Research Article

From the Kalman Filter to the Particle Filter: A Geometrical Perspective of the Curse of Dimensionality

Olivier Pannekoucke,[1,2] Pierrick Cébron,[1] Niels Oger,[1] and Philippe Arbogast[1]

[1]*CNRM, UMR 3589, 42 Av. Coriolis, 31057 Toulouse, France*
[2]*INPT-ENM, 42 Av. Coriolis, 31057 Toulouse, France*

Correspondence should be addressed to Olivier Pannekoucke; olivier.pannekoucke@meteo.fr

The aim of this contribution is to provide a description of the difference between Kalman filter and particle filter when the state space is of high dimension. In the Gaussian framework, KF and PF give the same theoretical result. However, in high dimension and using finite sampling for the Gaussian distribution, the PF is not able to reproduce the solution produced by the KF. This discrepancy is highlighted from the convergence property of the Gaussian law toward a hypersphere: in high dimension, any finite sample of a Gaussian law lies within a hypersphere centered in the mean of the Gaussian law and of radius square-root of the trace of the covariance matrix. This concentration of probability suggests the use of norm as a criterium that discriminates whether a forecast sample can be compatible or not with a given analysis state. The contribution illustrates important characteristics that have to be considered for the high dimension but does not introduce a new approach to face the curse of dimensionality.

1. Introduction

From the work of Bengtsson et al. [1] and Snyder et al. [2], the particle filter is known to suffer from the "curse of dimensionality" [3, 4], in the sense that when the problem is in high dimension, the size of the particle ensemble should be exponentially large leading to the failure of direct Monte-Carlo strategies.

More precisely, in the particle filter, a weight is computed for each particle from the observational error distribution. This weight measures the proximity of the particle to a given observations set. If the ensemble size does not follow an exponential increase with the dimension, then the maximum weight is systematically close to 1, meaning that only one particle is compatible with the observations (or at least the less far from the observation considering the other particles): this leads to the particle degeneracy [5]. While the observation error distribution is often assumed Gaussian in data assimilation, a change of observation error distribution by a larger tail distribution may be introduced to limit the particle degeneracy in realistic application [6]. But this replacement should not be used for theoretical study where it would

introduce incoherence between the way the observations are supposed to be and the way they are assimilated. Moreover, Snyder et al. [2] have shown that the particle degeneracy still occurs with larger tail distributions, for example, for the Cauchy distribution.

In the demonstration proposed by Snyder et al. [2], to study the particle filter, the curse of dimensionality is described in terms of weight, related to the observational space. Hence, it is still difficult to understand what happens in the state space, and also what is the real difference between the ensemble version of the Kalman equations (EnKF), introduced by Evensen [7], and the particle filter approach (PF), introduced by Gordon et al. [8]. This question is not to know whether the EnKF converges toward the PF in the general framework, where known results are existing for this issue: (a) the PF is known to converge toward the nonlinear filter in the limit of large number of particles in interaction, Del Moral [9]; (b) the EnKF is known to converge toward a different solution than the nonlinear filter [10, 11]. How can we feel the paradox that the two algorithms, the EnKF and the PF under Gaussian assumption, should deliver the same conclusion while, in the practical high dimension case, they do not?

In the present contribution, important characteristics are illustrated that have to be considered for the high dimension, but no new approach is introduced to face the curse of dimensionality. This work investigates another point of view that relies on the multidimensional spheres or hyperspheres. A similar but different approach has been considered by Chorin and Morzfeld [12]. This perspective, mentioned in the conclusion of Snyder et al. [2], helps to understand in a direct way the effect of the dimension on the difference between the Kalman filter and the particle approximation of the Bayes rule under Gaussian assumption. The main result used in this work is that a normalized Gaussian law in high dimension is not a simple extension of higher dimension of the usual scatter plot in 2d but converges toward the uniform law on a sphere in n dimension, a phenomenon known as the Poincaré Lemma [13]. This phenomenon occurs even from the "high dimension" 100 that is very small compared with the $\mathcal{O}(10^7)$ degrees of freedom encountered in geophysical applications: the constraints due to the dimension are quite strong for this area.

The understanding of the behaviour in high dimension is required when using alternative to classical data assimilation to update probabilistic information contained within an ensemble. Such an update can be motivated in order to increase the ensemble size by merging lagged ensemble [14, 15] with the question of how to merge ensemble from different forecast times [16]. This can also be due to the real time constraints: ensemble prediction system can now rely on an analysis ensemble [17], but in practice there can be an important delay between the operational analysis production (that relies on Kalman filter formalism) and the ensemble analysis production. As a result an analysis state exists that could be used to update the last forecast ensemble available, for example, by using the Particle filter approach [18, 19] where an adaptation of the metric in the computation of the weight is also introduced in the application of the particle filter strategy. This kind of situation is a motivation to clarify the differences between Kalman filter and particle filter, at least under Gaussian assumption.

To tackle this issue we first recall the behaviour of Gaussian distribution in high dimension in Section 2. In Section 3, we review the description of the filtering theory under Gaussian assumption, then detail the asymptotic consequences of the high dimension, and provide the constraints implied onto the background and the analysis distribution that constitutes the difference between the EnKF and the PF. The conclusions are reported in Section 4.

2. Behaviour of Gaussian Distribution in High Dimension

2.1. Concentration of Gaussian Law in High Dimension.
Human intuition of the high dimension behaviour is often an extrapolation of the low dimension experiences. This leads us to assume that graphical representations in low dimensions are also verified in high dimensions. This is particularly true for Gaussian distribution where the similitude of scatter plots in dimensions lower than 3 invites us to speculate that the distribution in high dimensions should not be so

different (see Figure 1). But actually this is completely wrong and leads to a misleading interpretation of what a Gaussian distribution is really, despite its intensive use in large systems as encountered in geophysical sciences and especially in data assimilation.

If $\varepsilon^b \in \mathbb{R}^n$ is a random Gaussian vector with zero mean and covariance matrix $\mathbf{B}$ (in data assimilation this ε^b is the statistical model for the background error), denoted by $\varepsilon^b \sim \mathcal{N}(0, \mathbf{B})$, then the asymptotic distribution of the norm $|\varepsilon^b|$ is given by (see Appendix A)

$$\left|\varepsilon^b\right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\operatorname{tr}(\mathbf{B})}, \frac{1}{2}\frac{\operatorname{tr}(\mathbf{B})}{d_{\mathrm{ef}}(\mathbf{B})}\right), \tag{1}$$

where $\operatorname{tr}(\cdot)$ denotes de trace operator and where

$$d_{\mathrm{ef}}(\boldsymbol{\Gamma}) = \frac{\operatorname{tr}(\boldsymbol{\Gamma})^2}{\operatorname{tr}(\boldsymbol{\Gamma}^2)} \tag{2}$$

is the effective dimension that provides a quantitative measure of the dimensionality [20].

This simple formula implies a deep contradiction between the natural intuition and the reality of Gaussian laws. But at the same time it provides the correct intuition as now detailed.

2.2. Asymptotic Convergence of Normal Law toward a Spherical Shell.
For the particular case where $\varepsilon^b \sim \mathcal{N}(0, \mathbf{I})$ is a normal law (reduced and centered Gaussian law), the effective dimension is $d_{\mathrm{ef}}(\mathbf{I}) = n$ and then

$$\left|\varepsilon^b\right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{n}, \frac{1}{2}\right). \tag{3}$$

This distribution means that, in high dimension, all samples of ε^b are concentrated within a spherical shell of radius $\sqrt{n}$ and thickness $1/\sqrt{2} \approx 0.71$. Note that the thickness of the shell is independent of the dimension n for the norm $|\varepsilon^b|$ (while this is not true for $|\varepsilon^b|^2$).

As a simplified representation, Figure 2 mimics the behaviour of the Gaussian sample as the dimension n increases. For each panel, a sample $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)^T$ of the normal law $\mathcal{N}(0, \mathbf{I})$ is represented by the point $(x, y) = |\boldsymbol{\zeta}|(\cos\theta, \sin\theta)$, where $\theta = \arg(z)$ is the argument of the complex number $z = |z|e^{i\theta} = \zeta_1 + i\zeta_2$ formed from the first two components of $\boldsymbol{\zeta}$ (here $i^2 = -1$). For $n = 2$ this representation verifies $(x, y) = (\zeta_1, \zeta_2)$, that is, the usual dispersion of Gaussian sample (see Figure 2(a)). But, for larger values of n, the natural intuition fails to find the spherical shell illustrated for Figures 2(b)–2(d). The apparent decreasing, for increasing n, of the theoretically constant thickness $1/\sqrt{2}$ is due to the scaling used for the representation and corresponds to the relative error of the fluctuation over the radius: $(1/\sqrt{2})/\sqrt{n}$.

It is impressive to realise the absence of any sample within the spherical shell. The position of a sample seems binary: either it is at the right distance in the state space, or this is not really a sample of the Gaussian that was supposed to be sampled.
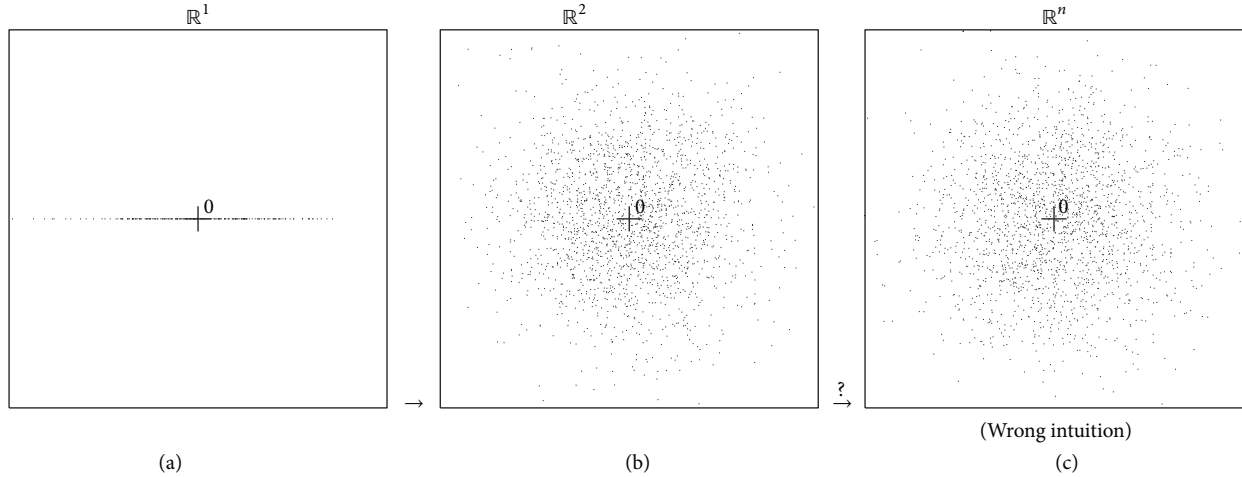
$\mathbb{R}^1$ (a) $\rightarrow$ $\mathbb{R}^2$ (b) $\xrightarrow{?}$ $\mathbb{R}^n$ (c) (Wrong intuition)

Figure 1: Wrong intuitive extrapolation of Gaussian samples distribution from dimension 1 (a) to dimension n (c).

This simplified illustration provides us with a graphical tool that supports our intuition of what are normal Gaussian laws in high dimensions. Now we develop what happens for general Gaussian law by taking into account the possible correlation between components of the random vector.

2.3. Asymptotic Convergence of Gaussian Law toward a Spherical Shell.
To illustrate the behaviour of Gaussian distribution $\varepsilon^b \sim \mathcal{N}(0, \mathbf{B})$ in high dimension one first needs to construct a nontrivial covariance matrix $\mathbf{B}$ and then to generate samples of ε^b.

In order to construct a correlation matrix $\mathbf{B}$, similar to the one encountered in data assimilation, we first consider an earth great circle of radius $a = 6400$ km, discretized with $n = 1000$ grid points $x_i = i \, dx$ with $dx = (1/n)2\pi a \approx 40$ km. Then, following the background covariance matrix modelling based on the diffusion equation [21, 22], which produces Gaussian correlation functions, a square-root matrix $\mathbf{B}^{1/2}$ is specified as such as $(\mathbf{B}^{1/2})_{ij} = \rho_{1/2}(x_i - x_j)$ with

$$\rho_{1/2}(x) = \left(\frac{2}{\pi}\right)^{1/4} \frac{\sqrt{dx}}{\sqrt{L_p}} e^{-x^2/L_p^2}, \tag{4}$$

where L_p is the correlation length-scale [23]. The resulting matrix $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$ is a covariance matrix where the diagonal terms representing the variance are all set to 1 provided that $\text{tr}(\mathbf{B}) = n$. This theoretical one-dimensional setting allows generating and computing all the quantities required from now.

As specified above, $\mathbf{B}$ is a correlation matrix and some correlation functions are represented, in Figure 3(a), for the varying length-scale values $L_p \in \{10, 100, 500, 1000\}$ km: the larger the length-scale value, the broader the correlation function. The effective dimension $d_{\text{ef}}(\mathbf{B}) = \text{tr}(\mathbf{B})^2/\text{tr}(\mathbf{B}^2)$ (Figure 3(b)) illustrates that the larger the length-scale value, the lower the number of significative principal direction of $\mathbf{B}$: the effective dimension is equal to 1000 (i.e., n) for $L_p =$

10 km, while it is below 100 for length-scale larger than $L_p = 500$ km.

In practice, samples of ε^b can be obtained as the transformation $\varepsilon^b = \mathbf{B}^{1/2}\zeta$ of samples of random vector following a normal law $\zeta \sim \mathcal{N}(0, \mathbf{I})$.

Similarly to the schematic representation of Figure 2, we illustrate the distribution of $N_e = 6400$ samples of $\varepsilon^b \sim \mathcal{N}(0, \mathbf{B})$ in Figure 4. For the short length-scale $L_p = 10$ km, the grid points are almost decorrelated and the samples are uniformly distributed within a spherical shell of radius $\sqrt{\text{tr}(\mathbf{B})} = \sqrt{n}$ and thickness $(1/\sqrt{2})(\sqrt{\text{tr}(\mathbf{B})}/\sqrt{d_{\text{ef}}(\mathbf{B})}) = 1/\sqrt{2}$, since $d_{\text{ef}}(\mathbf{B}) = n$ for this length-scale (see Figure 3). When the length-scale increases, the correlation between the first two components ε_1^b and ε_2^b of ε^b increases. For $L_p = 100$ km, the correlation is higher than 0.9 and the repartition of sampling points within the spherical shell becomes heterogeneous, with a concentration along the first principal diagonal. This heterogeneity is even more pronounced for $L_p = 500$ km, and it appears to be concentrated onto a subspace of an equatorial plane: this corresponds to the concentration of the sampling points within a spherical shell of the same radius $\sqrt{\text{tr}(\mathbf{B})} = \sqrt{n}$ but within a subspace of lower dimension (the effective dimension is the quantitative approximation of this dimension).

While the radius of the spherical shell remains at the fixed value $\sqrt{\text{tr}(\mathbf{B})}$, its thickness increases. This is the consequence of the effective dimension diminution as understood by the asymptotic variance formula $(1/\sqrt{2})(\sqrt{\text{tr}(\mathbf{B})}/\sqrt{d_{\text{ef}}(\mathbf{B})}) = (1/\sqrt{2})(\sqrt{n}/\sqrt{d_{\text{ef}}(\mathbf{B})})$: if $d_{\text{ef}}(\mathbf{B})$ decreases, then $1/d_{\text{ef}}(\mathbf{B})$ increases leading to an increase of the variance.

Hence, as the effective dimension decreases, the Gaussian distribution converges toward another Gaussian law of the same total variance but in a lower dimensional linear subspace. The graphical representation of the Gaussian law as a spherical shell is still verified.

Note that, in the continuous limit where the dimension tends to infinity (with vectors tending to functions), the above discussion remains true until the covariance matrix
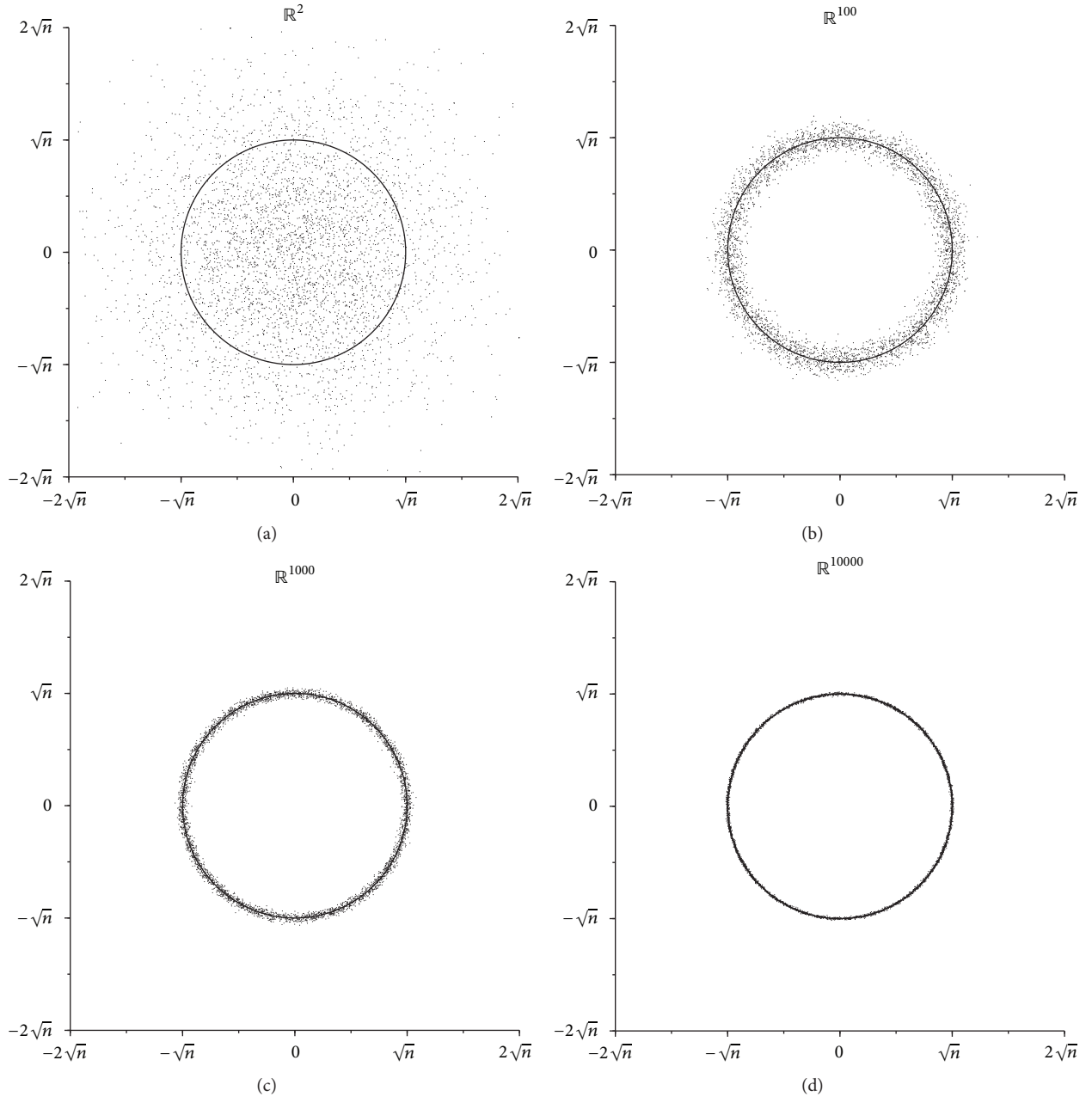
Figure 2: Schematic view of the concentration of $N_e = 6400$ samples of a normal law $\mathcal{N}(0, \mathbf{I})$ toward a spherical shell of radius $\sqrt{n}$ and thickness $1/\sqrt{2}$ as the dimension increases from $n = 2$ (a) to $n = 100$ (b), $n = 1000$ (c), and $n = 10000$ (d) (see text for the details of the construction).

converges toward a covariance operator of finite trace. In particular, the continuous limit of the reduced and normalized Gaussian vector, $\mathcal{N}(0, \mathbf{I})$, does not exist since it would imply an infinite trace operator (the radius $\sqrt{n}$ of its hypersphere increases to infinity). But the continuous limit of the correlated Gaussian vector $\mathcal{N}(0, \mathbf{B})$ exists (the trace of the limit operator is finite): for $L_p > 0$, the Fourier spectrum, which is also Gaussian, is summable. Figure 4(b) helps to interpret the finiteness of the covariance operator trace: all the pertinent information about the statistics is contained

within a subspace of finite dimension in the functional space.

From these results, the consequences of data assimilation are now described.

3. Consequences of the Concentration of the Measure in Data Assimilation

Data assimilation faces a paradox, recognized as resulting from the curse of dimensionality [2]: while the particle filter
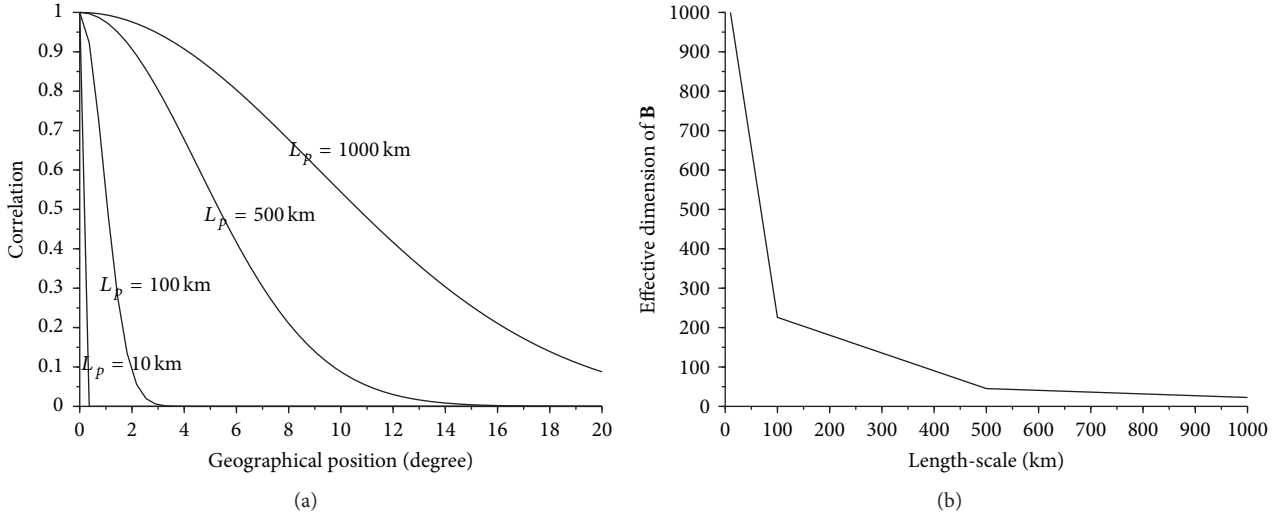
(a)

(b)

FIGURE 3: Correlation function (a) and effective dimension $d_{\text{ef}}(\mathbf{B}) = \text{tr}(\mathbf{B})^2/\text{tr}(\mathbf{B}^2)$ (b) for increasing length-scale parameter $L_p \in \{10, 100, 500, 1000\}$ km, for the dimension set to $n = 1000$.

and the ensemble Kalman filter are supposed to be two possible algorithms for the discretization of the nonlinear filtering under Gaussian assumption, the particle filter fails to produce the *a posteriori* distribution. This issue and the consequences of the concentration of Gaussian law are now explored.

3.1. Nonlinear Filtering. Data assimilation aims to provide the probability of the real state of a system knowing the past/present forecast and observations. The general framework formalism is, without any assumption of linearity of dynamics or Gaussianity of statistics, reduced to the Bayes rule

$$p\left(\mathcal{X}_q \mid \mathcal{Y}_q^o\right) \propto p\left(\mathcal{Y}_q^o \mid \mathcal{X}_q\right) p\left(\mathcal{X}_q\right), \qquad (5)$$

where $p(\mathcal{X}_q)$ is the (density of) probability distribution to find the true state $\mathcal{X}_q^t$ at time q in the vicinity of $\mathcal{X}_q$; $p(\mathcal{Y}_q^o \mid \mathcal{X}_q)$ is the (density of) probability distribution to measure $\mathcal{Y}_q^o$ at time q when the true state $\mathcal{X}_q^t$ is known. This corresponds to the analysis step (considering the Bayesian data assimilation setting). The vector space of $\mathcal{X}$ ($\mathcal{Y}^o$) is the state space (observational space), denoted by $\mathbb{R}^n$ ($\mathbb{R}^p$) where n (p) is the dimension.

The forecast step transports the conditional distribution from time q to time $q + 1$, by using the model propagator $\mathcal{X}_{q+1} = \mathcal{M}_{q+1 \leftarrow q}(\mathcal{X}_q)$. Since we are interested in comparing the respective analysis procedure proposed by the EnKF and by the PF, the forecast step is no more detailed.

3.2. Gaussian Assumption and Concentration Consequences. Under the usual additional Gaussianity assumption, the two distributions $p(\mathcal{X}_q)$ and $p(\mathcal{Y}_q^o \mid \mathcal{X}_q)$ take the form

$$p\left(\mathcal{X}_q\right) \propto \exp\left(-\frac{1}{2}\left\|\mathcal{X}_q - \mathcal{X}_q^b\right\|_{\mathbf{B}_q^{-1}}^2\right), \qquad (6)$$

where $\mathcal{X}_q^b$ is the background state at time q. $\varepsilon_q^b = \mathcal{X}_q^b - \mathcal{X}_q^t$ is the background error modeled as a Gaussian random vector of zero mean and covariance matrix $\mathbb{E}[\varepsilon_q^b \varepsilon_q^{bT}] = \mathbf{B}_q$ and with $\|\mathbf{u}\|_{\Gamma}^2 = \mathbf{u}^T \Gamma \mathbf{u}$; and

$$p\left(\mathcal{Y}_q^o \mid \mathcal{X}_q\right) \propto \exp\left(-\frac{1}{2}\left\|\mathcal{Y}_q^o - \mathbf{H}_q \mathcal{X}_q\right\|_{\mathbf{R}_q^{-1}}^2\right), \qquad (7)$$

where $\mathbf{H}_q$ is the observational operator that maps the state space into the observational space. In general, an observational operator is a nonlinear operator, but, here, it is assumed to be linear. $\varepsilon_q^o = \mathcal{Y}_q^o - \mathbf{H}_q \mathcal{X}_q^t$ is the observation error modeled as a random Gaussian vector of zero mean and covariance matrix $\mathbb{E}[\varepsilon_q^o \varepsilon_q^{oT}] = \mathbf{R}_q$. Note that $\mathbf{B}_q$ and $\mathbf{R}_q$ are random matrices that depend on the real state trajectory $\mathcal{X}_q^t$ and the observations assimilated. It results from the Gaussian assumption that $p(\mathcal{X}_q \mid \mathcal{Y}_q^o)$ is also Gaussian, written as

$$p\left(\mathcal{X}_q \mid \mathcal{Y}_q^o\right) \propto \exp\left(-\frac{1}{2}\left\|\mathcal{X}_q - \mathcal{X}_q^a\right\|_{\mathbf{A}_q^{-1}}^2\right), \qquad (8)$$

where $\mathcal{X}_q^a = \mathcal{X}_q^b + \mathbf{K}_q(\mathcal{Y}_q^o - \mathbf{H}_q \mathcal{X}_q^b)$ is the analysis state. $\varepsilon_q^a = \mathcal{X}_q^a - \mathcal{X}_q^t$ is the analysis error that is a Gaussian random vector of zero mean and covariance matrix $\mathbb{E}[\varepsilon_q^a \varepsilon_q^{aT}] = \mathbf{A}_q$ with

$$\mathbf{A}_q = \left(\mathbf{I} - \mathbf{K}_q \mathbf{H}_q\right) \mathbf{B}_q, \qquad (9)$$

where

$$\mathbf{K}_q = \mathbf{B}_q \mathbf{H}_q^T \left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q\right)^{-1} \qquad (10)$$
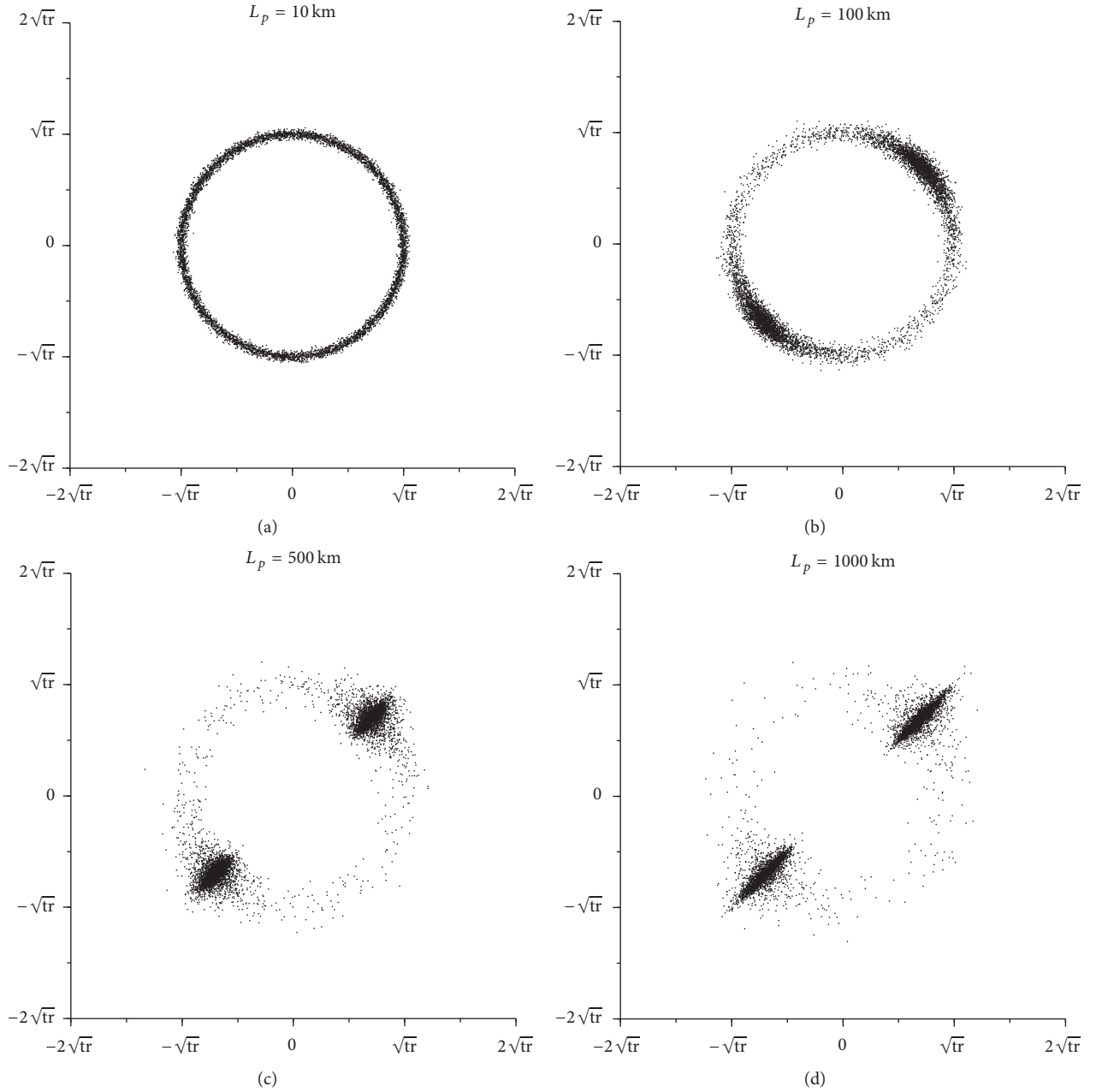
is the gain matrix.

FIGURE 4: Schematic view of the concentration of 6400 samples of a random vector $\varepsilon^b \sim \mathcal{N}(0, \mathbf{B})$ on a spherical shell (or a part of it) of constant radius $\sqrt{\mathrm{tr}(\mathbf{B})}(= \sqrt{n})$ and variable thickness $(1/\sqrt{2})(\sqrt{\mathrm{tr}(\mathbf{B})}/\sqrt{d_{\mathrm{ef}}(\mathbf{B})})(= (1/\sqrt{2})(\sqrt{n}/\sqrt{d_{\mathrm{ef}}(\mathbf{B})}))$ in function of the length-scale.

Direct consequences of the concentration of Gaussian law in high dimensions are that according to (1)

(a) $\quad \left| \mathcal{X}_q^b - \mathcal{X}_q^t \right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\mathrm{tr}\left(\mathbf{B}_q\right)}, \frac{1}{2} \frac{\mathrm{tr}\left(\mathbf{B}_q\right)}{d_{\mathrm{ef}}\left(\mathbf{B}_q\right)} \right),$

(b) $\quad \left| \mathcal{Y}_q^o - \mathbf{H}_q \mathcal{X}_q^t \right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\mathrm{tr}\left(\mathbf{R}_q\right)}, \frac{1}{2} \frac{\mathrm{tr}\left(\mathbf{R}_q\right)}{d_{\mathrm{ef}}\left(\mathbf{R}_q\right)} \right),$

(c) $\quad \left| \mathcal{X}_q^a - \mathcal{X}_q^t \right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\mathrm{tr}\left(\mathbf{A}_q\right)}, \frac{1}{2} \frac{\mathrm{tr}\left(\mathbf{A}_q\right)}{d_{\mathrm{ef}}\left(\mathbf{A}_q\right)} \right).$

$$(11)$$

Other consequences of the concentration property may be obtained, for example, for the observational variance tuning (see Appendix B).

Now we consider the discretization of the nonlinear filtering as it should be observed when Gaussian assumptions are verified.
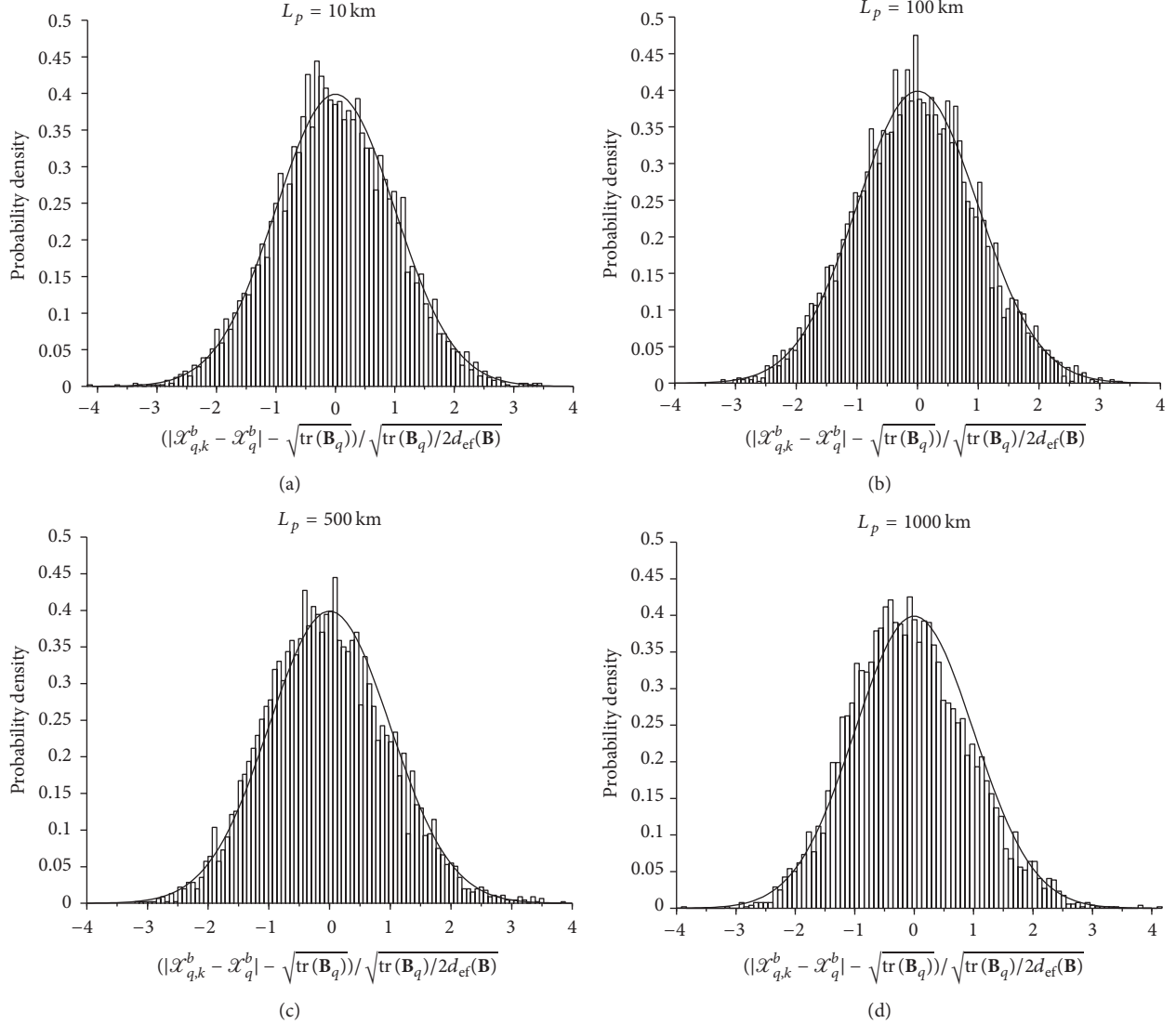
FIGURE 5: Histogram of distribution of the normalized distance $(|\mathcal{X}_{q,k}^b - \mathcal{X}_q^b| - \sqrt{\text{tr}(\mathbf{B}_q)})/\sqrt{\text{tr}(\mathbf{B}_q)/2d_{\text{ef}}(\mathbf{B})}$ normalized and compared with the theoretical asymptotic Gaussian law $\mathcal{N}(0,1)$ (solid line), estimated from an ensemble of $N = 6400$ samples and represented for the different length-scale values: (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

3.3. Discretization of the Nonlinear Filtering. Under Gaussian assumption, a discrete version of the prior distribution $p(\mathcal{X}_q)$ is given by an ensemble of samples $(\mathcal{X}_{q,k}^b)_{k \in [1,N_e]}$, such that

$$\mathcal{X}_{q,k}^b = \mathcal{X}_q^b + \varepsilon_{q,k}^b, \tag{12}$$

where $\varepsilon_{q,k}^b$ are independent sample of the Gaussian law $\mathcal{N}(0, \mathbf{B}_q)$. The empirical distribution is given by $p_q^e(\mathcal{X}_q) = (1/N_e) \sum_k \delta(\mathcal{X}_q - \mathcal{X}_{q,k}^b)$. The convergence of the empirical distribution toward the distribution $p(\mathcal{X}_q)$ has to be understood as the weak convergence, that is, for all f being of polynomial type, $\lim_{N_e \to \infty} P_q^e(f) = P_q(f)$, where $P_q(f) = \mathbb{E}_{P_q}[f(\mathcal{X}_q)] = \int_{\mathcal{X}_q} f(\mathcal{X}_q) p(\mathcal{X}_q) d\mathcal{X}_q$ and $P_q^e(f) = (1/N_e) \sum_k f(\mathcal{X}_{q,k}^b)$.

From the asymptotic concentration of Gaussian law in high dimension, it results that

$$\left|\mathcal{X}_{q,k}^b - \mathcal{X}_q^b\right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\text{tr}\left(\mathbf{B}_q\right)}, \frac{1}{2}\frac{\text{tr}\left(\mathbf{B}_q\right)}{d_{\text{ef}}\left(\mathbf{B}_q\right)}\right). \tag{13}$$

A numerical experiment supports this asymptotic distribution as illustrated in Figure 5. At first order, the histogram of the normalized quantity fits very well the theoretical normal law, whatever the length-scale L_p. As a result, the background samples $\mathcal{X}_{q,k}^b$ are all distributed within a spherical shell of radius $\sqrt{\text{tr}(\mathbf{B}_q)}$ and of thickness $\sqrt{(1/2)(\text{tr}(\mathbf{B}_q)/d_{\text{ef}}(\mathbf{B}_q))}$. One can notice a slight asymmetry as the length-scale increases (see Figures 5(c) and 5(d)). This

is related to the asymmetry of the chi-squared distribution which is weakly approximated by a Gaussian distribution (see Appendix A). This approximation is not accurate for the small dimension sizes as encountered for $L_p = 500$ km and $L_p = 1000$ km, where the effective dimensions are of orders $d_{\text{ef}}(500\,\text{km}) \approx 40$ and $d_{\text{ef}}(1000\,\text{km}) \approx 20$ (see Figure 3(b)).

Note that (13) is useful to understand the effect of the localization based on the Schur product, often used in EnKF to limit the spurious long distance correlations resulting from the sampling noise in the background covariance matrix estimation [24]. In this framework, a localized background covariance matrix $\mathbf{B}_q^{\text{loc}}$ is built from the element-wise product, $\circ$, of a correlation matrix $\mathbf{C}$ with the ensemble estimated background covariance matrix $\mathbf{B}_q^e$; that is, $\mathbf{B}_q^{\text{loc}} = \mathbf{C} \circ \mathbf{B}_q^e$. Since $\mathbf{C}$ is a covariance matrix, its diagonal coefficients are all equal to 1; as a result the traces of $\mathbf{B}_q^{\text{loc}}$ and $\mathbf{B}_q^e$ are equal. Moreover, since the correlation functions in $\mathbf{C}$ are chosen as compact support functions, the effective dimensions are different with $d_{\text{ef}}(\mathbf{B}^{\text{loc}}) > d_{\text{ef}}(\mathbf{B}^e)$. Hence, from (13), the variance associated with the localized matrix is smaller. As an illustration from Figure 5, the effect of the localization is equivalent to the transformation from the distribution in Figure 5(d) to the distribution in Figure 5(a). Thereafter, the question of the localization is no more considered.

At this stage, there is no difference between the EnKF and the PF strategies, and the difference takes place in the way the Bayes rule is applied.

3.4. Analysis Step in the EnKF.

In this section we do not detail the practical implementation of the EnKF but only the formalism (see, e.g., Houtekamer and Mitchell [24]; Evensen [25] for practical aspects).

For the EnKF, the analysis $\mathcal{X}_q^a = \mathcal{X}_q^b + \mathbf{K}_q(\mathcal{Y}_q^o - \mathbf{H}_q\mathcal{X}_q^b)$ is computed, and when using a "perturbation of observations" method [26], an ensemble of analysis perturbations is generated according to

$$\varepsilon_{q,k}^a = \varepsilon_{q,k}^b + \mathbf{K}\left(\varepsilon_{q,k}^o - \mathbf{H}\varepsilon_{q,k}^b\right), \tag{14}$$

where $\varepsilon_{q,k}^o$ is an ensemble of independent observational errors, randomly generated from the true covariance matrix $\mathbf{R}_q$ as $\varepsilon_{q,k}^o = \mathbf{R}_q^{1/2}\zeta_k$ with the random normal samples $\zeta_k \sim \mathcal{N}(0, \mathbf{R}_q)$. From computation [26], it follows that

$$\varepsilon_{q,k}^a \sim \mathcal{N}\left(0, \mathbf{A}_q\right). \tag{15}$$

The analysis ensemble is then constructed as

$$\mathcal{X}_{q,k}^a = \mathcal{X}_q^a + \varepsilon_{q,k}^a, \tag{16}$$

and then, from the application of asymptotic concentration of Gaussian law,

$$\left|\mathcal{X}_{q,k}^a - \mathcal{X}_q^a\right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{\text{tr}\left(\mathbf{A}_q\right)}, \frac{1}{2}\frac{\text{tr}\left(\mathbf{A}_q\right)}{d_{\text{ef}}\left(\mathbf{A}_q\right)}\right). \tag{17}$$

Similarly to the validation of the asymptotic distribution (13), a numerical experiment supports the asymptotic distribution (17) as illustrated in Figure 6. In this experiment,

an observation is placed at every 3 grid points, and the observational covariance matrix $\mathbf{R}_q$ is set as $\mathbf{R}_q = \sigma_o^2\mathbf{I}_p$, where $\sigma_o = 1$ and with $p = n/3$ being the dimension of the observational space. The matrix operations (trace, product,...) are computed directly within the numerical setting. At first order, the histogram of the normalized quantity fits very well to the theoretical normal law, whatever the length-scale L_p is. One can again notice a slight asymmetry as the length-scale increases (see Figures 6(c) and 6(d)).

We observe that the ensemble of analysis samples $\mathcal{X}_{q,k}^a$ are all distributed within a spherical shell of radius $\sqrt{\text{tr}(\mathbf{A}_q)}$ and of thickness $\sqrt{(1/2)(\text{tr}(\mathbf{A}_q)/d_{\text{ef}}(\mathbf{A}_q))}$. As expected, this shows that, from the distance view, the covariance matrix of the ensemble $\mathcal{X}_{q,k}^a$ is $\mathbf{A}_q$.

The EnKF transforms an ensemble of background samples $\mathcal{X}_{q,k}^b$ distributed on a spherical shell of radius $\sqrt{\text{tr}(\mathbf{B}_q)}$ into an ensemble of analysis samples $\mathcal{X}_{q,k}^a$ distributed on a spherical shell of radius $\sqrt{\text{tr}(\mathbf{A}_q)}$. This is the exact solution of the Bayes rule as supported by the numerical experiment. Hence, the statistics of the two distances appear to be a strong constraint; the strength increases with the dimension of the problem. This is the main property we want to investigate in order to explore the difference between the EnKF and the PF.

The analysis step for the PF is now detailed.

3.5. Analysis Step in the PF.

Several strategies exist for the particle discretization of the Bayes rule. However, as far as we know, most of these variants are suffering from the curse of dimensionality, for example, the importance sampling version as described by Snyder [27]. Hence, in this note, only the bootstrap filter is considered.

The analysis step in the PF [5, 6, 8, 9] follows a quite different way as it directly relies on the Bayes rule: for each member $\mathcal{X}_{q,k}^b$ one computes the weight

$$w_{q,k} = \frac{p\left(\mathcal{Y}_q^o \mid \mathcal{X}_{q,k}^b\right)}{\sum_m p\left(\mathcal{Y}_q^o \mid \mathcal{X}_{q,m}^b\right)}, \tag{18}$$

from which the a posteriori distribution is deduced as

$$P_q^e\left(\mathcal{X}_q \mid \mathcal{Y}_q^o\right) = \sum_k w_{q,k}\delta\left(\mathcal{X}_q - \mathcal{X}_{q,k}^b\right), \tag{19}$$

where $\delta(\mathcal{X})$ stands for the Dirac distribution positioned in 0. For the bootstrap particle filter [8], an ensemble of analyses $\mathcal{X}_{q,k}^a$ is generated as N_e random samples of the distribution P_q^e.

Hence, the difference between the EnKF and the PF is that, for the EnKF, background members are corrected to build analysis members, while, for PF, the analysis ensemble is a resampling of the background ensemble using the a posteriori weight. A natural question is to know when a background sample is an analysis sample of the particle filter. This is now addressed.
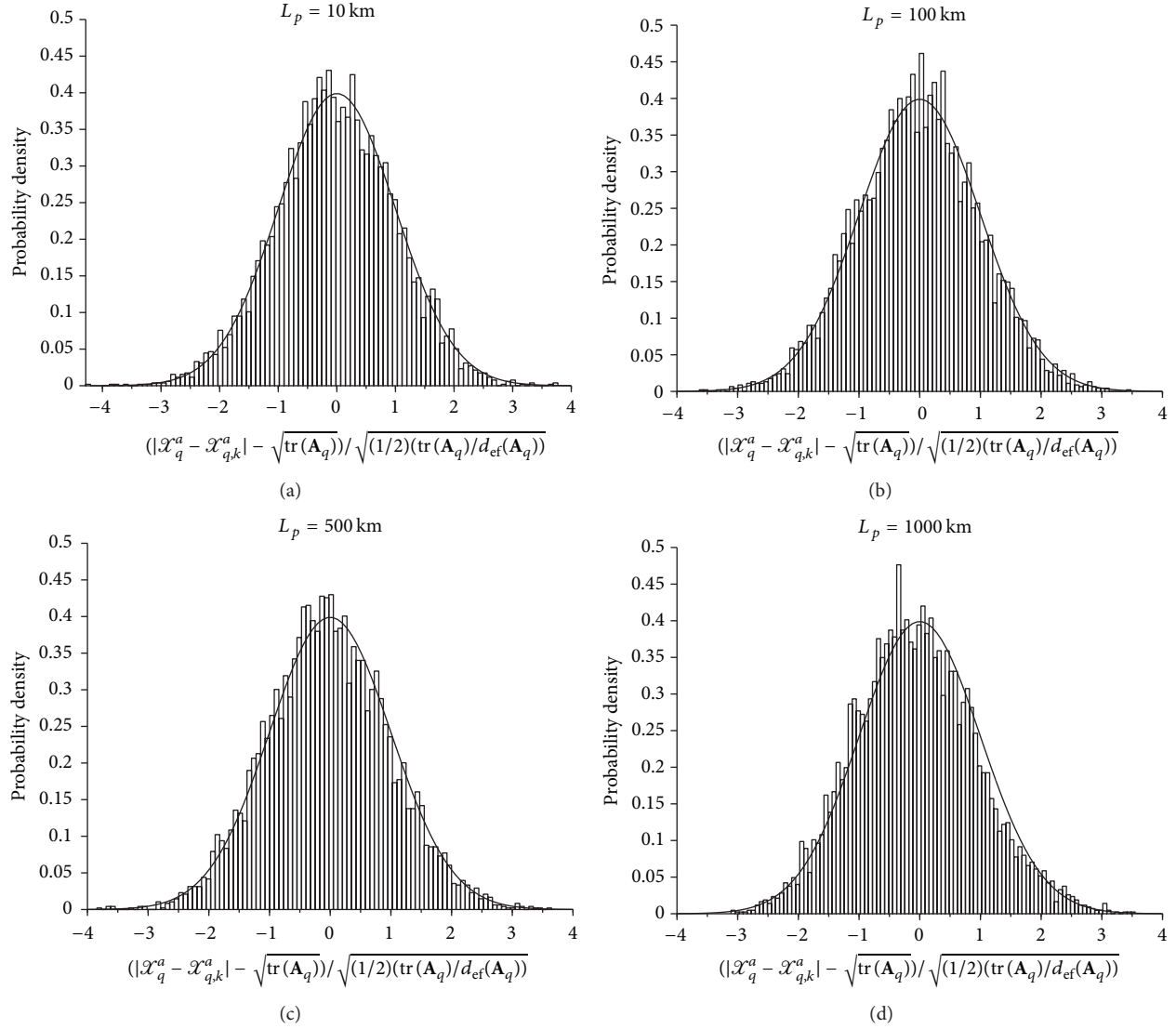
FIGURE 6: Histogram of distribution of the normalized distance $(|\mathcal{X}_q^a - \mathcal{X}_{q,k}^a| - \sqrt{\mathrm{tr}(\mathbf{A}_q)})/\sqrt{(1/2)(\mathrm{tr}(\mathbf{A}_q)/d_{\mathrm{ef}}(\mathbf{A}_q))}$ deduced from the EnKF and compared with the theoretical asymptotic Gaussian law $\mathcal{N}(0, 1)$ (solid line), estimated from an ensemble of $N = 6400$ samples and represented for the different length-scale values: (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

3.6. Compatibility Relation and Minimum Ensemble Size. We want to characterize in which case a background sample $\mathcal{X}_{q,k}^b$ can be considered as an analysis sample $\mathcal{X}_{q,k}^a$. This is achieved by giving some compatibility relations that are described in the first subsection. This is followed by two subsections that lead to formulating a constraint on the ensemble size.

3.6.1. Compatibility Relations. As shown for the EnKF, all the analysis members $\mathcal{X}_{q,k}^a$ are at a given fixed distance $\sqrt{\mathrm{tr}(\mathbf{A}_q)}$ from the analysis $\mathcal{X}_q^a$. More precisely, since the distance $|\mathcal{X}_{q,k}^a - \mathcal{X}_q^a|$ asymptotically follows the Gaussian distribution

(17), the maximum value $d_{a_q,a_{q,k}}^+$ that the distance can reach for an ensemble size of N_e is given by (see Appendix C)

$$
\begin{aligned}
d_{a_q,a_{q,k}}^+ &= \max_{k \in [1,N]} \left| \mathcal{X}_{q,k}^a - \mathcal{X}_q^a \right| \\
&= \sqrt{\mathrm{tr}\left(\mathbf{A}_q\right)} + \sqrt{\frac{1}{2} \frac{\mathrm{tr}\left(\mathbf{A}_q\right)}{d_{\mathrm{ef}}\left(\mathbf{A}_q\right)}} \sqrt{2 \log N_e}.
\end{aligned}
\tag{20}
$$

From the EnKF equations, $\mathcal{X}_{q,k}^b - \mathcal{X}_q^a = \varepsilon_{q,k}^b - \mathbf{K}_q(\mathcal{Y}_q^o - \mathbf{H}_q \mathcal{X}_q^b)$, where it results that $\mathcal{X}_{q,k}^b - \mathcal{X}_q^a$ is the random vector

$$
\mathcal{X}_{q,k}^b - \mathcal{X}_q^a \sim \mathcal{N}\left(-\delta \mathcal{X}_q^a, \mathbf{B}_q\right),
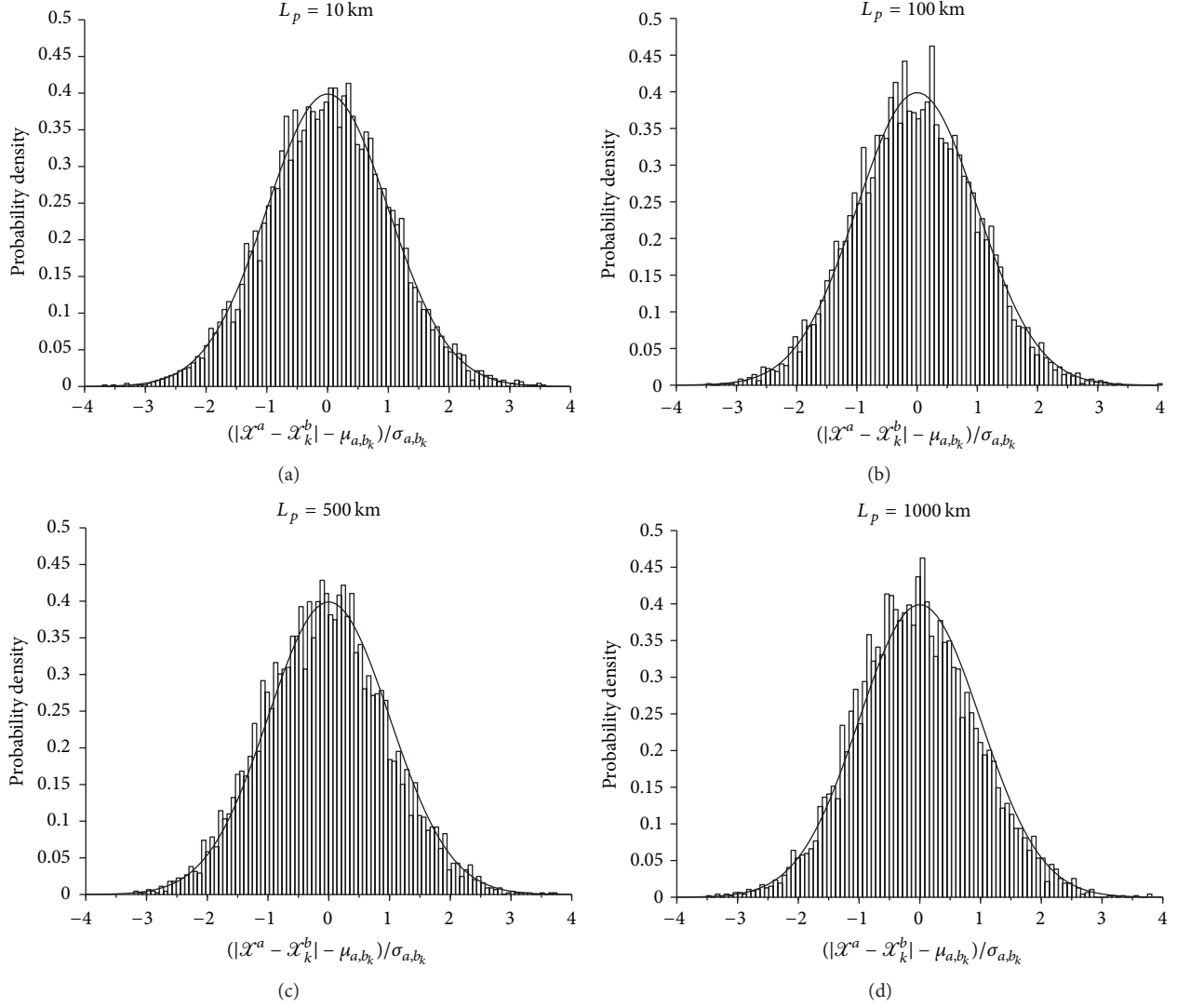\tag{21}
$$

FIGURE 7: Histogram of distribution of the normalized distance $(|\mathcal{X}^a - \mathcal{X}_k^b| - \mu_{a,b_k})/\sigma_{a,b_k}$ deduced from the EnKF (and valid here for the PF due to the equivalence in the Gaussian setting), compared with the theoretical asymptotic Gaussian law $\mathcal{N}(0, 1)$ (solid line), estimated from an ensemble of $N = 6400$ samples and represented for the different length-scale values: (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

where $\delta\mathcal{X}_q^a = \mathbf{K}_q(\mathcal{Y}_q^o - \mathbf{H}_q\mathcal{X}_q^b)$ is the analysis increment. We deduce (Appendix A) that the asymptotic distribution of the norm is the Gaussian law

$$\left|\mathcal{X}_{q,k}^b - \mathcal{X}_q^a\right| \underset{n \gg 1}{\sim} \mathcal{N}\left(\mu_{a_q,b_{q,k}}, \sigma_{a_q,b_{q,k}}^2\right),$$

$$\mu_{a_q,b_{q,k}} = \sqrt{\left|\delta\mathcal{X}_q^a\right|^2 + \operatorname{tr}\left(\mathbf{B}_q\right)},$$

$$\sigma_{a_q,b_{q,k}}^2 = \frac{1}{2} \frac{2\left\|\delta\mathcal{X}_q^a\right\|_{\mathbf{B}_q}^2 + \operatorname{tr}\left(\mathbf{B}_q^2\right)}{\left|\delta\mathcal{X}_q^a\right|^2 + \operatorname{tr}\left(\mathbf{B}_q\right)},$$

(22)

which is validated from numerical simulation in Figure 7. The minimum value $d_{a_q,b_{q,k}}^-$ that the distance $\left|\mathcal{X}_{q,k}^b - \mathcal{X}_q^a\right|$

can reach for an ensemble size of N_e is thus (see Appendix C)

$$d_{a_q,b_{q,k}}^-\left(\delta\mathcal{X}_q^a\right) = \min_{k \in [1,N]} \left|\mathcal{X}_{q,k}^b - \mathcal{X}_q^a\right|$$
$$= \mu_{a_q,b_{q,k}} - \sigma_{a_q,b_{q,k}}\sqrt{2\log N_e},$$

(23)

a quantity that depends on the analysis increment $\delta\mathcal{X}_q^a$.

Hence, for a background sample $\mathcal{X}_{q,k}^b$ to be likely an analysis sample $\mathcal{X}_{q,k}^a$, the following distance inequality should necessarily be verified:

$$d_{a_q,b_{q,k}}^-\left(\delta\mathcal{X}_q^a\right) \le \left|\mathcal{X}_{q,k}^b - \mathcal{X}_q^a\right| \le d_{a_q,a_{q,k}}^+.$$

(24)

If not, the two sample distributions of $\mathcal{X}_{q,k}^b$ and $\mathcal{X}_{q,k}^a$ are so different that no background sample can be retained as an
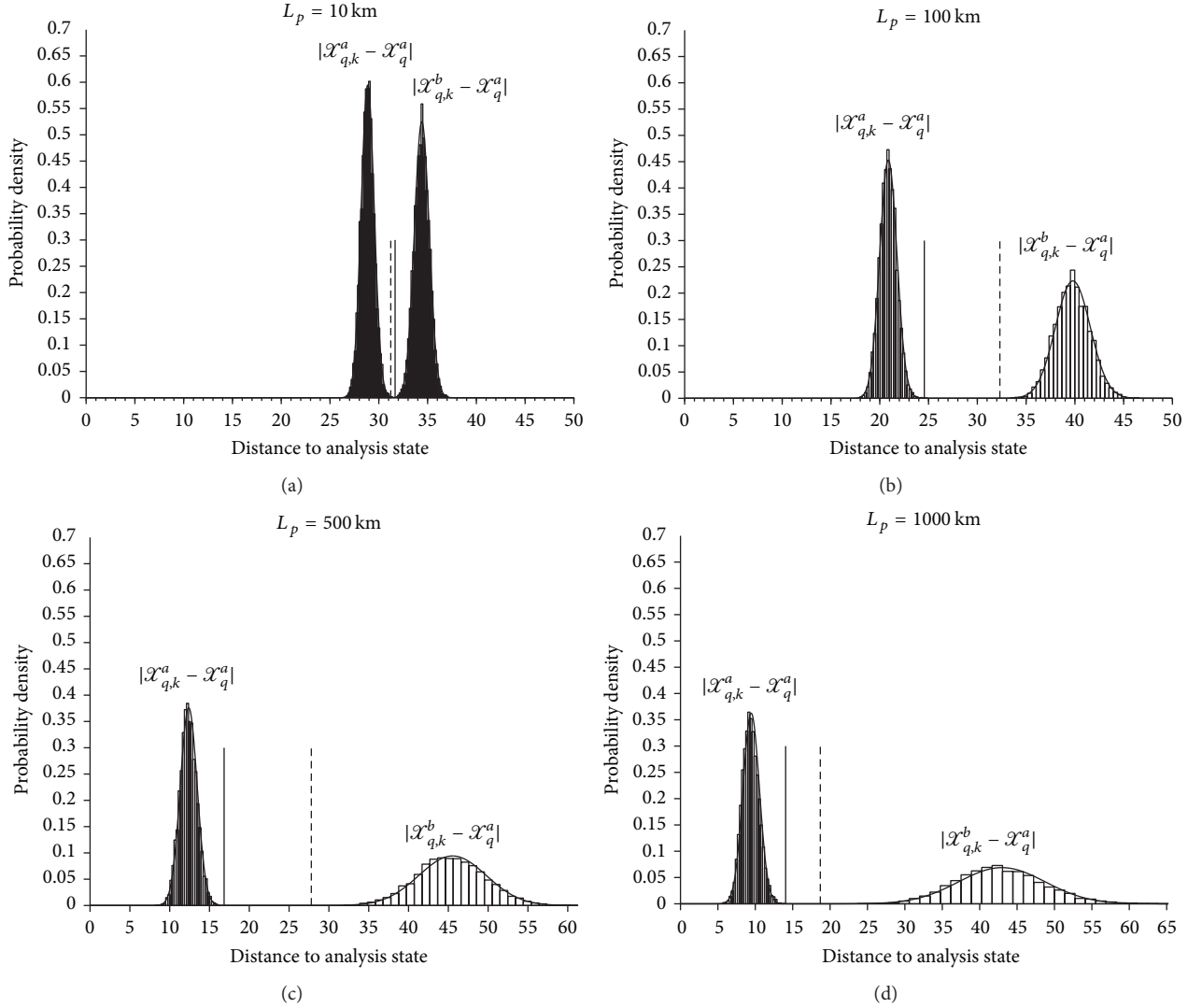
FIGURE 8: Histogram of distribution of $|\mathscr{X}^a_{q,k} - \mathscr{X}^a_q|$ deduced from the EnKF (and valid here for PF due to the equivalence in the Gaussian setting) (i.e., $|\varepsilon^a_{q,k}|$) compared to the one of $|\mathscr{X}^b_{q,k} - \mathscr{X}^a_q|$ for an ensemble size of $N = 6400$ samples. The maximum expected distance $d^+_{a_q,a_{q,k}}$ of $|\mathscr{X}^a_{q,k} - \mathscr{X}^a_q|$ is the vertical segment in solid line. The minimum expected distance $d^-_{a_q,b_{q,k}}$ of $|\mathscr{X}^b_{q,k} - \mathscr{X}^a_q|$ is the vertical segment in dashed line. The results are represented for the different length-scale values: (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

element of the analysis samples. Of course, this result only occurs for finite sample distributions. In finite dimension, if the size of ensemble is infinite, then the EnKF and the PF for this Gaussian case provide the same *a posteriori* distribution, corresponding to $d^-_{a_q,b_{q,k}}(\delta\mathscr{X}^a_q) = 0$ and $d^+_{a_q,a_{q,k}} = +\infty$.

This trade-off is now illustrated from a numerical experiment in Figure 8. Each panel compares the histogram of distribution of the distance $|\mathscr{X}^a_{q,k} - \mathscr{X}^a_q|$ and of the distance $|\mathscr{X}^b_{q,k} - \mathscr{X}^a_q|$, for a given finite ensemble of size $N_e = 6400$. The vertical solid segment specifies the position of the maximum of magnitude that can be reached by $|\mathscr{X}^a_{q,k} - \mathscr{X}^a_q|$, that is, $d^+_{a_q,a_{q,k}}$, while the vertical dashed segment is the position of the minimum of magnitude associated with $|\mathscr{X}^b_{q,k} - \mathscr{X}^a_q|$, that is, $d^-_{a_q,b_{q,k}}$. For the length-scale $L_p = 10$ km, the inequality $d^-_{a_q,b_{q,k}} \leq d^+_{a_q,a_{q,k}}$

implies almost surely the existence of a background sample $\mathscr{X}^b_{q,k}$ that can be considered as being an analysis sample. This means that, within a PF, a point $\mathscr{X}^b_{q,k}$, as (24) is verified, can be observed. However, this is no more the case for higher length-scale values where one observes that $d^+_{a_q,a_{q,k}} < d^-_{a_q,b_{q,k}}$.

The consequence is that the number of samples, N^-_q, required in order to inverse the inequality is the integer for which the equality applies; that is,

$$N^-_q\left(\delta\mathscr{X}^a_q\right)$$
$$= \exp\left(\frac{1}{2}\frac{\left(\mu_{a_q,b_{q,k}} - \sqrt{\operatorname{tr}\mathbf{A}_q}\right)^2}{\left(\sqrt{(1/2)\left(\operatorname{tr}\mathbf{A}_q/d_{\mathrm{ef}}\left(\mathbf{A}_q\right)\right)} + \sigma_{a_q,b_{q,k}}\right)^2}\right). \quad (25)$$

Hence, N_q^- is an exponentially large number, function of the analysis increment $\delta\mathscr{X}_q^a$. Note that this is also a function of $\mathbf{B}_q$ because of (22) and the dependence of $\mathbf{A}_q$ on $\mathbf{B}_q$.

The dependence in $\delta\mathscr{X}_q^a$ can be removed considering a climatological framework, as follows.

3.6.2. Climatological Approximation. In order to eliminate the dependence on the analysis increment, an average estimation of the minimum ensemble size $N_q^-(\delta\mathscr{X}_q^a)$ can be computed under the additional assumption that the statistics are sample of a climatology, eliminating the bottom index q. If $\mathbf{B}$ ($\mathbf{R}$) denotes the climatological background (observational) error covariance matrix, then we set $\mathbf{B}_q = \mathbf{B}$ and $\mathbf{R}_q = \mathbf{R}$. Hence, the analysis increment of a day is the random Gaussian vector

$$\delta\mathscr{X}_q^a \sim \mathcal{N}(0, \mathbf{KHB}), \qquad (26)$$

leading, in high dimension, to the asymptotic distribution

$$\left|\delta\mathscr{X}_q^a\right| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{\mathrm{tr}(\mathbf{KHB})}, \frac{1}{2}\frac{\mathrm{tr}(\mathbf{KHB})}{d_{\mathrm{ef}}(\mathbf{KHB})}\right). \qquad (27)$$

Moreover, the average values $\mu_{a,b_k}(\sigma_{a,b_k}^2)$ of $\mu_{a_q,b_{q,k}}(\sigma_{a_q,b_{q,k}}^2)$ can be deduced from the computation

$$\mathbb{E}\left[\left\|\delta\mathscr{X}_q^a\right\|_{\mathbf{B}}^2\right] = \mathrm{tr}\left(\mathbf{B}\mathbb{E}\left[\delta\mathscr{X}_q^a\delta\mathscr{X}_q^{aT}\right]\right) = \mathrm{tr}\left(\mathbf{BKHB}\right)$$

$$= \mathrm{tr}\left(\mathbf{KHB}^2\right), \qquad (28)$$

and they are given by

$$\mu_{a,b_k} = \sqrt{\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]},$$

$$\sigma_{a,b_k}^2 = \frac{1}{2}\frac{2\,\mathrm{tr}\left(\mathbf{KHB}^2\right)+\mathrm{tr}\left(\mathbf{B}^2\right)}{\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]}. \qquad (29)$$

Hence, the average value N^- of N_q^- is reduced to

$$N^- = \exp\left(\frac{1}{2}\frac{\left(\sqrt{\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]}-\sqrt{\mathrm{tr}\mathbf{A}}\right)^2}{\left(\sqrt{(1/2)\left(\mathrm{tr}\mathbf{A}/d_{\mathrm{ef}}(\mathbf{A})\right)}+\sqrt{(1/2)\left(\left(2\,\mathrm{tr}\left(\mathbf{KHB}^2\right)+\mathrm{tr}\left(\mathbf{B}^2\right)\right)/\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]\right)}\right)^2}\right). \qquad (30)$$

From general properties of Gaussian law, more than 99% of samples of $|\mathscr{X}_k^a - \mathscr{X}^a| \underset{n\gg 1}{\sim} \mathcal{N}(\sqrt{\mathrm{tr}\mathbf{A}}, (1/2)(\mathrm{tr}\mathbf{A}/d_{\mathrm{ef}}(\mathbf{A})))$ are within $\sqrt{\mathrm{tr}\mathbf{A}} \pm 3\sqrt{(1/2)(\mathrm{tr}\mathbf{A}/d_{\mathrm{ef}}(\mathbf{A}))}$; another ensemble size $N^{99\%}$ can be deduced, which corresponds to an ensemble size which is able to cover the analysis sample with more than 99% of probability; it writes (see Appendix C)

$$N^{99\%} = \exp\left(\frac{1}{2}\right.$$

$$\left.\cdot \frac{\left(\sqrt{\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]}-\sqrt{\mathrm{tr}\mathbf{A}}+3\sqrt{(1/2)\left(\mathrm{tr}\mathbf{A}/d_{\mathrm{ef}}(\mathbf{A})\right)}\right)^2}{(1/2)\left(\left(2\,\mathrm{tr}\left(\mathbf{KHB}^2\right)+\mathrm{tr}\left(\mathbf{B}^2\right)\right)/\mathrm{tr}\left[(\mathbf{I}+\mathbf{KH})\mathbf{B}\right]\right)}\right). \qquad (31)$$

This ensemble size may be considered as an upper bound for the maximal size required to sample the analysis distribution with the PF from the weighting and the resampling strategy.

The ensemble size N^- (solid line with diamond) and $N^{99\%}$ (solid line with triangle) are represented in Figure 9 as a function of the length-scale. The solid line with no mark represents the ensemble size of 6400. In this experiment, the ensemble size is maximum for medium length-scale values. For the small length-scale $L_p = 10$ km, when the length-scale is very small, the minimal ensemble size N^- is below 6400. This relatively small value is due to the observational network:

here only one per three grid points is observed; hence when there is no correlation in the background, the posterior distribution that is provided from the few observations is not so different from the prior distribution. When the length-scale is very large, for example, $L_p \in \{500, 1000\}$ km, then the effective dimension is small (see Figure 3), and only a relatively small ensemble size is required. The case $L_p = 100$ km is intermediate; the background error correlation implies a coupling between the points while the effective dimension remains important. Note that these results are in accordance with the one presented in Figure 8. $N^{99\%}$ follows a similar behaviour, but the gap with N^- is not constant and varies with the length-scale.

For the particular case where $\mathbf{B} = \mathbf{H} = \mathbf{R} = \mathbf{I}$, with $\mathbf{KH} = \mathbf{A} = \mathbf{I}/2$, these two bounds are given by $\log_{10}N^- = (1/(2\log 10))((\sqrt{3/2}-\sqrt{1/2})^2 n/((1/2)+\sqrt{2/3})^2) \approx 0.03n$ (this is in accordance with Snyder et al. [2], Figure 2) and $\log_{10}N^{99\%} = (1/(2\log 10))((\sqrt{3n/2}-\sqrt{n/2}+(3/2))^2/(2/3)) \approx 0.08n + 0.5\sqrt{n} + 0.73$, which are exponential quantities depending on the dimension n.

The magnitude reached by the minimal ensemble size N^- reminds one of the limits of the Monte-Carlo strategy when using the PF algorithm and the pitfall it presents.

3.6.3. Illustration of the Hyperspheres and Orthogonality Relation. From averaged values of (27) and (29), it is possible to represent the hyperspheres of the distributions; see
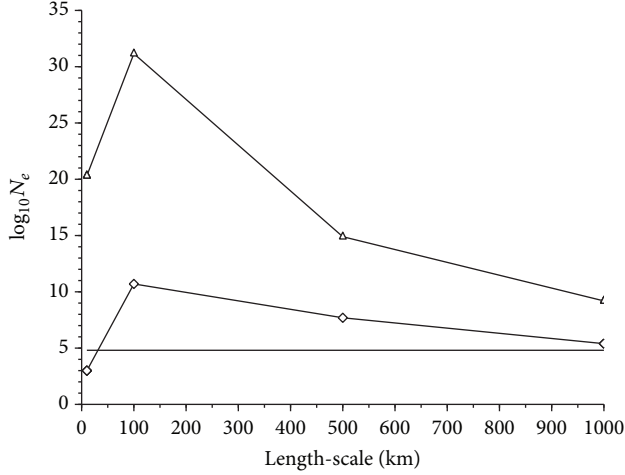
FIGURE 9: Illustration of the ensemble size deduced from the averaged values N^- (solid line with diamond marks) and $N^{99\%}$ (solid line with triangle marks), for the length-scale values $L_p \in \{10, 100, 500, 1000\}$ km, with the dimension set to $n = 1000$. The solid line with no mark represents the ensemble size 6400.

Figure 10. On all these panels, the $+$ symbol denotes the background $\mathcal{X}_q^b$ and $\times$ symbol denotes the average position of the analysis state $\mathcal{X}_q^a$ (the average value of $|\mathcal{X}_q^a - \mathcal{X}_q^b|$ is $\sqrt{\text{tr}(\mathbf{KHB})}$). The hypersphere representing the typical position of samples of the background distribution $\mathcal{N}(\mathcal{X}_q^b, \mathbf{B}_q)$ is the circle centered in $\mathcal{X}_q^b$ and of radius $\sqrt{\text{tr}(\mathbf{B}_q)}$, denoted by $\mathscr{C}_1 = \mathscr{C}(\mathcal{X}_q^b, \sqrt{\text{tr}(\mathbf{B}_q)})$. The typical position of samples of the analysis distribution $\mathcal{N}(\mathcal{X}_q^a, \mathbf{A}_q)$ is the circle with triangle centered in $\mathcal{X}_q^a$ and of radius $\sqrt{\text{tr}(\mathbf{A}_q)}$, denoted by $\mathscr{C}_2 = \mathscr{C}(\mathcal{X}_q^a, \sqrt{\text{tr}(\mathbf{A}_q)})$. The circle with diamond represents the hypersphere, $\mathscr{C}_3 = \mathscr{C}(\mathcal{X}_q^a, \sqrt{\text{tr}[(\mathbf{I} + \mathbf{K}_q\mathbf{H}_q)\mathbf{B}_q]})$, that contains the typical position deduced from the distribution of $|\mathcal{X}_q^a - \mathcal{X}_{q,k}^b|$. Since $\mathcal{X}_{q,k}^b$ should be typical of $\mathscr{C}_1$ and $\mathscr{C}_2$, all the typical positions of the point $\mathcal{X}_{q,k}^b$ view from the point $\mathcal{X}_q^a$ does not lie onto the whole hypersphere but only at the intersection of the two hyperspheres $\mathscr{C}_1$ and $\mathscr{C}_3$, that is, a hypersphere contained in the equatorial plane of $\mathscr{C}_1$ (here, the hypersphere of intersection, centered in $\mathcal{X}_q^b$, is reduced to the two points of intersection). In particular, it appears that the analysis increment $\delta\mathcal{X}_q^a$ is orthogonal to most of the background sample error $\varepsilon_{q,k}^b = \mathcal{X}_{q,k}^b - \mathcal{X}_q^b$.

This is another effect of the high dimension of the problem that can be understood from the computation of the cosine $\cos(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b})$ of the angle between the analysis increment $\delta\mathcal{X}_q^a$ and the background samples $\varepsilon_{q,k}^b$, where

$$\cos\left(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b}\right) = \frac{\langle \delta\mathcal{X}_q^a \mid \varepsilon_{q,k}^b \rangle}{\left|\delta\mathcal{X}_q^a\right| \cdot \left|\varepsilon_{q,k}^b\right|}. \tag{32}$$

The histogram of the cosine distribution (32) is illustrated in Figure 11 for the various length-scale experiments (these are the unnormalized distributions in order to appreciate the concentration of the cosine around its average value 0). For all the length-scales, the distributions are shown to be centered on zero, meaning that the angle $(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b})$ is $\pi/2$, or, said differently, $\delta\mathcal{X}_q^a \perp \varepsilon_{q,k}^b$, for almost all k. It appears that the orthogonality is more pronounced for the short length-scales.

The distribution of this angle cosine can be theoretically approximated as follows. Since in high dimension the average norm of $|\delta\mathcal{X}_q^a|$ was shown to be $\sqrt{\text{tr}\,\mathbf{KHB}}$, and the norm of $\varepsilon_{q,k}^b$ is concentrated around $\sqrt{\text{tr}\,\mathbf{B}}$, the angle cosine can be approximated as follows:

$$\cos\left(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b}\right) \approx \frac{\langle \delta\mathcal{X}_q^a \mid \varepsilon_{q,k}^b \rangle}{\sqrt{\text{tr}\,\mathbf{KHB}}\sqrt{\text{tr}\,\mathbf{B}}}. \tag{33}$$

Furthermore, the random scalar $\langle \delta\mathcal{X}_q^a \mid \varepsilon_{q,k}^b \rangle$ is a Gaussian random value since it appears as being the linear transformation of the Gaussian random vector $\varepsilon_{q,k}^b$. Hence, the distribution of the cosine angle can be asymptotically approximated by

$$\cos\left(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b}\right) \underset{n\gg 1}{\sim} \mathcal{N}\left(0, \frac{\left\|\delta\mathcal{X}_q^a\right\|_{\mathbf{B}}^2}{\text{tr}\,(\mathbf{KHB})\,\text{tr}\,(\mathbf{B})}\right). \tag{34}$$

The Gaussian approximation (34) is represented, for each length-scale, in Figure 11 by the solid line curve. It appears as being a good approximation of the angle cosine distribution $\cos(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b})$.

An averaged theoretical distribution for $\cos(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b})$ can be deduced from (28) leading to the Gaussian $\mathcal{N}(0, \text{tr}(\mathbf{KHB}^2)/\text{tr}(\mathbf{KHB})\,\text{tr}(\mathbf{B}))$. For the particular case where $\mathbf{B} = \mathbf{H} = \mathbf{R} = \mathbf{I}$ with $\mathbf{KH} = \mathbf{A} = \mathbf{I}/2$, it results that the distribution of the angle cosine is well approximated by

$$\cos\left(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b}\right) \underset{n\gg 1}{\sim} \mathcal{N}\left(0, \frac{1}{n}\right), \tag{35}$$

meaning that the orthogonality is more pronounced for small length-scales.

3.6.4. Recommendation for PF and Beyond the Gaussian Framework. A known result is that if the PF is quite interesting for non-Gaussian data assimilation; it faces the curse of dimensionality for problems of large size. The present contribution illustrates the geometrical interpretation of the curse of dimensionality, considering the favorable case where the distribution is Gaussian, so that the EnKF and the PF should produce the same analysis distribution. The limit of the use of PF for high dimension is that the distance between the background samples and the typical support of the analysis distribution would be too large to select some background samples as analysis samples. Of course, in the realm of geophysical data assimilation where n is of order
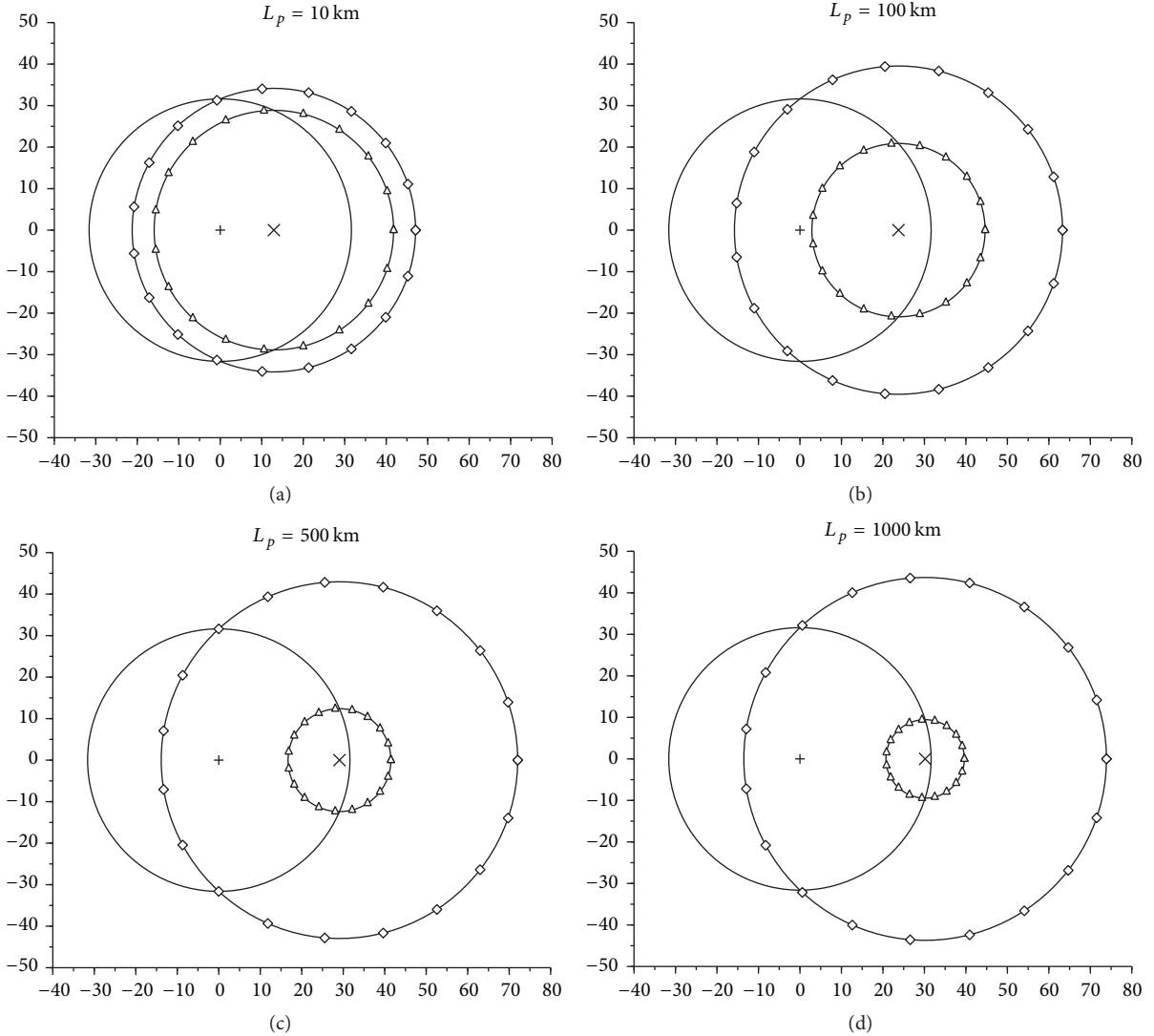
FIGURE 10: Illustration of the hyperspheres that contain the typical samples of the prior and *a posteriori* Gaussian distributions valid for the EnKF and the PF in the Gaussian setting, where + denotes the background $\mathcal{X}_q^b$, × denotes the analysis state $\mathcal{X}_q^a$, the circle without mark denotes the background distribution, the circle with triangle marks denotes the analysis distribution, and the circle with diamond marks is centered on the analysis state with a radius of the typical distance between the analysis state and the background samples. These hyperspheres are reproduced for the various length-scale parameters (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

10^8 (at the moment of writing this), we can expect the distributions to be highly non-Gaussian, and this could be an advantage for the PF. However, for such high dimension, 10^8, the diagnostic of the distance is entirely governed by the central limit theorem (or large deviation principle if we have to consider the correlation; see Appendix A), so the non-Gaussian distribution should lie within the hypersphere associated with the equivalent Gaussian distribution, that is, the Gaussian distribution whose covariance matrix is the covariance matrix of the non-Gaussian distribution (we assume that this is the last covariance matrix existing). Hence, even if non-Gaussian, the analysis distribution should lie within a hypersphere of smaller radius due to the data assimilation. As a consequence, the selection of background samples to produce analysis samples will face the same

difficulties as in the Gaussian case as illustrated in Sections 3.6.1 and 3.6.2.

In this contribution, and considering illustration in Figure 2, it appears that high dimension starts from surprisingly very small effective dimension: $d_{\text{ef}}(\mathbf{B}) = 100$ can be considered as high dimension. Hence, for problems of typical effective dimension larger than 100, we recommend not using the PF algorithm since it can not be used with a reasonable ensemble size as encountered in the EnKF (dozen of members when writing the manuscript). Note that this recommendation is stated considering the effective dimension of the problem, not the dimension n of $\mathbb{R}^n$ where the state vector lies.

As a consequence, and this is our strong recommendation, any new PF algorithm trying to tackle high dimension
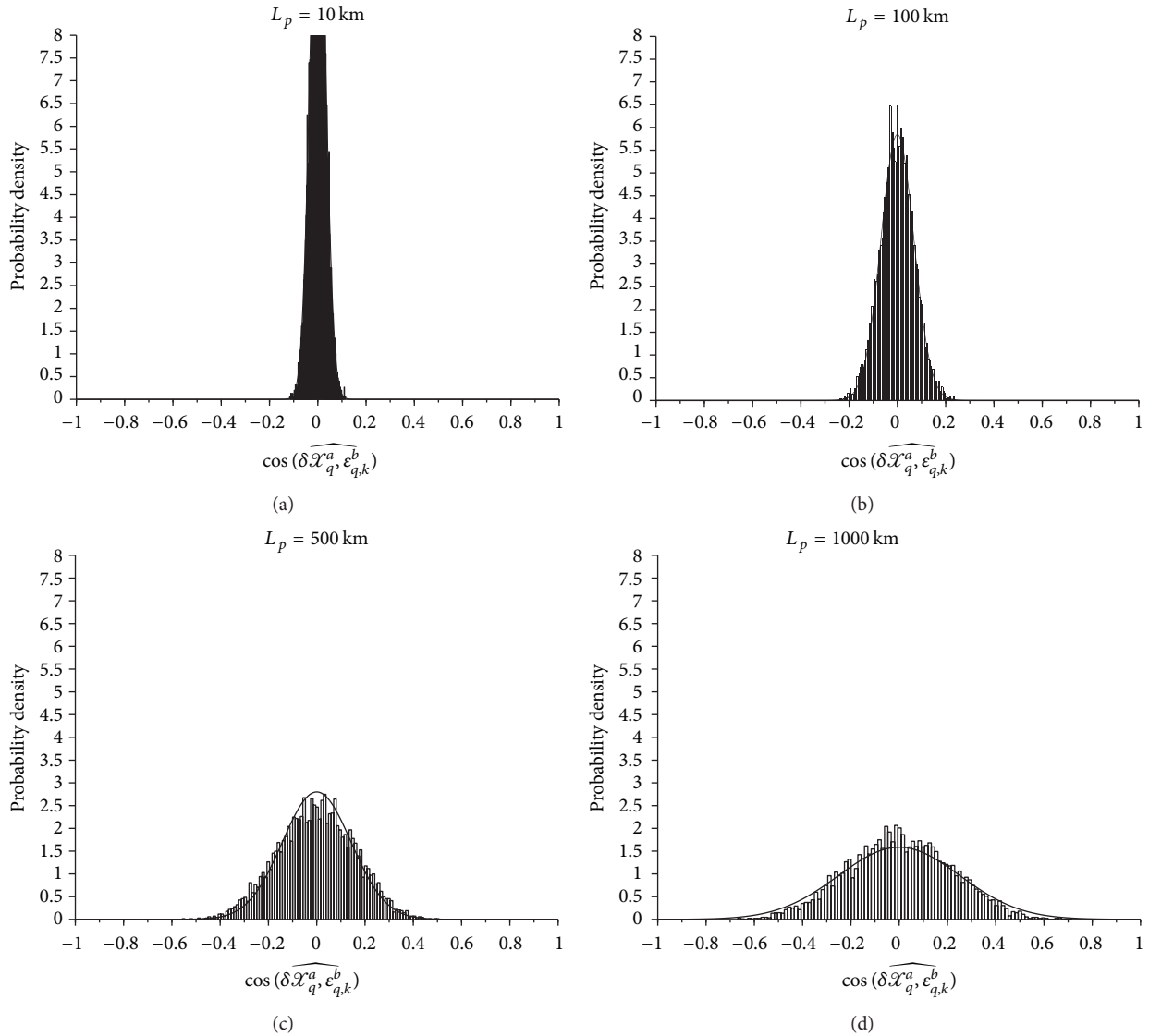
Figure 11: Histogram of $\cos(\widehat{\delta\mathcal{X}_q^a, \varepsilon_{q,k}^b})$, the cosine of the angle between the analysis increment $\delta\mathcal{X}_q^a$ and the background samples $\varepsilon_{q,k}^b$ deduced from the EnKF (and valid here for the PF due to the equivalence in the Gaussian setting), for the various length-scales (a) $L_p = 10$ km, (b) $L_p = 100$ km, (c) $L_p = 500$ km, and (d) $L_p = 1000$ km, for the dimension set to $n = 1000$.

issue should proof its ability to face the intrinsic limit, due to the dimension, as explored here in terms of distance. Said differently, the distance can be considered as a diagnostic to test the ability of a new algorithm to cope with the curse of dimension: if the new algorithm is suffering from the distance consequences, then the algorithm can not be used for the high dimension, and if the new algorithm is compatible with the distance consequences, then nothing more can be said from the distance point of view and other diagnostics than the distance should be employed to conclude.

4. Conclusion

In this work, the behaviour of the convergence of a Gaussian law toward a hypersphere (or a part of it) in high dimension

has been considered in order to discuss the difference between the ensemble Kalman filter and the particle filter.

Both algorithms correspond to the discretization of the Bayes rule (the nonlinear filter), but they are known to lead to separate results in practice. This is mainly due to the curse of dimensionality effect: particle filter requires an exponentially large number of samples as the dimension of the problem increases.

The consequence of the concentration of Gaussian probability distributions on hyperspheres has been verified within an experimental setting. It appears that the distance between the mean of a Gaussian and one of its samples is equal to the square-root of the trace of the covariance matrix: the fluctuation magnitude of the distance around this square-root value is very small as the dimension increases. The

computation of the variance of the distance provides an approximation of the effective dimension of the covariance matrix.

These properties suggest that the distance plays the role of a diagnostic tool determining if a background sample can be compatible, or not, with a given analysis state. It appears that the ensemble Kalman filter (EnKF) is able to transform a background sample from the background hypersphere to the analysis hypersphere. If the prior and the posterior distributions are too different, no background sample can be selected in the particle filter (PF) at least when the ensemble size is not an exponential function of the dimension of the state space. A formula for the minimal bound of the ensemble size has been obtained from the concentration behaviour.

Some recommendations have been stated considering the non-Gaussian case, where the geometrical constraints associated with the distance imply a test that should be verified by any PF implementation for problems in high dimension.

The concentration behaviour is particularly interesting with the emergence of the weighting ensemble strategy that relies on the particle filter formalism to update the probabilistic information contained in ensembles of different origin. When an analysis state is available, then the diagnosis of the distance between the members of the ensemble to the analysis state can be used to position the ensemble onto the right hypersphere.

Appendix

A. Asymptotic Distributions for Gaussian Vector Norm

In this section, we want to feature the behaviour of the norm of a Gaussian random vector in $\mathbb{R}^n$ when the dimension n is high.

If $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Gamma})$ denotes Gaussian random vector of zero mean and with covariance matrix $\boldsymbol{\Gamma}$, the behaviour of $|\mathbf{x}|^2 = \sum_i x_i^2$ can be deduced from generalization of the central limit theorem (when assuming that coordinate random value x_i is not too strongly correlated but dependent, the large deviation principle can be invoked under the form of the Gartner-Ellis theorem; see Dembo and Zeitouni [28]; the rigorous derivation is beyond the scope of the present manuscript): in the present case $|\mathbf{x}|^2$ follows a Gaussian law entirely featured by its mean and its variance. These two parameters are determined as described now.

If $\boldsymbol{\Gamma}^{1/2}$ denotes a (symmetric) square-root of the covariance matrix $\boldsymbol{\Gamma}$, that is, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Gamma}^{1/2}$, then a sample of $\mathbf{x}$ can be obtained as the transformation $\mathbf{x} = \boldsymbol{\Gamma}^{1/2}\boldsymbol{\zeta}$ of a Gaussian random vector $\boldsymbol{\zeta} \sim \mathcal{N}(0, \mathbf{I})$. Then, $\mathbb{E}[|\mathbf{x}|^2] = \mathbb{E}[\boldsymbol{\zeta}^T\boldsymbol{\Gamma}\boldsymbol{\zeta}]$ can be written as $\mathbb{E}[|\mathbf{x}|^2] = \sum_{ij} \boldsymbol{\Gamma}_{ij}\mathbb{E}[\zeta_i\zeta_j]$. With $\mathbb{E}[\zeta_i\zeta_j] = \delta_{ij}$, where δ_{ij} is the Kronecker symbol, we obtain $\mathbb{E}[|\mathbf{x}|^2] = \sum_i \boldsymbol{\Gamma}_{ii}$ that is also

$$\mathbb{E}\left[|\mathbf{x}|^2\right] = \text{tr}\left(\boldsymbol{\Gamma}\right), \tag{A.1}$$

where $\text{tr}(\boldsymbol{\Gamma})$ denotes the trace of the $\boldsymbol{\Gamma}$ matrix. From the Wick formula,

$$\mathbb{E}\left[\zeta_i\zeta_j\zeta_k\zeta_l\right] = \mathbb{E}\left[\zeta_i\zeta_j\right]\mathbb{E}\left[\zeta_k\zeta_l\right] + \mathbb{E}\left[\zeta_i\zeta_k\right]\mathbb{E}\left[\zeta_j\zeta_l\right]$$
$$+ \mathbb{E}\left[\zeta_i\zeta_l\right]\mathbb{E}\left[\zeta_k\zeta_j\right], \tag{A.2}$$

the computation of $\mathbb{E}[|\mathbf{x}|^4] = \sum_{ijkl} \boldsymbol{\Gamma}_{ij}\boldsymbol{\Gamma}_{kl}\mathbb{E}[\zeta_i\zeta_j\zeta_k\zeta_l]$ provides that $\mathbb{E}[|\mathbf{x}|^4] = 2\,\text{tr}(\boldsymbol{\Gamma}^2) + \text{tr}(\boldsymbol{\Gamma})^2$. Thus, the variance of $V(|\mathbf{x}|^2) = \mathbb{E}[|\mathbf{x}|^4] - \mathbb{E}[|\mathbf{x}|^2]^2$ is

$$V\left(|\mathbf{x}|^2\right) = 2\,\text{tr}\left(\boldsymbol{\Gamma}^2\right). \tag{A.3}$$

Hence, the law of $|\mathbf{x}|^2$ is given from the mean equation (A.1) and the variance equation (A.3), by

$$|\mathbf{x}|^2 \underset{n\gg 1}{\sim} \mathcal{N}\left(\text{tr}\left(\boldsymbol{\Gamma}\right), 2\,\text{tr}\left(\boldsymbol{\Gamma}^2\right)\right). \tag{A.4}$$

This leads to writing a sample of $|\mathbf{x}|^2$ as $|\mathbf{x}|^2 = \text{tr}(\boldsymbol{\Gamma}) + \sqrt{2\,\text{tr}(\boldsymbol{\Gamma}^2)}\gamma$, where $\gamma \sim \mathcal{N}(0, 1)$.

From the Taylor expansion $(1 + \alpha)^{1/2} = 1 + (1/2)\alpha + \mathcal{O}(\alpha^2)$, applied for $\alpha = \sqrt{2\,\text{tr}(\boldsymbol{\Gamma}^2)}\gamma / \text{tr}(\boldsymbol{\Gamma})$, one obtains the approximation $|\mathbf{x}| = \sqrt{\text{tr}(\boldsymbol{\Gamma})} + (1/2)(\sqrt{2\,\text{tr}(\boldsymbol{\Gamma}^2)}/\sqrt{\text{tr}(\boldsymbol{\Gamma})})\gamma$ and deduces that the asymptotic distribution of the norm $|\mathbf{x}|$ is

$$|\mathbf{x}| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{\text{tr}(\boldsymbol{\Gamma})}, \frac{1}{2}\frac{\text{tr}\left(\boldsymbol{\Gamma}^2\right)}{\text{tr}(\boldsymbol{\Gamma})}\right). \tag{A.5}$$

Introducing the effective dimension $d_{\text{ef}}(\boldsymbol{\Gamma}) = (\text{tr}(\boldsymbol{\Gamma})^2)/(\text{tr}(\boldsymbol{\Gamma}^2))$ of $\boldsymbol{\Gamma}$ [20], the asymptotic distribution takes the form

$$|\mathbf{x}| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{\text{tr}(\boldsymbol{\Gamma})}, \frac{1}{2}\frac{\text{tr}(\boldsymbol{\Gamma})}{d_{\text{ef}}(\boldsymbol{\Gamma})}\right). \tag{A.6}$$

If $\mathbf{x}_k$ denotes an ensemble of N samples of $\mathbf{x}$, an estimation of the effective dimension $d_{\text{ef}}(\boldsymbol{\Gamma})$ can be obtained from the computation of estimation of the mean norm $M_N = (1/N)\sum_k |\mathbf{x}_k|$ which is theoretically equal to $\sqrt{\text{tr}(\boldsymbol{\Gamma})}$, and the variance $V_N = (1/N)\sum_k (|\mathbf{x}_k| - M_N)^2$ theoretically equal to $(1/2)(\text{tr}(\boldsymbol{\Gamma})/d_{\text{ef}}(\boldsymbol{\Gamma}))$, where one obtains the estimation

$$d_{\text{ef}}\left(\boldsymbol{\Gamma}\right) = \frac{M_N^2}{2V_N}. \tag{A.7}$$

Note that when $\mathbf{x}$ is of covariance matrix $\mathbf{I}_n$, then $|\mathbf{x}|^2$ follows a chi-squared distribution which is weakly (accurately) approximated by a Gaussian law for small (large) dimension size n.

When the mean of $\mathbf{x}$ is no more zero but $\mathbb{E}[\mathbf{x}] = \mathbf{m}$, that is, $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Gamma})$, the previous route is modified leading to the corrected asymptotic distributions

$$|\mathbf{x}|^2 \underset{n\gg 1}{\sim} \mathcal{N}\left(|\mathbf{m}|^2 + \text{tr}(\boldsymbol{\Gamma}), 4\mathbf{m}^T\boldsymbol{\Gamma}\mathbf{m} + 2\,\text{tr}\left(\boldsymbol{\Gamma}^2\right)\right),$$

$$|\mathbf{x}| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{|\mathbf{m}|^2 + \text{tr}(\boldsymbol{\Gamma})}, \frac{1}{2}\frac{2\mathbf{m}^T\boldsymbol{\Gamma}\mathbf{m} + \text{tr}\left(\boldsymbol{\Gamma}^2\right)}{|\mathbf{m}|^2 + \text{tr}(\boldsymbol{\Gamma})}\right). \tag{A.8}$$

B. Observational Variance Tuning

The observational variance tuning is not a question directly related to the issue of difference between EnKF and PF for small ensemble size. However, this is of crucial importance in the realm of data assimilation where very few is known about the quality of the observations. Since the investigation of observational variance tuning illustrates an interesting use of the hypersphere constraints, we have chosen to discuss this application here.

The observational variance tuning aims to provide an inflation factor α_q such that $\mathbf{R}_q = \alpha_q \widetilde{\mathbf{R}}_q$, where $\mathbf{R}_q$ is the observational covariance matrix and $\widetilde{\mathbf{R}}_q$ is the observational covariance matrix as implemented in a given data assimilation. If an ensemble of background samples $\mathcal{X}_{q,k}^b = \mathcal{X}_q^b + \varepsilon_{q,k}^b$ is available, then $\mathbf{H}_q \varepsilon_{q,k}^b \sim \mathcal{N}(0, \mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T)$ verifies the asymptotic distribution

$$\left|\mathbf{H}_q \varepsilon_{q,k}^b\right|^2 \underset{n \gg 1}{\sim} \mathcal{N}\left(\operatorname{tr} \mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T, 2 \operatorname{tr}\left[\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T\right)^2\right]\right). \quad \text{(B.1)}$$

Moreover, a synthetic observational error $\varepsilon_{q,k}^o$ can be generated from $\widetilde{\varepsilon_{q,k}^o} = \widetilde{\mathbf{R}}_q^{1/2} \zeta_k$ with $\zeta_k \sim \mathcal{N}(0, \mathbf{I}_p)$, where p denotes the dimension of the observational space. Hence, the asymptotic distribution of the square norm is

$$\left|\widetilde{\varepsilon_{q,k}^o}\right|^2 \underset{n \gg 1}{\sim} \mathcal{N}\left(\operatorname{tr} \widetilde{\mathbf{R}}_q, 2 \operatorname{tr}\left[\widetilde{\mathbf{R}}_q^2\right]\right). \quad \text{(B.2)}$$

Under the usual unbiased assumption, the innovation $\mathbf{d}_q = \mathcal{Y}_q^o - \mathbf{H}_q \mathbf{x}_b$ verifies $\mathbf{d}_q = \varepsilon_q^o - \mathbf{H}_q \varepsilon_q^b \sim \mathcal{N}(0, \mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q)$, leading in high dimension to the asymptotic distribution

$$\left|\mathbf{d}_q\right|^2 \underset{n \gg 1}{\sim} \mathcal{N}\left(\operatorname{tr}\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q\right), 2\right.$$
$$\left. \cdot \operatorname{tr}\left[\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q\right)^2\right]\right). \quad \text{(B.3)}$$

Combining together (B.1), (B.2), and (B.3), it follows that $\widehat{\alpha}_{q,k} = (|\mathbf{d}_q|^2 - |\mathbf{H}_q \varepsilon_{q,k}^b|^2)/|\widetilde{\varepsilon_{q,k}^o}|^2$ verifies the (first-order) asymptotic distribution

$$\widehat{\alpha}_{q,k} \underset{n \gg 1}{\sim} \mathcal{N}\left(\alpha_q, 2\right.$$
$$\left. \cdot \frac{\operatorname{tr}\left[\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q\right)^2\right] + \operatorname{tr}\left[\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T\right)^2\right]}{\left(\operatorname{tr} \widetilde{\mathbf{R}}_q\right)^2}\right), \quad \text{(B.4)}$$

with the average value $\overline{|\mathbf{H}_q \varepsilon_{q,k}^b|^2}$, computed as the ensemble mean of the square norms $|\mathbf{H}_q \varepsilon_{q,k}^b|^2$; the asymptotic distribution of $\widehat{\overline{\alpha}}_{q,k} = (|\mathbf{d}_q|^2 - \overline{|\mathbf{H}_q \varepsilon_{q,k}^b|^2})/|\widetilde{\varepsilon_{q,k}^o}|^2$ is given by

$$\widehat{\overline{\alpha}}_{q,k} \underset{n \gg 1}{\sim} \mathcal{N}\left(\alpha_q, 2 \frac{\operatorname{tr}\left[\left(\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q\right)^2\right]}{\left(\operatorname{tr} \widetilde{\mathbf{R}}_q\right)^2}\right). \quad \text{(B.5)}$$

Hence, the quantities $\widehat{\alpha}_q$ and $\widehat{\overline{\alpha}}_q$ are two estimators for the inflation factor α_q.

C. Bounds for Gaussian Samples

The behaviour of the maximum of Gaussian samples of a random variable X following a Gaussian law $\mathcal{N}(m, \sigma^2)$ of mean m and variance σ^2 has been characterized by Berman [29] who has shown that

$$\lim_{N_e} \mathrm{P}\left(\max_{i \in [1, N_e]} |X_i - m| \in \left[T - \frac{\log \log N_e}{\log N_e}, T\right]\right) = 1, \quad \text{(C.1)}$$

where

$$T = \sigma \sqrt{2 \log N_e}. \quad \text{(C.2)}$$

This is a classical result from which one has to understand that the maximum distance $|X - m|$ can be reached for a sampling ensemble of size N_e is given by T and the maximum (minimum) sample value is $m + T$ $(m - T)$.

In reverse, for a given threshold T, the expected minimum ensemble size N^- required to observe a maximum distance $|X - m|$ of magnitude T is given by

$$N^- = \exp\left(\frac{1}{2} \frac{T^2}{\sigma^2}\right). \quad \text{(C.3)}$$

Competing Interests

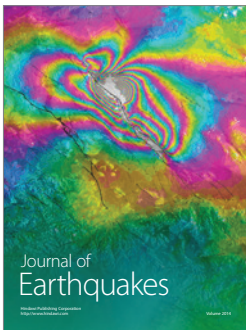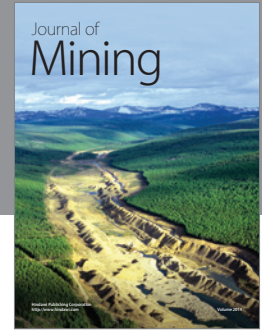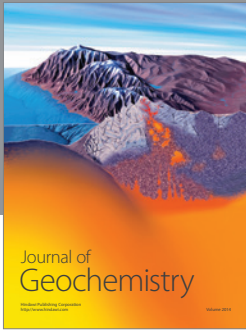The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] T. Bengtsson, P. Bickel, and B. Li, "Curseof-dimensionality revisited: collapse of the particle filter in very large scale systems," in *Collections*, pp. 316–334, Institute of Mathematical Statistics Collections, 2008.

[2] C. Snyder, T. Bengtsson, P. Bickel, and J. L. Anderson, "Obstacles to high-dimensional particle filtering," *Monthly Weather Review*, vol. 136, no. 12, pp. 4629–4640, 2008.

[3] R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, 1961.

[4] M. Bocquet, C. A. Pires, and L. Wu, "Beyond gaussian statistical modeling in geophysical data assimilation," *Monthly Weather Review*, vol. 138, no. 8, pp. 2997–3023, 2010.

[5] A. Doucet, N. de Freita, N. Gordon, and A. Smith, *Sequential Monte Carlo Methods in Practice*, Springer, Berlin, Germany, 2001.

[6] P. J. van Leeuwen, "A variance-minimizing filter for large-scale applications," *Monthly Weather Review*, vol. 131, no. 9, pp. 1175–2084, 2003.

[7] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99, no. 5, pp. 10143–10162, 1994.

[8] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.

[9] P. Del Moral, *Feynman-Kac Formulae: Ge-Nealogical and Interacting Particle Systems with Applications*, Springer, 2004.

[10] F. Le Gland, V. Monbet, and V.-D. Tran, "Large sample asymptotics for the ensemble Kalman filter," I. N. R. I. A 7014, 2009.

[11] J. Mandel, L. Cobb, and J. D. Beezley, "On the convergence of the ensemble Kalman filter," *Applications of Mathematics*, vol. 56, no. 6, pp. 533–541, 2011.

[12] A. Chorin and M. Morzfeld, "Conditions for successful data assimilation," *Journal Geophysical Research*, vol. 118, pp. 522–533, 2013.

[13] M. Talagrand, "A new look at independence," *The Annals of Probability*, vol. 24, no. 1, pp. 1–34, 1996.

[14] R. N. Hoffman and E. Kalnay, "Lagged average forecasting, an alternative to Monte Carlo forecasting," *Tellus A*, vol. 35, no. 2, pp. 100–118, 1983.

[15] A. R. Lawrence and J. A. Hansen, "A transformed lagged ensemble forecasting technique for increasing ensemble size," *Monthly Weather Review*, vol. 135, no. 4, pp. 1424–1438, 2007.

[16] Z. Ben Bouallègue, S. E. Theis, and C. Gebhardt, "Enhancing COSMO-DE ensemble forecasts by inexpensive techniques," *Meteorologische Zeitschrift*, vol. 22, no. 1, pp. 49–59, 2013.

[17] L. Descamps, C. Labadie, A. Joly, E. Bazile, P. Arbogast, and P. Cébron, "PEARP, the Météo-France short-range ensemble prediction system," *Quarterly Journal of the Royal Meteorological Society*, vol. 141, no. 690, pp. 1671–1685, 2015.

[18] M. Milan, D. Schüttemeyer, T. Bick, and C. Simmer, "A sequential ensemble prediction system at convection-permitting scales," *Meteorology and Atmospheric Physics*, vol. 123, no. 1-2, pp. 17–31, 2014.

[19] L. Raynaud, O. Pannekoucke, P. Arbogast, and F. Bouttier, "Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale," *Quarterly Journal of the Royal Meteorological Society*, vol. 141, no. 687, pp. 459–468, 2015.

[20] D. J. Patil, B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, "Local low dimensionality of atmospheric dynamics," *Physical Review Letters*, vol. 86, no. 26, pp. 5878–5881, 2001.

[21] A. Weaver and P. Courtier, "Correlation modelling on the sphere using a generalized diffusion equation," *Quarterly Journal Royal Meteorological Society*, vol. 127, no. 575, pp. 1815–1846, 2001.

[22] O. Pannekoucke and S. Massart, "Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 635, pp. 1425–1438, 2008.

[23] O. Pannekoucke, L. Berre, and G. Desroziers, "Background-error correlation length-scale estimates and their sampling statistics," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 631, pp. 497–508, 2008.

[24] P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble Kalman filter technique," *Monthly Weather Review*, vol. 126, no. 3, pp. 796–811, 1998.

[25] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, Germany, 2nd edition, 2009.

[26] G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724, 1998.

[27] C. Snyder, "Particle filters, the 'optimal' proposal and high-dimensional systems," in *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, ECMWF, Ed., pp. 1–10, ECMWF, Reading, UK, 2011.

[28] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Stochastic Modelling and Applied Probability, Springer, 1998.

[29] S. M. Berman, *Sojourners and Extremes of Stochastic Processes*, Wadsworth, Reading, Mass, USA, 1989.