# A STOCHASTIC LOW-ORDER MODELLING APPROACH FOR TURBULENT SHEAR FLOWS

**G. Lehnasch and J. Delville**
Department of Fluid Flow, Heat Transfer and Combustion
Institute Pprime - UPR 3346 - CNRS - Université de Poitiers - ENSMA
CEAT - F86036 Poitiers cedex, France
guillaume.lehnasch@ensma.fr

## ABSTRACT

A reduced-order modelling, based on the POD-Galerkin approach, is developped in a probabilistic framework in order to build robust low-dimensional dynamical systems able to reproduce the dominant long-term dynamics of turbulent flows. It combines a statistical inference procedure to determine joined probability distribution functions of the dynamical system coefficients and a Monte-Carlo approach to foresee its most probable time evolution. The method is shown to reproduce correctly the main dynamical features of a turbulent mixing layer.

## INTRODUCTION

The concept of low-order representation of turbulence has aroused considerable interest in the fluid mechanics community in view of addressing various practical issues, ranging from reduced-cost flow control to the generation of realistic turbulent boundary conditions. Dealing with reduced-order models leads to face the fundamental issue of identifying the low-dimensional deterministic order hidden by the apparent high-dimensional chaotic behaviour of turbulent flows. Among the various approaches described in the litterature, the so-called POD-Galerkin approach still appears rather appealing to prescribe a coherent spatio-temporal dynamics. In this framework, the large scale coherent spatial information of the flowfield is assumed to be mainly contained in the $n$ most energetic spatial modes $\Phi_i$ obtained by a Proper Orthogonal Decomposition (POD) of the fluctuating field, leading to the following expansion of the turbulent flowfield:

$$u_i(x,t) = \overline{u_i(x)} + \sum_{i=1}^{n} a_i(t)\Phi_i(x) \tag{1}$$

where the overbar denotes the mean. By direct projection of the Navier-Stokes equations onto this POD basis, some Low Order Dynamical Systems (LODS) can be obtained to describe the time evolution of the POD coefficients $a_i(t)$, which lead to a simple quadratic function. Due to the truncation of the expansion up to $n$ modes, these LODS are however intrinsically unstable (Noack et al., 2003). Various approaches have been proposed to overcome this problem with nuanced success. The addition of eddy-viscosity-like corrective terms was proposed to account for the neglected POD modes and recover the necessary dissipation at lower scales (Aubry et al., 1988; Rempfer and Fasel, 1994). Relevant alternatives may be found with the addition of an artificial linear coefficient as a penalty term in the model (Cazemier et al., 1998), a spectral vanishing viscosity term (Karamanos and Karniadakis, 2000) or just a dissipative term related to the numerical scheme (Iollo, 2000). A model of the missing fine scales, based on the residual of the Navier-Stokes operator, evaluated with the POD flow fields, was also recently proposed by Bergmann et al. (2009) while Noack et al. (2008) presented a Finite-Time Thermodynamics (FTT) formalism to model the mode dependant internal and external interactions and describe the long-term evolution of each mode energy. Alternative strategies rely on the direct calibration of the POD-Galerkin model coefficients based on various error criteria (Cordier et al., 2009), or empirical procedures based on neural networks ( Gillies, 1998; Lorang et al., 2006) to identify directly the temporal modes evolution. The results obtained by using these methods on rather simple flows are generally satisfying for short time prediction (typically of the order of the integral time length scale or a time period corresponding to the training signal duration). However, they often fail in extrapolating the long-term dynamical system evolution. Artificial amplification and phase-shifts of the temporal modes are generally observed during time integration, often followed by the system blow up.

The approach presented in this paper stems from the observation that measurement uncertainty or limited machine accuracy unavoidably limit the long-term physical representativity of purely deterministic LODS which cannot naturally surrogate the chaotic component of turbulent flows. A long term stabilization of LODS may be more naturally ensured by adding stochastic contributions to correct any drifting dynamical evolution. The present contribution aims at suggesting a new strategy in a probabilistic framework for building both robust and reliable POD-based stochastic reduced-order models. The key elements of the model are first presented. The capacity of this approach to stabilize POD-based LODS is then demonstrated on a low-Reynolds cylinder wake flow and a mixing layer.

## MODELLING STRATEGY

The principle of the proposed method is twofold. At first, a probabistic approach is introduced in the flow calibration procedure in order to take into account the possible bias of the available sample collection (or, in a similar way, the data uncertainty). A parametric bootstrap procedure is used for this purpose, which leads to compute a full set of joined probability functions of the LODS coefficients. Then, a Monte-Carlo procedure is built to select probable near-future evolution of the dynamical flow during time integration.

### Calibration procedure

The essential ingredients of the model are obtained by the application of a Snapshot POD (Sirovich, 1987) of the whole velocity field under consideration, leading to the expansion given by Eq. (1). The identification procedure here retained basically extends the global flow calibration method of Perret et al (2006). The LODS structure is assumed to have the quadratic polynomial form that would be obtained by a Galerkin projection. At each known temporal instant $t_n$ ($n < N_s$), $N_s$ being the total number of available snapshots, each time derivative $da_i/dt(t_n)$ is thus assumed to be related to the other modes $a_j(t_n)$ according to:

$$C_i + \sum_{j=1}^{N_T} L_{ij} a_j(t_n) + \sum_{j=1}^{N_T} \sum_{k=j}^{N_T} Q_{ijk} a_j(t_n) a_k(t_n) = \frac{da_i(t_n)}{dt} \quad (2)$$

This system may be written in condensed form by:

$$A\Theta^{(i)} = B^{(i)} \quad (3)$$

where $A$, that is constant for each modal equation, stands for the matrix of all the instantaneous temporal mode monomials $(1, a_i, a_i a_j)$, and $B^{(i)}$ is the vector of the instantaneous values of the time derivative $da_i(t_n)/dt$ for the $i^{th}$ POD mode. For each mode $i$, the model coefficients $\Theta^{(i)}$ form a vector of size $N_\Theta = 1 + N_T + N_T(N_T+1)/2$, $N_T$ being the number of modes retained in the POD expansion. It is denoted by:

$$\Theta^{(i)} = (C_i, L_{ij}, Q_{ijk})^T \quad (4)$$

A singular value decomposition (SVD) is applied to solve each overdetermined system and compute $\Theta^{(i)}$, relative to each modal equation. This enables to filter out the less important singular values which would be likely to contaminate the solution in case of ill-conditioning of $A$.

### Statistical inference of LODS coefficients

The variability of the model coefficients related to their uncertain estimation is evaluated by a *parametric boostrap* procedure. This bootstrap method, proposed by Efron (1979), initially aims at unbiasing statistics computed from samples of small size for which statistical convergence is unreachable. It consists in generating new artificial databases by randomly choosing new subsamples from the initial database, with replacement, in order to perform new estimate of the statistics. Instead of a single biased value, a probability distribution of this statistic is thus obtained, from which most probable unbiased statistics may be extracted. The parametric bootstrap is here applied to the identification of LODS parameters. The PDFs of $\Theta^{(i)}$ are build progressively by ressampling randomly vectors $(a_i(t_n), da_i(t_n)/dt), n = 1, N_s)$ used for the application of the calibration procedure previously described. The size $N_s$ of each subsampled dataset has been chosen of same size as the reference dataset for the present study. Considering initially a database of a few hundred snapshots, a few thousands iterations of the calibration procedure is found to be sufficient to make the PDFs of the model parameters converge.

The figure (1) illustrates the typical individual and joined probability distributions obtained for each model parameter (corresponding to the case of the cylinder wake flow presented in the following section). It should be noted that the reference value, obtained by considering classically only the whole original data set, significantly differs from the most probable value now identified. In addition, as illustrated in this particular example, the model parameters whose average is near zero may sometimes yield the same probability to be negative and positive and thus to have fully opposite effects. This clearly indicates that the straightforward application of the original calibration procedure (without ressampling) is likely to enforce systematically wrong intermodal contributions. Accordingly, it seems judicious to account for this intrinsic uncertainty in order to correct eventually the dynamical flow.

### Stochastic modelling

It is proposed in the following to include the PDFs of model coefficients previously identified in the low-order modelling process itself. As illustrated by Fig. (1), it is empirically found that the model coefficients are practically jointly normally distributed. This remains true as long as the cutoff eigenvalue retained for the calibration by the SVD procedure is high enough. Accordingly, a simple normal law $\mathcal{N}(\overline{\Theta}_i, \sigma_{ii})$ is retained to set up a probabilistic representation of each coefficient:

$$p(\Theta_i) = \frac{1}{\sqrt{2\pi\sigma_{ii}}} e^{\left(-\frac{(\Theta_i - \overline{\Theta}_i)^2}{\sigma_{ii}}\right)}$$

where $\overline{\Theta}$ stands for the vector of average coefficients $\overline{\Theta}_i$ (or most probable values for this case) and $\sigma_{ii}$ are the the standard deviations, ie the square roots of the diagonal terms of the covariance matrix, given by:

$$\Sigma = \left[\sigma_{ij}\right] = \left[\frac{1}{N_b - 1} \sum_{k=1}^{N_b} (\Theta_i^{(k)} - \overline{\Theta}_i)(\Theta_j^{(k)} - \overline{\Theta}_j)\right] \quad (5)$$

In Eq. (5), $N_b$ stands for the total number of ressampling steps carried out. The correlations between coefficients is introduced by a multivariate normal law $\mathcal{N}(\overline{\Theta}, \Sigma)$:

$$\mathcal{N}(\overline{\Theta}, \Sigma) = \frac{1}{(2\pi)^{Nc/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}\left(\Theta - \overline{\Theta}\right)^T \Sigma^{-1} \left(\Theta - \overline{\Theta}\right)} \quad (6)$$

where $|\Sigma|$ stands for the determinant of the covariance matrix given by Eq. 5 and $N_c = N_\Theta \times N_T$ corresponds to the total number of model coefficients.

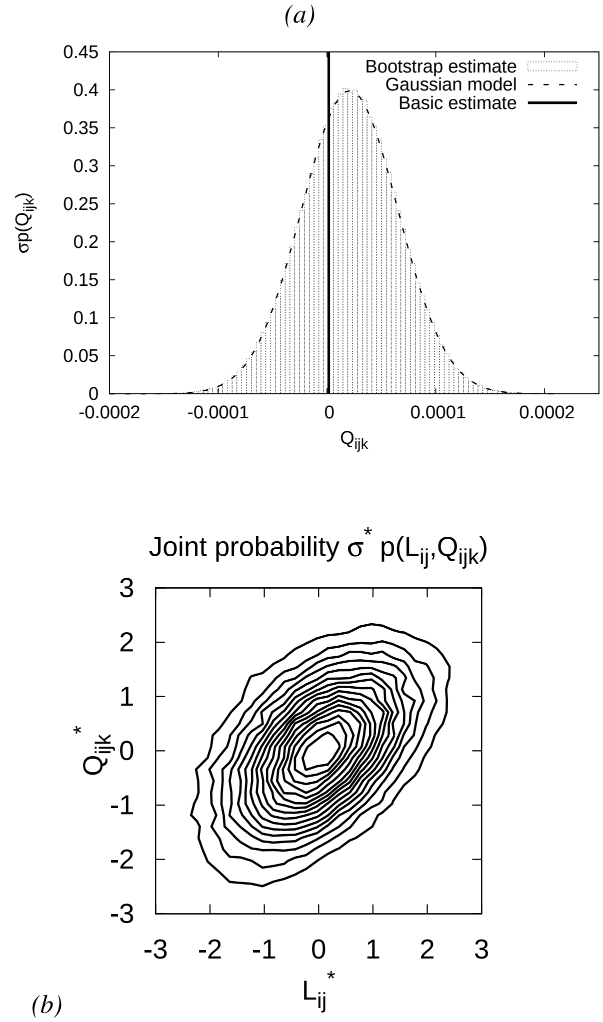


Figure 1. Example of stochastic modelling of the ROM coefficients for the cylinder wake flow: comparison between the bootstrap estimation of PDF, the gaussian model and the basic estimate for the coefficient corresponding to mode 1/mode 4 quadratic interaction ($(i,j,k) = (1,4,4)$) (*a*); example of a normalized joint probability distribution of reduced centered coefficients $L_{11}{}^*$ and $Q_{113}{}^*$ (isovalues from 0.01 to 0.16 by increment of 0.01) (*b*).

## Stochastic correction of the dynamical flow and Monte Carlo testing

In order to correct the dynamical flow, a stochastic selection of the model coefficients, respecting the multivariate normal law previously identified, can be carried out at any instant of the evolution of the LODS during time integration. Starting from sets of purely random uncorrelated pairs of numbers $(y_1, y_2)$, the classical method of Box and Muller (1958) is first used to make them follow the normal law, so that these numbers are recombined by following $y = \sqrt{-2\ln(y_1)}\cos(2\pi y_2)$. To obtain a random vector $Y = (y_1, \ldots, y_{N_T})^T$ whose components follow standard independant normal laws, this operation is simply marginally replicated for each component. The Cholesky decomposition of the covariance matrix of coefficients $\Sigma = LL^T$ is then computed to obtain the lower triangular matrix $L$. The inter-dependency and the centring of the model coefficients $\Theta$ can be thus finally obtained by a linear change of variable $\Theta = \overline{\Theta} + LY$.

The approach adopted for this study consists in considering that the dynamical flow remains locally uncertain and eventually requires real-time slight corrections, within the range of uncertainty evaluated, in order to ensure its existence at long term. Based on this stochastic generation of LODS coefficients, dynamical flow perturbations are used to generate an array of probable solutions at regular instants. This enables to scan different but nearly equiprobable short-term trajectories in phase space and to select *a priori* the trajectory which better respects a set of prescribed constraints. This procedure is illustrated in Fig. (2), where two trajectories are shown. In this example, both trajectories are drawn from the initial instant $t_0$ and enable a satisfying representation of the short-term evolution of the system (up to the mid-point $t_0 + T$), by comparison with the reference original signals. However, only one trajectory appears to respect the original dynamical flow topology in the following of the time integration up to a mid-term horizon $t_0 + 2T$, by keeping on evoluting within a bounded portion of the phase-space. Accordingly, only the LODS coefficients leading to an apparent stable trajectory are retained. For this preliminary study, the selection criterion simply consists in checking that the energy levels of the time signals should remain bounded within the reference phase subspace up to the prescribed time horizon.

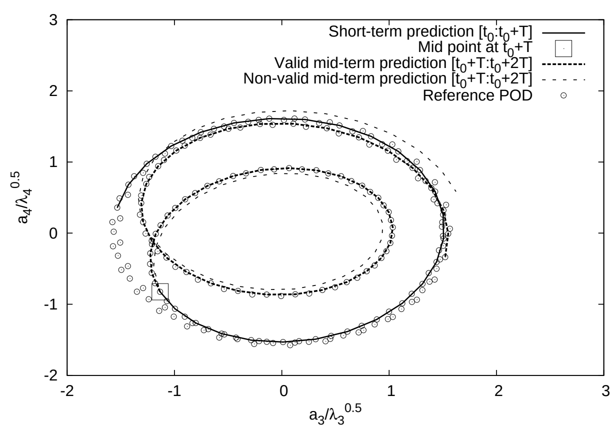


Figure 2. Principle of the Monte Carlo procedure retained.

# RESULTS

## ROM of a low-Reynolds cylinder wake flow

As a validation example, this strategy is first applied to model the direct numerical simulation data of a two-dimensional low-Reynolds cylinder wake flow studied by

Bergmann et al. (2005). This flow case has been retained by these authors as a prototype for separated flow. The "optimal" value chosen for the Reynolds number is slightly lower than 200 and corresponds about to the threshold value where a spanwise supercritical Hopf bifurcation occurs, associated with the emergence of three-dimensional effects. The 6 first POD modes have been considered, which is sufficient in this case to represent 99.9% of the turbulent kinetic energy. This simple flow case is already challenging and requires particular care to enable a stable long-term integration (Cordier et al., 2009). The present probabilistic calibration is based on 200 snapshots regularly sampled on about two periods of vortex shedding. The figure (3) illustrates the potential of the method to obtain a robust ROM. The phase trajectory of the POD coefficients in the phase plane $a_1 - a_2$ are here represented. Whereas such LODS obtained by classical flow calibration can be integrated in time, at best, up to a few dozens of pairing periods before blowing-up, it is shown here that the present stochastic LODS keeps on representing the expected dynamics for more than a thousand of pairing periods. In this case, the phase trajectories corresponds exactly to the limit cycles initially identified in phase planes, which fully demonstrate the viability of the approach for representating flow dynamics at long-term.

## ROM of a mixing layer

The present modelling approach is now applied on Large Eddy Simulation (LES) data of a spatially-growing transitional incompressible turbulent mixing layer flow performed in the following of the study of Comte et al. (1998). This LES has been performed with a prescribed Reynolds number equal to 150 (based on the initial vorticity thickness), at zero-molecular viscosity and with the aid of a Filtered Structure Function. The spatial developpement of the shear layer was triggered by an upstream modulation of a tanh streamwise profile by a translative instability. This reference simulation enables thus the observation of the initial shear flow dynamics with the growth of the convected instabilities, the formation of the Kelvin-Helmotz eddies and the regular pairing before the exit section of the computational domain. A thousand equidistantly sampled velocity snapshots from two thousand convective time units have been considered. The POD eigenspectrum of these data yields a typical stair-case pattern for the most energetic modes whose temporal evolution is dominated by a well defined main frequency. This constitutes the typical signature of the convection of coherent structures in the streamwise direction. In such a case, successive modes with similar energy yield a similar temporal evolution, but in quadrature. The first pair of modes holds the information of the growing Kelvin-Helmotz instability. The second pair mainly encompasses the information of vortex pairing. Thus, regular oscillations of the two first dominant modes can be observed, punctuated by regular transfer of energy towards modes 3 and 4 during the vortex pairing events. The relation between other neighboring modes is not yet well defined while the complexity of the temporal evolution of modes rapidly increases with the mode number. Any phase-shift or inaccuracy in the relative amplitudes is thus likely to lead rapidly to a loss of the spatio-temporal dynamics and to an unphysical behaviour of the corresponding surrogate flowfield.

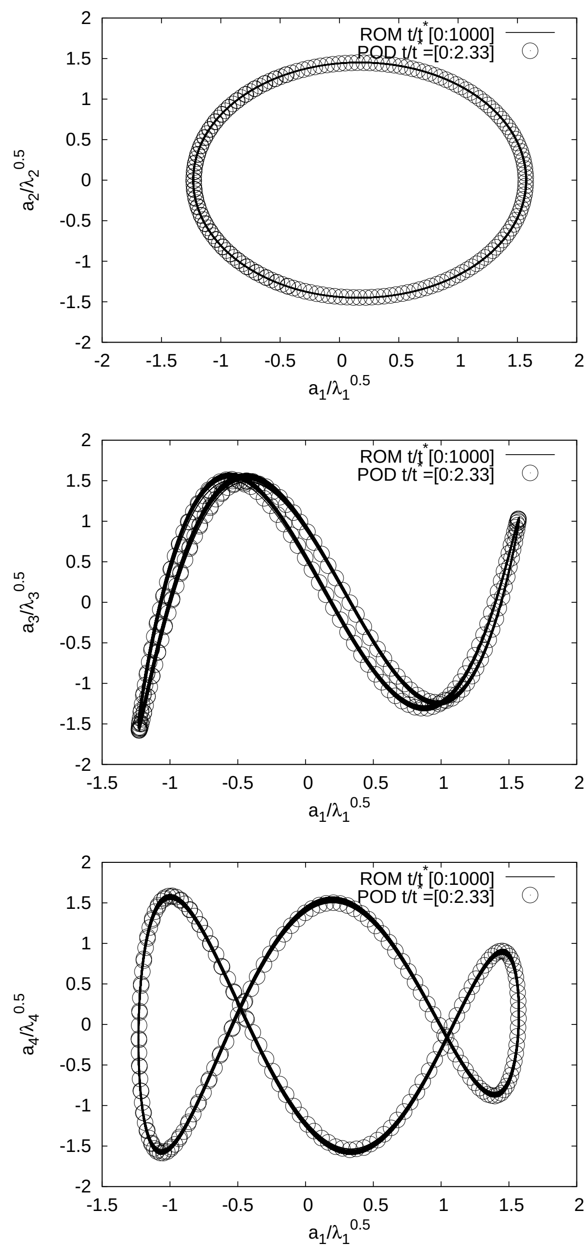


Figure 3. Stable model integration of a low-Reynolds cylinder wake: phase portraits $a_1 - a_2$, $a_1 - a_3$ and $a_1 - a_4$ of the POD coefficients predicted by the present model.

A stochastic ROM is calibrated based on the first 12 modes obtained by applying a snapshot POD on the three-dimensional LES data of the shear layer. This enables in this case to represent nearly 78% of the total turbulent kinetic energy of the whole simulated flow field. A characterization of the time signals obtained is first presented as a prerequesite to obtain a physically consistant model. As expected, the basic constraint used in the Monte-Carlo procedure does not allow for a fine selection of phase trajectories and the original temporal modes cannot be exactly followed during time integration. However, a correct phase-shift of modes working by pair is preserved and the expected regular energy decrease of

$(a_1, a_2)$ along with increase of the following pairs of modes, corresponding to pairing events, is well reproduced. This can be first observed for exemple by a comparison of phase portraits shown in Fig. 4 which fills the phase space in a similar way. The same drawing of an elliptical form with regular passage transitional passages near the center of the phase plane is observed in particular in phase space $a_1 - a_2$. The original and modelled power spectra of the second and seventh modes are given in Fig. (5) as exemples. They show that the dominant frequencies are also correctly reproduced. An artificial noise is typically observed in the high frequency domain of the ROM signals but its magnitude remains negligible by comparison with the peak amplitude. The secondary peaks of higher order modes in the low-frequency range appear also somewhat more difficult to reproduce.

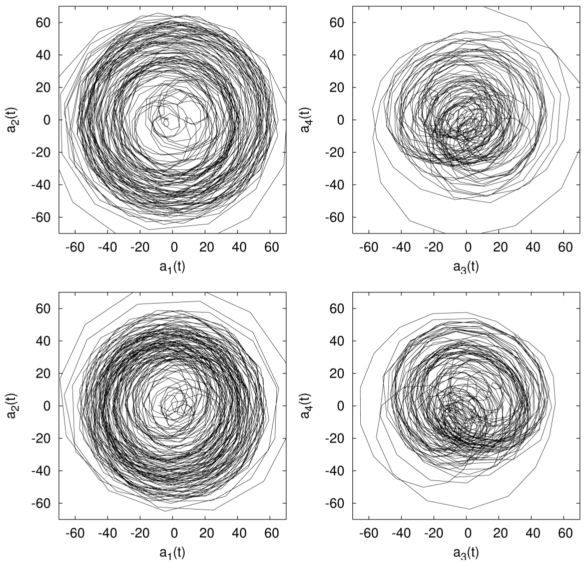


Figure 4. Comparison of original (top) and ROM (bottom).

The analysis of the energy levels, shown in Fig. (6) also confirms that the modes hierarchy is globally well preserved. It is worth noting nevertheless that these energy levels highly depend on the length of the temporal window of observation.

The realistic behaviour of the model is finally assessed by comparing the POD-truncated original flowfield with its surrogate obtained by summing the contribution of modes obtained by the ROM integration. The comparison of their respective animation reveals the model capacity to reproduce the main dynamical features. It enables not only to mimic the growth of Kelvin-Helmholtz rollers but is able to mimic their regular pairing as well as the growing of the secondary longitudinal instabilities. The figure (7) illustrates a snapshot of this animation of the global surrogate flowfield. It is worth noting that, as expected, using 12 modes only limites the representativity of form of eddies, in particular just after their pairing where the tridimensionalisation accelerates, involving a greater number of less energetic structures. In addition, the relation between higher modes is more difficult to determine and still requires improvements to limit the associated noise, observed at the shear layer edges. Howeve, the spatio-temporal correlations, shown in Fig. 8 for a probe located at the middle of the modelled mixing layer (half and two thirds of the transverse and streamwise extents $(L_y, L_x)$ of the computational domain respectively) clearly express that the main dynamical features are already correctly reproduced. The non-dimensionalisation used in this figure is based on the local vorticity thickness $\delta$ and the convective velocity $U$.

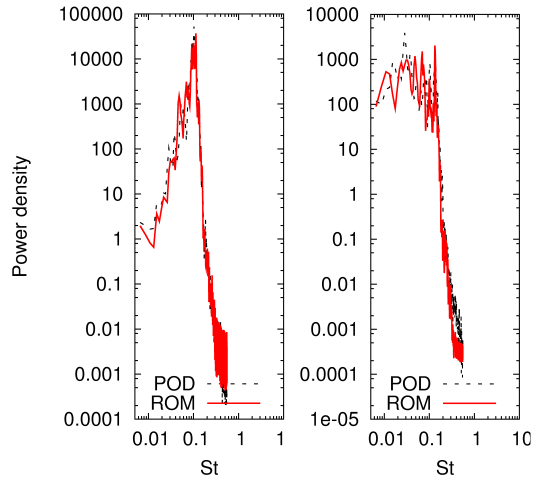


Figure 5. Power spectra of original POD-truncated and modelled coefficients $a_2$ and $a_7$.

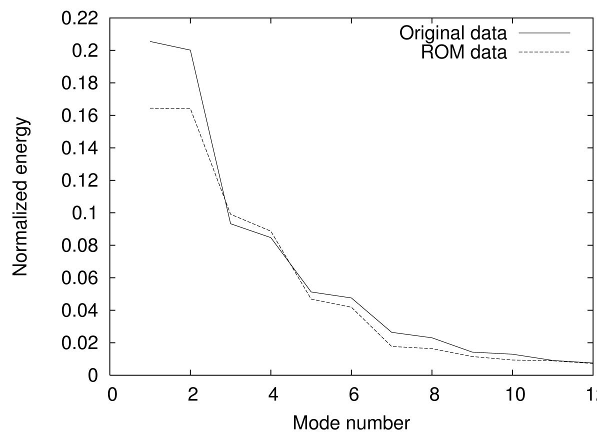


Figure 6. Original and modeled hierarchy of modes.

## CONCLUSIONS

A new modelling approach has been suggested in view of building robust and physical low-order representation of turbulent flowfield. It is based on the reformulation of classical POD-based LODS in a probabilistic framework, which enables dynamical flow perturbation and correction of phase trajectories during time integration. This approach is shown to enable the stabilization of a low-Reynolds cylinder wake flow and to lead to a correct representation of the main dynamics of a turbulent shear layer flow. The results obtained are all the more encouraging that they are only based on a very basic

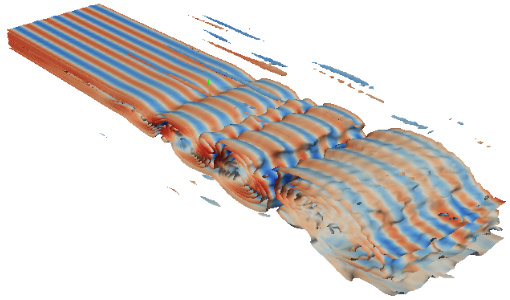

Figure 7. Snapshot of the flow rebuilt from the present 12 modes low-order model: isosurfaces of rotational magnitude colored by the longitudinal vorticity.

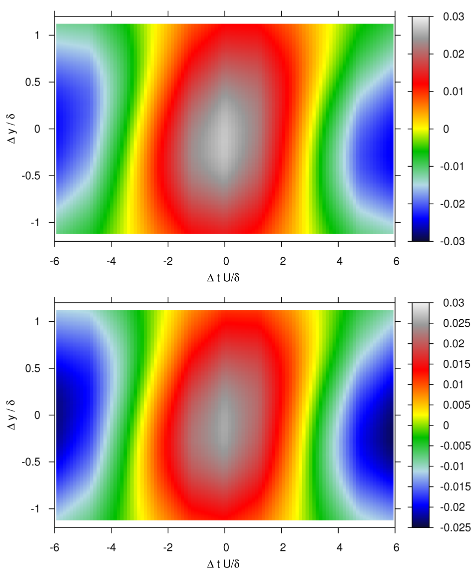


Figure 8. Transverse spatio-temporal correlation $R_{vv}(y,\Delta y,\Delta t)$ at the center of the mixing layer $(x,z) = (0.66L_x, 0.5L_z))$ rebuilt from POD signals (top) or model (bottom).

constraint (bounded energy levels) prescribed in order to select the most viable time evolution of the systems. The Monte-Carlo testing, based on more sophisticated LODS invariants, is likely to improve greatly the selection process. Various potential applications of this method may be considered. It could be in particular a good candidate for data/simulation coupling, improving the classical approaches to prescribe realistic boundary conditions for high-fidelity simulations.

## REFERENCES


Aubry, N., Holmes, P., Lumley, J. L. and Stone, E., 1988, ”The dynamics of coherent structures in the wall region of a turbulent boundary layer”, Journal of Fluid Mechanics, Vol. 192, pp. 115–173.

Bergmann, M., Cordier, L. and Brancher, J.P., 2005, ”Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model”, Physics of Fluids, Vol. 17.

Bergmann, M., Bruneau, C.-H. and Iollo, A., 2009, ”Enablers for robust POD models”, Journal of Computational Physics, Vol. 228, pp. 516-538.

Box, G. E. P. and Muller, M. E., 1958, ”A note on the generation of random normal deviates”, The Annals of Mathematical Statistics, Vol. 29, pp. 610-611.

Cazemier, W. and Verstapper, R.W.C.P. and Veldman, A.E.P., 1998, ”Proper orthogonal decomposition and low-dimensional models for driven cavity flows”, Physics of Fluids, Vol. 10, pp. 1685-1699.

Comte, P., Silvestrini, J.H. and Bgou, P., 1998, ”Streamwise vortices in large-eddy simulations of mixing layers”, Eur. J. Mech. B/Fluids, Vol. 17, pp. 615-637.

Cordier, L., Abou El Majd, B. and Favier, J., 2009, ”Calibration of POD reduced-order models by Tikhonov regularization”, Vol. 63, pp. 269-296.

Efron, B., 1979, ”Bootstrap methods: another look at the Jackknife”, Annals of Statistics, Vol. 7, pp. 1-26.

Gillies, E. A., 1998, ”Low-dimensional control of the circular cylinder wake”, Journal of Fluid Mechanics, Vol. 371, pp. 157–178.

Iollo, A., Lanteri, S. and Désidéri, J. A., 2000, ”Stability properties of POD-Galerkin approximations for the compressible Navier-Stokes equations”, Theoretical and Computational Fluid Dynamics, Vol. 13, pp. 377-396.

Karamanos, G.S. and Karniadakis, G.E., 2000, ”A spectral vanishing viscosity method for large-eddy simulations”, Journal of Computational Physics, Vol. 162, pp. 22-50.

Lorang, L., Podvin, B. and Le Quéré, P., 2006, ”Flow estimation using neural network”, $2^{nd}$ Flow Control Conference AIAA, 5-8 juin 2006, San Fransisco.

Noack, B.R. Afanasiev, K., Morzinski, M., Tadmor, G. and Thiele, F., 2003, ”A hierarchy of low-dimensional models for the transient and post-transient cylinder wake”, Journal of Fluid Mechanics, Vol. 497, pp. 335-363.

Noack, B.R., Schlegel, M., Ahlborn, B., Mutschke, G., Morzinski, M., Comte, P. and Tadmor, G., 2008, ”A finite-time thermodynamics of unsteady fluid flows”, Journal of Non-Equilibrium Thermodynamics, Vol. 33, pp. 103-148.

Perret, L., Collin, E. and Delville, J., 2006, ”Polynomial identification of POD based low-order dynamical system”, Journal of Turbulence, Vol. 7, pp. 1-15.

Rempfer, D. and Fasel, H.F., 1994, ”Evolution of three-dimensional coherent structures in a flat-plate boundary layer”, Journal of Fluid Mechanics, Vol. 260, pp. 351-375.

Sirovich, L., 1987, ”Turbulence and the dynamics of coherent structures”, Quarter. Appl Math. XLV, Vol. 3, pp. 561-590.