Principal components analysis in the frequency domain versus proper orthogonal decomposition for multiscale thermal fields

Sylvie Viguier-Pla[1&2], Stéphane Abide[1]
[1]*LAMPS - Université de Perpignan via Domitia, F-66860 Perpignan Cedex 9, France,*
viguier@univ-perp.fr, stephane.abide@univ-perp.fr
[2]*IMT, Université Paul Sabatier, F-31062 Toulouse Cedex 4, France*

Abstract. Nowadays direct numerical simulation of fluid system is known to provide a large amount of data that allows fine analysis of turbulent flows. We propose to use principal components analysis in the frequency domain to reduce data arising from simulations and to compare the results with that of proper orthogonal decomposition, namely principal components analysis. We compare the quality of reconstruction, and the specific features of data each analysis reveals.

Keywords. Direct numerical simulation, principal components analysis, time series, stationarity, random measure, spectral analysis

Résumé. La simulation directe numérique dans un système de mécanique des fluides est connue pour prodiguer beaucoup de données qui permettent une étude fine des phénomènes de turbulence. Nous proposons d'utiliser l'analyse en composantes principales dans le domaine des fréquences pour réduire ces données provenant de simulations, et de comparer les résultats avec ceux de l'analyse en décomposition propre, c'est-à dire l'analyse en composantes principales standard adaptée aux données de signal. Nous comparons les qualités de reconstitution, et les faits particuliers que chaque analyse révèle.

Mots-clés. Simulation numérique directe, analyse en composantes principales, séries chronologiques, stationnarité, mesure aléatoire, analyse spectrale

1 Introduction

Simulation of complex systems, such as fluid mechanics, lead to the production of a large amount of information. Their reduction is then of major importance to be able to carry out fine analyses of the underlying physical phenomena. In this context, several approaches have been developed from the last decade. One can mention the pioneer works concerning Proper Orthogonal Decomposition (POD) (Lumley, 1970), based on the Principal Components Analysis (PCA), or more recently the Dynamic Mode Decomposition (DMD), developed by Schmid (2010). All of them look for an efficient way to project data in order to facilitate analysis. Keeping in mind this purpose, PCA is devoted to the reduction of dimension, mainly in a context of independent observations. When data are from signal measurements, the independence is no more ensured in the time domain, but when the signal is a stationary process, its Fourier transformation gives independent observations. We propose to perform a method of PCA in the frequency domain, the FPCA, based on this property. This method, which has been first introduced by Brillinger (2001), especially for periodic data, extracts frequency information from the signals. We use an adaptation of this method for non-necessarily periodic series.

In the current studies, the POD has been largely investigated and extended (see for example Antoranz et al., 2018).

In this presentation, we first present the data of interest. Secondly, we present the POD method. Then we expose the FPCA method for non-periodic series and when frequencies are not known, and show how this method is numerically implemented, emphasing the encountered problems.

We apply the two methods on data from simulation. We end by showing the difficulties of each method, we compare the qualities of reconstitution and the phenomenon each method reveals at each step.

2 Description of the data of interest

In this work, POD and FCPA analyses are performed using data from Direct Numerical Simulation (DNS) of a flow in a baroclinic tank. A baroclinic tank allows to reproduce some features of atmospheric flows at a laboratory scale. It consists of an annular cavity filled with water with the two lateral walls with a fixed temperature such that an horizontal difference temperature is enforced. The whole cavity rotates at a constant rotation rate. The reader can find the detail of the configuration flow in Rodda et al. (2020). In this exploratory work, the three dimensional equations governing the fluid flow and the temperature field are solved numerically using high performance computing. For the sake of conciseness, the method is not reported here, however all details are presented in Abide (2017, 2018, 2020). For the PCA, we consider the output of our DNS-code which is a time series of temperature fields as shown in Fig. 1. We denote the temperature time series by $T = \{T_{ij}^n = T(t^n, r_i, \varphi_j)\}$ which corresponds to the temperature variable at the time step n on a point of (r_i, φ_j). A typical output is given Fig. 1 which represents the temperature at mid-radius.

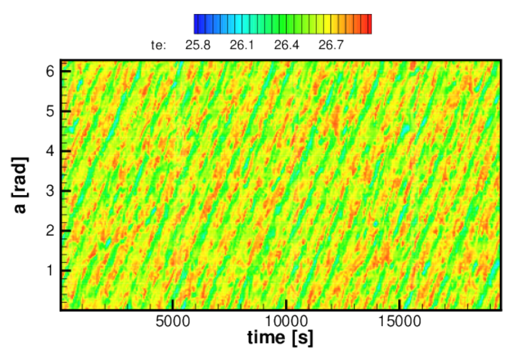

Figure 1: Time series of the temperature at the free surface of the baroclinic tank

Here according to the discretization parameters, the time series has 360 snapshots of temperature fields on 512 points (Figure 2). For the times $t = 1, \ldots, 360$ we extract these fields for 64 points along a circle into the cylinder. So our data is a matrix of 360 rows and 64 columns.

3 The Proper Orthogonal Decomposition

Let $\{u(x,t)\}$ be a stochastic process defined on $\mathbb{R}^n \times \mathbb{R}$, as for example a random $n-$dimensional field observed along the time. The POD is the search of a deterministic function $\phi(x)$ that best approximates the stochastic function in average.

Practically, it consists in considering a sample $(x_1, \ldots, x_n)$ of space points, and measures at times $t_1, \ldots, t_p$. The method is implemented via the principal components analysis of the matrix $U = (u(x_j, t_i))_{j=1,\ldots,n; i=1,\ldots,p}$.

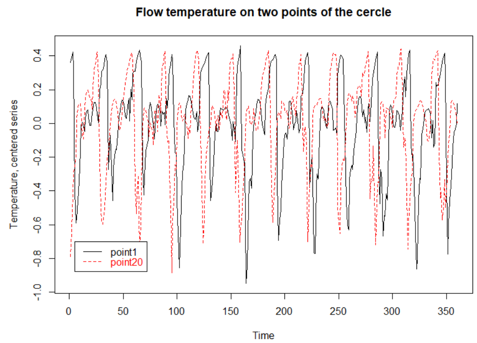

Figure 2: Initial data, points 1 and 20

.

4 Principal Components Analysis in the Frequency domain

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary $p-$dimensional random time series. The FPCA of $(X_n)_{n\in\mathbb{Z}}$ is the search of a $q-$dimensional series $(q<p)$ $(X'_n)_{n\in\mathbb{Z}}$, stationarily correlated with $(X_n)_{n\in\mathbb{Z}}$, as close as possible to it. As $(X_n)_{n\in\mathbb{Z}}$ and $(X'_n)_{n\in\mathbb{Z}}$ are stationary, there exist two unitary operators U and U' such that $X_n = U^n X_0$ et $X'_n = U'^n X'_0$. So the FPCA is the search of X'_0 and U' such that $X'_n = U'^n X'_0$ and $\|X_0 - X'_0\|$ is as small as possible.

The X_n's of the stationary series $(X_n)_{n\in\mathbb{Z}}$, are $p-$dimensional random vectors: $X_n = (x_n^1, \ldots, x_n^p)^t$. The stationarity is assumed in a broad sense, that is $\mathbb{E}(X_n\, {}^t\overline{X_m}) = \mathbb{E}(X_{n-m}\, {}^t\overline{X_0})$ for any pair (n,m) of elements from $\mathbb{Z}$. It is equivalent with the usual second order stationarity of each of its components $(x_n^i)_{n\in\mathbb{Z}}$ and with the pairwise correlated stationarity: $\mathbb{E}(x_n^i \overline{x}_m^j) = \mathbb{E}(x_{n-m}^i \overline{x}_0^j)$ for any (n,m,i,j) from $\mathbb{Z}\times\mathbb{Z}\times\{1,...,p\}\times\{1,...,p\}$.

This method is issued from the work of Boudou (1995). An example is implemented in Boudou, Caumont and Viguier-Pla (2004).

We define a $p-$dimensional random measure ($p-$r.m.) Z as a vector measure on $\mathcal{B}$, $\sigma-$field of $[-\pi,\, \pi[$, taking values in $L^2_{\mathbb{C}^p}(\mathcal{A})$, such that $E(Z\,A^t\overline{Z\,B}) = 0$, for any pair (A,B) of disjoint elements from $\mathcal{B}$.

This definition of the r.m. Z lets us define a random measure (r.m.) $\mu_Z(.) = \|Z(.)\|^2$, and the stochastic integral as the isometry

$$\begin{array}{rcl} L^2([-\pi,\,\pi[,\mathcal{B},\mu_Z) & \longrightarrow & \overline{\mathrm{vect}}\{ZA; A\in\mathcal{B}\} \\ 1_A & \mapsto & ZA \\ \varphi & \mapsto & \int \varphi \mathrm{d}Z. \end{array}$$

Under certain conditions, we define the image of (X_n) from a filter α (element of $L^2([-\pi,\,\pi[,\mathcal{B},\mu_Z))$ as a moving average.

If Z is a p-r.m. and μ the Lebesgue measure, then the family $\{C_m X_{n-m}\,;\, m\in\mathbb{Z}\}$, where $C_m = (2\pi)^{-1}(\int e^{i\,.\,m}\alpha(.)\,d\mu(.))$, is summable of sum $\int e^{i.n}\alpha\, dZ$.

We assume that the conditions are satisfied for the existence of the spectral density,

$$(2\,\pi)^{-1}\sum_{n\in\mathbb{Z}} e^{-i.n}\mathbb{E}\, X_n^t\overline{X_0}.$$

Theoretically, the FPCA needs to process the PCA of $E(Z\{\lambda\}^t\overline{Z\{\lambda\}}) = (2\,\pi)^{-1}\sum_{n\in\mathbb{Z}} e^{-i\lambda n}\mathbb{E}\, X_n^t\overline{X_0}$, for each λ from $[-\pi,\,\pi[$, this means an infinity of PCA's.

We overcome this difficulty by a discretization of the spectrum.

More precisely, if k is an integer, we consider the measurable application from $[-\pi,\, \pi[$ into itself:

$$f_k = \sum_{l=-k}^{k-1} \frac{\pi\, l}{k}\, 1_{B_{lk}}$$

where $B_{-k,k} = \{-\pi\}$, $B_{lk} = [\frac{\pi\, l}{k} - \frac{\pi}{k}, \frac{\pi\, l}{k}[$ for $l = -k+1, \ldots, -1$, $B_{0k} =]-\frac{\pi}{k}, \frac{\pi}{k}[$, and $B_{lk} = [\frac{\pi\, l}{k}, \frac{\pi\, l}{k} + \frac{\pi}{k}[$ for $l = 1, \ldots, k-1$.

The FPCA can be approximated by the spectral decomposition of the matrices $\mathbb{E}\,(Z\, B_{lk}\, {}^t\overline{Z\, B_{lk}})$; $l = -k+1, \ldots, k-1$, where Z is the random measure associated with $(X_n)_{n\in\mathbb{Z}}$.

Let $(X'_n)_{n\in\mathbb{Z}}$ be the $q-$dimensional solution of the $q-$order FPCA of $(X_n)_{n\in\mathbb{Z}}$. This series is of the form $X'_n = \sum_{m\in\mathbb{Z}} C'_m X_{n-m}$. It can be approximated via the discretization of the spectrum, by the series

$$X'^k_n = \sum_{m\in\mathbb{Z}} C'_{m,k} X_{n-m},$$

where

$$C'_{m,k} = (2\pi)^{-1} \sum_{l=-k+1}^{k-1} (\int_{B_{lk}} e^{i\lambda m}\, d\,\mu(\lambda)) \sum_{j=1}^{q} F_j\, {}^t\overline{A_{jlk}},$$

F_j being the j^{th} vector of the canonical basis of $\mathbb{C}^q$, and A_{jlk} being the j^{th} unitary eigenvector of $\mathbb{E}\,(Z\, B_{lk}\, {}^t\overline{Z\, B_{lk}})$.

The reconstituted series is then $(X''^k_n)_{n\in\mathbb{Z}}$, which can be writen

$$X''^k_n = \sum_{m\in\mathbb{Z}} C''_{m,k} X'^k_{n-m} = \sum_{m\in\mathbb{Z}} D_{m,k} X_{n-m}\ ,$$

where

$$C''_{m,k} = (2\pi)^{-1} \sum_{l=-k+1}^{k-1} (\int_{B_{lk}} e^{i\lambda m}\, d\mu(\lambda)) \sum_{j=1}^{q} A_{jlk}\, {}^t\overline{F_j}\,,$$

and

$$D_{m,k} = (2\pi)^{-1} \sum_{l=-k+1}^{k-1} (\int_{B_{lk}} e^{i\lambda m}\, d\mu(\lambda)) \sum_{j=1}^{q} A_{jlk}\, {}^t\overline{A_{jlk}}\,.$$

The matrices $\mathbb{E}(Z\, B_{jl}\, {}^t\overline{Z\, B_{jl}})$ can be estimated by $(2\,\pi\, m)^{-1} \sum_{u=1}^{m} \sum_{v=1}^{m} (\int_{B_{lk}} e^{i\lambda(u-v)}\, d\mu(\lambda)) X_v\, {}^t\overline{X_u}$.

Of course, the greater k is, the nearest the approximated FPCA is to theoretical FPCA defined above.

In the following, we will examine the norms of the $C'_{m,k}$, which are high when the gap of m in the linear combination of the reconstruction is high, what happens, for example, when the series is periodic of period m.

We will also compare the series before and after the FPCA, for various values of the dimension q of the reconstruction.

5 Results

5.1 Result of POD

We examine the modes of this analysis, which correspond to the principal components in PCA. The POD reconstructs 80% of the variance of the signal with 6 modes. Figure 3 gives the plot of the first and the second modes as a time series. It shows that the variability of the modes is decreasing slowly. These modes are not entirely uncorrelated with the part of the flow described by a particular space-only POD mode.

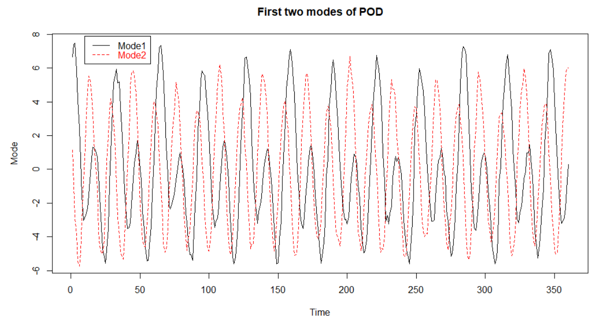

Figure 3: POD : modes 1 and 2

.

In Figure 4, the reconstruction is very slightly improved from 1 to 2 dimensions. At least, the essential of the variations is returned.

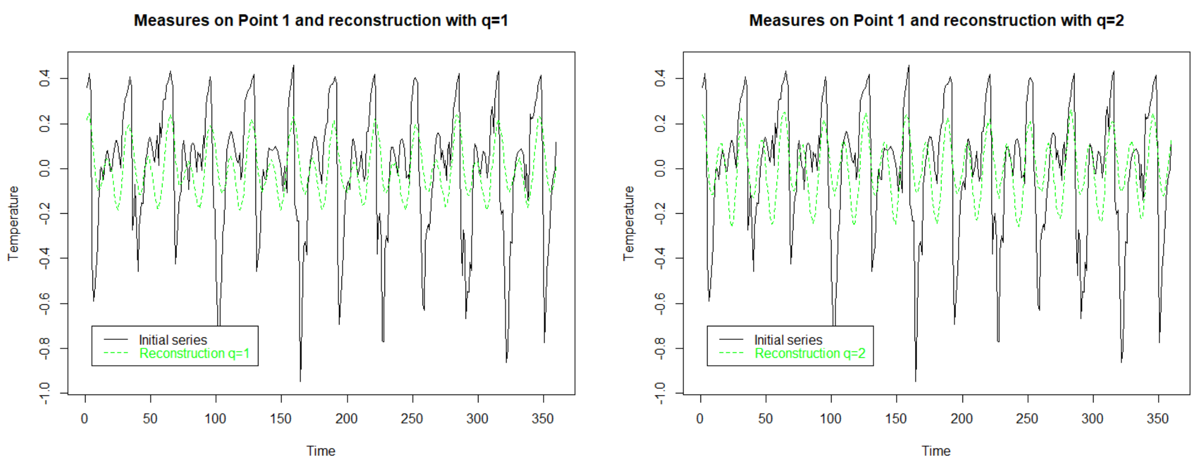

Figure 4: POD : initial data on Point 1 and reconstruction with one and two modes

.

5.2 Result of FPCA

Figure 5 gives the first and second dimension reconstruction of the first point.

In Table 1, we see that FPCA is best performing than POD for a same dimension reconstruction. The first dimension reconstruction reveals a phenomenom which has little amplitude variation, as the second dimension reconstruction is almost equal to the initial series.

The norms of the coefficients C'_{lk} (Figure 6) let us know that the series has got little periodicity. The fact that some coefficients different from $C_{0,k}$ are non null means that there is a time dependence, especially until lag 19.

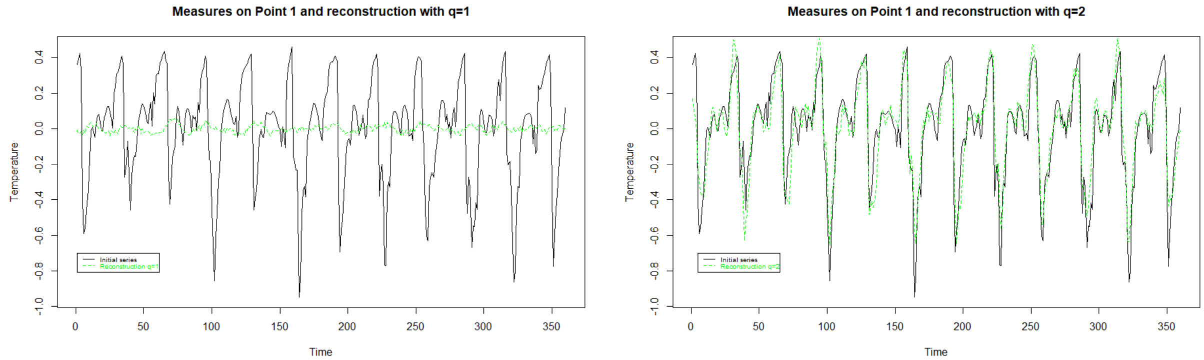

Figure 5: FPCA, $k = 20$: reconstruction on dimensions 1 and 2

.

Method	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 20$
POD	0.807	0.618	0.434	0.252	0.102
FPCA, $k = 10$	0.080	0.033	0.013	0.011	0.001
FPCA, $k = 20$	0.0007	0.0003	0.0002	0.0001	0.00002

Table 1: Table of mean square errors

.

6 Conclusion

The FPCA sounds interesting for several purposes in fluid mechanics. The summary needs few modes to give good quality of reconstitution compared to POD. The analysis of the coefficients of the reconstitution gives information above the periodic parts of the signal. Moreover, Boudou and Viguier-Pla (2006) have investigated the conditions where PCA and FPCA give the same results. This condition is the independence of data from time. The difference between POD and FPCA results give indications about how time-dependent are the data.

FPCA is here newly applied to data from fluid mechanics. This first study is giving promising results and offers new possibilities of exploration.

References

[1] Abide, S. and Binous, M.S. and Zeghmati, B. (2017) An efficient parallel high-order compact scheme for the 3D incompressible Navier-Stokes equations. *International Journal of Computational Fluid Dynamics.* **31**, 4-5, pp. 214-229

[2] Abide, S., Viazzo, S. and Raspo, I. (2018) Higher-order compact scheme for high-performance computing of stratified rotating flows. *Computers & Fluids.* **174** pp300-310

[3] Abide, S. (2020) Finite difference preconditioning for compact scheme discretizations of the Poisson equation with variable coefficients. *Journal of Computational and Applied Mathematics*, **379** 112872

[4] Antoranz, A., Ianiro, A., Flores, O., García-Villalba, M. (2018). Extended proper orthogonal decomposition of non-homogeneous thermal fields in a turbulent pipe flow. *Int. Journal of Heat and Mass Transfer*, 118, pp. 1264-1275.

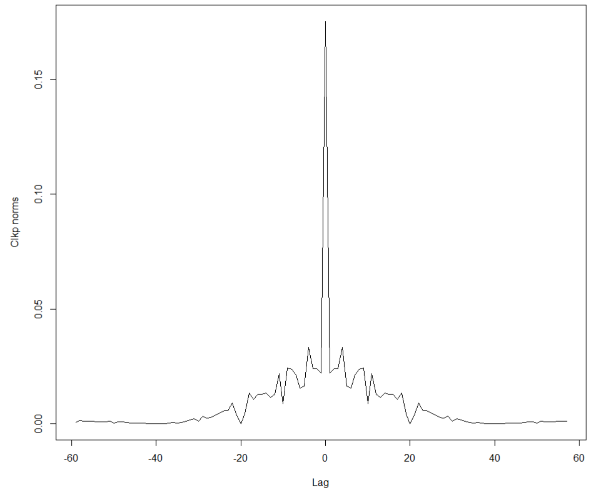

Figure 6: FPCA, $k = 20$: norms of the coefficients for the reconstruction of $\{X_n\}$

.

[5] Boudou, A. (1995). Mise en œuvre de l'analyse en composantes principales d'une série stationnaire multidimensionnelle. *Publications de l'Institut de Statistique de l'Université de Paris*, XXXIX, fasc.1, pp. 89-104.

[6] Boudou, A., Caumont, O. and Viguier-Pla, S. (2004). Principal components analysis in the frequency domain. *COMPSTAT 2004-Proceedings in Computational Statistics*, Physica, Heidelberg, pp. 729-736.

[7] Boudou, A., and Viguier-Pla, S. (2006). On proximity between P.C.A. in the frequency domain and usual P.C.A.. *Statistics*, **40**, pp. 447-464.

[8] Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* 2nd ed. Society for Industrial Applied Mathematics, Philadelphia. Initialy published in 1981.

[9] Lumley, J.L. (1970). *Stochastic tools in turbulence.* Academic Press, New-York.

[10] Rodda C., Hien, S., Achatz, U. and Harlander, U. (2020). A new atmospheric-like differentially heated rotating annulus configuration to study gravity wave emission from jets and fronts. *Experiments in Fluids* **61** 1 2

[11] Schmid, P.J. (2010). Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mechanics.* **656**, pp5-28.