ENSTA ParisTech université PARIS-SACLAY

THE UNIVERSITY of EDINBURGH

Graduation Project (PFE)/ Master 2 Internship

Field of Study: Applied Mathematics
Scholar Year: 2015-2016

Stochastic filtering with reduced-order forward models

Confidentiality Notice

Non-confidential report and publishable on the Internet

Author:
Emilien Flayac

ENSTA ParisTech Tutors:
Frédéric Jean

Promotion:
2016

Host Organism Tutor:
Michal Branicki

Internship from April 4th 2016 to September 30th 2016

Name of the host organism: University of Edinburgh
Address: Department of Mathematics
The King's Buildings
Peter Guthrie Tait Road Edinburgh
EH9 3FD
United Kingdom

Confidentiality Notice

This present document is not confidential. It can be communicated outside in paper format or distributed in electronic format.

Abstract

Estimation of chaotic non-linear dynamical systems embedded in highly dimensional vector space lead to huge computational cost and large model errors if treated naively. The main objective of the present work is to present filters run in low dimension subspaces determined by different reduction techniques with different forward model during the predictive phase. Moreover, the underlying dynamics is not always known. The results show that efficient reductions are possible. The quality of the filter logically depends on the amount of information known about the "truth dynamics"

Acknowledgment

I want to thank Frédéric Jean for having accepted to be my supervisor at ENSTA Paristech. I want to thank Michal Branicki for having accepted to be my supervisor at the University of Edinburgh and for having advised me during my internship..

Contents

Introduction

Estimating known or partially known highly dimensional non-linear chaotic potentially stochastic dynamical systems is a big challenge of modern applied mathematics. Indeed, these systems appear in several difficult physical problems like in weather forecasting fully depicted in[1] or in fluid mechanics with the Navier-Stokes equations. Sometimes one even gets a stream of observations from a real system without any knowledge about the existence or not a an underlying dynamical system. The main issues in these systems are numerous. First, even if the dynamics is known, the different time and space scales of these systems makes it difficult to solve precisely so their system of equations are often approximated by simplest linear or non-linear models. Of course, it is compulsory if nothing is known about a potential dynamics. These models create an error with respect to the truth called model error. The second issue is the influence of chaotic behaviours on model errors. By definition, a chaotic behaviour means that two solutions of the same chaotic system starting from two different initial conditions will diverge to each other in finite time. Another similar property of chaotic systems is that if you solve the system from one initial condition and that you run a model that approaches well the system for short times starting from the same point, then the two solutions will very often diverge from each other. This points out one property of chaotic systems, they are hard to modelize. Therefore, in order to estimate them properly, the best method that has been found is to use on-line observations even partial ones inside a filter. The last issue concerning the considered systems is their high number of variables (from 10^4 to 10^8 in weather forecasting for examples). The real system can also be of infinite dimension. This high dimension can make even the simplest filter tuned with the simplest forward model too costly to run. Then, the last part of the process is to find a relevant subspace of sufficiently low dimension on which to project the model. Thus, the main goal of this work is to estimate the state of one parlicular very interesting dynamical system: Lorenz'96 by filtering it with different forward models, different bases on which to project the models and different techniques to keep only the best vectors from the basis to reduce the dimension of the filter.

Part I

Filtering problem in discrete time

The structure of this part is highly inspired from [2]

I.1 Basic notions and concepts on probability theory

Here we present the notations and the basic notions used in the subsequent sections to describe the filtering problem in our particular case.

We will use the probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P})$, with $n \in \mathbb{N}$, where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-algebra on $\mathbb{R}^n$ and $\mathbb{P}$ is a probability measure on $\mathcal{B}(\mathbb{R}^n)$. Then, in the following, we always assume that any random variable Z is a measurable function and that it admits a probability density function (pdf) with respect to the Lebesgue measure denoted p_Z. This means that, for any $A \in \mathcal{B}(\mathbb{R}^n)$:

$$\mathbb{P}(Z^{-1}(A)) = \int_A p_Z(z)\mathrm{d}z$$

Where $p_Z : \mathbb{R}^n \to \mathbb{R}^+$ satisfies:

$$\int_{\mathbb{R}^n} p_Z(z)\mathrm{d}z = 1.$$

We also assume that all the random variables encountered are square integrable meaning that, for a random variable X in $\mathbb{R}^n$:

$$E[X^T X] < +\infty.$$

Moreover, if X and Z are two random variables, their joint probability density will be denoted $p_{X,Z}$ and the conditional pdf of X knowing Z, denoted $p_{X|Z}$ satisfies :

$$\begin{aligned} &\forall (x,z) \in (\mathbb{R}^n)^2,\ p_{X,Z}(x,z) = p_{X|Z}(x,z)p_Z(z), \\ &\text{and if } p_Z(z) \neq 0,\ \text{then } p_{X|Z}(x,z) = \frac{p_{X,Z}(x,z)}{p_Z(z)}. \end{aligned} \tag{I.1}$$

From (I.1), one can show the Bayes formula :

$\forall (x,z) \in (\mathbb{R}^n)^2$, and if $p_Z(z) \neq 0$,

$$p_{X|Z}(x,z) = \frac{p_{X,Z}(x,z)}{p_Z(z)}.$$

Now if one reverses the role of X and Z in (I.1), we get :

$$p_{X,Z}(x,z) = p_{Z|X}(z,x)p_X(x).$$

One can deduce from this the Bayes formula

$$p_{X|Z}(x,z) = \frac{p_X(x)}{p_Z(z)} p_{Z|X}(z,x). \tag{I.2}$$

It is also possible to link p_X and p_X to their joint pdf $p_{X,Z}$ by :

$$\forall (x,z) \in (\mathbb{R}^n)^2, p_X(x) = \int_{\mathbb{R}^n} p_{X,Z}(x,z)\mathrm{d}z, \tag{I.3}$$
$$\text{and} \quad p_Z(z) = \int_{\mathbb{R}^n} p_{X,Z}(x,z)\mathrm{d}x.$$

I.2 Statement of the filtering problem in discrete time

The filtering problem can be described as : given a set of observations $y = (y_i)_{i\in[\![1,N]\!]}$ with $N \in \mathbb{N}$ in $\mathbb{R}^m$ that are the realization of the random variables $(Y_i)_{i\in[\![1,N]\!]}$. $(Y_i)_{i\in[\![1,N]\!]}$ are obtained from the solution $X = (X_i)_{i\in[\![0,N]\!]}$ of a discrete dynamical system in $\mathbb{R}^n$ with $m \leq n$, and given the forward operator associated to this dynamical system, what is the best estimator of X_i knowing $(Y_1, \cdots, Y_i)$, $\forall i \in [\![1,N]\!]$?

In this problem, the estimator of X_i must depend only on the observations up to time i and the observations might be of lower dimension than the truth i.e. the available information about the state come from the present and the past but never from the future and it might be incomplete. What's more, we need to define the sens of "best estimator of X_i".

I.2.1 The forward dynamics

More formally, X which is called the truth, is defined as a random process in $\mathbb{R}^n$ which satisfies the following recursive stochastic equation :

$$X_i = \Phi(X_{i-1}) + \xi_i \ , \forall i \in [\![1,N]\!]. \tag{I.4}$$

Where $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is called forward operator and $\forall i \in [\![1,N]\!]$, ξ_i is a random variable independent of X_{i-1} whose pdf p_{ξ_i} is known. We also suppose that the pdf of X_0, p_{X_0}, is known.

One can also consider the deterministic case where $\forall i \in [\![1,N]\!]$, $\xi_i = 0$ meaning that :

$$X_i = \Phi(X_{i-1}), \ \forall i \in [\![1,N]\!] \tag{I.5}$$

In this case, the randomness of X is fully contained in X_0. We assume that Φ is bijective. It is not very restrictive for the following because Φ will the flow of a differential equation taken for a fixed duration therefore it will be naturally bijective. Then, the observations $(Y_i)_{i\in[\![1,N]\!]}$ are defined as a random process in $\mathbb{R}^m$ that satisfies:

$$Y_i = h(X_i) + \eta_i, \forall i \in [\![1,N]\!] \tag{I.6}$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$ is possibly non linear and $\forall i \in [\![1, N]\!]$, η_i is a random variable in $\mathbb{R}^m$ representing the noise on the observations. We assume that $\forall i \in [\![1, N]\!]$, η_i is independent of X_i and that its pdf p_{η_i} is known.

The last elements that need to be defined are the space in which we should look for a estimate and the criterion which allows us to choose the best estimator.

I.2.2 Minimum variance estimator

As, for $i \in [\![1, N]\!]$, $(Y_1, \cdots, Y_i)$ contains all the information available at time i, we will look for an estimate of X_i in $\mathbb{R}^n$ that is a function of $(Y_1, \cdots, Y_i)$. More precisely, if $\mathcal{G}_i$ is the σ-algebra generated by $(Y_1, \cdots, Y_i)$ then, the estimator $\widehat{X}_i$ of X_i will be $\mathcal{G}_i$-measurable.

Therefore, among all the $\mathcal{G}_i$-measurable random variables, we choose our estimate $\widehat{X}_i$ to be the one that minimizes the variance i.e. $\widehat{X}_i$ is the solution of :

$$\min_{Z \ \mathcal{G}_i\text{-meas}} E[\|X_i - Z\|^2]. \tag{I.7}$$

Definition I.2.1. *Conditional expectancy of square integrable random variables*
If X is a $\mathcal{B}(\mathbb{R}^n)$-measurable and square integrable random variable and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{B}(\mathbb{R}^n)$ then the conditional expectancy of X with respect to $\mathcal{G}$ is the only $\mathcal{G}$-measurable random variable Z satisfying :

$$\forall U \in L^2(\Omega, \mathbb{R}^n),\ \mathcal{G}\text{-measurable}, E[X^T U] = E[Z^T U]. \tag{I.8}$$

And it is denoted $E[X|\mathcal{G}]$

Proposition I.2.1. *$E[X|\mathcal{G}]$ is the solution of :*

$$\min_{Z \in L^2(\Omega, \mathbb{R}^n), \mathcal{G}\text{-meas}} E[\|X - Z\|^2] \tag{I.9}$$

Proof. We will show that (I.8) is equivalent to (I.9) and it will show the existence and uniqueness in I.2.1
Let X be a $\mathcal{B}(\mathbb{R}^n)$-measurable and square integrable random variable in $\mathbb{R}^n$. Let Z be a $\mathcal{G}$-measurable and square integrable random variable

Let's assume that Z satisfies (I.8).

For U $\mathcal{G}$-measurable and square integrable :

$$\begin{aligned} E[\|X - U\|^2] =& E[X^2 X] - 2E[X^2 U] + E[U^T U], \\ =& E[\|U - Z\|^2] + 2E[(Z - X)^T U] + E[X^T X] - E[U^T U]. \end{aligned} \tag{I.10}$$

By assumption, as U is $\mathcal{G}$-measurable

$$E[(Z - X)^T U] = 0.$$

Then,

$$E[\|X - U\|^2] = E[\|U - Z\|^2] + E[X^T X] - E[U^T U].$$

The right-hand side of the previous equality is minimal for $U = Z$. Thus Z minimizes the variance and satisfies (I.9).

Conversely, let's assume now that Z satisfies (I.9)

It is known that $(X, Y) \to E[X^T Y]$ is an inner product on the space of square integrable random variable so, by assumption, Z is the orthonormal projection of X on the space $\mathcal{G}$-measurable and square integrable random variables. Then we know that $X - Z$ is orthonormal (in the sense of the scalar product defined above) to all $\mathcal{G}$-measurable and square integrable random variables meaning that :

$$\begin{aligned} \forall U \ \mathcal{G}\text{-measurable and square integrable}, E[(X-Z)^T U] =& 0, \\ E[X^T U] =& E[Z^T U]. \end{aligned}$$

And Z satisfies (I.8) □

Actually, as all our random variables are square integrable, we know that :

$$\widehat{X}_i = E[X_i | \mathcal{G}_i].$$

As $\mathcal{G}_i$ is the is the σ-algebra generated by $(Y_1, \cdots, Y_i)$, we will denote it by:

$$\widehat{X}_i = E[X_i | Y_{1:i}]. \tag{I.11}$$

I.2.3 Filter equations

Thanks to (I.11), we have an expression of our estimator but it does not provide a way to actually compute it. The idea to do so is that, the filtering problem can be rewritten in a more general way with recursive equations on the conditional probability measure associated to the estimator. In our case, these equations only deal with probability densities by assumption. The starting point is that if $p_{X_i|Y_{1:i}}$ is the pdf of X_i conditionally to $(Y_1, \cdots, Y_i)$ wich is called the filtering density:

$$\widehat{X}_i = E[X_i | Y_{1:i}] = \int_{\mathbb{R}^n} x \, p_{X_i|Y_{1:i}}(x, Y_{1:i}) \mathrm{d}x.$$

Then, the whole problem is to express $p_{X_i|Y_{1:i}}$ as a function of $p_{X_{i-1}|Y_{1:i-1}}$, $\forall i \in [\![1, N]\!]$. This can be done in two phases. The first one is called the prediction phase. It consists in mapping the posterior density at time $i-1$ $p_{X_{i-1}|Y_{1:i-1}}$ to the prior density at time i $p_{X_i|Y_{1:i-1}}$ using our knowledge of the underlying dynamic :

For $x_i \in \mathbb{R}^n$

$$\begin{aligned} p_{X_i|y_{1:i-1}}(x_i, y_{1:i-1}) &= \int_{\mathbb{R}^n} p_{X_i, X_{i-1}|y_{1:i-1}}(x_i, x_{i-1}, y_{1:i-1}) \mathrm{d}x_{i-1}, && \text{from (I.3)} \\ &= \int_{\mathbb{R}^n} p_{X_i|X_{i-1}, Y_{1:i-1}}(x_i, x_{i-1}, y_{1:i-1}) p_{X_{i-1}|y_{1:i-1}}(x_{i-1}, y_{1:i-1}) \mathrm{d}x_{i-1}. && \text{from (I.1)} \end{aligned}$$

Thanks to (I.6), we can see that $Y_{1:i-1}$ does not contain more information than X_{i-1} so

$$\forall (x_i, x_{i-1}) \in (\mathbb{R}^n)^2, \ p_{X_i|X_{i-1}, Y_{1:i-1}}(x_i, x_{i-1}, y_{1:i-1}) = p_{X_i|X_{i-1}}(x_i, x_{i-1}).$$

And,

$$p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = \int_{\mathbb{R}^n} p_{X_i|X_{i-1}}(x_i, x_{i-1}) p_{X_{i-1}|Y_{1:i-1}}(x_{i-1}, y_{1:i-1}) \mathrm{d}x_{i-1}.$$

By (I.4), one has :

$$\xi_i = X_i - \Phi(X_{i-1}).$$

So,

$$p_{X_i|X_{i-1}}(x_i, x_{i-1}) = p_{\xi_i}(x_i - \Phi(x_{i-1})),\ \forall (x_i, x_{i-1}) \in (\mathbb{R}^n)^2.$$

Then, the prediction equation in terms of conditional pdfs is :

$$p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = \int_{\mathbb{R}^n} p_{\xi_i}(x_i - \Phi(x_{i-1})) p_{X_{i-1}|Y_{1:i-1}}(x_{i-1}, y_{1:i-1}) \mathrm{d}x_{i-1}$$

The second phase is called the analysis phase and consists in mapping $p_{X_i|Y_{1:i-1}}$ to $p_{X_i|Y_{1:i}}$ using the additional information given by Y_i. Therefore, for $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}^m$:

$$p_{X_i|Y_{1:i}}(x_i, y_{1:i-1}, y_i) = p_{X_i|y_{1:i-1}, Y_i}(x_i, y_{1:i-1}, y_i)$$

From the Bayes formula (I.2),

$$p_{X_i|Y_{1:i}}(x_i, Y_{1:i-1}, y_i) = \frac{p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1})}{p_{Y_i|Y_{1:i-1}}(y_i, y_{1:i-1})} p_{Y_i|Y_{1:i-1}, X_i}(y_i, y_{1:i-1}, x_i)$$

What's more, from (I.3)

$$p_{Y_i|Y_{1:i-1}}(y_i, y_{1:i-1}) = \int_{\mathbb{R}^n} p_{Y_i|Y_{1:i-1}, X_i}(y_i, y_{1:i-1}, x) p_{X_i|Y_{1:i-1}}(x, y_{1:i-1}) \mathrm{d}x$$

As X_i is a function of $Y_{1:i-1}$ according to (I.4)

$$p_{Y_i|Y_{1:i-1}, X_i}(y_i, y_{1:i-1}, x_i) = p_{Y_i|X_i}(y_i, x_i)$$

We also know from (I.6) that

$$\eta_i = X_i - h(Y_i)$$

Then,

$$p_{Y_i|X_i}(y_i, x_i) = p_{\eta_i}(y_i - h(x_i))$$

Finally, the analysis equation applied on Y_i can be written

$$p_{X_i|Y_{1:i}}(x_i, y_{1:i}) = \frac{p_{\eta_i}(y_i - h(x_i)) p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1})}{\int_{\mathbb{R}^n} p_{\eta_i}(y_i - h(x)) p_{X_i|Y_{1:i-1}}(x, y_{1:i-1}) \mathrm{d}x}$$

We assume in the following that the random variables $(\xi_i)_{\in[\![1,N]\!]}$ and $(\eta_i)_{\in[\![1,N]\!]}$ are independent and identically distributed (iid). Their common pdf will be denoted respectively by ρ_ξ and ρ_η. Thus, the filtering equations in the case of a stochastic dynamic based on iid stochastic processes can be written, for $i \in [\![1, N]\!]$, for $x_i \in \mathbb{R}^n$:

$$\text{(Prediction)} \quad p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = \int_{\mathbb{R}^n} p_\xi(x_i - \Phi(x_{i-1}))p_{X_{i-1}|Y_{1:i-1}}(x_{i-1}, y_{1:i-1})\mathrm{d}x_{i-1} \tag{I.12}$$

$$\text{(Analysis)} \quad p_{X_i|Y_{1:i}}(x_i, y_{1:i}) = \frac{p_\eta(y_i - h(x_i))p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1})}{\int_{\mathbb{R}^n} p_\eta(y_i - h(x))p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x} \tag{I.13}$$

In the deterministic case described by (I.5), the only change that occurs in the prediction equation is that ρ_ξ is replaced by δ which is defined by :

$$\forall x \in \mathbb{R}^n, \delta(x) = \begin{cases} 1 \text{ if } x = 0 \\ 0 \text{ otherwise} \end{cases}$$

Therefore, the prediction equation in the deterministic case is:

$$p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = \int_{\mathbb{R}^n} \delta(x_i - \Phi(x_{i-1}))p_{X_{i-1}|Y_{1:i-1}}(x_{i-1}, y_{1:i-1})\mathrm{d}x_{i-1}$$

as we supposed that Φ is bijective,

$$p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = p_{X_{i-1}|Y_{1:i-1}}(\Phi^{-1}(x_i), y_{1:i-1})$$

Finally the filtering equations in the deterministic case are :

$$\text{(Prediction)} \quad p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1}) = p_{X_{i-1}|Y_{1:i-1}}(\Phi^{-1}(x_i), y_{1:i-1}) \tag{I.14}$$

$$\text{(Analysis)} \quad p_{X_i|Y_{1:i}}(x_i, y_{1:i}) = \frac{p_\eta(y_i - h(x_i))p_{X_i|Y_{1:i-1}}(x_i, y_{1:i-1})}{\int_{\mathbb{R}^n} p_\eta(y_i - h(x))p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x} \tag{I.15}$$

I.3 The Kalman Filter

In this section, we will deal with the solution of the filtering equations in the case of gaussian pdfs, with a linear dynamic and linear observations. More precisely, we assume now that X_0 is a gaussian random variable with mean $\bar{X}_0$ and covariance matrix C_0. For $i \in [\![1, N]\!]$, ξ_i and η_i are also assumed to be gaussian with zero mean and respective covariance matrix Σ and Γ. Moreover, we supposed that Φ and h are linear meaning that :

$$\begin{aligned} \Phi(x) &= Ax, \ \forall x \in \mathbb{R}^n \\ h(x) &= Hx \end{aligned}$$

With $A \in M_n(\mathbb{R})$ invertible and $H \in M_{m,n}(\mathbb{R})$ with rank(H)=m.

From (I.4) or (I.5) and from the previous assumptions on the gaussianity of the pdfs and the linearity of the dynamic, for $i \in [\![1, N]\!]$, X_i is a linear combination of X_0 and $(Y_1, \cdots, Y_{i-1})$.As

X_0 and $(Y_1, \cdots, Y_{i-1})$ are independent gaussian random variables the pdf associated to X_i is gaussian and thanks to a property a the gaussian pdf, The conditional pdfs, $p_{X_i|Y_{1:i-1}}$ and $p_{X_i|Y_{1:i}}$ are gaussian.

Then for $(i,j) \in [\![1,N]\!]^2$, the mean and the covariance matrix of $p_{X_i|Y_{1:j}}$ will be respectively denoted $m_{i|j}$ and $C_{i|j}$

Now we will derive the equations in the stochastic case

I.3.1 Stochastic dynamics

The goal here is to express the filtering equations from (I.12) and (I.13) in terms of mean and covariance matrix.

We start with the prediction equation so we want to express $m_{i|i-1}$ in function of $m_{i-1|i-1}$ and $C_{i|i-1}$ in function of $C_{i-1|i-1}$. We know by assumption that, for $x \in \mathbb{R}^n$:

$$p_{X_i|Y_{1:i-1}}(x, y_{1:i-1}) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i|i-1}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - m_{i|i-1})^T C_{i|i-1}^{-1}(x - m_{i|i-1})\right)$$

And from the prediction equation (I.12)

$$p_{X_i|Y_{1:i-1}}(x, y_{1:i-1}) = \frac{1}{(2\pi)^n(\det(C_{i-1|i-1})\det(\Sigma))^{\frac{1}{2}}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right) \cdots \tag{I.16}$$

$$\cdots \exp\left(-\frac{1}{2}(x - Az)^T \Sigma^{-1}(x - Az)\right) \mathrm{d}z \tag{I.17}$$

By definition of the conditional expectancy,

$$m_{i|i-1} = \int_{\mathbb{R}^n} x p_{X_i|Y_{1:i-1}}(x, Y_{1:i-1}) \mathrm{d}x \tag{I.18}$$

From the previous equation,

$$m_{i|i-1} = \frac{1}{(2\pi)^n(\det(C_{i-1|i-1})\det(\Sigma))^{\frac{1}{2}}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right)$$
$$\int_{\mathbb{R}^n} x \exp\left(-\frac{1}{2}(x - Az)^T \Sigma^{-1}(x - Az)\right) \mathrm{d}x\mathrm{d}z$$

One can see that the interior integral is the expectancy of a gaussian random variable with mean Ay so, by linearity of the integral

$$m_{i|i-1} = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i-1|i-1}))^{\frac{1}{2}}} A \int_{\mathbb{R}^n} z \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right) \mathrm{d}z$$
$$m_{i|i-1} = A \int_{\mathbb{R}^n} z p_{X_{i-1}|Y_{1:i-1}}(z, y_{1:i-1}) \mathrm{d}z$$

With the same argument and by definition of $m_{i-1|i-1}$

$$m_{i|i-1} = Am_{i-1|i-1}. \tag{I.19}$$

By definition of the conditional covariance matrix,

$$\begin{aligned} C_{i|i-1} &= \int_{\mathbb{R}^n} (x - m_{i|i-1})(x - m_{i|i-1})^T p_{X_i|Y_{1:i-1}}(x, Y_{1:i-1})\mathrm{d}x, \\ &= \int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, Y_{1:i-1})\mathrm{d}x - \left(\int_{\mathbb{R}^n} xp_{X_i|Y_{1:i-1}}(x, Y_{1:i-1})\mathrm{d}x\right) m_{i|i-1}^T, \\ &\quad - m_{i|i-1}\left(\int_{\mathbb{R}^n} xp_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x\right)^T + \left(\int_{\mathbb{R}^n} p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x\right) m_{i|i-1}m_{i|i-1}^T \end{aligned}$$

From (I.18) and the definition of a pdf

$$C_{i|i-1} = \int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x - m_{i|i-1}m_{i|i-1}^T.$$

From (I.16)

$$\begin{aligned} \int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x = &\frac{1}{(2\pi)^n(\det(C_{i-1|i-1})\det(\Sigma))^{\frac{1}{2}}} \int_{\mathbb{R}^n} xx^T \exp\left(-\frac{1}{2}(x - Az)^T\Sigma^{-1}(x - Az)\right)\mathrm{d}x \\ &\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right)\mathrm{d}z \end{aligned}$$

With the change of variable $u = x - Az$,

$$\begin{aligned} \int_{\mathbb{R}^n} xx^T \exp\left(-\frac{1}{2}(x - Az)^T\Sigma^{-1}(x - Az)\right)\mathrm{d}x = &\int_{\mathbb{R}^n} uu^T \exp\left(-\frac{1}{2}u^T\Sigma^{-1}u\right)\mathrm{d}u \\ &+ \left(\int_{\mathbb{R}^n} u\exp\left(-\frac{1}{2}u^T\Sigma^{-1}u\right)\mathrm{d}u\right)(Az)^T \\ &+ Az\left(\int_{\mathbb{R}^n} u\exp\left(-\frac{1}{2}u^T\Sigma^{-1}u\right)\mathrm{d}u\right)^T \\ &+ (Az)(Az)^T \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}u^T\Sigma^{-1}u\right)\mathrm{d}u \end{aligned}$$

Moreover, thanks to the definition of expression of the mean and covariance of gaussian random variable,

$$\begin{aligned} \int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x = &\frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i-1|i-1}))^{\frac{1}{2}}} \int_{\mathbb{R}^n} (\Sigma + Azz^T A^T \times \cdots \\ &\cdots \times \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right)\mathrm{d}z \end{aligned}$$

With the same kind of change of variable and the same properties on gaussian pdfs as above

$$\int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x = \Sigma + AC_{i-1|i-1}A^T + (Am_{i-1|i-1})(Am_{i-1|i-1})^T$$

From (I.19)

$$\int_{\mathbb{R}^n} xx^T p_{X_i|Y_{1:i-1}}(x, y_{1:i-1})\mathrm{d}x = \Sigma + AC_{i-1|i-1}A^T + m_{i|i-1}{m_{i|i-1}}^T$$

Therefore,

$$C_{i|i-1} = AC_{i-1|i-1}A^T + \Sigma$$

Finally, the prediction equations in the case of gaussian pdfs and linear stochastic dynamic are

$$\begin{aligned} m_{i|i-1} &= Am_{i-1|i-1} \\ C_{i|i-1} &= AC_{i-1|i-1}A^T + \Sigma \end{aligned} \tag{I.20}$$

One can notice that if $C_{i-1|i-1}$ is symetric positive definite,$C_{i|i-1}$ is too.

The analysis equations consist in a relationships between $(m_{i|i-1}, C_{i|i-1})$ and $(m_{i|i}, C_{i|i})$. The equations given by (I.13) can also be explicited in the case of gaussian pdfs but similarly to what is done in [3] for the analysis step in the derivation of the Kalman filter, we will not deal with factor that are independent of x because there is a renormalization of the right-hand side.

This means that by definition, we have for $x \in \mathbb{R}^n$:

$$p_{X_i|Y_{1:i}}(x, y_{1:i}) \propto \exp\left(-\frac{1}{2}(x - m_{i|i})^T C_{i|i}^{-1}(x - m_{i|i})\right)$$

And from the analysis equation (I.13)

$$p_{X_i|Y_{1:i}}(x, y_{1:i}) \propto \exp\left(-\frac{1}{2}(Y_i - Hx)^T \Gamma^{-1}(Y_i - Hx)\right) \exp\left(-\frac{1}{2}(x - m_{i|i-1})^T C_{i|i-1}^{-1}(x - m_{i|i-1})\right)$$

To find the equations we only need to identify the linear and the quadratic terms in the previous equalities. Then, for $x \in \mathbb{R}^n$:

$$\begin{aligned} (x - m_{i|i})^T C_{i|i}^{-1}(x - m_{i|i}) &= x^T C_{i|i}^{-1} x - 2m_{i|i}^T C_{i|i}^{-1} x + m_{i|i}^T C_{i|i}^{-1} m_{i|i} \\ (x - m_{i|i-1})^T C_{i|i-1}^{-1}(x - m_{i|i-1}) &= x^T C_{i|i-1}^{-1} x - 2m_{i|i-1}^T C_{i|i-1}^{-1} x + m_{i|i-1}^T C_{i|i-1}^{-1} m_{i|i-1} \\ (y_i - Hx)^T \Gamma^{-1}(y_i - Hx) &= x^T H^T \Gamma^{-1} H x - 2Y_i^T \Gamma^{-1} H x + y_i^T \Gamma^{-1} y_i \end{aligned}$$

Thus, by identification,

$$C_{i|i}^{-1} = C_{i|i-1}^{-1} + H^T \Gamma^{-1} H \tag{I.21}$$

$$C_{i|i}^{-1} m_{i|i} = C_{i|i-1}^{-1} m_{i|i-1} + H^T \Gamma^{-1} Y_i \tag{I.22}$$

To go further, we need to use a lemma coming from linear algebra.

Lemma I.3.1. *For $A \in M_n(\mathbb{R})$, $U \in M_{n,m}(\mathbb{R})$, $C \in M_m(\mathbb{R})$ and $V \in M_{m,n}(\mathbb{R})$ If A and C are symetric positive definite then $A+UCV$ is invertible and*

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$

From (I.21), $C_{i|i}$ is symmetric positive definite as we know that $C_{i|i-1}$ is too. We can then apply Lemma I.3.1 and :

$$\begin{aligned}
C_{i|i} =& C_{i|i-1} - C_{i|i-1}H^T(\Gamma + HC_{i|i-1}H^T)^{-1}HC_{i|i-1} \\
=& \left(I_n - C_{i|i-1}H^T(\Gamma + C_{i|i-1}HC_{i|i-1}H^T)^{-1}H\right)C_{i|i-1} \\
=& \left(I_n - C_{i|i-1}H^T {S_i}^{-1}H\right)C_{i|i-1} \quad \text{with } S_i = \Gamma + C_{i|i-1}HC_{i|i-1}H^T \\
C_{i|i} =& \left(I_n - K_iH\right)C_{i|i-1} \quad \text{with } K_i = C_{i|i-1}H^T {S_i}^{-1}
\end{aligned} \tag{I.23}$$

By (I.22), we get :

$$m_{i|i} = C_{i|i}C_{i|i-1}^{-1}m_{i|i-1} + C_{i|i}H^T\Gamma^{-1}Y_i$$

From (I.23)

$$m_{i|i} = (I_n - K_iH)m_{i|i-1} + C_{i|i}H^T\Gamma^{-1}Y_i \tag{I.24}$$

From (I.21), we can also get,

$$C_{i|i}(C_{i|i-1}^{-1} + H^T\Gamma^{-1}H) = I_n$$

And

$$\begin{aligned}
C_{i|i}H^T\Gamma^{-1}H =& I_n - C_{i|i}C_{i|i-1}^{-1} \\
=& I_n - (I_n - K_iH) \\
=& K_iH
\end{aligned}$$

As rank(H)=m,

$$C_{i|i}H^T\Gamma^{-1} = K_i$$

Thus, combining (I.24) with the last equality, we get :

$$m_{i|i} = (I_n - K_iH)m_{i|i-1} + K_iHY_i$$

Finally, the analysis equations in the case of gaussian pdfs and linear stochastic dynamic are

$$\begin{aligned}
m_{i|i} =& (I_n - K_iH)m_{i|i-1} + K_iHY_i \\
C_{i|i} =& \left(I_n - K_iH\right)C_{i|i-1}
\end{aligned} \tag{I.25}$$

with $S_i = \Gamma + C_{i|i-1}HC_{i|i-1}H^T$ and $K_i = C_{i|i-1}H^T {S_i}^{-1}$.

(I.20) and (I.25) combined form the equation of the Kalman filter that allow us to compute the sought estimate in the case of a stochastic dynamic.

I.3.2 Deterministic dynamics

As the analysis phase is not affected by a change of dynamic (I.13) is identical to (I.15) so the analysis equations in terms of mean and covariance are the same in the deterministic case. However, the prediction equations will be slightly modified. From (I.14) we can get that, for $x \in \mathbb{R}^n$:

$$p_{X_i|Y_{1:i-1}}(x, y_{1:i-1}) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i|i-1}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(A^{-1}x - m_{i|i-1})^T C_{i|i-1}^{-1}(A^{-1}x - m_{i|i-1})\right)$$

As we did for the stochastic case :

$$m_{i|i-1} = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i-1|i-1}))^{\frac{1}{2}}} \int_{\mathbb{R}^n} x \exp\left(-\frac{1}{2}(A^{-1}x - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(A^{-1}x - m_{i-1|i-1})\right) \mathrm{d}x$$

With the change of variable $z = A^{-1}x$ and as det(A)=1 because A is invertible by assumption :

$$m_{i|i-1} = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i-1|i-1}))^{\frac{1}{2}}} A \int_{\mathbb{R}^n} z \exp\left(-\frac{1}{2}(z - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(z - m_{i-1|i-1})\right) \mathrm{d}z$$

$$m_{i|i-1} = A m_{i-1|i-1}$$

Similarly

$$C_{i|i-1} = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(C_{i-1|i-1}))^{\frac{1}{2}}} \int_{\mathbb{R}^n} x x^T \exp\left(-\frac{1}{2}(A^{-1}x - m_{i-1|i-1})^T C_{i-1|i-1}^{-1}(A^{-1}x - m_{i-1|i-1})\right) \mathrm{d}x$$

With the same change of variable we get

$$C_{i|i-1} = A C_{i-1|i-1} A^T$$

Finally, the prediction equation in the deterministic case are

$$\begin{aligned} m_{i|i-1} &= A m_{i-1|i-1} \\ C_{i|i-1} &= A C_{i-1|i-1} A^T \end{aligned} \tag{I.26}$$

The Kalman filter equations only apply in particular situations. It is clear from all the previous computations that without the assumption of linearity of the dynamic these equations do not stand. However, w will see in the section that, with non linear dynamics it is still possible to create a filter that will be look like Kalman filter.

I.4 The Approximate Gaussian Filters

Most of the complicated systems from which we want to estimate the state are non linear so it is not reasonable to keep the assumption of linearity of Φ. Therefore, if Φ is not linear, and even if the initial pdf and the pdfs of the noise, the conditional pdf of the state is no

longer gaussian. There is still an equation between $(m_{i|i-1}, C_{i|i-1})$ and $(m_{i-1|i-1}, C_{i-1|i-1})$ but it involves all the moments of the conditional pdf. Thus none of the previous apply.

The main idea of the approximate gaussian filter, as it is described in [3], is to assume that all the conditional pdfs stay gaussian and that we know a way to compute $(m_{i|i-1}, C_{i|i-1})$. From these new assumptions, the estimates $m_{i|i}$ is computed by minimizing the cost function g defined by, for $x \in \mathbb{R}^n$:

$$g(x) = (x - m_{i|i-1})^T C_{i|i-1}^{-1}(x - m_{i|i-1}) + (y_i - Hx)^T \Gamma^{-1}(y_i - Hx).$$

By taking the minimum of g, we find the same analysis equation of the mean that in the linear case that is :

$$m_{i|i} = (I_n - K_i H)m_{i|i-1} + K_i H y_i.$$

with $S_i = \Gamma + C_{i|i-1} H C_{i|i-1} H^T$ and $K_i = C_{i|i-1} H^T {S_i}^{-1}$. Then, we propagate the covariance matrix as in the linear case :

$$C_{i|i} = (I_n - K_i H) C_{i|i-1}.$$

The main issue of this filter is to determine $(m_{i|i-1}, C_{i|i-1})$. To do so, one can compute the differential equation satisfied by the mean and the covariance of a solution of the original dynamical system. In general, these equation might depend on all the moments of the probability measure associated to the solution so there is often no operator that maps $(m_{i-1|i-1}, C_{i-1|i-1})$ to $(m_{i|i-1}, C_{i|i-1})$.

Therefore, the fundamental idea is to look for an operator that maps $(m_{i-1|i-1}, C_{i-1|i-1})$ to $(m_{i|i-1}, C_{i|i-1})$ and that approximates the true dynamic. Such an operator is called a forward model and will be denoted Φ_{mod}. Hence the general equations of an approximate gaussian filter are :

$$\begin{aligned}(m_{i|i-1}, C_{i|i-1}) &= \Phi_{mod}(m_{i-1|i-1}, C_{i-1|i-1}) \\ m_{i|i} &= (I_n - K_i H)m_{i|i-1} + K_i H y_i \\ C_{i|i} &= (I_n - K_i H) C_{i|i-1}\end{aligned}$$

Of course the choice of a forward model is not unique and may change a lot the quality of the estimate.

I.5 The "truth" dynamics

When one gets observations from a real system, one often assumes, as in (I.6), that the observations are a noisy function of the solution of dynamical system. In other words, the real system is supposed to follow a particular dynamics but one can only partially measures the considered solution of the system that we call the truth. In the following, in order to simulate a real system we have to suppose that the observations come from a dynamical system which, in our case, is Lorenz'96 dynamic.The goal of this section is to present the reasons why this dynamics is well adapted to simulations. First, in order to present all the interesting properties of Lorenz'96 dynamic, we need to recall some results on dynamical systems and the notion of attractor.

I.5.1 Results on long time behaviour of dynamical systems

Let's consider a dynamical system in $\mathbb{R}^n$:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = g(u) \tag{I.27}$$

With $g : \mathbb{R}^n \to \mathbb{R}^n$

Definition I.5.1. *The dynamical system* (I.27) *has an absorbing bounded set* $\mathcal{B}_{abs}$ *iff* $\forall R > 0,\ \exists t(R)$ *such that:*

$$\forall t > t(R),\ \phi_t^g(B(0,R)) \subset \mathcal{B}_{abs}$$

Where ϕ_t^g *is the flow of* (I.27) *taken at time t and* $B(0,R)$ *is the ball of centre* 0 *and radius R.*

If the the dynamical system (I.27) has an absorbing bounded set we can define its global attractor $\mathcal{A}$ by

$$\mathcal{A} = \bigcap_{s\geq 0} \overline{\bigcup_{t\geq s} \phi_t^g(\mathcal{B}_{abs})}$$

Less formally, the global attractor of the system is the set in which all the solutions of the system will end after a certain time.

Property I.5.1. *If* $\exists(\alpha,\beta) \in (\mathbb{R}^{*+})^2$ *such that* $\forall u \in \mathbb{R}^n$, *:*

$$\langle g(v), v\rangle \leq \beta - \alpha(\|v\|^2)$$

where for $\langle\cdot,\cdot\rangle$ *is the euclidean scalar product of* $\mathbb{R}^n$ *Then the system* (I.27) *has a an absorbing bounded set and thus a global attractor.*

This property is called the absorbing ball property.

This property is important because it gives a computational way to show that a dynamical system admits a global attractor

I.5.2 Description of the underlying dynamic

We recall that the ambient space considered will be $\mathbb{R}^n$ with $n \in \mathbb{N}$.
Let $E = (e_1, e_2, \ldots, e_n)$ be the basis of $\mathbb{R}^n$ defined such that :

$$\forall (i,j) \in [\![1,n]\!]^2,\ e_i^{(j)} = \delta_{ij}$$

Where $e_i^{(j)}$ is the j^{th} component of e_i in $\mathbb{R}^n$ and δ_{ij} is the Kronecker symbol applied to i and j. Thus in the following, representing a dynamical system in the physical domain will correspond to representing it in the basis E. Then Lorenz 96' dynamic is defined in continuous time, in $\mathbb{R}^n$ by :

$$\frac{\mathrm{d}u^{(i)}}{\mathrm{d}t} = u^{(i-1)}(u^{(i+1)} - u^{(i-2)}) - u^{(i)} + F, \quad \forall i \in [\![1,n]\!] \tag{I.28}$$

Where $u^{(i)}$ represent the i^{th} component of u in the basis E and $F \in \mathbb{R}$ is the forcing parameter.

We set by definition in the equation $u^{(0)}{=}u^{(n)}$, $u^{(-1)}{=}u^{(n-1)}$ and $u^{(n+1)}{=}u^{(1)}$. It can be rewritten in a vectorial way by :

$$\frac{\mathrm{d}u}{\mathrm{d}t} = B(u,u) + Du + f \tag{I.29}$$

with

$$f = \begin{pmatrix} F \\ \vdots \\ F \end{pmatrix}_{n\times 1}$$

$$D = -I_n$$

and in the basis E,

$$\forall u, v \in \mathbb{R}^n, \quad B(u,v) = \frac{1}{2}\begin{pmatrix} u^{(n)}v^{(2)} + u^{(2)}v^{(n)} - u^{(n)}v^{(n-1)} - u^{(n-1)}v^{(n)} \\ u^{(1)}v^{(3)} + u^{(3)}v^{(1)} - u^{(1)}v^{(n)} - u^{(n)}v^{(1)} \\ \vdots \\ u^{(i-1)}v^{(i+1)} + u^{(i+1)}v^{(i-1)} - u^{(i-1)}v^{(i-2)} - u^{(i-2)}v^{(i-1)} \\ \vdots \\ u^{(n-1)}v^{(1)} + u^{(1)}v^{(n-1)} - u^{(n-1)}v^{(n-2)} - u^{(n-2)}v^{(n-1)} \end{pmatrix}_{n\times 1}$$

This system is deterministic so to be in the case described in (I.5), we assume that the initial condition of the system is a gaussian random variable. Thus it is possible to derive the equation of the mean and the covariance matrix satisfied by a solution. As u is defined with a random initial condition, one can define the mean $\bar{u}$ by

$$\bar{u} = \mathbb{E}[u]$$

Then one can define $Z = u - \bar{u}$.

Let's fix Z_i the i^{th} coordinate of Z in the basis E. Then,

$$\begin{aligned} u &= \bar{u} + Z \\ Z &= \sum_{i=1}^{n} Z_i e_i \end{aligned} \tag{I.30}$$

We can also define the covariance matrix R in the basis E by

$$\forall (i,j) \in [\![1,n]\!]^2, \; R_{ij} = \mathbb{E}[Z_i Z_j^*]$$

Where * stands for the conjugate. Then, the system of equations satisfied by $\bar{u}$ and R is :

$$\frac{\mathrm{d}\bar{u}}{\mathrm{d}t} = D\bar{u} + B(\bar{u},\bar{u}) + \sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij} B(e_i, e_j) + f \tag{I.31}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = L_v R + R L_v^{\dagger} + Q_F \tag{I.32}$$

Where † stands for the adjoint. L_v and Q_F are defined by :

$$\forall (i,j) \in [\![1,n]\!]^2,\ (L_v)_{ij} = (De_j + B(\bar{u}, e_j) + B(e_j, \bar{u})) \cdot e_i \tag{I.33}$$

$$(Q_F)_{ij} = \sum_{p=1}^{n}\sum_{q=1}^{n} (\mathbb{E}[Z_p Z_q Z_j] B(e_p, e_q) \cdot v_i + \mathbb{E}[Z_p Z_q Z_i] B(e_p, e_q) \cdot e_j) \tag{I.34}$$

where $\cdot$ stands for the euclidian scalar product. It is important to notice that Q_F depends on the third-order moments of Z that are non-zero because the pdf of u is non-gaussian in general. The simulation are computed by solving the system with a time step $\Delta t > 0$. Thus with the same notations as in the description of the filter equation is computed for $t = i\Delta t$, $i \in [\![1, N]\!]$. This dynamical system is a very well known "toy" model that has a lot of properties that we will use to test our different techniques.

It is also highly non-linear because of the presence of a quadratic term B. This term B conserves energy meaning that :

$$\forall u \in \mathbb{R}^n,\ B(u,u)^T u = 0 \tag{I.35}$$

Moreover, for a solution of (I.29) $u : \mathbb{R} \to \mathbb{R}^n$, let's define the function v_j by

$$\forall i \in [\![1,n]\!],\ v_j^{(i)} = u^{(\overline{i+j})}$$

With $\forall i \in \mathbb{N}$, $\overline{i} \equiv i[n]$.

Lorenz'96 dynamic is translation invariant meaning that for any solution u of (I.29), and $\forall j \in \mathbb{N}$, the corresponding v_j is still a solution of (I.29). This is due to the fact that, in (I.28), there are conditions on the first and the last components. Translation invariance will be important later because it gives sense in representing Lorenz'96 dynamic in the Fourier basis as described in [4].

What's more, Lorenz'96 dynamic has a global attractor and it has been proven that when the value of F increases the dimension of the attractor increases. This dimension can be seen in some way as the minimal dimension of the subspace on which to project without losing too many properties of the dynamics.

Nevertheless, the different forward models or techniques of reduction of dimension we used need different amount of information on the underlying dynamic. This means that the techniques we will present later are more or less efficient depending on how well you know your real system.

I.5.3 Observations

In the following, data will be available at two stages in the procedure. First, we consider that a set of snapshots is provided offline and is used to learn about the correct dynamics by computing the mean and the covariance matrix after a sufficiently long time. Then, observations coming from the same system are provided in real time during the filtering process. During the learning phase the observations are a collection of snapshots taken at time $i_1, \cdots, i_q$ with $q \in \mathbb{N}$ and $q \leq N$. So, they are denoted $y_{i_1}, \cdots, y_{i_q}$ From these snapshots one can compute the empirical mean and covariance matrix when the truth has reached the attractor i.e. after a sufficiently long time p:

$$\hat{u}_\infty = \frac{1}{n-q}\sum_{k=1}^{n} y_{i_k}$$
$$\hat{R}_\infty = \frac{1}{n-q}\sum_{k=p}^{n}(y_{i_k} - \hat{u}_\infty)(y_{i_k} - \hat{u}_\infty)^T$$

The main issues of the subsequent parts following are :

1. How to find the good modes that allow us to reduce the dimension of the system by projecting it only on a few of them.
2. Once the right basis has been chosen, how to select the modes to keep.
3. How to find an approximate forward operator projected on those modes that lead to good filtering results compared to the amount of data available about the dynamics

Part II

Description of the imperfect dynamics used in the forward models of filters

As it is said above, the several forward models we used needs different knowledge of the dynamics. The first one differs fundamentally from the two other ones because it needs complete knowledge of the true dynamics included the value of the parameter F. Besides, in all this part we consider a basis $\mathcal{V}$. of $\mathbb{R}^n$ (or potentially $\mathbb{C}^n$) and V its transition matrix represented in the basis E. Then, all the forward model are meant to be represented in the basis $\mathcal{V}$.

II.1 Recall on the singular value decomposition

Let's define $X \in M_{n,p}(\mathbb{R})$ with $p \geq n$ THe singular value decomposition is defined as follow:

$$X = U\Sigma V^T \tag{II.1}$$

Where Σ is diagonal with positive coefficients $n \times n$ matrix containing the singular values sorted in decreasing order U is a $n \times n$ orthonormal matrix, V a $n \times m$ orthonormal matrix. The coefficient of Σ are also the eigenvalues of $X^\dagger X$

II.2 Modified Quasilinear Gaussian approximation of Lorenz'96 dynamics

MQG model has been developped in [5] and [6]. The principle of the MQG model is to replace Q_F described in (I.34) by a matrix that only depends on the first and second order moments of u : $\bar{u}$ and R. This process is called a Gaussian closure.

II.2.1 Description of the model

The only data required by the model are the mean and the covariance matrix in the attractor, $\bar{u}_\infty$ and R_∞.

The first step to describe the new matrix Q_F as we want is to define its value when u has reached the attractor that we denote $Q_{F\infty}$. It is given by :

$$Q_{F\infty} = -(L_{v\infty}R_\infty + R_\infty L_{v\infty}^\dagger)$$

Where $L_{v\infty}$ is the value of L_v at $\bar{u}_\infty$

The second step is to decompose $Q_{F\infty}$ as the sum of a symmetric (or hermitian) positive semi-definite matrix $Q^+_{F\infty}$ of a symmetric (or hermitian) negative matrix semi-definite $Q^-_{F\infty}$ knowing that $Q_{F\infty}$ is symmetric (or hermitian).

To do so, one diagonalizes $Q_{F\infty}$ to get $P_\infty \in O_n(\mathbb{R})$ (or $P_\infty \in O_n(\mathbb{C})$) and D_∞ diagonal such that :

$$Q_{F\infty} = P_\infty^{-1} D_\infty P_\infty$$

Then one sets

$$\begin{aligned} Q^+_{F\infty} &= P_\infty^{-1} \max(D_\infty, 0) P_\infty \\ Q^-_{F\infty} &= P_\infty^{-1} \min(D_\infty, 0) P_\infty \end{aligned}$$

Where $\max(\cdot, 0)$ and $\min(\cdot, 0)$ applied on a matrix means that they are applied on each component of the matrix. The last definition is well defined because we know that the eigenvalues of Q_F are real.

The third step is to define the damping in the attractor and for any time. Its expression is :

$$N_\infty = \frac{1}{2}(Q^-_{F\infty} - qI_n)R_\infty^{-1} \text{ with } 0 < q << 1$$

so that

$$Q^-_{F\infty} = N_\infty R_\infty + R_\infty N_\infty^\dagger$$

Hence we can define the damping for any time N proportionally to N_∞ as follow :

$$N = \frac{f(R)}{f(R_\infty)} N_\infty$$

with $f(A) = \mathrm{tr}(A)^{\frac{1}{2}}$

Actually, the advised value of q is 0.1. Consequently, Q^-_F is defined by :

$$Q^-_F = NR + RN^\dagger$$

and Q^+_F is deduced as follow :

$$Q^+_F = -\frac{\mathrm{tr}(Q^-_F)}{\mathrm{tr}(Q^+_{F\infty})}(Q^+_{F\infty} + qI_n)$$

Finally, the approximate system of equations of mean and covariance matrix is :

$$\frac{\mathrm{d}\bar{u}}{\mathrm{d}t} = -\bar{u} + B(\bar{u}, \bar{u}) + \sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij} B(e_i, e_j) + f \tag{II.2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = L_v R + R L_v^\dagger + NR + RN^\dagger + Q^+_F \tag{II.3}$$

Therefore, the forward model associated to this dynamical system in the canonical basis E, denoted Φ^E_{MQG}, is the flow of the coupled equations (II.2) and (II.3) applied at time Δt.

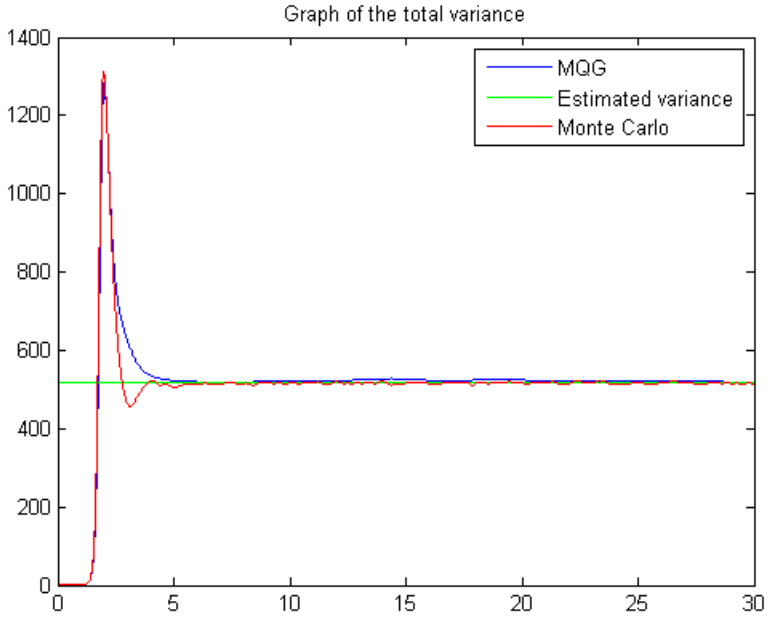
However later, one will want to project this system on another basis. If Φ^E_{MQG} is decomposed as follow, $\forall \bar{u} \in \mathbb{R}^n, \forall R \in M_n(\mathbb{R})$ symmetric positive definite :

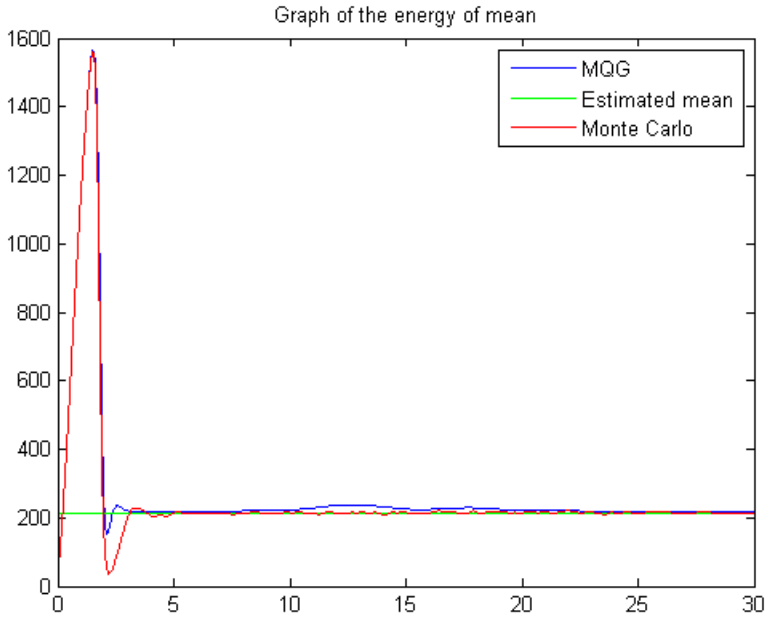
$$\Phi^E_{MQG}(\bar{u}, R) = (\Phi^E_{mean}(\bar{u}, R), \Phi^E_{cov}(\bar{u}, R)).$$

Then $\forall \bar{v} \in \mathbb{R}^n, \forall S \in M_n(\mathbb{R})$ symmetric positive definite

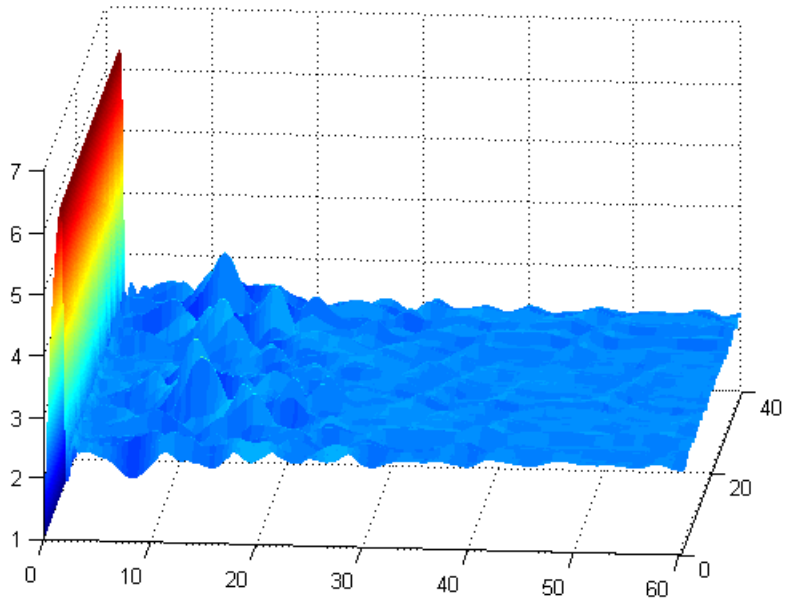
$$\Phi^V_{MQG}(\bar{v}, S) = (V^{-1}\Phi^E_{mean}(V\bar{v}, VSV^{-1}), V^{-1}\Phi^E_{cov}(V\bar{v}, VSV^{-1})V).$$

The following figures are the plots of the energy of the mean which is basically the square of the L_2-norm of $\bar{u}$ and of the trace of the covariance matrix called the total variance computed from Monte-Carlo simulations of Lorenz'96 dynamics, the MQG forward model as a function of time. The green lines correspond to the value for the empirical mean and covariance matrix : $\hat{u}_\infty$ and $\hat{R}_\infty$.

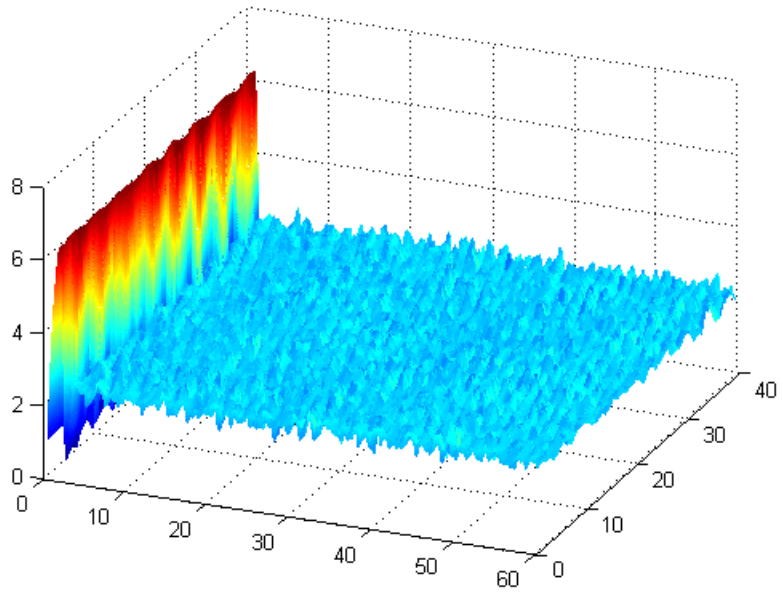




Here is a 3D representation of the mean computed by the MQG model.



Here is a 3D representation of the mean computed by Monte Carlo simulations.



II.3 Orstein-Ulenbeck processes

An Orstein-Ulenbeck process is a solution of the following linear stochastic differential equation in $\mathbb{R}$:

$$\mathrm{d}u_t = \gamma(\mu - u_t) + \sigma \mathrm{d}W_t. \tag{II.4}$$

where $\gamma > 0$, $\sigma > 0$, $\mu \in \mathbb{R}$ and $(W_t)_{t\geq 0}$ is a standard Wiener process. One can find an explicit expression for the solutions of (II.4). From this, the mean and variance of a solution of (II.4) $(\bar{u}_t, \bar{u}_t)_{t\geq 0}$ can also be explicitly computed :

$$\bar{u}_t = \mu(1 - \exp(-\gamma t)) + \exp(-\gamma t)\bar{u}_0, \tag{II.5}$$

$$r_t = \frac{\sigma^2}{2\gamma}(1 - \exp(-2\gamma t)) + \exp(-2\gamma t) r_0. \tag{II.6}$$

As the final goal is to obtain a forward model represented in the basis V, one considers that the coordinates of the prior estimate in the filter are n independent Orstein-Ulenbeck processes meaning that the covariance matrix is assumed diagonal in the basis $\mathcal{V}$.Thus the dynamical system used for the mean and covariance represented in the basis $\mathcal{V}$ is

$$\begin{aligned}
\bar{u}_{1t} &= \mu_1(1 - \exp(-\gamma_1 t)) + \exp(-\gamma_1 t)\bar{u}_{10}, \\
&\vdots \\
\bar{u}_{nt} &= \mu_n(1 - \exp(-\gamma_n t)) + \exp(-\gamma_n t)\bar{u}_{n0}, \\
r_{1t} &= \frac{\sigma_1^2}{2\gamma_1}(1 - \exp(-2\gamma_1 t)) + \exp(-2\gamma_1 t) r_{10}, \\
&\vdots \\
r_{nt} &= \frac{\sigma_n^2}{2\gamma_n}(1 - \exp(-2\gamma_n t)) + \exp(-2\gamma_n t) r_{n0}.
\end{aligned} \tag{II.7}$$

where $(\mu_i)_{i\in[\![1,n]\!]}$, $(\gamma_i)_{i\in[\![1,n]\!]}$ and $(\sigma_i)_{i\in[\![1,n]\!]}$ are parameters determined by $\hat{u}_\infty$ and $\hat{R}_\infty$ projected on $\mathcal{V}$ that are denoted $\hat{u}_\infty^V$ and $\hat{R}_\infty^V$. More precisely, for $i \in [\![1,n]\!]$, μ_i is the mean after a long time, γ_i is defined as the inverse of the correlation time of $(y_{i_1},\cdots,y_{i_q})$ projected on the basis $\mathcal{V}$: $T(y^{V(i)})$ with $y^V = V^{-1}y$. σ_i is related to $\hat{R}_\infty^V$ and γ_i. More formally,

$$\begin{aligned}\mu_i =& \hat{u}_\infty^{V(i)},\\ \gamma_i =& \frac{1}{T(y^{V(i)})},\\ \sigma_i =& \sqrt{2\gamma_i \hat{R}^V_{\infty ii}}.\end{aligned}$$

Finally, the sought forward model that is denoted Φ_{OU}^V is the flow of (II.7) applied at time Δt.

II.4 An equation free model : The dynamic mode decomposition (DMD)

The DMD method fully described in [7] and [8] does not use an equation defined in advance but assumes that there is an linear underlying dynamics and compute it hanks to snapshots. Actually, let's define Y_-^V the concatenation of all the snapshots projected on the basis $\mathcal{V}$ except the last one and Y_+^V the concatenation of all the snapshots projected on the basis $\mathcal{V}$ except the first one. Thus,

$$\begin{aligned}Y_-^V =& (V^{-1}y_{i_1},\cdots,V^{-1}y_{i_{q-1}}) = (y^V_{i_1},\cdots,y^V_{i_{q-1}})\\ Y_+^V =& (V^{-1}y_{i_2},\cdots,V^{-1}y_{i_q}) = (y^V_{i_2},\cdots,y^V_{i_q})\end{aligned}$$

The fundamental assumption is that $\exists A \in M_n(\mathbb{R})$ such that :

$$\forall k \in [\![1,q-1]\!],\ y^V_{i_{k+1}} = A y^V_{i_k}$$

Thus,

$$Y_+^V = AY_-^V$$

This method can be used if the dimension of the snapshots is n because A must be of the dimension of the state.. Then,

$$A \approx Y_+^V \widetilde{Y_-^V}$$

Where $\widetilde{Y_-^V}$ is the Moore-Penrose pseudo-inverse of Y_-^V. Then, using the singular value decomposition defined in (II.1) Y_-^V, one gets:

$$A \approx Y_-^V W \Sigma^{-1} U$$

Moreover, one projects A on the modes of U and gets :

$$C = U^\dagger Y_-^V W \Sigma^{-1}$$

Besides, one decomposes C with its eigenvalues Λ and its eigenvectors contained in Z. Then,

$$CZ = \Lambda Z.$$

Actually A and C have the same eigenvalues and there exists a relationship between the eigenvectors of A denoted Ψ and the eigenvectors of C, Z

$$\Psi = Y_-^V W \Sigma^{-1} Z$$

Therefore, as for any linear dynamical system, the eigenvector and eigenvalues of A allow us to find the forward model for the mean.Indeed, one can notice that the DMD method only provides a way to predict the mean of the dynamics thus the filter using the DMD method will keep the covariance matrix constant during the prediction phase : $\forall \bar{u} \in \mathbb{R}^n, \forall R \in M_n(\mathbb{R})$ symmetric positive definite :

$$\Phi_{DMD}^E(\bar{u}, R) = (\Psi \exp(\Lambda \Delta t) \widetilde{\Psi} \bar{u}, R)$$

Then Φ_{DMD}^V is found the same way as for the MQG model. In principle this model can be projected on any basis but the natural one is defined by he column vectors of Ψ and this one was used in the filter based on DMD method.

Part III

Choice of a basis and signal compression

III.1 Determination of relevant modes

III.1.1 Knowing properties of the correct dynamic

Let's first set M as the matrix representing the Discrete Fourier Transform from the canonical basis of $\mathbb{C}^n$ to the canonical basis of $\mathbb{C}^n$. M is defined by :

$$\forall (j,k) \in [\![1,n]\!]^2, M_{jk} = \frac{1}{\sqrt{n}} e^{-\frac{2i\pi}{n}(j-1)(k-1)}$$

We know that Lorenz'96 dynamics is translation invariant. From [4], translation invariant dynamical systems have a covariance matrix which is diagonal when represented in the Fourier basis. Thus, it is clearly a good idea to represent of the forward model in the Fourier and use the OU forward model for example. Indeed, with that model one assumes that the covariance matrix is diagonal in the chosen basis (the Fourier basis here) but as we know that the covariance matrix in the attractor is diagonal the loss of non-diagonal coefficients is minimal.

III.1.2 With no information of the correct dynamic

In the last section, one has chosen the Fourier basis denoted $\mathcal{F}$ because of some a priori knowledge about the "truth" dynamics but if one does not know anything about the "truth" dynamic there exists a basis computed thanks to a SVD on snapshots that uses proper orthonormal decomposition. To compute the POD modes, let's consider some of the snapshots $(y_{i_1}, \cdots, y_{i_r})$ with $r \leq q$represented in the basis E concatenated in a matrix Y. We suppose that $q > n$ to match the frame of (II.1).

To compute the POD modes, one has to compute the singular value decomposition of Y :

$$Y = U_{POD} \Sigma W^T$$

Where Σ is diagonal $n \times n$ matrix containing the singular values sorted in decreasing order U_{POD} is a $n \times n$ matrix whose columns in the basis V form the basis sought and W a $n \times m$ matrix that will not be used here. The modes computed in the DMD method are also modes that do not need any information on the underlying dynamics.

III.2 Dimension reduction of data

The main idea here is to find a criterion to select the best modes to keep among those of the chosen basis. Ideally, the modes that are kept preserves the behaviour of the system. In other words, the modes that are removed have a low impact

III.2.1 Energy or singular value criterion

The diagonal of $\hat{R}_\infty^M$ gives values for each modes of the basis $\mathcal{F}$. Thus, the reduction technique is to keep the modes with the biggest values. Something similar can be done with the POD and the DMD modes meaning that in both case values are also associated to the modes, by the singular values for the POD modes and the matrix Λ for the DMD modes.

III.2.2 Sparse representation of the observations

The idea is to find a sparse vector of coordinates in the chosen basis that approaches the snapshots in an optimal way. Then, one keeps the modes corresponding to non-zero coordinates in the sparse vector.

First, let's consider the snapshots not used to compute the POD modes i.e. $(y_{i_r+1}, \cdots, y_{i_q})$

Using several L_2 minimization problems with a L_1 constraint

The first idea to find a sparse representation of $(y_{i_r+1}, \cdots, y_{i_q})$ in chosen basis called again $\mathcal{V}$ to denote either the Fourier basis, the POD basis or the DMD basis, is to solve the following optimization problem for $k \in [\![r+1, q]\!]$:

$$(P_k): \quad \min_{a_k \in \mathbb{R}^n} \quad \|y_{i_k} - V a_k\|_2^2 \qquad \text{s.t.} \quad \|a_k\|_1 \leq t$$

With $t > 0$.

The solution of each problem (P_k), a_k is a sparse approximation of the vector of coordinates of y_{i_k} in the basis $\mathcal{V}$ because of the constraint on the L_1-norm. From all the vectors a_k, one can define a criterion vector c as follow :

$$\forall j \in [\![1, n]\!],\ c^{(j)} = \sum_{k=r+1}^{q} \mid a_k^{(j)} \mid$$

Using one L_2 minimization problem with a L_1 constraint

The major issue of the previous method is that each solution of (P_k) gives a sparse representation of one snapshot and these solutions may not be related because (P_k) does not take into account the dynamic during a whole interval but only at one time. Then, the position of the non-zero coefficients can differ from a solution to another and the criterion vector may not rank the modes properly.

Another idea is to look for one sparse vector a that suits all the snapshots . This can be done by minimizing the sum of the errors between a snapshot and its sparse representation with a constraint on the L_1-norm of the solution so that the number of non-zero components is minimal. Consequently, it can be written :

$$(P): \quad \min_{a\in\mathbb{R}^n} \quad \sum_{k=1}^{n} \|y_{i_k} - Va\|_2^2$$
$$\text{s.t.} \quad \|a\|_1 \leq t$$

Where $t \geq 0$ is a parameter that allows us to control the number of modes we want to keep by controlling the number of non-zero coefficients that appears in the solution. Actually, there is no theoretical result on the link between t and the number of non-zero coefficients of the solution. However, it seems numerically that the trend is the larger t is the more non-zero coefficients there will be in the solution. Therefore, in order to have a criterion to select the modes. One can solve several the problem (P) with increasing value of t. Then, the modes to keep are the ones that correspond to non-zeros coefficients and the lowest value of t. Indeed, the modes that appear with a low value of t are the modes that contribute first to the the L_1-norm of a and can reasonably be considered as the best modes.

Part IV

Results

IV.1 Measures of the distance between the truth and the estimate

IV.2 Summary of the procedure

1. Collecting data (simulation of the truth)
2. Selection of the basis in which the filter will be run
3. Selection of the best modes among the chosen basis
4. Selection of the forward model that will be used during the prediction phase of the filter + plus computation of the corresponding parameters projected on the reduced order set of modes.
5. Computation of the RMS error and pattern correlations

IV.3 Results of each filter

Let's assume that the whole procedure described above is done. Thus, the modes on which to project the forward models are chosen. Let's call V_{red} the concatenation of all the chosen modes i.e. $V_{red} \in M_{n,m}(\mathbb{R})$ (or $M_{n,m}(\mathbb{C})$). Let's define $(X_i)_{i\in[\![1,N]\!]}$ the truth and $(m_{i|i})_{i\in[\![1,N]\!]}$ the sequence of posterior mean of a filter. In order to measure the efficiency of the filters there exists two main instruments : the root mean square (RMS) error and the pattern correlation. To be able to compare correctly all the estimates and observations, they are projected back in the basis E so the reduced order estimates $(m_{i|i})_{i\in[\![1,N]\!]}$ of dimension m are projected back in the basis E and are denoted ${(m_{i|i})^E}_{i\in[\![1,N]\!]}$ One can stress the fact that the transition regime is not important in our case as the forward models have been designed with snapshots after a certain time. Then, the RMS error between the truth and the estimated state is defined by :

$$\mathrm{RMS}(m^E) = \frac{1}{N-p+1}\sum_{i=p}^{N} \|X_i - m^E_{i|i}\|_2$$

with $p \leq N$ The pattern correlation of one component j can be define by :

$$\text{patterncorr}_j(m) = \frac{\sum_{i=p}^{N}\left(X_i^{(j)} - \frac{1}{N-p+1}\sum_{l=p}^{N} X_l^{(j)}\right)\left(m_{i|i}^{(j)} - \frac{1}{N-p+1}\sum_{l=p}^{N} m_{l|l}^{E(j)}\right)}{\sqrt{\sum_{i=p}^{N}\left(X_i^{(j)} - \frac{1}{N-p+1}\sum_{l=p}^{N} X_l^{(j)}\right)^2}\sqrt{\sum_{i=p}^{N}\left(m_{i|i}^{E(j)} - \frac{1}{N-p+1}\sum_{l=p}^{N} m_{l|l}^{E(j)}\right)^2}}$$

One can also compute the projection of the new observations used in the filter on the subspace described by V_{red}. We will denote them $z^{V_{red}} = (z_1^{V_{red}}, \cdots, z_N^{V_{red}})$. Similarly to what is is with the estimates the observations are projected back in full dimension by the operation $V_{red} * z^{V_{red}}$

The most basic element to check is that :

$$\text{RMS}(m^E) \leq \text{RMS}(V_{red} * z^{V_{red}})$$

And

$$\text{patterncorr}_\text{j}(\text{m}^\text{E}) \leq \text{patterncorr}_j(V_{red} * z^{V_{red}})$$

This means that the estimates of the filter are better than the observations with the same dimension. It appears that every combination of forward model, basis and reduction techniques this inequality is satisfied.

Another inequality that is one of the main objective of the procedure is : The most basic element to check is that :

$$\text{RMS}(m^E) \leq \text{RMS}(z)$$

And

$$\text{pattercorr}_\text{j}(\text{m}^\text{E}) \leq \text{patterncorr}_j(z)$$

If this inequality is satisfied for one combination of techniques for one particular number of modes kept m, it means that the reduced order estimate is better than the observation with full dimension alone.

It also appears that using the Fourier is always better than using the POD modes but, as we said before, this is a logical result because the Fourier are naturally adapted to the systems. What has to be noticed is that the results of the POD modes are a bit worse than the results of the Fourier modes but still very close meaning that without any information about the true dynamics the final results would still be good.

Concerning the reduction techniques, the sparse reduction techniques don't seem to improve the reduction with respect to the "classical" reductions techniques based on energy and singular values.

Concerning the forward models, the DMD forward model with the DMD modes are quite precise in full dimension but reacts badly to the reduction of dimension meaning that their RMS increases a lot when the dimension of the estimates decreases. The reduced order OU forward model gives better results than the reduced order MQG forward model. Actually, the latter was not tested before. It was partially described in [6] with only a reduction on the covariance. As described in the section concerning MQG model we projected both the mean and the covariance matrix and as a matter of fact we reduced both by using the matrix V_{red}.

Conclusion

As a conclusion, one can say that the idea of reduced order estimates projected on a relevant basis with respect to the dynamics is promising because it proposes a solution to the curse of dimensionality and to the lack of information about the original system. More test can be run with different frequency of measurement, partial observations or other forward models. Actually, the OU forward model is restricted to stationary system whereas more complicated models as SPEKF which is a sort of extension of OU processes where the parameters follow a SDE are harder to implement but more efficient.

Bibliography

[1] Andrew J Majda, John Harlim, and Boris Gershgorin. Mathematical strategies for filtering turbulent dynamical systems. *Discrete Contin. Dyn. Syst*, 27(2):441–486, 2010.

[2] Brian DO Anderson and John B Moore. *Optimal filtering*. Courier Corporation, 2012.

[3] Andrew Stuart and Kostas Zygalakis. Data assimilation: A mathematical introduction. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2015.

[4] Philip Holmes, John L Lumley, and Gal Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry*.

[5] Themistoklis P Sapsis and Andrew J Majda. A statistically accurate modified quasilinear gaussian closure for uncertainty quantification in turbulent dynamical systems. *Physica D: Nonlinear Phenomena*, 252:34–45, 2013.

[6] Themistoklis P Sapsis and Andrew J Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, 110(34):13705–13710, 2013.

[7] JL Proctor, SL Brunton, BW Brunton, and JN Kutz. Exploiting sparsity and equation-free architectures in complex systems. *The European Physical Journal Special Topics*, 223(13):2665–2684, 2014.

[8] Jacob Grosek and J Nathan Kutz. Dynamic mode decomposition for real-time background/foreground separation in video. *arXiv preprint arXiv:1404.7592*, 2014.

[9] Steven L Brunton, Jonathan H Tu, Ido Bright, and J Nathan Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1716–1732, 2014.

[10] Boris Gershgorin, John Harlim, and Andrew J Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *Journal of Computational Physics*, 229(1):1–31, 2010.

[11] Boris Gershgorin and Andrew J Majda. Filtering a statistically exactly solvable test model for turbulent tracers from partial observations. *Journal of Computational Physics*, 230(4):1602–1638, 2011.

[12] Ross A Maller, Gernot Müller, and Alex Szimayer. Ornstein–uhlenbeck processes and extensions. In *Handbook of Financial Time Series*, pages 421–437. Springer, 2009.

Appendix

[5] [9] [3] [1] [4] [10] [11] [12] [6] [7] [8] [2]