



HAL
open science

A data-driven low-order model of sound source dynamics in a turbulent jet

Nishant Kumar, Franck Kerhervé, Laurent Cordier

► **To cite this version:**

Nishant Kumar, Franck Kerhervé, Laurent Cordier. A data-driven low-order model of sound source dynamics in a turbulent jet. 2024. hal-04819475

HAL Id: hal-04819475

<https://hal.science/hal-04819475v1>

Preprint submitted on 4 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A data-driven low-order model of sound source dynamics in a turbulent jet

N. Kumar ¹ and F. Kerhervé ² and L. Cordier ³

Abstract

Building on previous efforts dedicated to the eduction of flow motions associated with low-angle sound emission from turbulent jets [25], the current paper proposes a strategy for the reduced-order dynamical modelling of these motions. The technique is data-driven and comprises (i) a dynamic *Ansatz* obtained by Galerkin projection of the Navier-Stokes equations onto a set of POD orthonormal spatial basis functions coupled with (ii) a dual-Kalman filter which allows long term prediction of the state and correction in line of the model's parameters. Rather than directly computing the coefficients of the model equation driving the flow state, which is only possible when full flow information is available, the coefficients are first identified thanks to a regularised L_2 minimisation procedure. A twenty degree-of-freedom reduced-order dynamical model (ROM) is then established. The identification scheme is found to lead to bounded ROMs but with estimated trajectories drifting from the true ones. A data-driven correction based on a dual-ensemble Kalman filter is therefore implemented and an empirical non-linear eddy viscosity model is added to the ROM to solve its inherent lack of stability. This correction allows the estimated state to be corrected "on-line" when a measurement of the true state (here point velocity and far-field pressure) is available, as well as to find the optimal set of parameters associated with the eddy viscosity term. The described approach, though validated here using partial flow information from a numerical simulation, has been designed in the perspective of being implemented in an experimental context.

¹Institut PPRIME CNRS UPR 3346, Université de Poitiers, ENSMA (France)

²Institut PPRIME CNRS UPR 3346, Université de Poitiers, ENSMA (France)

³Institut PPRIME CNRS UPR 3346, Université de Poitiers, ENSMA (France)

1. Introduction

Description of turbulent flows using stochastic distribution of eddies has long been served as a paradigm [27]. Further experimental evidences of the role played by well-coherent large structures in the dynamics have then considerably changed this point of view [15]. As far as jet flows are concerned, evidences that part of the noise radiated is essentially driven by these large-scale well-coherent structures has lead to significant progress in the understanding of noise generation and new methodologies for noise prediction have been derived. Following the original works of Michalke [30], Jordan and co-workers [21, 10] have proposed wavepackets *Ansatz* as the main source of jet noise. These entities may be seen as a low-order description of jet flows. More generally, assumption that turbulent flows can thus be meaningfully reduced to simplified kinematics pushes forward for low-dimensional modelling of their dynamics. Where noise radiated is concerned, it is however essential that the low-dimensional model also satisfies the inhomogeneous linear wave equation.

A key objective of low-dimensional modelling is to obtain a dynamical model with as few degrees of freedom as possible, this being generally motivated also by the need to perform real-time sensing and control. The number of degrees of freedom is dictated by the number of basis functions deemed necessary for accurate estimation of the dynamics. This raises the question regarding what kind of basis functions are best suited to a given problem. The empirical eigenfunctions obtained by Proper Orthogonal Decomposition (hereafter POD) are frequently used, largely on account of the optimality property of this kind of decomposition in the sense of energy, and will be used here. However, rather than applying it to the full flow solution, it is here applied to the reduced complexity sound-producing flow skeleton identified in a previous work [25]. This last work showed that the number of modes necessary to capture a given percentage of fluctuation energy of the acoustically-important coherent structures is an order of magnitude less than the number necessary to capture the same percentage of fluctuation energy of the full field. The POD modes identified can thus be considered as acoustically-optimised and are somewhat equivalent to the Most Observable Decomposition modes of Jordan et al. [22]. While the kinematic features of the sound-producing flow skeleton identified were fully described in Kerhervé et al. [25], we here sought for a low-dimensional dynamical system driving the associated motions. The conventional Galerkin projection method is here employed. Projection of the Navier-Stokes equations on a truncated POD basis results in a Galerkin system which consists in a low-dimensional system of ordinary differential equations (ODE)

with a manageable limited rank. In the present case, the corresponding reduced-order model (POD-ROM) is built with the twenty first POD modes obtained and whose linear combination was found to reproduce the dynamics of the essential sound-producing flow motions. The main difficulty resides in the calibration of the POD-ROM itself. This methodology is effectively known to result, by construction, to unstable models: after a relatively short time horizon, the estimated state drifts from the true trajectory and corrections are required to guaranty not only the stability of the POD-ROM but also that the estimated state converges towards the true value. Keeping only the lower POD modes in the truncation is equivalent to conserve only the large-scale coherent structures associated with the production of turbulent kinetic energy while omitting the small-scales fluctuations responsible with dissipation. The dissipation loss due to the mode truncation is generally considered as the main origin of the lack of stability of the POD-ROM which therefore needs to be amended. Common approaches fall into two classes of strategies. The first class is closure model dependent and comes from the pioneering work of Aubry et al. [1] who proposed to add dissipation terms to the POD-ROM thanks to an empirical “eddy viscosity”. Strategies of eddy viscosity type have been since refined by numerous authors considering either linear [36, 18, 38] or non-linear additional terms [14, 34, 37]. As the Reynolds number increases, non-linear subscale turbulence models are found mandatory. In such cases, Östth et al. [34] recently showed that only non-linear modal eddy viscosity can guarantee stabilisation and robustness of the POD-ROM. Spectral eddy viscosity models inspired from LES models have been also proposed [8, 32, 39]. Energy conservation concept has also been used to derived closure modelling strategies. This has been initially proposed by Cazemier et al. [11] and do not require the specification of any free parameter. The second class is closure model free. The most recent strategies are formulated as constrained optimisation problem resulting in POD-ROM where part of the small scales are included [2]. In contrast to the first class, this type of strategy has the benefit to guaranty the power balance in the low-dimensional space and do not require a subgrid-turbulence model.

The lesson from these various studies is that while stability of the estimated POD-ROM can be maintained, accurate tracking over long time horizons is generally not achievable despite large efforts in the calibration procedure. In the present case, the calibration technique of Perret et al. [35] combined with a Tikhonov regularisation [19] such as proposed by Cordier et al. [13] is considered to obtain in a first step a stable dynamical system. To improve the predictability of the POD-ROM over long time horizons, different strategies are then presented, among which a sequential data-assimilation technique known as dual Ensemble Kalman

Filter (DualEnKF). Thanks to nowadays computational capabilities, Kalman filters have regain new interest in the fluid mechanics community and have been used to develop data-driven simulation [24, 41] or again to solve data-driven estimation [26] problems. Kalman filters have the ability to improve the state estimate by assimilating available informations. Since in the present paper the interest is essentially focused on low-angle sound emission, the acoustic pressure being radiated downstream will be considered as observation.

The paper is organised as follows. In §2, the flow configuration and the main elements of the database are described. Elaboration of a bounded reduced-order dynamical model is then presented in §3 and its limitation to short-term prediction only is discussed in §4. Strategies for long-term prediction are presented in details in §5. Main characteristics of the reconstructed spatio-temporal flow field dynamics associated with low angle radiation over long a time horizon is addressed in §6. Conclusions and perspectives finally closes the paper in §7.

2. Flow configuration & previous work

The flow investigated is a Mach 0.9 single-stream jet with Reynolds number, based on jet diameter D and exit velocity U , of 4×10^5 , obtained from Large Eddy Simulation (LES) by Bogey and Bailly [6]. Details of the simulation as well as validation of the flow and sound properties can be found in Bogey et al. [7] and Bogey and Bailly [5, 6]. As mentioned in the introduction, we here focus on the flow dynamics responsible for noise being radiated downstream. In the work of Kerhervé et al. [25], a methodology has been developed in order to educe from the overall flow field the sound-producing events. The core idea is to use the radiated pressure field to filter out flow mechanisms not involved in noise production. Linear Stochastic Estimation (hereafter LSE) was used to reconstruct a time-resolved conditional space-time flow field associated with the low-angle sound emission extracted from the overall radiated pressure field thanks to an angle-dependent wavenumber-frequency filter. Figure 1(a) shows the LES solution, while figure 1(b) shows, for the same time step: in zone Ω_A , the low-angle component of the sound field, and, in zone Ω_F , the associated conditional pressure field (a conditional velocity field, not shown, is also computed).

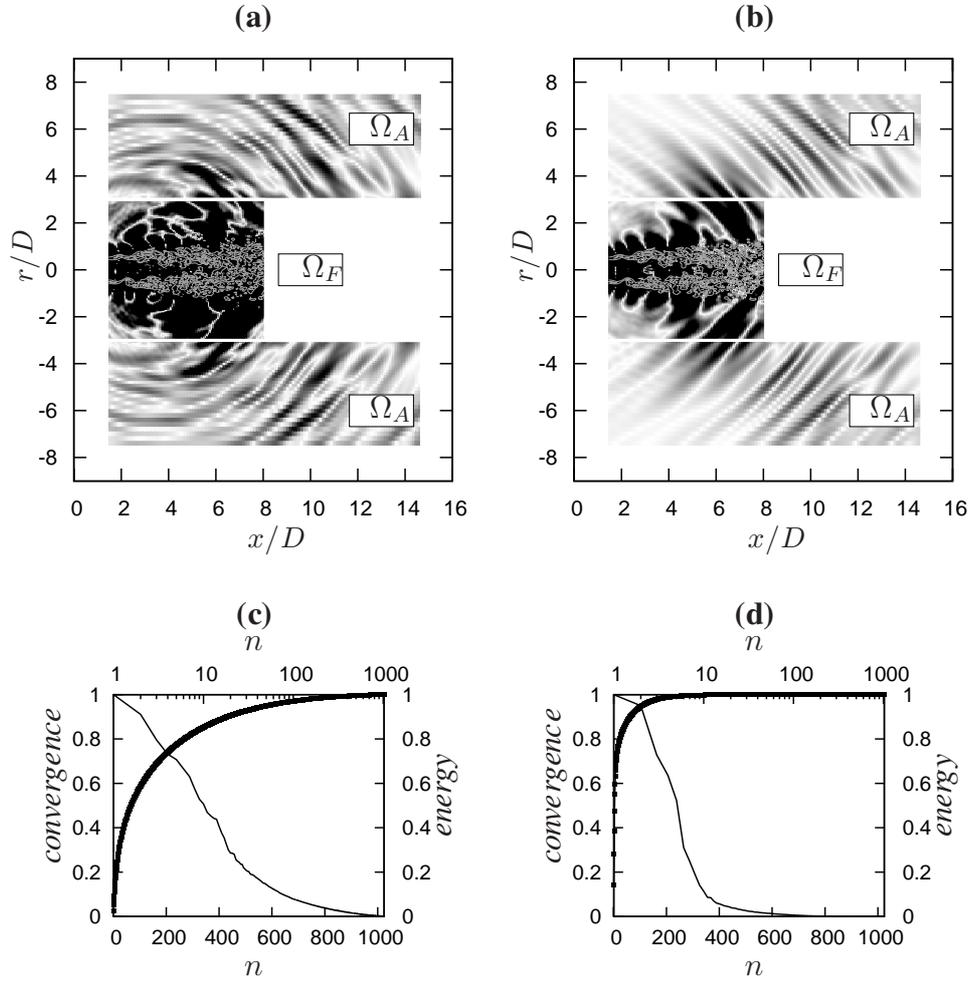


Figure 1: (a) Snapshot of vorticity and pressure fields of the LES solution. (b) Zone Ω_A : snapshot of the low-angle component of the LES sound field (obtained using a frequency-wavenumber filter); zone Ω_F : snapshot of pressure field associated with the sound-producing flow skeleton (this is a conditional field, computed by LSE from the information in Ω_A). (c) POD eigenspectrum (open symbols) and its convergence (solid symbols) of LES velocity fields corresponding to the sub-region Ω_F^{rot} of Ω_F (see (a) and the zoom given in figure 2). (d) POD eigenspectrum (open symbols) and its convergence (solid symbols) of the conditional field presented in (b).

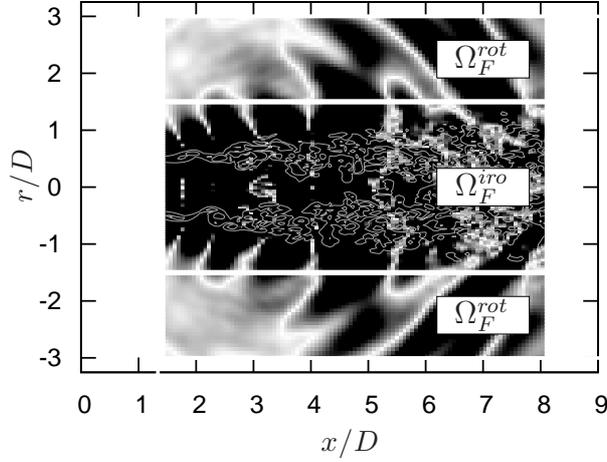


Figure 2: Zoom of region Ω_F , which is further broken down into rotational and irrotational regions, respectively Ω_F^{rot} and Ω_F^{irro} .

3. Elaboration of the dynamical system

3.1. Model equation

A reduced-order model describing the time evolution of the conditional field $\hat{\mathbf{q}}$, which can be here considered as the high-fidelity system, is sought. The Petrov-Galerkin projection method is implemented with POD modes as the basis of projection and the Navier-Stokes equations as the high-fidelity model equation. This results in a system of ordinary differential equations (ODE) with, in contrast to the Navier-Stokes equations, finite dimension. Unfortunately, these models are known to be intrinsically unstable mainly due to mode truncation. For this reason, the calibration method described by [13] is here implemented and is described in the following sections.

3.1.1. Proper Orthogonal Decomposition

Let $\{\mathbf{u}(\mathbf{x}, t_k)\}_{k=1, \dots, N_s}$ be a set of N_s velocity snapshots equally spaced over a time interval T_s with $\mathbf{x} \in \Omega_F^{rot}$. The velocity field can be decomposed onto an orthonormal basis including spatial basis functions and temporal coefficients $\{\Phi_i^p(\mathbf{x}), a_i^p(t)\}_{i=1, \dots, N_s}$ such that the normalised mean-square projection of the basis functions on the velocity field is maximised [29] and is given by,

$$\mathbf{u}(\mathbf{x}, t) = \sum_{i=1}^{N_s} a_i^p(t) \Phi_i^p(\mathbf{x}) \quad (1)$$

When the "POD snapshot" method [40] is implemented, the maximisation problem leads to an eigenvalue problem which reads as,

$$\mathcal{C}\mathbf{a}_i^p = \lambda_i \mathbf{a}_i^p \quad (2)$$

where superscript p relates with temporal POD coefficients obtained from the high-fidelity system (given here by the N_s velocity snapshots) and where $\mathcal{C} \in \mathbb{R}^{N_s \times N_s}$ is a correlation matrix averaged over the spatial direction whose elements can be expressed as,

$$\mathcal{C}_{ij} = \frac{1}{N_s} \langle \mathbf{u}(\mathbf{x}, t_i), \mathbf{u}(\mathbf{x}, t_j) \rangle_{\Omega_F^{rot}} = \sum_{k=1}^{n_c} \iint_{\Omega_F^{rot}} u_k(\mathbf{x}, t_i) u_k(\mathbf{x}, t_j) d\mathbf{x}, \quad (3)$$

with n_c the number of velocity components considered (here $n_c = 2$). In practice, when solving the eigenvalue problem (eq. 2), the temporal POD coefficients obtained are normalized so that the variance of each is equal to unity. Formally, these coefficients should be properly rescaled so that their variance be equal to the corresponding eigenvalue. However, for what follows, this fully suits the numerical constraints of linear regression problem.

In the following, the objective is to build a reduced-order model allowing long-term prediction. The overall database includes a total number of $N_T = 19000$ snapshots sampled at a Strouhal number based on the jet diameter of $St_D = 3.9$ (corresponding to a total duration of $tU/D = 4900$). The snapshot POD method recalled previously is here applied for a number of $N_s = 4096$ snapshots randomly sampled overall the database. Solving Eq. (2) results in N_s temporal POD modes, randomly sampled over time, which, when projected onto the initial snapshots randomly selected lead to N_s spatial eigenfunctions $\Phi_i(\mathbf{x})$. In order to build a ROM and to compare the estimated temporal POD modes overall the duration of the database, the eigenfunctions are reprojected into $u(\mathbf{x}, t)$ to obtain a set of N_s temporal POD functions sampled at the same rate than the velocity snapshots.

The POD eigenspectra obtained, respectively, for the LES field and that of the conditional acoustically-filtered velocity field are reported in Figure 1(c) & (d). While 20 modes suffice to recover 80 % of the fluctuation energy for the conditional field, over 200 are required to retain the same amount of energy for the LES field. As discussed in details in Kerhervé et al. [25], the acoustically-filtered field is found to comprise considerably lower degrees of freedom and suggest that a low-rank model can be found to properly estimate its dynamics.

3.1.2. POD reduced-order model based on Galerkin projection

A Galerkin projection of the Navier-Stokes equations onto the POD basis defined previously is then effected. After manipulations, this leads to a set of ordinary differential equations (ODEs) which reads as [20, 4, 13],

$$\dot{a}_i^P(t) = D_i + \sum_{j=1}^{N_{\text{gal}}} L_{ij} a_j^P(t) + \sum_{j=1}^{N_{\text{gal}}} \sum_{k=1}^{N_{\text{gal}}} Q_{ijk} a_j^P(t) a_k^P(t) + \mathcal{R}_i(t) \quad i = 1, \dots, N_{\text{gal}}, \quad (4)$$

where $\mathbf{y}_i = (D_i, L_{i1}, \dots, L_{iN_{\text{gal}}}, Q_{i11}, \dots, Q_{iN_{\text{gal}}N_{\text{gal}}})^\top \in \mathbb{R}^{N_{\mathbf{y}_i}}$ denotes an unknown vector of real coefficients, \dot{q} denotes the time derivative of q and \mathcal{R}_i the residual term which should be read as,

$$\mathcal{R}_i(t) = \sum_{j=N_{\text{gal}}+1}^{N_s} L_{ij} a_j^P(t) + \sum_{j=1}^{N_s} \sum_{k=1}^{N_s} Q_{ijk} a_j^P(t) a_k^P(t) \quad \text{for } \max(j,k) > N_{\text{gal}}+1 \quad (5)$$

As first approximation, we here consider the case where $\mathcal{R}_i = 0$ which is equivalent to neglect convective and viscous terms associated with unresolved modes $i > N_{\text{gal}}$. The POD-reduced order model can consequently be written as,

$$\dot{a}_i^R(t) = D_i + \sum_{j=1}^{N_{\text{gal}}} L_{ij} a_j^R(t) + \sum_{j=1}^{N_{\text{gal}}} \sum_{k=j}^{N_{\text{gal}}} Q_{ijk} a_j^R(t) a_k^R(t) \quad i = 1, \dots, N_{\text{gal}}, \quad (6)$$

where $a_i^R(t)$ denotes the estimated i -th temporal coefficient.

The coefficients D_i , L_{ij} and Q_{ijk} can be computed directly if full flow information is available, for instance from a sufficiently resolved numerical simulation. However, in situations where only incomplete flow information is available, such as typically in an experiment and in the case of incomplete numerical data, the coefficients must be determined by alternative means. This is discussed in §3.2.

Introducing,

$$\mathbf{a}^R(t) = \left(a_1^R(t), \dots, a_{N_{\text{gal}}}^R(t) \right)^\top$$

Eq (4) can be rewritten as,

$$\dot{a}_i^R(t) = f_i(\mathbf{a}^R(t), \mathbf{y}_i) \quad (7)$$

where f_i is the model equation for the i -th mode. The dynamical system (7) may be written in the more compact form,

$$\dot{\mathbf{a}}^R(t) = \mathbf{f}(\mathbf{a}^R(t), \mathbf{y}) \quad (8)$$

with $\mathbf{f} = (f_1, \dots, f_{N_{\text{gal}}})^\top \in \mathbb{R}^{N_{\text{gal}}}$ and $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{N_{\text{gal}}})^\top \in \mathbb{R}^{N_{\mathbf{y}}}$ the overall set of unknown coefficients where $N_{\mathbf{y}} = N_{\text{gal}}N_{\mathbf{y}_i}$.

In the following sections, this compact form is conserved to discuss the identification procedure for the coefficients vector \mathbf{y}_i .

3.2. Model identification

A number of studies have been performed with regards to identification of the model coefficients. These methods differentiate on the definition of the error to minimise. An intuitive choice is the error on the time derivative of the temporal coefficients since the objective of the reduced model is to predict as accurately as possible the dynamics of the system. This has been initially proposed by Perret et al. [35] and revisited by Cordier et al. [13] or more recently by Suzuki [42]. Here, the flow calibration method of Cordier et al. [13] is presented.

Let $e(\mathbf{y}, t)$ be the error function defined by,

$$\mathbf{e}(\mathbf{y}, t) = \dot{\mathbf{a}}^P(t) - \mathbf{f}(\mathbf{a}^P(t), \mathbf{y}), \quad (9)$$

The flow calibration procedure consists in minimising the cost functional $\mathcal{I}(\mathbf{y}) = \langle \|\mathbf{e}(\mathbf{y}, t)\|_2^2 \rangle_{T_s}$, where $\|\cdot\|_2$ denotes the L_2 norm, and $\langle \cdot \rangle_{T_s}$ a time-average operator defined as

$$\langle q(t) \rangle_{T_s} = \frac{1}{N} \sum_{k=1}^N q(t_k) \quad (10)$$

where $q(t_i)$ denote N discrete values of q equally distributed over time period $[0, T_s]$.

Equivalently, the functional \mathcal{I} can then be written as,

$$\mathcal{I}(\mathbf{y}) = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^{N_{\text{gal}}} [\dot{a}_i^P(t_k) - f_i(a_i^P(t_k), \mathbf{y})]^2. \quad (11)$$

Minimisation of \mathcal{I} leads to the following linear system [13],

$$A\mathbf{y} = \mathbf{b} \quad \text{with} \quad A \in \mathbb{R}^{N_{\mathbf{y}} \times N_{\mathbf{y}}} \quad \text{and} \quad \mathbf{b} \in \mathbb{R}^{N_{\mathbf{y}}}, \quad (12)$$

where,

$$A = \langle E^\top(t) E(t) \rangle_{T_s} \quad \text{and} \quad \mathbf{b} = -\langle E^\top(t) \mathbf{e}(\mathbf{0}, t) \rangle_{T_s}. \quad (13)$$

The time-dependent array, $E(t) \in \mathbb{R}^{N_{\text{gal}} \times N_{\mathbf{y}}}$, is defined as

$$E(t)\mathbf{y} = -\mathbf{f}(\mathbf{a}^P(t), \mathbf{y}). \quad (14)$$

The minimisation procedure is therefore equivalent to solving the linear system given in Eq. (12), where the unknown vector contains the coefficients of the dynamical system. Note that this procedure is similar to that proposed by Perret et al. [35] except that here the coefficients of the entire ODE system are calibrated in a one-shot procedure and is computationally more efficient.

3.3. Regularisation of the solution

The linear system Eq.(12) is generally ill-conditioned. It can be easily understood since the right-side term of the equation, \mathbf{b} , may be contaminated through $\mathbf{e}(\mathbf{0}, t) = \dot{\mathbf{a}}^P(t)$ by approximation errors related to the numerical evaluation of the time-derivatives of the temporal POD coefficients. Here, a 2nd-order centered finite-difference scheme is used for estimating these time derivatives.

To remedy to this potential ill-conditionness, a regularisation procedure is therefore necessary. In this work, a Tikhonov regularisation of zero-th order is implemented [19]. The basic idea of Tikhonov regularisation is to find a solution \mathbf{y}_ρ of the linear system Eq.(12) that minimizes the residual $\|A\mathbf{y}_\rho - \mathbf{b}\|_2$ without penalizing too much the value of $\|\mathbf{y}_\rho\|_2$. This is equivalent to minimise the following functional,

$$\phi_\rho(\mathbf{y}) = \|A\mathbf{y} - \mathbf{b}\|_2^2 + \rho\|\mathbf{y}\|_2^2, \quad (15)$$

where ρ is a regularisation parameter to be determined.

To understand the influence of the ill-conditionness of Eq.(12) on the solution of the linear system, the concept of filter factors is now introduced. To do so, the singular value decomposition of A is considered, and can be read as,

$$A = U\Sigma V^\top = \sum_{i=1}^{N_{\mathbf{y}}} \mathbf{u}_i \sigma_i \mathbf{v}_i^\top, \quad (16)$$

where U and V are orthogonal matrices containing left, \mathbf{u}_i , and right, \mathbf{v}_i , singular vectors, and where σ_i are the singular values of A arranged in decreasing order. The solution for \mathbf{y} can thus be written

$$\mathbf{y} = \sum_{i=1}^{N_{\mathbf{y}}} h_i \frac{1}{\sigma_i} \mathbf{u}_i^\top \mathbf{b} \mathbf{v}_i \quad \text{with} \quad h_i = 1 \quad \text{for} \quad i = 1, \dots, N_{\mathbf{y}}. \quad (17)$$

In this expression, some additional factors, h_i , have been artificially introduced. When these factors are fixed to unity, $|\mathbf{u}_i^\top \mathbf{b}|$ may not decrease sufficiently fast compared to $|\mathbf{u}_i^\top \mathbf{b}/\sigma_i|$ when σ_i becomes small. This can result in a solution with a large Euclidean norm [19, 13]. To minimise the contribution of modes i associated with small σ_i to the summation defined in Eq.(17), these components are "low-pass filtered" by modifying the value of h_i . Given a value of ρ , the solution of the linear system can be expressed by Eq.(17) with filter factors equal to,

$$h_i = \frac{\sigma_i^2}{\sigma_i^2 + \rho^2}. \quad (18)$$

Different strategies may be used to determine the parameter ρ . In the present case, the L-curve method described in Hansen [19] is used. This consists in an iterative procedure to identify $\mathbf{y}_\rho = \arg \min_{\mathbf{y}} \phi_\rho$ which balances values of the two residuals $\|A\mathbf{y}_\rho - \mathbf{b}\|_2^2$ and $\|\mathbf{y}_\rho\|_2^2$.

4. Bounded reduced-order model

4.1. Minimisation of cost functional \mathcal{I}

As discussed previously, identification of the dynamical model given by Eq.(8) is equivalent to solve the linear system Eq.(12). The matrix A and vector \mathbf{b} are evaluated according to Eq. (13). Performance of the minimisation procedure is evaluated in figure 3 where time derivative \dot{a}_i^P of the POD temporal coefficients compared to that obtained by the identified dynamical model $\mathbf{f}(\mathbf{a}^P(t), \mathbf{y})$ for the first and fourth POD modes are reported. Identification of \mathbf{f} is in good agreement with the original data and, as expected, very similar results are observed without and with application of the Tikhonov regularisation. However, as we will see in the next section, the very small differences that do exist between the two estimates have important consequences when the dynamical system is integrated in time.

4.2. Short-term prediction

Once the coefficients D_i , L_{ij} and Q_{ijk} have been determined, temporal integration of the dynamical system Eq. (4) is performed using a 4-th order Runge-Kutta scheme. The initial condition at $t = 0$ corresponds to the initial sample of the data sequence considered for the POD analysis. The time step used for the integration is equal to the time resolution of the data described in §3.

Two solutions are obtained: one where the coefficients of the model are calibrated using Tikhonov regularisation, and a second where no regularisation is

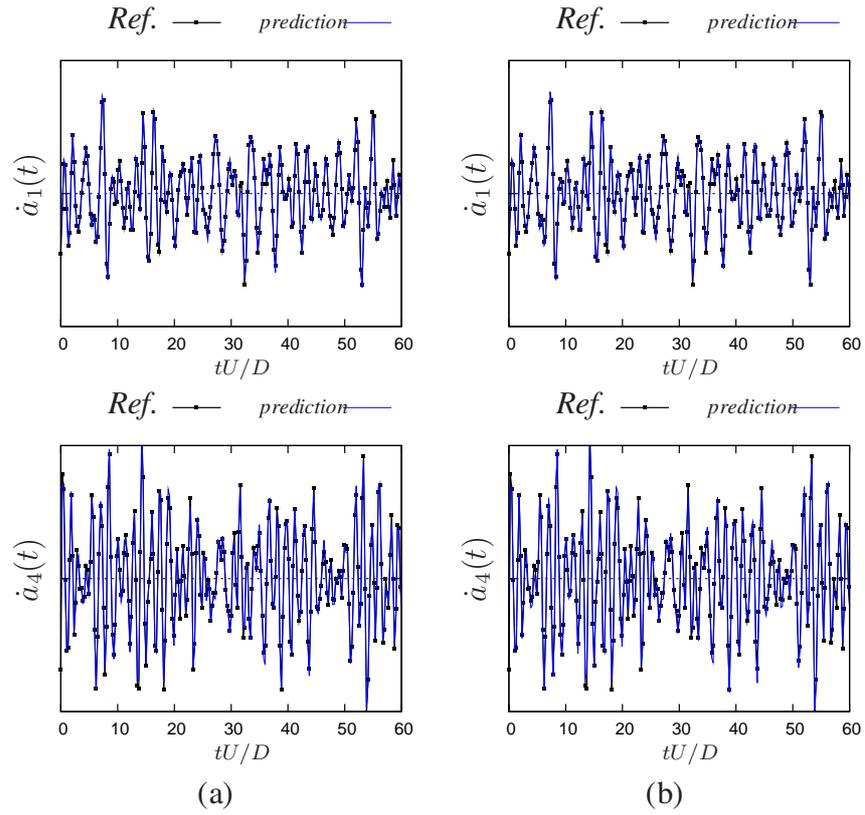


Figure 3: Visual evaluation of error function $e(\mathbf{y}, t)$: comparison between (line with dots) time derivative of POD temporal coefficient $\dot{a}_i^P(t)$ and (blue line) identified model $f_i(\mathbf{a}^P(t), \mathbf{y})$ for (top) mode 1 and (bottom) mode 4. (a) Without and (b) with Tikhonov regularisation.

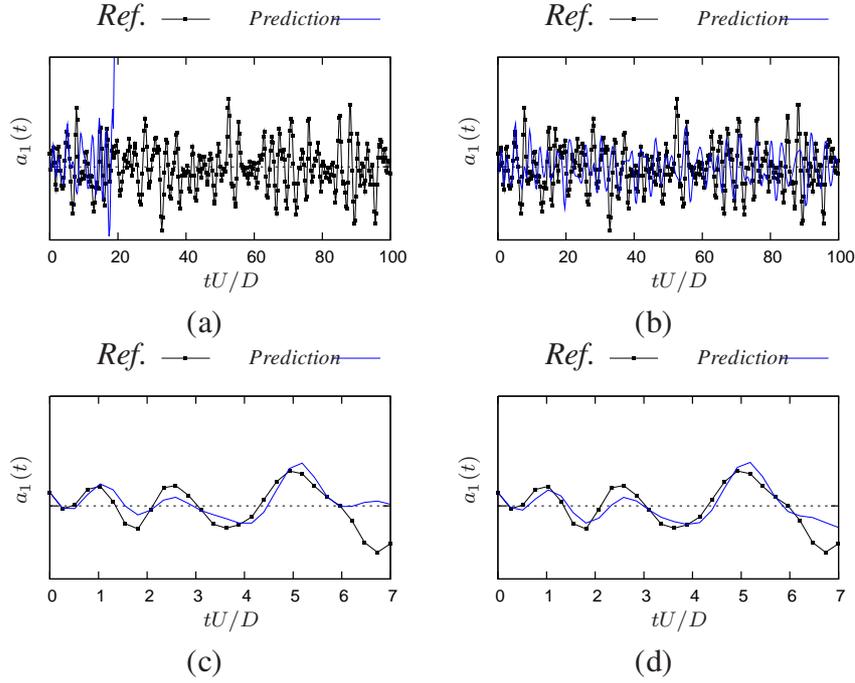


Figure 4: Time histories of the first POD temporal coefficients obtained directly from POD (line with dots) and predicted by the dynamical model (blue line) without (a,c) and with (b,d) Tikhonov regularisation. (top) Time histories over 100 time units and (bottom) zoom over the first 7 time units.

used. Time histories of the corresponding first temporal POD modes are shown in figure 4 for both solutions. The dynamical system obtained without regularisation is unstable and diverges after 20 time units, whereas the other is stable. Both systems nonetheless provide good predictions beyond a short time horizon. The stable system begins to drift from the trajectory described by the original data after about six convective time units as illustrated in figures 4(c & d). This result is not surprising as discussed in the introduction and as reported by many others. In the following sections, different strategies for long-term predictions will thus be considered.

5. Strategies for long-term predictions

5.1. Dependence of the initial condition and intrinsic stabilisation

It is of practical interest to establish if the identified dynamical system can give correct predictions when the time integration is started in different points on the

POD attractor. As illustrated in Figure 5 where estimated temporal coefficients $a_1(t)$ is illustrated when starting the time integration at different time steps with, as initial condition the true value, the dynamical model performs equally well. Furthermore, when an initial condition which was not contained in the data set for the identification procedure is used, as it is the case for figure 5(d), the model continues to provide correct prediction over the short time horizon. This is of considerable importance because it means that the model coefficients identified from a given set of realisations (learning sequence) can be reemployed to predict another sequence of realisations. Note that this remains valid only if the matrices A and b remains statistically well representatives of the new realisations. In the present case, the results obtained are an indication that these two matrices are sufficiently well estimated and that a longer learning data sequence would not have been necessary.

Since a correct prediction can be obtained on a short term duration by assimilating a new initial condition, a basic procedure for long-term prediction is now evaluated. This procedure consists in injecting new initial conditions every 6 time units. This procedure is somewhat similar to the intrinsic stabilisation scheme proposed by Kalb and Deane [23]. The results are shown in figure 6 for the first POD temporal coefficient (similar results are obtained for the higher modes). The long-term prediction is found as expected to be satisfactory. This procedure is however of limited practical interest. It requires temporal POD coefficients to be estimated when a new initial condition is necessary. This can be obtained only if velocity snapshots at the corresponding instants are available from measurements. In addition, measurement noise or again incoming perturbation may deteriorate the prediction since the model is not designed to take into account such deviations from the original trajectory. In the next section, we thus introduce additional correction to the model in perspective of long-term prediction to take into account for such deviations.

5.2. Long-term prediction with data assimilation

Despite efforts in the calibration procedure, low-order dynamical systems inherently drift from the true trajectory when integrated in time if no correction is applied. The model equation do not perfectly drives the real flow system examined as already mentioned. Injecting periodically a new initial condition as performed previously is the zero-th step of more advanced data assimilation technique. Suzuki [42] recently proposed a data-driven approach where a proportional feedback term is added to the model equation. When the gain of the feedback term is properly selected –typically of the order of the larger coefficient of the

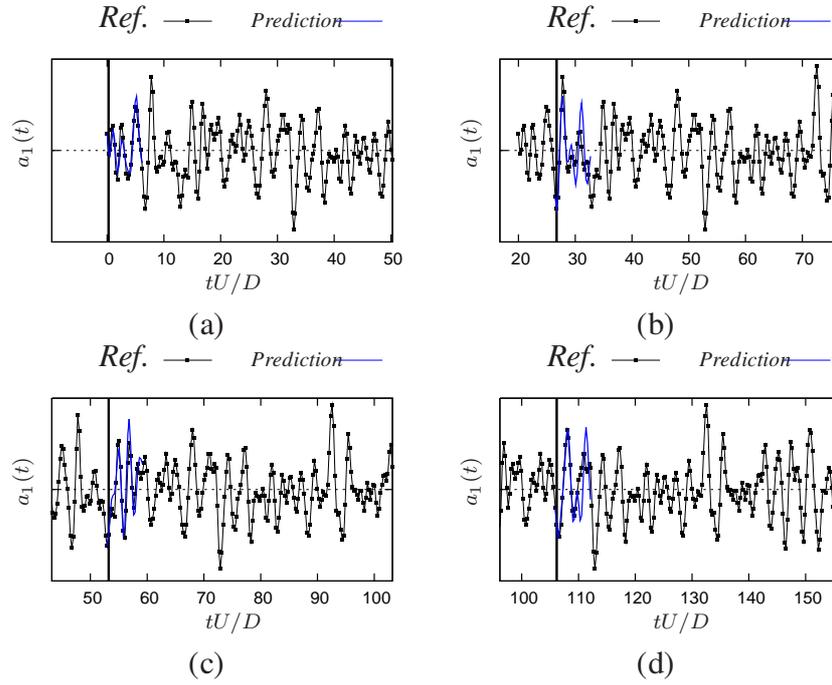


Figure 5: Time histories of the first POD temporal coefficients (line with dots) and its prediction (blue line) obtained from different initial conditions.

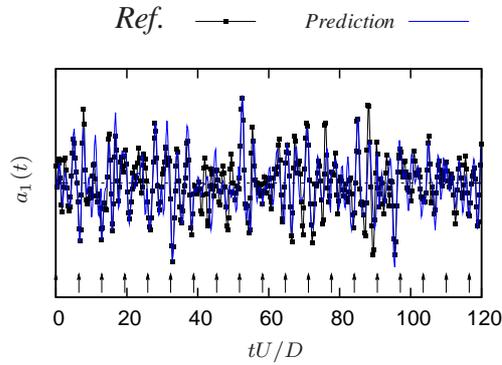


Figure 6: Time histories of the first POD temporal coefficients (line with dots) and its prediction (blue line) obtained by intrinsic stabilisation. Vertical oriented arrows indicate the time steps where the initial condition is updated by observing the original POD coefficient.

parameter vector– the estimates is found well robust. This approach and the one previously proposed, have disadvantage to require perfect knowledge of the real trajectory of the state vector. In practice, it may not be possible to access easily

on-line to this quantity. In contrast, the user may instantaneously access to “observable” quantities such as radiated pressure or again velocity at given locations. Among the range of sequential data assimilation techniques used for example in meteorology for long-term weather prediction, we here introduce the Ensemble Kalman filter (EnKF) detailed in Appendix A. The later is an extended version of the more conventional Kalman Filter (KF) and is appropriate for system described by non-linear model equation. Unlike the KF, the EnKF may be interpreted as the statistically suboptimal estimate of the underlying system state providing given noisy and/or inaccurate observations. The EnKF generates an analysis ensemble, that is, an ensemble of model states that reflects both an estimate of the true state (through its mean) and the uncertainty of this estimate (through its spread). This probabilistic approach allows to deal with non-linear model equations and the final estimated state vector may be seen as a higher order linearisation of the true state. Like the KF, a recursive resolution is employed. At each time step, a model equation is first used to generate an *a priori* (or forecast) background ensemble of estimated states and then the available observations are assimilated to build the *a posteriori* (or analysis) ensemble thanks to an optimal filter. Extensive details on the EnKF can be find in the literature (see Evensen [17] for a thorough review). Its implementation thus requires a model equation of the system state, here provided by the reduced-order model of Eq. (8), and an observation equation relating the state vector and the measurements available. For the observation equation, we propose to start here with a fictive velocity sensor located at $x/D = 3$ & $y/D = 0.5$ and a fictive pressure sensor in Ω_F^{rot} located at $x/D = 7$ & $y/D = 2.5$ for local surveys of the velocity and the radiated pressure (hereafter referred to $u(t)$ and $p(t)$ respectively). This choice is motivated by the perspective of a practical implementation. The observation equation can thus be read, in a discrete form, as,

$$\mathbf{z}(t_k) = H\mathbf{a}^R(t_k) \quad \text{with} \quad \mathbf{z}(t_k) = \begin{bmatrix} u(t_k) \\ p(t_k) \end{bmatrix} \in \mathbb{R}^2 \quad (19)$$

Matrix $H \in \mathbb{R}^{2 \times N_{gal}}$ maps the estimated state vector $\mathbf{a}^R(t)$ to the available observation vector $\mathbf{z}(t)$. Its first row contains the values of the spatial eigenfunctions, calculated *a priori*, at the sensor location. For the mapping between the estimated temporal POD coefficients and the pressure in the radiated field, we here propose to use a linear stochastic estimate of the pressure field and which can be written as,

$$\tilde{p}(t + \tau_{ac}) = \sum_{i=1}^{N_{gal}} \alpha_i a_i^R(t) \quad (20)$$

where $\{\alpha_i\}_{i=1, N_{gal}}$ defines a set of coefficients which minimises $\|p(t) - \tilde{p}(t)\|_2^2$ in the L_2 norm sense. The time-delay τ_{ac} introduced in the last expression takes into account for acoustic propagation time. This quantity needs to be optimised but we will not discuss this point here. The second row of the observation matrix H contains the values of the coefficients α_i . Finally, the non-linear discrete system considered for application of the EnKF may thus be read as,

$$\begin{cases} \mathbf{a}^R(t_k) &= \mathbf{F}(\mathbf{a}^R(t_{k-1}), \mathbf{y}) + \mathbf{u}_k \\ \mathbf{z}(t_k) &= H\mathbf{a}^R(t_k) + \mathbf{v}_k \end{cases} \quad (21)$$

where \mathbf{u}_k & \mathbf{v}_k are white zero-mean uncorrelated process and observation noises respectively with known covariance matrices. These noise components are implemented to take into account for errors in the model equations and errors in the measurements. These can also be interpreted as levels of trust in both the model and the available observations. The time-discrete form of the model equation used in the implementation of the EnKF is here obtained thanks to a 4-th order Runge-Kutta scheme applied to Eq. (8). The white gaussian noises are initialised identically as $\mathcal{N}(0, 0.01\mathbb{I}_{N_{gal}})$ and a number of 500 members is chosen for the state ensemble.

Results of the procedure are illustrated in figure 7 for the first temporal POD mode. The top part of the figure shows the predictive variable $\mathbf{z}(t)$ obtained with the observation equation of Eq. (21) allowing for measurement noise, while the bottom part shows the time history of the estimated first POD coefficient. In contrast with intrinsic stabilization or feedback procedures discussed above, it is here important to point out that values of the true temporal POD coefficients are not used. Only the observations are assimilated. When comparing the results of figure 5 obtained without assimilation, the time history of the temporal POD coefficient is now well recovered over a long time duration while the assimilation procedure is maintained. The error of the reconstruction is found to slightly increase with the mode number. In order to increase the level of accuracy of the prediction, additional fictive sensors can be used for observation. Figure 8 shows the results of the assimilation procedure for the first four temporal POD coefficients when adding velocity sensors distributed along the mixing layer axis ($y/D = 0.5$) between $x/D = 2$ and $x/D = 6.5$ by step of $0.5D$. The estimates obtained without assimilation are also reported for comparison. As illustrated, time histories of the temporal coefficients are well reproduced with very small discrepancies. While the trajectory estimated without assimilation rapidly drifts from the correct one, that estimated with the EnKF fits well with the original data thanks only to a limited number of observations and a stable model equation.

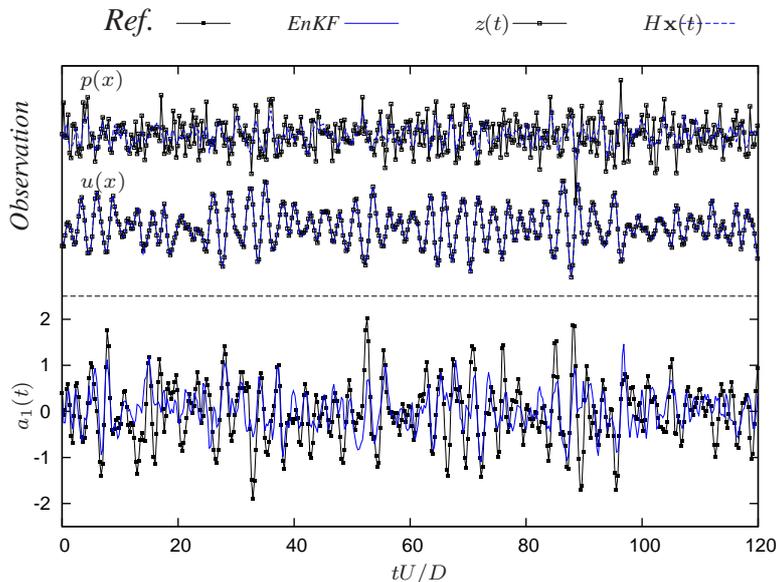


Figure 7: (Bottom part) Time histories of the first POD temporal coefficient (line with dots) and its prediction using the EnKF (blue line). (Top part) Time histories of the observations and their estimates through 19 used for assimilation.

5.3. Dual parameters & state estimation

As mentioned in the introduction, the drift of POD-ROMs typically observed is believed to be due to the omission of the residual term \mathcal{R}_i in Eq. 4. Empirical subscale linear and non-linear models have been proposed in the literature[34]. The modal linear eddy viscosity model proposed by Östth et al. [34] is here considered, mainly to demonstrate the potential of the assimilation scheme detailed below. The residual term in Eq. (4) is hence written as,

$$\mathcal{R}_i(t) = \nu_i^t \sum_{j=1}^{N_{\text{gal}}} L_{ij} a_j^R(t) \quad (22)$$

Traditionally, ν_i^t can be obtained by solution matching [14] or again thanks to a modal power balance [33]. In this study, $\theta = (\nu_1^t, \dots, \nu_{N_{\text{gal}}}^t)^\top$ is considered as an unknown parameter vector, independent of time, that we want to estimate simultaneously with the state of the system.

While the ensemble Kalman filter presented previously offers the potential to correct online the state estimate, it does not allow for an improvement of the model

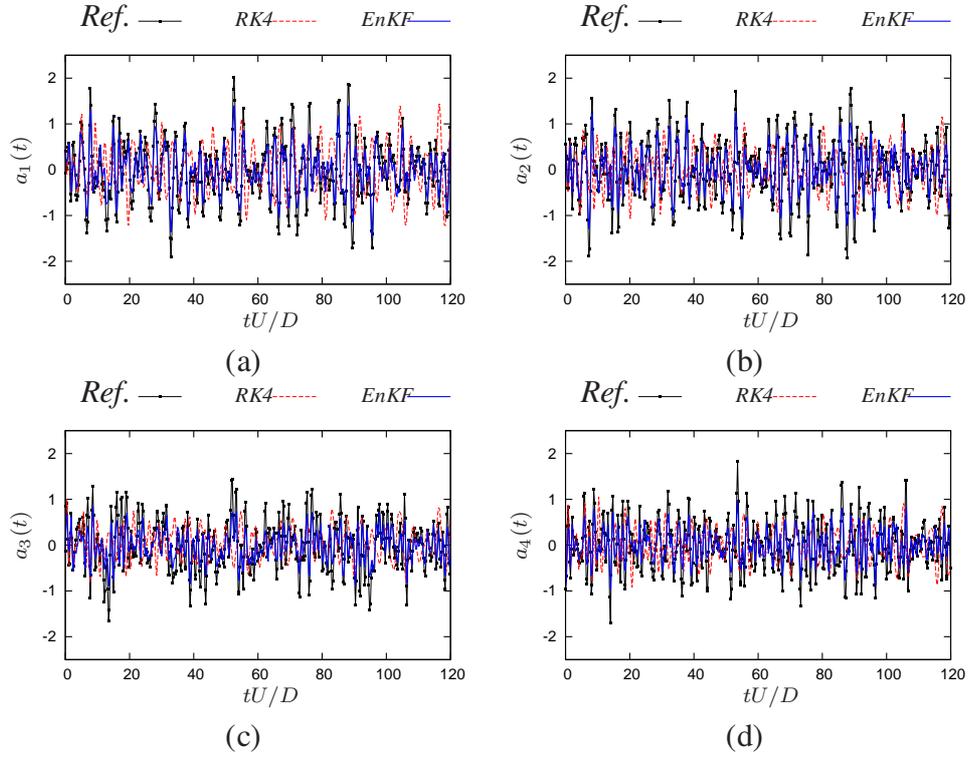


Figure 8: Time histories of the first four POD temporal coefficients (line with dots) and that of their estimates without (red dashed line) and with (blue line) data assimilation using the Ensemble Kalman Filter of Eq. (21). The observation vector includes far-field pressure sensor and longitudinal velocity sensors along the shear layer axis.

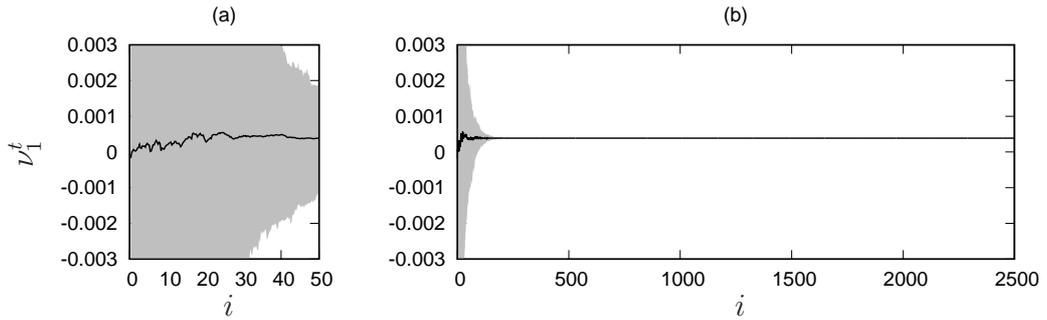


Figure 9: Time evolution of the estimated ν_1^t parameter of Eq. (22) obtained with Dual-EnKF. (a) Close view of the first 50 time units, (b) Complete view overall the learning sequence.

itself through, for example, the parameters vector θ . We therefore here introduce a dual state-parameter prediction technique based on EnKF, known as dual-EnKF and initially proposed by Moradkhani et al. [31]. The dual-EnKF requires separate state-space representation for the state and parameters through two parallel filters. The equation of evolution for the parameter vector is set up artificially assuming a random walk with a zero-mean white gaussian noises, hereafter \mathbf{w} . At each time step, the usual EnKF as described in Appendix A is applied twice: first to update the parameter vector and secondly to update the state vector using the analysis parameter vector. This dual approach is generally found to overperform joint Kalman filters where state and parameters are updated at the same time step, by overcoming two limitations [3]. The first one is that in a joint approach, the number of parameters that can be estimated is restricted to the number of measurements available. The second is that if the sensitivity matrix of the state-to-output equations is ill-conditioned, the joint estimation results in deteriorated accuracy for parameter estimation. Finally, the non-linear discrete system considered can be read as,

$$\begin{cases} \boldsymbol{\theta}_k & = \boldsymbol{\theta}_{k-1} + \mathbf{w}_k \\ \mathbf{a}^R(\boldsymbol{\theta}_k; t_k) & = \mathbf{F}(\mathbf{a}^R(\boldsymbol{\theta}_k, t_{k-1}), \mathbf{y}) + \mathbf{u}_k \\ \mathbf{z}(t_k) & = H\mathbf{a}^R(\boldsymbol{\theta}_k; t_k) + \mathbf{v}_k \end{cases} \quad (23)$$

Demonstration of the dual-EnKF applied on the standard Lorenz model equation to recover the coefficients of the model equation is discussed in Appendix A. For the present purpose, the white gaussian noises are initialised as $\mathbf{u}_0 = \mathbf{v}_0 = \mathcal{N}(0, 0.01\mathbb{I}_{N_{\text{gal}}})$ and $\mathbf{w}_0 = \mathcal{N}(0, \mathbb{I}_{N_{\text{gal}}})$. A number of 500 members for the ensemble is chosen and was found to give satisfactory results. Note that the same observation vector than that used in §5.2 is considered. The parameter vector $\boldsymbol{\theta}$ is initialised with null values. The entire time sequence is divided into a learning sequence of duration 265 time units during which the observations are assimilated at each time step, followed with a validation sequence during which observations are considered available every 50 time steps only (corresponding to time intervals of $\Delta t_z U/D \approx 13$).

The value of ν_1^t over the learning window is reported in figure 9. The shaded area corresponds to the variance in the ensemble members (or confidence interval) while the black line represents the mean ensemble, i.e. the estimated parameter value. The variance of the ensemble is arbitrarily initialised as large. The parameter value is found to converge smoothly in a given region of the parameter space towards a solution with negligible variance or, in other words, uncertainty. Similar results are observed for the other parameters as reported in figure 10(a-c).

Time evolution of the temporal coefficient $a_1(t)$ during the learning window is reported in figure 11(a). The error between the reference and estimated trajectories can be quantified with the normalised root-mean-square error (hereafter NRMSE) defined as the quadratic error between the state estimate and the state reference given by,

$$\text{NRMSE}(t) = \sqrt{\frac{\sum_{i=1}^{N_{\text{gal}}} [a_i^P(t) - a_i^R(t)]^2}{\sum_{i=1}^{N_{\text{gal}}} [a_i^P(t)]^2}} \quad (24)$$

As shown in figure 11(b), the NRMSE remains low during the learning window. At the end of the learning window, as mentioned previously, the observations are then assimilated every 13 time units only. Time evolution of the estimated coefficient $a_1(t)$ after the end of the learning window and of the NRMSE is reported in figure 12(a&b). The NRMSE only slightly increases and the reference trajectory for $a_1(t)$ is well recovered. Similar results are obtained for the other modes as illustrated in figure 13 where estimated trajectories for $a_2(t)$, $a_4(t)$, $a_6(t)$ & $a_{10}(t)$ are shown. While not shown here also, the NRMSE is found to slightly increase with the mode number but the estimation remains satisfactory. While not shown here, if the assimilation of new observations is stopped, the estimated trajectories remain bounded but the reference dynamics is lost as observed initially in §4.1. At this stage, this raises the question of the reliability of POD-ROMs to be used for long-term prediction without correction of the estimate thanks, for example, to assimilation of new observations.

6. Reconstruction of the spatio-temporal flow field dynamics associated with low-frequency sound emission

In this last part, we test the degree to which the corrected ROM can reproduce the dynamics of the reference spatio-temporal flow field associated with sound-producing mechanisms.

Unrolled phase portrait over a long time horizon using the two first POD temporal coefficients estimated is first reported in figure 14. The phase portrait is found to be well bounded thanks to stabilization of the dynamical system, and to properly follow the original trajectory. Using the distance defined by $d(t) = \sqrt{a_1(t)^2 + a_2(t)^2}$ to identify the location of a point along the phase portrait with coordinates $(a_1(t), a_2(t))$ at any time, the trajectory described by the portrait may be seen as an orbit with center close to coordinate (0,0) essentially bounded. However, the trajectory is found to jumps intermittently to another orbit characterised by a larger diameter. The phase portrait shown in figure 14 is

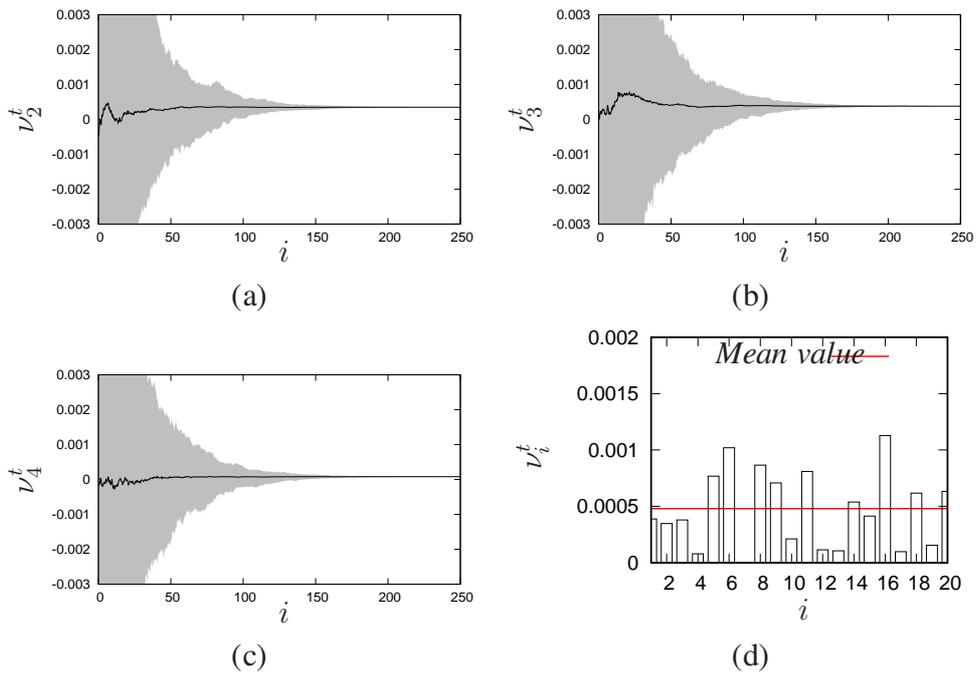


Figure 10: (a,b & c) Evolution of the first estimated parameters during the assimilation window. (d) Last values of the parameter vector at the end of the assimilation window with (red line) mean value of θ .

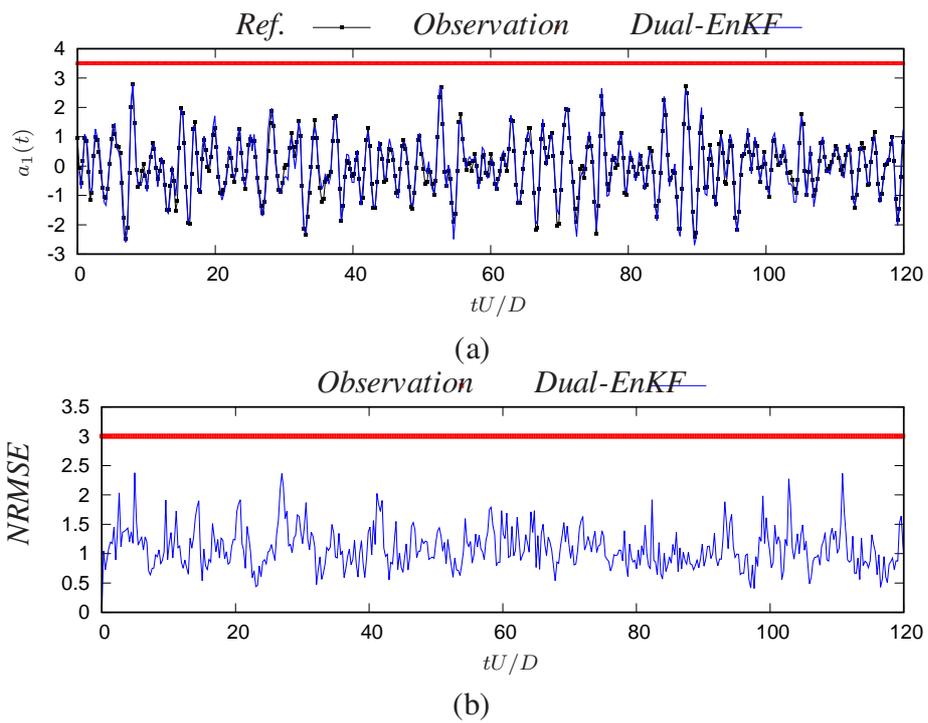


Figure 11: Time evolution during the learning window of (top) the first temporal POD coefficient estimated thanks to the dual-EnKF scheme and (b) the associated root-mean-square error defined in Eq. ???. The red dots indicates when observations are assimilated. (Black square dots) Original POD coefficient, (orange line) coefficient estimated with EnKF and (blue line) coefficient estimated with dual-EnKF.

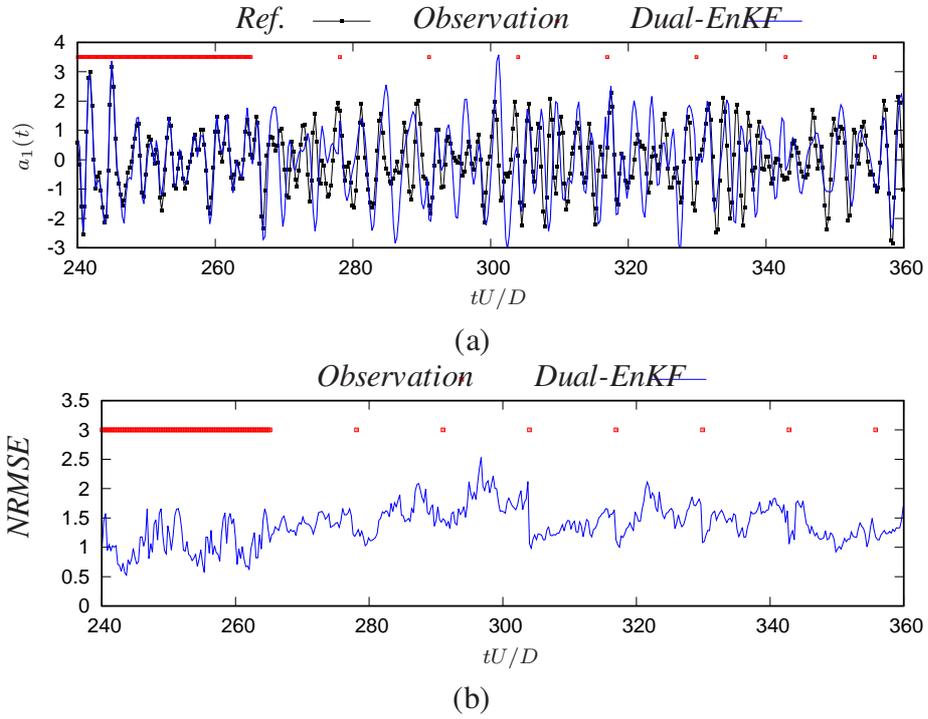


Figure 12: Time evolution during the testing window of (top) the first temporal POD coefficient estimated thanks to the dual-EnKF scheme and (b) the associated root-mean-square error defined in Eq. ?? . The learning window ends at $tU/D = 265$ above which observations are assimilated every 13 time units. Same legends as figure 11.

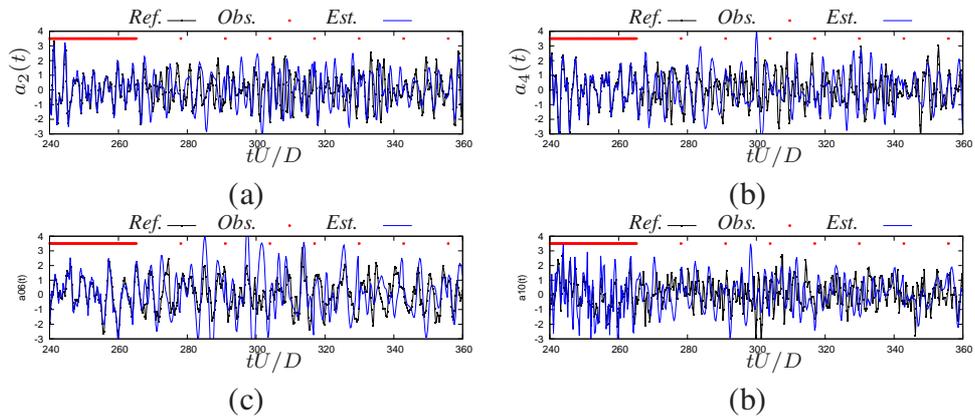


Figure 13: Time evolution during the testing window of different POD temporal POD coefficients: (a) $a_2(t)$, (b) $a_4(t)$, (c) $a_6(t)$ and (d) $a_{10}(t)$. Same legends as figure 11.

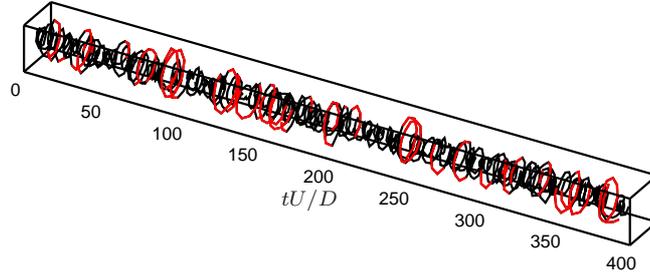


Figure 14: Unrolled phase portrait of POD mode 1 versus POD mode 2. When the trajectory of the portrait is found to follow an orbit by a larger distance from the centerline, the line is colored in red.

coloured in red when the diameter of the trajectory is above a given threshold to highlight these intermittent jumps. These jumps are connected to the sound producing intermittent events discussed by Cavalieri et al. [9] and highlighted in Kerhervé et al. [25] for the present data.

The full space-time velocity flow field can be obtained by combining the spatial POD eigenfunction extracted from the original LSE data with the temporal POD coefficients estimated. A quantitative evaluation of the error committed in the flow estimate is provided by comparing the centerline velocity spectra at two different stations from the exit nozzle for the original and estimated reduced-order flows as illustrated in figure 15 using the temporal coefficients estimated with the Dual-EnKF discussed previously. The overall frequency content below $St_D \sim 1$ of the reference data is well reproduced. Above $St_D \sim 1$, the discrepancies with the reference data are due, first, to the POD truncation and, secondly, to the higher discrepancies observed in the higher modes between the reference and estimated temporal coefficients (see figure 13). Spurious frequency peaks are observed for the estimated field. These peaks are harmonics of the low frequency peak around $St_D \approx 0.08$ which is exactly the frequency associated with the time interval of the assimilation of new observations ($\Delta t_z U/D = 13$). Another estimation was effected with assimilation at time intervals $\Delta t_z U/D = 4$. The same frequency spectra for the obtained field are also plotted in figure 15. The spurious frequency peaks have almost disappeared, with one reminiscent close to $St_D \approx 0.25$, i.e. again the frequency associated to the time interval of the assimilation. In conclusion, the energy content of the dominated coherent structures of interest is well reproduced by the low-order modelling introduced.

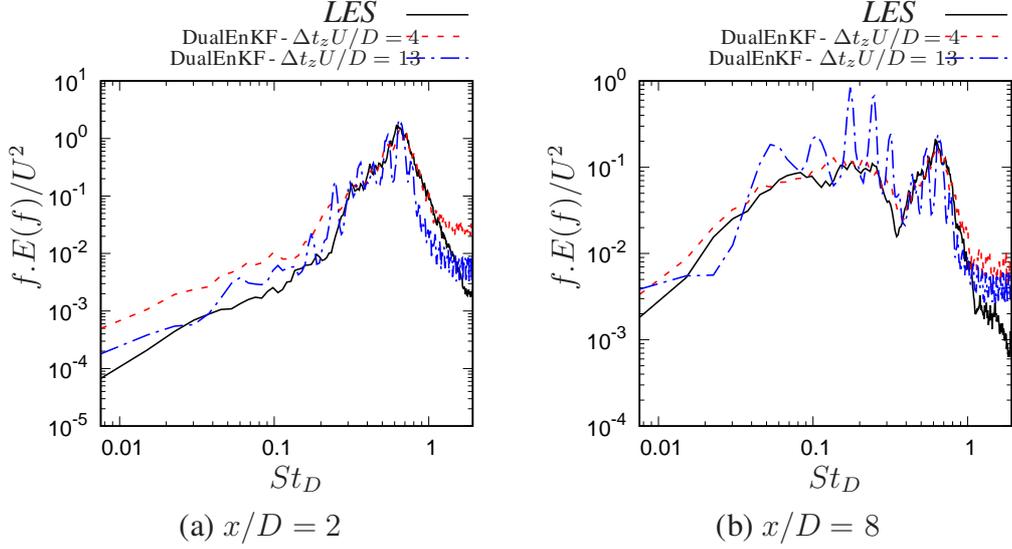


Figure 15: Energy spectrum of the fluctuating axial velocity component for a point located along the jet centerline at (a) $x/D = 2$ and (b) $x/D = 8$ for the LES and modelled fields using the long-term prediction procedure.

7. Conclusion

A data-driven strategy for the elaboration of POD reduced-order models able to predict the flow states over a long-time horizon is proposed. The methodology combines (i) a regularised least-square identification of the coefficients of the POD-ROM, (ii) a modal linear eddy viscosity model to mimic the high-order modes and parametrised with an unknown parameters vector, and (iii) a sequential data assimilation technique known as the Dual Ensemble Kalman Filter (DualEnKF). The later is used to propagate and update sequentially the estimated state and to discover, simultaneously, the parameters vector of the residual term by assimilating observations at a regular time steps. A Mach 0.9 turbulent jet, simulated by Large Eddy Simulation by Bogey and Bailly [6], is considered as a test case. The present work follows that of Kerhervé et al. [25] who proposed a methodology to educe the flow motions associated with sound-producing events and responsible for low-angle sound radiation. The data-driven strategy is here applied with the objective to build a low-order dynamical system of these specific motions. A twenty-mode POD-ROM is therefore identified and used in combination with the DualEnKF to recover the dynamics of the sound-producing flow motions over long term horizon based on a limited number of observations.

The later have been chosen to replicate what can be done in experiments such as, for instance, velocity point and far-field pressure measurements. The regularised identification procedure allows to obtain a bounded dynamical system but whose trajectory drifts from the true one after some time steps. When combined with the DualEnKF, the trajectory is found to be corrected and to replicate, with small errors, that of the true state. Further works in the selection of the observation vector, in the estimation of the model and error covariances or again in the identification of reduced-order models which are non-linear in the coefficients are some of the ways forwards which may improve the proposed strategy.

- [1] 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* **192**, 115–173.
- [2] Balajewicz, M. J., Dowell, E. H. and Noack, B. R. [2013], ‘Low-dimensional modelling of high-reynolds-number shear flows incorporating constraints from the navier–stokes equation’, *Journal of Fluid Mechanics* **729**, 285–308.
- [3] Bavdekar, V. A., Prakash, J., Shah, S. L. and Gopaluni, R. B. [2013], Constrained dual ensemble kalman filter for state and parameter estimation, in ‘2013 American Control Conference’, IEEE, pp. 3093–3098.
- [4] Bergmann, M. and Cordier, L. [2008], ‘Control of the circular cylinder wake by Trust-Region methods and POD Reduced-Order Models’, *Research Report 6552, INRIA 06*, URL <https://hal.inria.fr/inria-00284258>.
- [5] Bogey, C. and Bailly, C. [2005], ‘Effects of inflow conditions and forcing on subsonic jet flows and noise’, *AIAA J.* **43(5)**, 1000–1007.
- [6] Bogey, C. and Bailly, C. [2006], ‘Computation of a high Reynolds number jet and its radiated noise using large eddy simulation based on explicit filtering’, *Comp. Fluids* **35(10)**, 1344–1358.
- [7] Bogey, C., Bailly, C. and Juvé, D. [2003], ‘Noise investigation of a high subsonic moderate Reynolds number jet using a compressible LES’, *Theor. Comput. Fluid Dyn.* **16(4)**, 273–297.
- [8] Borggaard, J., Iliescu, T. and Wang, Z. [2011], ‘Artificial viscosity proper orthogonal decomposition’, *Mathematical and Computer Modelling* **53(1)**, 269–279.

- [9] Cavalieri, A., Jordan, P., Agarwal, A. and Gervais, Y. [2011], ‘Jittering wave-packet models for subsonic jet noise’, *J. Sound Vib.* **330**, 4474–4492.
- [10] Cavalieri, A. V., Jordan, P. and Lesshafft, L. [2019], ‘Wave-packet models for jet dynamics and sound radiation’, *Applied Mechanics Reviews* **71**(2).
- [11] Cazemier, W., Verstappen, R. and Veldman, A. [1998], ‘Proper orthogonal decomposition and low-dimensional models for driven cavity flows’, *Physics of Fluids (1994-present)* **10**(7), 1685–1699.
- [12] Chen, M., Liu, S., Tieszen, L. and Hollinger, D. [2008], ‘An improved state-parameter analysis of ecosystem models using data assimilation’, *ecological modelling* **219**(3-4), 317–326.
- [13] Cordier, L., Abou El Majd, B. and Favier, J. [2010], ‘Calibration of POD reduced-order models using tikhonov regularization’, *Int. J. Num. Meth. in Fluids* **63**(2), 269–296.
- [14] Cordier, L., Noack, B. R., Tissot, G., Lehnasch, G., Delville, J., Balajewicz, M., Daviller, G. and Niven, R. K. [2013], ‘Identification strategies for model-based control’, *Experiments in fluids* **54**(8), 1–21.
- [15] Crow, S. and Champagne, F. [1971], ‘Orderly structures in jet turbulence’, *J. Fluid Mech.* **48**(3), 547–591.
- [16] Evensen, G. [2003], ‘The ensemble kalman filter: Theoretical formulation and practical implementation’, *Ocean dynamics* **53**(4), 343–367.
- [17] Evensen, G. [2009], *Data assimilation: the ensemble Kalman filter*, Springer Science & Business Media.
- [18] Galletti, B., Bruneau, C., Zannetti, L. and Iollo, A. [2004], ‘Low-order modelling of laminar flow regimes past a confined square cylinder’, *Journal of Fluid Mechanics* **503**, 161–170.
- [19] Hansen, P. [1994], ‘Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems’, *Numerical algorithms* **6**(1), 1–35.
- [20] Holmes, P., Lumley, J. and Berkooz, G. [1998], ‘Turbulence, coherent structures, dynamical systems and symmetry’, *Cambridge University Press* .

- [21] Jordan, P. and Colonius, T. [2013], ‘Wave packets and turbulent jet noise’, *Annual review of fluid mechanics* **45**, 173–195.
- [22] Jordan, P., Schlegel, M., Stalnov, O., Noack, B. and Tinney, C. [2007], ‘Identifying noisy and quiet modes in jet’, *13th AIAA/CEAS Aeroacoustics Conference, Paper 2007-3602* .
- [23] Kalb, V. and Deane, A. [2007], ‘An intrinsic stabilization scheme for proper orthogonal decomposition based low-dimensional models’, *Physics of fluids* **19**, 054106.
- [24] Kato, H., Yoshizawa, A., Ueno, G. and Obayashi, S. [2015], ‘A data assimilation methodology for reconstructing turbulent flows around aircraft’, *Journal of Computational Physics* **283**, 559–581.
- [25] Kerhervé, F., Jordan, P., Cavalieri, A., Delville, J., Bogey, C. and Juvé, D. [2012], ‘Educing the source mechanism associated with downstream radiation in subsonic jets’, *J. Fluid Mech.* **710**, 606–640.
- [26] Leroux, R., Chatellier, L. and David, L. [2014], ‘Bayesian inference applied to spatio-temporal reconstruction of flows around a naca0012 airfoil’, *Experiments in Fluids* **55**(4), 1–19.
- [27] Lighthill, M. [1952], ‘On sound generated aerodynamically: I. general theory’, *Proc. R. Soc. A* **222**, 564–587.
- [28] Liu, J. and West, M. [2001], Combined parameter and state estimation in simulation-based filtering, in ‘Sequential Monte Carlo methods in practice’, Springer, pp. 197–223.
- [29] Lumley, J. [1967], ‘The structure of inhomogeneous turbulent flows’, *Atmospheric Turbulence and Wave Propagation (Moscow, Nauka)* pp. 166–178.
- [30] Michalke, A. [1970], ‘A wave model for sound generation in circular jets’.
- [31] Moradkhani, H., Sorooshian, S., Gupta, H. V. and Houser, P. R. [2005], ‘Dual state–parameter estimation of hydrological models using ensemble kalman filter’, *Advances in Water Resources* **28**(2), 135–147.

- [32] Noack, B., Papas, P. and Monkewitz, P. [2002], Low-dimensional galerkin model of a laminar shear-layer, Technical report, Tech. Rep. 2002-01. Laboratoire de Mecanique des Fluides, Departement de Genie Mecanique, Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- [33] Noack, B. R., Papas, P. and Monkewitz, P. A. [2005], ‘The need for a pressure-term representation in empirical galerkin models of incompressible shear flows’, *Journal of Fluid Mechanics* **523**, 339.
- [34] Östh, J., Noack, B. R., Krajnović, S., Barros, D. and Borée, J. [2014], ‘On the need for a nonlinear subscale turbulence term in pod models as exemplified for a high-reynolds-number flow over an ahmed body’, *Journal of Fluid Mechanics* **747**, 518–544.
- [35] Perret, L., Collin, E. and Delville, J. [2006], ‘Polynomial identification of POD based low-order dynamical system’, *J. Turb.* **7**, 1–15.
- [36] Podvin, B. [2009], ‘A proper-orthogonal-decomposition–based model for the wall layer of a turbulent channel flow’, *Physics of Fluids (1994-present)* **21**(1), 015111.
- [37] Protas, B., Noack, B. R. and Östh, J. [2015], ‘Optimal nonlinear eddy viscosity in galerkin models of turbulent flows’, *Journal of Fluid Mechanics* **766**, 337–367.
- [38] Rempfer, D. and Fasel, H. F. [1994], ‘Evolution of three-dimensional coherent structures in a flat-plate boundary layer’, *Journal of Fluid Mechanics* **260**, 351–375.
- [39] Sirisup, S. and Karniadakis, G. [2004], ‘A spectral viscosity method for correcting the long-term behavior of pod models’, *Journal of Computational Physics* **194**(1), 92–116.
- [40] Sirovich, L. [1987], ‘Turbulence and the dynamics of coherent structures. i. coherent structures’, *Quarterly of applied mathematics* **45**(3), 561–571.
- [41] Suzuki, T. [2012], ‘Reduced-order kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation’, *Journal of Fluid Mechanics* **709**, 249–288.

- [42] Suzuki, T. [2014], ‘Pod-based reduced-order hybrid simulation using the data-driven transfer function with time-resolved ptv feedback’, *Experiments in Fluids* **55**(8), 1–17.
- [43] West, M. [1993], ‘Mixture models, monte carlo, bayesian updating, and dynamic models’, *Computing Science and Statistics* pp. 325–325.

Appendix A. Ensemble and dual-Ensemble Kalman filters

Appendix A.1. Ensemble Kalman filter

The main elements for the implementation of the Dual-Ensemble Kalman Filter for state and parameters estimation are here detailed and follows the original formulation of Evensen [16] and Moradkhani et al. [31]. In the following, the dimension of the state space is n while the true state of the system at discrete time t_k is denoted by $\mathbf{x}_k \in \mathbb{R}^n$. Note that for practical implementation, the time is discretised, which justifies the subscript notation. Let the dynamics of the system be described by a non-linear operator $\mathcal{M}(\mathbf{x}_k, \theta)$ where θ forms a vector of time-invariant parameters of dimension n_b . In addition, the dimension of the observation space is n_o while the observation vector at time t_k is denoted by $\mathbf{y}_k \in \mathbb{R}^{n_o}$. We here consider the case where the observation and the state vectors are related through $\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k)$ where \mathcal{H} can stand for a linear or non-linear operator. Kalman filtering is a two steps procedure alternating forecast and analysis of the system state. In the following, the forecast and analysis of the state at time t_k are denoted by \mathbf{x}_k^f and \mathbf{x}_k^a respectively. In the sequential procedure, the dynamical model f is first used to evolve the state estimate forward in time to form the forecast estimate \mathbf{x}_k^f . When an observation is available, a correction of \mathbf{x}_k^f is effected to generate the analysis estimate \mathbf{x}_k^a such that the mean-square error between the prediction $\mathcal{H}(\mathbf{x}_k^f)$ and the observation \mathbf{y}_k is minimised.

When the dynamics of the state is described by a linear model, the linear Kalman filter (KF) offers the statistically optimal estimate of the underlying system state providing given noisy observations and model. The generalization to non-linear system is known as the Extended Kalman filter (EKF) which requires the Jacobian of the dynamics matrix. This filter mimics the classical Kalman filter by propagating a matrix (the surrogate covariance) that is analogous to the error covariance in the linear case. Another approach which belongs to the category of particle filters and widely used in weather forecasting where the models are high order and nonlinear, the initial states are uncertain, is known as the Ensemble

Kalman filter (EnKF). While the KF uses a single state estimate, the EnKF uses a statistical sample of state estimates, called an ensemble. It is a suboptimal estimator based upon Monte Carlo or ensemble generations where the approximation of the forecast state error covariance matrix is made by propagating an ensemble of model states using the updated states (ensemble members) from the previous time step. In contrast to the standard Kalman filter, propagation of covariance matrices is not necessary. This results in a gain of computational time which is however compensated by an additional cost since one need to maintain the ensemble members throughout the time marching.

Let the discrete-time non-linear system be expressed as,

$$\begin{cases} \mathbf{x}_k &= \mathcal{M}(\mathbf{x}_{k-1}, \theta) + \eta_{k-1} \\ \mathbf{y}_k &= \mathcal{H}(\mathbf{x}_k) + \epsilon_k \\ \eta_k &\sim \mathcal{N}(0, \Sigma_k^{\mathcal{M}}) \\ \epsilon_k &\sim \mathcal{N}(0, \Sigma_k^{\mathcal{H}}) \end{cases} \quad (\text{A.1})$$

where $\eta_k \in \mathbb{R}^n$ and $\epsilon_k \in \mathbb{R}^{n_o}$ are white, zero-mean, uncorrelated process and observation noises respectively with known covariance matrices $\Sigma_k^{\mathcal{M}}$ and $\Sigma_k^{\mathcal{H}}$. The EnKF propagates ensembles of state vectors in parallel such that each state vector represents one realisation of generated model replicates. In the following, ensembles at a given time step are denoted as X_k with superscripts f or a for forecast and analysis respectively. These ensemble are formed with a number of n_q replicates $\mathbf{x}_k^{(i)}$ of the state vector,

$$X_k = (\mathbf{x}_k^{(1)}, \dots, \mathbf{x}_k^{(n_q)})$$

For a given ensemble, the state estimate, hereafter $\bar{\mathbf{x}}_k$, must be interpreted as the most expected value of the ensemble:

$$\bar{\mathbf{x}}_k = \frac{1}{n_q} \sum_{i=1}^{n_q} \mathbf{x}_k^{(i)} \quad (\text{A.2})$$

The EnKF uses the standard Kalman filter equations to propagate the members of the ensembles. The forecast ensemble members are first obtained by propagating the analysis ensemble through the nonlinear dynamics,

$$\mathbf{x}_k^{f,(i)} = \mathcal{M}(\mathbf{x}_{k-1}^{a,(i)}, \theta) + \eta_k^{(i)} \quad \text{with} \quad \eta_k^{(i)} = \mathcal{N}(0, \Sigma_k^{\mathcal{M}}) \quad (\text{A.3})$$

A linear correction is then used according to standard Kalman filter to update the members of the forecast ensemble as

$$\mathbf{x}_k^{a,(i)} = \hat{\mathbf{x}}_k^{f,(i)} + K_k \left(\mathbf{y}_k^{(i)} - \hat{\mathbf{y}}_k^{(i)} \right) \quad (\text{A.4})$$

where $\hat{\mathbf{y}}_k^{(i)}$ stands for the predicted variable given by,

$$\hat{\mathbf{y}}_k^{(i)} = \mathcal{H}(\mathbf{x}_k^{f,(i)}) \quad (\text{A.5})$$

and $\mathbf{y}_k^{(i)}$ for the replicate members of the observation vector \mathbf{y}_k generated by,

$$\mathbf{y}_k^{(i)} = \mathbf{y}_k + \epsilon_k^{(i)} \quad \text{with} \quad \epsilon_k^{(i)} = \mathcal{N}(0, \Sigma_k^{\mathcal{H}}) \quad (\text{A.6})$$

The Kalman gain K_k is computed as,

$$K_k = \Sigma_k^{xy,f} [\Sigma_k^{yy} + \Sigma_k^{\mathcal{H}}]^{-1}$$

where Σ_k^{yy} stands for the forecast error covariance matrix of the prediction ensemble and given by,

$$\Sigma_k^{yy} = \frac{1}{n_q - 1} \sum_{i=1}^{n_q} \hat{\mathbf{y}}_k^{(i)} \left(\hat{\mathbf{y}}_k^{(i)} \right)^\top$$

and $\Sigma_k^{xy,f}$ for the forecast cross-covariance matrix between the forecast ensemble $\mathbf{x}_k^{f,(i)}$ and prediction $\hat{\mathbf{y}}_k^{(i)}$ and given by,

$$\Sigma_k^{xy,f} = \frac{1}{n_q - 1} \sum_{i=1}^{n_q} \mathbf{x}_k^{f,(i)} \left(\hat{\mathbf{y}}_k^{(i)} \right)^\top$$

In contrast to Kalman filter, propagation of the covariance matrices is not necessary, resulting in a gain in computational cost. A key of the EnKF however resides in the perturbation of the forcing and observation data by adding noise. The noise variances must be representatives of the uncertainty in these data.

Appendix A.2. Dual Ensemble Kalman filter

When there is no guarantee that the model parameters does not change with time or when the values chosen for these parameters represent only an initial guess, simultaneous state-parameter estimation may be necessary. An extension of the EnKF to state-parameter estimation has been proposed by Moradkhani et al. [31] and is known as dual-EnKF. The parameters are treated similarly to the state variables except that their time evolution is assumed to follow a random walk such that,

$$\theta_{k+1}^{f,(i)} = \theta_k^{a,(i)} + \tau_k^i \quad (\text{A.7})$$

Let θ_k denotes the ensemble of parameter members $\theta^{(i)}_k$ at time k , with subscript a or f for analysis and forecast estimates respectively, with associated covariance

Σ_k^θ . This artificial evolution is however known to result in an over-dispersion of parameter members leading to a loss of continuity between two consecutive time steps [12]. This flaw can be corrected thanks to kernel smoothing with location shrinkage [43, 28] where the individual ensemble members are drawn from a truncated multivariate normal distribution (*TMVN*),

$$\theta_k^{f,(i)} = \text{TMVN}(a\theta_{k-1}^{a,(i)} + (1-a)\bar{\theta}_k^a, h^2 \Sigma_k^{\theta,a}) \quad (\text{A.8})$$

where $h^2 = 1 - a^2$, $a = (3\delta - 1)/2\delta$ and $\delta \in \mathbb{R}$ is the smoothing parameter, typically between 0.95 and 0.99 [28]. In the dual-EnKF, the standard EnKF equations are first applied to update the parameters ensemble and to obtain the analysis parameter estimate accordingly:

$$\begin{cases} \theta_k^{f,(i)} = \text{TMVN}(a\theta_{k-1}^{a,(i)} + (1-a)\bar{\theta}_k^a, h^2 \Sigma_k^{\theta,f}) \\ \mathbf{x}_k^{f,(i)} = \mathcal{M}(\mathbf{x}_{k-1}^{a,(i)}, \theta_k^{f,(i)}) \\ \hat{\mathbf{y}}_k^{(i)} = \mathcal{H}(\mathbf{x}_k^{f,(i)}) \\ \theta_k^{a,(i)} = \theta_k^{f,(i)} + K_k^\theta \left(\mathbf{y}_k^{(i)} - \hat{\mathbf{y}}_k^{(i)} \right) \\ K_k^\theta = \Sigma_k^{\theta y} [\Sigma_k^{yy} + \Sigma_k^{\mathcal{H}}] \end{cases} \quad (\text{A.9})$$

For the same time step, the EnKF equations are then applied to update the state variables taking into account for the update parameters ensemble according to,

$$\begin{cases} \mathbf{x}_k^{f,(i)} = \mathcal{M}(\mathbf{x}_{k-1}^{a,(i)}, \theta_k^{a,(i)}) \\ \hat{\mathbf{y}}_k^{(i)} = \mathcal{H}(\mathbf{x}_k^{f,(i)}) \\ \mathbf{x}_k^{a,(i)} = \mathbf{x}_k^{f,(i)} + K_k^x \left(\mathbf{y}_k^{(i)} - \hat{\mathbf{y}}_k^{(i)} \right) \\ K_k^x = \Sigma_k^{xy,f} [\Sigma_k^{yy} + \Sigma_k^{\mathcal{H}}] \end{cases} \quad (\text{A.10})$$

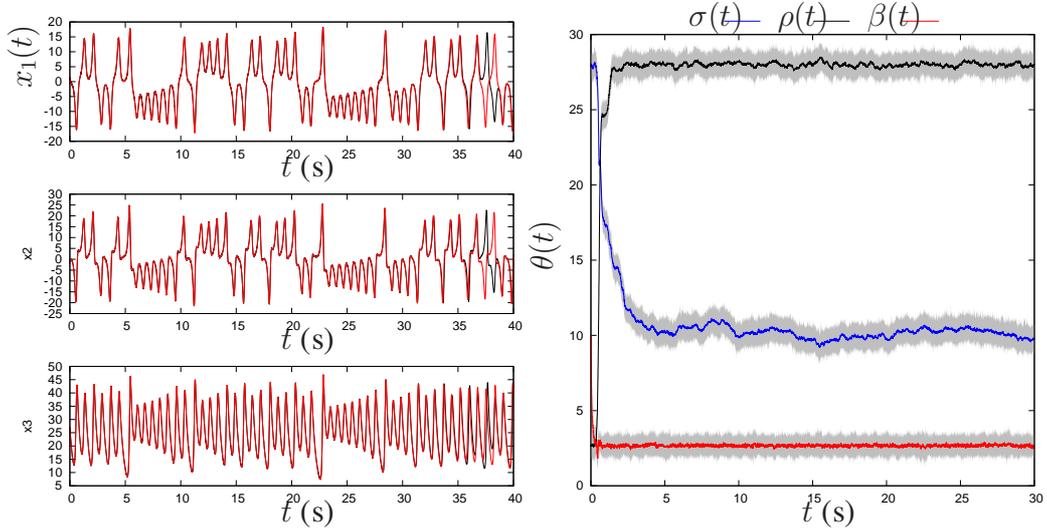
It is noteworthy that, theoretically, the two previous sequences can be reversed.

To illustrate the capability of the Dual-EnKF technique to recover both the state variable and the model's parameter vector, we applied here it to the Lorenz-63 model,

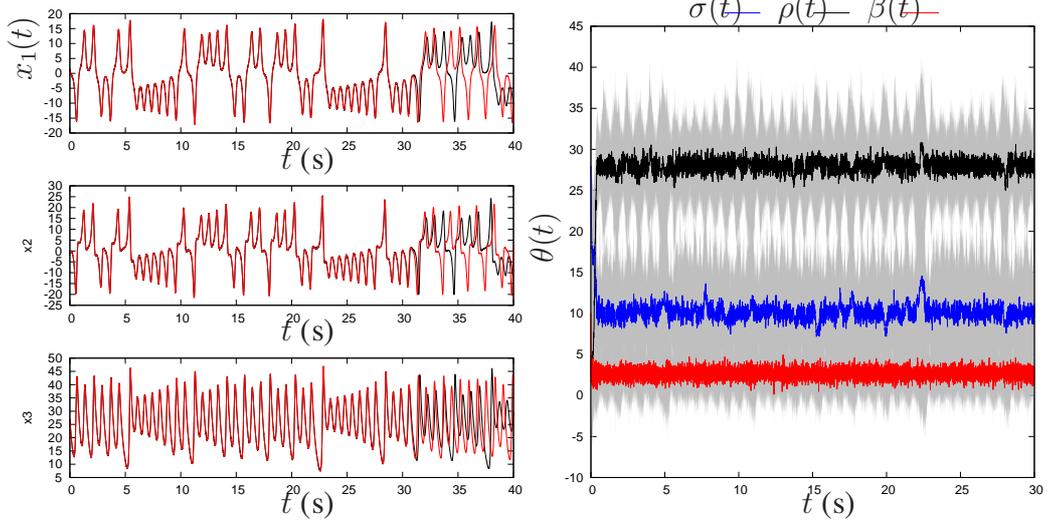
$$\begin{cases} \dot{x}_1 = \sigma(x_2 - x_1) \\ \dot{x}_2 = x_1(\rho - x_3) - x_2 \\ \dot{x}_3 = x_1x_2 - \beta x_3 \end{cases} \quad (\text{A.11})$$

where $\mathbf{x} = (x_1, x_2, x_3)^\top$ denotes the three-dimensional state vector and $\theta = (\sigma, \beta, \rho)^\top$ the parameter vector. This model equation a non-linear dynamical system. For $\theta = (10, 8/3, 28)$, the system is known to converge towards a chaotic

solution known as strange attractor. The initial condition considered in the present case is $\mathbf{x}(t = 0) = \mathbf{x}_0 = (5, 5, 5)^\top$. The model of Eq. (A.11) can easily be rewritten in the general form given in Eq. (A.1). In the present case, the operator \mathcal{H} is defined as the identity matrix. The parameter vector θ is initialised with incorrect values such that $\theta_0 = (8/3, 28, 10)^\top$, and random noise is introduced in the initial condition \mathbf{x}_0 . Observations are assimilated until $t^* = 30$. During this learning sequence, the parameter vector is updated. At the end of the sequence, the last value obtained for the parameter vector is conserved and time integration is maintained for the state variable up to $t^* = 40$. To test the procedure with regards to observation noise, two cases are considered: (case A) $\Sigma^y = 0.01\mathbf{I}_3$ and (case B) $\Sigma^y = \mathbf{I}_3$. Results of the estimation procedure for the state and parameter vectors are reported in figure A.16. Due to incorrect initial parameter vector and initial condition, discrepancies between the true and estimated trajectories of the state variables are observed for the two cases in the first time steps. The estimated state trajectory and the parameter vector are then rapidly corrected and fast convergence towards the true system is observed. Even with large noisy observations –figure A.16(b)– the system’s trajectory is well recovered while observations are assimilated thanks to the dual-EnKF. When assimilation is stopped, the estimated state’s trajectory remains well estimated over duration which decreases as the observation noise level increases due to small errors in the parameter vector but continue to exhibit similar dynamics.



(a) Case A: $\Sigma^y = 0.01\mathbf{I}_3$



(b) Case B: $\Sigma^y = \mathbf{I}_3$

Figure A.16: Results of the dual state-parameter ensemble Kalman filter applied to the Lorenz'63 model equation with observation noise covariance (a) $\Sigma^y = 0.01\mathbf{I}_3$ and (b) $\Sigma^y = \mathbf{I}_3$. Left column: Time history of (black) true and (red) estimated state variables. Right column: Time history of the estimated parameters through the learning sequence. The learning sequence stops at $t = 30$.