Proceedings of the Combustion Institute 40 (2024) 105524

Contents lists available at ScienceDirect

Proceedings of the Combustion Institute

journal homepage: www.elsevier.com/locate/proci

Data-driven identification of precursors of flashback in a lean hydrogen reheat combustor

Mihnea Floris [a], Tadikonda Shiva Sai [b,1], Dibyajyoti Nayak [b,1], Ivan Langella [a], Konduri Aditya [b,*], Nguyen Anh Khoa Doan [a]

[a] Faculty of Aerospace Engineering, Delft University of Technology, Delft, Netherlands
[b] Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, India

ARTICLE INFO

Keywords:
Flashback
Hydrogen combustion
Precursor identification
Clustering
Featurization

ABSTRACT

In this work, we propose a data-driven framework to identify precursors of extreme events in turbulent reacting flows. Specifically, we tackle the problem of flashback prediction in a lean hydrogen reheat combustor. Our framework is composed of two parts. The first consists in the use of a co-kurtosis based approach to identify the components of the thermochemical and flow state which are the most relevant for the onset of flashback. This allows for an efficient low-dimensional representation. From this reduced representation, a modularity-based clustering algorithm is then employed to segregate between clusters which contain normal and extreme (flashbacking) states, and the cluster located in-between these states, which are the precursor states of extreme events. We show that this method is able to identify the most important features at the onset of flashback in the considered reheat combustor and then provide precursor states based on those. The prediction time obtained with the identified precursors is relatively large when compared to the duration over which the combustor is stable. Additional analyses on the specific choice of features for the precursor identification and the sampling locations are made. The robustness of the method when using shorter time series to identify the precursor is also investigated. Results show that the method is generally robust with respect to such changes. A first step towards practical measurements is also attempted with wall pressure measurements, which shows only a moderate reduction in prediction time. This work proposes for the first time a data-driven technique to automatically identify precursors of flashback in hydrogen combustion opening the path for such applications on other extreme events in reacting flows.

1. Introduction

Many combustion systems can exhibit extreme events which are events where the combustor exhibits large excursions away from its design point, potentially resulting in severe damages [1]. Examples of such extreme events include engine unstart in scramjet, where the shock necessary for flame stabilization is ejected from the scramjet [2], flame blow-out where the flame extinguishes [3] or flashback in hydrogen combustors, where the flame suddenly moves upstream into the mixing duct [4].

While one can try to design the combustor to avoid such events, it is rarely possible to make such designs entirely fail-proof [1]. First, turbulent systems are chaotic, which implies that identifying and understanding the infinitesimal perturbations at the origin of extreme events is challenging. Second, combustors are high-dimensional systems, which span a wide range of spatio-temporal scales and are often coupled with other components. Therefore, extreme events can originate from many different paths which cannot all be accounted for at the design phase. Given these complexities and potential grave consequences, it is necessary to develop methods that can reliably predict the occurrence of such events in combustors. However, achieving the prediction of extreme events has been quite limited owing to the complexity of their origin and the limited observations we have of those [5]. For some specific extreme events in reacting flows, such as thermoacoustic instabilities, some precursors could be identified from physical considerations [6]. However, for many others, reliable techniques are still needed and advances on this understanding is still limited in the literature. Such advances can be provided by machine learning techniques given their ability in analysing and identifying patterns in large datasets [7,8].

For methods based on machine learning, two different capabilities are needed: (i) featurization to obtain a latent reduced representation

* Corresponding author.
E-mail address: konduriadi@iisc.ac.in (K. Aditya).
[1] Equal contributions to this paper.

https://doi.org/10.1016/j.proci.2024.105524
Received 4 December 2023; Accepted 25 June 2024
Available online 26 July 2024

and (ii) precursor identification in that reduced representation. Regarding featurization, the high-dimensionality of reacting flows makes their analysis at the onset of extreme events very challenging, Thus data-driven techniques capable of identifying which features (or reduced number of combined features) are the most relevant in the early stage of extreme events, is crucial. Principal Component Analysis (PCA) has for example been used to project the thermochemical state onto a lower dimension [9,10]. However, PCA is only a linear method that identifies a reduced space based on maximizing the explained variance and extreme states may not be easy to spot with such a method [11–13].

For precursor identification, different approaches have been developed in the past years which are mostly *ad-hoc* methods. For example, for thermoacoustic instabilities, precursor identification techniques have relied on the multifractality concept [6]. However, such methods are by definition not generalizable to other forms of extreme events in reacting flows. More general alternatives have reframed the precursor identification problem into an optimization problem with partial differential equations as constraints. In this case, precursor states are identified as the states which show the largest "growth" (according to a quantity linked to extreme events) [14]. This method has so far only been applied to non-reacting flows and is an expensive technique requiring an adjoint-based solver to solve the optimization problem. Another line of methods, of interest in this work, has relied on a graph-based interpretation of the evolution of a flow combined with community clustering to identify precursors of extreme events [15,16]. This was shown to be effective and relatively cheap on non-reacting flows and be applicable for various extreme events. However, so far, no precursor identification technique has shown to be applicable for reacting flows in a general manner.

In this work, we propose and demonstrate a combination of data-driven techniques, based on co-kurtosis [11–13] and modularity clustering [15,16], which can help in the physical investigation of the cause of extreme events and in the identification of their precursors. This framework is applied to the case of a lean hydrogen turbulent reheat combustor which exhibits intermittent flashback. The remainder of this paper is organized as follows. Section 2 provides details on the case considered. Section 3 details the featurization method based on the co-kurtosis approach and the precursor identification technique based on modularity clustering used in this work. Section 4 presents the features identified and the resulting identified precursors, including the time warning they provide alongside further investigations on the performance of the clustering technique. The final section summarizes the main results and outlines avenues for future work.

2. Test case: reheat hydrogen combustor with intermittent flashback

The configuration considered here is a simplified version of the Ansaldo Energia GT36 reheat combustor, as in [17], where both autoignition and deflagration play a role in the flame stabilization. Such reheat configurations are typically designed to exploit autoignition flame properties to ensure stability in a high velocity flows and have been studied in past work like [17,18] using physics-based approaches.

The mid-z plane is shown in Fig. 1 alongside the boundary conditions considered. Compared to [17], a higher pressure of 20 atm is considered instead of 1 atm and the inflow temperature is increased from 1100 K to 1180 K. The inflow velocity profile is set as a uniform profile with a magnitude of 200 m/s. The turbulence intensity at the inlet is set to 10% using the method in [19]. The reactant mixture consists of premixed hydrogen/air with an equivalence ratio of 0.35. With these conditions, intermittent flashback is observed as shown in Fig. 2 which will be discussed below.

Our numerical set-up is very similar to the LES set-up used in [20] which studied the same configuration (at low pressure) and whose LES results were validated against the DNS data of [17]. Compared to [20], we appropriately refined the mesh to resolve at least 80% of the subgrid

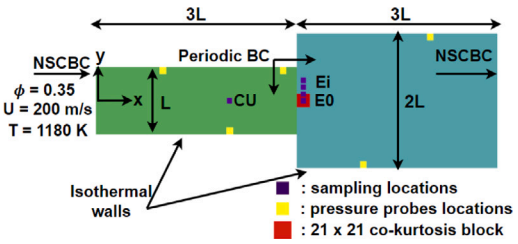

Fig. 1. Geometry of the reheat combustor shown in the mid-z plane. $L = 1$ cm [20].

kinetic energy (Pope's criterion) everywhere in the domain, except near walls where wall functions are used and performed several sensitivity analyses. This resulted in a mesh with approximately 10 million cells which is considered to be sufficient to appropriately resolve the main combustion dynamics. The simulation is performed using LES with the artificial thickened flame model [21]. NSCBC boundary conditions are applied at the inlet and outlet to properly account for pressure waves propagation and non-slip isothermal walls at a temperature $T = 750$ K are considered. The Converge CFD solver is used with the PISO loop, in conjunction with the SOR method and Rhie-Chow interpolation to avoid pressure–velocity decoupling. The hydrogen-air chemical kinetics is evaluated using the 9-species 21-reactions mechanism of [22].

The simulation is run for a time long enough to collect eight flashback events. It should be noted that those eight flashback events are not identical to one another. Indeed, our set-up has partially non-reflective boundary conditions, which will have a "rubberband" effect on the pressure waves. These waves are partially reflected at the boundaries subsequently inducing random pressure/velocity fluctuations. This ensures some stochasticity making each flashback event different. This will be shown later in Fig. 7 where the duration of each flashback event is shown to be different. 2D snapshots of the mid z-plane are saved at a frequency of 1 MHz (to well capture the flashback dynamics) resulting in 2500 snapshots for the analysis in Section 4.

From physical analysis, the following sequence leading to flashback was observed. First, the flame is stabilized on the centreline by means of autoignition, with its base located at the exit of the mixing tube, where the flame is stabilized in an ignition-assisted mode due to the recirculation zones at the expansion. This is the designed point of operation, which results from the balance between the autoignition and flow-through times. The process of autoignition expands locally the gases, inducing pressure waves both downstream and upstream at the combustor walls. The pressure waves converge at the centreline, forming a constructive interference pattern, which leads to a positive temperature fluctuation due to compressive heating (Fig. 2b). This fluctuation ultimately gives rise to an early autoignition event (Fig. 2b where the slightly higher temperature spot can be seen in the mixing tube accompanied by the resulting pressure waves). Subsequently, the pressure waves travel upstream into the combustion chamber heating the inflowing mixture, inducing their early ignition (through a decrease of their ignition delay time), eventually resulting in reactions in the mixing tube.

3. Methodology

To predict the occurrence of flashback in the reheat combustor configuration, two methods are combined. First, a featurization approach based on co-kurtosis, as introduced in [11–13], is applied to derive the features that have the largest contribution towards the occurrence of flashback. Thereafter, based on the findings, the precursor identification technique based on modularity clustering [15,16] is used to identify the precursor states. This workflow is illustrated in Fig. 3. The aforementioned techniques are elaborated in the subsequent sections.

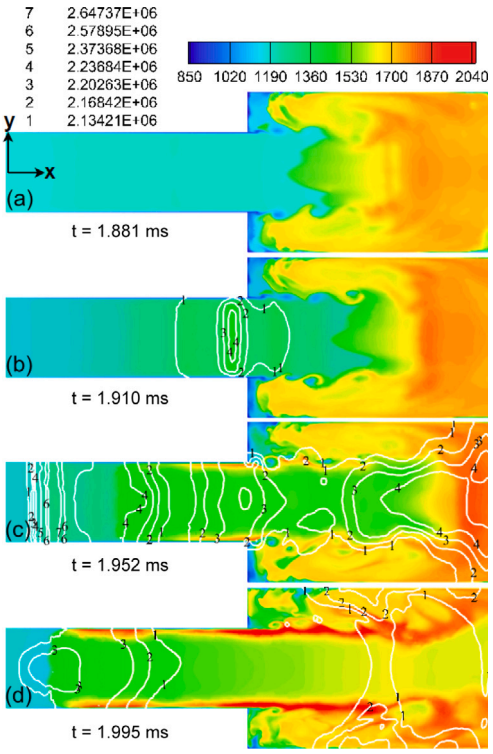

Fig. 2. Contours of temperature during a typical flashback event (top to bottom). White lines: pressure isolines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

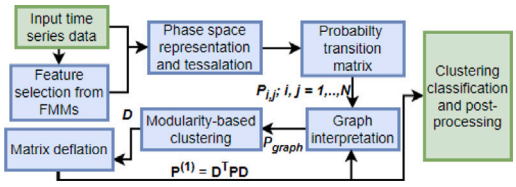

Fig. 3. A schematic illustrating the precursor identification workflow.

3.1. Featurization method: co-kurtosis PCA

The co-kurtosis PCA technique is chosen here for the featurization step as its use for extreme event/outlier detection was well documented in previous studies [11–13]. In [11], the robustness of this approach was also assessed on several test cases with extreme events and it was shown that the co-kurtosis PCA was consistently more sensitive than standard PCA to the presence of outliers (that signify the inception of extreme events). Details on the co-kurtosis PCA procedure are now provided.

Consider a feature vector $\mathbf{V}$ of dimension N_f with n_s samples. The joint fourth order moment tensor, i.e., co-kurtosis, can be represented using index notation as:

$$\tau_{ijkl} = \frac{1}{n_s}\sum_{n_s} v_i v_j v_k v_l, \quad 1 \le i,j,k,l \le N_f \tag{1}$$

where $v_i \in \mathbf{V}$. We are interested in computing the principal vectors of the co-kurtosis tensor that, in analogy to PCA, align themselves in the directions in which the outliers are located. By subtracting the excess variance, we obtain the final expression for computing the co-kurtosis tensor as:

$$\left[C_4^y\right]_{i_1 i_2 i_3 i_4} = \mathbb{E}[y \otimes y \otimes y \otimes y] - \mathbb{E}[y_{i_1} y_{i_2}]\,\mathbb{E}[y_{i_3} y_{i_4}] - \mathbb{E}[y_{i_1} y_{i_3}]\,\mathbb{E}[y_{i_2} y_{i_4}] - \mathbb{E}[y_{i_1} y_{i_4}]\,\mathbb{E}[y_{i_2} y_{i_3}]$$

where, $1 \le i_1 \ldots i_4 \le q$ and $\mathbb{E}$ is the expectation operator. It should be noted that the data used in the definition of joint moments is zero-centred around the mean. It was shown that the cumulant tensor C_4^y can be reshaped into a matrix M^y which results in a simple singular value decomposition (SVD) problem [23,24]:

$$\mathrm{mat}(C_4^y) \equiv M^y = \sum_{i=1}^{q} \kappa_i\, a_i \otimes \mathrm{vec}(a_i \otimes a_i \otimes a_i) \tag{2}$$

where the vectors a_i can be determined from the SVD of M^y. Note that, mat and vec denote operations of matricising and vectorizing a tensor respectively.

We perform a featurization step by defining a feature moment metric $F_i^{j,n}$ for each feature i in a given sub-domain j and time-step n (see Eq. (3)), which can be used to quantify the changes in the principal values and vectors:

$$F_i^{j,n} = \frac{\sum_{k=1}^{N_f} \lambda_k (\hat{e}_i . \hat{v}_k)^2}{\sum_{k=1}^{N_f} \lambda_k} \tag{3}$$

Note that $\hat{e}_i \cdot \hat{v}_k$ corresponds to the i-th entry in the k-th vector $\hat{v}_k$. By definition, the set of vectors $\hat{v}_k$ is orthonormal. Second, the sum of all feature moment metric (FMM) values for each feature i in a given sub-domain j and time-step n yields unity, i.e., $\sum_{i=1}^{N_f} F_i^{j,n} = 1, \forall j,n$. The FMM value for a given feature i represents the fraction of the overall moment (kurtosis or variance) contained in feature i and thus can be considered as a representative of moment distribution. Qualitatively, this represents the contribution of a feature i towards the occurrence of an anomalous/extreme state. Therefore, this method can provide the most important features to identify precursors of extreme states.

3.2. Precursor identification: modularity clustering

To identify precursors of extreme events (i.e. states which will lead towards an extreme state within a given time), the modularity-based clustering approach first introduced in [25] and then further developed in [16] is used. The initial step consists in extracting time series of representative features from the dataset (in this case, those observed to be of importance from the co-kurtosis featurization) and then interpreting them in a phase space. This phase space trajectory is then tessellated into hypercubes covering the entire phase space, thus representing the system's dynamics as the transitions between the different hypercubes, essentially reducing the complexity of the system. At this stage, the specific hypercubes covering extreme states are also identified based on a user-defined threshold. A probability transition matrix, $\boldsymbol{P}$, is then computed, where the probability of transitioning from hypercube B_i to hypercube B_j (between two consecutive snapshots) is calculated as follows:

$$P_{ij} = \frac{m\left(B_i \cap \mathcal{F}^1\left(B_j\right)\right)}{m\left(B_i\right)} \quad i,j = 1,\ldots,N \tag{4}$$

Here, N is the total number of hypercubes needed to represent the trajectory, $m(B_i)$ represents the number of phase space points (i.e. snapshots) laying in hypercube B_i and $\mathcal{F}^1$ is the temporal forward operator. P_{ij} can then be interpreted as a weighted and directed graph, where the nodes of the graph represent the hypercubes, the graph edges are the possible transitions and the edge weights are values given by the actual probabilities of transitioning from one hypercube to another. To make this graph human-tractable, modularity-based clustering can be applied as proposed in [26]. This algorithm focuses on optimizing a metric called *modularity*, which gauges how effectively a network is divided into distinct clusters (of nodes). This metric is premised on the notion that a favourable cluster division is one characterized by a reduction in inter-cluster edges beyond what would be statistically expected, as opposed to simply having fewer edges overall. The greater the deviation from a random network created with the same degree sequence as the

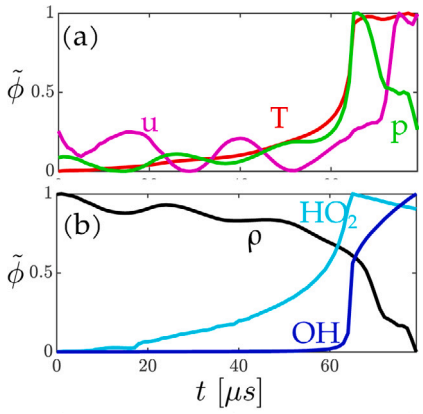

Fig. 4. Evolution of normalized scalars of interest during a typical flashback event.

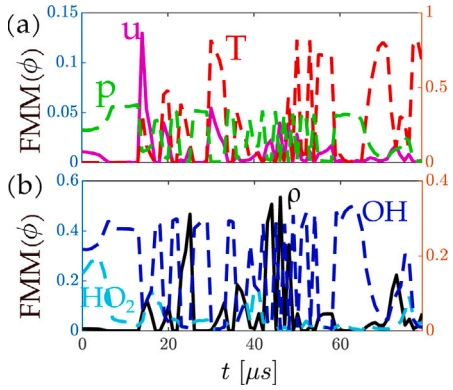

Fig. 5. Evolution of the six most important FMMs during flashback of Fig. 4. Full line: left axis, dashed: right axis.

original graph, the higher the resulting modularity score. The division of the network into clusters is done iteratively until no further division increases the modularity metric.

Once the clusters have been found, a matrix deflation step is made, where a cluster affiliation matrix $\boldsymbol{D}$ is created by assigning the nodes of the original graph to the identified clusters. Then, a new transition probability matrix $\boldsymbol{P}_{(1)}$ is calculated by deflating the original matrix $\boldsymbol{P}$:

$$\boldsymbol{P}_{(1)} = \boldsymbol{D}^T \boldsymbol{P} \boldsymbol{D} \tag{5}$$

This matrix is again interpreted as a graph and the process repeats until an interpretable (but still detailed enough) graph is obtained. In the graph based on the deflated transition probability matrix, we can then distinguish between the extreme and precursor clusters and compute statistics related to the transition time between these two types of clusters. This time represents the time available to react to an impeding extreme event when entering precursor clusters.

4. Results

4.1. State-space featurization

In this section, the co-kurtosis analysis is performed to identify the important features for the onset of flashback. First, the full thermochemical state (all nine species mass fractions, temperature, pressure and density) and components of velocity are collected at the centreline close to the area jump (square E0 in Fig. 1) for a total of $N_f = 15$ features. That sampling location is picked as it is located close to the observed early autoignition kernel discussed in Section 2 and shown in Fig. 2. Typical time-series of quantities which will be further investigated (temperature, streamwise velocity u, pressure p, density ρ and some species mass fractions) are shown in Fig. 4 during a flashback event. The quantities are normalized by their maximum and minimum to have an evolution between 0 and 1 (only for plotting purposes): $\tilde{\phi} = (\phi - \min(\phi))/(\max(\phi) - \min(\phi))$ (ϕ being one of the considered features). One can observe the flashback when the temperature increases abruptly around $t = 60$ μs which indicates that the flame front has progressed upstream of the sampling location.

Before applying the co-kurtosis PCA method, each feature is normalized by subtracting their respective means and dividing by their absolute maximum to ensure an equitable contribution of each variable to the FMMs. Then, the co-kurtosis analysis described in Section 3.1 is applied using these normalized features on the full time series (which contains 8 flashback events). The FMMs are then obtained using Eq. (3), which allows to identify the features that contribute the most to changes towards extreme states (here, the flashbacking state).

In Fig. 5, the FMM evolution of the six most important features during flashback is plotted for the same time instants as in Fig. 4. As the FMMs (of all features) sum up to unity, it is possible to rank what feature contributes more to state changes based on its FMM values. To more clearly identify the most important ones, we further computed a moving time average of the FMM of each feature (not shown here for brevity) which resulted in the selection of the temperature T, the streamwise velocity u, the density ρ, the pressure p and the species mass fractions of HO_2 and OH as the most important features for flashback. This analysis allows to discard the other features of the thermochemical and flow states as they are less relevant for this flashback according to their lower FMM contributions (and are not shown here for brevity).

From a physical perspective, the features identified confirm the physical intuition given by the observed flashback mechanism (detailed in Section 2): there is an early autoignition in the sampling region that can be associated with changes in HO_2 (a precursor of autoignition), which is picked up by the FMM. The effect of the pressure waves at the origin of that early kernel also appears in the FMM of pressure (by the large contribution of the pressure FMM) as well as its resulting effect on the streamwise velocity. The other quantities identified by the FMM (ρ, T and OH) are quantities which are directly resulting from those initial small changes in the flow state.

4.2. Precursor identification

In this section, we only retain the features identified previously, namely T, p, u, ρ, HO_2, OH to apply the precursor identification technique. We retain only six features to limit the computational cost of the clustering algorithm. To more easily capture the full range of the species mass fractions, we consider their logarithms instead of their actual values (i.e., we use $\ln(Y_{HO_2})$ and $\ln(Y_{OH})$). This allows to use an equidistant phase space tessellation (in logarithmic space) which better covers the full range of variations of the species mass fractions (which evolve exponentially over very short times). This improves the convergence of the clustering algorithm. From the time-series of those features sampled at location E0, we now consider their evolution in a phase space. For illustration purposes, a projection in the T-p plane of the evolution is shown in Fig. 6 (while we actually still retain the six mentioned features). It can be seen that the evolution shows two different patches: one at high temperature corresponding to when flashback has occurred and the flame is upstream of E0, and one at low temperature corresponding to when the flame is located in the combustion chamber. The aim of the clustering algorithm is to identify which parts of the lower patch corresponds to precursors of extreme events.

We apply here the modularity-based clustering technique described in Section 3.2, using a uniform tessellation consisting of 20^{N_f} hypercubes (N_f being the number of features used), to identify different

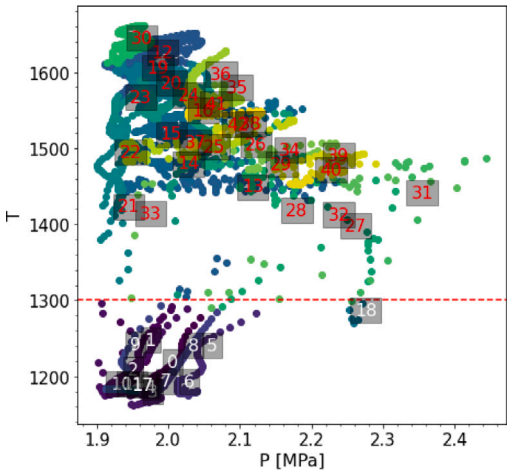

Fig. 6. Projected evolution in the T-p space of the features sampled at location E0. Dot colours indicate the cluster index in which that specific snapshot belongs. The cluster numbers are shown and located at the centroid of that cluster (red numbers indicate extreme clusters). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clusters of states that can be categorized into *normal*, *precursor* and *extreme* states. The horizontal line in Fig. 6 indicates the temperature above which we consider that the clusters are extreme (here $T = 1300$ K). The resulting clusters identified are shown in Fig. 6, where the colour of the points indicates the cluster to which a given state (along the evolution) belongs. In the bottom half several clusters are identified. The following clusters are identified as precursor: 0, 1, 2, 5, 8, 9 and 10. From direct inspection of the phase-space trajectory in Fig. 6, it would not be obvious that those specific states are precursors because of the density of states that overlaps and how normal clusters are located directly next to precursor clusters. This shows the advantage the clustering provides in analysing the evolution in this phase space.

From the obtained clusters, it becomes possible to estimate the mean time between the moment the combustor enters the precursor cluster and when it enters the extreme cluster. We call this time the *prediction time*, t_{pred}. In addition, we also assess the number of false positives (FP), which is the number of times the combustor enters what is considered a precursor cluster without being followed by an actual extreme state. It should be noted that we do not discuss false negative here as, in our approach, each cluster of states that can transition to flashback is marked as a precursor. Therefore, there is no extreme state without a preceding precursor and there are thus no false negative.

The prediction time when considering all six features is shown in the first line of Table 1 with a value of $t_{pred} = 15.7$ µs. While this prediction time may seem small, it actually corresponds to a state with a temperature of around 1218 K sampled at location E0, which is a very small change in the mixture given that the inflow temperature is 1180 K. Furthermore, this combustor is strongly unstable: between two flashback events, the combustor actually only spends approximately 60 µs in a "non flashbacking" state and therefore, this prediction time constitutes 25% of that "stable" time. Additionally, the precursor does not show any false positive meaning that it did not wrongly integrate normal states within precursor clusters. The prediction time is further illustrated in Fig. 7 for the best combination of features (to be discussed next). We show the temperature evolution (sampled at E0) with a background colour indicating to what type of cluster the combustor is considered to belong (blue: normal, orange: precursor, red: extreme). One can see that the algorithm correctly identifies a precursor state ahead of the large temperature increase, showing that it could potentially be used as a warning indicator of flashback.

We now analyse the influence of the number of features on t_{pred}. The different combinations tested are shown in Table 1. When only

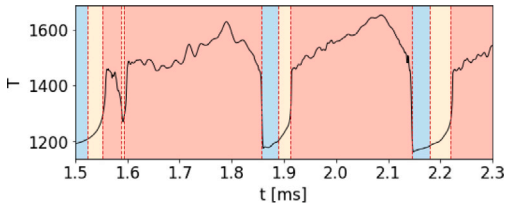

Fig. 7. Temperature time series sampled at the E0 location with background colour indicating the type of cluster the combustor is in (blue: normal; orange: precursor; red: extreme) when using [T, u, p, ρ, HO_2] for clustering. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Prediction time (t_{pred}) and false positives (FP) for different combinations of the important features. Sampling location E0 ($x = 3.1$ and $y = 0.0$ cm in the mid-z plane). FP reported as number of events.

Features used	t_{pred} [µs]	FP
T, u, p, ρ, HO_2, OH	15.7	0
T, u, p, ρ, HO_2	32.1	0
T, u, p, HO_2, OH	16.0	0
T, u, p, ρ, OH	20.9	0
T, u, ρ HO_2, OH	17.1	0
T, p, ρ, HO_2, OH	14.1	1
T, u, p, HO_2	15.8	0
T, u, p, ρ	18.1	0
T, u, p, OH	24.5	0
T, u, ρ, HO_2	12.6	0
T, u, HO2, OH	12.5	0
T, u, ρ, OH	15.0	0
T, p, ρ, HO_2	18.6	1
T, p, HO_2, OH	11.1	1
T, p, ρ, OH	13.1	1
T, ρ, HO_2, OH	8.38	0

four features are used (last ten lines), the prediction time is seen to generally decrease. This could be expected as, in this case, the clustering algorithm has fewer features available to segregate between normal and precursor states. However, when using five features, a different observation can be made for the case [T, u, p, ρ, HO_2], which shows a larger t_{pred} than when using all six features. This could be explained by the limited amount of data available for the clustering algorithm which only contains 8 flashback events. Indeed, when using all the features, the probability transition matrix used for the graph representation (explained in Section 3.2) may not have fully converged (as there are fewer points per tessellation hypercube). This would decrease the performance of the clustering technique in identifying correctly all the clusters. Conversely, decreasing the number of features may yield better converged transition statistics which then allows improving the prediction time obtained from the clustering. Therefore, a balance has to be struck between number of features used and the length of the available dataset.

In addition to the above analysis, we also investigate the importance of the sampling location with locations offset from the centreline (Ei locations in Fig. 1) and a location in the mixing tube (CU location in Fig. 1) for the best case of Table 1. The resulting prediction times are shown in Table 2, where it can be observed that the prediction time degrades, but not significantly when considering offset locations. This makes sense given that the location E0 was chosen as it was observed to be the region of most early autoignition events that lead to flashback. Regarding the CU location, a large decrease in t_{pred} is observed due to the large velocity of the flashbacking flame when it is travelling upstream (at nearly the speed of sound), which makes any long prediction time impossible.

Table 2
t_{pred} and FP for the $[T, p, u, \rho, HO_2]$ case sampled at different locations. FP reported as number of events.

Location [cm]	t_{pred} [μs]	FP
E0: x = 3.1; y = 0.0	32.1	0
E1: x = 3.1; y = 0.1	20.6	2
E2: x = 3.1; y = 0.2	32.0	1
E3: x = 3.1; y = 0.3	20.4	0
CU: x = 2.0; y = 0.0	13.9	0

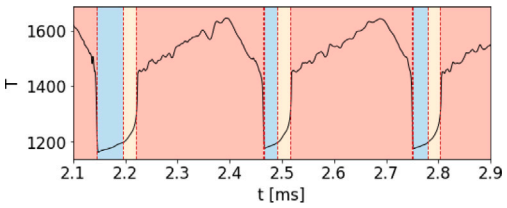

Fig. 8. Temperature evolution during 3 flashback events not used during the clustering procedure. Background colour indicates the combustor state (blue background: normal; orange: precursor; red: extreme). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Robustness analysis of precursor identification

Here, we analyse the robustness of the identified precursor to future flashback events not used for the clustering algorithm. Specifically, we consider a time series that only contains the first three flashback events to obtain the precursor clusters (using the features of the best case in Table 1: T, u, p, ρ, HO_2 sampled at E0). We use at least three flashback events as we observed that using only two did not give reliable precursors. Using those three flashback events results in a prediction time (on those three events) of 26 μs, which is a decrease of 19% as compared to using the full time series. Then, the centroid of all identified precursor clusters are recorded. Fig. 6 indicates, for example, the centroids of each cluster (the number of each cluster is plotted at the centroid of that specific cluster). Subsequently, for the rest of the time series (i.e. during the rest of the time series that contains the five next flashbacks), we classify the state of the combustor by computing the distance between the normalized state at that time instant, $\underline{\tilde{\phi}}(t)$, and the centroid of the clusters previously identified: $\epsilon_i = \|\underline{\tilde{\phi}}(t) - \underline{\tilde{\phi}}_{cent,i}\|$, where $\underline{\tilde{\phi}}_{cent,i}$ indicates the location in phase space of the centroid of cluster i. The combustor state is attributed to the cluster with the smallest distance ϵ_i and the combustor state is then classified as normal, precursor or extreme depending on the type of cluster. This classification of the combustor state is shown in Fig. 8 for flashbacks not used during clustering. The identified precursor still shows a similar prediction time on the flashback events not seen during training, with a prediction time of 23.6 μs (computed on the five flashback events not used to obtain the precursor states). This shows that the algorithm is able to somewhat identify useful precursors even when the dataset is relatively short. This indicates that despite potentially not having a fully converged probability transition matrix (due to the shortness of the dataset), the clustering algorithm is able to already roughly identify the region of the phase space where most precursor states lie.

4.4. Precursors from pressure time series

Here, we assess the performance of the clustering-based precursor identification technique when using measurable quantities. Specifically, we use pressure measurements at five different locations of the combustor (yellow squares in Fig. 1), in addition to the temperature time series at location E0, which is used to identify extreme clusters. We apply the same methodology as previously and obtain a prediction time of 23.75 μs, which implies a decrease of 25% as compared to the previous best case in Table 1. This reduction in prediction time could be expected as using only pressure measurements constitutes a limitation in the information that the clustering can use to identify precursors. Indeed, as observed in the previous section and in the FMM analysis, other features (such as HO_2) are also important in the flashback mechanism in this reheat combustor. Nonetheless, the order of magnitude of the prediction time remains similar as in the previous best precursor, indicating that the clustering algorithm is still able to identify some specific combinations of pressure signals that can serve as a precursor to flashback in this reheat combustor.

5. Conclusions

We have proposed and combined two different data-driven techniques for the analysis and prediction of extreme events in turbulent reacting flows, specifically the prediction of flashback in a lean premixed hydrogen reheat combustor. First, we used a co-kurtosis based featurization method to identify the most important thermochemical and flow features that change prior to flashback. It is observed that this method is able to identify the important features of flashback, inline with our physical understanding of the mechanism of flashback in this reheat combustor. Second, we used those features to identify the precursors of flashback using a clustering-based method. This combination of featurization with clustering differentiates our approach from similar combustion mode/reaction rate analysis techniques such as Chemical Explosive Mode Analysis (CEMA) or Computational Singular Perturbation (CSP). Indeed, while in the featurization steps the full thermochemical states was used, like in CEMA or CSP, we then combine the results of co-kurtosis PCA with a clustering algorithm that accounts for the preceding dynamics to identify precursor states before even the onset of flashback. We showed that the identified precursor states provide some prediction time before the flashback, up to 50% of the time over which the combustor is stable between two flashback events, depending on the specific choice of features and sampling locations. The robustness of the identified precursor was also analysed by obtaining it based on few flashback events and we observed that it was reliable for future unseen flashback events. A first step towards practical measurement configurations was finally made by considering wall pressure measurements, where only a moderate decrease in prediction time was observed. This last test shows the novelty of our approach as such an analysis, based purely on pressure/temperature times series, would not be possible with CEMA or CSP.

In future works, the robustness of the proposed method to stronger forms of fluctuations (like temperature/equivalence ratio fluctuations) or when applied to a less ideal case (with partially premixed mixture) will be assessed. Furthermore, the applicability of the proposed methods will be explored on other extreme events in turbulent reacting flows, such as blow-off. The performance of the proposed algorithms under more practical conditions, such as fewer measurements or the effect of noise, will also be investigated.

Novelty and significance statement

In this work, we tackle the problem of extreme event prediction in turbulent reacting flows. We demonstrate the capabilities of two novel techniques, namely co-kurtosis principal component analysis for the featurization problem and modularity-based clustering for the precursor identification. Our approaches are applied to the case of intermittent flashback in a reheat lean premixed turbulent hydrogen combustor which is relevant for the decarbonisation of our society. Our results show for the first time a robust data-driven workflow that allows to identify precursor of flashback, potentially enabling its prevention. This is significant both from a fundamental perspective given the advances this represent in the analysis of complex chaotic dynamics, as well as from a practical perspective given the potential risk that flashback poses for hydrogen combustion.

CRediT authorship contribution statement

Mihnea Floris: Generated the data, Analyzed data, Performed research, Wrote paper. **Tadikonda Shiva Sai:** Analyzed data, Performed research, Wrote paper. **Dibyajyoti Nayak:** Analyzed data, Performed research, Wrote paper. **Ivan Langella:** Designed and supervised research, Wrote and reviewed paper. **Konduri Aditya:** Designed and supervised research, Wrote and reviewed paper. **Nguyen Anh Khoa Doan:** Designed and supervised research, Wrote and reviewed paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

KA is a recipient of the Arcot Ramachandran Young Investigator award, IISc. The authors acknowledge support from the IISc-TU Delft joint seed fund 2022. Part of the simulations presented in this work were performed using the Dutch national supercomputer Snellius under NWO project n. 2021.043 and with the support of the SURF Cooperative using grant no. EINF-5775.

References

[1] M. Hassanaly, V. Raman, Classification and computation of extreme events in turbulent combustion, Prog. Energy Combust. Sci. 87 (2021) 100955.
[2] J.L. Wagner, K.B. Yuceil, A. Valdivia, N.T. Clemens, D.S. Dolling, Experimental investigation of unstart in an inlet/isolator model in mach 5 flow, AIAA J. 47 (2009) 1528–1542.
[3] B. Wu, X. Zhao, B.R. Chowdhury, B.M. Cetegen, C. Xu, T. Lu, A numerical investigation of the flame structure and blowoff characteristics of a bluff-body stabilized turbulent premixed flame, Combust. Flame 202 (2019) 376–393.
[4] T. Lieuwen, V. McDonell, E.L. Petersen, D. Santavicca, Fuel flexibility influences on premixed combustor blowout, flashback, autoignition, and stability, J. Eng. Gas. Turbine Power 130 (2008) 011506.
[5] J. Jiang, Z.G. Huang, C. Grebogi, Y.C. Lai, Predicting extreme events from data using deep machine learning: When and where, Phys. Rev. Res. 4 (2022) 023028.
[6] V. Nair, R.I. Sujith, Multifractality in combustion noise: Predicting an impending combustion instability, J. Fluid Mec. 747 (2014) 635–655.
[7] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence modeling in the age of data, Annu. Rev. Fluid Mech. 51 (2019) 357–377.
[8] M. Ihme, W.T. Tong, A.A. Mishra, Combustion machine learning : Principles , progress and prospects, Prog. Energy Combust. Sci. 91 (2022) 101010.
[9] G. D'Alessio, A. Attili, A. Cuoci, H. Pitsch, A. Parente, Analysis of turbulent reacting jets via principal component analysis, in: Data Analysis for Direct Numerical Simulations of Turbulent Combustion: From Equation-Based Analysis to Machine Learning, Springer International Publishing, 2020, pp. 233–251.
[10] G. D'Alessio, S. Sundaresan, M.E. Mueller, Automated and efficient local adaptive regression for principal component-based reduced-order modeling of turbulent reacting flows, Proc. Combust. Inst. 39 (2023) 5249–5258.
[11] K. Aditya, H. Kolla, W.P. Kegelmeyer, T.M. Shead, J. Ling, W.L. Davis, Anomaly detection in scientific data using joint statistical moments, J. Comput. Phys. 387 (2019) 522–538.
[12] A. Jonnalagadda, S. Kulkarni, A. Rodhiya, H. Kolla, K. Aditya, A co-kurtosis based dimensionality reduction method for combustion datasets, Combust. Flame 250 (2023) 112635.
[13] D. Nayak, A. Jonnalagadda, U. Balakrishnan, H. Kolla, K. Aditya, A co-kurtosis PCA based dimensionality reduction with nonlinear reconstruction using neural networks, Combust. Flame 259 (2024) 113192.
[14] M. Farazmand, T.P. Sapsis, A variational approach to probing extreme events in turbulent dynamical systems, Sci. Adv. 3 (2017) 1–8.
[15] P.J. Schmid, A. Garciá-Gutierrez, J. Jiménez, Description and detection of burst events in turbulent flows, J. Phys. Conf. Ser. 1001 (2018).
[16] U. Golyska, N.A.K. Doan, Clustering-based identification of precursors of extreme events in chaotic systems, in: Lecture Notes in Computer Science - ICCS2023, vol. 10476, Springer, Cham, 2023, pp. 313–327.
[17] K. Aditya, A. Gruber, C. Xu, T. Lu, A. Krisman, M.R. Bothien, J.H. Chen, Direct numerical simulation of flame stabilization assisted by autoignition in a reheat gas turbine combustor, Proc. Combust. Inst. 37 (2) (2019) 2635–2642.
[18] F. Gant, B. Bunkute, M.R. Bothien, Reheat flames response to entropy waves, Proc. Combust. Inst. 38 (2021) 6271–6278.
[19] A. Yoshizawa, K. Horiuti, A statistically-derived subgrid-scale kinetic energy model for the large-eddy simulation of turbulent flows, J. Phys. Soc. Japan 54 (8) (1985) 2834–2839.
[20] B. Kruljevic, A. Cabello López, I. Langella, A. Ciani, M. Duesing, LES/Thickened flame model of reheat hydrogen combustion with water/steam injection, in: Proceedings of the ASME Turbo Expo 2023: Turbomachinery Technical Conference and Exposition, vol. Volume 3B: Combustion, Fuels, and Emissions, 2023, V03BT04A042.
[21] J.P. Legier, T. Poinsot, D. Veynante, Dynamically thickened flame LES model for premixed and non-premixed turbulent combustion, in: Proceedings of the Summer Program, Center for Turbulence Research, 2000, pp. 157–168.
[22] J. Li, Z. Zhao, A. Kazakov, F.L. Dryer, An updated comprehensive kinetic model of hydrogen combustion, Int. J. Chem. Kinet. 36 (10) (2004) 566–575.
[23] L. De Lathauwer, B. De Moor, J. Vandewalle, Independent component analysis and (simultaneous) third-order tensor diagonalization, IEEE Trans. Signal Process. 49 (10) (2001) 2262–2271.
[24] A. Anandkumar, R. Ge, D. Hsu, S.M. Kakade, M. Telgarsky, Tensor decompositions for learning latent variable models, J. Mach. Learn. Res. 15 (2014) 2773–2832.
[25] P.J. Schmid, A. García-Gutierrez, J. Jiménez, Description and detection of burst events in turbulent flows, J. Phys. Conf. Ser. 1001 (2018) 012015.
[26] M.E.J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. 103 (23) (2006) 8577–8582.