Cluster-Based Reduced-Order Modelling and Control for Chaotic Systems with Extreme Events

Yang, Yuxuan; Doan, Nguyen Anh Khoa

DOI
10.1007/978-3-031-97567-7_1

Publication date
2025

Document Version
Final published version

Published in
Computational Science – ICCS 2025 Workshops - 25th International Conference, 2025, Proceedings

Citation (APA)
Yang, Y., & Doan, N. A. K. (2025). Cluster-Based Reduced-Order Modelling and Control for Chaotic Systems with Extreme Events. In M. Paszynski, A. S. Barnard, & Y. J. Zhang (Eds.), *Computational Science – ICCS 2025 Workshops - 25th International Conference, 2025, Proceedings* (pp. 3-17). (Lecture Notes in Computer Science; Vol. 15910 LNCS). Springer. https://doi.org/10.1007/978-3-031-97567-7_1

Cluster-Based Reduced-Order Modelling and Control for Chaotic Systems with Extreme Events

Yuxuan Yang[1,2] and Nguyen Anh Khoa Doan[1(✉)]

[1] Faculty of Aerospace Engineering, Delft University of Technology, 2629HS Delft, The Netherlands
n.a.k.doan@tudelft.nl

[2] College of Design and Engineering, National University of Singapore, Singapore 117575, Singapore

Abstract. Chaotic systems with extreme events present significant challenges in terms of prediction and control due to their complex nonlinear dynamics and potential high dimensionality. We investigate here the use of cluster-based reduced-order modelling (ROM) and control techniques applied to such systems. In this approach, we first model the dynamics of the system by identifying clusters of similar states and only model the transition between clusters. Then, based on those identified clusters, we define a per-cluster control parameter. This effectively neglects the specific dynamics within a given cluster while retaining the main dynamics of the full-order model. The considered test case is the Moehlis-Faisst-Eckhart (MFE) system which exhibits extreme events in the form of quasi-relaminarization events. The influence of the number of clusters and the order of modelling on the accuracy of the resulting reduced-order cluster-based model is explored. A cluster-based control strategy is also proposed and applied to the MFE system to prevent extreme events. This strategy manages to achieve the objectives with a large reduction in extreme events in the controlled MFE system, decreasing the amount of time spent in extreme state by 90% and the mean kinetic energy by 20%. This work highlights the potential of cluster-based reduced-order modelling and control.

Keywords: Chaotic System · Reduced-order Modelling · Extreme Events · Clustering · System Control

1 Introduction

Most physical systems of importance to our society, from climate systems to biological ones, exhibit a chaotic dynamics that involves nonlinear interactions between multiple spatio-temporal scales [21]. This makes their study and prediction particularly challenging, a difficulty which is compounded by their high dimensionality. In addition, many of these chaotic systems can also exhibit

M. Paszynski et al. (Eds.): ICCS 2025, LNCS 15910, pp. 3–17, 2025.
https://doi.org/10.1007/978-3-031-97567-7_1

extreme events, which are abrupt short-lasting changes in the systems' state, often accompanied by grave consequences. Examples of such chaotic systems include extreme atmospheric events in our atmosphere or rogue waves in the ocean [22,23]. Therefore, there is a need to be able to predict and control these chaotic systems and their extreme events.

Traditionally, the prediction of these systems is based on numerically solving the equations governing their dynamics [19], using some form of spatio-temporal discretization schemes. However, given their high dimensionality, this approach requires large computational resources, impeding our ability to obtain quick predictions. Additionally, developing control strategies based on the full order model would become excessively challenging. Therefore, reduced-order modelling (ROM) techniques have been developed in the past decades to obtain models which can provide quick forecast of the essential evolution of the full chaotic system and be used as a basis for the development of control strategies. More recently, advances in machine learning have greatly contributed to improving the accuracy of such ROMs [24].

Reduced-order modelling techniques can generally be divided into two categories. The first category requires the knowledge of the governing equations of the system from which appropriate physics-based simplification can be made. Alternatively, the simplification can be obtained by projecting a set of truncated modes onto the governing equations, as is done in the proper orthogonal decomposition (POD)-Galerkin approach [9].

More recently, neural networks have also been used for reduced-order modelling [3]. Deep convolutional autoencoders involve neural networks that are used to obtain a reduced representation of the high dimensional data onto a latent space. These have shown a strong ability to reduce the spatial dimensions with little error when applied to turbulent flows. In addition, echo state networks have been used to learn the dynamics of the flow in that reduced space [17]. Nonetheless, these kinds of approaches may however not be easily applicable to the prediction of extreme events in chaotic systems, and nor be used as a basis for their control and prevention. This is because extreme events are short in duration, occur infrequently, and thus, are challenging to capture using such machine learning methods as they require large amount of data which may be a limiting factor considering the rarity of extreme events [4].

An alternative data-driven approach is called Cluster-based modelling, first developed by Kaiser et al. [11]. In contrast to the previous approaches, it aims to reduce the complexity of the system dynamics by only considering transitions through sets of flow states identified with clustering techniques. These clusters are typically associated with coherent structures and, therefore provide an approach in identifying relevant states that the flow takes during its evolution [7]. This represents an advantage of cluster-based modelling compared to other deep learning approaches as they identify "clusters of flow states" that can be directly targeted for control instead of relying on a full-order representation of the flow in which to devise control. Originally, this approach, called cluster-based Markov modelling (CMM), only models the transitions between clusters in a statisti-

cal sense, but recent advances like the cluster-based network model (CNM) [12] have refined such an approach by incorporating time-resolved data and modelling the underlying dynamics via a directed network. Extensions include tCNM [10], optimizing centroids for better performance, and HiCNM [5], which analyzes multi-frequency and multi-attractor behavior. In addition, cluster-based reduced order modelling has been extended for control applications, where the control strategies is defined per-cluster, as in [15,25]. This cluster-based approach appears promising for the reduced-order modelling and control of extreme events considering it is based on identifying clusters of states, which may be able to segregate between extreme and normal flow states.

In this work, we investigate this approach by using cluster-based reduced order modelling approaches for the statistical prediction and control of extreme events in chaotic systems. This is applied to the Moehlis-Faisst-Eckhart (MFE) system, which is a model of a turbulent shear flows where extreme events take the form of quasi-relaminarisation events [14].

The paper is structured as follows. First, the main features of the MFE system are briefly described. Then, the methodology for modelling and control is presented, detailing its various steps, including clustering, cluster-based Markov model, cluster-based network model, and cluster-based control. Section 4 presents the results of applying these methods to the MFE system. The paper is concluded with a summary of the main findings and directions for future work.

2 MFE System

The Moehlis-Faisst-Eckhart (MFE) system is a low-dimensional model for turbulent shear flows [14]. The MFE system models a sinusoidal shear flow in a domain $L_x \times L_y \times L_z$ (with $L_y = 2$) where the fluid flows between two free-slip walls at $y = -1$ and $y = 1$ and is subjected to a sinusoidal body force. The MFE system models this flow using nine modes whose amplitudes, a_i, are governed by ordinary differential equations provided hereunder in Eqs. (1) to (9). These modes describe the basic mean velocity profile and its modification, downstream vertices, streaks, and instabilities of streaks, with other modes being a consequence of their nonlinear interactions. The velocity field can be recovered from the modal amplitude as $\mathbf{u}(\mathbf{x}, t) = \sum_{k=1}^{9} a_i(t)\mathbf{u_i}(\mathbf{x})$ where $\mathbf{u_i}(\mathbf{x})$ are the spatial modes whose expressions can be found in [14].

$$\frac{da_1}{dt} = \frac{\beta^2}{Re} - \frac{\beta^2}{Re}a_1 - \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\alpha\beta\gamma}}a_6a_8 + \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\beta\gamma}}a_2a_3, \tag{1}$$

$$\begin{aligned}\frac{da_2}{dt} = &-\left(\frac{4\beta^2}{3} + \gamma^2\right)\frac{a_2}{Re} + \frac{5\sqrt{2}}{3\sqrt{3}}\frac{\gamma^2}{\kappa_{\alpha\gamma}}a_4a_6 - \frac{\gamma^2}{\sqrt{6}\kappa_{\alpha\gamma}}a_5a_7 \\ &- \frac{\alpha\beta\gamma}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\alpha\beta\gamma}}a_5a_8 - \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\beta\gamma}}a_1a_3 - \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\beta\gamma}}a_3a_9,\end{aligned} \tag{2}$$

$$\frac{da_3}{dt} = -\frac{\beta^2+\gamma^2}{Re}a_3 + \frac{2}{\sqrt{6}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}}(a_4a_7 + a_5a_6) + \frac{\beta^2(3\alpha^2+\gamma^2) - 3\gamma^2(\alpha^2+\gamma^2)}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta\gamma}}a_4a_8, \tag{3}$$

$$\begin{aligned}\frac{da_4}{dt} =& -\frac{3\alpha^2+4\beta^2}{3Re}a_4 - \frac{\alpha}{\sqrt{6}}a_1a_5 - \frac{10}{3\sqrt{6}}\frac{\alpha^2}{\kappa_{\alpha\gamma}}a_2a_6 \\ &- \sqrt{\frac{3}{2}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}}a_3a_7 - \sqrt{\frac{3}{2}}\frac{\alpha^2\beta^2}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta\gamma}}a_3a_8 - \frac{\alpha}{\sqrt{6}}a_5a_9,\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{da_5}{dt} =& -\frac{\alpha^2+\beta^2}{Re}a_5 + \frac{\alpha}{\sqrt{6}}a_1a_4 + \frac{\alpha^2}{\sqrt{6}}a_1a_4 + \frac{\alpha^2}{\sqrt{6}\kappa_{\alpha\gamma}}a_2a_7 \\ &- \frac{\alpha\beta\gamma}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\alpha\beta\gamma}}a_2a_8 + \frac{\alpha}{\sqrt{6}}a_4a_9 + \frac{2}{\sqrt{6}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}}a_3a_6,\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{da_6}{dt} =& -\frac{3\alpha^2+4\beta^2+3\gamma^2}{3Re}a_6 + \frac{\alpha}{\sqrt{6}}a_1a_7 + \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\alpha\beta\gamma}}a_1a_8 \\ &+ \frac{10}{3\sqrt{6}}\frac{\alpha^2-\gamma^2}{\kappa_{\alpha\gamma}}a_2a_4 - 2\sqrt{\frac{2}{3}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\beta\gamma}}a_3a_5 + \frac{\alpha}{\sqrt{6}}a_7a_9 + \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\alpha\beta\gamma}}a_8a_9,\end{aligned} \tag{6}$$

$$\frac{da_7}{dt} = -\frac{\alpha^2+\beta^2+\gamma^2}{Re}a_7 - \frac{\alpha}{\sqrt{6}}(a_1a_6 + a_6a_9) + \frac{1}{\sqrt{6}}\frac{\gamma^2-\alpha^2}{\kappa_{\alpha\gamma}}a_2a_5 + \frac{1}{\sqrt{6}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa\beta\gamma}a_3a_4, \tag{7}$$

$$\frac{da_8}{dt} = -\frac{\alpha^2+\beta^2+\gamma^2}{Re}a_8 + \frac{2}{\sqrt{6}}\frac{\alpha\beta\gamma}{\kappa_{\alpha\gamma}\kappa_{\alpha\beta\gamma}}a_2a_5 + \frac{\gamma^2(3\alpha^2-\beta^2+3\gamma^2)}{\sqrt{6}\kappa_{\alpha\gamma}\kappa_{\beta\gamma}\kappa_{\alpha\beta\gamma}}a_3a_4, \tag{8}$$

$$\frac{da_9}{dt} = -\frac{9\beta^2}{Re}a_9 + \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\beta\gamma}}a_2a_3 - \sqrt{\frac{3}{2}}\frac{\beta\gamma}{\kappa_{\alpha\beta\gamma}}a_6a_8. \tag{9}$$

In the above, Re is the Reynolds numbers and the other parameters are defined as provided in Eqs. (10) and (11).

$$\kappa_{\alpha\gamma} = \sqrt{\alpha^2+\gamma^2},\ \kappa_{\beta\gamma} = \sqrt{\beta^2+\gamma^2},\ \kappa_{\alpha\beta\gamma} = \sqrt{\alpha^2+\beta^2+\gamma^2}, \tag{10}$$

$$\alpha = \frac{2\pi}{L_x},\ \beta = \frac{\pi}{2},\ \gamma = \frac{2\pi}{L_z}. \tag{11}$$

For our considered case, the size of the domain is set to $L_x = 4\pi$, $L_z = 2\pi$ and the Reynolds number is set to 800 to ensure a complete turbulent system, as done in [20]. The MFE system was simulated using the Runge-Kutta 4 method,

for a duration of 50000 time units, using a timestep $\Delta t = 0.25$. The evolution of the nine modal coefficients is shown in Fig. 1 where their chaotic evolution can clearly be seen. The associated time evolution of the kinetic energy $k = \frac{1}{2}\sum_{i=1}^{9} a_i^2$ and dissipation rate $\epsilon = \int_V |\nabla \times \mathbf{u}|^2 dV$, with V being the volume of the domain, are shown in Fig. 2. In this figure, extreme events can be readily observed, corresponding to the sudden spikes in kinetic energy and dissipation rate. These two quantities will be used as indicators of extreme events.

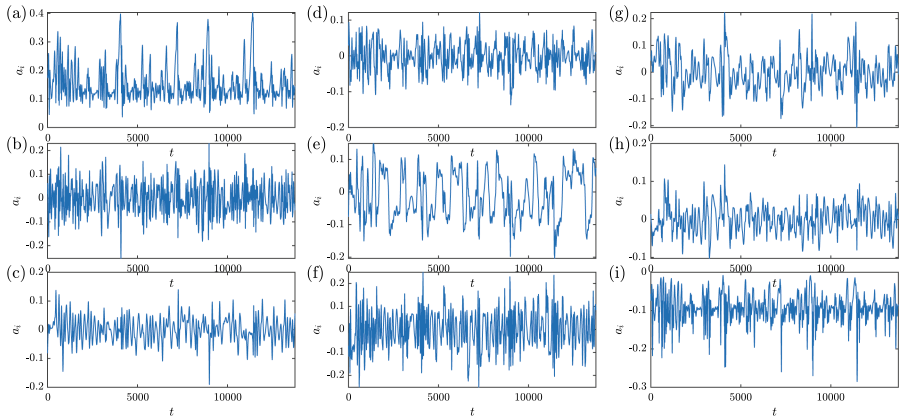

Fig. 1. Time series of the nine modal coefficients, (a) a_1 to (i) a_9 of the MFE system.

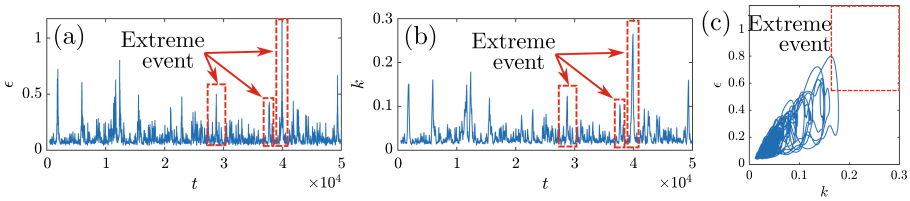

Fig. 2. Time evolution of (a) ϵ and (b) k in the MFE system and (c) its representation in the k-ϵ space. Some extreme events are highlighted using red arrow and boxes. (Color figure online)

Projecting the MFE system onto the $k-\epsilon$ phase space reveals extreme events as excursions far from the origin with high kinetic energy and dissipation rates, as shown in Fig. 2c. In this space, the control objective will be defined as preventing evolutions of the system that evolve towards that high kinetic energy/high dissipation region, effectively maintaining the system closer to the low energy/low dissipation region. This will be elaborated in Sect. 4.2.

3 Methodology

From the data obtained by simulating the MFE system, clustering using k-means++ is applied in the phase space composed by the nine modal coefficients, a_1 to a_9. The clustered system is then modeled using two methods: Cluster-based Markov Modelling (CMM) [11] and Cluster-based Network Modelling (CNM) [12]. Additionally, the impact of modelling parameters on CNM results is analyzed. Furthermore, a cluster-based control algorithm is developed based on the cluster-based model to achieve a controlled chaotic system where extreme events are prevented. Each step is now described in details.

3.1 Clustering

First, the k-means++ clustering algorithm [13] is used in this study to cluster the snapshots in the phase space. The k-means algorithm is an unsupervised classification algorithm where samples (or snapshots) are partitioned into a prescribed N_{cl} number of clusters. Every cluster, denoted as $\mathcal{C}_k$, is characterized by its centroid, c_k, computed as the mean over all samples within that cluster. These centroids will serve to simplify the dynamics in the phase space, as the dynamics and evolution within a cluster will be neglected and the model will solely focus on the transition between clusters.

The k-means algorithm can sometimes converge to a local optimum, making initialization crucial. K-means++ improves this by spreading initial centroids, reducing convergence time despite higher initial computational cost [1].

To select an appropriate number of clusters, N_{cl}, we use the elbow method based on the within-cluster sum of square (WCSS, noted J_m) [8], defined in Eq. (12). Practically, as the number of clusters N_{cl} increases, J_m decreases as a result of a decrease in the variance within a larger number of clusters, with an initial steep decrease in J_m for small initial increment in N_{cl}. However, as N_{cl} further increases, this decrease slows down as each increment in the number of clusters results in smaller decrease in within-cluster variance, as will be shown in Fig. 4. This results in a so-called elbow curve when plotting J_m as a function of N_{cl} which allows to pick N_{cl}.

$$J_m(N_{cl}) = \frac{1}{N} \sum_{k=1}^{N_{cl}} \sum_{s \in \mathcal{C}_k} \|s - c_k\|^2 \tag{12}$$

In Eq. (12), N represents the total number of samples in the dataset.

3.2 Cluster-Based Reduced-Order Modelling

Two different cluster-based reduced-order modelling techniques are considered in this study: Cluster-based Markov Modelling (CMM) [11] and Cluster-based Network Modelling (CNM) [12]. Both are now briefly explained.

Cluster-based Markov model simplifies the evolution of the dynamical system into transitions between each cluster that follows a Markov process.

The transitions are elucidated as a cluster transition matrix (CTM), which serves as the propagator in terms of probability.

Since we assume that the dynamical system is a Markov process, the state in the next time step is only dependent on the state at the current time step. Firstly, the state at a given time t is defined as a vector, $\mathbf{p}(t)$, that represents the probability of the snapshots to fall in each cluster:

$$\mathbf{p}(t) = [p_1, ..., p_{N_{cl}}]^T. \tag{13}$$

The CTM is then noted as $\mathbf{P}$ where its element P_{ij} represents the probability of transitioning from cluster $\mathcal{C}_i$ to $\mathcal{C}_j$ in one forward time-step. The elements of the resulting CTM can thus be estimated as

$$P_{ij} = \frac{N_{ij}}{N_i}, \tag{14}$$

where N_{ij} is the number of snapshots that move from cluster $\mathcal{C}_i$ to cluster $\mathcal{C}_j$, and N_i is the number of snapshots in cluster $\mathcal{C}_i$. With this CTM, the dynamic property can be identified. The probability of state at time step $n+1$ is then equal to the state at time step n times the CTM $\mathbf{P}$.

$$\mathbf{p}^{n+1} = \mathbf{p}^n \mathbf{P}, \quad n = 0, 1, 2... \tag{15}$$

The long-term behavior can then be analyzed by iteratively multiplying by the CTM. The asymptotic probability distribution is therefore obtained as

$$\mathbf{p}^\infty := \lim_{n\to\infty} \mathbf{P}^n \mathbf{p}^0. \tag{16}$$

In contrast to the method above that relies on transition probabilities, cluster-based network (CNM) model resolves the deterministic transition time and allows the model to predict the state as a function of time [12]. Discarding the critical transition time step in CMM, CNM defines two important time parameters: the residence time in a cluster and the transition time between two clusters. First, the residence time in a given cluster, after the n-th transitions between clusters, is defined as

$$\tau_n = t_{n+1} - t_n. \tag{17}$$

In the above, t_n represents the time instant in which the system has entered the considered cluster, the subscript n indicating that it is the n-th transition from a cluster to another the system has seen so far. t_{n+1} then indicates the exit time when the system exits that considered cluster (and therefore enters another cluster). This is illustrated in Fig. 3.

Let j and i be the indices of the clusters after t_n and t_{n+1} respectively. Then the transition time from j to i is defined as half of the residence times in each cluster:

$$\tau_{ij} = \frac{\tau_n + \tau_{n+1}}{2}. \tag{18}$$

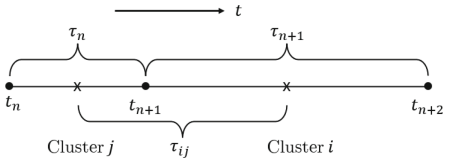

Fig. 3. Sketch of times and periods employed in the cluster-based network model [12].

The direct transition time T_{ij} from cluster j to cluster i is defined as the average of all values of τ_{ij} observed in the considered dataset:

$$T_{ij} = \langle \tau_{ij} \rangle. \tag{19}$$

In CNM, the cluster transition matrix is also further transformed to account for multiple previous states, quantified by the model order L. This order represents the number of clusters visited before the current state to be considered in the transition dynamics. This leads to a more complex CTM with more rows than columns, where the "from states" are sequences of L clusters influencing the transition, and the calculation follows a similar process to that in Eq. (14).

For each cluster transition, the next cluster is selected based on the CTM probability, and the direct transition time is added to the total time, with states linearly interpolated between the centroids during this period. The prediction for $t \in [t_n, t_{n+1}]$ is then given by

$$\boldsymbol{u}(t) = \alpha_n(t) c_{k_n} + [1 - \alpha_n(t)]\, c_{k_{n+1}}, \quad \alpha_n = \frac{t_{n+1} - t}{t_{n+1} - t_n}. \tag{20}$$

where c_{k_n} and $c_{k_{n+1}}$ represent the centroids of the cluster that the system enters at time $t = t_n$ and $t = t_{n+1}$, respectively.

3.3 Cluster-Based Control (CBC) Algorithm

The specific control actuation for the MFE will be inspired by what has been done in past literature where the control method was based on a modification of the Reynolds number in time, dependent on the flow state as in [18]. Physically, such a control actuation would correspond to a modification of the magnitude of the volume forcing term in this shear flow modelled by the MFE system. In this work, we will set up the actuation as a coefficient, b, in front of the Reynolds number in the governing equations presented in Sect. 2. The value of this coefficient b will depend on the system state in the phase space and the control parameters, b_k which are defined per cluster $\mathcal{C}_k$, i.e.: $b(t) = \mathcal{F}(\mathbf{a}(t); b_k)$.

Specifically, following the method proposed in [15], the coefficient b is defined here as a proportional feedback dependent on the current state in the feature space, $\mathbf{a}(t)$, and is defined as:

$$b\left(\mathbf{a}(t); b_k\right) = \frac{\sum_{k=1}^{N_{cl}} b_k \mathrm{e}^{-\|\mathbf{a}(t) - c_k\|^2 / J_i}}{\sum_{k=1}^{N_{cl}} \mathrm{e}^{-\|\mathbf{a}(t) - c_k\|^2 / J_i}}, \tag{21}$$

where J_i denotes the inter-clusters variance defined as:

$$J_i = \frac{1}{N} \sum_{k=1}^{N_{cl}} N_k \|c_k - \bar{c}\|^2, \tag{22}$$

with $\bar{c} = \frac{1}{N} \sum_{k=1}^{N_{cl}} N_k c_k$ and N being the total number of samples in the dataset, and N_k the number of samples in cluster $\mathcal{C}_k$. The value of b_k is constant per each cluster $\mathcal{C}_k$, with values clipped to a maximum of 5. The objective of the control strategy is to find optimal values of b_k in Eq. (21), given an adequately defined objective function, $\mathcal{J}$.

To solve this nonlinear multivariate optimization problem, following [15,25], the simplex search method described by Nelder and Mead [16] is chosen given its effectiveness in optimizing smooth functions without requiring gradient calculations. This method relies on a linear combination of $N_c + 1$ set of parameters, called vertices, in a N_c dimensional optimization space. In our case, N_c corresponds to the number of cluster N_{cl} and a vertex would therefore represent a set of b_k values for control. The vertices form a simplex (in the control space) that is reflected and stretched following the local gradients with four operations: reflection, expansion, contraction and shrink. These operations generate, respectively 1, 1, 1, and N_c new vertices. In [16], the four operations are applied sequentially following a deterministic procedure based on the performance of the newly generated vertices. After each operation, the new vertices replace the least performing ones in the simplex if they perform better (i.e. if the new vertices have a better score as measured by the objective function $\mathcal{J}$). The process is repeated until a stopping criterion is reached. Once the amplitudes b_k are determined, the controller can guide the system to achieve the control objective. Interested readers are referred to [16,25] for additional details.

4 Results

In this section, we discuss first the accuracy of the different cluster-based reduced order modelling techniques (CMM and CNM) in function of the number of clusters N_{cl} and the order of the model L for the CNM. Subsequently, we analyze the effectiveness of the cluster-based control in mitigating the occurrence of extreme events.

4.1 Cluster-Based ROM Results

First, as discussed in Sect. 3.1, an optimal number of clusters N_{cl} is determined for the MFE system. The evolution of the within-cluster sum of squares, J_m, with N_{cl} is shown in Fig. 4. It can be observed that the reduction in J_m becomes much less pronounced after N_{cl} exceeds 80, indicating diminishing returns when considering more clusters. Therefore, the optimal number of clusters N_{cl} is selected at 80 in what follows (unless mentioned otherwise), which balances model complexity and accuracy. This number of clusters N_{cl} will also be used for further

cluster-based control in Sect. 4.2. It should be noted that although the modelling and clustering cost increases with the number of clusters, the computing time never exceeded a few minutes on our workstation equipped with an Intel Xeon W7-2475X CPU.

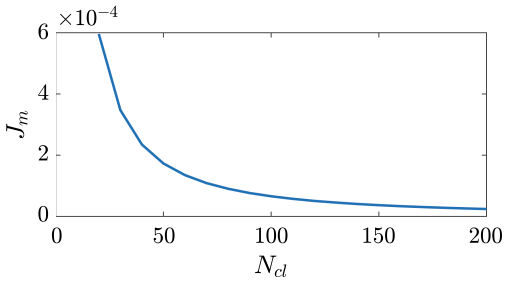

Fig. 4. Evolution of J_m with changes in N_{cl} in the clustering of the MFE system.

Starting with CMM-based model, the associated cluster transition matrix, as defined in Eq. 14, is shown in Fig. 5a using a logarithmic scale, with darker shades representing higher values. The logarithmic scale is used due to the dominance of the diagonal terms as between two consecutive timesteps the system is very likely to stay within the same cluster. This also highlights a potential drawback of the CMM method, as predictions based on the CTM may often result in the system getting stuck in the same cluster.

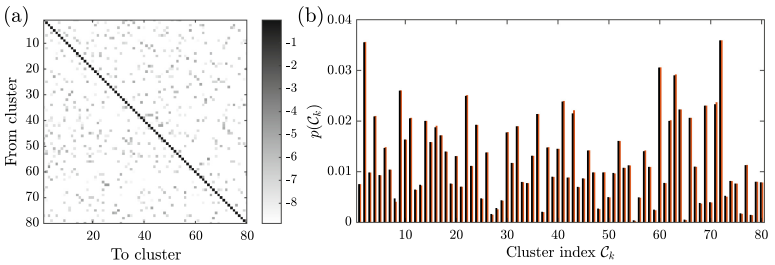

Fig. 5. (a) Transition probability matrix (in log-scale) of the MFE system obtained after clustering. (b) PDF of the system to be in a given cluster obtained from the training data (black) and by the CMM (red). (Color figure online)

Figure 5b compares the actual distribution of the MFE system across clusters (black bars) with the distribution obtained by using the CMM model to make a long-term prediction of 10000 time units (red bars). The strong agreement between the two confirms the statistical accuracy of the CMM model, validating the clustering process.

In the context of the CNM method, two main parameters are studied here to analyze their effects on the model: the order of the model L and the number of clusters N_{cl}, which are evaluated from a statistical perspective by analyzing the predicted distribution of the modes.

The distribution of a_1, a_2 and a_3 for fixed order of model and different number of clusters can be found in Fig. 6. These are obtained by using the CNM model to make a long-term prediction of the evolution of the MFE system, from which the probability density function (pdf) of the various modal coefficients are estimated. For $N_{cl} = 40$, compared to the real distribution, noticeable peaks can be observed in the predicted ones (orange curves). These corresponds to the centroids of the clusters which are over-represented. Furthermore, the predicted distributions are more narrow indicating that the CNM model lacks the ability to accurately capture the tail statistical distribution of the system's states from a continuous perspective. This is intrinsically linked to the small number of clusters used and the fact that the CNM interpolates the system state between the centroids of the clusters therefore not navigating the entire phase space. With the increase of N_{cl}, the agreement between the orange and blue curves improves as a result of the finer description of the phase space through more clusters, as can be observed when comparing Fig. 6a to Fig. 6b and c, which corresponds to cases with $N_{cl} = 80$ and 160 respectively.

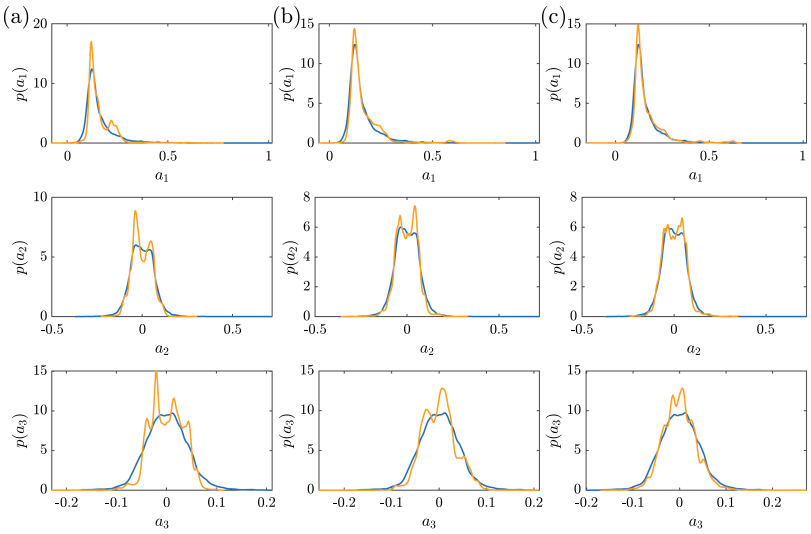

Fig. 6. Distribution of a_i for CNM with (a) $N_{cl} = 40$ and $L = 3$, (b) $N_{cl} = 80$ and $L = 3$ and (c) $N_{cl} = 160$ and $L = 3$ (blue: true distribution, orange: CNM). (Color figure online)

The effect of the variation of the model order L is shown in Fig. 7, where $L = 3$, 20 and 40 are considered. It is observed that there is no noticeable

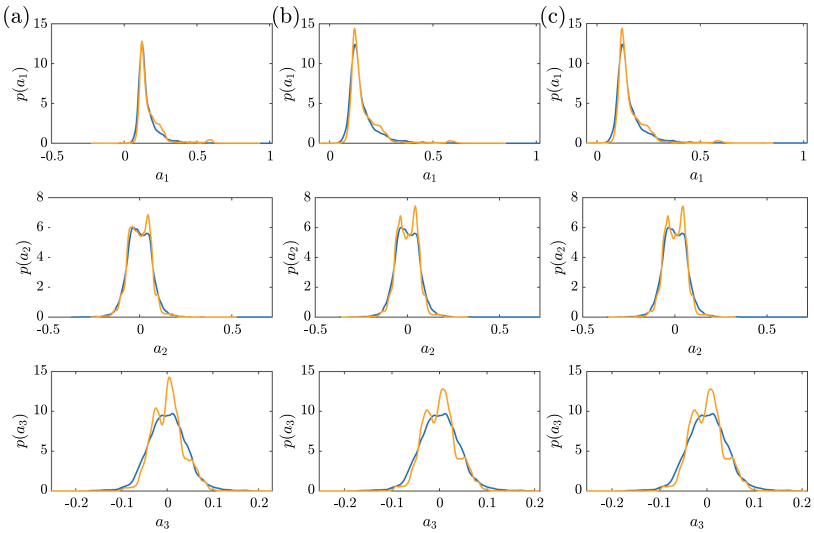

Fig. 7. Distribution of a_i for CNM with (a) $N_{cl} = 80$ and $L = 3$, (b) $N_{cl} = 80$ and $L = 20$, and (c) $N_{cl} = 80$ and $L = 40$ (blue: true distribution, orange: CNM). (Color figure online)

improvement in the predicted distribution of the modes a_i, indicating that knowing just in which clusters the system was in just a few preceding transition is sufficient in enabling the prediction of which cluster the system will enter.

Finally, Fig. 8 shows time series of the MFE system predicted by the CNM with $N_{cl} = 80$ and $L = 20$. Comparing Fig. 8 to Fig. 1, it can be observed that the CNM reproduces qualitatively well the evolution of the MFE system beyond the evidence provided in the distributions predicted by the CNM in Fig. 7. It should be noted that no comparison is done with the CMM approach as CMM only models the transition between cluster from a statistical view and does not provide a time-prediction of the evolution of the system during those transitions.

4.2 Cluster-Based Control Results

As the objective of our control test is to mitigate the occurrence of extreme events, we define the objective function as the minimization of the time-averaged dissipation rate of the flow during its controlled evolution:

$$\mathcal{J}(t) = \frac{1}{t_c} \int_0^{t_c} \epsilon(t) dt. \tag{23}$$

In Eq. (23), t_c represents the time over which the MFE is simulated for a given set of control parameters b_k and from which the cost function is estimated. It should be noted that having a cost function purely based on the energy dissipation rate should also act in preventing the possibility of high kinetic energy states, as it is those specific states that then yield high energy dissipation rate values.

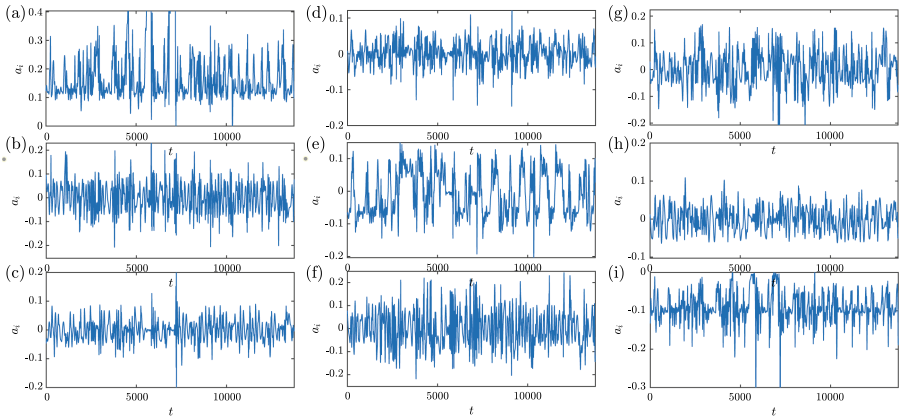

Fig. 8. Evolution of the MFE modal coefficients (a) a_1 to (i) a_9 predicted by the CNM with $N_{cl} = 80$ and $L = 20$.

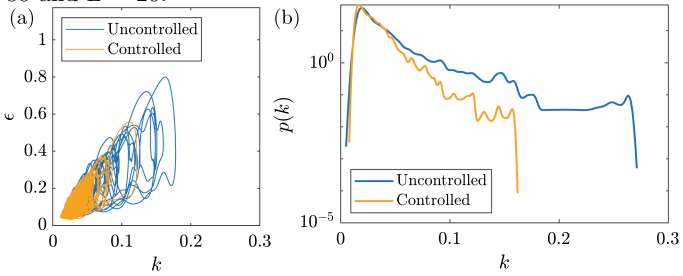

Fig. 9. (a) k-ϵ plot for dissipation rate control. (b) Probability density function of the uncontrolled and controlled MFE system.

Using the downhill simplex method described in [16], the cluster-specific control parameters b_k are determined, with a test duration of $t_c = 20LT$ (Lyapunov time), where the Lyapunov time, LT is defined as in [2] and has a value of ~ 40.98 for the MFE system [6]. Solving this optimization problem took approximately 6 h using a workstation equipped with an Intel Xeon W7-2475X CPU.

The evolutions of the uncontrolled and controlled MFE systems projected in the k-ϵ plane are shown in Fig. 9a, demonstrating the control algorithm's effectiveness in reducing extreme events. The controlled system mostly evolves in the low-k and low-ϵ region and there are no large "loops" towards the high k/high ϵ region. This is also confirmed by the distribution of k shown in Fig. 9b for the uncontrolled and controlled MFE systems. It can be seen that the tail of the pdf of k is greatly reduced in the controlled case confirming further that the control approach achieved a reduction in extreme events in the MFE system.

We further analyze this effectiveness in time and statistically. First, we define the system to be in an extreme state if its kinetic energy, k, is above 0.1. This threshold value is chosen as most extreme events exhibit peaks beyond this value, as can be seen in Fig. 2. Second, we compute the percentage of time the uncontrolled and controlled systems spend in such a state (of $k > 0.1$).

From this approach, the uncontrolled system is estimated to spend 2.96% of its evolution in an extreme state while the controlled one only spends 0.29% demonstrating the ability in the identified control in decreasing the occurrence of extreme events. Finally, the time-averaged kinetic energy of the uncontrolled and controlled systems are 0.0332 and 0.0263 respectively, further showing that the control approach effectively maintains the system in a k state preventing extreme events.

5 Conclusion

In this work, we investigated the use of cluster-based reduced-order modelling and control for chaotic systems that exhibit extreme events. Two kinds of cluster-based ROM techniques were applied on the MFE system, and satisfactory modelling accuracy were obtained with both methods. The influence of key parameters on the cluster-based reduced-order modelling was evaluated from a statistical perspective. Compared to earlier work [11], a quantitative approach to determine the optimal number of clusters N_{cl} was used, ensuring a balance between model accuracy and computational efficiency. The control results of the MFE system validate the effectiveness of the cluster-based control algorithm. By minimizing the mean dissipation rate, the approach significantly reduces extreme events and keeps both kinetic energy and dissipation rate at lower levels. Quantitative results show a substantial decrease in mean kinetic energy and a significant reduction in extreme events. These findings underscore the potential of cluster-based control methods in controlling chaotic systems, especially when analytical models are difficult to obtain.

Future research will focus on addressing the main limitation of this approach which lies in extending it to higher dimensional flow where obtaining an appropriate state space in which to perform the clustering is crucial. This will be done by using deep learning-based reduced-order representation techniques such as deep autoencoder. Additionally, determining the optimal model order for cluster-based modelling and control, either through governing equations or data-driven machine learning techniques, represents another important avenue for further exploration.

References

1. Arthur, D., Vassilvitskii, S., et al.: k-means++: the advantages of careful seeding. Soda 7, 1027–1035 (2007)
2. Boffetta, G., Cencini, M., Falcioni, M., Vulpiani, A.: Predictability: a way to characterize complexity. Phys. Rep. **356**(6), 367–474 (2002)
3. Brunton, S.L., Noack, B.R., Koumoutsakos, P.: Machine learning for fluid mechanics. Annu. Rev. Fluid Mech. **52**(1), 477–508 (2020)
4. Chen, N., Majda, A.J.: Predicting observed and hidden extreme events in complex nonlinear dynamical systems with partial observations and short training time series. Chaos **30**, 033101 (2020)

5. Deng, N., Noack, B.R., Morzyński, M., Pastur, L.R.: Cluster-based hierarchical network model of the fluidic pinball-cartographing transient and post-transient, multi-frequency, multi-attractor behaviour. J. Fluid Mech. **934**, A24 (2022)
6. Doan, N., Polifke, W., Magri, L.: Short- and long-term predictions of chaotic flows and extreme events: a physics-constrained reservoir computing approach. Proc. R. Soc. Lond. A **477**, 20210135 (2021)
7. Fernex, D., Noack, B.R., Semaan, R.: Cluster-based network modeling—from snapshots to complex dynamical systems. Sci. Adv. **7**(25), eabf5006 (2021)
8. Goutte, C., Toft, P., Rostrup, E., Nielsen, F.Å., Hansen, L.K.: On clustering fMRI time series. Neuroimage **9**(3), 298–310 (1999)
9. Holmes, P.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press (2012)
10. Hou, C., Deng, N., Noack, B.R.: Trajectory-optimized cluster-based network model for the sphere wake. Phys. Fluids **34**(8) (2022)
11. Kaiser, E., et al.: Cluster-based reduced-order modelling of a mixing layer. J. Fluid Mech. **754**, 365–414 (2014)
12. Li, H., Fernex, D., Semaan, R., Tan, J., Morzyński, M., Noack, B.R.: Cluster-based network model. J. Fluid Mech. **906**, A21 (2021)
13. Lloyd, S.: Least squares quantization in PCM. IEEE Trans. Inf. Theory **28**(2), 129–137 (1982)
14. Moehlis, J., Faisst, H., Eckhardt, B.: A low-dimensional model for turbulent shear flows. New J. Phys. **6**(1), 56 (2004)
15. Nair, A.G., Yeh, C.A., Kaiser, E., Noack, B.R., Brunton, S.L., Taira, K.: Cluster-based feedback control of turbulent post-stall separated flows. J. Fluid Mech. **875**, 345–375 (2019)
16. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965)
17. Racca, A., Doan, N., Magri, L.: Predicting turbulent dynamics with the convolutional autoencoder echo state network. J. Fluid Mech. **975**, A2 (2023)
18. Racca, A., Magri, L.: Data-driven prediction and control of extreme events in a chaotic flow. Phys. Rev. Fluids **7**(10), 104402 (2022)
19. Reynolds, O.: Iv. on the dynamical theory of incompressible viscous fluids and the determination of the criterion. Philos. Trans. R. Soc. Lond. A **186**, 123–164 (1895)
20. Schmid, P., Schmidt, O., Towne, A., Hack, M.: Analysis and prediction of rare events in turbulent flows. In: Proceedings of the Summer Program. Center for Turbulence Research (2018)
21. Strogatz, S.H.: Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering. CRC Press (2018)
22. Thompson, V., et al.: The 2021 Western North America heat wave among the most extreme events ever recorded globally. Sci. Adv. **8**(18), eabm6860 (2022)
23. Tomassini, L., Gerber, E.P., Baldwin, M.P., Bunzel, F., Giorgetta, M.: The role of stratosphere-troposphere coupling in the occurrence of extreme winter cold spells over northern Europe. J. Adv. Model. Earth Syst. **4**(4) (2012)
24. Vinuesa, R., Brunton, S.L.: Emerging trends in machine learning for computational fluid dynamics. Comput. Sci. Eng. **24**(5), 33–41 (2022)
25. Wang, X., Deng, N., Cornejo Maceda, G.Y., Noack, B.R.: Cluster-based control for net drag reduction of the fluidic pinball. Phys. Fluids **35**(2) (2023)