cdot \mathbf{\Gamma} \cdot \mathbf{V} \in \mathbb{R}^{n \times n}, \tag{2.9}$$

where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is an unitary lower-triangular matrix, and $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix. Even more, when only local effects are considered during the estimation of $\widehat{\mathbf{B}}^{-1}$, in addition, the matrix $\mathbf{V}$ is sparse with only a few non-zero elements per row. Typically, the number of non-zero elements are some function of the radius of influence during the estimation of background error correlations. For instance, in the one-dimensional case, the radius of influence denotes the maximum number of non-zero elements, per row, in $\mathbf{V}$. The EnkF-MC is then obtained by plugging in the estimator (2.9) in (2.4). Given the structure of the Cholesky factors, the EnKF-MC can be seen as a matrix-free implementation of the EnKF.

Recall that, the precision analysis covariance matrix reads,

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \in \mathbb{R}^{n \times n}. \tag{2.10}$$

and since $\mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \in \mathbb{R}^{n \times n}$ can be written as a sum of m rank-one matrices, the factors (2.9) can be updated in order to obtain an estimate of the inverse analysis covariance matrix. In the next section, we propose an ensemble Kalman filter implementation based on this idea.

2.2 Proposed Method

Before we start, we make the assumptions [TAB+03, MB06] that, in practice, the data error covariance matrix $\mathbf{R}$ has a simple structure, the observation operator $\mathbf{H}$ is sparse and therefore, it can be applied efficiently, and that the number of model components n is several times the ensemble size N. We want to estimate the moments of the analysis distribution,

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}^a, \mathbf{A}),$$

based on the background ensemble (2.1), where $\mathbf{x}^a$ is the analysis state and $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the analysis covariance matrix. Consider the estimate of the inverse background error covariance matrix (2.9), the precision analysis covariance matrix (2.10) can be approximated as follows,

$$\mathbf{A}^{-1} \approx \widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{B}}^{-1} + \mathbf{X} \cdot \mathbf{X}^T, \tag{2.11}$$

where $\mathbf{X} = \mathbf{H}^T \cdot \mathbf{R}^{-1/2} \in \mathbb{R}^{n \times m}$. The matrix (2.11) can be written as follows,

$$\widehat{\mathbf{A}}^{-1} = \mathbf{V}^T \cdot \mathbf{\Gamma} \cdot \mathbf{V} + \sum_{i=1}^{m} \mathbf{x}_i \cdot \mathbf{x}_i^T,$$

where $\mathbf{x}_i$ denotes the i-th column of matrix $\mathbf{X}$, for $1 \leq i \leq m$. Consider the sequence of factors updates

$$\begin{aligned}
\mathbf{V}^{(i)^T} \cdot \mathbf{\Gamma}^{(i)} \cdot \mathbf{V}^{(i)} &= \left[\mathbf{V}^{(i-1)}\right]^T \cdot \mathbf{\Gamma}^{(i-1)} \cdot \mathbf{V}^{(i-1)} + \mathbf{x}_i \cdot \mathbf{x}_i^T \\
&= \left[\mathbf{V}^{(i-1)}\right]^T \cdot \left[\mathbf{\Gamma}^{(i-1)} + \mathbf{p}_i \cdot \mathbf{p}_i^T\right] \cdot \mathbf{V}^{(i-1)} \\
&= \left[\tilde{\mathbf{L}}^{(i-1)} \cdot \mathbf{V}^{(i-1)}\right]^T \cdot \tilde{\mathbf{D}}^{(i-1)} \cdot \left[\tilde{\mathbf{L}}^{(i-1)} \cdot \mathbf{V}^{(i-1)}\right]
\end{aligned}$$

where $\mathbf{V}^{(i-1)} \cdot \mathbf{p}_i = \mathbf{x}_i \in \mathbb{R}^{n\times 1}$, for $1 \leq i \leq m$, $\widehat{\mathbf{B}}^{-1} = \left[\mathbf{V}^{(0)}\right]^T \cdot \mathbf{\Gamma}^{(0)} \cdot \mathbf{V}^{(0)}$, and

$$\mathbf{\Gamma}^{(i-1)} + \mathbf{p}_i \cdot \mathbf{p}_i^T = \left[\tilde{\mathbf{L}}^{(i)}\right]^T \cdot \tilde{\mathbf{D}}^{(i)} \cdot \tilde{\mathbf{L}}^{(i)} \in \mathbb{R}^{n\times n}\,. \tag{2.12}$$

We can make use of the Dolittle's method in order to compute the factors $\tilde{\mathbf{D}}^{(i)}$ and $\tilde{\mathbf{L}}^{(i)}$ in (2.12), it is enough to note that,

$$\underbrace{\begin{bmatrix} 1 & \tilde{l}_{21} & \tilde{l}_{31} & \dots & \tilde{l}_{n1} \\ 0 & 1 & \tilde{l}_{32} & \dots & 0 \\ 0 & 0 & 1 & \dots & \tilde{l}_{n2} \\ \vdots & \vdots & \vdots & \ddots & \tilde{l}_{n3} \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\tilde{\mathbf{L}}^{(i)\,T}} \cdot \underbrace{\begin{bmatrix} \tilde{d}_1 & 0 & 0 & \dots & 0 \\ 0 & \tilde{d}_2 & 0 & \dots & 0 \\ 0 & 0 & \tilde{d}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix}}_{\tilde{\mathbf{D}}^{(i)}} \cdot \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \tilde{l}_{21} & 1 & 0 & \dots & 0 \\ \tilde{l}_{31} & \tilde{l}_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \tilde{l}_{n1} & \tilde{l}_{n2} & \tilde{l}_{n3} & \dots & 1 \end{bmatrix}}_{\tilde{\mathbf{L}}^{(i)}}$$

$$= \underbrace{\begin{bmatrix} d_1 + p_1^2 & p_1 \cdot p_2 & p_1 \cdot p_3 & \dots & p_1 \cdot p_n \\ p_2 \cdot p_1 & d_2 + p_2^2 & p_2 \cdot p_3 & \dots & p_2 \cdot p_n \\ p_3 \cdot p_1 & p_3 \cdot p_2 & d_3 + p_3^2 & \dots & p_3 \cdot p_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_n \cdot p_1 & p_n \cdot p_2 & p_n \cdot p_3 & \dots & d_n + p_n^2 \end{bmatrix}}_{\mathbf{\Gamma}^{(i)} + \mathbf{p}_i \cdot \mathbf{p}_i^T}\,.$$

After some math simplifications, the next equations are obtained,

$$\tilde{d}_k = p_k^2 + d_k - \sum_{q=k+1}^{n} \tilde{d}_q \cdot \tilde{l}_{qi}^2\,, \tag{2.13a}$$

and

$$\tilde{l}_{kj} = \frac{1}{\tilde{d}_k} \cdot \left[p_k \cdot p_j - \sum_{q=k+1}^{n} \tilde{d}_q \cdot \tilde{l}_{qi} \cdot \tilde{l}_{qj} \right], \tag{2.13b}$$

for $1 \leq k \leq n$, and $1 \leq j \leq k-1$. The set of equations (2.13) can be used in order to derive an algorithm for rank-one update of Cholesky factors, the updating process is shown in the Algorithm 3.

Algorithm 3 can be used in order to update the factors of $\widehat{\mathbf{B}}^{-1}$ for all column vectors in $\mathbf{X}$, this process is detailed in the Algorithm 2. Once the updating process has been performed, the resulting factors form an estimate of the inverse analysis covariance matrix,

$$\widehat{\mathbf{A}}^{-1} = \left[\mathbf{V}^{(m)}\right]^T \cdot \mathbf{\Gamma}^{(m)} \cdot \mathbf{V}^{(m)} \in \mathbb{R}^{n\times n}\,. \tag{2.14a}$$

From this covariance matrix, the posterior mode of the distribution can be approximated as follows,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \boldsymbol{\omega} \in \mathbb{R}^{n\times 1}\,, \tag{2.14b}$$

Algorithm 3 Rank-one update for the factors $\mathbf{V}^{(i-1)}$ and $\mathbf{\Gamma}^{(i-1)}$.

function UPD_CHOLESKY_FACTORS($\mathbf{V}^{(i-1)}$, $\mathbf{\Gamma}^{(i-1)}$, $\mathbf{x}_i$)
  Compute $\mathbf{p}_i$ from ${\mathbf{V}^{(i)}}^T \cdot \mathbf{p}_i = \mathbf{x}_i$.
  for $k = n \rightarrow 1$ **do**
    Compute $\tilde{d}_k$ via equation (2.13a).
    Set $l_{kk} \leftarrow 1$.
    for $j = 1 \rightarrow k-1$ **do**
      Compute $\tilde{l}_{kj}$ according to (2.13b).
    end for
  end for
  Set $\mathbf{V}^{(i)} \leftarrow \tilde{\mathbf{L}}^{(i-1)} \cdot \mathbf{V}^{(i-1)}$ and $\mathbf{\Gamma}^{(i)} \leftarrow \tilde{\mathbf{D}}^{(i)}$.
  return $\mathbf{V}^{(i)}$, $\mathbf{\Gamma}^{(i)}$
end function

Algorithm 4 Computing the factors $\mathbf{V}^{(m)}$ and $\mathbf{\Gamma}^{(m)}$ of $\widehat{\mathbf{A}}^{-1} = {\mathbf{V}^{(m)}}^T \cdot \mathbf{\Gamma}^{(m)} \cdot \mathbf{V}^{(m)}$.

function COMPUTE_ANALYSIS_FACTORS($\mathbf{V}^{(0)}$, $\mathbf{\Gamma}^{(0)}$, $\mathbf{H}$, $\mathbf{R}$)
  Set $\mathbf{X} \leftarrow \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$.
  for $i = 1 \rightarrow m$ **do**
    Set $[\mathbf{V}^{(i)}, \mathbf{\Gamma}^{(i)}] \leftarrow$ UPD_CHOLESKY_FACTORS $\left(\mathbf{V}^{(i-1)}, \mathbf{\Gamma}^{(i-1)}, \mathbf{x}_i\right)$
  end for
  return $\mathbf{V}^{(m)}$, $\mathbf{\Gamma}^{(m)}$
end function

where

$$\left[\mathbf{V}^{(m)}\right]^T \cdot \mathbf{\Gamma}^{(m)} \cdot \mathbf{V}^{(m)} \cdot \boldsymbol{\omega} = \mathbf{q} \tag{2.14c}$$

with $\mathbf{q} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathbf{H} \cdot \overline{\mathbf{x}}^b\right] \in \mathbb{R}^{n\times 1}$. Notice, the linear system (2.14c) involves lower and upper triangular matrices and therefore, $\overline{\mathbf{x}}^a$ can be estimated without the needing of matrix inversion. Once the posterior mode is computed, the analysis ensemble is built about it. Note that, $\widehat{\mathbf{A}}$ reads,

$$\widehat{\mathbf{A}} = \left[\mathbf{V}^{(m)}\right]^{-1} \cdot \left[\mathbf{\Gamma}^{(m)}\right]^{-1} \cdot \left[\mathbf{V}^{(m)}\right]^{-T}$$

and therefore a square root of $\widehat{\mathbf{A}}$ can be approximated as follows,

$$\widehat{\mathbf{A}}^{1/2} = \left[\mathbf{V}^{(m)}\right]^{-1} \cdot \left[\mathbf{\Gamma}^{(m)}\right]^{-1/2} \in \mathbb{R}^{n\times n}, \tag{2.15}$$

which can be utilized in order to build the analysis ensemble,

$$\mathbf{X}^a = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \mathbf{\Delta X}^a, \tag{2.16}$$

where $\mathbf{\Delta X}^a \in \mathbb{R}^{n\times N}$ is given by the solution of the linear system,

$$\mathbf{V}^{(m)} \cdot \left[\mathbf{\Gamma}^{(m)}\right]^{1/2} \cdot \mathbf{\Delta X}^a = \mathbf{W} \in \mathbb{R}^{n\times N}, \tag{2.17}$$

and the columns of $\mathbf{W} \in \mathbb{R}^{n\times N}$ are formed by samples from a multivariate standard normal distribution. Again, since $\mathbf{V}^{(m)}$ is lower triangular, the solution of (2.17) can be obtained readily.

2.2.1 Exploiting the structure of V

Readily, the computational efforts of the Algorithms 3 and 4 are bounded by $O\left(m \cdot n^3\right)$ long computations which can be quite impractical given grid resolutions in current operational data assimilation models. However, the use of conditional independence of background errors among distant model components can be exploited in order to reduce the computational complexity during the rank-one updates, for instance, we can consider only local neighbourhoods for each model component in order to approximate the analysis error correlations. This implies that, error correlations of distant model components regarding some radius of influence δ can be neglected. Note that, this preserves the structure of $\mathbf{V}$ in (2.9) and potentially, the impact of spurious correlations. In the one-dimensional case, for instance, the set of equations (2.13) can be written as follows,

$$\tilde{d}_k = p_k^2 + d_k - \sum_{q=k+1}^{k+r} \tilde{d}_q \cdot \tilde{l}_{qi}^2, \tag{2.18a}$$

and

$$\tilde{l}_{kj} = \frac{1}{\tilde{d}_k} \cdot \left[p_k \cdot p_j - \sum_{q=k+1}^{k+r} \tilde{d}_q \cdot \tilde{l}_{qi} \cdot \tilde{l}_{qj}\right]. \tag{2.18b}$$

Algorithm 5 Modified rank-one update for the factors $\mathbf{V}^{(i-1)}$ and $\mathbf{\Gamma}^{(i-1)}$.

function UPD_CHOLESKY_FACTORS($\mathbf{V}^{(i-1)}$, $\mathbf{\Gamma}^{(i-1)}$, $\mathbf{x}_i$)
  Compute $\mathbf{p}_i$ from ${\mathbf{V}^{(i)}}^T \cdot \mathbf{p}_i = \mathbf{x}_i$.
  for $k = n \rightarrow 1$ **do**
    Compute $\tilde{d}_k$ via equation (2.18a).
    Set $l_{kk} \leftarrow 1$.
    for $j = k - r \rightarrow k - 1$ **do**
      Compute $\tilde{l}_{kj}$ according to (2.18b).
    end for
  end for
  Set $\mathbf{V}^{(i)} \leftarrow \tilde{\mathbf{L}}^{(i-1)} \cdot \mathbf{V}^{(i-1)}$ and $\mathbf{\Gamma}^{(i)} \leftarrow \tilde{\mathbf{D}}^{(i)}$.
  return $\mathbf{V}^{(i)}$, $\mathbf{\Gamma}^{(i)}$
end function

Algorithm 4 can be modified in order to account for only local effects, this modification is reflected in Algorithm 5. The computational effort of the Algorithm 5 is then $O\left(m \cdot n \cdot \delta^2\right)$ which is linear regarding the number of observed components m and the model dimension n. For general dimensions, one can think in some computational effort of the form $O\left(m \cdot n \cdot f\left(\delta\right)^2\right)$ where $f\left(\delta\right)$ is some function of the radius of influence with, intuitively, $f\left(\delta\right) \ll n$.

Putting it all together, the posterior ensemble Kalman filter based on a modified Cholesky decomposition (P-EnKF) proceeds as follows:

1. Based on the samples (2.1), compute an estimate of $\mathbf{B}^{-1} = \mathbf{V}^T \cdot \mathbf{\Gamma} \cdot \mathbf{V}$ based on the modified Cholesky decomposition [NRSD15].
2. Update $\mathbf{V}$ and $\mathbf{\Gamma}$ according to the Algorithm 4.
3. Compute the analysis state (2.14b).
4. Build the analysis ensemble (2.16).
5. Propagate the analysis members until new observations are available.

2.3 Experimental Results

In this section, we assess the accuracy of the P-EnKF and compare it against that of the LETFK implementation proposed by Hunt in [OHS+04]. The numerical model is the Lorenz 96 model [FVE04] which mimics the behaviour of the atmosphere. This model is described by the next set of ordinary differential equations:

$$\frac{dx_k}{dt} = -x_{k-1} \cdot (x_{k-2} - x_{k+1}) - x_k + F, \quad \text{for } 1 \leq k \leq n\,, \tag{2.19}$$

where n is the number of model components and F is an external force. It is well-known that when F equals 8.0, the Lorenz 96 model exhibits a chaotic behaviour [KP10] which makes it attractive as a toy problem for testing weather prediction methods [FHH07]. The experimental settings are described below:

- An initial random solution $\mathbf{x}^{+}_{-3}$ is propagated for a while in time making use of the model (2.19) and a 4th order Runge Kutta method in order to obtain a vector state $\mathbf{x}^{*}_{-2}$ whose physics are consistent with the dynamics of such numerical model. This vector state serves as our reference solution.

- The reference solution is perturbed making use of samples from a Normal distribution with parameters $\mathcal{N}(\mathbf{0},\ \sigma_B \cdot \mathbf{I})$. Three different values for σ_B are considered during the numerical experiments $\sigma_B \in \{0.05,\ 0.10,\ 0.15\}$. This perturbed state is propagated in time in order to make it consistent with the physics and dynamics of the numerical model (2.19). From here, an initial background state $\mathbf{x}^{b}_{-1}$ is obtained.

- A similar procedure is performed in order to build a perturbed ensemble about $\mathbf{x}^{b}_{-1}$. The ensemble members are propagated in time from where an ensemble of model realizations $\left\{\mathbf{x}^{b[i]}_{0}\right\}_{i=1}^{N}$ of (2.19) is obtained.

- The assimilation windows consists of 15 equidistant observations. The frequency of observations is 0.5 time units which represents 3.5 days in the atmosphere.

- The dimension of the vector state is $n = 40$. The external force of the numerical model is set to $F = 8.0$.

- The number of observed components is 50% the dimension of the vector state.

- Three ensemble sizes are tried during the experiments $N \in \{20,\ 40,\ 60\}$.

- As a measure of quality, the L-2 norm of the analysis state and the reference solution is computed across assimilation steps.

- 100 runs are performed for each pair $(N,\ \sigma_B)$. For each run, a different initial random vector is utilized in order to build the initial perturbed reference solution $\mathbf{x}^{+}_{-3}$ (before the model is applied $\mathbf{x}^{*}_{-2}$). This yields to different initial ensembles as well as synthetic data for the different runs of each configuration (pair).

The average of the error norms of each pair $(N,\ \sigma_B)$ for the LETKF and the P-EnKF implementations are shown in the Table 2.1. As can be seen, in average across 100 of runs, the performance of the proposed EnKF implementation outperforms that of the LETKF in terms of $L-2$ norm of the error. Even more, the P-EnKF seems to be invariant to the initial background error σ_B since, in all cases, when the ensemble size is increased a better estimation of the reference state $\mathbf{x}^{*}$ at different observation times is obtained. This can also obey to the estimation of background error correlations via the modified Cholesky decomposition [BL08a] since it is drsaticaly improved whenever the ensemble size is increased as is pointed out by Bickel and Levina in [BP14]. In such case, the error decreases by $\mathcal{O}(\log(n)/N)$. This is crucial in the P-EnKF formulation since estimates of the precision analysis covariance matrix are obtained by rank-one updates on the inverse background error covariance matrix. On the other hand, in the LETKF context, increasing the ensemble size can improve the accuracy of the method but, that is not better than the one shown by the P-EnKF.

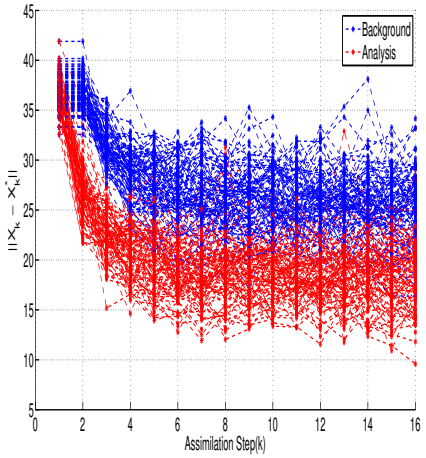
Some plots of the L-2 norm of error for the P-EnKF and the LETKF across different configurations and runs can be seen in figure 2.2. Note that, the error of the P-EnKF decreases aggressively since the earlier iterations. In the LETKF context, the accuracy is similar to that of the P-EnKF only at the end of the assimilation window.

σ_B	N	LETKF	P-EnKF
0.05	20	22,6166	21,2591
	40	20,5671	18,2548
	60	20,0567	17,8824
0.10	20	23,1742	21,0725
	40	20,9513	18,3542
	60	18,5048	17,8240
0,15	20	24,8201	20,9059
	40	21,1314	18,1731
	60	20,8487	17,7590

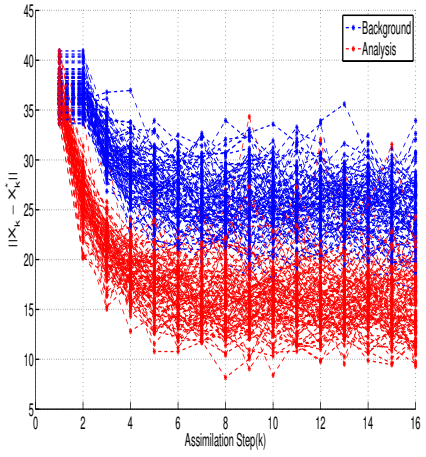
Table 2.1: Average of $L-2$ norm of errors for 100 of runs of each configuration (σ_B, N) for the compared filter implementations.

2.4 Conclusions

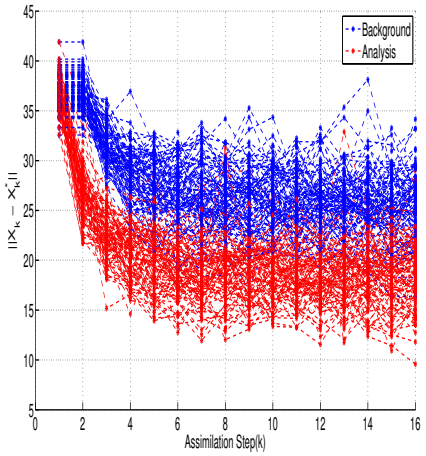
We propose a posterior ensemble Kalman filter based on a modified Cholesky decomposition. The proposed method estimates the posterior moments of the error distribution based on an ensemble of model realizations. An estimate of the inverse background error covariance matrix via a modified Cholesky decomposition is updated making use of rank-one matrices with information brought by the data error correlations in order to estimate the precision covariance matrix. This matrix is utilized in order to compute the posterior mode of the error distribution and then, samples are taken about it. This implementation is matrix-free making it attractive for practical implementations. Experimental settings are performed making use of the Lorenz 96 model and different observations and ensemble configurations. The results obtained by the proposed method are compared against those obtained by the local ensemble transform Kalman filter (LETKF). The results reveal that, the use of the proposed implementation can mitigate the impact of sampling errors and even more, the accuracy of the proposed EnKF implementation is similar to that of the LETKF.



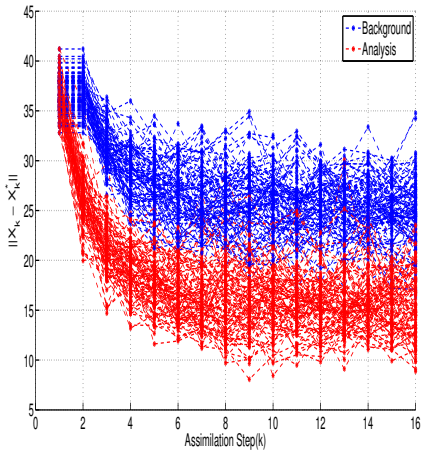
(a) LETKF $\sigma_B = 0.05,\ N = 60$



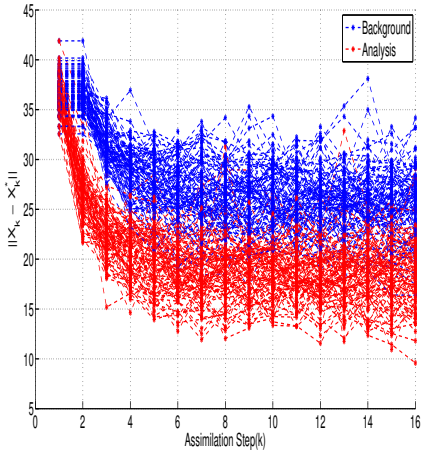
(b) P-EnKF $\sigma_B = 0.05,\ N = 60$



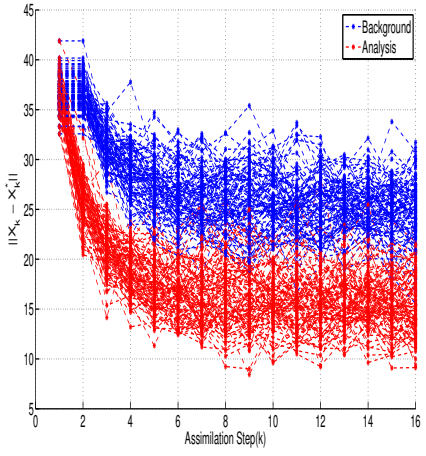
(c) LETKF $\sigma_B = 0.10,\ N = 60$



(d) P-EnKF $\sigma_B = 0.10,\ N = 60$



(e) LETKF $\sigma_B = 0.15,\ N = 60$



(f) P-EnKF $\sigma_B = 0.15,\ N = 60$

Figure 2.2: $L-2$ norm of the error for the LETKF and the P-EnKF implementations at different observation times. For each configuration, 100 of runs are performed. The assimilation window consists of 15 equidistant observations.

3 A Matrix-Free Posterior Ensemble Kalman Filter Implementation Based On A Modified Cholesky Decomposition

In this chapter, we propose a posterior ensemble Kalman filter (EnKF) based on a modified Cholesky decomposition. The main idea behind our approach is to estimate the moments of the analysis distribution based on an ensemble of model realizations. The method proceeds as follows: initially, an estimate of the precision background error covariance matrix is computed via a modified Cholesky decomposition and then, based on rank-one updates, the Cholesky factors of the inverse background error covariance matrix are updated in order to obtain an estimate of the inverse analysis covariance matrix. The special structure of the Cholesky factors can be exploited in order to obtain a matrix-free implementation of the EnKF. Once the analysis covariance matrix is estimated, the posterior mode of the distribution can be approximated and samples about it are taken in order to build the posterior ensemble. Experimental tests are performed making use of the Lorenz 96 model in order to assess the accuracy of the proposed implementation. The results reveal that, the accuracy of the proposed implementation is similar to that of the well-known local ensemble transform Kalman filter and even more, the use of our estimator reduces the impact of sampling errors during the assimilation of observations.

3.1 The Need of Covariance Matrix Localization

Data Assimilation is the process by which imperfect numerical forecasts and sparse observational networks are fused in order to estimate the state $\mathbf{x}^* \in \mathbb{R}^{n\times 1}$ of a system [Lor86, SB11a] which (approximately) evolves according to some model operator,

$$\mathbf{x}_p^* = \mathcal{M}_{t_{p-1}\rightarrow t_p}\left(\mathbf{x}_{p-1}^*\right),\ \text{for } 1 \leq p \leq G, \tag{3.1}$$

where, for instance, $\mathcal{M} : \mathbb{R}^{n\times 1} \rightarrow \mathbb{R}^{n\times 1}$ is a numerical model which mimics the ocean and/or the atmosphere dynamics, n is the number of model components, G is the number of observations (which form the assimilation window), and p denotes time index at time t_p. Sequential and smoothing methods are commonly utilized in order to perform the estimation process [BHC+10a, BHC+10b, CSS05]. In the context of sequential data assimilation, when Gaussian assumptions are done over background and observational errors, based on Bayes rule, the posterior mode of the error distribution can be computed as follows:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\cdot\mathbf{H}^T\cdot\mathbf{R}^{-1}\cdot\left[\mathbf{y} - \mathcal{H}\left(\mathbf{x}^b\right)\right] \in \mathbb{R}^{n\times 1}, \tag{3.2a}$$

where $\mathbf{x}^a \in \mathbb{R}^{n\times 1}$ is known as the analysis state, the analysis covariance matrix reads,

$$\mathbf{A} = \left[\mathbf{B}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \in \mathbb{R}^{n\times n}, \tag{3.2b}$$

m is the number of observed components from the model domain, $\mathcal{H} : \mathbb{R}^{n\times 1} \rightarrow \mathbb{R}^{m\times 1}$ is the observation operator, $\mathbf{B} \in \mathbb{R}^{n\times n}$ is the unknown background error covariance matrix, and $\mathbf{R} \in \mathbb{R}^{m\times m}$ stands for the data error covariance matrix. Likewise, $\mathcal{H}'(\mathbf{x}) \approx \mathbf{H}^T \in \mathbb{R}^{n\times m}$ is a linearized observation operator (with the linearization performed about the background state). Typically, the moments of the prior distribution,

$$\mathbf{x} \sim \mathcal{N}\left(\mathbf{x}^b, \mathbf{B}\right), \tag{3.3}$$

can be estimated based on an ensemble of model realizations [NRS15]. However, since ensemble members come at high computational costs owing to current operational data assimilation settings (i.e., numerical grid resolutions), ensemble sizes are bounded by the hundreds while their underlying error distributions range in the order of billions [And12]. Consequently, sampling errors can impact the quality of the analysis state [JFW14]. In practice, localization methods can be utilized in order to mitigate the impact of sampling errors during the assimilation steps [Bue05]. For instance, in the EnKF implementation based on a modified Cholesky decomposition (EnKF-MC) [NRSD15, NRSD17], the covariance matrix estimator proposed by Bickel and Levina in [BL08b] and the conditional independence of model components regarding their spatial distances are exploited in order to obtain sparse Cholesky factors of the precision background error covariance matrix, to reduce the computational cost of the analysis step, and to mitigate the impact of spurious correlations during the assimilation of observations. Given the relation between $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ in (3.2b) and by using the Bickel and Levina estimator, we think that sparse estimators of the analysis covariance matrix can be obtained without the needing of actually computing (3.2b), and therefore, efficient assimilation steps can be proposed.

This chapter is organized as follows: in section 3.2, efficient implementations of the EnKF which account for localization are discussed, section 3.3 presents a matrix-free posterior ensemble Kalman filter implementation based on a modified Cholesky decomposition; in section 3.4, experimental tests are performed making use of the Lorenz-96 model in order to assess the accuracy of the proposed method, and finally, section 3.5 states the conclusions of this research.

3.2 Preliminaries

3.2.1 The ensemble Kalman filter

The ensemble Kalman filter (EnKF) is a sequential Monte-Carlo method for parameter and state estimation in highly non-linear models [Eve03]. The popularity of the EnKF owes to his simple formulation and relatively ease implementation [Lor03b]. In the EnKF context, an ensemble of model realizations is utilized,

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \ldots, \mathbf{x}^{b[N]}\right] \in \mathbb{R}^{n\times N}, \tag{3.4a}$$

in order to estimate the moments of the background error distribution (3.3) via the empirical moments of the ensemble,

$$\mathbf{x}^b \approx \overline{\mathbf{x}}^b = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}^{b[e]} \in \mathbb{R}^{n \times 1}, \tag{3.4b}$$

and,

$$\mathbf{B} \approx \mathbf{P}^b = \frac{1}{N-1} \cdot \mathbf{\Delta X} \cdot \mathbf{\Delta X}^T \in \mathbb{R}^{n \times n}, \tag{3.4c}$$

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

$$\mathbf{\Delta X} = \mathbf{X}^b - \overline{\mathbf{x}}^b \cdot \mathbf{1}^T \in \mathbb{R}^{n \times N}, \tag{3.4d}$$

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

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^a \cdot \mathbf{\Delta Y} \in \mathbb{R}^{n \times N}, \tag{3.5a}$$

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

$$\mathbf{\Delta Y} = \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} \cdot \mathbf{1}^T + \mathbf{R}^{1/2} \cdot \mathbf{E} - \mathbf{H} \cdot \mathbf{X}^b \right] \in \mathbb{R}^{n \times N}, \tag{3.5b}$$

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

$$\mathbf{P}^a = \left[\left[\mathbf{P}^b \right]^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \right]^{-1} \in \mathbb{R}^{n \times n}, \tag{3.5c}$$

is the analysis ensemble covariance matrix.

3.2.2 Localization methods

As we mentioned before, the ensemble size is much lower than the model resolution ($N \ll n$) and as a consequence, $\mathbf{P}^b$ is rank-deficient which implies that, the precision covariance matrix (3.5c) can not be computed. In practice, localization methods are commonly utilized in order to artificially increase the rank of $\mathbf{P}^b$ and to mitigate the impact of spurious correlations during the analysis steps [HWS01]. In general, we can think in two different flavours of localization: covariance matrix localization [CJAS10, CEK+13], and spatial localization [Kep00, NRS15].

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

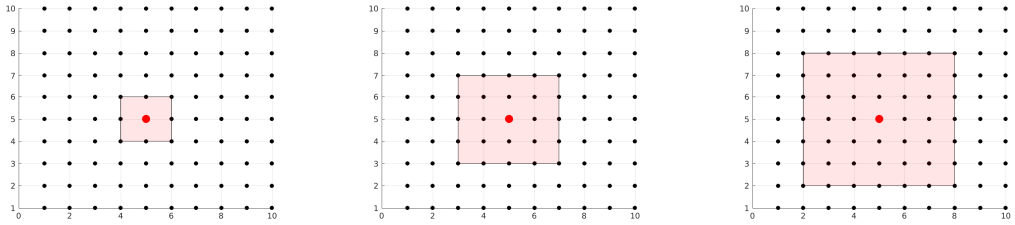
$$\widehat{\mathbf{P}}^b = \mathbf{\Lambda} \circ \mathbf{P}^b \in \mathbb{R}^{n \times n}, \tag{3.6}$$

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

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

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

In spatial localization schemes, each model component is surrounded by a local box of radius δ and all the information contained within such box is utilized in order to perform a local assimilation step [Bue11]. All local analysis components are then mapped back to the global domain from where a global analysis state is estimated. Some examples of spatial localization are shown in figure 3.1 for a two dimensional domain. In general, radius sizes can vary among different model components.



(a) $\delta = 1$ (b) $\delta = 2$ (c) $\delta = 3$

Figure 3.1: Local domains for different radii of influence δ. The red dot is the model component to be assimilated, the red square denotes components within the scope of δ, and components outside the region are unused during the local assimilation process.

It is important to note that, covariance matrix localization and spatial localization are equivalent computations [SB11b] and their use during the assimilation of observations should rely, mainly, on computational aspects.

3.2.3 Efficient EnKF implementations: accounting for localization

One of the best EnKF formulations in the context of spatial localization is the local ensemble transform Kalman filter (LETKF) which belongs to the family of deterministic EnKF formulations [TAB+03, BT99, OHS+04, HKS07]. This method has been sucesfully tested in operational data assimilation centres for weather forecast [OHS+08]. In the LETKF, the mode of the analysis distribution is estimated in the ensemble space as follows,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{\Delta X}\cdot\mathbf{w}^a \in \mathbb{R}^{n\times 1}\,, \tag{3.7a}$$

where $\mathbf{w}^a = \widetilde{\mathbf{P}}^a\cdot[\mathbf{H}\cdot\mathbf{\Delta X}]^T\cdot\left[\mathbf{y} - \mathcal{H}\left(\overline{\mathbf{x}}^b\right)\right] \in \mathbb{R}^{N\times 1}$, and $\widetilde{\mathbf{P}}^a \in \mathbb{R}^{N\times N}$ is a projection of the analysis covariance matrix (3.5c) onto such space wherein the ensemble covariance matrix (3.4c) is well-conditioned. The analysis ensemble is then built about the estimated state (3.7a) as follows,

$$\mathbf{X}^a = \overline{\mathbf{x}}^a\cdot\mathbf{1}^T + \sqrt{N-1}\cdot\mathbf{\Delta X}\cdot\widetilde{\mathbf{P}}^{a\,1/2} \in \mathbb{R}^{n\times N}\,. \tag{3.7b}$$

The assimilation step (3.7) is applied to each model component for a given radius of influence δ from where the global analysis state is obtained. Hence, since the most expensive computation is the inversion of $\widetilde{\mathbf{P}}^a$, the computational effort of the LETKF reads,

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

where φ denotes the local box sizes.

Another efficient EnKF implementation based on covariance matrix localization is the ensemble Kalman filter based on a modified Cholesky decomposition (EnKF-MC) [NRSD15]. In this context, the posterior ensemble is computed as follows:

$$\mathbf{X}^a = \mathbf{X}^b + \widehat{\mathbf{A}} \cdot \Delta\mathbf{Y} \in \mathbb{R}^{n\times N}, \tag{3.8a}$$

where the analysis error covariance matrix reads,

$$\widehat{\mathbf{A}} = \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \in \mathbb{R}^{n\times n}, \tag{3.8b}$$

and $\widehat{\mathbf{B}}^{-1} \in \mathbb{R}^{n\times n}$ is an estimate of the precision background error covariance matrix via the Bickel and Levina estimator [BL08b],

$$\widehat{\mathbf{B}}^{-1} = \mathbf{V}^T \cdot \mathbf{\Gamma}^{-1} \cdot \mathbf{V} \in \mathbb{R}^{n\times n}, \tag{3.8c}$$

the Cholesky factor $\mathbf{V} \in \mathbb{R}^{n\times n}$ is a lower triangular matrix,

$$\{\mathbf{V}\}_{i,j} = \begin{cases} -\beta_{i,j} & , j \in P(i,\delta) \\ 1 & , i = j \\ 0 & , otherwise \end{cases}, \tag{3.9}$$

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

$$\mathbf{x}^T_{[i]} = \sum_{j\in P(i,\delta)} \beta_{i,j} \cdot \mathbf{x}^T_{[j]} + \boldsymbol{\gamma}_i \in \mathbb{R}^{N\times 1},\ 1 \le i \le n, \tag{3.10}$$

where $\mathbf{x}^T_{[i]} \in \mathbb{R}^{N\times 1}$ denotes the i-th row (model component) of the ensemble (3.4a), components of vector $\boldsymbol{\gamma}_i \in \mathbb{R}^{N\times 1}$ are samples from a zero-mean Normal distribution with unknown variance σ^2, and $\mathbf{\Gamma} \in \mathbb{R}^{n\times n}$ is a diagonal matrix whose diagonal elements read,

$$\{\mathbf{\Gamma}\}_{i,i} = \widehat{\mathbf{var}}\left(\mathbf{x}^T_{[i]} - \sum_{j\in P(i,\delta)} \beta_{i,j} \cdot \mathbf{x}^T_{[j]}\right)^{-1} \approx \mathbf{var}\left(\boldsymbol{\gamma}_i\right)^{-1} = \frac{1}{\sigma^2} > 0, \text{ with } \{\mathbf{\Gamma}\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}^T_{[1]}\right)^{-1} \tag{3.11}$$

where $\mathbf{var}(\bullet)$ and $\widehat{\mathbf{var}}(\bullet)$ denote the actual and the empirical variances, respectively. Likewise, $P(i,\ \delta)$ stands for the predecessors of model component i with regard to δ and some ordering of components. For instance, figure 3.2 shows an example for a two dimensional domain when $\delta = 1$, $i = 6$, and model components are labelled by using column-major format. Since the number of predecessors of model components depend on δ, the resulting factor (3.9) can be sparse, this implies huge savings in terms of memory usage under current operational data

assimilation settings wherein n can be very large. Besides, $\widehat{\mathbf{B}}^{-1}$ can be represented in terms of his Cholesky factors and therefore, efficient manners to compute the ensemble (3.8a) can be derived [NRSA14]. Since the number of predecessors is some multiple of the local box sizes, the computational effort of the analysis step in the EnKF-MC can be expressed as follows,

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

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

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

(a) In blue, local box for the model component 6 when $\delta = 1$.

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

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

Figure 3.2: Local model components (local box) and local predecessors for the model component 6 when $\delta = 1$. Column-major ordering is utilized to label the model components.

We think that, by updating the Cholesky factors of (3.8b) it is possible to obtain an approximation of the precision analysis covariance matrix in terms of sparse Cholesky factors which can be exploited during the assimilation step (3.8a) in order to avoid the explicit solution of the linear system,

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

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

3.3 A Posterior Ensemble Kalman Filter Based On Modified Cholesky Decomposition

Before we start the derivation of our filter implementation, we assume that, the observational operator is nearly linear and/or it can be easily applied [SEB10], the data error covariance matrix possesses a simple structure and/or it can be easily decomposed [Eve09], and that observational networks are sparse [And01].

To be concise, we want to obtain an estimate of the precision analysis covariance matrix (3.8b) of the form,

$$\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} = \mathbf{V}^T \cdot \mathbf{\Gamma} \cdot \mathbf{V} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} = \widehat{\mathbf{V}}^T \cdot \mathbf{\Gamma} \cdot \widehat{\mathbf{V}} \in \mathbb{R}^{n \times n}, \tag{3.12}$$

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

error distribution,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \Delta\mathbf{y} = \overline{\mathbf{x}}^b + \Delta\mathbf{x}^a\,, \tag{3.13}$$

via the solution of lower and upper triangular systems of equations,

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

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

$$\mathbf{x} \sim \mathcal{N}\left(\overline{\mathbf{x}}^a, \widehat{\mathbf{A}}\right)\,, \tag{3.14}$$

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

1. drawing samples from the Normal distribution (3.14),

$$\mathbf{X}^a = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \widehat{\mathbf{A}}^{1/2} \cdot \mathbf{E} = \overline{\mathbf{x}}^a \cdot \mathbf{1}_N^T + \mathbf{V}^a\,, \tag{3.15}$$

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

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

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

2. using the synthetic data (3.5b),

$$\mathbf{X}^a = \overline{\mathbf{x}}^b \cdot \mathbf{1}_N^T + \widehat{\mathbf{A}} \cdot \Delta\mathbf{Y} = \overline{\mathbf{x}}^b \cdot \mathbf{1}_N^T + \widehat{\mathbf{V}}^a\,, \tag{3.16}$$

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

$$\widehat{\mathbf{V}}^T \cdot \widehat{\mathbf{\Gamma}} \cdot \widehat{\mathbf{V}} \cdot \widehat{\mathbf{V}}^a = \Delta\mathbf{Y}\,.$$

The approximation (3.15) is named *the posterior ensemble Kalman filter* (P-EnKF) since the analysis ensemble is built based on samples from the posterior distribution (3.14) while the approximation (3.16) is called *the synthetic posterior ensemble Kalman filter* (P-EnKF-S) since synthetic data is utilized in order to compute the analysis increments as is commonly done in stochastic EnKF formulations.

Details about the computations of $\widehat{\mathbf{V}}$ and $\mathbf{\Gamma}$ are discussed next.

3.3.1 Computing the Cholesky factors of the precision analysis covariance

By letting $\boldsymbol{\Omega} = \mathbf{H}^T \cdot \mathbf{R}^{-1/2} \in \mathbb{R}^{n\times m}$ in (3.12), the precision analysis covariance matrix can be written as follows,

$$\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{B}}^{-1} + \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}^T = \mathbf{V}^T\cdot\boldsymbol{\Gamma}\cdot\mathbf{V} + \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}^T = \mathbf{V}^T\cdot\boldsymbol{\Gamma}\cdot\mathbf{V} + \sum_{j=1}^{m}\cdot\boldsymbol{\omega}^{[j]}\cdot\left[\boldsymbol{\omega}^{[j]}\right]^T, \tag{3.17}$$

where $\boldsymbol{\omega}^{[j]} \in \mathbb{R}^{n\times 1}$ is the j-th column of matrix $\boldsymbol{\Omega} \in \mathbb{R}^{n\times m}$, for $1 \le j \le m$. Consider the sequence of matrices,

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

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

$$\begin{aligned}
\widehat{\mathbf{A}}^{(j)} &= \left[\mathbf{V}^{(j-1)}\right]^T\cdot\boldsymbol{\Gamma}^{(j-1)}\cdot\left[\mathbf{V}^{(j-1)}\right] + \boldsymbol{\omega}^{[j]}\cdot\left[\boldsymbol{\omega}^{[j]}\right]^T\\
&= \left[\mathbf{V}^{(j-1)}\right]^T\cdot\left[\boldsymbol{\Gamma}^{(j-1)} + \boldsymbol{\gamma}^{(j)}\cdot\left[\boldsymbol{\gamma}^{(j)}\right]^T\right]\cdot\left[\mathbf{V}^{(j-1)}\right],
\end{aligned} \tag{3.18}$$

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

$$\boldsymbol{\Gamma}^{(j-1)} + \boldsymbol{\gamma}^{(j)}\cdot\left[\boldsymbol{\gamma}^{(j)}\right]^T = \left[\widetilde{\mathbf{V}}^{(j-1)}\right]^T\cdot\boldsymbol{\Gamma}^{(j)}\cdot\left[\widetilde{\mathbf{V}}^{(j-1)}\right], \tag{3.19}$$

from equation (3.18), $\widehat{\mathbf{A}}^{(j)}$ can be written as follows,

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

where $\mathbf{V}^{(j)} = \widetilde{\mathbf{V}}^{(j-1)}\cdot\mathbf{V}^{(j-1)} \in \mathbb{R}^{n\times n}$. In equation (3.19), the components of the factors $\widetilde{\mathbf{V}}^{(j-1)}$ and $\widetilde{\boldsymbol{\Gamma}}^{(j)}$ can be easily computed from the elements of $\boldsymbol{\Gamma}^{(j-1)}$ and $\boldsymbol{\gamma}^{(j)}$ based on the Dolittle's method for matrix factorization, it is enough to note that,

$$\left\{\left[\widetilde{\mathbf{V}}^{(j-1)}\right]^T\cdot\boldsymbol{\Gamma}^{(j)}\cdot\left[\widetilde{\mathbf{V}}^{(j-1)}\right]\right\}_{i,k} = \delta_{i,k}\cdot\left\{\boldsymbol{\Gamma}^{(j-1)}\right\}_{i,i} + \left\{\boldsymbol{\gamma}^{(j)}\right\}_i\cdot\left\{\boldsymbol{\gamma}^{(j)}\right\}_k,$$

from which, the next equations are then obtained,

$$\left\{\boldsymbol{\Gamma}^{(j)}\right\}_{n,n} = \left[\left\{\boldsymbol{\gamma}^{(j)}\right\}_n\right]^2 + \left\{\boldsymbol{\Gamma}^{(j-1)}\right\}_{n,n}, \tag{3.20a}$$

$$\left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{i,k} = \frac{1}{\left\{\boldsymbol{\Gamma}^{(j)}\right\}_{i,i}} \cdot \left[\left\{\boldsymbol{\gamma}^{(j)}\right\}_i \cdot \left\{\boldsymbol{\gamma}^{(j)}\right\}_k - \sum_{q \in P(i,\,\delta)} \left\{\boldsymbol{\Gamma}^{(j)}\right\}_{q,q} \cdot \left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{q,i} \cdot \left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{q,k}\right], \quad (3.20b)$$

and

$$\left\{\boldsymbol{\Gamma}^{(j)}\right\}_{i,i} = \left[\left\{\boldsymbol{\gamma}^{(j)}\right\}_i\right]^2 + \left\{\boldsymbol{\Gamma}^{(j-1)}\right\}_{i,i} - \sum_{q \in P(i,\,\delta)} \left\{\boldsymbol{\Gamma}^{(j)}\right\}_{q,q} \cdot \left[\left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{q,i}\right]^2, \quad (3.20c)$$

for $n-1 \geq i \geq 1$ and $k \in P(i,\,\delta)$, where $\delta_{i,j}$ is the Kronecker delta function, and the diagonal elements of $\widetilde{\mathbf{V}}^{(j-1)}$ are all ones. In algorithm 6 the updating process for a column $\boldsymbol{\omega}^{[j]}$ is detailed while the computations of factors $\widehat{\mathbf{V}}$ and $\boldsymbol{\Gamma}$ for $\widehat{\mathbf{A}}^{-1}$ are shown in algorithm 7. We also report an estimate of the number of multiplications performed by the different steps of the algorithms. φ denotes the maximum number of predecessors across all model components. Typically, it will be a function of the radius of influence δ with $\varphi \ll n$. Note that, since the index k is constrained to the predecessors of component i, the structure of $\mathbf{V}^{(j-1)}$ is replicated in $\widetilde{\mathbf{V}}^{(j-1)}$ and therefore, the structure of $\mathbf{V}^{(j-1)}$ is preserved in $\mathbf{V}^{(j)}$. Consequently, the sparsity pattern of $\widehat{\mathbf{V}}$ equals that of $\mathbf{V}$.

Algorithm 6 Rank-one update for the factors $\mathbf{V}^{(j-1)}$ and $\boldsymbol{\Gamma}^{(j-1)}$.

1: **function** UPD_CHOLESKY_FACTORS($\mathbf{V}^{(j-1)}$, $\boldsymbol{\Gamma}^{(j-1)}$, $\boldsymbol{\omega}^{[j]}$) ▷ **COST**
2: Compute $\mathbf{p}^{(j)}$ from $\left[\mathbf{V}^{(j-1)}\right]^T \cdot \mathbf{p}^{(j)} = \boldsymbol{\omega}^{[j]}$. ▷ $\mathcal{O}(\varphi \cdot n)$
3: Compute $\left\{\boldsymbol{\Gamma}^{(j)}\right\}_{n,n}$ via equation (3.20a). ▷ $\mathcal{O}(1)$
4: **for** $i = n-1 \rightarrow 1$ **do** ▷ $n-1$ times lines 5–9, $\mathcal{O}\left(\varphi^2 \cdot n\right)$
5: Set $\left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{i,i} \leftarrow 1$. ▷ $\mathcal{O}(1)$
6: **for** $k \in P(i,\,\delta)$ **do** ▷ φ times line 7, $\mathcal{O}\left(\varphi^2\right)$
7: Compute $\left\{\widetilde{\mathbf{V}}^{(j-1)}\right\}_{i,k}$ according to (3.20b). ▷ $\mathcal{O}(\varphi)$
8: **end for**
9: Compute $\left\{\boldsymbol{\Gamma}^{(j)}\right\}_{i,i}$ via equation (3.20c). ▷ $\mathcal{O}(\varphi)$
10: **end for**
11: Set $\mathbf{V}^{(j)} \leftarrow \widetilde{\mathbf{V}}^{(j-1)} \cdot \mathbf{V}^{(j-1)}$. ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
12: **return** $\mathbf{V}^{(j)}$, $\boldsymbol{\Gamma}^{(j)}$
13: **end function**

3.3.2 Computational cost of the analysis step

Once the Cholesky factors of the precision analysis covariance matrix (3.12) are estimated, the posterior state (3.13) can be computed and the analysis ensemble (3.15) can be built. Putting it all together, the assimilation step of the P-EnKF is detailed in algorithm 8. Based on the algorithms 6 and 7, it is clear that, the computational effort of the P-EnKF is,

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

Algorithm 7 Computing the factors $\widehat{\mathbf{V}}$ and $\boldsymbol{\Gamma}$ of $\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{V}}^T \cdot \boldsymbol{\Gamma} \cdot \widehat{\mathbf{V}}$.

1: **function** COMPUTE_ANALYSIS_FACTORS($\mathbf{V}^{(0)}$, $\boldsymbol{\Gamma}^{(0)}$, $\mathbf{H}$, $\mathbf{R}$) ▷ **COST**
2: Set $\boldsymbol{\Omega} \leftarrow \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$. ▷ $\mathcal{O}(m \cdot n)$
3: **for** $j = 1 \rightarrow m$ **do** ▷ m times line 4, $\mathcal{O}\left(\varphi^2 \cdot m \cdot n\right)$
4: Set $[\mathbf{V}^{(j)}, \boldsymbol{\Gamma}^{(j)}] \leftarrow$ UPD_CHOLESKY_FACTORS $\left(\mathbf{V}^{(j-1)}, \boldsymbol{\Gamma}^{(j-1)}, \boldsymbol{\omega}^{[j]}\right)$ ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
5: **end for**
6: **return** $\mathbf{V}^{(m)}$ as $\widehat{\mathbf{V}}$, $\boldsymbol{\Gamma}^{(m)}$ as $\boldsymbol{\Gamma}$.
7: **end function**

which is linear with regard to the number of model components n and the number of observed components from the model domain m. This computational effort obeys, mainly, to the special structure of the Cholesky factors $\mathbf{V}^{(j)}$ and $\widetilde{\mathbf{V}}^{(j)}$, for $1 \leq j \leq m$, and how these structures can be exploited in practice in order to reduce the overall computational effort of the analysis step, for instance, the backward substitution in line 2 of algorithm 6 can be efficiently performed as follows,

$$\left\{\boldsymbol{\gamma}^{(j)}\right\}_{n-i+1} = \boldsymbol{\omega}^{[j]} - \sum_{k \in P(i,\delta)} \left\{\boldsymbol{\gamma}^{(j)}\right\}_k \cdot \left\{\left[\mathbf{V}^{(j-1)}\right]\right\}_{i,k}, \text{ for } 1 \leq i \leq n,$$

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

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

where, evidently, the maximum number of elements in $P(i, \delta) \cap P(k, \delta)$ is bounded by φ and therefore, the total number of multiplications is bounded by $\mathcal{O}\left(\varphi^2 \cdot n\right)$. Notice, since the matrices $\mathbf{V}^{(j)}$ and $\widetilde{\mathbf{V}}^{(j)}$ are sparse and lower triangular, their non-zero bands can represented by vectors. In this manner, for instance, the computations derived for the elements of $\mathbf{V}^{(j)}$ and $\widetilde{\mathbf{V}}^{(j)}$ can be performed on the elements of such vectors. Efficient matrix storage schemes are now proposed by the literature [SGV05, LT16] as well as scientific computational languages which exploit such structures in order to speed-up matrix computations and to save memory usage [DDO13]. From here, matrix-free implementations of the P-EnKF can be easily derived in order to make it practical under operational data assimilation settings. Readily, the computational

Algorithm 8 Assimilation of observations via the posterior ensemble Kalman filter (3.15).

1: **function** ANALYSIS_P-ENKF($\mathbf{X}^b$, $\mathbf{V}$, $\boldsymbol{\Gamma}$, $\mathbf{y}$, $\mathbf{H}$, $\mathbf{R}$) ▷ Having, $\widehat{\mathbf{B}}^{-1} = \mathbf{V}^T \cdot \boldsymbol{\Gamma} \cdot \mathbf{V}$.
2: Set $[\widehat{\mathbf{V}}, \boldsymbol{\Gamma}] \leftarrow$ COMPUTE_ANALYSIS_FACTORS ($\mathbf{V}$, $\boldsymbol{\Gamma}$, $\mathbf{H}$, $\mathbf{R}$).
3: Compute $\overline{\mathbf{x}}^a$ according to equation (3.13).
4: Compute the posterior ensemble $\mathbf{X}^a$ based on equation (3.15).
5: **return** $\mathbf{X}^a$.
6: **end function**

effort of the P-EnKF-S is similar to that of the P-EnKF. We detail the assimilation step of the P-EnKF-S in algorithm 9.

Algorithm 9 Assimilation of observations via the posterior ensemble Kalman filter (3.16) .

function ANALYSIS_P-ENKF($\mathbf{X}^b$, $\mathbf{V}$, $\boldsymbol{\Gamma}$, $\mathbf{y}$, $\mathbf{H}$, $\mathbf{R}$) ▷ Having, $\widehat{\mathbf{B}}^{-1} = \mathbf{V}^T \cdot \boldsymbol{\Gamma} \cdot \mathbf{V}$.
  Set $[\widehat{\mathbf{V}}, \boldsymbol{\Gamma}] \leftarrow$ COMPUTE_ANALYSIS_FACTORS ($\mathbf{V}$, $\boldsymbol{\Gamma}$, $\mathbf{H}$, $\mathbf{R}$).
  Compute the posterior ensemble $\mathbf{X}^a$ based on equation (3.16).
  return $\mathbf{X}^a$.
end function

3.3.3 Inflation aspects

While localization methods reduce the impact of spurious correlations, covariance inflation mitigates the impact of under-estimation of sample variances [Wes16, LW17, LMQ16, PXJ+17]. Typically, ensemble members are inflated prior the forecast step in order to enrich the background error information for the next assimilation cycle and to reduce the odds of ensemble collapsing. For instance, after the assimilation step of EnKF, the ensemble members (3.5a) are inflated by a factor of $\rho > 1$ about the analysis mean,

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

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

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

and therefore, the inflation can be performed before the analysis members are drawn from the distribution (3.14),

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

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

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

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

3.3.4 Main differences between the EnKF-MC, the P-EnKF, and the P-EnKF-S

In essence, the EnKF-MC, the P-EnKF, and the P-EnKF-S are stochastic filters based on a modified Cholesky decomposition [BL08b]. Nevertheless, the manner how they operate is quite different. Consider again the analysis covariance matrix (3.12),

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

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

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

where the columns of $\mathbf{E} \in \mathbb{R}^{n \times N}$ follows a standard Normal distribution. $\widehat{\mathbf{Z}}$ can be computed by using the iterative Sherman Morrison formula [RSA15] in order to avoid the direct inversion of $\widehat{\mathbf{A}}$ while the trivial linear system,

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

is the one to solve in order to compute the analysis increments for the background members in the P-EnKF formulation. Likewise, the linear system to solve in the P-EnKF-S case depends on the Cholesky factors and the synthetic data (3.5b),

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

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

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

while the analysis mean,

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

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

3.4 Experimental Results

In this section, we tests the proposed EnKF implementations and compare our results with those obtained by the EnKF-MC and the LETKF implementations discussed in section 3.2. We make use of the Lorenz-96 model [Lor05] as our surrogate model during the experiments. The Lorenz-96 model is described by the following set of ordinary differential equations [FHH07]:

$$\frac{dx_j}{dt} = \begin{cases} (x_2 - x_{n-1}) \cdot x_n - x_1 + F & \text{for } j = 1, \\ \left(x_{j+1} - x_{j-2}\right) \cdot x_{j-1} - x_j + F & \text{for } 2 \leq j \leq n-1, \\ (x_1 - x_{n-2}) \cdot x_{n-1} - x_n + F & \text{for } j = n, \end{cases} \tag{3.21}$$

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 [KP10, GM05]. A time unit in the Lorenz-96 represents 7 days in the atmosphere. The experimental settings are described below:

- An initial random solution is integrated over a long time period in order to obtain an initial condition $\mathbf{x}^*_{-2} \in \mathbb{R}^{n\times 1}$ dynamically consistent with the model (3.21).
- A perturbed background solution $\mathbf{x}^{*b}_{-2}$ is obtained at time t_{-2} by drawing a sample from the Normal distribution,

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

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

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

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

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

  where the number of observed components from the model space is $m = 30$. The components are randomly chosen at the different assimilation steps. Thus, $\mathbf{H}_k$ is randomly formed at the different assimilation cycles.
- Values of the inflation factor ρ are ranged in $1 \leq \rho \leq 1.1$.
- We try different values for the radii of influence δ, these range in $1 \leq \delta \leq 7$.
- The ensemble size for the benchmarks is $N = 20$. These members are randomly chosen from the pool $\widehat{\mathbf{X}^b}_0$ for the different pairs $(\delta,\ \rho)$ in order to form the initial ensemble $\mathbf{X}^b_0$ for the assimilation window. Evidently, $\mathbf{X}^b_0 \subset \widehat{\mathbf{X}^b}_0$.
- The assimilation steps are also performed by the EnKF with full size of $\widehat{\mathbf{X}^b}_0$ in order to obtain a reference solution regarding what to expect from the EnKF formulations. Note that, the ensemble size is large enough in order to dissipate the impact of sampling errors. No inflation is needed as well.
- The L-2 norm of the error is utilized as a measure of accuracy at the assimilation step p,

$$\lambda_p = \sqrt{\left[\mathbf{x}^*_p - \mathbf{x}^a_p\right]^T \cdot \left[\mathbf{x}^*_p - \mathbf{x}^a_p\right]} , \tag{3.22}$$

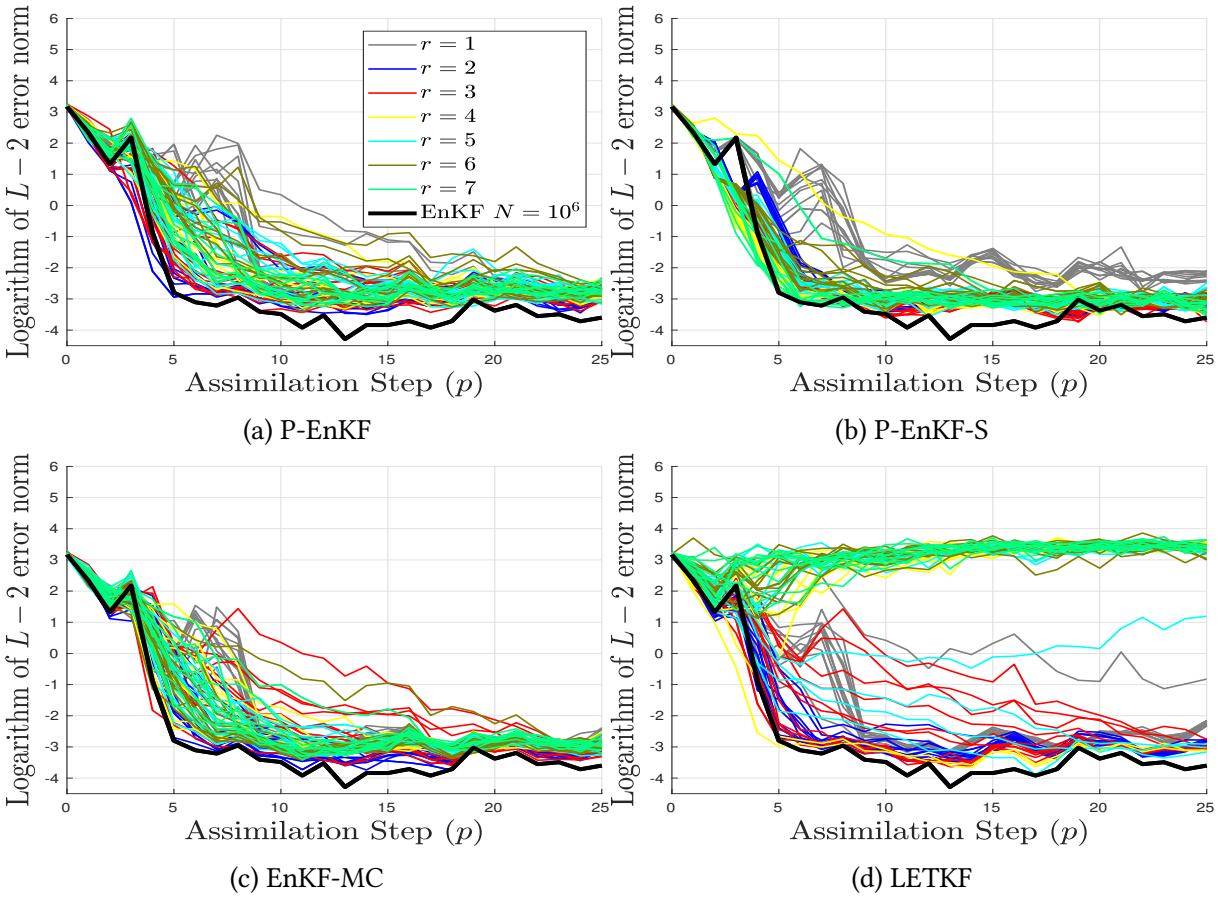
  where $\mathbf{x}^*_p$ and $\mathbf{x}^a_p$ are the reference and the analysis solutions, respectively.

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

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

In figure 3.3, the results are shown in terms of $L-2$ error norms and assimilation steps (p) for the compared methods, we group the results by radius sizes. As can be seen, there are parameter values for δ and ρ in which the results proposed by the filters are comparable with those obtained by the EnKF with a large ensemble size of $N = 10^6$. This supports the theory that, all filters can provide good approximations of posterior states with just a few ensemble members. However, the performance of the LETKF degrades as the radius size is increased, this can be explained as follows: in the LETKF, local background error correlations are estimated based on a local ensemble covariance matrix, therefore, sampling noise can easily impact the sample covariance matrix when the ensemble size is not much larger than the number of components within the local boxes. Such cases are not evident in the EnKF implementations based on a modified Cholesky decomposition. All these filters perform remarkably well and in all cases, the initial error (uncertainty) is decreased. The importance of having filters which are not so sensible to the parameters δ and ρ is that, in practice, such parameters can be hard to tune. Thus, one can prefer filter formulations whose accuracy is not highly impacted by changes in those parameters.

We can analyze the results in terms of RMSE values as well. Those are reported in figure 3.4. Note that, as we mentioned before, the accuracy of the LETKF can be sensible to the radius size regardless the inflation factor. After $\delta = 3$, the quality of the analysis obtained by this filter is clearly impacted by sampling errors. On the other hand, the proposed filters behave similarly to the EnKF-MC since their background error correlations are estimated making use of the same estimator (3.8c). However, we can see that, there are always configurations of the parameters for all filters which provide accurate enough approximations.



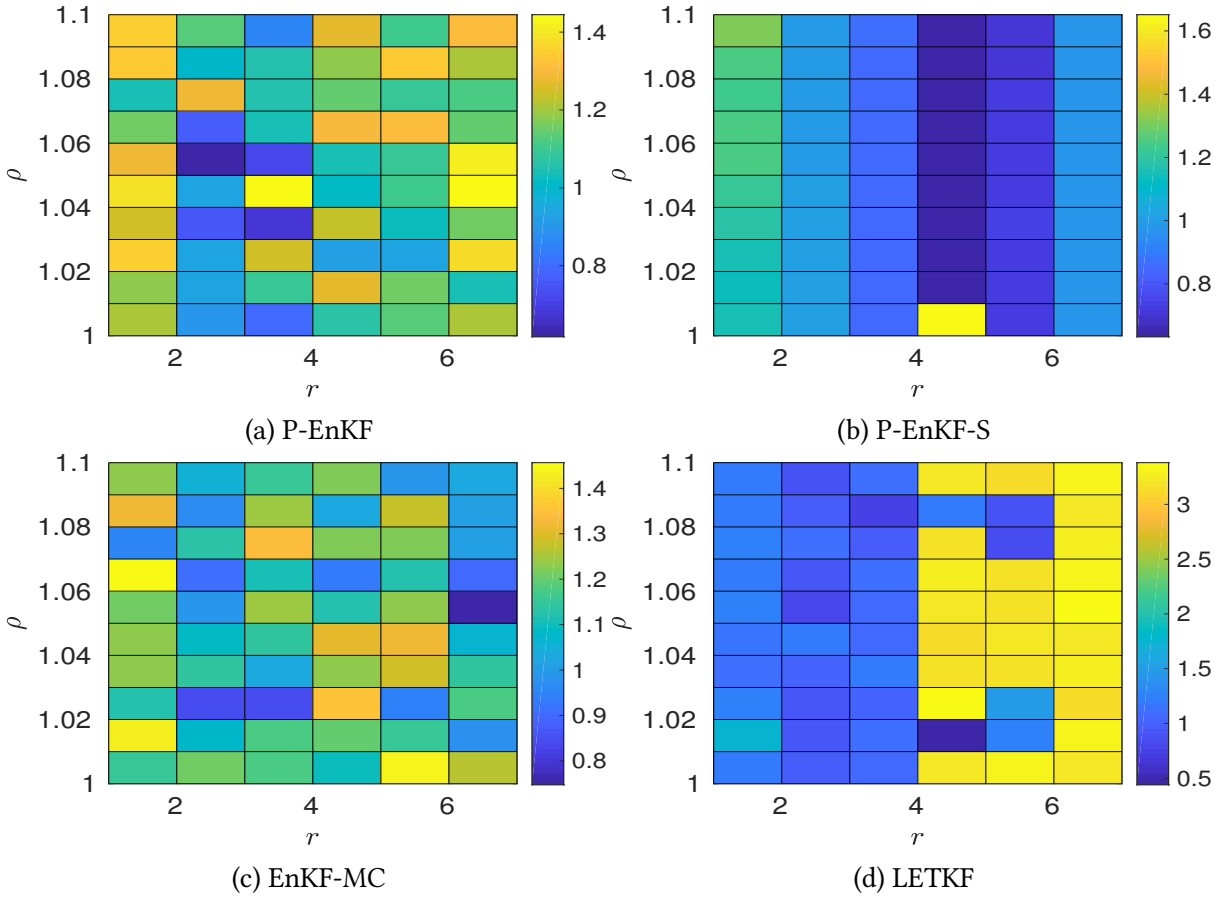
(a) P-EnKF

(b) P-EnKF-S

(c) EnKF-MC

(d) LETKF

Figure 3.3: Experimental results with the Lorenz-96 model (3.21). The results are grouped by values of δ. inflation factors are ranged in $1 \leq \rho \leq 1.1$ for each group.



(a) P-EnKF (b) P-EnKF-S

(c) EnKF-MC (d) LETKF

Figure 3.4: Experimental results with the Lorenz-96 model (3.21). The RMSE values are shown for the compared filter implementations for different values of δ and ρ.

3.5 Conclusions

This chapter proposes a posterior ensemble Kalman filter based on a modified Cholesky decomposition which works as follows: the precision background error covariance matrix is estimated in terms of Cholesky factors via a modified Cholesky decomposition, the factors of the posterior precision covariance are then obtained by rank-one updates over the background factors and from there, the posterior mode of the error distribution can be estimated. By using the analysis factors, the posterior ensemble can be built by either sampling from the posterior distribution or making use of synthetic data. Besides, sparse estimators of the precision analysis covariance can be obtained by exploiting the conditional independence of model components regarding some radius of influence as is done in the EnKF-MC context. The computational effort of the proposed method is similar to that of the EnKF-MC. Experimental tests are performed in order to assess the accuracy of the proposed method as the inflation factor and the radius of influences are varied. The numerical model operator utilized during the experiments is the Lorenz-96 model. The results reveal that, the accuracy in terms of root-mean-square-error of the proposed method is similar to that of one of the best EnKF implementations from the current literature.

Besides, the results obtained by the proposed implementation are comparable to those of the EnKF with large ensemble sizes.

4 A Four Dimensional Variational Data Assimilation Framework for Wind Energy Potential Estimation

In this paper, we propose a Four-Dimensional Variational (4D-Var) data assimilation framework for wind energy potential estimation. The framework is defined as follows: we choose a numerical model which can provide forecasts of wind speeds then, an ensemble of model realizations is employed to build control spaces at observation steps via a modified Cholesky decomposition. These control spaces are utilized to estimate initial analysis increments and to avoid the intrinsic use of adjoint models in the 4D-Var context. The initial analysis increments are mapped back onto the model domain from which we obtain an estimate of the initial analysis ensemble. This ensemble is propagated in time to approximate the optimal analysis trajectory. Wind components are post-processed to get wind speeds and to estimate wind energy capacities. A matrix-free analysis step is derived from avoiding the direct inversion of covariance matrices during assimilation cycles. Numerical simulations are employed to illustrate how our proposed framework can be employed in operational scenarios. A catalogue of twelve Wind Turbine Generators (WTGs) is utilized during the experiments. The results reveal that our proposed framework can properly estimate wind energy potential capacities for all wind turbines within reasonable accuracies (in terms of Root-Mean-Square-Error) and even more, these estimations are better than those of traditional 4D-Var ensemble-based methods. Moreover, large variability (variance of standard deviations) of errors are evidenced in forecasts of wind turbines with the largest rate-capacity while homogeneous variability can be seen in wind turbines with the lowest rate-capacity.

4.1 Hybrid 4D-Var Methods for Wind Power Estimation

Data Assimilation (DA) is the process by which an imperfect numerical forecast $\{\mathbf{x}_k^b\}_{k=0}^G$ is adjusted according to real noisy observations $\{\mathbf{y}_k\}_{k=0}^G$ [NR18, NRCB18], where $\mathbf{x}_k^b \in \mathbb{R}^{n\times 1}$ and $\mathbf{y}_k \in \mathbb{R}^{m\times 1}$ are the background state and the observations at step k, for $0 \leq k \leq G$, respectively, n is the model size (model resolution), m denotes the number of observations per assimilation step, and G is the size of the assimilation window (the number of times wherein observations are available). In strong constraint Four-Dimensional Variational (4D-Var) methods, cost functions of the form [Lor03b, Lor03a]:

$$\mathcal{J}\left(\mathbf{x}_0\right) = \frac{1}{2}\cdot\left\|\mathbf{x}_0 - \mathbf{x}_0^b\right\|_{\mathbf{B}_0^{-1}}^2 + \frac{1}{2}\cdot\sum_{k=0}^{G}\left\|\mathbf{y}_k - \mathcal{H}\left(\mathbf{x}_k\right)\right\|_{\mathbf{R}_k^{-1}}^2, \tag{4.1}$$

are employed to perform the assimilation process, where $\mathbf{B}_0 \in \mathbb{R}^{n\times n}$ and $\mathbf{R}_k \in \mathbb{R}^{m\times m}$ are the covariance matrix of the initial background errors and the estimated data error covariance matrix at step k, respectively. Likewise,

$$\mathbf{x}_f = \mathcal{M}_{t_{f-1}\to t_f}\left(\mathbf{x}_{f-1}\right)\,,\ \text{for } 1 \le f \le G\,, \tag{4.2}$$

where $\mathcal{M} : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ is a numerical model which, for instance, mimics the behavior of the ocean and/or the atmosphere and even more, we assume that the model can predict wind components (or wind speeds). We then seek the initial condition which best fit the data:

$$\mathbf{x}_0^a = \arg\min_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0)\,,\ \text{subject to (4.2)}\,. \tag{4.3}$$

To solve the optimization problem (4.3), for instance, we can make use of adjoint models (i.e., transpose of linearization of the numerical model) or an ensemble of model realizations. Regardless of which one is chosen, the initial condition (4.3) can provide a forecast of relevant physical variables (depending on the numerical model) such as wind components, temperature, and humidity. Forecasts of wind components can be exploited in the context of Wind Turbine Generators (WTGs) to estimate the potential energy capacities of wind turbines and then to employ green sources of energy in cities, countries, and even more, remote places wherein these are the unique option. Hence, we can make use of numerical models and 4D-Var optimization problems to estimate wind components (in particular, wind speeds since they are key inputs to size WTG), then by using parameters from wind turbines such as cut-in speed, cut-out speed, and rated wind speed, we can estimate their potential wind power capacity for a specific place. WTG parameters allow choosing the best WTG for a specific place according to its wind-speed values. This can be exploited in places such as the Latin American and the Caribbean (LAC) countries since these are widely known for their large power generation capacity using renewable energies. This makes them highly attractive for clean energy investment [FAF+17]. Recent studies indicate that the full deployment of this capacity can be almost seven times larger than the current world installed one, and even more, it can constitute a near-zero carbon emissions option for developing countries [DRB08]. This could provide substantial societal benefits, including energy security, local and global environmental benefits, domestic job creation, and improved balance of payments, amongst others [VIR+14]. Given LAC geography, wind turbines (wind farms) can be exploited as clean energy sources. The economic benefits of wind farms are better than those of traditional sources such as solar farms, which make the first option desirable for long-term plans at national levels. The proper planning and scheduling of wind power systems can lead to almost no impact on LAC ecosystems, neither visual or audible. Besides, this can serve as a complement of the hydro-dominated electricity grids [MWOH17], as the winds are stronger during the dry season when hydroelectric generation is most limited. Moreover, wind farms can provide a full or complementary source of energy in some areas of difficult access; the application of wind turbines is primarily in windmills that are used to generate electricity [GJSAGJ17]. These wind turbines can be used to avail off-grid electricity in remote regions (i.e., some islands).

The structure of this Chapter is as follows: Section 4.2 discusses DA formulations and wind turbine generators (WTG). In Section 4.3, we propose a novel framework for electrical power estimation via WTGs and 4D-Var ensemble DA. In Section 4.4, numerical simulations are

employed to show how our framework can be employed; we generate synthetic scenarios by using an Atmospheric General Circulation Model. Lastly, Section 4.5 states the conclusions of this research.

4.2 Preliminaries

In this section, we briefly discuss concepts related to 4D-Var ensemble DA and wind turbine generators. We primarily focus on the necessary topics for the derivation of our 4D-Var ensemble-based method.

4.2.1 Data Assimilation

To solve the optimization problem (4.3), we can employ adjoint models to approximate gradients and to conduct optimization steps via, for instance, line-search or trust region methods. However, these models can be labor-intensive to develop and computationally expensive to run. For instance, the adjoint model of the High Resolution Limited Area Modelling (HIRLAM) 4D-Var [GB14, SUL+09] was developed in 10 years in which most of the time was spent to detect and to fix errors in the tangent and the adjoint models [Gus07]. To avoid the use of adjoint models, we can employ an ensemble of model realizations [NRS15] as follows [HM98, SKW18]:

$$\mathbf{X}_k^b = \left[\mathbf{x}_k^{b[1]},\, \mathbf{x}_k^{b[2]},\, \ldots,\, \mathbf{x}_k^{b[N]}\right] \in \mathbb{R}^{n\times N} \tag{4.4a}$$

where $\mathbf{x}_k^{b[e]} \in \mathbb{R}^{n\times 1}$ stands for the e-th ensemble member, for $1 \leq e \leq N$, at time k, for $0 \leq k \leq G$. Then, the ensemble mean:

$$\overline{\mathbf{x}}_k^b = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}_k^{b[e]} \in \mathbb{R}^{n\times 1}\,, \tag{4.4b}$$

and the ensemble covariance matrix:

$$\mathbf{P}_k^b = \frac{1}{N-1} \cdot \Delta\mathbf{X}_k^b \cdot \left[\Delta\mathbf{X}_k^b\right]^T \in \mathbb{R}^{n\times n}\,, \tag{4.4c}$$

act as estimates of the forecast state $\mathbf{x}_k^b$ and the forecast error covariance matrix $\mathbf{B}_k$, respectively, where the matrix of member deviations reads:

$$\Delta\mathbf{X}_k^b = \mathbf{X}_k^b - \overline{\mathbf{x}}_k^b \cdot \mathbf{1}^T \in \mathbb{R}^{n\times N}\,. \tag{4.4d}$$

The model trajectory in (4.1) can be constrained to the space spanned by the background ensemble members (4.4a), this is:

$$\mathbf{x}_k = \overline{\mathbf{x}}_k^b + \Delta\mathbf{X}_k \cdot \mathbf{w}\,, \tag{4.5}$$

where $\mathbf{w} \in \mathbb{R}^{N\times 1}$ is a vector in redundant coordinates to be determined later. This is equivalent to:

$$\mathbf{x}_k - \overline{\mathbf{x}}_k^b \in \mathbf{range}\left\{\Delta\mathbf{X}_k\right\} \approx \mathbf{range}\left\{\mathbf{B}_k^{1/2}\right\}\,,$$

therefore, the estimation of analysis increments is performed onto the sub-space given by a low-rank square root approximation of the background error covariance matrices (4.4c) at observation times (times where observations are available). By replacing (4.5) into the equation (4.1) one obtains [GAVL15]:

$$\begin{aligned} \mathcal{J}(\mathbf{x}_0) &= \mathcal{J}\left(\overline{\mathbf{x}}_0^b + \mathbf{\Delta X}_0 \cdot \mathbf{w}\right) \\ &= \widehat{\mathcal{J}}(\mathbf{w}) = \frac{(N-1)}{2} \cdot \|\mathbf{w}\|^2 + \frac{1}{2} \cdot \sum_{k=0}^{G} \|\mathbf{d}_k - \mathbf{Q}_k \cdot \mathbf{w}\|^2_{\mathbf{R}_k^{-1}}, \end{aligned} \tag{4.6}$$

where $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \cdot \overline{\mathbf{x}}_k^b \in \mathbb{R}^{m \times 1}$ is the innovation vector and $\mathbf{Q}_k = \mathbf{H}_k \cdot \mathbf{\Delta X}_k^b \in \mathbb{R}^{m \times N}$. Note that, this cost function does not rely in the numerical model (4.2) anymore. The optimal value of the control variable $\mathbf{w}$ is then seek:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \widehat{\mathcal{J}}(\mathbf{w}). \tag{4.7a}$$

The gradient of (4.6) equals:

$$\begin{aligned} \nabla_{\mathbf{w}} \widehat{\mathcal{J}}(\mathbf{w}) &= (N-1) \cdot \mathbf{w} - \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot [\mathbf{d}_k - \mathbf{Q}_k \cdot \mathbf{w}] \\ &= \left[(N-1) \cdot \mathbf{I} + \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Q}_k\right] \cdot \mathbf{w} - \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{d}_k \in \mathbb{R}^{N \times 1}, \end{aligned} \tag{4.7b}$$

and from here, the optimal weight (4.7a) can be approximated as follows:

$$\mathbf{w}^* = \left[(N-1) \cdot \mathbf{I} + \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Q}_k\right]^{-1} \cdot \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{d}_k \in \mathbb{R}^{N \times 1}, \tag{4.7c}$$

from which the initial analysis state can be estimated:

$$\overline{\mathbf{x}}_0^a = \overline{\mathbf{x}}_0^b + \mathbf{\Delta X}_0^b \cdot \mathbf{w}^*. \tag{4.7d}$$

Since in (4.5), $\mathbf{x}_k^a$ represents an approximation rather than an exact analysis trajectory, the initial analysis is recovered and then, it is evolved in time by using the numerical model (4.2) from which we obtain an estimate of the optimal trajectory of (4.3):

$$\overline{\mathbf{x}}_f^a = \mathcal{M}_{t_{f-1} \to t_f}\left(\overline{\mathbf{x}}_{f-1}^a\right), \text{ for } 1 \leq f \leq G. \tag{4.7e}$$

Notice, in the 4D-EnKF, all computations are performed onto the ensemble space (4.5) and therefore, the computational cost of estimating (4.7d) is linearly bounded regadring n and m [NRS16]:

$$O\left(N \cdot n \cdot m + N^2 \cdot m\right).$$

Readily, posterior ensemble members at the initial time can be estimated via the implicit covariance matrix in (4.7c):

$$\overline{\mathbf{x}}_0^{a(e)} = \overline{\mathbf{x}}_0^b + \Delta\mathbf{X}_0^b \cdot \mathbf{w}^{(e)}\,,\ \text{for } 1 \leq e \leq N\,,$$

where:

$$\mathbf{w}^{(e)} \sim \mathcal{N}\left(\mathbf{w}^*,\ \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{G} \mathbf{Q}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{Q}_k\right]^{-1}\right).$$

In practice, model dimensions range in the order of millions while ensemble sizes are constrained by the hundreds and as a direct consequence, undersampling degrades the quality of analysis corrections onto the space spanned by (4.4d). To counteract the effects of sampling noise, localizations methods are commonly employed [GKM+11, CO10], in practice. For instance, methods such as covariance matrix localization (**B**-localization) [LWB18], domain localization, and observation localization (**R**-localization) [And01, HZS18, And19] are employed in operational DA scenarios. Yet another possible choice is to make use of precision matrix estimation. In this context, for instance, the use of the *spatial-predecessors* concept can be employed to obtain sparse estimators of precision matrices [LRZ+08]. The predecessors of model component i, from now on $\Pi(i,\ \delta)$, for $1 \leq i \leq n$ and a radius of influence $\delta \in \mathbb{Z}^+$, are given by the set of components whose labels are lesser than that of the i-th one. Of course, this will depend on the format employed to label components on a numerical grid. In practice, column major and row major format are commonly employed. This idea is exploited in the EnKF formulation proposed in [NRSD17, NRSD18] wherein the following estimator is employed to approximate precision matrices [BL+08c]:

$$\widehat{\mathbf{B}}_k^{-1} = \widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1} \cdot \widehat{\mathbf{V}}_k \in \mathbb{R}^{n\times n}\,, \tag{4.8a}$$

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

$$\left\{\widehat{\mathbf{V}}_k\right\}_{i,g} = \begin{cases} -\beta_{i,g,k} & ,\ g \in P(i,\delta) \\ 1 & ,\ i = g \\ 0 & ,\ otherwise \end{cases}\,, \tag{4.8b}$$

whose (sub-diagonal) elements $\beta_{i,g,k}$ are estimated by fitting linear models:

$$\mathbf{x}_{[i]_k}^T = \sum_{g\in\Pi(i,\,\delta)} \beta_{i,g,k} \cdot \mathbf{x}_{[g]_k}^T + \boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N\times 1}\,,\ 1 \leq i \leq n\,, \tag{4.8c}$$

where $\mathbf{x}_{[i]_k}^T \in \mathbb{R}^{N\times 1}$ denotes the model component i from the ensemble (4.4a). Likewise, $\boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N\times 1} \sim \mathcal{N}\left(\mathbf{0},\ \sigma^2\cdot\mathbf{I}\right)$, where the variance σ_k^2 is unknown, and the diagonal matrix $\mathbf{\Gamma}_k \in \mathbb{R}^{n\times n}$ holds the variance of residuals:

$$\begin{aligned} \{\mathbf{\Gamma}_k\}_{i,i} &= \widehat{\mathbf{var}}\left(\mathbf{x}_{[i]_k}^T - \sum_{g\in\Pi(i,\,\delta)} \beta_{i,g,k}\cdot\mathbf{x}_{[j]_k}^T\right)^{-1} \\ &\approx \mathbf{var}\left(\boldsymbol{\gamma}_{ik}\right)^{-1} = \frac{1}{\sigma_k^2} > 0\,,\ \text{with } \{\mathbf{\Gamma}_k\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[1]_k}^T\right)^{-1}\,, \end{aligned} \tag{4.8d}$$

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

4.2.2 Wind Energy Potential

The effects of climate change have triggered alarms to employ alternatives and to reduce Carbon Dioxide ($CO2$) emissions around the world. In many countries, regulation and $CO2$ reduction goals promote the substitution of fossil energy sources with Renewable Energy Sources (RES) [KST17]. For instance, China, the largest energy consumer worldwide, has an economic motivation to execute such substitution [Liu17]: traditional power systems (mainly composed of nuclear, hydro, and thermal generators) are drastically decreasing, and now, they are trying to integrate RES as a shock absorber of this situation. However, RES integration is not straightforward since it brings new issues and challenges that need to be analyzed and addressed. One of the main challenges comes from the intermittency of RES [VDVDH17]. Intermittency combines variability and uncertainty. The former is produced by the movement of large cloud systems owing to high and low-pressure areas. Uncertainty, also known as unpredictability, comes from the forecast error, which in turn depends on the numerical model (4.2). Thus, uncertainty amplification relies on model errors (i.e., physics simplifications to make numerical models computationally feasible to run). For instance, if the accuracy of the numerical model is poor, and no Data Assimilation is performed, the bias on the resulting estimate will be large concerning the actual wind speed. Thus, wind speeds can be poorly estimated, and as a direct consequence, wind energy potentials can be underestimated. Hence, Data Assimilation can be employed in this context to mitigate the impact of poor potential energy estimations via real noisy observations of wind speeds. The potential energy $p(v)$ in MegaWatts (MW) of a wind turbine given a wind speed v (km/h) can be estimated as follows:

$$p(v) = \begin{cases} P_{nom} \cdot \left(\frac{v^3 - v_c^3}{v_r^3 - v_c^3}\right) & v_c \leq v \leq v_r \\ P_{nom} & v_r \leq v \leq v_f \\ 0 & \text{otherwise} \end{cases}, \tag{4.9}$$

where v_c, v_r, v_f, and v_p are the cut-in wind speed, the rated wind speed, the cut-out wind speed, and the rated power of wind turbine, respectively. Table 4.1 shows the 12 wind turbine generators types assumed and utilized in many case studies [XB10]. The outage rate of each wind turbine reads 0.04. Commonly, the useful life of a wind turbine is about 25 years, this does not depend on its size. We also report the capital cost, and the maintenance and operating cost for each turbine, these are taken from [W$^+$08, Mas13].

Type	Rated capacity (MW)	v_c (km/h)	v_r (km/h)	v_f (km/h)	Capital cost	$M\&O$
WTG 1	0.5	10	40	80	1350	36
WTG 2	0.5	10	45	70	1350	36
WTG 3	1	12	40	80	1250	35
WTG 4	2	12	30	55	1120	30
WTG 5	1	13	33	60	1220	33
WTG 6	1	14	40	90	1250	32
WTG 7	2	15	33	50	1100	35
WTG 8	2	15	33	60	1100	30.5
WTG 9	1	15	37	70	1200	32
WTG 10	1	18	48	70	1250	32
WTG 11	2	18	45	70	1100	30
WTG 12	2	18	35	75	1100	30

Table 4.1: WTG Unit Parameters

Based on the Table 4.1, places with wind speeds below 10 km/h do not have the chance to generate electrical power from wind speeds since these are lower than the minimum cut-in wind speed across all Wind Turbine Generators (WTGs). Similarly, places with wind speeds greater than 90 km/h cannot produce electrical power because wind speeds exceed the maximum cut-out wind speed (i.e., WTG 6). Although, we do not consider economic impacts of wind-farm placements, it is important to note that wind-speed constraints have economic implications. For instance, for a place with bimodal wind speeds of 40 km/h and 11 km/h, WTGs 1 and 2 can be employed while the rest of them must be discarded in spite of the last are cheaper.

4.3 Proposed Framework

In this section, we develop an adjoint-free 4D-Var framework for potential energy estimation. The framework is divided into four stages. First, we build an ensemble of snapshots at observation times by employing a numerical model which can forecast wind components. Second, these snapshots are employed to build control spaces via a modified Cholesky decomposition. Third, the control spaces are utilized to obtain initial conditions whose wind forecasts fit a set of time spaced observations. Lastly, forecasts of wind components are employed to estimate forecasts of wind speeds, which in turn allow us to forecast potential energies of Wind Turbine Generators (WTGs). Since, in practice, model resolutions range in the order of the millions, we develop a matrix-free analysis formulation to avoid the direct inversion of linear systems during assimilation steps. All these stages are clearly detailed next.

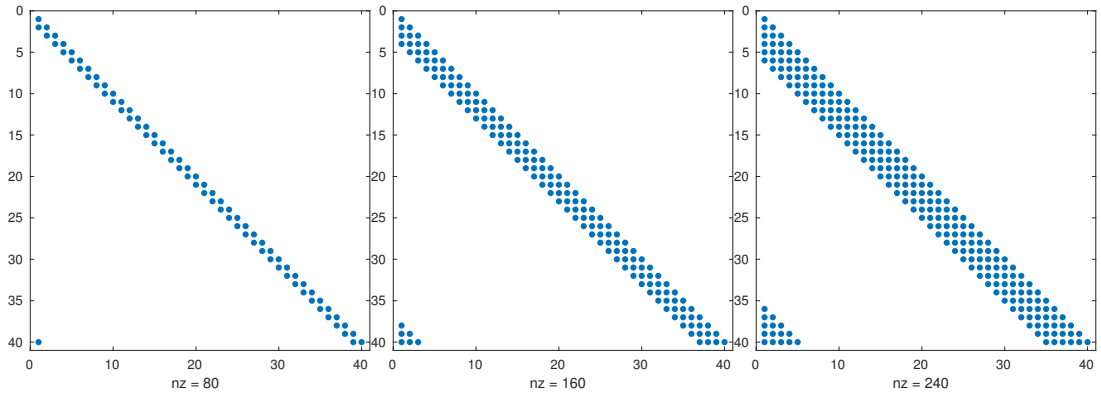
4.3.1 Building an Ensemble of Snapshots

Initially, we choose a numerical model which mimics the dynamics of wind components in places of interest. For this purpose, numerical models such as the Atmospheric General Circulation Model (AT-GCM Speedy) [AKW11] and the Weather Research Forecast (WRF) Model [MDG+05, MCD+01] can be employed. Once the numerical model is chosen, snapshots

of an ensemble of model realizations (4.4a) are taken at $G + 1$ observation times. At step k, for $0 \leq k \leq G$, the background ensemble $\mathbf{X}_k^b$ (4.4a) is employed to estimate a full-rank square-root approximation of the precision matrix of background errors $\mathbf{B}_k^{-1}$ via a modified Cholesky decomposition (4.8a):

$$\widehat{\mathbf{B}}_k^{-1/2} = \widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1/2} \in \mathbb{R}^{n \times n} . \tag{4.10}$$

At this step, we choose a radius of influence (localization radius) δ to compute the factor $\widehat{\mathbf{V}}_k^T$. Beyond the scope of this radius (and the predecessors of model components) all components of $\widehat{\mathbf{V}}_k^T$ are assumed zero. We exploit the fact that, when the error correlations of two model components are conditionally independent (given a radius of influence δ), their corresponding entry in the precision matrix of background errors is zero. This results in a sparse Cholesky factor $\widehat{\mathbf{V}}_k^T$ and even more, a localized square-root precision matrix. In this manner, the impact of sampling errors can be mitigated in the square-root approximations (4.10). Some structures of $\widehat{\mathbf{V}}_k$ are shown in figure 4.1 for a one dimensional grid and different values of δ, cyclic boundary conditions are assumed for physics/dynamics.



(a) $\delta = 1$ (b) $\delta = 3$ (c) $\delta = 5$

Figure 4.1: Structure of the Cholesky factor $\widehat{\mathbf{V}}_k$ as a function of the localization radius δ.

The square-root approximations (4.10) serve as control spaces onto which analysis increments can be estimated, therefore, the analysis increment at observation time k can be written as follows:

$$\mathbf{x}_k - \overline{\mathbf{x}}_k^b \in \mathbf{range} \left\{ \widehat{\mathbf{B}}_k^{1/2} \right\} ,$$

or equivalently:

$$\mathbf{x}_k = \overline{\mathbf{x}}_k^b + \widehat{\mathbf{B}}_k^{1/2} \cdot \boldsymbol{\alpha} \in \mathbb{R}^{n \times 1} , \tag{4.11}$$

where $\widehat{\mathbf{B}}_k^{1/2} = \left[\widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1/2} \right]^{-1} \in \mathbb{R}^{n \times n}$, and $\boldsymbol{\alpha} \in \mathbb{R}^{n \times 1}$ is a vector in redundant coordinates to be determined later. We assume that:

$$\mathbf{range} \left\{ \left[\widehat{\mathbf{V}}_k^T \cdot \widehat{\mathbf{\Gamma}}_k^{-1/2} \right]^{-1} \right\} \approx \mathbf{range} \left\{ \mathbf{B}_k^{1/2} \right\} .$$

Note that, since the square root approximations (4.10) are full-rank, the dimension of the spaces (4.11) equal those of the range of $\mathbf{B}^{1/2}$. This differs from what is usually employed in the literature: a control space whose dimension equals the ensemble size (4.5) and therefore, analysis increments can be highly impacted by sampling noise. We then expect to capture all error dynamics onto the spaces (4.10).

Since the initial background error covariance matrix $\mathbf{B}_0$ onto the control space (4.11) is nothing but the identity matrix, the following error statistics hold for the prior weights $\boldsymbol{\alpha}^{b(e)}$:

$$\boldsymbol{\alpha}^{b(e)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \ , \ \text{for } 1 \leq e \leq N \, .$$

Due to this, the 4D-Var cost function (4.1) onto the control space (4.11) can be written as follows:

$$\mathcal{J}(\mathbf{x}_0) \ = \ \mathcal{J}\left(\overline{\mathbf{x}}_0^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}\right) = \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) = \frac{1}{2} \cdot \|\boldsymbol{\alpha}\|^2 + \frac{1}{2} \cdot \sum_{k=0}^{G} \left\| \widetilde{\mathbf{d}}_k - \widetilde{\mathbf{Q}}_k \cdot \boldsymbol{\alpha} \right\|_{\mathbf{R}_k^{-1}}^2 , \quad (4.12)$$

where $\widetilde{\mathbf{d}}_k = \mathbf{y}_k - \mathbf{H} \cdot \overline{\mathbf{x}}_k^b \in \mathbb{R}^{m \times 1}$, and $\widetilde{\mathbf{Q}}_k = \mathbf{H} \cdot \widehat{\mathbf{B}}_k^{1/2} \in \mathbb{R}^{m \times n}$. Again, this cost function does not rely on the numerical model (4.2).

4.3.2 Adjoint-Free 4D-Var Optimization

Once the control spaces are estimated across observation times, the adjoint-free optimization problem to solve reads:

$$\boldsymbol{\alpha}^a = \arg\min_{\boldsymbol{\alpha}} \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) \ . \quad (4.13a)$$

The gradient of this cost function can be written as follows:

$$\nabla_{\boldsymbol{\alpha}} \widetilde{\mathcal{J}}(\boldsymbol{\alpha}) = \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right] \cdot \boldsymbol{\alpha} - \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{d}}_k \, ,$$

whose root reads:

$$\boldsymbol{\alpha}^a = \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1} \cdot \left[\sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{d}}_k \right] , \quad (4.13b)$$

and therefore an estimate of the initial analysis state (4.3) can be computed as follows:

$$\overline{\mathbf{x}}_0^a = \overline{\mathbf{x}}_0^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}^a \, , \quad (4.13c)$$

whose model trajectory provides a forecast which accounts for the given data into the assimilation window. Note that, the closed form expression (4.13b) for the optimal weights (4.13a) is possible since we consider linear observation operators in our formulation. The posterior ensemble onto the control space can then be built by using a square root approximation of the information matrix in (4.13b), it can be easily shown that the posterior error statistics read:

$$\boldsymbol{\alpha}^{a[e]} \sim \mathcal{N}\left(\boldsymbol{\alpha}^a, \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1} \right) , \ \text{for } 1 \leq e \leq N \, , \quad (4.14)$$

with corresponding analysis members in the model space:

$$\mathbf{x}_0^{a[e]} = \overline{\mathbf{x}}^b + \widehat{\mathbf{B}}_0^{1/2} \cdot \boldsymbol{\alpha}^{a[e]} .$$

Then, the analysis members of the initial ensemble are propagated in time

$$\mathbf{x}_f^{a[e]} = \mathcal{M}_{t_{f-1} \rightarrow t_f} \left(\mathbf{x}_f^{a[e]} \right) , \text{ for } 1 \leq f \leq G \text{ and } 1 \leq eleN ,$$

from which an estimate of the optimal trajectory

$$\mathbf{x}_k^a \approx \overline{\mathbf{x}}_k^a = \frac{1}{N} \cdot \sum_{e=1}^{N} \mathbf{x}_k^{a[e]} , \tag{4.15}$$

and his uncertainty (i.e., by employing a modified Cholesky decomposition on the ensemble members at time k) can be obtained. Note that, the posterior mode (4.13c) can be written as follows:

$$\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_0^b = \widehat{\mathbf{B}}_0^{1/2} \cdot \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1} \cdot \left[\sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{d}}_k \right] ,$$

which is nothing but a linear transformation of the prior increment to the posterior one. In this sense, the analysis step is similar to that of square root filter formulations. However, we compute the analysis increments of the initial ensemble members by using synthetic data, which is statistically consistent with the posterior error distribution:

$$\overline{\mathbf{x}}_0^{a[e]} - \overline{\mathbf{x}}_0^{b[e]} = \left[\widehat{\mathbf{V}}^T \cdot \widehat{\boldsymbol{\Gamma}}^{-1/2} \right] \cdot \left[\boldsymbol{\alpha}^a + \left[\mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k \right]^{-1/2} \cdot \boldsymbol{\xi}^{[e]} \right] , \text{ with } \boldsymbol{\xi}^{[e]} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) .$$

Readily:

$$\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_0^b = \mathbb{E} \left(\frac{1}{N} \cdot \sum_{e=1}^{N} \overline{\mathbf{x}}_0^{a[e]} - \overline{\mathbf{x}}_0^{b[e]} \right) ,$$

and therefore, in spite of the posterior mode of the analysis distribution can be estimated via a linear transformation of the initial background increments, the analysis increments of the initial ensemble are actually computed by employing synthetic data. This places our proposed filter formulation into the family of stochastic formulations of data assimilation methods.

Notice, given the special structure of our estimator $\mathbf{B}_k^{-1/2}$, the Woodbury matrix identity can be exploited to avoid direct inversions [RSA15]. We denote this filter implementation *Four Dimensional Variational Data Assimilation via a Modified Cholesky Decomposition* (4D-Var-MC).

4.3.3 Post-Processing of Data, Potential Energy Estimation

Once the model trajectory is computed for each ensemble member, we proceed to map wind fields to wind energy potentials in two steps:

1. whenever is necessary, the wind components of ensemble members are mapped to wind-speeds,
2. this subset of information is exploited to estimate the wind energy potential of each analysis ensemble member:

$$\widehat{\mathbf{x}}_k^{a[e]} = w\left(\mathbf{x}_k^{a[e]}\right), \text{ for } 1 \leq e \leq N, \text{ and } 0 \leq k \leq G, \tag{4.16}$$

where $w : \mathbb{R}^{n\times 1} \to \mathbb{R}^{h\times 1}$ is a function that maps model states to potential energy states (this is, for each ensemble member, its wind-speed components are mapped to wind energy potentials), where h is the number of wind-speed components (with $h \leq n$), and $\widehat{\mathbf{x}}_k^{a[e]} \in \mathbb{R}^{h\times 1}$ is the k-th transformed member. The mapping process depends on the wind turbine employed, for instance, one can consider the wind turbines discussed in Section 4.2.2.

Note that, the empirical moments of the samples (4.16) can be exploited to estimate mean and standard deviations of wind energy potential capacities. Besides, covariances of such samples can be estimated via a modified Cholesky decomposition to understand better (and to estimate) their uncertainties.

4.3.4 Further Comments: Matrix-Free Formulation of the 4D-Var-MC

In practice, the number of model components n range in the order of the millions and therefore, matrix computations can be constrained by computational resources. For instance, the direct inversion of (4.13b) is prohibitive. Thus, it is mandatory to count with a matrix-free implementation of any data assimilation process. Following the ideas discussed in [NR17], we can develop a matrix-free equation for the analysis step of the 4D-Var-MC implementation. We can proceed as follows, consider:

$$\boldsymbol{\Omega} = [\boldsymbol{\Omega}_0,\ \boldsymbol{\Omega}_1,\ \ldots,\ \boldsymbol{\Omega}_G] \in \mathbb{R}^{n\times O}$$

where $\boldsymbol{\Omega}_k = \widetilde{\mathbf{Q}}_k \cdot \mathbf{R}_k^{-1/2} \in \mathbb{R}^{n\times m}$, and $O = m \cdot G$, the precision matrix in (4.14) can be written as follows,

$$\widehat{\mathbf{A}}^{-1} = \mathbf{I} + \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}^T = \mathbf{T}^T\cdot\mathbf{C}\cdot\mathbf{T} + \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}^T = \mathbf{T}^T\cdot\mathbf{C}\cdot\mathbf{T} + \sum_{o=1}^{O}\cdot\boldsymbol{\omega}^{[o]}\cdot\left[\boldsymbol{\omega}^{[o]}\right]^T, \tag{4.17}$$

where $\boldsymbol{\omega}^{[o]} \in \mathbb{R}^{n\times 1}$ is the o-th column of matrix $\boldsymbol{\Omega}$, for $0 \leq o \leq O$, and $\mathbf{I} = \mathbf{T}^T\cdot\mathbf{C}\cdot\mathbf{T}$ is the Cholesky decomposition of $\mathbf{I}$ (all factors equal the identity matrix). Consider the sequence of matrices,

$$\begin{aligned}
\widehat{\mathbf{A}}^{(0)} &= \left[\mathbf{V}^{(0)}\right]^T\cdot\boldsymbol{\Gamma}^{(0)}\cdot\left[\mathbf{V}^{(0)}\right] = \mathbf{T}^T\cdot\mathbf{C}\cdot\mathbf{T} = \mathbf{I},\\
\widehat{\mathbf{A}}^{(1)} &= \widehat{\mathbf{A}}^{(0)} + \boldsymbol{\omega}^{[1]}\cdot\left[\boldsymbol{\omega}^{[1]}\right]^T = \left[\mathbf{V}^{(1)}\right]^T\cdot\boldsymbol{\Gamma}^{(1)}\cdot\left[\mathbf{V}^{(1)}\right],\\
\widehat{\mathbf{A}}^{(2)} &= \widehat{\mathbf{A}}^{(1)} + \boldsymbol{\omega}^{[2]}\cdot\left[\boldsymbol{\omega}^{[2]}\right]^T = \left[\mathbf{V}^{(2)}\right]^T\cdot\boldsymbol{\Gamma}^{(2)}\cdot\left[\mathbf{V}^{(2)}\right],\\
&\vdots\\
\widehat{\mathbf{A}}^{(O)} &= \widehat{\mathbf{A}}^{(O-1)} + \boldsymbol{\omega}^{[O]}\cdot\left[\boldsymbol{\omega}^{[O]}\right]^T = \left[\mathbf{V}^{(O)}\right]^T\cdot\boldsymbol{\Gamma}^{(O)}\cdot\left[\mathbf{V}^{(O)}\right] = \widehat{\mathbf{V}}^T\cdot\widehat{\boldsymbol{\Gamma}}\cdot\widehat{\mathbf{V}} = \widehat{\mathbf{A}}^{-1},
\end{aligned}$$

where $\mathbf{V}^{(0)} \in \mathbb{R}^{n\times n}$ and $\mathbf{\Gamma}^{(0)} \in \mathbb{R}^{n\times n}$ are the factors of the Cholesky decomposition of the identity matrix $\mathbf{I}$. Among iteration steps o, for $0 \leq o \leq O$, one can see that:

$$\begin{aligned}\widehat{\mathbf{A}}^{(o)} &= \left[\mathbf{V}^{(o-1)}\right]^T \cdot \mathbf{\Gamma}^{(o-1)} \cdot \left[\mathbf{V}^{(o-1)}\right] + \boldsymbol{\omega}^{[o]} \cdot \left[\boldsymbol{\omega}^{[o]}\right]^T \\ &= \left[\mathbf{V}^{(o-1)}\right]^T \cdot \left[\mathbf{\Gamma}^{(o-1)} + \boldsymbol{\gamma}^{(o)} \cdot \left[\boldsymbol{\gamma}^{(o)}\right]^T\right] \cdot \left[\mathbf{V}^{(o-1)}\right], \end{aligned} \tag{4.18}$$

where $\left[\mathbf{V}^{(j-1)}\right]^T \cdot \boldsymbol{\gamma}^{(j)} = \boldsymbol{\omega}^{[j]} \in \mathbb{R}^{n\times 1}$. Via the Cholesky factors of,

$$\mathbf{\Gamma}^{(o-1)} + \boldsymbol{\gamma}^{(o)} \cdot \left[\boldsymbol{\gamma}^{(o)}\right]^T = \left[\widetilde{\mathbf{V}}^{(o-1)}\right]^T \cdot \mathbf{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)}\right], \tag{4.19}$$

and, by considering equation (4.18), the matrix $\widehat{\mathbf{A}}^{(j)}$ can be decomposed as follows,

$$\widehat{\mathbf{A}}^{(o)} = \left[\widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}\right]^T \cdot \mathbf{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}\right] = \left[\mathbf{V}^{(o)}\right]^T \cdot \mathbf{\Gamma}^{(o)} \cdot \left[\mathbf{V}^{(o)}\right],$$

where $\mathbf{V}^{(o)} = \widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)} \in \mathbb{R}^{n\times n}$. By taking a close look at equation (4.19), the elements of factors $\widetilde{\mathbf{V}}^{(o-1)}$ and $\widetilde{\mathbf{\Gamma}}^{(o)}$ can be easily related to those of $\mathbf{\Gamma}^{(o-1)}$ and $\boldsymbol{\gamma}^{(o)}$ via the Dolittle's method for matrix factorization, for instance, we can note that:

$$\left[\widetilde{\mathbf{V}}^{(o-1)}\right]^T \cdot \mathbf{\Gamma}^{(o)} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)}\right]_{i,b} = \delta_{i,b} \cdot \mathbf{\Gamma}^{(o-1)}_{i,i} + \boldsymbol{\gamma}^{(o)}_{i} \cdot \boldsymbol{\gamma}^{(o)}_{b},$$

and therefore, the next relations hold:

$$\mathbf{\Gamma}^{(o)}_{n,n} = \left[\boldsymbol{\gamma}^{(o)}_{n}\right]^2 + \mathbf{\Gamma}^{(o-1)}_{n,n}, \tag{4.20a}$$

$$\widetilde{\mathbf{V}}^{(o-1)}_{i,b} = \frac{1}{\mathbf{\Gamma}^{(o)}_{i,i}} \cdot \left[\boldsymbol{\gamma}^{(o)}_{i} \cdot \boldsymbol{\gamma}^{(o)}_{b} - \sum_{q\in\Pi(i,\,\delta)} \mathbf{\Gamma}^{(o)}_{q,q} \cdot \widetilde{\mathbf{V}}^{(o-1)}_{q,i} \cdot \widetilde{\mathbf{V}}^{(o-1)}_{q,b}\right], \tag{4.20b}$$

and

$$\mathbf{\Gamma}^{(o)}_{i,i} = \left[\boldsymbol{\gamma}^{(o)}_{i}\right]^2 + \mathbf{\Gamma}^{(o-1)}_{i,i} - \sum_{q\in\Pi(i,\,\delta)} \mathbf{\Gamma}^{(o)}_{q,q} \cdot \left[\widetilde{\mathbf{V}}^{(o-1)}_{q,i}\right]^2, \tag{4.20c}$$

for $1 \leq i \leq n-1$ and $b \in \Pi(i,\,\delta)$, where the Kronecker delta function $\delta_{i,j}$ equals 1 for $i = j$ and 0 otherwise, and the diagonal entries of matrix $\widetilde{\mathbf{V}}^{(o-1)}$ are all equal to one. In algorithm 10, we show how the Cholesky factors $\mathbf{V}^{(o)}$ and $\mathbf{\Gamma}^{(o)}$ can be updated with the information brought by the rank-one matrix $\boldsymbol{\omega}^{[o]} \cdot \left[\boldsymbol{\omega}^{[o]}\right]^T$, the general updating process of factors $\widehat{\mathbf{V}}^{(0)}$ and $\mathbf{\Gamma}^{(0)}$ for the estimation of $\widehat{\mathbf{A}}^{-1}$ are detailed in algorithm 11. The number of long computations is reported as well for each step of our proposed updating process. We let by φ the largest number of non-zero elements per row in the $\mathbf{V}^{(o)}$ factor. This value will depend on chosen radius of influence δ during assimilation steps, and intuitively $\varphi \ll n$. Note that, $\mathbf{V}^{(o-1)}$ and $\widetilde{\mathbf{V}}^{(o-1)}$ hold the same

structure since this is given by the predecessors of i. Thus, the structure (form) of $\mathbf{V}^{(o-1)}$ is preserved in $\mathbf{V}^{(o)}$. Consequently, we can hold a desired structure in the resulting estimator $\widehat{\mathbf{V}}$, for instance, this can be equal to that of $\mathbf{V}$. Note that, the number of long computations in the 4D-Var-MC reads:

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

which increases linearly regarding the number of model components. This makes the proposed filter implementation attractive in operational scenarios where the number of model components ranges in the order of millions.

Algorithm 10 Rank-one update of factors $\mathbf{V}^{(o-1)}$ and $\mathbf{\Gamma}^{(o-1)}$ via Doolittle's method.

1: **function** UPDATE_RANK_ONE($\mathbf{V}^{(o-1)}$, $\mathbf{\Gamma}^{(o-1)}$, $\boldsymbol{\omega}^{[o]}$) ▷ **COST**
2: Solve $\left[\mathbf{V}^{(o-1)}\right]^T \cdot \mathbf{p}^{(o)} = \boldsymbol{\omega}^{[o]}$. ▷ $\mathcal{O}(\varphi \cdot n)$
3: Compute $\mathbf{\Gamma}_{n,n}^{(o)}$ via equation (4.20a). ▷ $\mathcal{O}(1)$
4: **for** $i = n-1 \rightarrow 1$ **do** ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
5: Let $\widetilde{\mathbf{V}}_{i,i}^{(o-1)} \leftarrow 1$. ▷ $\mathcal{O}(1)$
6: **for** $k \in \Pi(i, \delta)$ **do** ▷ $\mathcal{O}\left(\varphi^2\right)$
7: Compute $\widetilde{\mathbf{V}}_{i,k}^{(o-1)}$ according to (4.20b). ▷ $\mathcal{O}(\varphi)$
8: **end for**
9: Compute $\mathbf{\Gamma}_{i,i}^{(o)}$ via equation (4.20c). ▷ $\mathcal{O}(\varphi)$
10: **end for**
11: Let $\mathbf{V}^{(o)} \leftarrow \widetilde{\mathbf{V}}^{(o-1)} \cdot \mathbf{V}^{(o-1)}$. ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
12: **return** $\mathbf{V}^{(o)}$, $\mathbf{\Gamma}^{(o)}$
13: **end function**

Algorithm 11 Computing the posterior factors $\widehat{\mathbf{V}}$ and $\mathbf{\Gamma}$ of $\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{V}}^T \cdot \mathbf{\Gamma} \cdot \widehat{\mathbf{V}} = \mathbf{I} + \sum_{k=0}^{G} \widetilde{\mathbf{Q}}_k^T \cdot \mathbf{R}_k^{-1} \cdot \widetilde{\mathbf{Q}}_k$.

1: **function** COMPUTE_POSTERIOR_CHOLESKY_FACTORS($\mathbf{V}^{(0)}$, $\mathbf{\Gamma}^{(0)}$, $\mathbf{H}$, $\mathbf{R}$) ▷ **COST**
2: Let $\Omega \leftarrow \left[\mathbf{H}_0^T \cdot \mathbf{R}_0^{-1/2}, \mathbf{H}_1^T \cdot \mathbf{R}_1^{-1/2}, \ldots, \mathbf{H}_G^T \cdot \mathbf{R}_G^{-1/2}\right]$. ▷ $\mathcal{O}(O \cdot n)$
3: Let $O \leftarrow m \cdot G$
4: **for** $j = 1 \rightarrow O$ **do** ▷ O times line 4, $\mathcal{O}\left(\varphi^2 \cdot O \cdot n\right)$
5: Let $[\mathbf{V}^{(j)}, \mathbf{\Gamma}^{(j)}] \leftarrow$ UPDATE_RANK_ONE $\left(\mathbf{V}^{(j-1)}, \mathbf{\Gamma}^{(j-1)}, \boldsymbol{\omega}^{[j]}\right)$ ▷ $\mathcal{O}\left(\varphi^2 \cdot n\right)$
6: **end for**
7: **return** $\mathbf{V}^{(O)}$ as $\widehat{\mathbf{V}}$, $\mathbf{\Gamma}^{(O)}$ as $\mathbf{\Gamma}$.
8: **end function**

Now, we are ready to test our proposed framework.

4.4 Numerical Results

In this section, we employ our proposed framework by using the Atmospheric General Circulation Model (AT-GCM) Speedy [BKKM04]. This model is a general circulation model that

mimics the behavior of the atmosphere across different pressure levels [Miy11]. The number of numerical layers in this model is 7, and we employ a T-30 spectral model resolution (96×48 grid components) for the space discretization of each model layer [Mol03, KMB06]. The number of physical variables is 5. These are detailed in the Table 4.2 with their corresponding units and number of numerical layers.

Name	Notation	Units	Number of Layers
Temperature	T	K	7
Zonal Wind Component	u	m/s	7
Meridional Wind Component	v	m/s	7
Specific Humidity	Q	g/kg	7
Pressure	T	K	1

Table 4.2: Physical variables of the AT-GCM Speedy model.

Note that the total number of model components to be estimated reads $n = 133,632$. We let the number of model realizations (ensemble size) as $N = 30$ for all experimental scenarios. In this case, the model resolution is approximately 4,454 times larger than the sample size ($n \gg N$), which is very common in operational DA scenarios. Additional details of the experimental settings are described below, some of them are similar to those detailed in [MKI14]:

- Starting with a system in equilibrium, the model is integrated over a long time period to obtain an initial condition whose dynamics are consistent with those of the SPEEDY model.
- The initial condition is perturbed N times and propagated over a long-time period from which the initial background ensemble is obtained.
- We employ the trajectory of the initial condition as the reference one. This reference trajectory serves to build synthetic observations. Besides, we will consider that the actual potential capacities of WTGs are based on this solution.
- We let the standard deviations of errors in the observations as follows:
  - Temperature $1\,K$.
  - Zonal Wind Component $1\,m/s$.
  - Meridional Wind Component $1\,m/s$.
  - Specific Humidity $10^{-3}\,g/kg$.
  - Pressure $100\,hPa$.
- 50% of model components are observed during assimilation steps. This linear observation operator is shown in figure 4.2.
- Observations are available every six hours (6 h).

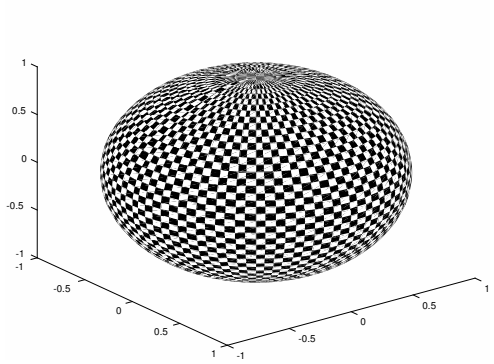


Figure 4.2: Linear observation operator during assimilation steps. Shaded regions denote observed components (observations) from the model state. The operator is replicated across all numerical layers.

- The experiments are performed under perfect model assumptions.
- The number of assimilation steps reads $G = 15$. Thus, the simulation times is 7.5 days.
- We use the wind turbines discussed in section 4.2.2 for computing wind potential energies.
- To estimate wind speeds, the wind fields (zonal and meridional components) are taken from the numerical grid at the pressure level 100 *mb*.
- Our results are compared with those obtained by the 4D-EnKF formulation.
- We employ the $L-2$ error norm as a measure of accuracy for the estimation of wind energy potential:

$$\zeta_k = \left\| p_k(v) - p_k^*(v) \right\| , \tag{4.21}$$

  where $p^*(v)_k$ is the reference wind energy potential, and $p_k(v)$ is the estimated one by a filter implementation. Likewise, k stands for observation time and v for wind speed.
- The Root-Mean-Square-Error (RMSE) provides an estimate of the performance of a filter for a given assimilation window:

$$RMSE = \sqrt{\frac{1}{G} \cdot \zeta_k^2} .$$

- We estimate the potential energy capacities of Wind Turbines Generators (WTGs) discussed in section 4.2.2.
- Our numerical results are compared with those of the 4D-EnKF formulation discussed in Section 4.2.

4.4.1 Results with $p = 50\%$ of observations from the model state

The $L - 2$ error norms (4.21) of wind-energy-potential estimations for an ensemble size of $N = 20$ is shown in figure 4.3. We employ a log scale in the y axis to render the text easier to read. As can be seen, for all WTGs, the compared filter implementations provide better estimates of potential generations than those obtained by pure forecasts, as should be expected. Thus, regardless of the employed DA method, the accuracy of forecasts can be improved by injecting real observations of the dynamical system. This can be beneficial for taking actions on whether to employ or not green sources of energy during, for instance, industrial operations. In all cases, on average, the estimated analysis trajectories in the 4D-Var-MC context outperform those computed by the 4D-EnKF formulation. In the 4D-Var-MC, the dimension of control spaces equals those of model one; therefore, we have enough degrees of freedom to capture most of the directions where errors grow faster. This allows our proposed implementation to properly correct initial background states with the information brought by observations in time. Besides, the initial analysis state (initial condition of the initial value problem) relies on the quality of the estimated background error correlations:

$$\left\|\Delta \mathbf{x}_0^a\right\|_{\mathbf{B}_0^{-1}} \approx \left\|\mathbf{x}^a - \mathbf{x}^b\right\|_{\widehat{\mathbf{B}}_0^{-1}} \approx \left\|\mathbf{x}^a - \mathbf{x}^b\right\|^2_{\left[\widehat{\mathbf{V}}_0^T \cdot \widehat{\mathbf{\Gamma}}_0^{-1} \cdot \widehat{\mathbf{V}}_0\right]^{-1}} .$$

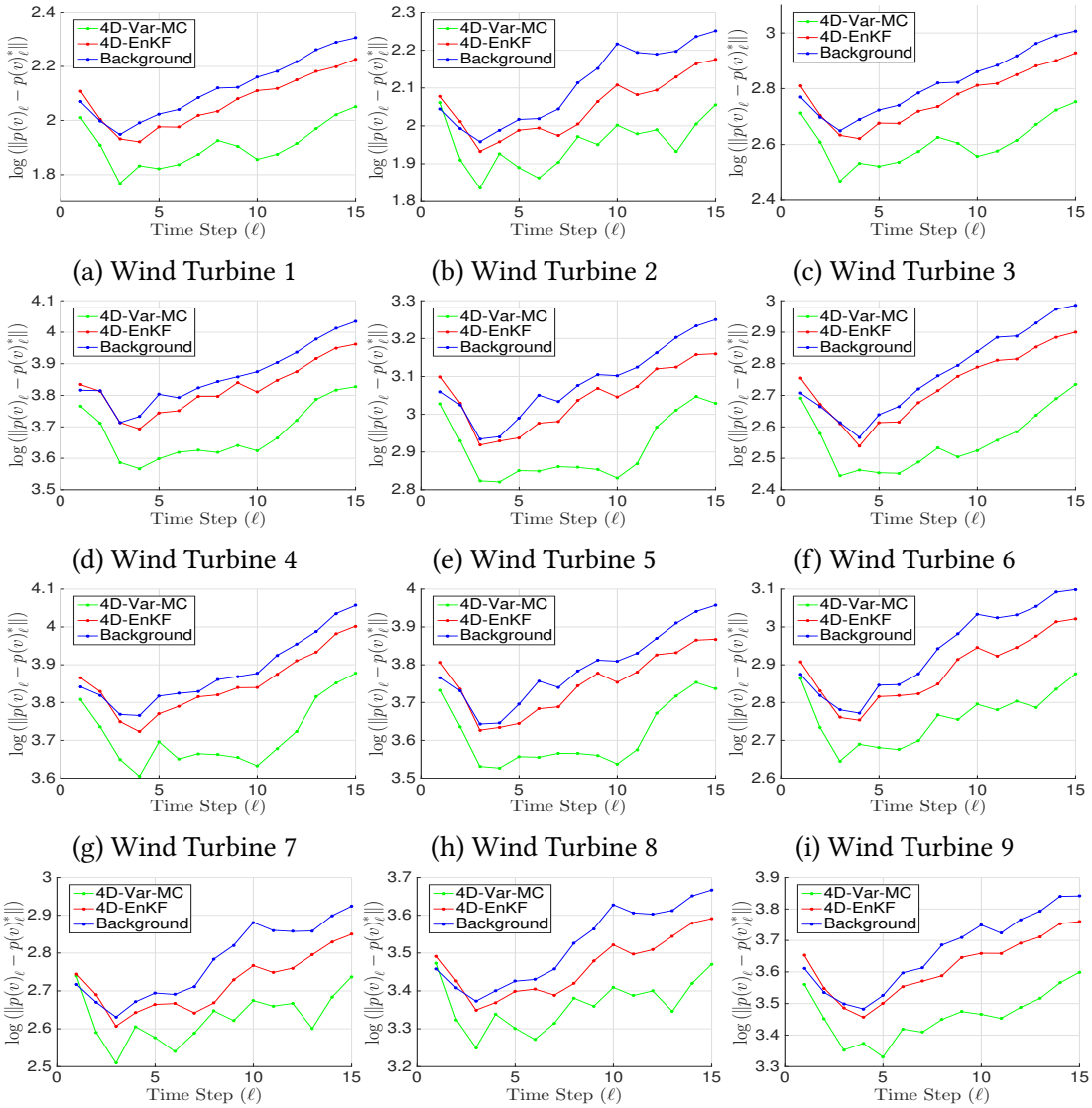
As is proven in [NRSD18, Theorem 1], the precision matrix estimator (4.8a) converges to the actual precision matrix $\mathbf{B}_0^{-1}$ as long as $\log(n)/N$ goes to zero. This value, under the current experimental settings, reads ~ 0.170, which can explain as well why the accuracy of the 4D-EnKF-MC method is better than that of the 4D-EnKF. On the other hand, the control space in the 4D-EnKF formulation relies on the ensemble size, whose dimension is much lesser than that of the model one. Consequently, this sub-space can be highly sensitive to sampling noise, which can create spurious correlations among distant model components. Besides, there is no guaranty that such sub-space can capture the leading directions where errors grow faster. This results in the poor estimation of the analysis increments of the initial ensemble mean and, as a direct consequence, the analysis members of the initial ensemble. For this reason, the benefits of increasing the number of model realizations are just evident in the 4D-EnKF context; for instance, the accuracy of this formulation improves drastically as the ensemble size increases. The Table 4.3 provides an overview of the compared filter implementations in terms of performance (RMSE values) and all parameter configurations, RMSE values are computed based on the analysis trajectory (estimated initial condition). It is clear that, on average, our proposed filter implementation outperforms the traditional 4D-EnkF one in terms of accuracy. In general, both filters formulations can improve their performance as the ensemble size is increased.

In figure 4.4, snapshots of the estimated initial wind-energy-potential are shown for the proposed 4D-Var-MC method. Their corresponding standard deviations of errors (based on analysis ensembles) are shown in figure 4.5. Recall that this initial state is our estimate of the initial condition in the optimization problem (4.3). As can be expected, most of the wind-energy-potential is produced on the ocean where wind speeds get the largest rise for all wind turbines. This serves as a validation test since no wind farms (turbines) can be placed under such a place. However, countries well-known for their potential capacities are just evident in these results, for instance, countries such as those from Latin American and the Caribbean and

Africa. Moreover, by taking a close look at standard deviations, one can see that forecasts are obtained with low uncertainties for all filter implementations. This, together with the RMSE results, show that the proposed framework can be employed to estimate wind energy potentials with high accuracy and low variations.

Notice, green sources of energy such as those based on wind speeds are impacted by three observable conditions (which can be implicitly evidenced in our numerical results): variability, unpredictability, and placement. Variability obeys to the fact that as time moves forward, wind speeds can drastically vary, this impacts the potential energy that can be generated by WTGs. Regardless the DA method employed to estimate wind potential capacities, unpredictability is always present: numerical forecasts are imperfect and even more, uncertain. Placement of WTGs is crucial, as can be seen in figure 4.4, no all WTGs can properly work in different zones of the globe, this is, electrical power via WTGs can drastically vary from one place to other. We can stood out the importance of employing WTGs as sources of energy based on wind speeds in different regions of the globe but, we cannot argue which turbine is better than others. To do this analysis, we should consider other relevant factors such as wind speed variability, WTGs constraints, and economic considerations, the last two are out of the scope of our analysis.

Consider, again, the WTG parameters reported in the Table 4.1 and the initial snapshots reported in figures 4.4 and 4.5. Note that, in most places in the globe, WTGs 1 and 2 can be installed to guaranty electrical power from wind speeds; these WTGs are the ones with rate-capacity 0.5. Note that, as the cut-in wind speed and the rate-capacity increase, the electrical energy generation of WTGs can be impacted. For instance, near the poles, the power generation is almost null for WTGs with the largest cut-in wind speed parameters. WTGs with rated-capacity of 1 are the ones that have large variability across different places in the world. Lastly, the largest amount of energy across different places in the domain can be obtained for WTGs with the largest rate-capacity values. However, WTGs with low rate-capacity values are the ones whose numerical forecasts are obtained with lesser variability (i.e., in figure 4.5 the standard deviation of errors has a homogeneous behavior across different regions of the domain). As the rate-capacity increases, the variance of the standard deviation of errors increases as well. This means, more variability of errors can be evidenced across different parts of the world. Thus, WTGs with large rate capacities provide forecasts with a large amount of clean energy, but these come with large uncertainties in certain regions of the world, which can difficult decision making.



(a) Wind Turbine 1 (b) Wind Turbine 2 (c) Wind Turbine 3

(d) Wind Turbine 4 (e) Wind Turbine 5 (f) Wind Turbine 6

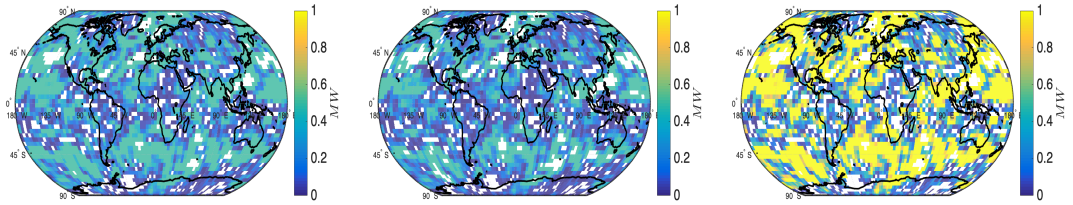
(g) Wind Turbine 7 (h) Wind Turbine 8 (i) Wind Turbine 9

(j) Wind Turbine 10 (k) Wind Turbine 11 (l) Wind Turbine 12

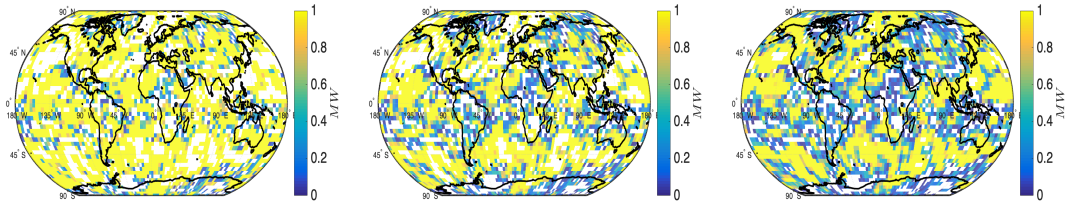
Figure 4.3: Error norms of wind energy potential estimations for the compared filter implementations. The ensemble size reads $N = 20$. 12 wind turbines are employed for the experiments. Units are in MW.

	N			
	20		40	
Wind Turbine Generator (WTG)	4D-EnKF	4D-Var-MC	4D-EnKF	4D-Var-MC
WTG 1	0.11713	0.09927	0.11211	0.10098
WTG 2	0.11481	0.10391	0.11143	0.10596
WTG 3	0.23608	0.20008	0.22597	0.20354
WTG 4	0.67597	0.58524	0.65088	0.59049
WTG 5	0.31010	0.27093	0.29692	0.27475
WTG 6	0.22808	0.19058	0.21876	0.19488
WTG 7	0.69412	0.60609	0.67170	0.61034
WTG 8	0.62901	0.54901	0.60232	0.55692
WTG 9	0.26503	0.23305	0.25554	0.23756
WTG 10	0.22425	0.20466	0.21797	0.20872
WTG 11	0.47221	0.42631	0.45824	0.43497
WTG 12	0.55006	0.47031	0.52981	0.47967

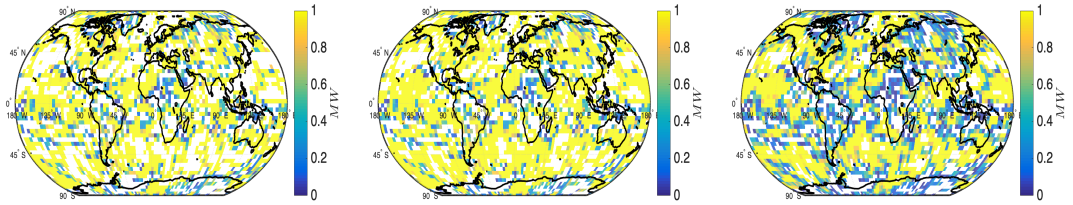
Table 4.3: Root-Mean-Square-Error values of wind energy potential estimations. Two ensemble sizes are tried during the experiments.



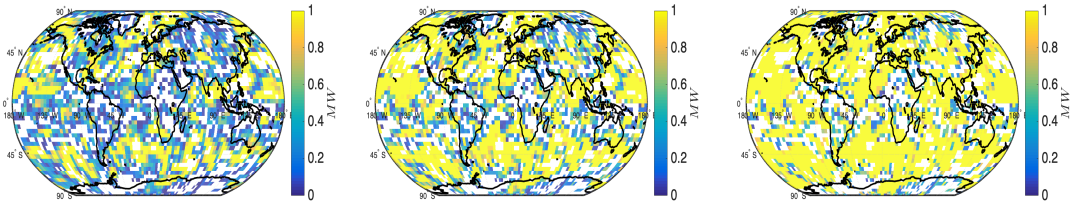
(a) WTG 1 (b) WTG 2 (c) WTG 3



(d) WTG 4 (e) WTG 5 (f) WTG 6

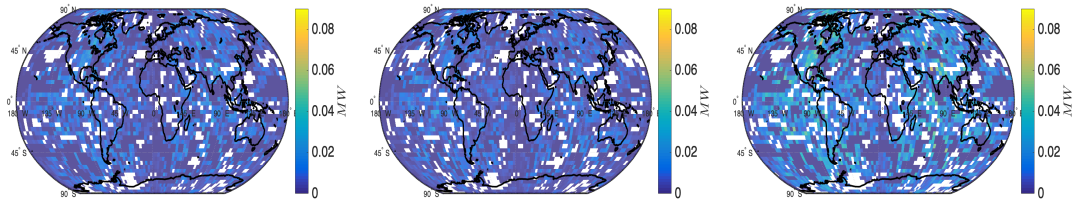


(g) WTG 7 (h) WTG 8 (i) WTG 9

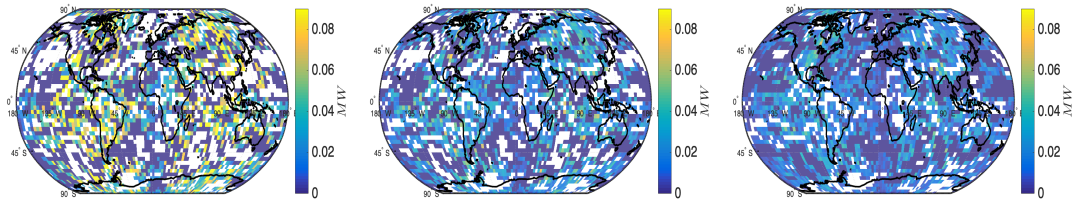


(j) WTG 10 (k) WTG 11 (l) WTG 12

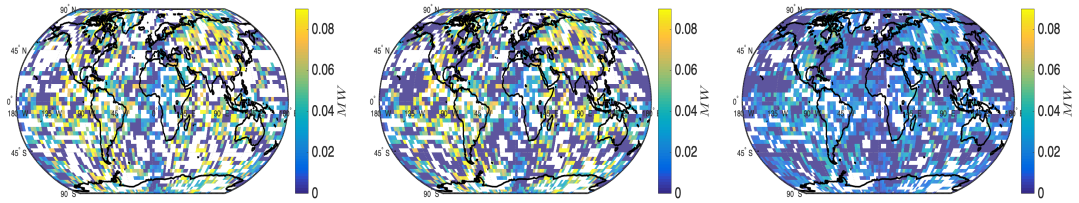
Figure 4.4: Mean of wind energy potentials for the 4D-Var-MC implementations. The number of ensemble members $N = 20$. White regions denote no wind-energy-potential generation.



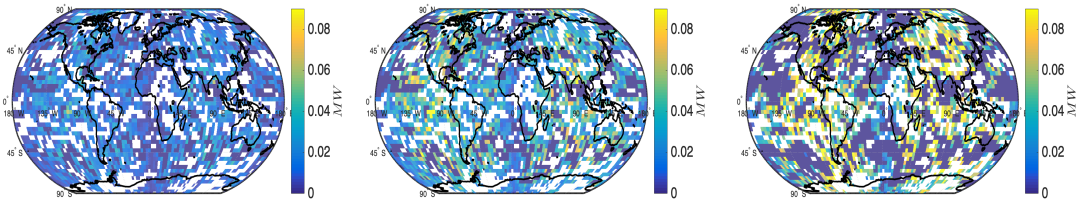
(a) WTG 1 (b) WTG 2 (c) WTG 3



(d) WTG 4 (e) WTG 5 (f) WTG 6



(g) WTG 7 (h) WTG 8 (i) WTG 9



(j) WTG 10 (k) WTG 11 (l) WTG 12

Figure 4.5: Mean of wind energy potentials for the 4D-Var-MC implementations. The number of ensemble members $N = 20$. White regions denote no wind-energy-potential generation.

4.4.2 Single observations across observation times

In this section, we briefly discuss the performance of our proposed 4D-EnKF-MC method by using a single observation test. We hold the same experimental settings as those in Section 4.4.1 and report the estimation errors in the initial conditions via $L-2$ norms (4.21). The single observation, across all observation times, is placed as is shown in figure 4.6.

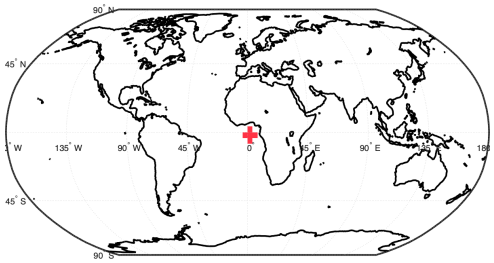
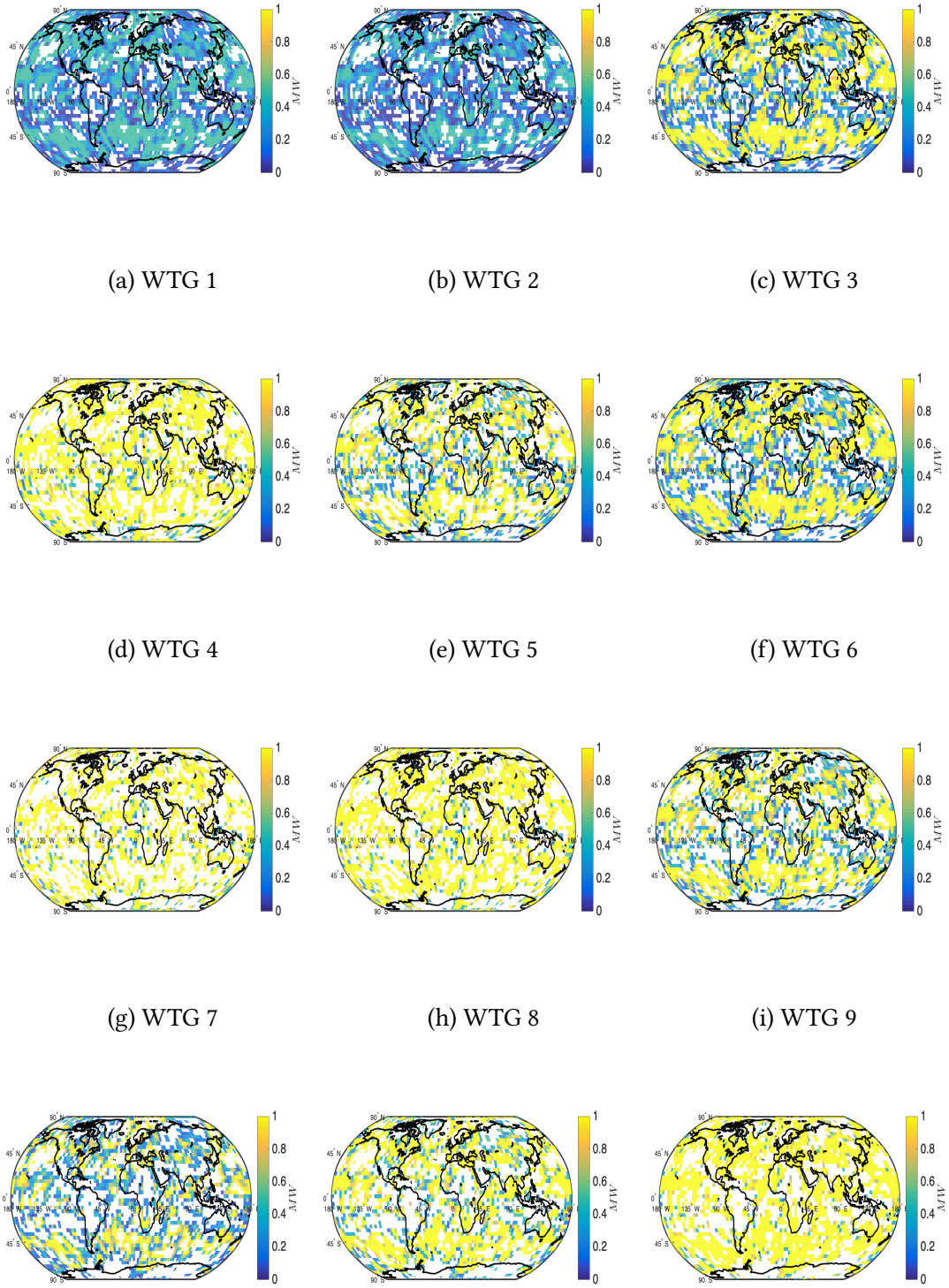


Figure 4.6: Observation operator across assimilation steps. A single observation (red cross) is placed during the experiments.

In the Table 4.4, we can clearly seen the advantages of employing the control space (4.11) instead of the traditional approach based on the ensemble sub-space (4.5). For instance, for WTG with low rate-capacity, error differences are of one order of magnitude. In figures 4.7 and 4.8, we report the potential energy estimation and the uncertainty of each component (as the standard deviation of errors from the initial members of the analysis ensemble). As can be seen, low-rate capacity WTGs such as the WTG 1 and the WTG2 provide estimates whose error dispersion is small. Again, as the rate-capacity increases, the spread of ensemble members grow. The accuracy of the proposed method obeys to the fact that the precision matrix is full-rank, well-conditioned, and even more localized. Thus, the impact of spurious correlations is mitigated in the analysis increments of the initial ensemble.

	Data Assimilation Method	
Wind Turbine Generator (WTG)	4D EnKf	4D EnKf-Cho
WTG 1	10.4452	8.5893
WTG 2	10.3149	8.4186
WTG 3	21.0570	17.3117
WTG 4	58.1186	49.0419
WTG 5	27.2986	22.9304
WTG 6	20.4388	16.4551
WTG 7	59.4296	50.3057
WTG 8	55.3220	46.4524
WTG 9	23.5265	19.2920
WTG 10	20.2342	16.5172
WTG 11	42.4308	34.6328
WTG 12	48.3756	40.4795

Table 4.4: $L-2$ error norms of wind energy potential estimations at the initial analysis member. Two ensemble sizes are tried during the experiments.



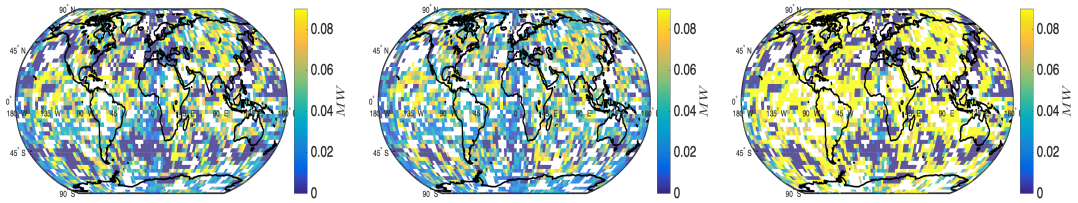
(a) WTG 1 (b) WTG 2 (c) WTG 3

(d) WTG 4 (e) WTG 5 (f) WTG 6

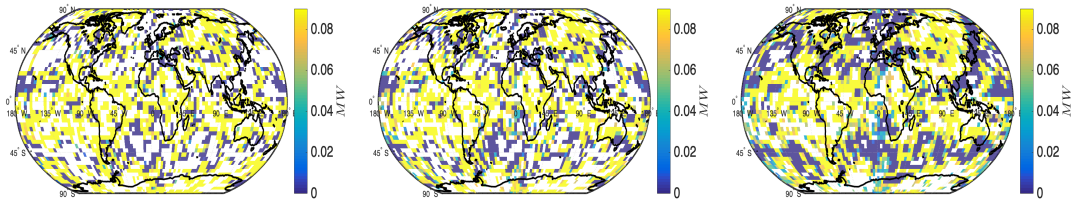
(g) WTG 7 (h) WTG 8 (i) WTG 9

(j) WTG 10 (k) WTG 11 (l) WTG 12

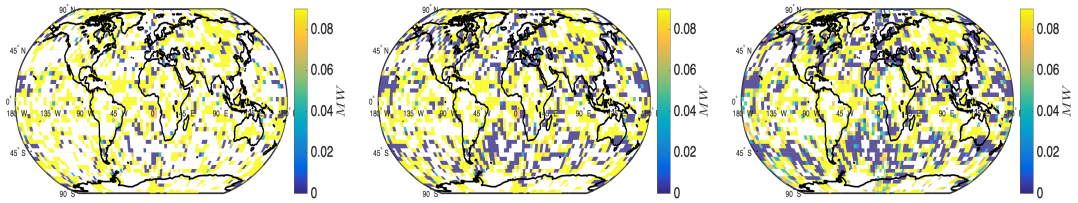
Figure 4.7: Mean of wind energy potentials for the 4D-Var-MC implementations. The number of ensemble members $N = 20$. White regions denote no wind-energy-potential generation. The number of observations reads 1.



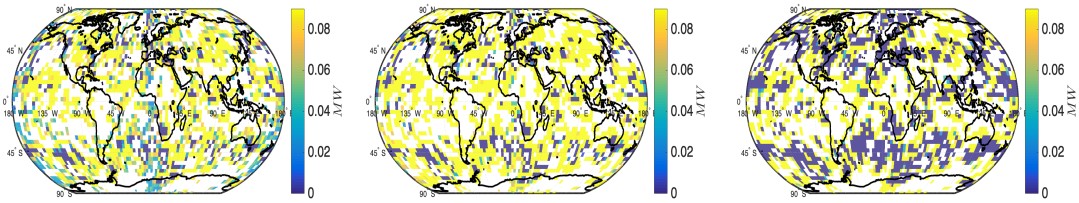
(a) WTG 1 (b) WTG 2 (c) WTG 3



(d) WTG 4 (e) WTG 5 (f) WTG 6



(g) WTG 7 (h) WTG 8 (i) WTG 9



(j) WTG 10 (k) WTG 11 (l) WTG 12

Figure 4.8: Mean of wind energy potentials for the 4D-Var-MC implementations. The number of ensemble members $N = 20$. White regions denote no wind-energy-potential generation. The number of observations reads 1.

4.5 Conclusions

We propose a 4D-Var ensemble-based data assimilation framework for wind energy potential estimation. In this formulation, in the 4D-Var context, the intrinsic need of adjoint models is avoided via the use of an ensemble of model realizations. These ensembles are employed to build control spaces onto which analysis increments are estimated. Control spaces are built via a modified Cholesky decomposition. The particular structure of this estimator allows for a matrix-free implementation of the proposed filter formulation. Experimental tests are performed, making use of wind turbines catalogs and the Atmospheric General Circulation Model Speedy. The results reveal that our proposed framework can properly estimate wind energy potential capacities within reasonable accuracies in terms of Root-Mean-Square-Error, and even more, these estimations are better than those of traditional 4D-Var ensemble-based methods. Besides, Wind Turbine Generators (WTGs) with low rate-capacity are the ones which provide homogeneous behavior of error estimations around the globe. As the rate-capacity increases, the potential energy increases as well, but the error dispersion of ensemble members grow, which can difficult decision-making processes. Of course, rate-capacity is just a single parameter of many in the WTG context, and we do not consider, for instance, economic aspects in our study, which can be crucial for deciding whether or not to employ green sources of energy.

5 Conclusions

To be concise, the contributions of the current work are detailed next:

1. We propose a posterior ensemble Kalman filter based on a modified Cholesky decomposition. The proposed method estimates the posterior moments of the error distribution based on an ensemble of model realizations. An estimate of the inverse background error covariance matrix via a modified Cholesky decomposition is updated making use of rank-one matrices with information brought by the data error correlations in order to estimate the precision covariance matrix. This matrix is utilized in order to compute the posterior mode of the error distribution and then, samples are taken about it. This implementation is matrix-free making it attractive for practical implementations. Experimental settings are performed making use of the Lorenz 96 model and different observations and ensemble configurations. The results obtained by the proposed method are compared against those obtained by the local ensemble transform Kalman filter (LETKF). The results reveal that, the use of the proposed implementation can mitigate the impact of sampling errors and even more, the accuracy of the proposed EnKF implementation is similar to that of the LETKF.

2. This work proposes a posterior matrix-free ensemble Kalman filter based on a modified Cholesky decomposition which works as follows: the precision background error covariance matrix is estimated in terms of Cholesky factors via a modified Cholesky decomposition, the factors of the posterior precision covariance are then obtained by rank-one updates over the background factors and from there, the posterior mode of the error distribution can be estimated. By using the analysis factors, the posterior ensemble can be built by either sampling from the posterior distribution or making use of synthetic data. Besides, sparse estimators of the precision analysis covariance can be obtained by exploiting the conditional independence of model components regarding some radius of influence as is done in the EnKF-MC context. The computational effort of the proposed method is similar to that of the EnKF-MC. Experimental tests are performed in order to assess the accuracy of the proposed method as the inflation factor and the radius of influences are varied. The numerical model operator utilized during the experiments is the Lorenz-96 model. The results reveal that, the accuracy in terms of root-mean-square-error of the proposed method is similar to that of one of the best EnKF implementations from the current literature. Besides, the results obtained by the proposed implementation are comparable to those of the EnKF with large ensemble sizes.

3. We propose a 4D-Var ensemble-based data assimilation framework for wind energy potential estimation. In this formulation, in the 4D-Var context, the intrinsic need of adjoint models is avoided via the use of an ensemble of model realizations. These ensembles are employed to build control spaces onto which analysis increments are

estimated. Control spaces are built via a modified Cholesky decomposition. The particular structure of this estimator allows for a matrix-free implementation of the proposed filter formulation. Experimental tests are performed, making use of wind turbines catalogs and the Atmospheric General Circulation Model Speedy. The results reveal that our proposed framework can properly estimate wind energy potential capacities within reasonable accuracies in terms of Root-Mean-Square-Error, and even more, these estimations are better than those of traditional 4D-Var ensemble-based methods. Besides, Wind Turbine Generators (WTGs) with low rate-capacity are the ones which provide homogeneous behavior of error estimations around the globe. As the rate-capacity increases, the potential energy increases as well, but the error dispersion of ensemble members grow, which can difficult decision-making processes. Of course, rate-capacity is just a single parameter of many in the WTG context, and we do not consider, for instance, economic aspects in our study, which can be crucial for deciding whether or not to employ green sources of energy.

6 Bibliography

[AA99] Jeffrey L. Anderson and Stephen L. Anderson. A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999.

[AKW11] Javier Amezcua, Eugenia Kalnay, and Paul D Williams. The effects of the raw filter on the climatology and forecast skill of the speedy model. *Monthly Weather Review*, 139(2):608–619, 2011.

[And01] Jeffrey L Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.

[And07] Jeffrey L Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007.

[And12] Jeffrey L. Anderson. Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Monthly Weather Review*, 140(7):2359–2371, 2012.

[And19] Jeffrey L Anderson. A nonlinear rank regression method for ensemble kalman filter data assimilation. *Monthly Weather Review*, 147(8):2847–2860, 2019.

[Ban16] RN Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 2016.

[BHC+10a] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. Intercomparison of variational data assimilation and the ensemble kalman filter for global deterministic nwp. part i: Description and single-observation experiments. *Monthly Weather Review*, 138(5):1550–1566, 2010.

[BHC+10b] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. Intercomparison of variational data assimilation and the ensemble kalman filter for global deterministic nwp. part ii: One-month experiments with real observations. *Monthly Weather Review*, 138(5):1567–1586, 2010.

[BKKM04] Annalisa Bracco, Fred Kucharski, Rameshan Kallummal, and Franco Molteni. Internal variability, external forcing and climate trends in multi-decadal agcm ensembles. *Climate Dynamics*, 23(6):659–678, 2004.

[BL08a] Peter J Bickel and Elizaveta Levina. Covariance regularization by thresholding. *The Annals of Statistics*, pages 2577–2604, 2008.

[BL08b] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 02 2008.

[BL+08c] Peter J Bickel, Elizaveta Levina, et al. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008.

[BP14] David R. Bickel and Marta Padilla. A Prior-free Framework of Coherent Inference and Its Derivation of Simple Shrinkage Estimators. *Journal of Statistical Planning and Inference*, 145(0):204–221, 2014.

[BT99] Craig H. Bishop and Zoltan Toth. Ensemble Transformation and Adaptive Observations. *Journal of the Atmospheric Sciences*, 56(11):1748–1765, 1999.

[Bue05] Mark Buehner. Ensemble-derived Stationary and Flow-dependent Background-error Covariances: Evaluation in a Quasi-operational NWP Setting. *Quarterly Journal of the Royal Meteorological Society*, 131(607):1013–1043, 2005.

[Bue11] Mark Buehner. Evaluation of a Spatial/Spectral Covariance Localization Approach for Atmospheric Data Assimilation. *Monthly Weather Review*, 140(2):617–636, 2011.

[CEK+13] Abhishek Chatterjee, Richard J. Engelen, Stephan R. Kawa, Colm Sweeney, and Anna M. Michalak. Background Error Covariance Estimation for Atmospheric CO2 Data Assimilation. *Journal of Geophysical Research: Atmospheres*, 118(17):10,140–10,154, 2013.

[CJAS10] Haiyan Cheng, Mohamed Jardak, Mihai Alexe, and Adrian Sandu. A Hybrid Approach to Estimating Error Covariances in Variational Data Assimilation. *Tellus A*, 62(3):288–297, March 2010.

[CO10] Yan Chen and Dean S Oliver. Cross-covariances and localization for enkf in multiphase flow data assimilation. *Computational Geosciences*, 14(4):579–601, 2010.

[CSS05] A Caya, J Sun, and C Snyder. A comparison between the 4dvar and the ensemble kalman filter techniques for radar data assimilation. *Monthly Weather Review*, 133(11):3081–3094, 2005.

[CTW+94] Thomas J Conway, Pieter P Tans, Lee S Waterman, Kirk W Thoning, Duane R Kitzis, Kenneth A Masarie, and Ni Zhang. Evidence for interannual variability of the carbon cycle from the national oceanic and atmospheric administration/climate monitoring and diagnostics laboratory global air sampling network. *Journal of Geophysical Research: Atmospheres*, 99(D11):22831–22855, 1994.

[DDO13] Jack J Dongarra, James W Demmel, and Susan Ostrouchov. Lapack: A linear algebra library for high-performance computers. In *Computational Statistics: Volume 1: Proceedings of the 10th Symposium on Computational Statistics*, page 23. Springer Science & Business Media, 2013.

[DRB08] Pablo Del Río and Mercedes Burguillo. Assessing the impact of renewable energy deployment on local sustainability: Towards a theoretical framework. *Renewable and sustainable energy reviews*, 12(5):1325–1344, 2008.

[Eve03] Geir Evensen. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics*, 53(4):343–367, 2003.

[Eve06] Geir Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[Eve09] Geir Evensen. The ensemble kalman filter for combined state and parameter estimation. *IEEE Control Systems*, 29(3), 2009.

[EWH13] Ahmed H. Elsheikh, Mary F. Wheeler, and Ibrahim Hoteit. An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *Journal of Computational Physics*, 242:696 – 714, 2013.

[FAF+17] Marianne Fay, Luis Alberto Andres, Charles Fox, Ulf Narloch, and Michael Slawson. *Rethinking Infrastructure in Latin America and the Caribbean: Spending Better to Achieve More*. World Bank Publications, 2017.

[FHH07] Elana J Fertig, John Harlim, and Brian R Hunt. A comparative study of 4d-var and a 4d ensemble kalman filter: Perfect model simulations with lorenz-96. *Tellus A*, 59(1):96–100, 2007.

[FVE04] Ibrahim Fatkullin and Eric Vanden-Eijnden. A computational strategy for multiscale systems with applications to lorenz 96 model. *Journal of Computational Physics*, 200(2):605–638, 2004.

[GAVL15] Michael Goodliff, Javier Amezcua, and Peter Jan Van Leeuwen. Comparing hybrid data assimilation methods on the lorenz 1963 model with increasing non-linearity. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):26928, 2015.

[GB14] Nils Gustafsson and J Bojarova. Four-dimensional ensemble variational (4d-en-var) data assimilation for the high resolution limited area model (hirlam). *Nonlinear Processes in Geophysics*, 21(4):745–762, 2014.

[GJSAGJ17] Stephany Griffith-Jones, Stephen Spratt, Rodrigo Andrade, and Edward Griffith-Jones. Investment in renewable energy, fossil fuel prices and policy implications for latin america and the caribbean. 2017.

[GKM+11] Steven J Greybush, Eugenia Kalnay, Takemasa Miyoshi, Kayo Ide, and Brian R Hunt. Balance and ensemble kalman filter localization techniques. *Monthly Weather Review*, 139(2):511–522, 2011.

[GM05] Georg A Gottwald and Ian Melbourne. Testing for chaos in deterministic systems with noise. *Physica D: Nonlinear Phenomena*, 212(1):100–110, 2005.

[GMC+06] S. Gillijns, O.B. Mendoza, J. Chandrasekar, B. L R De Moor, D.S. Bernstein, and A Ridley. What is the Ensemble Kalman Filter and How Well Does It Work? In *American Control Conference, 2006*, pages 6 pp.–, June 2006.

[Gus07] Nils Gustafsson. Discussion on '4d-var or enkf?'. *Tellus A: Dynamic Meteorology and Oceanography*, 59(5):774–777, 2007.

[HKS07] Brian R Hunt, Eric J Kostelich, and Istvan Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, 230(1):112–126, 2007.

[HM98] Peter L Houtekamer and Herschel L Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[HWS01] Thomas M Hamill, Jeffrey S Whitaker, and Chris Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129(11):2776–2790, 2001.

[HZS18] Yunfeng Han, Jucheng Zhang, and Dajun Sun. Error control and adjustment method for underwater wireless sensor network localization. *Applied Acoustics*, 130:293–299, 2018.

[JFW14] Poterjoy Jonathan, Zhang Fuqing, and Yonghui Weng. The Effects of Sampling Errors on the EnKF Assimilation of Inner-Core Hurricane Observations. *Monthly Weather Review*, 142(4):1609–1630, 2014.

[Kep00] Christian L. Keppenne. Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. *Monthly Weather Review*, 128(6):1971–1981, 2000.

[KMB06] Fred Kucharski, Franco Molteni, and Annalisa Bracco. Decadal interactions between the western tropical pacific and the north atlantic oscillation. *Climate dynamics*, 26(1):79–91, 2006.

[KP10] A Karimi and Mark R Paul. Extensive chaos in the lorenz-96 model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):043105, 2010.

[KST17] Jakob Kopiske, Sebastian Spieker, and George Tsatsaronis. Value of power plant flexibility in power systems with high shares of variable renewables: A scenario outlook for germany 2035. *Energy*, 137:823–833, 2017.

[Ler07] P. F. J. Lermusiaux. Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D Nonlinear Phenomena*, 230:172–196, June 2007.

[Liu17] Zhenling Liu. China's strategy for the development of renewable energies. *Energy Sources, Part B: Economics, Planning, and Policy*, 12(11):971–975, 2017.

[LKM09] Hong Li, Eugenia Kalnay, and Takemasa Miyoshi. Simultaneous estimation of covariance inflation and observation errors within an ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(639):523–533, 2009.

[LMQ16] Yoonsang Lee, Andrew J Majda, and Di Qi. Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence. *Monthly Weather Review*, (2016), 2016.

[Lor86] A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[Lor03a] Andrew C Lorenc. Modelling of error covariances by 4d-var data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 129(595):3167–3182, 2003.

[Lor03b] Andrew C Lorenc. The potential of the ensemble kalman filter for nwp—a comparison with 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003.

[Lor05] Edward N. Lorenz. Designing chaotic models. *Journal of the Atmospheric Sciences*, 62(5):1574–1587, 2005.

[LR99] P. F. J. Lermusiaux and A. R. Robinson. Data assimilation via error subspace statistical estimation. part i: Theory and schemes, 1999. Accessed: 08-29-2015.

[LRZ+08] Elizaveta Levina, Adam Rothman, Ji Zhu, et al. Sparse estimation of large covariance matrices via a nested lasso penalty. *The Annals of Applied Statistics*, 2(1):245–263, 2008.

[LT16] Daniel Langr and Pavel Tvrdik. Evaluation criteria for sparse matrix storage formats. *IEEE Transactions on parallel and distributed systems*, 27(2):428–440, 2016.

[LW17] Lili Lei and Jeffrey S Whitaker. Evaluating the tradeoffs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *Journal of Advances in Modeling Earth Systems*, 2017.

[LWB18] Lili Lei, Jeffrey S Whitaker, and Craig Bishop. Improving assimilation of radiance observations by implementing model space localization in an ensemble kalman filter. *Journal of Advances in Modeling Earth Systems*, 10(12):3221–3232, 2018.

[Mas13] Gilbert M Masters. *Renewable and efficient electric power systems*. John Wiley & Sons, 2013.

[MB06] Jan Mandel and Jonathan D Beezley. *Predictor-corrector and morphing ensemble filters for the assimilation of sparse data into high-dimensional nonlinear systems*. University of Colorado at Denver and Health Sciences Center, Center for Computational Mathematics, 2006.

[MCD+01] J Michalakes, S Chen, J Dudhia, L Hart, J Klemp, J Middlecoff, and W Skamarock. Development of a next-generation regional weather research and forecast model. In *Developments in Teracomputing*, pages 269–276. World Scientific, 2001.

[MDG+05] J Michalakes, J Dudhia, D Gill, T Henderson, J Klemp, W Skamarock, and W Wang. The weather research and forecast model: software architecture and performance. In *Use of high performance computing in meteorology*, pages 156–168. World Scientific, 2005.

[Miy11] Takemasa Miyoshi. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter. *Monthly Weather Review*, 139(5):1519–1535, 2011.

[MKI14] Takemasa Miyoshi, Keiichi Kondo, and Toshiyuki Imamura. The 10,240-member ensemble kalman filtering with an intermediate agcm. *Geophysical Research Letters*, 41(14):5264–5271, 2014.

[Mol03] Franco Molteni. Atmospheric simulations using a gcm with simplified physical parametrizations. i: Model climatology and variability in multi-decadal experiments. *Climate dynamics*, 20(2):175–191, 2003.

[MWOH17] Francisco David Munoz, Sonja Wogrin, Shmuel S Oren, and Benjamin F Hobbs. Economic inefficiencies of cost-based electricity market designs. In *Heading Towards Sustainable Energy Systems: Evolution or Revolution?, 15th IAEE European Conference, Sept 3-6, 2017*. International Association for Energy Economics, 2017.

[NR17] Elias Nino-Ruiz. A matrix-free posterior ensemble kalman filter implementation based on a modified cholesky decomposition. *Atmosphere*, 8(7):125, 2017.

[NR18] Elias D Nino-Ruiz. Implicit surrogate models for trust region based methods. *Journal of Computational Science*, 2018.

[NRCB18] Elias D Nino-Ruiz, Haiyan Cheng, and Rolando Beltran. A robust non-gaussian data assimilation method for highly non-linear models. *Atmosphere*, 9(4):126, 2018.

[NRMC17] 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.

[NRMR18] Elias D Nino-Ruiz and Luis E Morales-Retat. A tabu search implementation for adaptive localization in ensemble-based methods. *Soft Computing*, pages 1–17, 2018.

[NRS15] Elias D Nino-Ruiz and Adrian Sandu. Ensemble kalman filter implementations based on shrinkage covariance matrix estimation. *Ocean Dynamics*, 65(11):1423–1439, 2015.

[NRS16] 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.

[NRSA14] EliasD. Nino Ruiz, Adrian Sandu, and Jeffrey Anderson. An Efficient Implementation of the Ensemble Kalman Filter Based on an Iterative Sherman–Morrison Formula. *Statistics and Computing*, pages 1–17, 2014.

[NRSD15] 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.

[NRSD17] 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.

[NRSD18] Elias D Nino-Ruiz, Adrian Sandu, and Xinwei Deng. An ensemble kalman filter implementation based on modified cholesky decomposition for inverse covariance matrix estimation. *SIAM Journal on Scientific Computing*, 40(2):A867–A886, 2018.

[NSD16] Elias D Nino, Adrian Sandu, and Xinwei Deng. An ensemble kalman filter implementation based on modified cholesky decomposition for inverse covariance matrix estimation. *arXiv preprint arXiv:1605.08875*, 2016.

[OHS+04] Edward Ott, Brian R. Hunt, Istvan Szunyogh, Aleksey V. Zimin, Eric J. Kostelich, Matteo Corazza, Eugenia Kalnay, D. J. Patil, and James A. Yorke. A local ensemble kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004.

[OHS+08] Edward Ott, Brian Hunt, Istvan Szunyogh, Aleksey V Zimin, Eic J. Kostelich, Matteo Corazza, Eugenia Kalnay, D. J. Patil, and James A. Yorke. A Local Ensemble Transform Kalman Filter Data Assimilation System for the NCEP Global Model. *Tellus A*, 60(1):113–130, 2008.

[PXJ+17] Bryan J Putnam, Ming Xue, Youngsun Jung, Nathan A Snook, and Guifu Zhang. Ensemble probabilistic prediction of a mesoscale convective system and associated polarimetric radar variables using single-moment and double-moment microphysics schemes and enkf radar data assimilation. *Monthly Weather Review*, (2017), 2017.

[RSA15] Elias D Nino Ruiz, Adrian Sandu, and Jeffrey Anderson. An efficient implementation of the ensemble kalman filter based on an iterative sherman–morrison formula. *Statistics and Computing*, 25(3):561–577, 2015.

[SB11a] Pavel Sakov and Laurent Bertino. Relation between two common localisation methods for the enkf. *Computational Geosciences*, 15(2):225–237, 2011.

[SB11b] Pavel Sakov and Laurent Bertino. Relation between two common localisation methods for the enkf. *Computational Geosciences*, 15(2):225–237, 2011.

[SEB10] Pavel Sakov, Geir Evensen, and Laurent Bertino. Asynchronous data assimilation with the enkf. *Tellus A*, 62(1):24–29, 2010.

[SGV05] FS Smailbegovic, Georgi N Gaydadjiev, and Stamatis Vassiliadis. Sparse matrix storage format. In *Proceedings of the 16th annual workshop on circuits, systems and signal processing*, pages 445–448, 2005.

[SKW18] Jonathan R Stroud, Matthias Katzfuss, and Christopher K Wikle. A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. *Monthly Weather Review*, 146(1):373–386, 2018.

[SUL+09] M Stengel, Per Undén, Magnus Lindskog, Per Dahlgren, Nils Gustafsson, and R Bennartz. Assimilation of seviri infrared radiances with hirlam 4d-var. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 135(645):2100–2109, 2009.

[TAB+03] Michael K. Tippett, Jeffrey L. Anderson, Craig H. Bishop, Thomas M. Hamill, and Jeffrey S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490, Jul 2003.

[TNC19] The Nature Conservancy en Colombia, 2019.

[VDVDH17] RA Verzijlbergh, LJ De Vries, GPJ Dijkema, and PM Herder. Institutional challenges caused by the integration of renewable energy sources in the european electricity sector. *Renewable and Sustainable Energy Reviews*, 75:660–667, 2017.

[VIR+14] Walter Vergara, Paul Isbell, Ana R Rios, José Ramon Gómez, and Leandro Alves. Societal benefits from renewable energy in latin america and the caribbean. Technical report, Inter-American Development Bank, 2014.

[W+08] Ryam Wiser et al. Annual report on us wind power installation, cost, and performance trends: 2007. Technical report, EERE Publication and Product Library, 2008.

[Wes16] Philip M Westgate. A covariance correction that accounts for correlation estimation to improve finite-sample inference with generalized estimating equations: a study on its applicability with structured correlation matrices. *Journal of statistical computation and simulation*, 86(10):1891–1900, 2016.

[XB10] Kaigui Xie and Roy Billinton. Determination of the optimum capacity and type of wind turbine generators in a power system considering reliability and cost. *IEEE Transactions on Energy Conversion*, 26(1):227–234, 2010.

[Zup09] Milija Zupanski. Theoretical and Practical Issues of Ensemble Data Assimilation in Weather and Climate. In SeonK. Park and Liang Xu, editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, pages 67–84. Springer Berlin Heidelberg, 2009.