Spectral analysis of multidimensional thermal fields

Mélanie Dreina, Sylvie Viguier-Pla and Stéphane Abide

Abstract The study of multidimensional data along time often needs data reduction methods to reduce dimension, and to study some particular phenomena. In a context of fluid mechanics, we propose to compare a version of functional principal components analysis (FPCA), named proper orthogonal decomposition (POD), with two methods based on the spectral decomposition of data Fourier transform, the spectral proper orthogonal decomposition (SPOD) and principal components analysis in the frequency domain (PCAFD). In this context of, both POD and SPOD have been proposed, while PCAFD is been newly applied to this domain. Thus, we provide a discussion on the contribution of PCAFD to deal with multiscale physics.

1 Introduction

Simulation of multidimensional data, such as in fluid mechanics, or observation of such data, as for example ocean temperature, lead 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. Naturally, principal components analysis (PCA) is the basic method which has been declined by several approaches since the last decade. The pioneer work concerning fluid mechanics has

Mélanie Dreina
LAMPS, Université de Perpignan, 52 avenue Paul Alduy 66000 Perpignan Cedex 9, France, e-mail: melanie.dreina@univ-perp.fr

Sylvie Viguier-Pla
LAMPS, Université de Perpignan, 52 avenue Paul Alduy 66000 Perpignan Cedex 9, France & IMT, Université Paul Sabatier, 118 route de Narbonne 66860 Toulouse Cedex 4, France, e-mail: viguier@univ-perp.fr

Stéphane Abide
LJAD, Université Côte d'Azur, 28 avenue Valrose 06108 Nice Cedex 02, France, e-mail: Stephane.ABIDE@univ-cotedazur.fr

been introduced by Lumley (1970), with Proper Orthogonal Decomposition (POD), which can be apparented with Functional Principal Components Analysis (FPCA). The Spectral POD (SPOD), also proposed by Lumley (2007), which performs the Fourier transform on data before POD has been brought to the fore by Towne et al. (2018), and put into practice by various authors, as Schmidt and Colonius (2020), Xiao et al. (2021) or Nekkanti and Schmidt (2021). More recently the Dynamic Mode Decomposition (DMD) proposed by Schmid (2010), is performed by a time domain PCA, with time-dependent assumptions. We will focus our work on PCA in the frequency domain, introduced by Brillinger (1981) generalized to a wide context by Boudou (1995), performed for a periodic flow by Boudou et al. (2004), and generalized for a cyclostationary flow by Boudou and Viguier-Pla (2020). This technique of using reduction methods in the frequency domain is knowning a large interest, as it is better appropriate for time dependent series than working in the time domain (see also Hörmann et al, 2015).

In this study, we propose to compare results from POD, SPOD and an improved method of PCA in the frequency domain, that we name PCAFD, which is not restrictive on the structure of the multidimensional signal spectrum. In this work, we first present the data of interest. Secondly, we present the three compared methods, that is POD, SPOD and PCAFD. In the third part, we apply the three methods on our simulated data. 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 data

The comparison of POD, SPOD and PCAFD 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 simulations $\{u(t, x_1, x_2); (t, x_1, x_2) \in \mathbb{R}_+ \times [a; b] \times [c; d]\}$, where t is time index, x_1 and x_2 are coordinates of the point where the random process u is determined by Direct Numerical Simulation (DNS). For the sake of illustration, a snapshot, that is an image at a time index t, and time series for three points across time, are given in Fig. 1. Details of the DNS solver used in this work are presented in Abide et al. (2017, 2018).

3 The methods of dimension reduction

Proper Orthogonal Decomposition. Let $\{u(t, x)\}$ be a stochastic process defined on $\mathbb{R}\times\mathbb{R}^n$, 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 $(x_1, \ldots, x_n)$

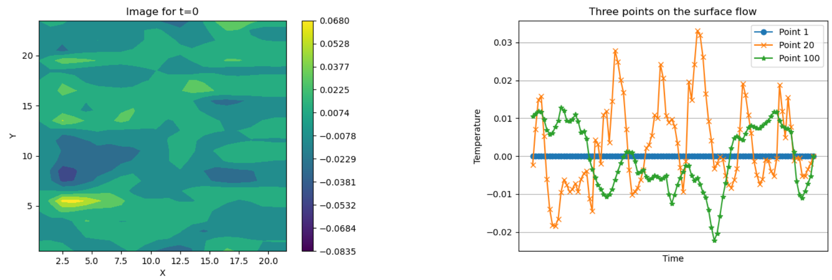

Fig. 1 First snapshot and time series of the temperature fluctuations

of space points, and measures at times $t_1, \ldots, t_p$. The method is implemented via PCA of the matrix $U = (u(t_i, x_j))_{i=1,\ldots,p; j=1,\ldots,n}$. Each principal component of such a PCA is named a mode.

Spectral Proper Orthogonal Decomposition. SPOD is designed for statistically stationary flows, from which it is aimed to extract coherent structures. The eigenvectors of the cross-spectral density (CSD) matrix at individual frequencies represent SPOD modes, and the eigenvalues represent the energy associated with each mode at a given frequency. We perform SPOD with the open access python script proposed and successfully applied to several examples by He et al. (2021) [https://github.com/HexFluid/spod_python (accessed December 2024)]. This script follows the main steps of the method. First it builds a matrix with the spatio-temporal data. Let the vector $u_k \in \mathbb{R}^n$ be the k^{th} time snapshot after subtracting the time-averaged data. The chronologically sorted spatio-temporal data matrix is:

$$Q = U^T = \left[u_1, u_2, \ldots, u_p\right] \in \mathbb{R}^{n \times p},$$

where p is the number of snapshots. Second, the data matrix is splitted 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 j^{th} block matrix is given by

$$\hat{Q}^{(j)} = \left[\hat{u}_1^{(j)}, \hat{u}_2^{(j)}, \ldots, \hat{u}_{N_f}^{(j)}\right] \in \mathbb{C}^{n \times N_f}.$$

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

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

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 is the weight matrix for scaling the various flow variables (we chose $W = I$). Finally, the eigen-decomposition is performed by the CSD matrix S_k for each frequency. Similarly to other versions of 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.

Principal Components Analysis in the Frequency Domain. Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary $p-$dimensional random time series. The PCAFD 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 PCAFD 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: $X_n = \left(x_n^1, \ldots, x_n^p\right)^t$. 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 PCAFD 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 λ 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 PCAFD can be approximated by a spectral decomposition of each spectral density M_{lk} defined on B_{lk}; $l = -k + 1, \ldots, k - 1$. The matrices M_{lk} can be estimated by

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

Let $(X'_n)_{n\in\mathbb{Z}}$ be the $q-$dimensional solution of the $q-$order PCAFD 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

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

where

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

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 M_{lk}.

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

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

where

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

and

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

Of course, the greater k is, the closest the approximated PCAFD is to the theoretical PCAFD defined above.

We can examine the norms of the $C'_{m,k}$, which are high when the autocorrelation of order m is high, what happens, for example, when the series is periodic of period m. We can also compare the series before and after PCAFD, for various dimensions q of the reconstruction.

4 Results and discussion

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 using to 2 dimensions. At least, the essential of the variations is returned, that is those for median temperature.

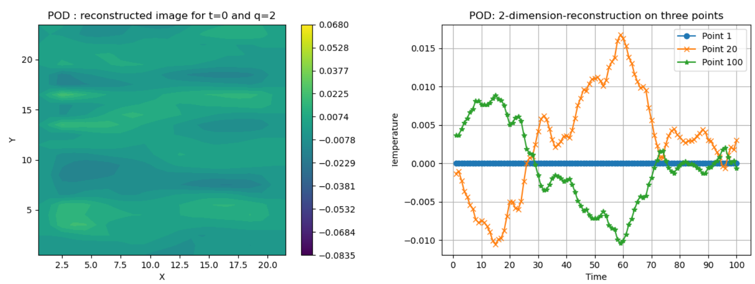

Fig. 2 POD. Reconstruction of image at $t = 0$ and of three points variations with two modes

.

Analysis with SPOD. In Figure 3, we can see that the variations most reconstructed are those for extreme temperatures at $t = 0$. The variations of points 20 and 100 are slightly more complex than the ones from POD, but the same variations are first retrieved.

Analysis with PCAFD. Figure 4 gives the first snapshot and trajectories of three points for a two-dimensions reconstruction. By comparison with the initial first snapshot, we can recognise the main variations of the flow. As for the trajectories, PCAFD retrieves more complexity than the previous methods. It takes into account more frequencies in the first modes.

Now we can focus on the possible periodicities in the signal, by looking at the norms of the coefficients of the reconstruction of (X_n). In Figure 5, the norms of the C'_{lk} show some phenomenum of period around 26. So we guess that any frequency of the form $p\frac{2\pi}{26}$ in $[-\pi;\pi[$ is of high energy. The frequency $\frac{4\pi}{26}$ falls in the interval

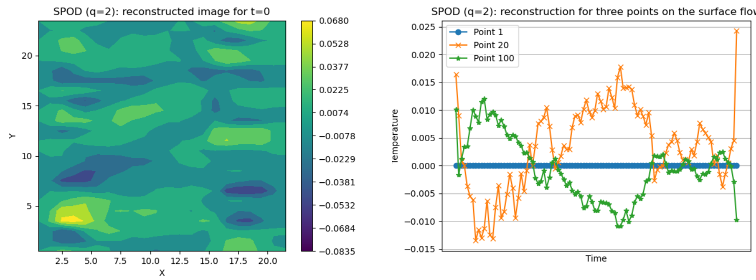

Fig. 3 SPOD. Reconstruction of image at $t = 0$ and of three points variations with two modes

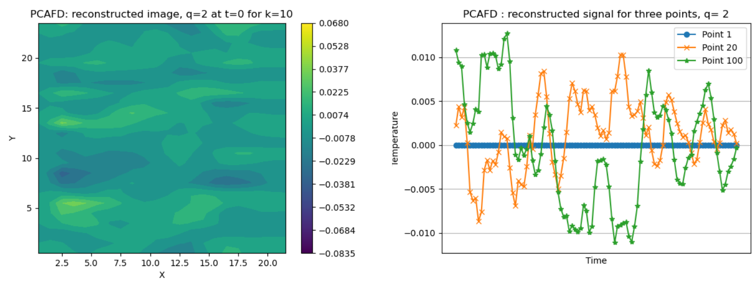

Fig. 4 PCAFD, $k = 10$. Reconstruction on dimension 2

B_{lk}, where $l = 1$ for $k = 10$. In the second diagram of Figure 5, we show the three selected trajectories for frequencies in $B_{1,10}$ and $q = 2$. Note that this period depends on the time discretization. This technique offers a way to analyze what is specific to a given frequency, when the bandwidth is small enough.

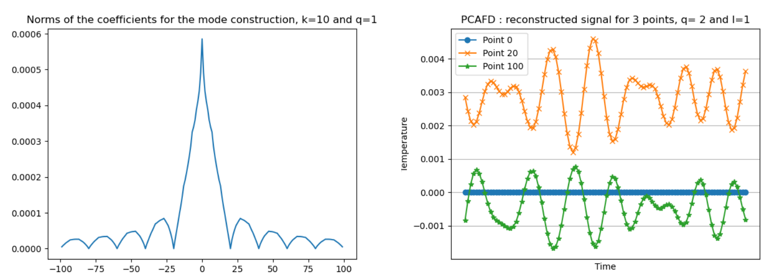

Fig. 5 PCAFD, $k = 10$. Norms of the C'_{lk} and reconstruction for 3 points on $B_{1,10}$

5 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 6 presents the relative error computed for the three methods POD, SPOD and PCAFD. One can note that PCAFD is able to reconstruct the data with fewer modes than the other methods. In this way, PCAFD overcomes the POD and SPOD in its ability to retrieve data. Moreover, the control parameter k greatly improves 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 PCAFD 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. PCAFD 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 $k = 20$.

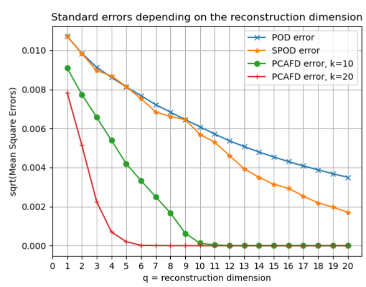

Fig. 6 Standard deviation of errors of reconstruction for dimensions 1 to 20

.

6 Conclusion

The PCAFD 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 for the extraction of some particular phenomena. Moreover, Boudou and Viguier-Pla (2006) have investigated the conditions where PCA and PCAFD give the same results. This condition is the independence of data from time, and a consequence of this independence is that PCAFD and POD become equivalent. The difference between POD, SPOD and PCAFD results give indications about how time-dependent are the data.

PCAFD is newly compared to SPOD, which is supposed to proceed with the same way of dealing with the frequency domain, 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 PCAFD 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.: 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 (2017)
2. Abide, S., Viazzo, S. and Raspo, I.: Higher-order compact scheme for high-performance computing of stratified rotating flows. *Computers & Fluids*, **174** pp. 300-310 (2018)
3. Boudou, A.: 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 (1995)
4. Boudou, A., Caumont, O., Viguier-Pla, S.: Principal components analysis in the frequency domain. *COMPSTAT 2004-Proceedings in Computational Statistics*, Physica, Heidelberg, pp. 729-736 (2004)
5. Boudou, A., Viguier-Pla, S.: Principal components analysis of a Cyclostationary Random Function. *IWFOS 2020-Special issue*, pp. 43-50. Springer, Heideberg (2020)
6. Hörmann, S., Kidziński, L., Hallin, M.: Dynamic Functional Principal Components. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* vol. 77, issue 2, pp 319-348. (2002) doi: 10.1111/rssb.12076
7. He, X., Fang, Z., Rigas, G., Vahdati, M.: Spectral proper orthogonal decomposition of compressor tip leakage flow. *Physics of Fluids* vol. 33, 105105 (2021) doi: 10.1063/5.0065929
8. Lumley, J.L.: *Stochastic Tools in Turbulence*. Courrier Corporation. Initially published in 1970 in Academic Press, New-York (1970, 2007)
9. Nekkanti, A., and Schmidt, O.T.: Frequency–Time Analysis, Low-Rank Reconstruction and Denoising of Turbulent Flows Using SPOD. *Journal of Fluid Mechanics*, **926**, A26 (2021)
10. Schmidt, O.T., Colonius, T.: Guide to Spectral Proper Orthogonal Decomposition. *AIAA Journal*, **58**, 3 pp. 1023-1033 (2020)
11. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, **656**, pp. 5-28 (2010)
12. Sergent, A., Xin, Shihe, Joubert, P., Le Quéré, P., Salat, J., Penot, F.: Resolving the stratification discrepancy of turbulent natural convection in differentially heated air-filled cavities - Part I: Reference solutions using Chebychev spectral methods. *International Journal of Heat and Fluid Flow*, **39**, pp. 1-14 (2013)
13. Towne, A., Schmidt, O.T., Colonius, T.: Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics*, **847**, pp. 821-867 (2018)
14. Trias, F.X., Soria, M., Oliva, A., Pérez-Segarra, C.D.: Direct numerical simulations of two- and three-dimensional turbulent natural convection flows in differentially heated cavity of aspect ratio 4. *Journal of Fluid Mechanics*, **586**, pp. 259-293 (2007)