Data-driven fault detection and estimation in thermal pulse combustors

S Chakraborty[1], **S Gupta**[1], **A Ray**[1]*, and **A Mukhopadhyay**[2]
[1]Department of Mechanical Engineering, Pennsylvania State University, PA, USA
[2]Department of Mechanical Engineering, Jadavpur University, Kolkata, India

Abstract: This paper presents the development of a dynamic data-driven statistical method for: (a) early detection of incipient faults and (b) parameter estimation for prognosis of forthcoming failures and operational disruptions (e.g. flame extinction) in thermal pulse combustors. From these perspectives, reduction in the tailpipe friction coefficient is estimated from time-series data of pressure oscillations. The algorithms for parameter estimation are built upon the principles of *Symbolic Dynamics*, *Information Theory* and *Statistical Pattern Recognition*. The proposed algorithms have been tested on an experimentally validated simulation model of a generic thermal pulse combustor.

Keywords: anomaly detection, parameter estimation, inverse problem, symbolic dynamic filtering, thermal pulse combustor

1 INTRODUCTION

Thermal pulse combustors have several advantages, such as significantly higher thermal efficiency, higher heat transfer rate, and lower pollutant emission, over steady-flow combustors. On the other hand, pulse combustors are prone to self-sustained pressure oscillations due to strong coupling between thermo-fluid dynamics in the combustor and the tailpipe [1]. Consequently, small parametric and non-parametric changes in the combustion process may cause unpredictable anomalies (i.e. deviations from the nominal behaviour) and thereby lead to significant degradation in the combustor performance over time [2]. The resulting evolution of anomalies is often very difficult to detect from measurements of the process variables unless the embedded statistical information is extracted via analytical tools of signal processing and pattern identification [3, 4].

The desired goal of thermal pulse combustors is to attain constant-amplitude self-sustained oscillations, which is considered as the nominal behaviour [2].

However, a change in the value of the friction coefficient, possibly due to wear, fatigue, and corrosion, results in progressive deterioration of flow conditions that adversely affect the stable operation of the combustor [5]. Results from an experimentally validated model [1] show that even a slight variation in this parameter may result in instabilities culminating in flame extinction or a steady flame. The objective of this paper is to estimate the tailpipe friction coefficient of thermal pulse combustors via statistical analysis of time-series data of pressure oscillations.

Pressure oscillation monitoring and estimation of the tail-pipe friction coefficient for prediction of flame extinction in thermal pulse combustors are addressed in this paper as solutions of the following two inter-related problems:

(a) the forward (analysis) problem of anomaly detection and generation of the respective statistical information;
(b) the inverse (synthesis) problem of identification of anomalous parameter(s) (e.g. statistical quantification of the friction coefficient) through assimilation and analysis of the above-mentioned statistical information.

As a partial solution to the forward problem, previous publications [6, 7] have reported detection

Corresponding author: Department of Mechanical Engineering,
Pennsylvania State University, 329 Reber Building, University Park,
PA 16802, USA. email: axr2@psu.edu

of performance degradation in aircraft gas turbine engines; however, estimation of the fault parameters was not addressed. This paper addresses both the forward and inverse problems for early detection of combustion instability and estimation of the tailpipe friction coefficient. The underlying concepts are tested on an experimentally validated model of a generic thermal pulse combustor model.

The paper is organized as follows. Section 2 briefly describes the thermal pulse combustor kinetics and presents a lumped parameter model. Section 3 reviews underlying concepts and essential features of symbolic dynamic filtering (SDF) for anomaly detection [8]. Section 4 presents formulation of the forward and inverse problems. Section 5 presents the results of SDF-based anomaly detection and statistical analysis for parameter estimation. The paper is summarized and concluded in section 6 along with recommendations for future research.

2 MODEL OF A GENERIC THERMAL PULSE COMBUSTOR

While the details of the non-linear dynamic model of a generic thermal pulse combustor are available in earlier publications [1, 9], this section presents a summary of the governing equations for completeness and self-sufficiency of the paper.

The main feature of a thermal pulse combustor is self-sustained pressure oscillations that accrue from strong coupling between combustion dynamics and gas flow in the tailpipe. The mechanism of thermal and pressure pulses has been explained in literature [1]. The drag force is represented in the present model by the tailpipe friction coefficient [10]. A schematic view of the combustion zone and the tailpipe is shown in Fig. 1.

The non-linear dynamic model is described in terms of four first-order coupled differential equations, resulting in four dimensionless state variables: temperature ($\widetilde{T}$), pressure ($\widetilde{P}$), fuel mass fraction (Y_f) and exit velocity ($\widetilde{u}$)

$$\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}t} = \frac{\gamma \widetilde{T}}{\widetilde{P}}\left[\frac{1}{\tau_f} + \frac{1}{\tau_h} + \frac{1}{\tau_c}\right] - \frac{\widetilde{T}^2}{\widetilde{P}}\left[\frac{1}{\tau_f} + \frac{(\gamma - 1)Z_e}{\rho_0} + \frac{\gamma}{\widetilde{T}_w \tau_h}\right] \qquad (1)$$

$$\tau_f = \rho_0/Z_i, \quad \tau_h = L_{c1}\rho_0 C_p/h\widetilde{T}_w, \quad \tau_c = \frac{\rho_0 C_p T_0}{\dot{R}_f \Delta h_c} \qquad (2)$$

$$\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}t} = \gamma\left[\frac{1}{\tau_f} + \frac{1}{\tau_h} + \frac{1}{\tau_c}\right] - \gamma\widetilde{T}\left[\frac{Z_e}{\rho_0} + \frac{1}{\widetilde{T}_w \tau_h}\right] \qquad (3)$$

$$\frac{\mathrm{d}Y_f}{\mathrm{d}t} = \frac{\widetilde{T}}{\widetilde{P}}\frac{1}{\tau_f}[Y_{fi} - Y_f] - \frac{C_p T_0}{\Delta h_c}\frac{\widetilde{T}}{\widetilde{P}}\left(\frac{1}{\tau_c}\right) \qquad (4)$$

$$\frac{\mathrm{d}\widetilde{u}}{\mathrm{d}t} = \frac{RT_0 \tau_f}{L_{c2}L_{tp}}\frac{\widetilde{T}_e}{\widetilde{P}_e}(\widetilde{P}_e - 1) - \frac{f|\widetilde{u}|\widetilde{u}L_{c2}}{2D_{tp}\tau_f} \qquad (5)$$

The combustion time (τ_c) is calculated based on a single-step Arrhenius kinetics for stoichiometric mixtures at inlet

$$\dot{R}_f = -Bv_0 T^{1/2}\rho^2 Y_f^2 \exp(-T_a/T) \qquad (6)$$

Equations (2) and (6) yield the chemical reaction time

$$\tau_c = \left[\frac{B'\Delta h_c}{C_p T_0}\frac{\widetilde{P}^2}{\widetilde{T}^{3/2}}Y_f^2 \exp(-\widetilde{T}_a/\widetilde{T})\right]^{-1} \qquad (7)$$

where the constant terms are merged into a single parameter B'. An expression for Z_e, needed to close the system of equations, is obtained from the conservation of mass within the tailpipe as

$$Z_e = \rho_0 \frac{\widetilde{u}}{\tau_f}\frac{\widetilde{P}_e}{\widetilde{T}_e} \qquad (8)$$

The pressure and temperature in the tailpipe are related to the combustor variables through isentropic
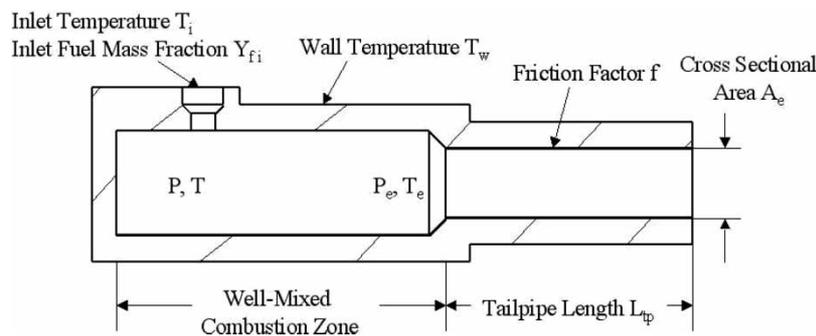
Fig. 1 Schematic of a generic thermal pulse combustor with specified inlet and tailpipe parameters

Table 1 Model parameters

Parameter	Value	Parameter	Value
A_s	$0.0167\,\text{m}^2$	h	$120\,\text{W/m}^2\,\text{K}$
V	$1.985 \times 10^{-4}\,\text{m}^3$	P_0	$1 \times 10^5\,\text{Pa}$
B'	3.85×10^8	T_a	$50\,\text{K}$
C_p	$1200\,\text{J/kg K}$	T_0	$300\,\text{K}$
D_{tp}	$0.0178\,\text{m}$	T_w	$1000\,\text{K}$
L_{c1}	$0.0119\,\text{m}$	γ	1.27
L_{c2}	$0.7434\,\text{m}$	ρ_0	$1.12\,\text{kg/m}^3$
L_{tp}	$0.61\,\text{m}$	t_f	$0.027\,\text{s}$
Y_{fi}	0.06	Δh_c	$4.6 \times 10^7\,\text{J/kg}$

relations as

$$\widetilde{T}_e = \widetilde{T} - \frac{\widetilde{u}^2 L_{c2}^2}{2 C_p T_0 \tau_f^2}, \quad \widetilde{P}_e = \widetilde{P}\left(\frac{\widetilde{T}_e}{\widetilde{T}}\right)^{\gamma/(\gamma-1)} \tag{9}$$

The above pulse combustor model follows the type and geometry as reported in [1]; the kinetic parameters of the model are listed in Table 1. To initiate the reaction, the initial temperature is raised to five times the ambient temperature. The choice of the initial time step affects only the initial transients that last for a period $t \leqslant 0.2\,\text{s}$. The anomaly detection analysis, presented in this paper, is based on time series data of combustor pressure oscillations in the time interval of $1.0 \leqslant t \leqslant 1.2\,\text{s}$. The data sampling rate is $\sim 0.1\,\text{MHz}$. Thus, the generated data set is sufficiently large for post-processing analysis.

3 REVIEW OF SYMBOLIC DYNAMIC FILTERING

This section reviews the underlying concepts and essential features of symbolic dynamic filtering (SDF) for anomaly detection. While the details and applications of SDF are reported in previous publications [8, 11], the key features of the SDF procedure are briefly summarized here for clarity and completeness of this paper.

Detection of anomaly patterns is formulated as a two-time-scale problem. The *fast-time* scale is related to response time of the process dynamics (e.g. combustor pressure oscillations). Over the span of a given time-series data sequence, dynamic behaviour of the system is assumed to remain invariant, i.e. the process is quasi-stationary at the fast-time scale. In other words, the variations in the behaviour of system dynamics is assumed to be negligible on the fast-time scale. The *slow-time* scale is related to the time span over which parametric or non-parametric changes may occur and exhibit non-stationary dynamics that can be associated with the evolving anomalies (e.g. degradation of the friction coefficient in the tailpipe wall leading to flame extinction).

In general, a long time span in the fast scale is a tiny (i.e. several order of magnitude smaller) interval in the slow scale. For example, change in friction coefficient in a pulse combustor, causing a detectable change in the dynamics of the system, occurs on the slow scale (possibly in the order of several hundreds of operation hours); the combustor behaviour is essentially invariant on the fast scale over which data acquisition is done (approximately in the order of milliseconds). Nevertheless, the notion of fast- and slow-time scales is dependent on the specific application, operating conditions, and environment. The concept of two time scales is illustrated in Fig. 2.

3.1 Symbolic dynamics, encoding, and state machine

This section briefly describes the concepts of *symbolic dynamics*, encoding non-linear system dynamics from observed time-series data, state machine construction, and computation of the state probability vectors that are representatives of the evolving characteristics of the dynamical system.

Let $\Omega \in \mathbb{R}^n$ be a compact (i.e. closed and bounded) region, within which the trajectory of the dynamical system is circumscribed as illustrated in Fig. 3. The region Ω is partitioned into a finite number of (mutually exclusive and exhaustive) cells, so as to obtain a coordinate grid. Let the cell, visited by the trajectory at a time instant, be denoted as a random variable taking a symbol value from the alphabet Σ. An orbit of the dynamical system is described by the time-series data as a sequence $\{x_0, x_1, \ldots, x_k, \ldots\}$, where each $x_i \in \Omega$. Thus, the trajectory of the orbit passes through or touches one of the cells of the partition. Each initial state $x_0 \in \Omega$ generates a sequence of symbols defined by a mapping from the phase space into the symbol space as

$$x_0 \rightarrow s_0 s_1 s_2 \cdots s_k \cdots \tag{10}$$

where each s_i, $i = 0, 1, \ldots$, takes a symbol from the alphabet Σ. The mapping in equation (10) is called *symbolic dynamics* as it attributes a legal (i.e. physically admissible) symbol sequence to the system dynamics starting from an initial state. Symbolic dynamics can be viewed as coarse graining of the
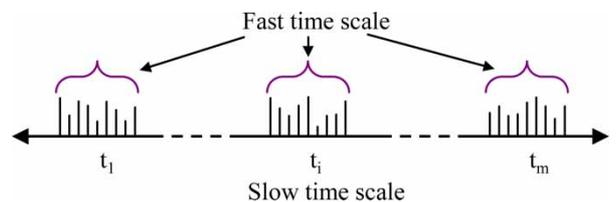
Fig. 2 Pictorial view of the two time scales: *slow-time* scale of anomaly evolution and *fast-time* scale of data acquisition
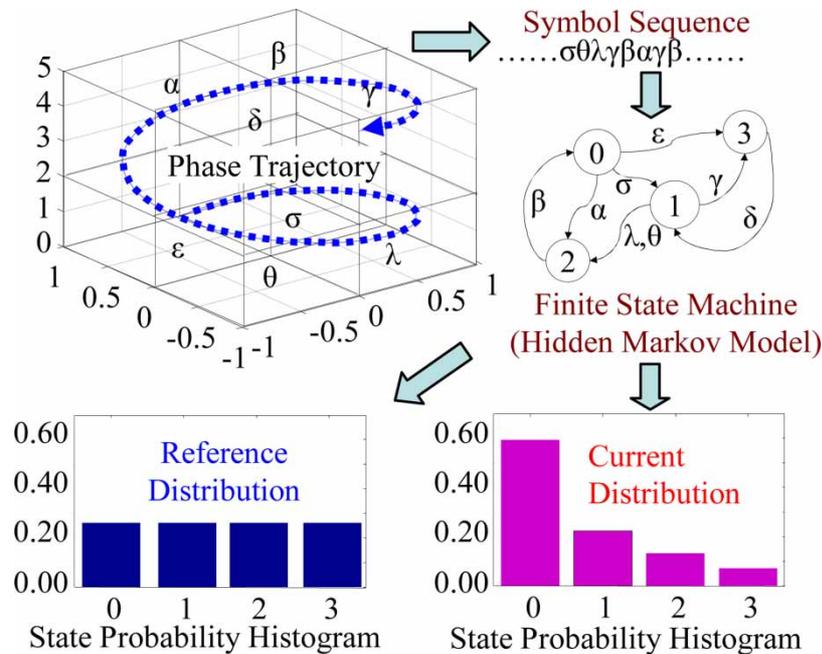
Fig. 3   Concept of symbolic dynamic filtering

phase space, which is subjected to (possible) loss of information resulting from granular imprecision of partitioning boxes. However, the essential robust features (e.g. periodicity and chaotic behaviour of an orbit) are expected to be preserved in the symbol sequences through an appropriate partitioning of the phase space [**12**, **13**]. Figure 3 shows the conversion of the symbol sequence into a finite-state machine and generation of state probability vectors [**8**] as briefly explained in the following section.

3.2 Summary of anomaly detection procedure

This section summarizes the anomaly detection algorithms based on the SDF method [**8**, **11**].

1. *Time-series data acquisition at the nominal condition, at time epoch t_0.* Sensor data (e.g. time series of pressure oscillations in a thermal pulse combustor) are collected when the system is assumed to be operating in the nominal condition.
2. *Generation of wavelet transform coefficients, obtained with an appropriate choice of the wavelet basis* [**14**]. A crucial step in SDF is partitioning of the phase space for symbol sequence generation [**13**]. Since the structure of phase space may become very complex for high-dimensional systems or may even be unknown for unmodelled systems, an alternative way is to extract information from the time-series data of appropriate sensors. Once the time-series data set is collected, it is transformed to the wavelet domain by appropriate choice of scale(s) and basis [**11**]. A wavelet-based partitioning method [**11**] has been adopted for construction

of symbol sequences from the time-series data. The wavelet transform [**14**] largely alleviates the difficulties of phase-space partitioning in case of high dimensions and is particularly effective for noisy data from high-dimensional dynamical systems [**11**]

3. *Maximum entropy partitioning of the wavelet space at the nominal condition* [**11**] *and generation of symbol sequence.* In this method of data space partitioning, the regions with more information are partitioned finer and those with sparse information are partitioned coarser. This is achieved by maximizing the Shannon entropy, which is defined as

$$S = -\sum_{i=1}^{|\Sigma|} p_i \log(p_i)$$

where p_i is the probability of the ith segment of the partition and summation is taken over all segments. The partitioning is fixed for subsequent time epochs. Each segment of the partitioning is assigned a particular symbol and a symbol sequence is generated.

4. *Generation of statistical information at the reference condition.* A finite-state machine is constructed from the symbol sequence for alphabet size $|\Sigma|$ and the window length D. (Note: the structure of the finite-state machine is also fixed for subsequent slow-time epochs.) The probability vector $\boldsymbol{p}^0$ is calculated at the (slow-scale) time epoch t_0 of the nominal condition, whose elements represent the state probabilities of the hidden Markov model constructed from the symbol sequence [**8**]. As a

consequence of maximum entropy, $\boldsymbol{p}^0$ has uniform distribution, i.e. each element has equal probability (see reference [**11**] for details).

5. *Time-series data acquisition at subsequent slow-time epochs of (possible) anomaly evolution.* Ensemble of data is collected from the same sensor(s) as in the nominal condition at time epoch t_0, at subsequent (slow-scale) time epochs, $t_1, t_2, \ldots t_k, \ldots$, and converted to the wavelet domain to generate respective symbol sequences based on the fixed partitioning at the time epoch t_0.

6. *Generation of statistical information at (possibly) anomalous conditions.* The probability vectors $\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^k, \ldots$ at (slow-scale) time epochs, $t_1, t_2, \ldots t_k, \ldots$ are calculated from the respective symbol sequences.

7. *Computation of scalar anomaly measures.* The anomalies are quantified as pattern changes (i.e. deviations from the nominal pattern) and are characterized by a scalar-valued function, called *Anomaly Measure* ψ. An appropriate distance function $d(\bullet, \bullet)$ is chosen to calculate the scalar anomaly measures $\psi^1, \psi^2, \ldots, \psi^k, \ldots$ at time epochs, $t_1, t_2, \ldots t_k, \ldots$ based on the distance of these probability vectors with respect to the nominal condition (see reference [**8**] for details) such that

$$\psi^k \equiv d(\boldsymbol{p}^k, \boldsymbol{p}^0)$$

Capability of SDF has been demonstrated for anomaly detection at early stages of gradually evolving anomalies by real-time experimental validation. Application examples include active electronic circuits [**15**] and fatigue damage monitoring in polycrystalline alloys [**16**]. Applications of SDF have also been reported for fault detection and isolation in gas turbine engines [**6**, **7**]. In this regard, major advantages of SDF are listed below

(a) robustness to measurement noise and spurious signals [**11**];
(b) adaptability to low-resolution sensing due to the coarse graining in space partitions [**8**];
(c) capability for early detection of anomalies because of sensitivity to signal distortion and real-time execution on commercially available inexpensive platforms [**15**].

4 FORMULATION OF FORWARD AND INVERSE PROBLEMS

The issues of combustion instability monitoring and estimation of the tailpipe friction coefficient are addressed as two inter-related problems: 1) the *forward (or analysis) problem* and 2) the *inverse (or synthesis) problem*. In general, the *forward problem* consists of prediction of anomalies from simulation

or experimental data, given *a priori* knowledge of the underlying model parameters. In the absence of a reliable model, this problem requires generation of behavioural patterns of the system evolution through off-line analysis of an ensemble of the observed time-series data. On the other hand, *the inverse problem* consists of off-line assimilation of the statistical information produced in the forward problem, followed by on-line estimation of the critical parameter(s) (such as friction coefficient) that characterize the possible anomalies based on the observed data. Inverse problems may become ill-posed and are challenging due to the absence of a unique solution or non-existence of an exact solution [**17**]. This section delineates the objectives, description, and the solution procedure of the forward and the inverse problems.

4.1 The forward problem

The objective of the forward (analysis) problem is to extract behavioural information of the dynamic system as it responds to parameter variations in the slow scale as a result of evolving anomaly or anomalies. Specifically, the forward problem aims at detecting the deviations in the statistical patterns in the time-series data, generated at different time epochs in the slow-time scale, from the nominal behaviour pattern. The solution procedure of the forward problem is summarized below.

F1. Collection of fast-scale data at different slow-scale time epochs to characterize the statistical behaviour of the system with anomaly evolution.

F2. Analysis of these data sets using the SDF method as discussed in section 3 to generate pattern vectors defined by the probability distributions at the corresponding slow-time epochs. The profile of anomaly measure is then obtained from the evolution of the statistical pattern vector with respect to that at the nominal condition.

F3. Generation of a family of such anomaly measure profiles by repeating the numerical experiments or laboratory experiments under conditions similar to the extent of inherent component and parameter uncertainties. This procedure provides a statistical description of anomaly evolution (i.e. reduction of friction coefficient) in the presence of parameter uncertainties (such as fluctuations in combustor wall temperature) in the dynamical system under observation. This step is required in systems where there is a source of parametric or non-parametric uncertainty. In the case of a pulse combustor, degradation in the tailpipe friction coefficient is considered as the evolution of anomaly and fluctuations in the combustor wall temperature, and the fuel/air ratio inflow is considered as possible sources of uncertainties.

4.2 The inverse problem

The objective of the inverse problem is estimation of the varying system parameter(s) that could be responsible for anomaly growth. This issue is directly related to the problem of failure prognosis. Although the physics-based model of generic thermal pulse combustors is experimentally validated, it is deterministic, i.e. the model only represents mean-value characteristics of the uncertain combustion dynamics. Therefore, pertinent statistical information must be generated to obtain reliable confidence bounds of the estimated parameters, which is necessary for robust decision-making and control. For parameter estimation in a given combustor that has its own signature, it is necessary to make decisions based on the measurement data collected from this specific combustor and the statistical information that has been already generated off-line.

Analysis of the observed time-series data and information derived from the forward problem are combined to generate parameter estimates. For example, in the current problem, degradation of flow characteristics through the combustor tail-pipe due to change in the friction coefficient has been selected as the parameter for anomaly detection. However, uncertainties such as fluctuations in the combustor wall temperature lead to an ensemble of anomaly evolution profiles under otherwise identical operating conditions. Therefore, as a precursor to the solution of the inverse problem, generation of an ensemble of time-series data sets is required during the forward problem for multiple numerical experiments or laboratory experiments conducted under identical operating conditions similar to the extent of inherent component and parameter uncertainties. With the current technique, estimates of the friction coefficient can be obtained at a slow-time epoch within a certain confidence interval. The statistical information derived from the ensemble of anomaly measure profiles, generated in the forward problem, are used to obtain the estimates of the friction coefficient as time evolves. The solution procedure of the inverse problem consists of the following steps.

I1. Generation of a pattern matrix Υ from the family of anomaly measure profiles derived in step F3 of the forward problem (see details in section 5.2.1).

I2. Hypothesis of a probability distribution for the anomaly (for example, deviation in the friction coefficient f) by using statistical goodness of fit [18], and estimation of the distribution parameters from the resulting pattern matrix.

I3. Collection of fast-scale data at the current slow-scale time epoch as in step F1 of the forward problem.

I4. Analysis of time-series data sets as described in previous sections to calculate the anomaly measure at the current time epoch. The procedure is similar to step F2 of the forward problem. As such, the information available at any particular instant is the value of the anomaly measure calculated at that particular instant;

I5. Detection and estimation of an anomaly (if any) based on the computed anomaly measure in step I4 and the information derived in steps I1 and I2 of the inverse problem.

5 CONCEPT VALIDATION, RESULTS, AND DISCUSSION

This section presents a solution procedure for both forward and inverse problems as well as generates and discusses the pertinent results of statistical analysis.

5.1 Solution methodology of the forward problem and results

As discussed earlier, the primary objective of the forward problem is to identify the behavioural pattern of damage evolution in a complex dynamical system involving the uncertainties (if any) which can be both parametric or non-parametric in nature. The solution procedure of a part of the forward problem, i.e. early detection of degradation of tailpipe friction coefficient using SDF of pressure oscillations, is briefly reviewed here for completeness.

5.1.1 Data analysis using SDF

The governing equations of the pulse combustor model (see reference [1]) are solved as a system of coupled non-linear ordinary differential equations. Time-series data of pressure oscillations are generated for the time period $1.0\,\text{s} \leqslant t \leqslant 1.2\,\text{s}$ which gives enough time for the initial transients to die out. The combustor pressure exhibits self-sustained oscillations at the tailpipe friction coefficient $f = 0.0300$ [1], which is considered to be the desired nominal behaviour. The friction coefficient is monotonically decreased at decrements of $\Delta f = 0.0005$ from the nominal value of $f = 0.0300$ to represent a gradual reduction in the drag force at consecutive slow-scale time epochs. Each member of this ensemble of time-series data (at different values of f) is analysed using the SDF-based anomaly detection procedure to generate the corresponding state probability vectors $\{\boldsymbol{p}^k\}$. Anomaly measures $\{\psi^k\}$ are then calculated as distances from the reference $\boldsymbol{p}^0$.

The alphabet size for partitioning and the window depth are chosen to be $|\Sigma| = 8$ and $D = 1$, respectively. The wavelet basis for the partitioning is chosen to be *MexicanHat*; this choice is made because of the following facts.

1. The characteristics of pressure wave signals are such that the Mexican Hat basis yields an acceptable solution with fewer coefficients in comparison with different wavelet bases such as the Daubechies family [**11**, **14**].
2. The code for the Mexican Hat basis is available in many commercial wavelet software libraries.

The length of each symbol sequence used in this paper is 20 000, which satisfies the stopping rule presented in reference [**6**] for tolerance $\eta = 4.0 \times 10^{-4}$ and the number of states of the D-Markov machine, which is equal to eight in this case.

5.1.2 Generation of statistical patterns

It is noted that, in a practical operation of the pulse combustor, it is very difficult to maintain consistent operating conditions, i.e. exact parameter values devoid of fluctuations. For example, the combustor wall temperature, fuel–air ratio, heat transfer coefficient from the wall, and fuel–mass input to the combustor are parameters that affect the dynamic characteristics of the pulse combustor system, but it is virtually impossible to control them accurately. Noting this, it has been argued that a statistical description of the anomaly measure is an appropriate indicator of the anomaly state of the combustor.

To implement this logic, a family of anomaly profiles were generated by repeating the simulation experiments multiple times under varying operating conditions. In the present simulation experiments, combustor wall temperature was selected as the source of system uncertainty. It has been shown by extensive simulation and laboratory experimentation [**1**] that corresponding to a wall temperature of 1000 K, pressure oscillations of considerable amplitude are observed in the combustion chamber. At lower wall temperatures, the flame simply dies out, whereas at higher temperatures, the oscillations stabilize to a steady flame. This shows that the pulse combustor dynamics are sensitive to the combustor wall temperature. Moreover, in the operation of a pulse combustor, fluctuations within ±0.2 per cent in the wall temperature are expected even in very well-regulated combustor systems.

To capture the effects of this uncertainty on the system behaviour, and to investigate the robustness of the SDF technique to such parameter fluctuations, $\ell = 50$ simulation experiments have been conducted on the simulation test-bed of a generic pulse combustor [**1**] with the wall temperature varying from 998 to 1002 K around the nominal value of 1000 K. Each simulation experiment was conducted for degrading friction coefficient values at different slow-time epochs as described earlier. The SDF analysis of time-series data sets for different values of
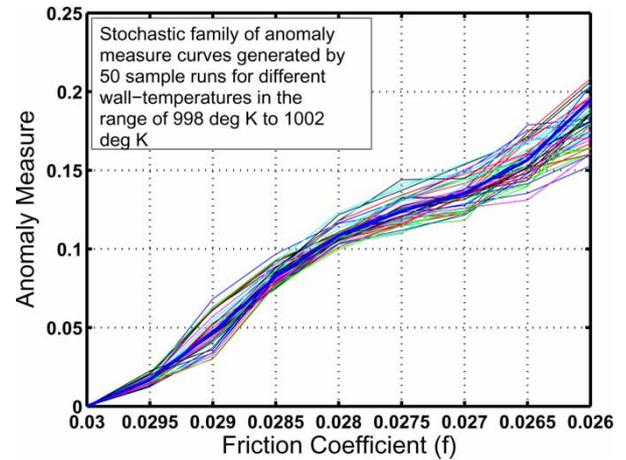
Fig. 4 Ensemble of anomaly measure curves at different wall temperatures

friction coefficient generated during each simulation experiment provided an anomaly measure profile. Figure 4 exhibits the family of profiles of the normalized anomaly measures generated from $\ell = 50$ simulation experiments plotted against the friction coefficient f. The region towards the right of $f = 0.0260$ (not shown in Fig. 4) is the flame extinction region with no pressure oscillations and the region towards the left (shown in Fig. 4) consists of self-oscillations modes. Although the transition and flame extinction regions offer insights into interesting phenomena such as chaos and bifurcation, these regions have not been investigated in the present work. The rationale is that these phenomena are of no significant consequence from the standpoint of early detection of anomalies, which is the focus of this paper. For example, from the perspectives of life-extending control and resilient control, the operating region should maintain a reasonably safe distance from the potential bifurcation points to avert potentially catastrophic events such as flame extinction.

Richards *et al.* [**1**] have studied the effects of friction coefficient, heat loss from the combustion zone, and flowrate on the performance of the combustor. It would be an interesting study to determine and isolate the effects of these multiple sources of uncertainties and anomalies on the computed measure, and that is part of the proposed future work.

The family of anomaly measure profiles, generated in the forward problem, is used in the analysis of the inverse problem to obtain an estimate of the friction coefficient f at a particular time epoch as presented below.

5.2 Solution methodology of the inverse problem and results

The objective of the inverse problem is the identification of anomalies and estimation of the fault

parameters based on the family of curves generated in the forward problem. It is essential to detect the evolving damage and to estimate the fault parameters during the operating period of the mechanical system, so that appropriate remedial action(s) can be taken before the onset of widespread damage leading to complete failure.

In the simulation experiments, sets of time-series data, collected in the fast scale of pressure oscillations, are generated at the desired slow-scale time epochs and are analysed using SDF to compute the anomaly measure. However, based on this derived value of the anomaly measure, the friction coefficient f cannot be exactly determined due to the variations observed in the family of anomaly profiles as reported in Fig. 4. Since a unique map from the measure values to the parameter space does not exist due to inherent uncertainties in the combustion dynamics, the friction coefficient f is treated as a random variable, where the computed values can be treated as the (noise-corrupted) observables.

5.2.1 Generation of the pattern matrix

It is seen from the family of profiles in Fig. 4 that the lower bound for the anomaly measure at the critical value of the friction coefficient parameter (i.e. $f = 0.026$) is $\psi = 0.15$, which signifies a transition from the self-excited oscillations to extinction. Hence, conservatively, the working range for the friction-coefficient estimation problem has been selected from the range of $\psi = 0$–0.15 for generation of the pattern matrix. This range of ψ is then partitioned into $m = 100$ uniformly spaced intervals. The family of anomaly measure profiles shown in Fig. 4 is then arranged in the form of a pattern matrix Υ of dimension $\ell \times m$, such that each column of Υ lists the values of the friction-factor f measured for ℓ samples at the corresponding anomaly measure interval. As such, the elements of any particular column of Υ thus describes the distribution of the random variable f for the specific interval of anomaly measure associated with that column.

5.2.2 Estimation of the friction coefficient with confidence bounds

In order to provide statistical information about the distribution of the friction coefficient f for different values of the anomaly measure ψ, the two-parameter lognormal distribution [19] is hypothesized for each column of Υ. The lognormal probability density function of the random variable f is defined as

$$p_f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(\frac{-(\ell n(x) - \mu)^2}{2\sigma^2}\right)\mathcal{U}(x)$$

where $\mathcal{U}(\bullet)$ is the standard Heaviside unit step function; and μ and σ are, respectively, the mean and standard deviation of the Gaussian distributed random variable $\ell n(f)$. The rationale for selecting lognormal distribution of f, as opposed to other distributions (e.g. normal or Weibull), is stated below.

1. *Lognormal* distribution is one directional on the positive axis.
2. The shape of lognormal distribution is suitable to model gradual failures in mechanical structures.
3. Many standard statistical tools become available for further analysis by modelling the random variable $\ell n(f)$ as Gaussian.

This makes the *lognormal* distribution a natural choice for failure analysis of the combustor tailpipe.

The goodness of fit of the estimated friction coefficient f has been examined by both χ^2 and Kolmogorov–Smirnov (KS) tests [19] and the number of bins were chosen to be $r = 7$ for statistical analysis of the data set in each column of Υ, so that each bin contains about seven data points. Thus, with the two-parameter distribution (i.e. $k = 2$), the degree of freedom for statistical goodness of fit becomes $r - k - 1 = 4$. The χ^2-test and the KS test showed that, for all of the $m = 100$ anomaly measure intervals, the hypothesis of the two-parameter lognormal distribution passed the 5 per cent significance level which is a conventional standard; the hypothesis test passed the 10 per cent significance level for 95 per cent of all anomaly measure intervals. Details of these statistical tests are provided in standard text books such as [19].

Once the lognormal distributions of the family of friction coefficients are obtained, the interval bounds at different confidence levels are computed from the properties of the lognormal distribution [18, 19]. Confidence level signifies the probability that the estimated parameter will lie within the corresponding confidence interval. As an example, for a confidence level of 95 per cent, the probability that the actual parameter will lie between the specified confidence interval is 95 per cent.

5.2.3 Validation and discussion of the results

Table 2 lists the following statistical information on evolution of the friction coefficient f for different values of the anomaly measure ψ.

1. Two parameters μ and σ of the lognormal distribution of the friction coefficient f, which are, respectively, the mean and standard deviation of $\ell n(f)$. Then, the mean and variance of f are $\mu_f = \exp(\mu + \sigma^2/2)$ and $\sigma_f^2 = (\exp(\sigma^2) - 1)\mu_f^2$, respectively.
2. Maximum-likelihood estimate $\widehat{f} = \exp(\mu - \sigma^2)$ of the friction coefficient f, i.e. where the probability density function p_f attains the maximum value.

Table 2 Probability distribution parameters of the friction coefficient at different confidence levels

| | Parameter estimates of lognormal distribution | | Maximum-likelihood estimate of friction coefficient $\widehat{f}$ | Confidence interval bounds of $\widehat{f}$ | | | | | |
| | | | | 75% confidence | | 85% confidence | | 95% confidence | |
Anomaly measure	Mean of $\ell n(f)$ (μ)	Standard deviation of $\ell n(f)$ (σ)		Lower	Upper	Lower	Upper	Lower	Upper
0.0125	−3.5193	1.9162×10^{-3}	0.029 62	0.029 56	0.029 69	0.029 54	0.029 70	0.029 51	0.029 73
0.0250	−3.5284	2.2490×10^{-3}	0.029 35	0.029 28	0.029 43	0.029 26	0.029 45	0.029 22	0.029 48
0.0375	−3.5357	3.5640×10^{-3}	0.029 14	0.029 02	0.029 26	0.028 99	0.029 29	0.028 94	0.029 34
0.0500	−3.5418	3.6366×10^{-3}	0.028 96	0.028 84	0.029 08	0.028 81	0.029 11	0.028 76	0.029 17
0.0625	−3.5477	3.7828×10^{-3}	0.028 79	0.028 67	0.028 92	0.028 64	0.028 95	0.028 58	0.029 01
0.0750	−3.5541	3.4618×10^{-3}	0.028 61	0.028 50	0.028 72	0.028 47	0.028 75	0.028 42	0.028 80
0.0875	−3.5611	3.6418×10^{-3}	0.028 41	0.028 29	0.028 53	0.028 26	0.028 59	0.028 21	0.028 61
0.1000	−3.5695	3.2168×10^{-3}	0.028 17	0.028 07	0.028 28	0.028 04	0.028 30	0.027 99	0.028 35
0.1125	−3.5816	6.3905×10^{-3}	0.027 83	0.027 63	0.028 04	0.027 58	0.028 09	0.027 49	0.028 18
0.1250	−3.5974	10.5580×10^{-3}	0.027 40	0.027 06	0.027 73	0.026 98	0.027 81	0.026 83	0.027 97
0.1375	−3.6127	10.4300×10^{-3}	0.026 98	0.026 66	0.027 30	0.026 58	0.027 39	0.026 43	0.027 54
0.1500	−3.6253	8.7480×10^{-3}	0.026 64	0.026 37	0.026 91	0.026 31	0.026 98	0.026 19	0.027 10

Maximum-likelihood estimate $\widehat{f} = \exp(\mu - \sigma^2)$, mean $\mu_f = \exp(\mu + \sigma^2/2)$, and variance $\sigma_f^2 = (\exp(\sigma^2) - 1)\mu_f^2$.

3. Confidence interval bounds for the friction coefficient f, at three different confidence levels of 95, 85, and 75 per cent.

Figure 5 exhibits the plots of confidence interval bounds at these three confidence levels along with the plot of the mean of the friction coefficient f. Figure 5 also shows the plots of the anomaly measure ψ for three additional test cases, marked as validation runs 1, 2, and 3. These three test cases are not included in $\ell = 50$ test cases that generated the results in Table 2. It is seen that the profiles of these test cases lie within the 85 per cent confidence interval for the entire range of the anomaly measure.

For the computed anomaly measure ψ at a (slow-scale) time epoch, Table 2 yields an estimate of the mean and standard deviation of the lognormal-distributed friction coefficient f. For example, if the computed value of ψ corresponds to one of the pre-selected values of anomaly that have been used for generating the pattern matrix Υ, then mean and standard deviation of f can be directly generated from Table 2. In general, these parameters can be computed as a function of ψ by interpolation of the columns of the pattern matrix Υ.

For the purpose of illustration, the interval bounds at 95 per cent confidence level are shown in Fig. 5 at anomaly measure $\psi = 0.05$; the corresponding lower
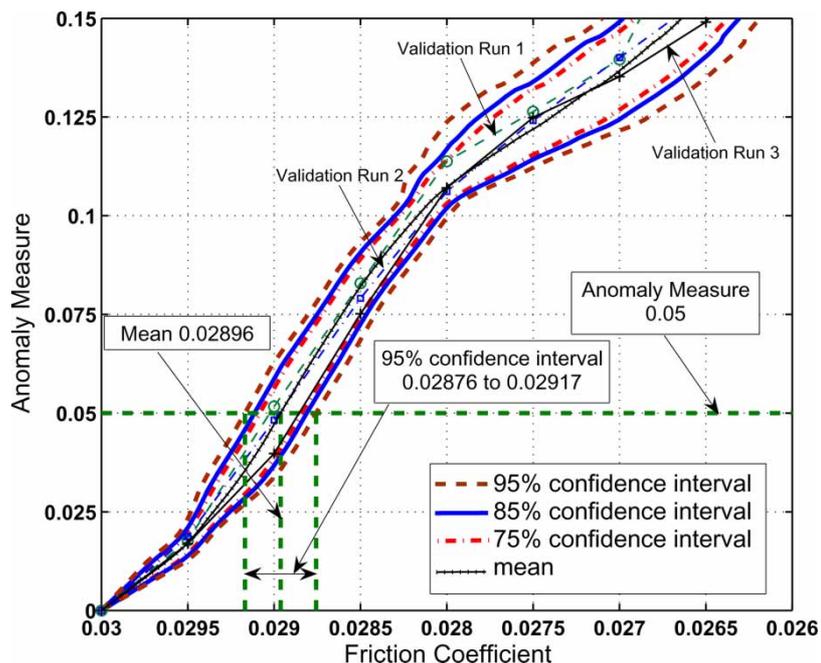
Fig. 5 Anomaly evolution and estimation of the friction coefficient

bound is 0.028 76 and the upper bound is 0.029 17. An estimate $\widehat{f}$ of the friction coefficient is chosen as the maximum-likelihood estimate, i.e. at the point of maximum probability, as seen in Table 2. It should be noted that, unlike the normal distribution, the mean for lognormal distribution does not correspond to the point of maximum probability. Specifically, for lognormal distribution, $\widehat{f}/\mu_f = \exp(-3/2\sigma^2) < 1$. Since $\widehat{f}$ may serve as a control parameter for mitigation of flame extinction and other failures, it is logical to have a more conservative estimate at the point of maximum probability rather than the mean. In general, an estimate $\widehat{f}$ of the friction coefficient should be chosen by the problem at hand.

As reported in Fig. 5, the variance of the estimate $\widehat{f}$ of friction coefficient grows with decreasing friction-coefficient, as indicated by the width of confidence intervals for any particular value of the anomaly measure. This observation is explained by the fact that, with decreasing friction coefficient, the combustor system slowly approaches a highly oscillatory near-extinction regime, characterized by increasingly chaos-like pressure fluctuations. As a result, the confidence in the estimated friction coefficient from such data is expected to deteriorate. However, near the nominal condition, the pressure fluctuations are much more well-behaved, which yields a reasonably accurate estimate of the combustor's health status, and consequently an early warning on probability of flame extinction can be issued with a high level of confidence.

The information from Table 2 and Fig. 5 (including the estimate of f and different confidence intervals) can be utilized to monitor the combustor tailpipe friction coefficient and to generate early warnings of flame extinction.

6   SUMMARY, CONCLUSIONS, AND FUTURE WORK

Recent literature has reported potential applications of SDF [8] for early detection of forthcoming failures in human-engineered complex systems [16]. The work reported in this paper formulates and validates a dynamic data-driven statistical method for anomaly detection in thermal pulse combustors and estimation of the tailpipe friction coefficient as a failure precursor; the objective is the prognosis of forthcoming operational disruptions (e.g. flame extinction) well in advance of their occurrence. The detection procedure is demonstrated by simulation on an experimentally validated model of a thermal pulse combustor [1]. The algorithm of the SDF-based anomaly detection is built upon the principles of *Information Theory, Symbolic Dynamics*, and *Statistical Pattern Recognition*. Specifically, the paper delineates a method of

utilizing time series of pressure oscillation signals for estimating the friction coefficient of the tailpipe wall of thermal pulse combustors, along with corresponding statistical confidence bounds.

The friction coefficient is a representation of the drag force in the tailpipe, which can be directly measured using sensors (e.g. load cell or strain gauge) and hence this parameter estimation method can be experimentally validated. Since the transition to unstable combustion dynamics, resulting from variations in several different parameters, is found to follow similar routes, SDF is potentially a useful tool for detecting various types of anomalies in experimental combustors.

Further theoretical, computational, and experimental work is necessary before the SDF-based anomaly detection tool can be considered for incorporation into the instrumentation and control system of commercial-scale pulse combustors. For example, the lumped parameter model needs to be calibrated by a computational fluid dynamic (CFD) model of a thermal pulse combustor before experimental validation of anomaly detection is planned and executed on a laboratory-scale combustor.

From the perspectives of anomaly detection in combustion and other complex dynamical systems, future research is recommended in the following areas:

(a) comparative evaluation of SDF with existing pattern recognition tools;
(b) identification and estimation of multiple parameters in the presence of multi-source multi-time-scale anomalies;
(c) quality assurance and automated calibration of the sensor time-series data;
(d) application of the above algorithms to information-based real-time control.

REFERENCES

1 **Richards, G. A., Morris, G. J., Shaw, D. W., Kelley, S. A.,** and **Welter, M. J.** Thermal pulse combustion. *Combust. Sci. Technol.*, 1993, **94**, 57–85.
2 **Rhode, M. A., Rollins, R. W., Markworth, A. J., Edwards, K. D., Nguyen, K., Daw, C. S.,** and **Thomas, J. F.** Controlling chaos in a model of thermal pulse combustion. *J. Appl. Phys.*, 1995, **78**, 2224–2232.

3 **Duda, R., Hart, P.,** and **Stork, D.** *Pattern classification*, 2001 (Wiley Interscience, New York, NY).

4 **Kantz, H.** and **Schreiber, T.** *Nonlinear time series analysis*, 2nd edition, 2004 (Cambridge University Press, Cambridge, UK).

5 **Datta, S., Mukhopadhyay, A.,** and **Sanyal, D.** Effect of tailpipe friction factor on the nonlinear dynamics of a thermal pulse combustor. In Proceedings of the 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Sacramento, CA, USA, 2006, paper no. AIAA-2006-4396.

6 **Gupta, S., Ray, A., Sarkar, S.,** and **Yasar, M.** Fault detection and isolation in aircraft gas turbine engines. Part I: underlying concept. *Proc. IMechE, Part G: J. Aerospace Engineering*, 2008, **222**(G3), 307–308.

7 **Sarkar, S., Yasar, M., Gupta, S., Ray, A.,** and **Mukherjee, K.** Fault detection and isolation in aircraft gas turbine engines. Part II: validation on a simulation test bed. *Proc. IMechE, Part G: J. Aerospace Engineering*, 2008, **222**(G3), 319–330.

8 **Ray, A.** Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Process.*, 2004, **84**(7), 1115–1130.

9 **Gupta, S., Ray, A.,** and **Mukhopadhyay A.** Anomaly detection in thermal pulse combustors using symbolic time series analysis. *Proc. IMechE, Part I: J. Systems and Control Engineering*, 2006, **220**(I5), 339–351.

10 **In, V., Spano, M. L., Neff, J. D., Ditto, W. L., Daw, C. S., Edwards, K. D.,** and **Nguyen, K.** Maintenance of chaos in a computational model of thermal pulse combustor. *Chaos*, 1997, **7**, 605–613.

11 **Rajagopalan, V.** and **Ray, A.** Symbolic time series analysis via wavelet-based partitioning. *Signal Process.*, 2006, **86**(11), 3309–3320.

12 **Badii, R.** and **Politi, A.** *Complexity hierarchical structures and scaling in physics*, 1997 (Cambridge University Press, Cambridge, UK).

13 **Buhl, M.** and **Kennel, M. B.** Statistically relaxing to generating partitions for observed time-series data. *Phys. Rev. E*, 2005, **71**(4), 046213.

14 **Mallat, S.** *A wavelet tour of signal processing 2/e*, 1998 (Academic Press, Boston, MA).

15 **Rao, C., Ray, A., Sarkar, S.,** and **Yasar, M.** Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal Image Video Process.*, 2008, DOI: 101007/s11760-008-0061-8.

16 **Gupta, S., Ray, A.,** and **Keller, E.** Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. *Mech. Syst. Signal Process.*, 2007, **21**(2), 866–884.

17 **Tarantola, A.** *Inverse problem theory*, 2005 (Society for Industrial and Applied Mathematics, Philadelphia, PA).

18 **Snedecor, G. W.** and **Cochran, W. G.** *Statistical methods, 8/e*, 1989 (Iowa State University Press, Ames, IA, USA).

19 **Bishop, C. M.** *Neural networks for pattern recognition*, 1995 (Oxford University Press Inc., New York, NY).

APPENDIX

Notation

A_e	tailpipe cross-sectional area (m^2)
A_s	combustor surface area (m^2)
B	pre-exponential factor for single step chemical kinetics (m^{3}kg^{-1}K$^{-1/2}$s^{-1})
C_p	specific heat at constant pressure (J/kg K)
D	window length on a symbolic sequence
D_{tp}	diameter of the tailpipe (m)
f	friction coefficient (dimensionless)
$\widehat{f}$	maximum-likelihood estimate of friction coefficient
h	convective heat transfer coefficient (W/m^2 K)
k	number of parameters in a statistical distribution
ℓ	number of simulations performed
L_{c1}	(V/A_s), first characteristic length (m)
L_{c2}	(V/A_e), second characteristic length (m)
L_{tp}	length of the tailpipe (m)
m	number of partitions used to form the pattern matrix
$\dot{m}_e$	mass flowrate at combustor exit (kg/s)
$\dot{m}_i$	mass flowrate at combustor inlet (kg/s)
p_j	probability of jth segment of partition
$\boldsymbol{p}^0$	probability vector at nominal condition
$\boldsymbol{p}^k$	probability vector at kth (slow-scale) time epoch
P	pressure in the combustion zone (Pa)
$\widetilde{P}$	(P/P_0), normalized pressure (dimensionless)
P_e	pressure at the tailpipe entrance (Pa)
$\widetilde{P}_e$	(P_e/P_0), normalized tailpipe pressure (dimensionless)
P_0	ambient pressure (Pa)
r	number of bins in histogram construction
R	($\gamma - 1$)C_p/γ, gas constant (J/kg K)
$\dot{R}_f$	fuel reaction rate (kg/m^3 s)
S	Shannon entropy of the symbol sequence
SDF	symbolic dynamic filtering
t	time in the fast scale of process dynamics
t_k	kth time epoch in the slow scale
T	temperature in the combustion zone (K)
$\widetilde{T}$	(T/T_0), normalized temperature (dimensionless)
T_a	activation temperature (K)
T_e	temperature at the tailpipe entrance (K)
$\widetilde{T}_e$	(T_e/T_0), normalized tailpipe temperature (dimensionless)
T_w	wall temperature in the combustion zone (K)
$\widetilde{T}_w$	(T_w/T_0), normalized wall temperature (dimensionless)
T_0	ambient temperature (K)
u	gas velocity in the tailpipe (m/s)
$\widetilde{u}$	$u/(L_{c2}/\tau_f)$ (dimensionless)
V	volume of the combustor (m^3)
Y_f	average fuel mass fraction in the combustor chamber (dimensionless)
Y_{fi}	fuel mass fraction at the combustor inlet (dimensionless)

Z_e	$(\dot{m}_e/V)$ (kg/m^3 s)		ρ	density in the combustion zone (kg/m^3)
Z_i	$(\dot{m}_i/V)$ (kg/m^3 s		ρ_0	ambient density (kg/m^3)
			σ	standard deviation of a random variable
γ	ratio of specific heats (dimensionless)			
Δf	decrement in f from one slow-time epoch to another		Σ	alphabet or set of symbols for partitioning
			$\lvert\Sigma\rvert$	cardinality of the alphabet Σ
Δh_c	enthalpy of combustion (J/kg)		τ_c	characteristic chemical reaction time (s)
η	tolerance in steady-state probability variation		τ_f	characteristic flow time (s)
			τ_h	characteristic heat transfer time (s)
μ	mean of a random variable		Υ	pattern matrix
ν_0	stoichiometric oxygen–fuel ratio by mass (dimensionless)		χ^2	statistical test
			ψ^k	anomaly measure at kth time epoch