# Analysing discrete and continuous spectrum and dimension reduction for thermal fields

Mélanie Dreina[1], Sylvie Viguier-Pla[1&2], Stéphane Abide[3]
[1]*LAMPS - Université de Perpignan via Domitia, F-66860 Perpignan Cedex 9, France*,
[2]*IMT, Université Paul Sabatier, F-31062 Toulouse Cedex 4, France*
[3]*Université Côte d'Azur, CNRS, LJAD F-06108 Nice Cedex 2, France*
melanie.dreina@univ-perp.fr, viguier@univ-perp.fr, stephane.abide@univ-cotedazur.fr

**Abstract**. The aim of this work is to compare three methods of data reduction in the context of heat transfer. This follows the well-known practice of observing unsteady phenomena according to space or time energetic arguments. Especially, the present study focuses on the efficiency of the Proper Orthogonal Decomposition (POD), Spectral Proper Orthogonal Decomposition (SPOD) and Principal Components Analysis in the Frequency domain (FPCA). In several previous works, both POD and SPOD have been proposed in the context of fluid mechanics while FPCA is been newly applied to this domain. Thus, in this work we provide a discussion on the contribution of the FPCA to deal with multiscale physics.

**Keywords.** Direct numerical simulation, principal components analysis, time series, stationarity, spectral analysis, thermal field

**Résumé**. Le but de ce travail est de comparer trois méthodes de réduction des données dans un contexte de transfert de chaleur. Nous nous plaçons dans le cas bien connu où nous observons des phénomènes instables dans le temps et l'espace. Plus précisément, nous nous intéressons à l'analyse en composantes principales (appelée dans le domaine de la mécanique des fluides la décomposition orthogonale aux valeurs propres ou POD), à la POD spectrale (SPOD), et à l'analyse en composantes principales dans le domaine des fréquences. Les méthodes POD et SPOD ont été proposées dans un contexte de mécanique des fluides, alors que la FPCA est nouvellement appliquée à ce domaine. Ainsi, dans ce travail, nous proposons une discussion sur la capacité de la méthode FPCA à se positionner dans une analyse physique multi-échelle.

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

# 1 Introduction

Simulation of complex systems, such as in fluid mechanics, leads to the production of a large amount of information. Therefore, dimension reduction is of major importance to be able to carry out fine analyses of the underlying physical phenomena. In this context, several approaches have been developed since the last decades. 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) proposed by Schmid (2010). All of them look for an efficient way to give an alternative representation of 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 time-dependent measurements, the independence is no more ensured in

the time domain, but when the signal is a stationary process, its Fourier Transform gives independent observations. The Spectral POD (SPOD) is a new method introduced by Lumley (2007) in order to rank the modes according to their energy level having a characteristic frequency. Finally, PCA in the frequency domain (Brillinger, 2001) aims a similar purpose to the SPOD, with PCA of the Fourier Transform for each identified frequency. Boudou (1995) has proposed a generalisation of this PCA in the frequency domain, that we name FPCA, and which has first been performed in Boudou et al. (2004) for periodic flows.

In this study, we propose to compare results from POD, SPOD and an improved method of FPCA, which is not restrictive on the structure of the multidimensional signal spectrum . In this presentation, we first present the data of interest. Secondly, we present the three compared methods, that is POD, SPOD and FPCA. In the third part, we apply the three methods on data from simulation. We end by showing the difficulties of each method, we compare the qualities of reconstruction and the phenomenon each method reveals at each step.

## 2 Description of the interest data

The comparison of POD, SPOD and FPCA methods is carried out on the spatio-temporal serie basis of a natural convection flow temperature field (Sergent et al., 2013, Trias et al., 2007). In particular, we simulate the thermal coupling between a fluid and a solid wall, imposing continuity of the temperature field at the fluid/solid interface. The data considered here is therefore a sampling of a variable $T(t, x, y)$ determined by Direct Numerical Simulation. For the sake of illustration, a snapshot and a time series are given in Fig. 1. Details of the DNS solver used in this work are presented in Abide (2017, 2018).

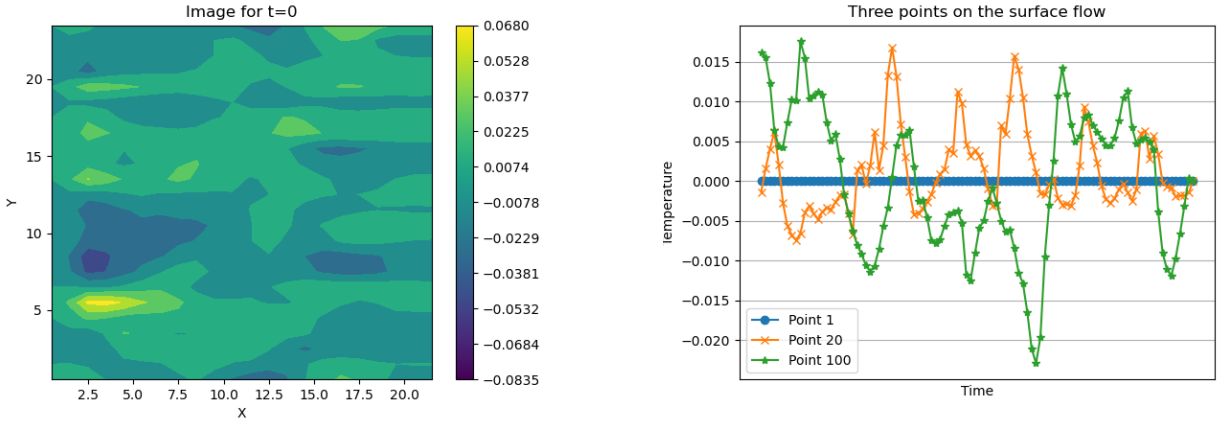


Figure 1: First snapshot and time series of the temperature fluctuations

.

The direct numerical simulation of turbulent natural convection at $Ra = 10^9$ (Trias et al., 2007) has been carried out on a grid $48 \times 65$ over 1000 time steps. In the following, comparisons between POD, SPOD and FPCA are based on a sub-sampling of the wall temperature by 100 snapshots of dimension $24 \times 22$ points.

# 3 The methods of dimension reduction

## 3.1 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 of 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 (PCA) of the matrix $U = (u(x_j, t_i))_{j=1,\ldots,n;i=1,\ldots,p}$. Each principal component of such a PCA is named a mode.

## 3.2 Spectral Proper Orthogonal Decomposition

Spectral proper orthogonal decomposition (SPOD) is a frequency domain variation of POD recently brought to the fore by Towne et al. (2018). This method is designed for statistically stationary flows. It is aimed to extract coherent structures from flow data. For example, it has been applied to extract the spatio-temporal modes of a jet and wind turbine flows in He et al. (2021). The main contribution of the SPOD compared to the POD is that the modes vary in both spatial and temporal dimensions, and are orthogonal under a space-time inner product, as opposed to being purely spatial. Consequently, these modes are optimal for representing spatio-temporal coherence within the data. In mathematical terms, the SPOD modes represent eigenvectors of the cross-spectral density (CSD) matrix at individual frequencies, where the eigenvalues denote the energy associated with each mode at a given frequency. A more detailed description of the method is given by Schmidt & Colonius (2020). We perform the SPOD with the open access python script proposed and successfully applied to several example by He et al. (2021) [https://github.com/HexFluid/spod_python (accessed February 2024)]. This script is built up as follows.

The first step is to build a matrix for the spatio-temporal data. Let the vector $q_k \in \mathbb{R}^{N_q}$ be the $k^{th}$ time snapshot after subtracting the time-averaged data. The chronologically sorted spatio-temporal data matrix is:

$$\mathcal{Q} = [q_1, q_2, \ldots, q_{N_t}] \in \mathbb{R}^{N_q \times N_t},$$

where $N^t$ is the number of snapshots. Secondly, the data matrix is decomposed into $N_b$ blocks using the Welch periodogram method and the discrete Fourier transform is applied to each block to pass into the frequency domain. At this stage, to prevent loss of precision due to spectral leakage, each data block is processed with a Hamming window and then overlapped with neighbouring blocks. The matrix for the $j^{th}$ block is

$$\hat{Q}^{(j)} = \left[\hat{q}_k^{(j)}, \hat{q}_2^{(j)}, \ldots, \hat{q}_{N_f}^{(j)}\right] \in \mathbb{C}^{N_q \times N_f}$$

Then, according to the frequency, the matrices are reshaped so that the matrix for the $k^{th}$ frequency is

$$\hat{Q}_k = \left[\hat{q}_k^{(1)}, \hat{q}_k^{(2)}, \ldots, \hat{q}_k^{(N_b)}\right] \in \mathbb{C}^{N_q \times N_b}.$$

The weighted cross-spectral density (CSD) matrix for the $k^{th}$ frequency, denoted as $S_k$ is obtained as follows:

$$S_k = \frac{1}{N_b} W^{1/2} \hat{Q}_k^* \hat{Q}_k W^{1/2} \in \mathbb{C}^{N_b \times N_b},$$

where $W$ represents the weight matrix for scaling the various flow variables. The specific definition of $W$ determines the physical interpretation of the energy associated with the SPOD modes. Finally, the eigen-decomposition is performed on the weighted cross-spectral density (CSD) matrix $S_k$ for each frequency. The resulting modes are used for a variety of purposes such as classification and

reduced order modelling. Similar to other versions of Proper Orthogonal Decomposition (POD), SPOD determines an orthogonal basis for the data, meaning that a subset of these modes captures a proportion of the total energy (variance) within the data compared to any other orthogonal basis. The function used for the reconstruction is based on Nekkanti & Schmidt (2021), and is available in the python script above mentioned.

## 3.3 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 $\mathbb{Z} \times \mathbb{Z} \times \{1, ..., p\} \times \{1, ..., p\}$.

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 $(2\,\pi)^{-1} \sum_{n\in\mathbb{Z}} e^{-i\lambda n} \mathbb{E}\, X_n^t \overline{X_0}$, for each $\lambda$ from $[-\pi,\, \pi[$, this means an infinity of PCA's. We overcome this difficulty by a discretization of the spectrum $[-\pi,\, \pi[$.

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 spectral density $M_{lk}$ defined on each $B_{lk}$; $l = -k+1, \ldots, k-1$. The matrices $M_{lk}$ can be estimated by

$$(2\,\pi\, m)^{-1} \sum_{u=1}^{m} \sum_{v=1}^{m} (\int_{B_{lk}} e^{i\lambda(u-v)}\, d(\lambda)) X_v\, {}^t\overline{X_u}$$

.

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

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

$F_j$ being the $j^{\text{th}}$ vector of the canonical basis of $\mathbb{C}^q$, and $A_{jlk}$ being the $j^{\text{th}}$ unitary eigenvector of $M_{lk}$.

The reconstructed series is then $(X_n''^k)_{n\in\mathbb{Z}}$, which can be writen

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

where

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

and

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

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

We can examine the norms of the $C'_{m,k}$, which are high when the gap $m$ in the linear combination of the reconstruction is high, what happens, for example, when the series is periodic of period $m$. We can also compare the series before and after the FPCA, for various dimension $q$ values of the reconstruction.

# 4 Results and discussion

## 4.1 Analysis with POD

We examine the modes of this analysis, which match with the principal components in usual PCA. In Figure 2, the reconstruction is very slightly improved from 1 to 2 dimensions. At least, the essential of the variations is returned.

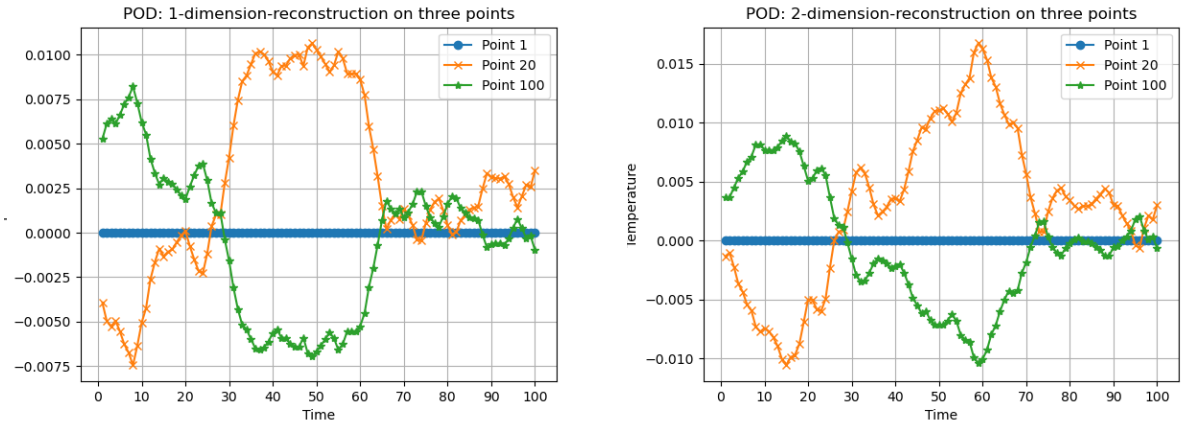


Figure 2: POD : Reconstruction of three points variations with one and two modes

.

The variations most reconstructed are those for median temperature, as we see in Figure 3.

## 4.2 Analysis with SPOD

In Figure 4, we can see that the variations of points 20 and 100 are slightly more complex than the ones from POD, but the same variations are first retrieved.

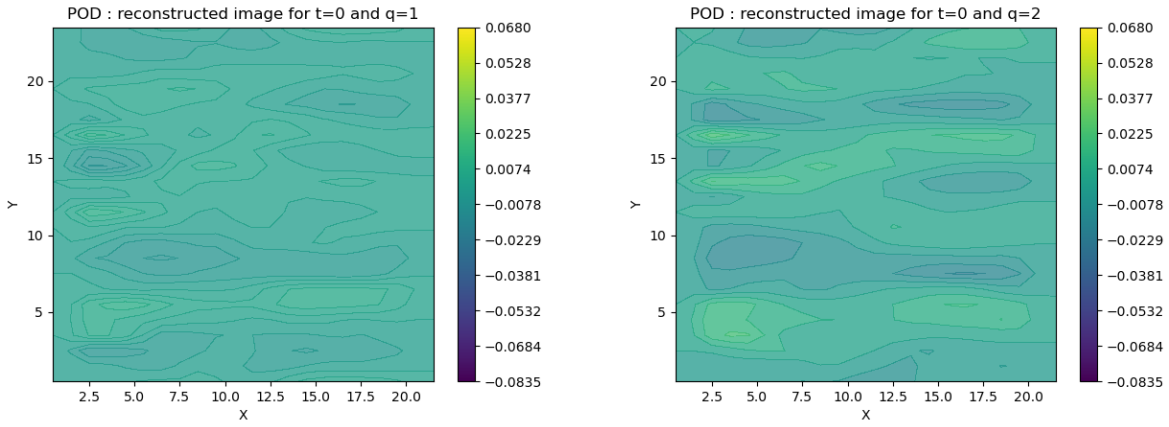


Figure 3: POD : Reconstruction of image at t=0 with one and two modes

.

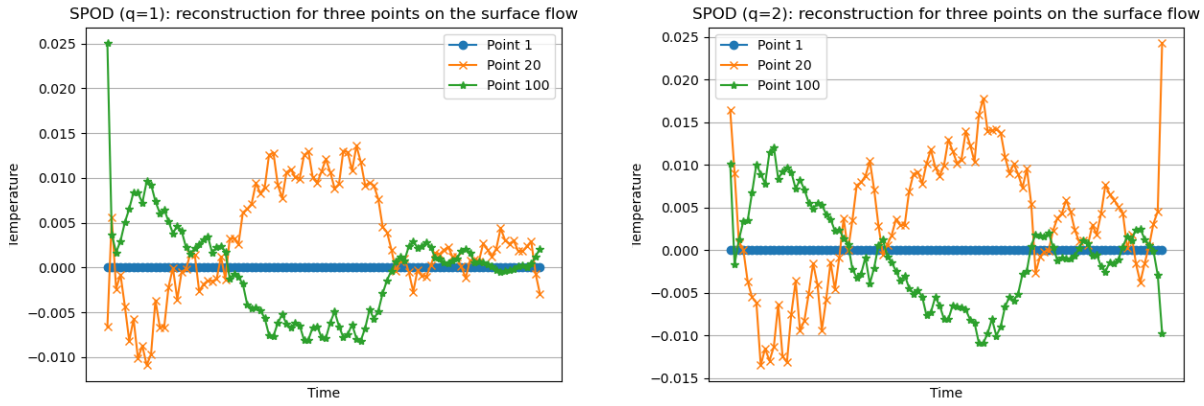


Figure 4: SPOD : Reconstruction of three points variations with one and two modes

.

The variations most reconstructed are those for extreme temperatures at $t = 0$, as we see in Figure 5.

## 4.3 Analysis with FPCA

As for the trajectories of points 20 and 100, FPCA retrieves more complexity than the previous methods. It takes into account more frequencies in the first modes (Figure 6).

Figure 7 gives the first snapshot for a one and two-dimensions reconstruction. By comparison with the initial first snapshot, we can recognise the main variations of the flow, yet for $q = 1$.

## 4.4 Comparisons

One way to assess the efficiency of decomposition method relies on its ability to reconstruct the initial signal with few modes. To this end, we evaluate the error in reconstruction with respect to the mode numbers. Figure 8 presents the relative error computed for the three methods POD, SPOD and FPCA. One can note that FPCA is able to reconstruct the data with fewer modes than the other methods. In this way, FPCA overcomes the POD and SPOD in its ability to retrieve data. Moreover,

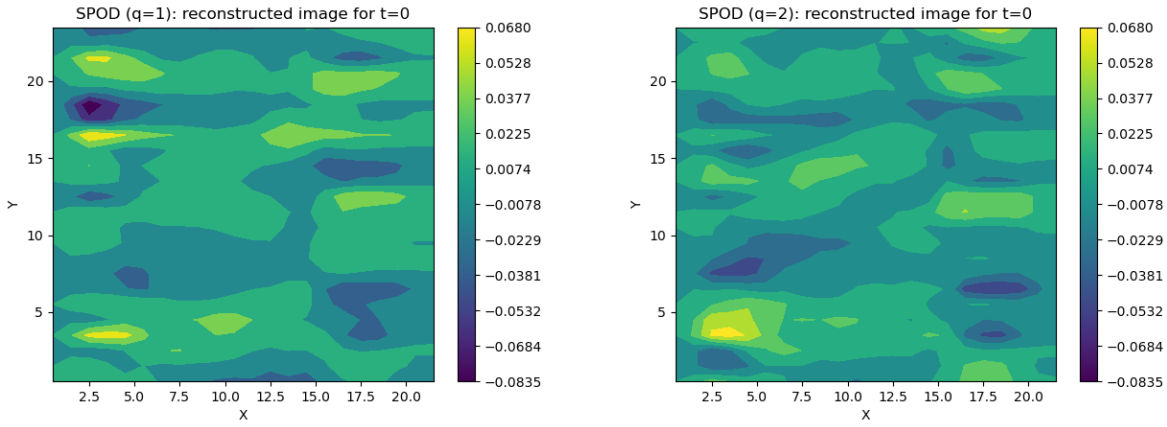


Figure 5: SPOD : Reconstruction of image at t=0 with one and two modes

.

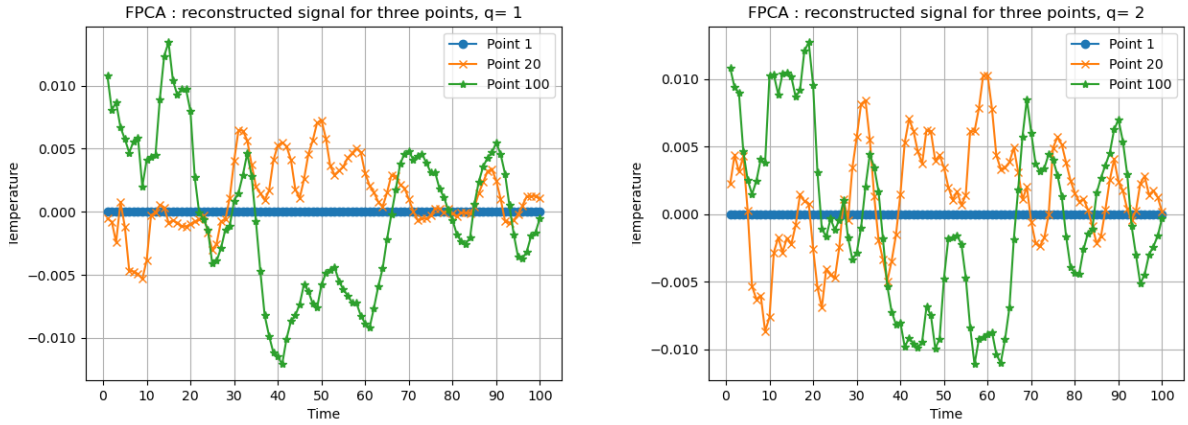


Figure 6: FPCA, $k = 10$ : reconstruction on dimensions 1 and 2

.

the control parameter $k$ improves greatly the decomposition efficiency. When $k$ is small, the number of subdivisions of the frequency spectrum is small, so few frequencies are taken into account. The higher $k$ is, the higher is the number of considered potential frequencies. The method FPCA has been performed with $k = 10$ and $k = 20$. This comparison of two values of $k$ illustrates the fact that the higher $k$ is, the smaller the error is, for a fixed value of $q$. As POD and SPOD present similar errors in the first dimensions, SPOD tends to be better with dimension getting higher. FPCA has more little errors, and the quality of reconstruction is almost perfect as soon as the dimension reaches $q = 10$ when $k = 10$, and $q = 6$ when $k = 20$.

# 5 Conclusion

The FPCA sounds interesting for several purposes in fluid mechanics. The summary needs few modes to give good quality of reconstruction compared to POD and SPOD. We can analyze the coefficients of the reconstruction for information above the periodic parts of the signal, and we can select part of the spectrum part for the extraction of some particular phenomena. 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, and a consequence of this independence is that FPCA and POD become equivalent. The difference between POD, SPOD and FPCA results give

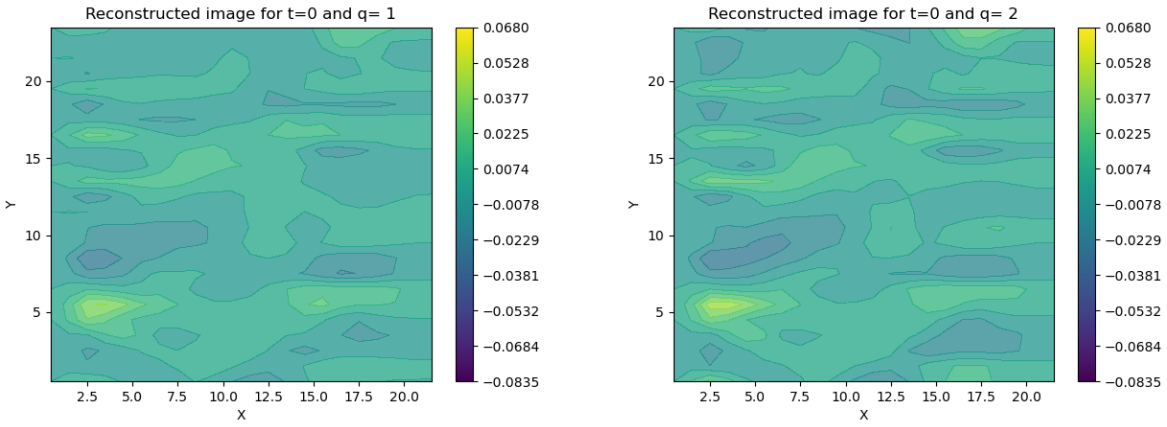


Figure 7: FPCA, $k = 10$ : reconstruction on dimensions 1 and 2

.

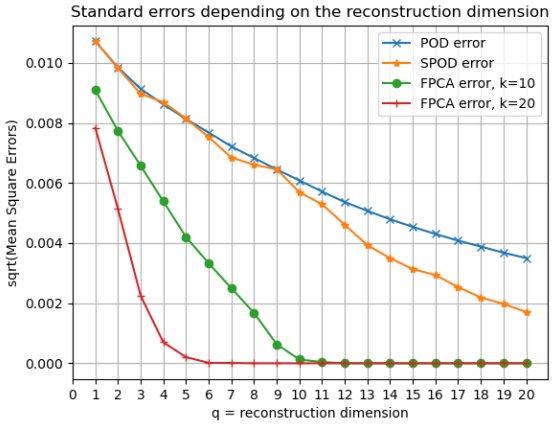


Figure 8: Standard deviation of errors of reconstruction for dimensions 1 to 20

.

indications about how time-dependent are the data.

FPCA is compared for the first time to SPOD, which is supposed to proceed with the same way of dealing with the frequency domain, and this on data simulated from fluid mechanics models. However, we must also analyze the computational efficiency of each method, and the ability of these methods to apply to large volumes of data. As FPCA has got longer execution time, one of the challenges is to adapt its algorithms to this context.

# 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** pp. 300-310.

[3] 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.

[4] 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.

[5] 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.

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

[7] He, X., Fang, Z., Rigas, G., and Vahdati, M. (2021). Spectral Proper Orthogonal Decomposition of Compressor Tip Leakage Flow. *Physics of Fluids*, **33 (10)**.

[8] Lumley, J.L. (2007). *Stochastic Tools in Turbulence*. Courrier Corporation. Initially published in 1970 in Academic Press, New-York.

[9] Nekkanti, A., and Schmidt, O.T. (2021). Frequency–Time Analysis, Low-Rank Reconstruction and Denoising of Turbulent Flows Using SPOD. *Journal of Fluid Mechanics*, **926**, A26.

[10] Schmid, P.J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, **656**, pp. 5-28.

[11] Schmidt, O.T., and Colonius, T. (2020). Guide to Spectral Proper Orthogonal Decomposition. *AIAA Journal*, **58 (3)**, pp. 1023–33.

[12] Sergent, A., Xin, S., Joubert, P., Le Quéré, P., Salat, J., & Penot, F. (2013). Resolving the stratification discrepancy of turbulent natural convection in differentially heated air-filled cavities *International Journal of Heat and Fluid Flow*, **39**, pp. 1-14.

[13] Towne, A., Schmidt, O.T., and Colonius, T. (2018). Spectral Proper Orthogonal Decomposition and Its Relationship to Dynamic Mode Decomposition and Resolvent Analysis. *Journal of Fluid Mechanics*, **847**, pp. 821–67.

[14] Trias, F. X., Soria, M., Oliva, A., & Pérez-Segarra, C. D. (2007). Direct numerical simulations of two-and three-dimensional turbulent natural convection flows in a differentially heated cavity of aspect ratio 4. *Journal of Fluid Mechanics*, **586**, pp. 259-293.