Analysis and prediction of rare events in turbulent flows

By P. J. Schmid†, O. T. Schmidt‡, A. Towne AND M. J. P. Hack

Turbulent shear flows are characterized by an interplay of many scales that describe persistent, quasi-invariant motion as well as violent, intermittent events. A data-driven computational framework, based on the decomposition of an embedded phase-space trajectory together with a community-identification step, will be introduced to properly describe and analyze these slow-fast dynamics. The framework combines elements of dynamic system theory with network analysis, and is applied to data-sequences from a reduced model of the turbulent self-sustaining process (SSP) in wallbounded shear flows. Its effectiveness in detecting and quantifying structures and in laying the foundation for their targeted manipulation will be assessed.

1. Introduction and motivation

Numerical simulations and experiments of turbulent fluid motion are characterized by a vast range of temporal and spatial scales. Nonetheless, over the past decades, we have succeeded in systematizing and deconstructing turbulent flows into widely acknowledged coherent structures (e.g., streaks, hairpin vortices, etc.) interspersed with short-lived, violent events (e.g. bursts). The search for the essential mechanisms underlying and sustaining turbulence, e.g., by studying minimal-channel units (Jimenez & Moin 1991) or otherwise restricting scales, has identified persistent, quasi-periodic motion (of large-scale and very large-scale, streaky structures) as well as intermittent, instability-triggered breakdown and reorganization of these structures. Characteristic turbulence features, such as production, dissipation and intermittency, can be associated with these events and the statistics of their occurrence in simulations and experiments. Further progress in analyzing and manipulating turbulent fluid motion has to recognize these dynamic structures, and our current computational and mathematical tools have to be adapted to them.

From a control-theoretic point of view, while it seems feasible to control (or delay) transitional flows, it appears too ambitious (and misguided) to aim at relaminarizing high-Reynolds number flows. From minimal-unit investigations (Jimenez & Moin 1991), we have learned that a significant part of turbulence activity is concentrated in the intermittent events (bursts). It appears more promising then, to concentrate on these extreme events and design strategies to reliably predict and perhaps, manipulate them.

The starting point of our analysis is a description of time-evolving processes using a dynamical systems approach in phase space combined with set-theoretic methods to characterize the dynamics Schmid et al. (2017). More specifically, the dynamic evolution of the flow fields will be described by a path in a suitably chosen phase space that converges towards and diverges from coherent building blocks. A hierarchy of these building blocks

† Department of Mathematics, Imperial College London, UK
‡ Department of Mechanical and Aerospace Engineering, University of California San Diego

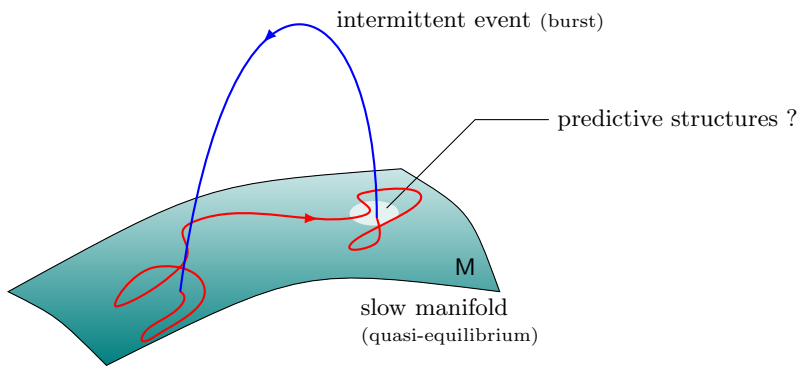


FIGURE 1. Sketch of slow-manifold dynamics and burst event. The flow progresses along the (red) phase-space trajectory on the slow manifold M, but intermittently deviates (blue trajectory) to return back to M. The structures near the detachment point could be predictive of an imminent burst event.

will be identified using network and cluster analysis, and the phase-space trajectory will be categorized by its residence time close to coherent states in phase space. To this end, transition probabilities will have to be computed between regions of phase space. An initial discretization of the full phase-space trajectory into finite-sized hypercubes will be computed. From there, a transition probability matrix between occupied sections of phase space will be determined. This procedure and the information contained in the transition matrix are reminiscent of a probability density analysis within a Fokker-Planck description of dynamical systems.

Persistent dynamics are given by high probabilities (near certainties) of remaining within a given group of phase-space sectors, while intermittent events are characterized by rare transitions to corners of phase space that have low residence time and are otherwise weakly connected to persistent structures. The demarcation of persistent from intermittent structures is accomplished by reorganizing the initial discretization of the phase space into communities. This step is achieved by interpreting the phase-space trajectory (and the transition matrix) as a directed network graph (Chartrand & Zhang 2012), with the transition probability as a weight between its nodes. Community-clustering algorithms are then applied to regroup phase space into similar groups (coherent structures) and outlier groups (transitory, intermittent structures). This process will provide a novel and objective manner of identifying low-dimensional mechanisms that interact to form the overall dynamics and reproduce the processed data sequence. In other words, we will objectively decompose the flow into a slow manifold (in wall-bounded turbulence consisting of streaks, hairpin vortices, etc.) and a fast manifold orthogonal to the trajectory's tangent space in the slow manifold (that models the intermittent bursts). Concentrating on the slow manifold, we can also identify structures that have a higher probability of detaching from the slow manifold and use these structures as precursors (or predictors) of impending violent events (see Figure 1). These predictive structures, paired with a dynamic programming (model-predictive control) approach, can then guide a control strategy to prevent intermittent bursts and remain/return to the slow manifold.

The use of phase-space embedding and set-theoretic tools has recently been proposed by Kaiser et al. (2014) and applied to a mixing layer. While this work introduced some

of the techniques (such as phase-space tesselation) used in our work, it concentrated on the bimodal nature of mixing layers where two states with nearly equal catchment basins are connected by a saddle-point structure. Moreover, the clustering of phase-space regions in Kaiser et al. (2014) was accomplished by the k-means algorithm, which is unsuitable for our application as it would not be able to objectively identify intermittent events. A similar attempt, using network-theoretic tools, was made to recast interacting vortices in terms of weighted graphs, with the weights given by the Biot-Savart-induced velocity (Nair & Taira 2015).

The description of evolutionary processes in terms of transition probabilities and Markov matrices has been pioneered over the past years within the Lagrangian-coherent-structure and stochastic PDE community, (see, e.g., Froyland & Padberg 2009). These techniques – so far applied only to simple and low-dimensional problems – have shown great promise, but have not yet been embraced by the turbulence community.

The detection and description of rare events have also been attempted by a dynamically orthogonal (DO) mode approach, (see, e.g., Farazmand & Sapsis 2017). This method uses a dynamically changing, linear-tangent approximation to the phase-space trajectory to detect a substantial (and drastic) deviation from an otherwise smoothly varying manifold described by a few DO-modes. While elegant and effective in its approach, the computational cost of computing the DO-modes is quite significant.

2. Formulation and computational setup

We will follow the above-mentioned procedure and recast our data sequence as a trajectory in a high-dimensional phase space spanned by appropriate coherent flow fields (given by, e.g., POD modes) or by characteristic scalar flow variables (such as dissipation, production and intermittency). This phase space is then discretized and the resulting trajectory is converted into a Markov matrix describing the probabilities of transition between various phase-space elements over one time-step. This Markov matrix is then interpreted as a directed, weighted graph (Chartrand & Zhang 2012) consisting of all occupied phase-space elements; it fully encodes the dynamics of the data sequence within a probabilistic setting. In a final step, the original network is reorganized into larger communities (Leicht & Newman 2008) to iteratively isolate persistent motion from rare outlier events. Each of the procedural steps will be further explained below.

2.1. *Dimensionality reduction by phase-space embedding*

Starting with a sequence of temporally equispaced flow-field snapshots, we identify a set of flow fields that allow an accurate, but approximate, description of the data sequence in a lower-dimensional space. While the choice of these flow fields is dependent on the specific application, we propose proper orthogonal decomposition modes or variants thereof (see Lumley 1970; Sirovich 1987; Berkooz et al. 1993) as a reasonable starting point. By truncating the basis and projection of the original flow fields onto the retained POD modes, we arrive at a lower-dimensional system of time-evolving expansion coefficients. For complex fluid systems and/or sophisticated temporal behavior (e.g., bi- or multi-modality), a more judicious choice of embedding basis has to be made.

2.2. *Constructing the Markov matrix*

Following the lower-dimensional representation of the phase-space trajectory, we proceed by discretizing the phase space into k-dimensional hypercubes, with k denoting the dimensionality of the phase space. Traditionally, this step is performed using a subdivision

algorithm, where successive cuts by $(k-1)$-dimensional hyperplanes are made followed by a reorganization/reassignment of all trajectory points into the resulting two hyperspaces. The cutting dimension is then rotated through all phase-space coordinates and periodically continued, until all hypercubes contain less than a user-specified number of trajectory points. This technique is the method of choice for low- to moderate-dimensional phase space, but will not produce favorable scaling properties for higher dimensions. Consequently, we will opt for a search algorithm that scales linearly with the number of processed snapshots and scales weakly (or negligibly) with the dimensionality k of the phase space. In this search algorithm, we follow the phase-space trajectory and assign the trajectory points to hypercubes of a prescribed size; points leaving a current hypercube require a search through previously created hypercubes and, in the case of a failed search, the creation of a new one. To accelerate the search through previous hypercubes, hashing can be introduced; while this may be necessary for very large-scale applications, no particular advantage has been found so far for our initial applications.

With the initial discretization given, we can compute the transition probability matrix. We label our hypercubes by B_i with $i = 1, \ldots, N$, and N denoting the number of hypercubes containing the phase-space trajectory in its entirety. We also need to introduce a suitable density measure $m(\mathsf{B}_i)$ for each box B_i which we take as the number of instances (phase-space points) it contains. We then can define the transition probability matrix P as

$$\mathsf{P}_{ij} = \frac{m(\mathsf{B}_i \cap \mathcal{F}^{-1}(\mathsf{B}_j))}{m(\mathsf{B}_i)} \qquad i, j = 1, \ldots, N, \tag{2.1}$$

where $\mathcal{F}^{-1}$ stands for the temporal backstep operator, i.e., we evaluate its argument at the previous time-step (see Kaiser et al. (2014)).

In the above expression, the numerator counts the number of occurrences where phase-space points transition into box B_i from box B_j over one time step. The denominator simply determines the phase-space points in box B_i. The ratio represents an approximation of the (Markov) transition probability between box B_j and B_i and signifies the probability that any point in B_i may have come from box B_j over one time-step. The full matrix P then provides an approximation of the (Markov) transition probability matrix between all boxes in our discretization of phase space and is referred to as the Ulam-Galerkin method (Ulam 1960). Mathematically, the above matrix P is a finite-dimensional approximation of the Frobenius-Perron operator (Lasota & Mackey 1994) (i.e., the adjoint of the Koopman operator), analogous to the fact that the matrix arising from the Dynamic Mode Decomposition is a finite-dimensional approximation of the Koopman operator.

Note that the discretization of phase space (i.e., the user-specified size of the initial hypercubes) remains as a critical parameter in our analysis. It has to be chosen sufficiently small to accurately approximate the phase-space trajectory and sufficiently large to contain enough trajectory points for an accurate representation of the transition probabilities; a judicious balance between these two extremes has to be found.

The Markov matrix P encodes the full dynamics, albeit in a probabilistic manner.

2.3. *Community clustering*

The Markov matrix P describes the transition probabilities between our initial discretization of the entire phase-space trajectory. It is a row-stochastic matrix of size $N \times N$. We choose to interpret P as an adjacency matrix (Chartrand & Zhang 2012) of a weighted, directed graph, where each node represents a box B_i and each edge is given by a connec-

tion between the two respective boxes $\mathsf{B}_i, \mathsf{B}_j$, with the transition probability P_{ij} as the edge's weight. The sparse nature of P and the low connectivity of the associated graph reflect the fact that in general only a few boxes B_j feed into a given box B_i.

This type of graph can be reorganized into larger communities of similar (to be defined) clusters. In addition, it can be analyzed using graph-theoretic measures (Chartrand & Zhang 2012), such as the graph Laplacian or the structure of near-invariant modes (Froyland & Padberg 2009). For our application, we are interested in detecting communities in the network. A community is defined as a collection of nodes from the graph that shows strong intraconnectivity (within the community) but weak interconnectivity (to other communities). Multiple algorithms exist for the detecting of graph communities within large networks, many of them relying on a measure of interconnectivity referred to as modularity given by

$$Q = \frac{1}{N} \sum_{i,j} \left[\mathsf{P}_{ij} - \frac{k_i^{\text{in}} k_j^{\text{out}}}{N} \right] \delta_{c_i, c_j} \tag{2.2}$$

with $k_i^{\text{in,out}}$ as the in- or out-degrees of the nodes and $\{c_i\}$ as the i-th community. $\delta_{i,j}$ denotes the Kronecker symbol. Based on this definition, we look for a division of the graph into communities $\{c_i\}$ such that Q attains a maximum value. This undertaking constitutes a combinatorial optimization problem which can be solved by using simulated annealing or a greedy algorithm. Instead, we use an approximation proposed by Leicht & Newman (2008) which incrementally computes a maximum modularity Q by exploiting spectral properties of the graph. This algorithm is computationally efficient and produces essentially identical results as more costly algorithms.

Once the communities are established, we deflate the Markov matrix P by lumping identified communities into a single node. This is accomplished using the deflation transformation

$$\mathsf{P}^{(1)} = \mathsf{D}^T \mathsf{P} \mathsf{D}, \tag{2.3}$$

with

$$\mathsf{D}_{i,j} = \begin{cases} 1 & \text{if } i \in c_j \\ 0 & \text{otherwise} \end{cases}, \qquad i = 1, \ldots, N \qquad j = 1, \ldots, |\{c\}|. \tag{2.4}$$

The deflated matrix $\mathsf{P}^{(1)}$ contains the full (statistical) dynamics between the identified communities. By repeated application of the community-detection-deflation procedure, we arrive at a progressively compressed network that increasingly consists of fewer, but larger, communities. By maintaining the identities of the original snapshots within each community, we can reorganize the original transition probability matrix P into block-diagonal form: the (rather dense) blocks describe the motion within the community, while the (very sparse) off-diagonal blocks represent the exchange between communities. The reordering of the graph nodes that results in an optimal block-diagonal form of P is the desired result of our iterative process.

Before demonstrating the full algorithm on an example, it is worth keeping in mind that the above methodology can equally be applied to fluid behavior that, rather than being dominated by rare events, is characterized by a co-existence of multiple metastable equilibrium states with intermittent switching between them. Both for the rare-event scenario and the multi-modal setting, traditional analysis techniques often fail to capture the essential features of the flow; the above formalism may then present a better alternative.

3. Application to a self-sustaining process (SSP) model

We will demonstrate and test the above algorithm on a simple model of the self-sustaining process in wall-bounded shear flows. Wall-bounded shear flow at sufficient Reynolds numbers is widely acknowledged to support a feedback loop that maintains turbulent flow motion. It consists of streamwise vortices which, in the presence of spanwise shear, generate streamwise elongated structures (streaks). These streaks are prone to three-dimensional instabilities which cause eventual breakdown. During the reorganization phase, streamwise vortices are generated, and the cycle begins again Hamilton et al. (1995). The various elements of this model have been identified in fully developed and minimal-unit turbulence and have been proposed as the engine underlying the sustenance of turbulent fluid motion.

3.1. *The Moehlis-Faisst-Eckhart model*

Rather than processing high-dimensional numerical simulations of fluid flows, we benchmark our proposed method by considering a lower-dimensional model that captures the essential components of the self-sustaining process (SSP). This choice sidesteps the embedding of the flow-field snapshots in a suitable basis.

Various models have been proposed over the years (Waleffe 1995; Moehlis et al. 2004); we will concentrate on the model by Moehlis et al. (2004). It consists of nine time-dependent ordinary differential equations for the coefficients of the various components involved in the SSP: the base flow and its modifications, the streamwise rolls and streaks, and the three-dimensional perturbations. The system is formulated for plane channel flow and approximate trigonometric perturbation shapes are assumed for the wall-normal coordinate direction. Fourier transforms in the homogeneous streamwise and spanwise directions then allows for the expression of the interaction coefficients in a compact analytic form. The full system for the coefficients $\{a_i(t)\}$ is given by

$$\dot{a}_1 + \zeta_1 a_1 = \zeta_1 + \xi_{11} a_6 a_8 + \xi_{12} a_2 a_3, \tag{3.1a}$$
$$\dot{a}_2 + \zeta_2 a_2 = \xi_{21} a_4 a_6 - \xi_{22} a_5 a_7 - \xi_{23} a_5 a_8 - \xi_{24} a_1 a_3 - \xi_{25} a_3 a_9, \tag{3.1b}$$
$$\dot{a}_3 + \zeta_3 a_3 = \xi_{31}(a_4 a_7 + a_5 a_6) + \xi_{32} a_4 a_8, \tag{3.1c}$$
$$\dot{a}_4 + \zeta_4 a_4 = -\xi_{41} a_1 a_5 - \xi_{42} a_2 a_6 - \xi_{43} a_3 a_7 - \xi_{44} a_3 a_8 - \xi_{45} a_5 a_9, \tag{3.1d}$$
$$\dot{a}_5 + \zeta_5 a_5 = \xi_{51} a_1 a_4 + \xi_{52} a_2 a_7 - \xi_{53} a_2 a_8 + \xi_{54} a_4 a_9 + \xi_{55} a_3 a_6, \tag{3.1e}$$
$$\dot{a}_6 + \zeta_6 a_6 = \xi_{61} a_1 a_7 + \xi_{62} a_1 a_8 + \xi_{63} a_2 a_4 - \xi_{64} a_3 a_5 + \xi_{65} a_7 a_9 + \xi_{66} a_8 a_9, \tag{3.1f}$$
$$\dot{a}_7 + \zeta_7 a_7 = -\xi_{71}(a_1 a_6 + a_6 a_9) + \xi_{72} a_2 a_5 + \xi_{73} a_3 a_4, \tag{3.1g}$$
$$\dot{a}_8 + \zeta_8 a_8 = \xi_{81} a_2 a_5 + \xi_{82} a_3 a_4, \tag{3.1h}$$
$$\dot{a}_9 + \zeta_9 a_9 = \xi_{91} a_2 a_3 - \xi_{92} a_6 a_8, \tag{3.1i}$$

where the parameters in the above equations are given as

$$\zeta_1 = \frac{\beta^2}{Re}, \quad \zeta_2 = \frac{1}{Re}\left(\frac{4\beta^2}{3} + \gamma^2\right), \quad \zeta_3 = \frac{\beta^2 + \gamma^2}{Re}, \quad \zeta_4 = \frac{3\alpha^2 + 4\beta^2}{3Re},$$
$$\zeta_5 = \frac{\alpha^2 + \beta^2}{Re}, \quad \zeta_6 = \frac{3\alpha^2 + 4\beta^2 + 3\gamma^2}{3Re}, \quad \zeta_7 = \zeta_8 = \frac{\alpha^2 + \beta^2 + \gamma^2}{Re},$$
$$\zeta_9 = 9\zeta_1 \tag{3.2}$$

and

$$\xi_{11} = \xi_{62} = \xi_{66} = \xi_{92} = \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\alpha\beta\gamma}}, \quad \xi_{12} = \xi_{24} = \xi_{25} = \xi_{91} = \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\beta\gamma}},$$
$$\xi_{21} = \frac{5\sqrt{2}}{3\sqrt{3}}\frac{\gamma^2}{\kappa_{\alpha\gamma}}, \quad \xi_{22} = \frac{\gamma^2}{\sqrt{6}\kappa_{\alpha\gamma}}, \quad \xi_{23} = \xi_{53} = \frac{1}{\sqrt{6}}\kappa_2, \quad \xi_{31} = \xi_{55} = \frac{2}{\sqrt{6}}\kappa_1,$$
$$\xi_{32} = \frac{\beta^2(3\alpha^2+\gamma^2) - 3\gamma^2(\alpha^2+\gamma^2)}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta}}, \quad \xi_{41} = \xi_{45} = \xi_{51} = \xi_{54} = \xi_{61} = \xi_{65} = \xi_{71} = \frac{\alpha}{\sqrt{6}},$$
$$\xi_{42} = \frac{10\alpha^2}{3\sqrt{6}\kappa_{\alpha\gamma}}, \quad \xi_{43} = \sqrt{\frac{3}{2}}\kappa_1, \quad \xi_{44} = \sqrt{\frac{3}{2}}\frac{\alpha^2\beta^2}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta\gamma}}, \quad \xi_{52} = \frac{\alpha^2}{\sqrt{6}\kappa_{\alpha\gamma}},$$
$$\xi_{63} = \frac{10}{3\sqrt{6}}\frac{\alpha^2-\gamma^2}{\kappa_{\alpha\gamma}}, \quad \xi_{64} = 2\sqrt{\frac{2}{3}}\kappa_1, \quad \xi_{72} = \frac{1}{\sqrt{6}}\frac{\gamma^2-\alpha^2}{\kappa_{\alpha\gamma}}, \quad \xi_{73} = \frac{1}{\sqrt{6}}\kappa_1,$$
$$\xi_{81} = \frac{2}{\sqrt{6}}\kappa_2, \quad \xi_{82} = \frac{\gamma^2(3\alpha^2-\beta^2+3\gamma^2)}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta\gamma}} \tag{3.3}$$

with

$$\kappa_{\alpha\gamma} = \sqrt{\alpha^2+\gamma^2}, \qquad \kappa_{\beta\gamma} = \sqrt{\beta^2+\gamma^2}, \qquad \kappa_{\alpha\beta\gamma} = \sqrt{\alpha^2+\beta^2+\gamma^2} \tag{3.4}$$

and

$$\kappa_1 = \frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}}, \qquad \kappa_2 = \frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\alpha\beta\gamma}}. \tag{3.5}$$

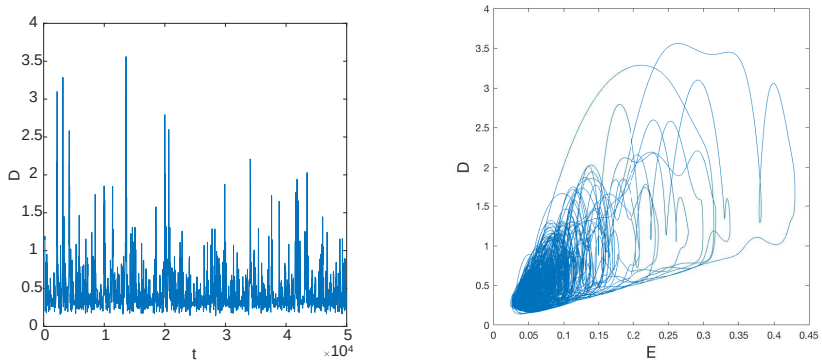
The remaining, user-specified coefficients are the Reynolds number Re, and the streamwise and spanwise wavenumbers α and γ, respectively. The wavenumber β, which determines the length scale in the wall-normal direction, is taken as constant and set to $\pi/2$.

For the test case below, we choose a box size of horizontal extent $L_x = L_y = 4\pi$, which yields $\alpha = \gamma = 1/2$, and a Reynolds number of $Re = 800$. Furthermore, we produce $5 \cdot 10^6$ snapshots based on the above system of ordinary differential equations. In order to help with visualization, we choose as observables the energy and dissipation of the perturbations, which allows for the uncompressed display of the system's dynamics. The direct processing of the nine-dimensional coefficient space will give qualitatively similar results.

Figure 2 shows a time-trace of the dissipation (subfigure a) as well as energy-dissipation phase-space trajectory (subfigure b), after eliminating an initial transient period. The intermittency of the dissipation signal reveals the occurrence of rare and violent events; these events correspond to the breakdown of coherent structures and are responsible for the bulk of the overall turbulent activity. The same characteristics are also observed in phase space where predominantly low-energy, low-dissipation behavior is complemented by rare excursions to higher energy and higher dissipation.

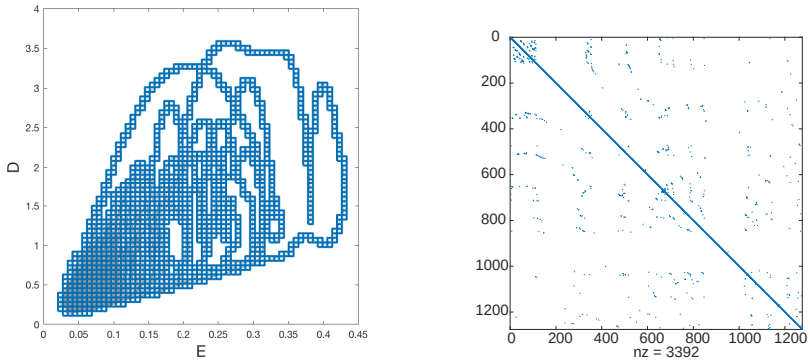
The phase space is then discretized using the search algorithm outlined above (see Figure 3a). The final discretization consists of 1275 cells, ensuring a proper resolution of the phase-space trajectory while maintaining significant subcell resolution to determine the transition probabilities with sufficient fidelity. The fill pattern of the initial transition matrix (a 1275×1275 row-stochastic matrix) is displayed in Figure 3b. It shows a strong diagonal dominance, but also noticeable entries in the off-diagonals (representing cell-to-cell transitions).

The task is then to detect communities in the corresponding network graph and to



(a) (b)

FIGURE 2. Dissipation as a function of time (a) and energy-dissipation phase-space trajectory (b) resulting from the integration of the Moehlis-Faisst-Eckhardt model for $\alpha = \gamma = 1/2$ and $Re = 800$.



(a) (b)

FIGURE 3. (a) Discretized energy-dissipation phase-space trajectory. (b) Fill pattern of the transition probability matrix P using 1275 cells.

recluster the original cells into these communities. This is performed iteratively. First, the 1275 cells are clustered into 88 communities (not shown), after which a repeated application of the algorithm detects 14 communities (see Figure 4a) and finally 4 communities (see Figure 4b). Further application of the clustering algorithm does not yield fewer communities; the procedure then terminates.

At this stage the dynamics, represented by the phase-space trajectory, has been dissected into four parts. This is indicated by the four diagonal blocks (see Figure 4b) that describe the dynamics within each identified community. The sparse entries in blocks off the block-diagonals establish connections between the detected communities. Their location in the transition probability matrix also hints at possible exit and entry points for a community-to-community transition. These cells are particularly interesting as they may provide predictive structures that are prone to triggering a rare event. These structures could be the target of control efforts to prevent bursts and instead remain in the low-energy, low-dissipation corner of phase space.

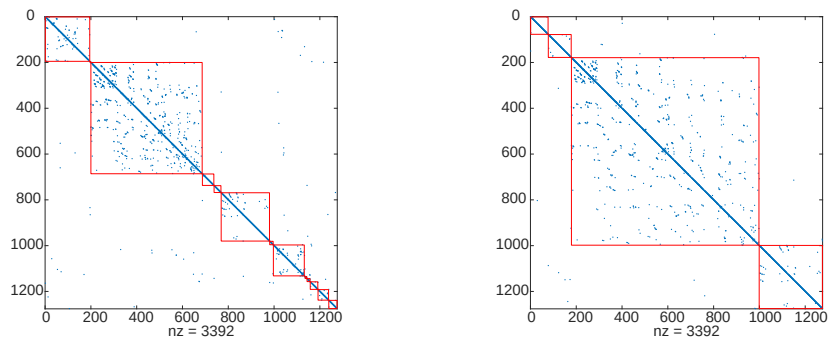


FIGURE 4. Fill pattern of the transition probability matrix P (a) after two iterations of the clustering algorithm, showing 14 identified communities (indicated by red diagonal blocks); (b) after one more iteration, displaying the final result of 4 communities.

4. Conclusions

A computational framework has been developed that allows the analysis of datasets that generally describe the interplay of distinct dynamic behaviors – among them bi-modality (as observed in the wake of bluff bodies), multi-modality or intermittent bursts from otherwise persistent motion (as observed in wall-bounded turbulence). It relies on the formulation of the flow-field evolution as a trajectory in an appropriate phase-space and the description of the dynamics in terms of a transition probability matrix between discrete phase-space elements. The matrix is then interpreted as the adjacency matrix of a directed, weighted graph, and community-detection algorithms are brought to bear to extract persistent communities (structures) and to delineate them from volatile events (bursts). The methodology has been applied to a simple nine-dimensional model of the self-sustaining process in wall-bounded flows and has shown its capability in isolating a hierarchy of dynamic communities. While still in its initial stage of development, we will continue to refine the algorithmic components and apply them to more complex and large-scale data sequences. The data-driven approach, however, holds promise for an objective and effective treatment of complex dynamic behavior, intermittent features and multimodal phenomena frequently found in fluid dynamics.

Acknowledgments

Fruitful discussions with Profs. Qiqi Wang and Gianluca Iaccarino during the group meetings are gratefully acknowledged. O.T.S. gratefully thanks Prof. Tim Colonius.

REFERENCES

BERKOOZ, G., HOLMES, P. & LUMLEY, J. L. 1993 The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.*, **25**, 539–575.

CHARTRAND, G. & ZHANG, P. 2012 *A First Course in Graph Theory.* Dover Publishing

FARAZMAND, M. & SAPSIS, T. 2017 A variational approach to probing extreme events in turbulent dynamical systems. *Sci. Adv.* **3**, e1701533.

FROYLAND, G. & PADBERG, K. 2009 Almost-invariant sets and invariant manifolds – connecting probabilistic and geometric descriptions of coherent structures in flows. *Physica D: Nonlinear Phenomena*, **238**, 1507–1523.

Hamilton, J. M., Kim, J. & Waleffe, F. 1995 Regeneration mechanisms of near-wall turbulence structures. *J. Fluid Mech.*, **287**, 317–348.

Jimenez, J. & Moin, P. 1991 The minimal flow unit in near-wall turbulence. *J. Fluid Mech.*, **225**, 213–240.

Kaiser, E., Noack, B., Cordier, L., Spohn, A., Segond, M., Abel, M., Daviller, G., Östh, J., Krajnovic, S. & Niven, R. 2014 Cluster-based reduced-order modelling of a mixing layer. *J. Fluid Mech.*, **754**, 365–414.

Lasota, A. & Mackey, M. C. 1994 *Chaos, Fractals, and Noise: Stochastic aspects of dynamics.* Springer Verlag.

Leicht, E. A. & Newman, M. E. 2008 Community structure in directed networks. *Phys. Rev. Lett.*, **100**, 118703.

Lumley, J. L. 1970 *Stochastic Tools in Turbulence.* Academic Press.

Moehlis, J., Faisst, H. & Eckhardt, B., 2004 A low-dimensional model for turbulent shear flows. *New J. Phys.*, **6**, P10008.

Nair, A. & Taira, K. 2015 Network-theoretic approach to sparsified discrete vortex dynamics. *J. Fluid Mech.*, **768**, 549–571.

Schmid, P. J., Garcia-Gutierrez, A.& Jimenez, J., 2017 Description and detection of burst events in turbulent flows, *J. Phys.: Conf. Series*, **1001**, 012015.

Sirovich, L. 1987 Turbulence and the dynamics of coherent structures. I. Coherent structures. *Q. Appl. Math.*, **XLV**, 561–571.

Ulam, S. M. 1960 *A Collection of Mathematical Problems.* InterScience Publisher, New York.

Waleffe, F. 1995 Transition in shear flows. Nonlinear normality versus non-normal linearity. *Phys. Fluids*, **7**, 3060–3066.