From the Kalman Filter to the Particle Filter: A geometrical perspective of the curse of dimensionality

O. Pannekoucke, P. Cébron, N. Oger and Ph. Arbogast

CNRM/GAME, Météo-France/CNRS, UMR 3589,
team GMAP/RECYF
Toulouse, France.
corresponding author: olivier.pannekoucke@meteo.fr

December 3, 2014

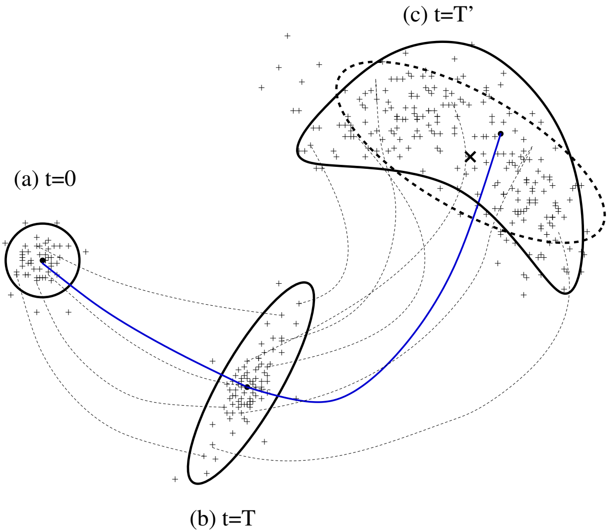
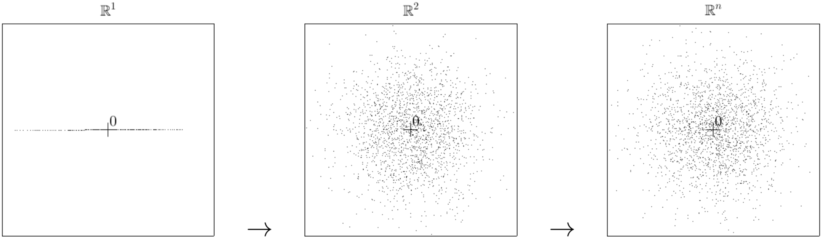

(c) t=T'
(a) t=0
(b) t=T

Table of contents

What is a Gaussian distribution $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$ in high dimension ?

$\mathbb{R}^1$ $\mathbb{R}^2$ $\mathbb{R}^n$, $n \gg 1$



here each point represents a sample of $\mathbf{x}$.

The answer to the previous question can be obtained while considering the distribution of the norm $|\mathbf{x}|$, given by

$$|\mathbf{x}| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{Tr(\mathbf{I})}, \frac{1}{2}\frac{Tr(\mathbf{I})}{d_{ef}(\mathbf{I})}\right), \tag{1}$$

where $Tr(\cdot)$ denotes de trace operator, and where

$$d_{ef}(\mathbf{I}) = \frac{Tr(\mathbf{I})^2}{Tr(\mathbf{I}^2)} = n, \tag{2}$$

is the effective dimension that provides a quantitative measure of the dimensionality. [Patil et al., 2001].

Said differently, the asymptotic approximation

$$|\mathbf{x}| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{n}, \frac{1}{2}\right), \tag{3}$$

tells that the distance between any sample $\mathbf{x}_k$ to zero is approximately constant, equal to $\sqrt{n}$, with a fluctuation of order $\frac{1}{\sqrt{2}} \approx 0.7$, independant of the dimension.

Concentration property of Gaussian law

For $\mathbf{x} \in \mathbb{R}^n$, when $n \gg 1$, any sample $\mathbf{x}_k$ of $\mathcal{N}(0, \mathbf{I})$ is within an hyper-sphere of radius $\sqrt{n}$, with fluctuation ± 0.7.

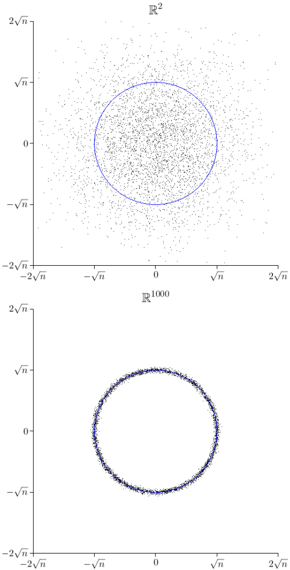

$\mathbb{R}^2$
$\mathbb{R}^{100}$
$\mathbb{R}^{1000}$
$\mathbb{R}^{10000}$
$2\sqrt{n}$
$\sqrt{n}$
0
$-\sqrt{n}$
$-2\sqrt{n}$

Table of contents

What happends for correlated Gaussian law ?
If we consider the distribution $\varepsilon^b \sim \mathcal{N}(0, \mathbf{B})$ that corresponds in data assimilation to the statistical model for the background error under Gaussian assumption. The asymptotic distribution of the norm is given by

$$|\varepsilon^b| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{Tr(\mathbf{B})}, \frac{1}{2}\frac{Tr(\mathbf{B})}{d_{ef}(\mathbf{B})}\right). \quad (4)$$

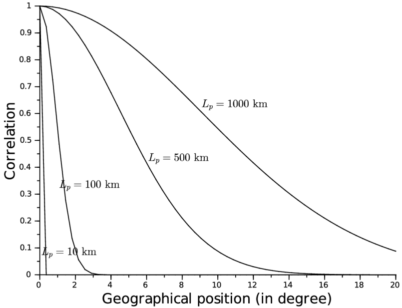
In practice, samples of ε^b can be obtained as the transformation $\varepsilon^b = \mathbf{B}^{1/2}\zeta$ of samples of random vector following a normal law $\zeta \sim \mathcal{N}(0, \mathbf{I})$. One needs to compute $\mathbf{B}^{1/2}$.

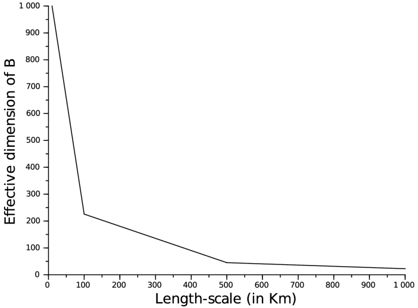
Note that in the continuous limit, where vectors tends to functions, the covariance matrix tends to a covariance operator of finite trace
There is no continous limit for the iid random vector $\mathcal{N}(0, \mathbf{I})$.

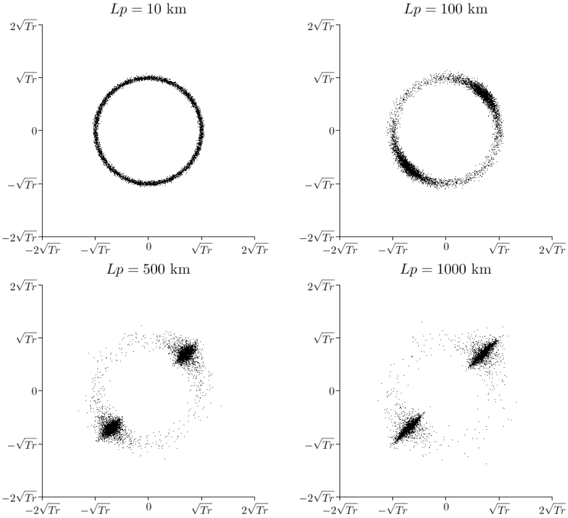
For experimentations over a 1D great circle on Earth with $n = 1000$ points, the square-root covariance matrix is $\left(\mathbf{B}^{1/2}\right)_{ij} = \rho_{1/2}(x_i - x_j)$ with

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

where L_p is the correlation length-scale.







$n = 1000$ with $N_e = 6400$ samples

Table of contents

Analysis step

The general framework formalism is, without any assumption of linearity of dynamics or Gaussianity of statistics, reduced to the Baye's rule

$$p(\mathbf{x}_q/\mathbf{y^o}_q) \propto p(\mathbf{y^o}_q/\mathbf{x}_q)p(\mathbf{x}_q), \tag{5}$$

where

- $p(\mathbf{x}_q)$ is the (density of) probability distribution to find the true state $\mathbf{x^t}_q$ at time q in the vicinity of $\mathbf{x}_q$,
- $p(\mathbf{y^o}_q/\mathbf{x}_q)$ is the (density of) probability distribution to measure $\mathbf{y^o}_q$ at time q when the true state $\mathbf{x^t}_q$ is known,
- $p(\mathbf{x}_q/\mathbf{y^o}_q)$ is the posterior distribution.

Table of contents

The update of the prior distribution

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

by using the observational distribution

$$p(\mathbf{y}^o{}_q/\mathbf{x}_q) \propto \exp\left(-\frac{1}{2}||\mathbf{y}^o{}_q - \mathbf{H}_q\mathbf{x}_q||^2_{\mathbf{R}_q^{-1}}\right), \quad (7)$$

is the posterior distribution (analysis distribution)

$$p(\mathbf{x}_q/\mathbf{y}^o{}_q) \propto \exp\left(-\frac{1}{2}||\mathbf{x}_q - \mathbf{x}^a{}_q||^2_{\mathbf{A}_q^{-1}}\right), \quad (8)$$

where

$$\left\{\begin{array}{l} \mathbf{x}^a{}_q = \mathbf{x}^b{}_q + \mathbf{K}_q(\mathbf{y}^o{}_q - \mathbf{H}_q\mathbf{x}^b{}_q), \\ \mathbf{A}_q = (\mathbf{I} - \mathbf{K}_q\mathbf{H}_q)\mathbf{B}_q, \\ \mathbf{K}_q = \mathbf{B}_q\mathbf{H}_q^T(\mathbf{H}_q\mathbf{B}_q\mathbf{H}_q^T + \mathbf{R}_q)^{-1}. \end{array}\right. \quad (9)$$

$$p(\mathbf{x}_q/\mathbf{y}^o{}_q) \propto \exp\left(-\frac{1}{2}||\mathbf{x}_q - \mathbf{x}^a{}_q||^2_{\mathbf{A}_q^{-1}}\right),$$

where

$$\begin{cases} \mathbf{x}^a{}_q = \mathbf{x}^b{}_q + \mathbf{K}_q(\mathbf{y}^o{}_q - \mathbf{H}_q\mathbf{x}^b{}_q), \\ \mathbf{A}_q = (\mathbf{I} - \mathbf{K}_q\mathbf{H}_q)\mathbf{B}_q, \\ \mathbf{K}_q = \mathbf{B}_q\mathbf{H}_q^T(\mathbf{H}_q\mathbf{B}_q\mathbf{H}_q^T + \mathbf{R}_q)^{-1}. \end{cases}$$

For the particular Gaussian framework:
the EnKF and the PF provide the same exact posterior distribution.

For the numerical computation: only one per three grid points are observed.

Table of contents

Under Gaussian assumption:
the EnKF and the PF should deliver the same background distribution in the limit of very large sample.

The discretization of the distribution means there exists an ensemble of background samples ${\mathbf{x}^b}_{q,k} \sim \mathcal{N}(0, \mathbf{B}_q)$ meaning that

$$|{\mathbf{x}^b}_{q,k} - {\mathbf{x}^b}_q| \underset{n \gg 1}{\sim} \mathcal{N}\left(\sqrt{Tr(\mathbf{B}_q)}, \frac{1}{2}\frac{Tr\mathbf{B}}{d_{ef}(\mathbf{B}_q)}\right) \quad (10)$$

Any sample ${\mathbf{x}^b}_{q,k}$ is within the hyper-sphere of radius $\sqrt{Tr(\mathbf{B}_q)}$ (up to a small fluctuation)

Numerical validation of the asymptotic distribution (6400 samples, $n = 1000$)

$$\left(|\mathbf{x}^b{}_{q,k} - \mathbf{x}^b{}_q| - \sqrt{Tr(\mathbf{B}_q)}\right) / \sqrt{\frac{1}{2}\frac{Tr(\mathbf{B}_q)}{d_{ef}(\mathbf{B}_q)}} \underset{n \gg 1}{\sim} \mathcal{N}(0,1)$$

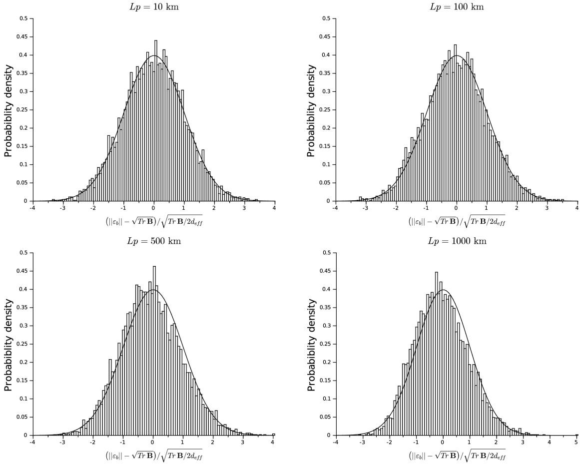


Table of contents

From the EnKF equations, an ensemble of analysis samples $\mathbf{x}^a{}_{q,k} = \mathbf{x}^a{}_q + \varepsilon^a{}_{q,k}$ can be constructed from $\mathbf{x}^a{}_q = \mathbf{x}^b{}_q + \mathbf{K}_q(\mathbf{y}^o{}_q - \mathbf{H}_q\mathbf{x}^b{}_q)$ by using the "perturbation of observation" method

$$\varepsilon^a{}_{q,k} = \varepsilon^b{}_{q,k} + \mathbf{K}(\varepsilon^o{}_{q,k} - \mathbf{H}\varepsilon^b{}_{q,k}), \tag{11}$$

with the theoretical distribution $\varepsilon^a{}_{q,k} \sim \mathcal{N}(0, \mathbf{A}_q)$, hence

$$|\mathbf{x}^a{}_{q,k} - \mathbf{x}^a{}_q| \underset{n\gg 1}{\sim} \mathcal{N}\left(\sqrt{Tr(\mathbf{A}_q)}, \frac{1}{2}\frac{Tr(\mathbf{A}_q)}{d_{ef}(\mathbf{A}_q)}\right). \tag{12}$$

Numerical validation of the asymptotic distribution (from EnKF eq., 6400 samples, $n = 1000$) $\left(|\mathbf{x}^a{}_q - \mathbf{x}^a{}_{q,k}| - \sqrt{Tr(\mathbf{A}_q)}\right) / \sqrt{\frac{1}{2}\frac{Tr(\mathbf{A}_q)}{d_{ef}(\mathbf{A}_q)}} \underset{n \gg 1}{\sim} \mathcal{N}(0,1)$

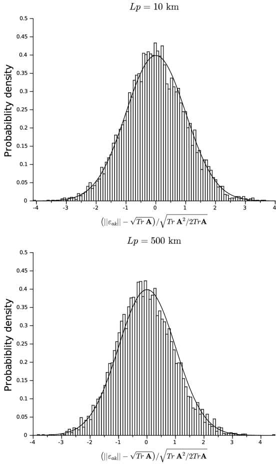


Table of contents

The analysis step in the PF [Gordon et al., 1993, Doucet et al., 2001, Del Moral, 2004, van Leeuwen, 2003] follows a quite different way as it directly relies on the Baye's rule:
for each member ${\mathbf{x}^b}_{q,k}$ one compute the weight

$$w_{q,k} = \frac{p({\mathbf{y}^o}_q/{\mathbf{x}^b}_{q,k})}{\sum_m p({\mathbf{y}^o}_q/{\mathbf{x}^b}_{qm})} \tag{13}$$

from which the *a posteriori* distribution is deduced as

$$P^e_q(\mathbf{x}_q/{\mathbf{y}^o}_q) = \sum_k w_{q,k}\delta(\mathbf{x}_q - {\mathbf{x}^b}_{q,k}). \tag{14}$$

after resampling, analysis sample are members chosen within the background ensemble samples
(this is a shortcut but .. [Snyder, 2011])
When does a background sample can be considered as a possible analysis sample candidate ?

Table of contents

Minimal/maximum sample for Gaussian law

If X is a random variable following a Gaussian law $\mathcal{N}(m, \sigma^2)$ of mean m and variance σ^2, then a classical result is that the maximum distance $|X - m|$ that can be reached for a sampling ensemble of size N_e is given by [Mallat, 1999]

$$T = \sigma\sqrt{2\log N_e}, \tag{15}$$

and the maximum (minimum) sample value is $m + T$ $(m - T)$.
(see also Appendix A in [Pannekoucke et al., 2014])

The maximal bound for the distance $|\mathbf{x}^a{}_{qk} - \mathbf{x}^a{}_q|$ is then

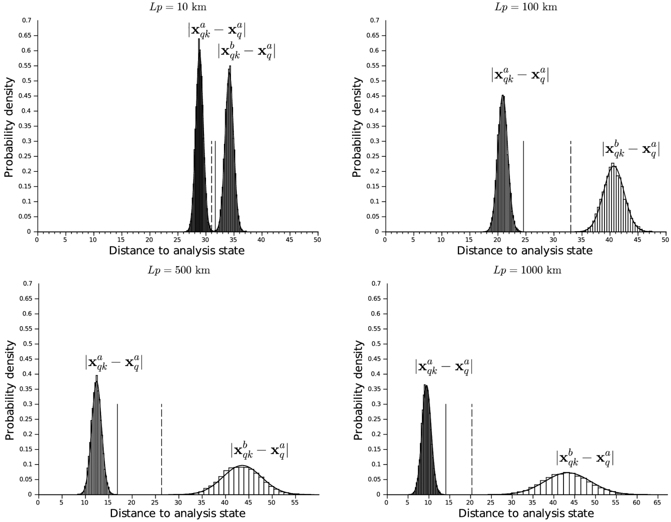
$$d^+_{a_q,a_{qk}} = \underset{k\in[1,N]}{\mathrm{Max}}\ |\mathbf{x}^a{}_{qk} - \mathbf{x}^a{}_q| = \sqrt{Tr(\mathbf{A}_q)} + \sqrt{\frac{1}{2}\frac{Tr(\mathbf{A}_q)}{d_{ef}(\mathbf{A}_q)}}\sqrt{2\log N_e}. \quad (16)$$

and the minimal bound for $|\mathbf{x}^b{}_{qk} - \mathbf{x}^a{}_q|$ is

$$d^-_{a_q,b_{qk}}(\delta\mathbf{x}^a{}_q) = \underset{k\in[1,N]}{\mathrm{Min}}\ |\mathbf{x}^b{}_{qk} - \mathbf{x}^a{}_q| = \mu_{a_q,b_{qk}} - \sigma_{a_q,b_{qk}}\sqrt{2\log N_e} \quad (17)$$

with

$$\left\{\begin{array}{lcl} |\mathbf{x}^b{}_{qk} - \mathbf{x}^a{}_q| & \underset{n\to\infty}{\sim} & \mathcal{N}(\mu_{a_q,b_{qk}}, \sigma^2_{a_q,b_{qk}}), \\ \mu_{a_q,b_{qk}} & = & \sqrt{|\delta\mathbf{x}^a{}_q|^2 + Tr(\mathbf{B}_q)}, \\ \sigma^2_{a_q,b_{qk}} & = & \frac{1}{2}\frac{2||\delta\mathbf{x}^a{}_q||^2_{\mathbf{B}_q} + Tr(\mathbf{B}_q^2)}{|\delta\mathbf{x}^a{}_q|^2 + Tr(\mathbf{B}_q)}. \end{array}\right. \quad (18)$$



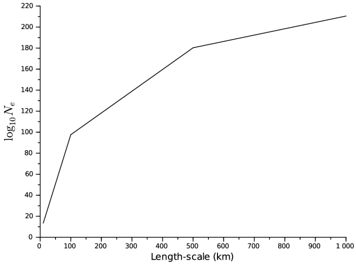
$n = 1000$ with $N_e = 6400$ samples, unsing EnKF Eq.

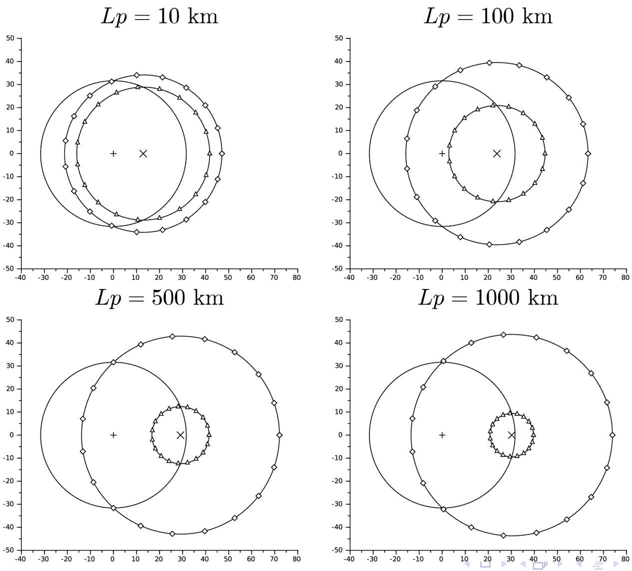
It results that:
there exists a background sample $\mathbf{x}^b{}_{qk}$ that can be considered as an analysis sample when

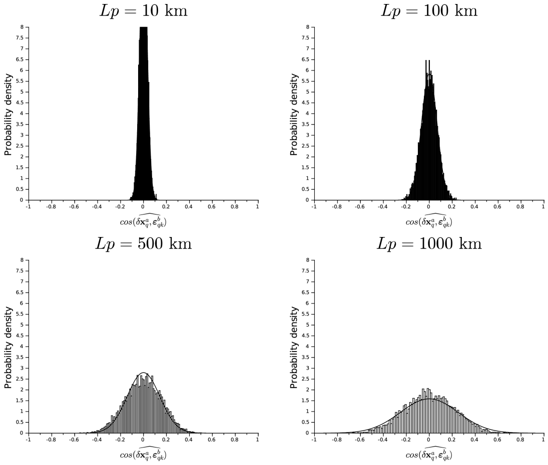
$$d^-_{a_q,b_{qk}}(\delta\mathbf{x}^a{}_q) \leq |\mathbf{x}^b{}_{qk} - \mathbf{x}^a{}_q| \leq d^+_{a_q,a_{qk}}. \quad (19)$$

This implies in average that the minimal ensemble size required to verify Eq.(19) is

$$N_e^- = \exp\left[2\frac{d_{ef}(\mathbf{A})}{Tr(\mathbf{A})}\left(Tr(\mathbf{B}) - \sqrt{Tr(\mathbf{B})^2 - Tr(\mathbf{KHB})^2}\right)\right]. \quad (20)$$




$Lp = 10$ km
$Lp = 100$ km
$Lp = 500$ km
$Lp = 1000$ km

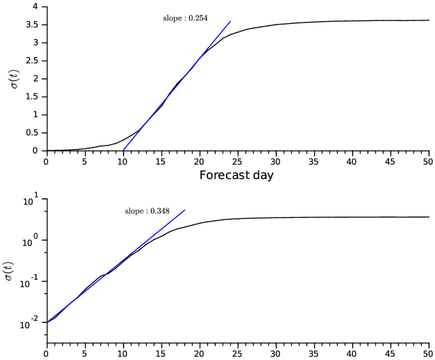


$$\cos\left(\widehat{\delta \mathbf{x}^a{}_q, \varepsilon^b{}_{q,k}}\right) \underset{n\gg 1}{\sim} \mathcal{N}\left(0, \frac{||\delta\mathbf{x}^a{}_q||^2_{\mathbf{B}}}{Tr(\mathbf{KHB})\,Tr(\mathbf{B})}\right). \qquad (21)$$

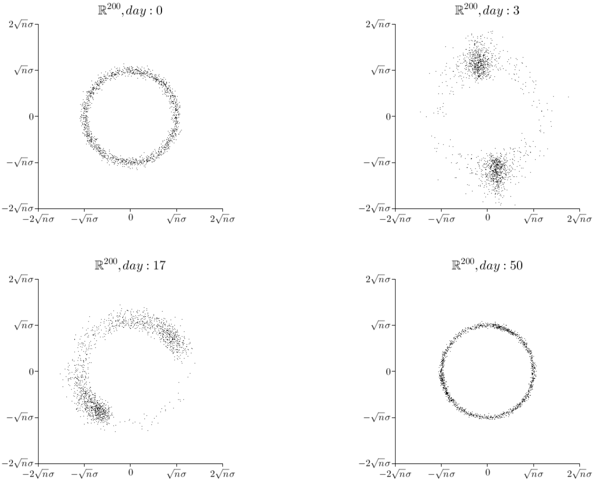
Non-linear dynamics in the Lorenz'96, $n = 200$

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \quad (22)$$

Illustration of time evolution of the averaged standard deviation:



Non-linear dynamics in the Lorenz'96, $n = 200$



- Gaussian samples lie within an hyper-sphere as the dimension increases:
  the intuition of the law dimension doesn't resists to the high dimension
- Under Gaussian framework, many asymptotic formula can be given, and numerically verified, for the norm of random perturbation:
  the distance plays the role of a diagnostic tool telling if a sample can be compatible, or not, with a given probability distribution,
  all or nothing!
- The difference between EnKF and PF, in the high dimension and under Gaussian assumption, is instructive:
  For non exponential ensemble size, background samples are not compatible with analysis samples
- These points are guidelines for the general non-linear framework (from numerical experiments ..)

Del Moral, P. (2004).
Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.
Springer.

Doucet, A., de Freita, N., Gordon, N., and Smith, A. (2001).
Sequential Monte Carlo Methods in Practice.
Springer.

Gordon, N., Salmond, D., and Smith, A. (1993).
Novel approach to nonlinear/non-gaussian bayesian state estimation.
Radar Signal Processing, IEE Proceedings F, 140(2):107–113.

Mallat, S. (1999).
A Wavelet Tour of Signal Processing.
Academic Press.

Pannekoucke, O., Raynaud, L., and Farge, M. (2014).
A wavelet-based filtering of ensemble background-error variances.
Quarterly Journal Royal Meteorological Society, 140:316–327.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J., and Ott, E. (2001).
Local low dimensionality of atmospheric dynamics.
Phys. Rev. Lett., 86:5878–5881.

Snyder, C. (2011).
Particle filters, the “optimal” proposal and high-dimensional systems.
In ECMWF, editor, *Proc. ECMWF Seminar on Data assimilation for atmosphere and ocean,*, pages 1–10, Reading, UK. ECMWF, ECMWF.

van Leeuwen, P. J. (2003).
A variance-minimizing filter for large-scale applications.
Monthly Weather Review, 131(9):2071–2084.