Mechanical Systems and Signal Processing 159 (2021) 107746

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Measure invariance of ergodic symbolic systems for low-delay detection of anomalous events

Najah F. Ghalyan [a,b], Asok Ray [c,*]

[a] Department of Mechanical Engineering, Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Mechanical Engineering, University of Kerbala, Kerbala 56001, Iraq
[c] Department of Mechanical Engineering and Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 23 May 2020
Received in revised form 24 January 2021
Accepted 9 February 2021

Communicated by Jerome Antoni

Keywords:
Symbolic dynamics
Measure invariance
Spectral analysis
Markov modeling
Ergodicity
Low-delay detection

ABSTRACT

This paper investigates the spectral properties of ergodic measure of symbolic systems for applications to signal processing, pattern recognition, and anomaly detection in uncertain dynamical systems. The underlying algorithm is built upon the concept of ergodic sequences of measure-preserving transformations (MPT) on probability spaces, where non-stationary probabilistic finite state automata (PFSA) are constructed using short-length windows of time-series measurements. The resulting PFSA are non-homogeneous Markov, in which spectral properties of the MPT depend on several parameters that include the window length. The paper also develops an MPT-based metric to quantify the divergence of the evolving PFSA from that of the nominal PFSA; this information is then used for pattern classification and anomaly detection with low-delay tolerance. The MPT-based methodology has been validated with experimental data generated from a laboratory apparatus that deals with detection of thermoacoustic instabilities in combustion processes, which are known to have chaotic characteristics on a time scale of milliseconds. In this application, the concepts of MPT and ergodicity have been used to develop a novel symbolic time series analysis (STSA)-based detection method, whose performance is validated by comparison with two well-known techniques, namely, hidden Markov model (HMM) and cumulative sum (CUSUM), on the same experimental data. The results consistently show superior performance of the proposed MPT-based STSA.

1. Introduction

The topic of low-delay detection has attracted the attention of many researchers over the last several decades. For example, Shiryaev [1] established Bayesian methods to address the problem of fast sequential detection of changes (e.g., disruptions) in time series of independent and identically distributed (iid) observations by using the theory of optimal stopping. Subsequently, Tartakovsky and Veeravalli [2] established a general asymptotic change-point detection theory that is not limited to iid observations.

The above discussion largely applies to cases where (i) the change in the time series occurs as a result of abrupt variations in the underlying model structure [3], and (ii) the distribution of observations before a change point occurs and the distribution after a change point are both known, in which case the problem can be treated as a standard (e.g., binary) hypothesis

* Corresponding author.
E-mail address: axr2@psu.edu (A. Ray).

https://doi.org/10.1016/j.ymssp.2021.107746
0888-3270/

testing. Nevertheless, in many applications, changes in the time series occur due to gradual degradation in the underlying dynamical system. Moreover, as it is realistic in practice, the probability distribution after occurrence of the change is typically unknown. In this context, the method of hidden Markov model (HMM) [4,5] has been widely used for detection of both abrupt changes and gradually evolving anomalies in diverse applications, where the post-change distribution is unknown [6]. Another well-known technique in this regard is the cumulative sum (CUSUM), developed by Page [7], which has been widely used for change point detection [3,8,9]. CUSUM is a sequential method and it requires a cumulative statistic which is thresholded to decide whether a change has occurred or not. An alternative approach for detection of change points involves: (i) selection of a cost function to measure goodness-of-fit of the generated time series to the null model, and (ii) identification of signal segments for which the goodness of fit falls below a threshold. In this context, Bai and Perron [10] considered a time series generated by a linear regression model that undergoes multiple structural changes (i.e., breaks) at unknown times with the sum of squared residuals between the regression model outputs and the observed time series as the goodness of fit. Later, Bai and Perron [11] introduced a computationally-efficient algorithm that finds the global minimum of the sum of squared residuals by dynamic programming. However, the above technique [11] considers the entire time series for change point detection and thus imposes no constraint on the amount of allowed *delay* in making a decision on detection. Nevertheless, in real-time applications, the decision on detection must be made with *low* delay (e.g., by using short observation windows).

The concept of symbolic time series analysis (STSA) (e.g., [12–14]) has been used to construct Markov models for anomaly and change detection from observed windows of sensor data. In the framework of STSA, a (finite-length) time series is partitioned for conversion into a string of symbols from a (finite-cardinality) alphabet $\mathcal{A}$ [15–18]. Subsequently, probabilistic finite state automata (PFSA) are constructed from the symbol strings [19–21], in which the probability distribution of the *emitted* symbols depends upon the immediately preceding D symbols, where D is a positive integer called Markov depth. Such a PFSA is called a *D-Markov machine*, which has found diverse applications in pattern classification and anomaly detection (e.g., [14,22]). The main distinction between HMM and D-Markov machines is that the algebraic structure of state transition in HMM is allowed to be non-deterministic, while it is restricted to be deterministic in D-Markov machines [20,21,14]; this restriction yields significant computational advantages of PFSA [19] over HMM [4]. Moreover, training in HMM is typically done statistically by using an expectation maximization tool (e.g., Baum-Welch algorithm [4,5]). In addition to its iterative computation cost, such a method only guarantees convergence to a point, having zero gradient with respect to the HMM parameters, and therefore this point could be only a local optimum. In contrast, the deterministic algebraic structure of D-Markov machines makes the modeling process simpler and less prone to the local optimum issue, where the model can be trained using the frequency counting method [21], for example.

In the above setting, the selection of the window length, L, of the time series used to construct the PFSA in D-Markov machines largely depends on the Markov depth D, the alphabet size $|\mathcal{A}|$, and the nature of the particular underlying process generating the time series [21]. To find a lower bound on the window length of the time series required to estimate the PFSA parameters, one may consider an increasing sequence of L. Under the assumption of asymptotic statistical stationarity [21], the computed state transition probability matrix converges to a constant matrix; as a consequence, L may become arbitrarily large [14]. Thus, one may choose the smallest L at which the stochastic matrix tends to be approximately time-invariant; the resulting model could be treated as a time-homogeneous Markov chain [5]. However, this scenario would typically require a large value of L, which could be infeasible in many physical applications, where decisions need to be made with low-delay tolerance.

The notion of measure-preserving transformation (MPT) has evolved in the discipline of Statistical Mechanics to represent Hamiltonian systems (e.g., Louisville Theorem [23]), where the total energy of the dynamical system is invariant. A key concept in this regard is that, even though a measure-preserving dynamical system is described by a transformation with time-varying eigenvalues, the absolute values of the eigenvectors remain time-invariant under the ergodicity assumption [24]. Based on this rationale, Ghalyan and Ray [25] introduced a methodology for constructing PFSA, from short-length time series, which generate non-homogeneous Markov chain models for describing the uncertain dynamics of the physical process. As a result, the time-invariance of the eigenvectors, which reflects measure-invariance and ergodicity of the dynamical system, is used to decide the window length of the time series required to construct PFSA. Unlike the standard STSA ([14,21]) which requires increasing the window length until the resulting PFSA is no longer significantly changing, the MPT-based STSA framework proposed in [25] can be used to select a minimum window length for which the eigenvectors are nearly constant. The rationale is that anomalies could often be associated with a variation in the system's total energy, which naturally makes the system no longer measure-invariant and thus the eigenvectors are no longer time-invariant. Along this line, Ghalyan and Ray [25] proposed a metric of variability of eigenvectors, as a measure of anomaly, for detection of fatigue damage in polycrystalline alloys using ultrasonic time series measurements.

The current paper is a major extension of the authors' previous work [25] and provides a significantly more detailed exposition to the subject matter by focusing on the fundamental principles of symbol-sequence generation [24] and their relationships with the dynamics of the underlying physical process from a measure-theoretic perspective. Although the main results in both the previous paper [25] and the current paper are built upon the concepts of invariance and ergodicity of the probability measure of the generated symbolic systems, the previous paper [25] does not precisely define a probability measure of symbolic systems. In contrast, the current paper develops a concrete mathematical framework by using the concept of cylinders [24] later in the paper, which is used to define the probability measure for the generated symbolic systems. In essence, the current paper lays a mathematical foundation, which is necessary not only to thoroughly understand the theory and the application results, but also to gainfully make future extensions of the reported work. To this end, a new detection methodology has been developed in the framework of MPT-based STSA, which is experimentally validated on an application, different from the one reported in [25] by comparison with two well-known techniques, namely, HMM and CUSUM, serving as benchmarks. The results show consistent superiority of the proposed MPT-based STSA over both HMM and CUSUM.

In particular, the application investigated in this paper is directly related to real-time monitoring and active control of thermoacoustic instabilities (TAI) in combustion processes. TAI are typically caused by spontaneous excitation of natural modes of acoustic waves [26], which induces large-amplitude self-sustained chaotic pressure oscillations in the combustion chamber, potentially leading to damage in its mechanical structures. A critical issue here is that TAI may develop within a very short time (e.g., tens of milliseconds), for which a real-time algorithm for early detection is required for fast actuation of control signals to circumvent any potential damage.

Contributions: Major contributions of the paper are summarized below:

1. *A measure-theoretic spectral analysis of ergodic symbolic systems*: A measure-theoretic framework is rigorously established for spectral analysis of ergodic symbolic systems, defined on a shift space whose topological basis is described by relying on the concept of *cylinders* [24]. This setting provides a thorough understanding of the application results in the paper, and lays a concrete mathematical foundation for potential future work in signal analysis and pattern recognition from observed time series measurements.
2. *Time-inhomogeneous Markov-chain modeling for pattern classification with low-delay tolerance*: The proposed methodology facilitates low-delay detection of anomalies in uncertain dynamical systems using short-length time series of sensor data. To this end, an MPT-based STSA algorithm has been developed on ergodic sequences of measure-preserving transformations, with a detailed exposition to the underlying theory.
3. *Validation of the proposed detection methodology*: The underlying algorithm is experimentally validated on a laboratory-scale apparatus for detection of TAI in combustion systems using time series of pressure measurements. The detection performance[1] is compared with that of both HMM and CUSUM to serve as benchmarks.

Organization: The paper is organized in six sections including the present section. Section 2 briefly provides the background information on measure-preserving transformation (MPT) and ergodicity. These background information are required for a full understanding of the material in the subsequent section, namely Section 3, which succinctly presents the mathematical principles of symbolic time series analysis (STSA). This section also develops a measure-theoretic framework for spectral analysis of ergodic symbolic systems generated from observed time series, which plays a central role in developing the detection algorithm presented in this paper. Section 4 presents the technical approach and develops the algorithm. Section 5 validates the proposed concept and the underlying algorithm with experimental data. Section 6 summarizes and concludes the paper along with recommendations for future research.

2. Background information

This section provides a brief analysis of the spectral properties of measure-preserving transformation (MPT) sequences with ergodic probability measure, which form the backbone of the statistical pattern classification and anomaly detection methodology presented in this paper. A key concept here is that, even though a measure-preserving dynamical system is described by a transformation with time-varying eigenvalues, the absolute values of the eigenvectors remain time-invariant under the ergodicity assumption [25]; the details are extensively reported by Cornfeld et al. [24]. The following definitions, which are available in standard literature, are presented below for completeness of the paper and ease of its readability.

Definition 1. Let Ω be a nonempty set. A collection $\mathbb{T}$ of subsets of Ω is called a topology and the members of $\mathbb{T}$ are called $\mathbb{T}$-open (or open) sets, provided that the following three conditions are satisfied.

- $\varnothing \in \mathbb{T}$ and $\Omega \in \mathbb{T}$
- Closure under finite intersection, i.e., $\bigcap_{k=1}^{n} E_k \in \mathbb{T}$ if $E_k \in \mathbb{T}$ for $k = 1, \ldots, n \quad \forall n \in \mathbb{N} \triangleq \{1, 2, 3, \ldots\}$.
- Closure under arbitrary (i.e., finite, countable, or uncountable) union, i.e., $\bigcup_{\alpha} E_\alpha \in \mathbb{T}$ if $E_\alpha \in \mathbb{T}$ $\forall \alpha \in \mathbb{I}$ that is an arbitrary index set.

The pair $(\Omega, \mathbb{T})$ is said to form a topological space, where the complement of an open set is called a closed set, i.e., if S is an open set, then $\Omega \setminus S$ is a closed set.

[1] The *Receiver Operating Characteristic* (ROC) curves [27] have been used in this paper for assessing the detection performance of MPT-based STSA relative to that of HMM [4,5] and CUSUM [7].

Definition 2. Let $(\Omega, \mathbb{T})$ be a topological space. Then, a basis of $(\Omega, \mathbb{T})$ is a (nonempty) collection $\mathcal{B}$ of (nonempty) subsets of Ω (called basis elements), provided that the following two conditions are satisfied.

- For every $\omega \in \Omega$, there exists $B \in \mathcal{B}$ such that $\omega \in B$.
- If ω belongs to the intersection of two arbitrary basis elements $B_1, B_2 \in \mathcal{B}$, then there exists a basis element B_3 containing ω such that $B_3 \subset B_1 \cap B_2$.

Definition 3. Let Ω be a nonempty set. A collection $\mathcal{E}$ of subsets of Ω is called a σ-algebra and the members of $\mathcal{E}$ are called $\mathcal{E}$-measurable (or measurable) sets provided that the following three conditions are satisfied.

- $\Omega \in \mathcal{E}$.
- If $E \in \mathcal{E}$, then $\Omega \setminus E \in \mathcal{E}$
- A countable union of measurable sets is measurable, i.e., if $\{E_k\}$ is a countable collection of members of $\mathcal{E}$, then $\bigcup_k E_k \in \mathcal{E}$.

The pair $(\Omega, \mathcal{E})$ is said to form a measurable space.

Definition 4. Let $(\Omega, \mathcal{E})$ be a measurable space. Then, the value of the set function defined as $P : \mathcal{E} \to [0, 1]$ is called a (probability) measure provided that the following two conditions are satisfied.

- $P[\Omega] = 1$.
- If $\{E_k\}$ is a countable collection of members of $\mathcal{E}$, then

$$P\left[\bigcup_k E_k\right] \leqslant \sum_k P[E_k]$$

and the equality holds if the members of $\{E_k\}$ are pairwise disjoint, i.e., if $E_i \cap E_j = \emptyset \; \forall \; i \neq j$.

The triple $(\Omega, \mathcal{E}, P)$ is called a probability space.

Definition 5. If two measurable sets $E, F \in \mathcal{E}$ are such that $P[E \; \Delta \; F] = 0$, then it is said that $E = F$ P-almost everywhere (abbreviated as P-ae). Note: Δ is the symmetric difference on sets such that $A \Delta B \triangleq (A \setminus B) \cup (B \setminus A)$.

Remark 6. All measurable sets that are equal P-ae form an equivalence class; members of this equivalence class are P-almost equal sets.

Definition 7. Let $(\Omega, \mathcal{E}, P)$ be a probability space and let $T : (\Omega, \mathcal{E}, P) \to (\Omega, \mathcal{E}, P)$ be a transformation. Then, T is called measurable if $T^{-1}E \in \mathcal{E} \; \forall E \in \mathcal{E}$.

Definition 8. A measurable set $E \in \mathcal{E}$ is called T-invariant if $P\left[E \; \Delta \; T^{-1}E\right] = 0$, which implies that $Tx \in E$ for P-almost all $x \in E$. Furthermore, a function $f : \Omega \to [0, \infty)$ is called T-invariant if $f(Tx) = f(x)$ for P-almost all $x \in \Omega$. A measurable transformation T is called a measure-preserving transformation (MPT) if $P\left[T^{-1}E\right] = P[E] \quad \forall E \in \mathcal{E}$.

Definition 9. A measure-preserving transformation T is called an endomorphism if T is surjective (i.e., onto). If T is bijective (i.e., one-to-one and onto) and preserves measure, then it is called an automorphism.

Remark 10. The concept of MPT has been widely used to investigate the asymptotic properties of random sequences in statistical mechanics [24]. For an endomorphism T on a (finite) measure space $(\Omega, \mathcal{E}, P)$, every measurable set E has the recurrence property in the sense that once E is visited, it would be revisited infinitely many times [24]; that is, if $x \in E$, then there are infinitely many values of n such that $T^{(n)}x \in E$.

Definition 11. Let S be a (nonempty) set with a binary operation $\circ$. Then, S is called a semigroup if the following two conditions are satisfied:

- Closure: $\circ: \ S \times S \to S$.
- Associativity: $(x \circ y) \circ z = x \circ (y \circ z) \quad \forall\ x, y, z \in S$.

Definition 12. [28] Let $\left\{T^{(n)}\right\}$ be a one-parameter semigroup of endomorphisms on a probability space $(\Omega, \mathcal{E}, P)$. Then, a function $f \in L_1(P)$ is said to be an eigenfunction of $\left\{T^{(n)}\right\}$ with the eigenvalue $\lambda^{(n)}$ if f is a non-zero function and $f\left(T^{(n)}\right) = \lambda^{(n)} f$ P-ae for all $n \in \mathbb{N}$.

Definition 13. [24] Let $\left\{T^{(n)}\right\}$ be a one-parameter semigroup of endomorphisms on a probability space $(\Omega, \mathcal{E}, P)$. Then, the sequence $\left\{T^{(n)}\right\}$ is said to be ergodic if each $T^{(n)}$-invariant set $E \in \mathcal{E}$ is trivial, i.e., either $P[E] = 0$ or $P[E] = 1$ $\forall n \in \mathbb{N}$. Equivalently, the measure P is said to be $\left\{T^{(n)}\right\}$-ergodic.

Remark 14. The concept of ergodicity has been widely used in Statistical Mechanics and probabilistic modeling of dynamical systems [24]. In an ergodic process, it is sufficient to have a single adequately-long realization in order to characterize the statistics of the underlying process. Given any discrete-time realization $\{x_n\}$ of an ergodic process $\{X_n : X_n \in L_1(P)\}$, as $n \to \infty$, the time average $\frac{1}{n}\sum_{j=1}^{n} x_j$ converges P-ae and in $L_1(P)$ to the ensemble average $\int_\Omega X_n dP$ [29].

Another useful formulation of ergodicity is as follows: Given a probability space $(\Omega, \mathcal{E}, P)$, the sequence of endomorphisms $\left\{T^{(n)}\right\}$ is ergodic if and only if every invariant measurable function is equal to a constant P-ae on Ω. Based on this formulation, it can be tested whether a sequence $\left\{T^{(n)}\right\}$ is ergodic or not by looking at its eigenfunction f corresponding to the eigenvalue $\lambda^{(n)} = 1$, for which $f(x) = f\left(T^{(n)}x\right)$ for P-almost all $x \in \Omega$ and for all $n \in \mathbb{N}$. Hence, f is an invariant function under $\left\{T^{(n)}\right\}$ and therefore is a constant P-ae if and only if $\left\{T^{(n)}\right\}$ is ergodic. In this context, an ergodic sequence of endomorphisms has an important property given by the following theorem [24].

Theorem 15. *Let $(\Omega, \mathcal{E}, P)$ be a probability space, and let $\left\{T^{(\theta)}\right\}$ be a sequence of endomorphisms, where $\theta \in [0, \infty)$. Then, $\left\{T^{(\theta)}\right\}$ is ergodic if and only if the absolute value of every eigenfunction is a constant P-ae. That is, if f is an eigenfunction of the endomorphism $T^{(\theta)}$, then $|f(x)|$ is a constant for P-almost all $x \in \Omega$.*

Proof. The proof of the theorem is given in [24].

3. Symbolic time series analysis

Before delving into the technical approach and the algorithm therein, it is necessary to provide an adequate mathematical background for describing dynamical stochastic systems from observed time series in a symbolic setting, which is a major theme of the current paper. It is shown in this section how these symbolic systems can significantly reduce the complexity of testing measure-invariance and ergodicity properties as well as generate Markov models of the underlying dynamical systems from observed time series data.

3.1. Measure-invariant symbolic systems

Let a dynamical system on the probability space $(\Omega, \mathcal{E}, P)$ be described by a quadruple $(\Omega, \mathcal{E}, P, T)$, where the transformation $T : (\Omega, \mathcal{E}, P) \to (\Omega, \mathcal{E}, P)$ is a P-measurable mapping of Ω onto itself. A symbolic representation of the dynamical system can be generated by partitioning the space $(\Omega, \mathcal{E}, P)$.

Definition 16. A *partition* of the probability space $(\Omega, \mathcal{E}, P)$ is a (non-empty) finite-cardinality family of pairwise disjoint (non-empty) members of $\mathcal{E}$, whose union is Ω.

Given a partition $\alpha = \{A_1, A_2, \ldots, A_{|\alpha|}\}$ of $(\Omega, \mathcal{E}, P)$, an alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\alpha|}\}$ is constructed, where the symbols a_i's bear a one-to-one correspondence to the members A_i's of α. Let $\mathcal{F}$ be the σ-algebra generated by $\mathcal{A}$, and let v be a probability measure such that $v(a_i) = P(A_i)$ $\forall a_i \in \mathcal{A}$, which defines the measure space $(\mathcal{A}, \mathcal{F}, v)$.

Let $\mathcal{A}^{\mathbb{N}}$ denote the set of all one-sided semi-infinite symbol sequences, i.e., $\mathcal{A}^{\mathbb{N}} \triangleq \{s_1, s_2, \ldots : s_k \in \mathcal{A} \text{ and } k \in \mathbb{N}\}$, where $\mathbb{N} \triangleq \{1, 2, 3, \ldots\}$. Hence, the transformation $T : (\Omega, \mathcal{E}, P) \to (\Omega, \mathcal{E}, P)$ and a partition α of the space $(\Omega, \mathcal{E}, P)$ together generate a (left) shift operator $\Sigma : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$ defined as:

$$\Sigma(s_1, s_2, \ldots) = (s_2, s_3, \ldots) \tag{1}$$

Definition 17. Given a block of symbols $(\sigma_1, \ldots, \sigma_N)$, where $\sigma_i \in \mathcal{A}$ and $N \in \mathbb{N}$, a cylinder $C^n_{\sigma_1,\ldots,\sigma_N}, n \in \mathbb{N}$, generated by $(\sigma_1, \ldots, \sigma_N)$ is the collection of all members $\mathbf{S} \in \mathcal{A}^{\mathbb{N}}$ such that the symbol block $(\sigma_1, \ldots, \sigma_N)$ occurs at the location n, i.e.,

$$C^n_{\sigma_1,\ldots,\sigma_N} = \left\{\mathbf{S} \in \mathcal{A}^{\mathbb{N}} : s_n = \sigma_1, \ldots, s_{n+N-1} = \sigma_N\right\} \tag{2}$$

and is called a centered cylinder if it has the form $C^1_{\{\sigma_1,\ldots,\sigma_N\}}$.

Remark 18. A cylinder is both an open and a closed subset of $\mathcal{A}^{\mathbb{N}}$ [30]. Moreover, the centered cylinders $C^1_{\{s_1,\ldots,s_N\}}, N \in \mathbb{N}$, form a topological basis for $\mathcal{A}^{\mathbb{N}}$ (see Definition 2).

Let $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi)$ denote the Cartesian product of countably infinitely many copies of the measurable space $(\mathcal{A}, \mathcal{F})$, where $\mathcal{F}_\Pi$ is the product σ-algebra generated by the cylinders $C^n_{\sigma_1,\ldots,\sigma_N}$ for all $n, N \in \mathbb{N}$, corresponding to all feasible initial conditions of the dynamical system $(\Omega, \mathcal{E}, P, T)$ [31]. Let m define a probability measure on $\mathcal{F}_\Pi$, given by:

$$m\left(C^n_{\{\sigma_1,\ldots,\sigma_N\}}\right) \triangleq \nu(s_n = \sigma_1, \ldots, s_{n+N-1} = \sigma_N) \tag{3}$$

That is, the probability of a symbol block is equal to the measure of the cylinder generated by that symbol block. In this way, a probability measure space $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m)$ and a shift system $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m, \Sigma)$ are defined. It follows from Eq. (3) that the symbolic representation of the dynamical system $(\Omega, \mathcal{E}, P, T)$ is stationary if and only if the measure m of the generated cylinders is n-invariant. In this context, a *Stationary Symbolic System* is the shift system $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m, \Sigma)$ whose cylinders are n-invariant.

A partition α of the probability space of the dynamical system $\left\{\left(\Omega, \mathcal{E}, P, T^{(n)}\right) : n \in \mathbb{N}\right\}$, where $\left\{T^{(n)}\right\}$ is a transformation sequence, generates a $(\mathcal{E} - \mathcal{F}_\Pi)$-measurable coding map [24] (also called partitioning map) $\Phi^\alpha : \Omega \to \mathcal{A}^{\mathbb{N}}$, defined for all $\omega \in \Omega$, as follows:

$$(\Phi^\alpha(\omega))_n = a_i \in \mathcal{A} \text{ if and only if } T^{(n)}(\omega) \in A_i \tag{4}$$

i.e., the n-*th* element of the symbol sequence $\Phi^\alpha(\omega), \omega \in \Omega$, is $a_k \in \mathcal{A}$ if and only if $T^{(n)}(\omega) \in A_k, A_k \in \alpha$.

Let $X_{n\in\mathbb{N}}$ be a stochastic process defined by the dynamical system $\left\{\left(\Omega, \mathcal{E}, P, T^{(n)}\right) : n \in \mathbb{N}\right\}$. Then, $\mathbf{S}_n \triangleq (\Phi^\alpha(X_{n\in\mathbb{N}}))_n$ is a random variable on the probability space $(\mathcal{A}, \mathcal{F}, \nu)$, and $\{\mathbf{S}_n\}$ defines a *symbolic* stochastic process on the shift space $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m)$. Therefore, for every stochastic process $\{X_{n\in\mathbb{N}}\}$, there exists a symbolic representation $\{\mathbf{S}_{n\in\mathbb{N}}\}$ that can be identified by finding a partition of the measure space of the original stochastic process[2]. If the underlying transformations $\left\{T^{(n)}\right\}$ are measure-preserving, the resulting symbolic representation is called a *Measure-Invariant Symbolic System*.

It is noted that the measure-invariance property of a sequence of transformations does not guarantee stationarity of the corresponding symbolic stochastic process. Therefore, the probabilistic finite state automata (PFSA) (see Section 3.2) generated by partitioning the probability space of the dynamical system $\left\{\left(\Omega, \mathcal{E}, P, T^{(n)}\right) : n \in \mathbb{N}\right\}$, are non-stationary in general. Consequently, the resulting state transition probability matrices of D-Markov machines (see Section 3.3) are non–homogeneous in general.

3.2. Probabilistic finite state automata

As seen in Section 3.1, a partition of the probability space $(\Omega, \mathcal{E}, P)$ of a given dynamical system generates a symbol string that is used to construct a probabilistic finite state automaton (PFSA). The PFSA model describes the statistics of the underlying stochastic process on the shift space $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m)$. The probability of the generated PFSA states is given by the measure of the corresponding cylinders in $(\mathcal{A}^{\mathbb{N}}, \mathcal{F}_\Pi, m)$. The following definitions, which are available in standard literature (e.g., [14,21]), are recalled here for completeness of the paper.

[2] If the stochastic process is defined over both positive and non-positive times, i.e., $\{X_n : n \in \mathbb{Z}\}$, a coding map $\Phi^\alpha : \Omega \to \mathcal{A}^{\mathbb{Z}}$ can be used to generate a symbolic representation given by the *two-sided shift system* $(\mathcal{A}^{\mathbb{Z}}, \mathcal{F}_\Pi, m, \Sigma)$. In this case, the centered cylinders take the form $C^{-N}_{\{\sigma_{-N},\ldots,\sigma_{-1},\sigma_0,\sigma_1,\ldots,\sigma_N\}}$.

Definition 19. A finite state automaton (FSA) G, having a deterministic algebraic structure, is a triple $(\mathcal{A}, \mathcal{Q}, \delta)$ where:

- $\mathcal{A}$ is a (nonempty) finite alphabet, i.e., $|\mathcal{A}| \in \mathbb{N}$.
- $\mathcal{Q}$ is a (nonempty) finite set of states, i.e., $|\mathcal{Q}| \in \mathbb{N}$.
- $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is a state transition map.

Definition 20. A symbol block, also called a word, is a finite-length string of symbols belonging to the alphabet $\mathcal{A}$, where the length of a word $w \triangleq s_1 s_2 \ldots s_\ell$ with $s_i \in \mathcal{A}$ is $|w| = \ell$, and the length of the empty word ϵ is $|\epsilon| = 0$. The parameters of FSA are extended as:

- The set of all words, constructed from symbols in $\mathcal{A}$ and including the empty word ϵ, is denoted as $\mathcal{A}^{\star}$.
- The set of all words, whose suffix (respectively, prefix) is the word w, is denoted as $\mathcal{A}^{\star}w$ (respectively, $w\mathcal{A}^{\star}$).
- The set of all words of (finite) length ℓ, where $\ell \in \mathbb{N}$, is denoted as $\mathcal{A}^{\ell}$.

Remark 21. A symbol string (or word) w that occurs at time n generates a cylinder C_w^n in the symbol-sequence space $\mathcal{A}^{\mathbb{N}}$, and the probability of a string w is given by the measure of that cylinder, i.e. $P(w) = m(C_w^n)$ (see Section 3.1).

Definition 22. A probabilistic finite state automaton (PFSA) $\mathcal{K}$ is a pair (G, π), where:

- Deterministic FSA G is the underlying algebraic structure of PFSA $\mathcal{K}$.
- The morph function (also known as the symbol generation probability function) $\pi : \mathcal{Q} \times \mathcal{A} \to [0, 1]$ satisfies the condition: $\sum_{\sigma \in \mathcal{A}} \pi(q, \sigma) = 1$ for all $q \in \mathcal{Q}$.

The state transition probability mass function $\kappa : \mathcal{Q} \times \mathcal{Q} \to [0, 1]$ is constructed by combining δ and π, which can be structured as a $|\mathcal{Q}| \times |\mathcal{Q}|$ matrix $\mathcal{T}$. In that case, the PFSA can be described as the triple $\mathcal{K} = (\mathcal{A}, \mathcal{Q}, \mathcal{T})$.

It is noted that the $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition probability matrix $\mathcal{T}$ is stochastic [32] (i.e., each element of $\mathcal{T}$ is non-negative and each row sum is unity). Ergodicity of the symbolic dynamical system is equivalent to irreducibility of $\mathcal{T}$ [32], which implies that $\mathcal{T}$ has exactly one eigenvalue at unity (i.e., $\lambda = 1$) and that the rest of the eigenvalues are either on or within the unit circle with center at 0 (i.e., $|\lambda| \leqslant 1$). The (sum-normalized) eigenvector v_1 corresponding to the unity eigenvalue (i.e., $\lambda = 1$) represents the stationary state probability vector of the Markov chain [32]. In the probability space $(\Omega, \mathcal{E}, P)$ of STSA, Ω is the (finite) set of states $\mathcal{Q}$. With symbols $s \in \mathcal{A}$ occurring randomly, the state transition map $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ in Definition 19 becomes a random transformation $T : (\mathcal{Q}, \mathcal{E}, P) \to (\mathcal{Q}, \mathcal{E}, P)$ such that $T(q)$ yields a $\mathcal{Q}$-valued random variable for each $q \in \mathcal{Q}$. The state transition probability mass function $\kappa : \mathcal{Q} \times \mathcal{Q} \to [0, 1]$ satisfies the following condition:

$$P[T(q) \in E] \triangleq \sum_{\tilde{q} \in E} \kappa(q, \tilde{q}) \quad \forall q \in \mathcal{Q} \quad \forall \mathrm{E} \in \mathcal{E}$$

where P denotes the transition probability with respect to the underlying probability space $(\mathcal{Q}, \mathcal{E}, P)$, which implies that the random variable $T(q)$ has the probability mass function $\kappa(q, \cdot)$. Then, the $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition probability matrix $\mathcal{T}$ is a stochastic representation of the random transformation T. A simple example from a previous publication [25] is presented below to elucidate the concept of MPT.

Example 23. In the probability space $(\mathcal{Q}, \mathcal{E}, P)$, let $\mathcal{Q} = \{q_1, q_2\}$ with the σ-algebra $\mathcal{E} = 2^{\mathcal{Q}} \triangleq \{\emptyset, \{q_1\}, \{q_2\}, \mathcal{Q}\}$, and $P : \mathcal{E} \to [0, 1]$. Let $\left\{T^{(k)}\right\}$ be a sequence of mappings $(\mathcal{Q}, \mathcal{E}, P) \to (\mathcal{Q}, \mathcal{E}, P)$, such that the representation of $T^{(k)}$ by a $|\mathcal{Q}| \times |\mathcal{Q}|$ stochastic matrix (see Definition 22) is: $\mathcal{T}^{(k)} = \begin{bmatrix} p^{(k)} & 1 - p^{(k)} \\ 1 - \tilde{p}^{(k)} & \tilde{p}^{(k)} \end{bmatrix}$, where $p^{(k)}, \tilde{p}^{(k)} \in [0, 1)$ can be arbitrary for any given k. The eigenvalues of each stochastic (ergodic) matrix $\mathcal{T}^{(k)}$ are: $\lambda_0^{(k)} = 1$ and $\lambda_1^{(k)} = (p^{(k)} + \tilde{p}^{(k)} - 1)$; and the corresponding absolute-sum-normalized left eigenvectors are: $v_0^{(k)} = \left[\frac{1 - \tilde{p}^{(k)}}{2 - p^{(k)} - \tilde{p}^{(k)}} \quad \frac{1 - p^{(k)}}{2 - p^{(k)} - \tilde{p}^{(k)}}\right]$ that are k-variant, in general, and $v_1^{(k)} = [0.5 \;\; -0.5]$ that are k-invariant. Hence, $P\left[\left(T^{(k)}\right)^{-1} \mathcal{Q}\right] = P[\mathcal{Q}] = 1$ and $P\left[\left(T^{(k)}\right)^{-1} \emptyset\right] = P[\emptyset] = 0$. Furthermore, the preimage of $\{q_1\}$ (i.e., $\left(T^{(k)}\right)^{-1}\{q_1\}$) is either $\{q_1\}$ itself or $\{q_2\}$; similar results hold for $\{q_2\}$. Then, it follows that $P\left[\left(T^{(k)}\right)^{-1} E\right] = P[E] \; \forall E \in \mathcal{E}$ (i.e., the system is measure-preserving) if and only if $P[\{q_1\}] = P[\{q_2\}] = 0.5$, i.e., $\tilde{p}^{(k)} = p^{(k)}$ in the matrix $\mathcal{T}^{(k)}$ $\forall k$. Moreover, with the restriction $p^{(k)}, \tilde{p}^{(k)} \in [0,1)$, it also follows that $\left\{T^{(k)}\right\}$ is ergodic, because the only $T^{(k)}$-invariant measurable sets (i.e., members of $\mathcal{E}$) are $\emptyset$ and $\mathcal{Q}$. Then, it follows from Definitions 8 and 13 that, in this example, the sequence $\left\{T^{(k)}\right\}$ is measure-preserving and ergodic if and only if $\tilde{p}^{(k)} = p^{(k)}$ and $p^{(k)} \in [0,1)$ $\forall k$.

Following Definition 13 and Definition 22, a PFSA sequence $\{\mathcal{K}^{(n)}\} \triangleq \left\{\left(\mathcal{A}, \mathcal{Q}, \mathcal{T}^{(n)}\right)\right\}$ on a probability space $(\mathcal{Q}, \mathcal{E}, P)$ is measure-preserving and ergodic if $\mathcal{T}^{(n)}$ is measure-preserving and ergodic $\forall n \in \mathbb{N}$. A straightforward result, which is central to the current paper, follows from Theorem 15 for PFSA and is presented as the following corollary.

Corollary 24. *Let* $\left\{\mathcal{K}^{(n)}\right\}$ *be a sequence of measure-preserving and ergodic PFSA on a probability space* $(\mathcal{Q}, \mathcal{E}, P)$*. Then, the (sum-normalized) left eigenvector* $v_0^{(n)}$*, corresponding to the eigenvalue* $\lambda_0^{(n)} = 1$*, is uniformly distributed, i.e.,*

$$v_0^n = \left[\frac{1}{|\mathcal{Q}|}, \ldots, \frac{1}{|\mathcal{Q}|}\right] \quad \forall n$$

Remark 25. Following Remark 14, given any discrete-time realization $\{x_n\}$ of an ergodic process $\{X_n : X_n \in L_1(P)\}$, as $n \to \infty$, the time average $\frac{1}{n}\sum_{j=1}^{n} x_j$ converges P-ae and in $L_1(P)$ to the ensemble average $\int_\Omega X_n dP$. On the other hand, ergodicity of the corresponding symbolic system is much easier to check; it is equivalent to irreducibility of the state transition probability matrix. Following Corollary 24, both ergodicity and measure-invariance can be checked from the state transition probability matrix.

3.3. D-Markov machines

When constructing a D-Markov machine, the generation of the next symbol is assumed to depend only on a *finite* history of at most D consecutive symbols, i.e., a symbol block not exceeding the specified length D. In this context, a D-Markov machine [14,21] is defined as follows.

Definition 26. A D-Markov machine is a PFSA, in the sense of Definition 22, which generates symbols that solely depend on the (most recent) history of at most D consecutive symbols, where the positive integer D is called the *depth* of the machine. Equivalently, a D-Markov machine is a stochastic process $S = \ldots s_{-1}s_0s_1\ldots$, where the probability of occurrence of a new symbol depends only on the last consecutive (at most) D symbols, i.e.,

$$P[s_n | \ldots s_{n-D} \ldots s_{n-1}] = P[s_n | s_{n-D} \ldots s_{n-1}] \tag{5}$$

Consequently, for $w \in \mathcal{A}^D$ (see Definition 20), the equivalence class $\mathcal{A}^{\star}w$ of all (finite-length) words, whose suffix is w, is qualified to be a D-Markov state that is denoted as w.

The PFSA of a D-Markov machine generates symbol strings having the form $\{s_1s_2\ldots s_\ell\}$, where $s_j \in \mathcal{A}$ and ℓ is a positive integer. Both morph function π and state transition probability matrix $\mathcal{T}$ implicitly support the fact that *PFSA* satisfies the Markov condition, where generation of a symbol depends only on the current state that is a symbol string of at most D consecutive symbols [21]. For the proposed D-Markov-based methodology, there are four primary choices as enumerated below:

1. **Alphabet size ($|\mathcal{A}|$)**: Larger is the alphabet size more distinct are the different regimes, but more training data would be needed. A critical step of PFSA construction is selection of the alphabet size $|\mathcal{A}|$; while there are several algorithms for the choice of optimal alphabet size (e.g., [17,21]), the alphabet size has been chosen to be very small (e.g., $|\mathcal{A}| = 2$ and $|\mathcal{A}| = 3$) in this paper so that the number of states $|\mathcal{Q}|$ is also small to limit the required window length L. However, $|\mathcal{A}|$ could be larger for other applications.
2. **Partitioning Method**: While there are many data partitioning techniques, maximum entropy partitioning (MEP) [15,16,21] and K-means partitioning [33] have been chosen here as they are commonly used for partitioning of time series.
3. **Depth (D) in the D-Markov machine**: In this paper, $D = 1$ and $D = 2$ have been chosen for PFSA construction to limit the window length L. Higher values of D may lead to better results at the expense of increased computational time due to larger dimension of the feature space and need for more data [21].
4. **Choice of Feature**: The feature needs to be the one that adequately captures the texture of the signal and that is not computationally expensive.

3.4. Anomaly detection in the setting of a standard STSA

D-Markov machines provide an efficient and convenient way of modeling a dynamical system from the perspectives of discrete-time measurements and their discrete-space symbolization. In this setting, a time series $\{x_n\}$ is first symbolized using symbols from a finite-cardinality alphabet $\mathcal{A}$. The resulting symbol string $\{s_n\}$ is then used to construct a PFSA, which in turn generates low-dimensional feature vectors that are used for anomaly detection. The process is detailed in the following procedures:

1. *Choose* a nominal time series, for which the system is under a nominal operating condition.
2. *Construct* a partitioning for the nominal time series to generate a symbol string which is used to construct a nominal PFSA model. The state transition probability matrix of the generated PFSA model is constructed (e.g. by frequency counting [21]) which is used to produce a quasi-stationary probability vector μ^0 that represents the nominal pattern.
3. *Select* a new time series block up to the current time n and symbolize it using the learned nominal partition. This yields a new PFSA with a new quasi-stationary probability vector that represents the feature vector μ^n of the system at the time epoch n.
4. *Compute* the scalar-valued anomaly metric $\theta(n)$ at a time epoch n by having a string of divergences between the nominal feature μ^0 and current feature μ^m, and by sliding the block of data N times as:

$$\theta(n) = \frac{1}{N}\sum_{m=n}^{n+N-1} d_{KL}(\mu^0, \mu^m) \tag{6}$$

where $d_{KL}(\bullet, \bullet)$ is the Kullback–Leibler divergence [34] and the positive integer N serves the sole purpose of data smoothing.

4. Technical approach

Earlier publications [17,21] have reported identification of two afore-mentioned parameters of D-Markov machines, namely, Markov depth D and the alphabet size $|\mathcal{A}|$ for construction of PFSA. This paper proposes a methodology for identifying the remaining aforesaid parameter, namely, the minimum window length L, for symbolic analysis of observed time series, which is required to train Markov models generated from short-length windows of time series. In this framework, the underlying stochastic process is assumed to be an ergodic sequence of endomorphisms (i.e., surjective measure-preserving transformations (MPT)) [24]; and the developed model is a time-inhomogeneous Markov chain, in general, which is constructed by relying on the spectral properties of the underlying physical process.

Based on the above rationale, this paper presents a methodology for constructing PFSA from short-length time series, which generate non-homogeneous Markov chain models for describing the uncertain dynamics of the physical process. As a result, the time-invariance of the eigenvectors, which reflects measure-invariance of the dynamical system, is used to decide the window length of the time series required to construct PFSA. Unlike the standard STSA [14,21] that requires increasing the window length until the resulting PFSA is no longer significantly changing, the proposed MPT-based STSA selects a minimum window length for which absolute values of the eigenvectors are (nearly) uniformly distributed. The rationale is that anomalies are usually associated with a change in the system's total energy, which naturally makes the system no longer measure-invariant and thus the eigenvectors absolute values are no longer time-invariant. Along this line, a metric of variability for absolute values of the eigenvectors is proposed as a measure of anomaly. This paper also develops an MPT-based metric to quantify the divergence of evolving PFSA from the nominal PFSA; this information is then used for pattern classification and anomaly detection with low-delay tolerance.

Let a dynamical system be represented by a sequence of measure-preserving transformations $\{\mathcal{T}^n\}$ acting on the probability space generated by STSA (see Example 23 in Section 3.2). A major issue in the standard STSA-based anomaly detection is that it may require a long time series to construct the stationary state transition probability matrix $\mathcal{T}$, from which the stationary probability vector v is generated to serve as a feature vector. However, many real-life applications may require low-delay detection of anomalous events, for which short-length time series of measurements must be used. For example, thermoacoustic instabilities (TAI) in combustion systems typically develop within tens of milliseconds, which mandate an early and fast detection using short-length sensor time series [35,36].

A modification of the standard STSA technique (see Section 3.4) is now proposed as MPT-based STSA by relying on variations in the absolute values of eigenvectors. Table 1 lists major differences between a standard STSA [21] and the MPT-based STSA.

For MPT-based STSA, Algorithm 1 presents a procedure that provides: (i) identification of an optimal length of data windows for anomaly detection, and (ii) decision-making on whether the system is nominal or anomalous by using an optimal window length. The threshold parameter τ_1 is used to select an optimal widow length L. As a result, Algorithm 1 provides a method for detecting anomalous patterns using short-length time series by constructing a D-Markov machine model of the underlying stochastic process, given the alphabet size $|\mathcal{A}|$ and the Markov depth D. Thus, an anomaly is detected online when the metric ρ, which represents ergodicity and measure-preservation, exceeds the threshold parameter τ_2. Fig. 1 presents a flowchart of Algorithm 1.

Table 1
Comparison of standard STSA and MPT-based STSA.

	Standard STSA	MPT-based STSA
1	Time-invariance of the nominal-phase PFSA	Time-invariance of absolute values of left eigenvectors of the nominal-phase PFSA, which are time-varying, in general
2	Generation of homogeneous Markov-chain models	Generatation of non–homogeneous Markov-chain models
3	Anomaly quantification by divergence of the current PFSA from the nominal PFSA	Anomaly quantification by variability of left eigenvectors of the evolving PFSA
4	Requirement of relatively long time series for modeling the underlying process dynamics	Requirement of relatively short time series for modeling the underlying process dynamics
5	Less robust to parametric changes in data partitioning and detection system (e.g., alphabet size $\lvert\mathcal{A}\rvert$ and Markov depth D)	More robust to parametric changes in data partitioning and detection system (e.g., alphabet size $\lvert\mathcal{A}\rvert$ and Markov depth D)

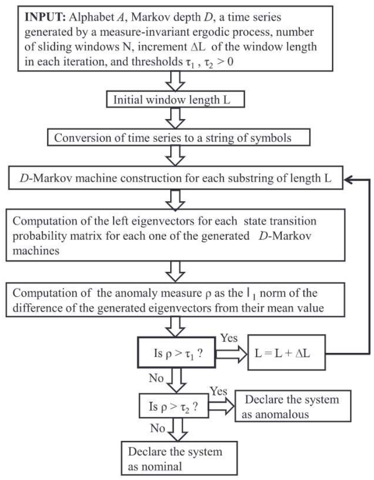


Fig. 1. Flowchart for Algorithm 1.

Algorithm 1: MPT-Based Anomaly Detection

INPUT: Alphabet $\mathcal{A}$, Markov depth D, a training time series $\{x_n^{trn}\}$ generated in the nominal phase, and a family of test time series $\{x_{n+1:n+L}^{tst}\}$, number of sliding windows $N \in \mathbb{N}$, increment of the window length ΔL in each iteration in the training phase, and two thresholds $\tau_1, \tau_2 > 0$.
OUTPUT: A decision on whether the system is nominal or anomalous.
1: Choose an initial window size L.
2: Convert time series blocks $\{x_{n+1:n+L}^{trn}\}$, $n = 0, 1, \ldots, N-1$, into symbol strings $\{s_{n+1:n+L}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
3: Using frequency counting, construct a D-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\left\{\mathcal{T}^{(n)}\right\}$.
4: Find the left eigenvectors $\{v(n)\}$, corresponding to eigenvalue $\lambda_0 = 1$, for each one of the state transition probability matrices.
5: $\overline{v} \leftarrow \frac{1}{N}\sum_{n=0}^{N-1}|v(n)|$, where $|v(n)|$ is the vector of absolute values of components of $v(n)$.
6: $\theta \leftarrow \frac{1}{N}\sum_{n=0}^{N-1}|(|v(n)| - \overline{v})|$.
7: $\rho \leftarrow ||\theta||_{\ell_1}$.

Algorithm 1 (*Continued*)

8: **if** $\rho > \tau_1$
9: $L \leftarrow L + \Delta L$ and GOTO 2.
10: **else**
11: Repeat Steps 2 to 7 by replacing the training time series $\{x^{trn}_{n:n+L-1}\}$ with the test time series $\{x^{tst}_{n:n+L-1}\}$.
12: **if** $\rho > \tau_2$
13: declare the system as anomalous.
14: **else**
15: declare the system as nominal.
16: **end if**
17: **end if**

Remark 27. The threshold parameters τ_1 and τ_2 are inter-dependent in the sense that a smaller τ_1 would result in a larger window length L (i.e., a larger delay), which would require a smaller τ_2 under a fixed maximum allowable false positive rate (FPR) for similar quality of decisions on anomaly detection. It is noted that both parameters, namely, tolerated delay and maximum allowable FPR, are generally application-dependent. In this regard, Pareto optimization for selection of the (user-set) threshold parameters τ_1 and τ_2 is recommended as a topic of future research in Section 6.

Remark 28. Algorithm 1 utilizes the left eigenvector corresponding to the unique eigenvalue $\lambda_0 = 1$. However, depending on the application, other eigenvectors could also be information-bearing for discrimination of anomalous behavior from the nominal.

5. Experimental validation: results & discussion

This section validates the MPT-based STSA methodology (see Section 4) on experimental data, generated from a laboratory apparatus for predicting thermoacoustic instabilities (TAI) in combustion systems. The objective here is to evaluate the performance of the proposed MPT-based STSA, presented in Algorithm 1, for anomaly detection using short-length time series and low-dimensional feature vectors, and its robustness to changes in the parameters of data partitioning and the algorithm. In this regard, performance of MPT-based STSA is compared with those of HMM (e.g., [5]) and CUSUM [7] on the same sets of experimental data.

5.1. The laboratory apparatus and experimental data

Fig. 2 depicts the experimental apparatus for emulation of thermoacoustic instabilities (TAI) on an electrically heated Rijke tube apparatus [37], which has been widely used in the study of TAI, because electrical heating is easier to operate in the laboratory environment compared to a fuel burning-based combustor, and yet it can emulate the salient properties of TAI in real-life combustors [38]. The Rijke tube apparatus in Fig. 2 consists of a 1.5 m long horizontal tube with an external cross-section of 10 cm $\times$ 10 cm with a wall thickness of 6.5 mm. It is equipped with an air-flow controller that regulates the flow of air at atmospheric pressure through the tube. It has a heating element placed at quarter length of the tube from the air-input end. A programmable DC power supply controls the power input to the heater [37]. The experiments were conducted in the following manner:

1. For every run, the air flow-rate was set at a constant value. Different runs were performed with flow-rates ranging from 130 liters per minute (LPM) to 250 LPM at increments of 20 LPM.
2. First the Rijke tube system was heated to a steady state with a primary heater power input of $\sim$200 W.
3. Then the power input was abruptly increased to a higher value that shows limit cycle behavior as depicted in the stability chart reported by Mondal et al. [37].

Fifteen (15) experiments have been conducted on the Rijke tube, where the process starts with stable behavior, gradually becoming anomalous and eventually unstable. For each experiment, a time series of pressure oscillations (sampled at 8192 Hz) was collected over 30 s, followed by a filtering process in order to mitigate the effects of environmental acoustics. Two samples of pressure time series are presented in Fig. 3.

When the signal space is partitioned in the stable part, the resulting symbol sequences can be described by a (generally) time-varying, measure-preserving and ergodic PFSA, where there is no significant energy change in the system, and each

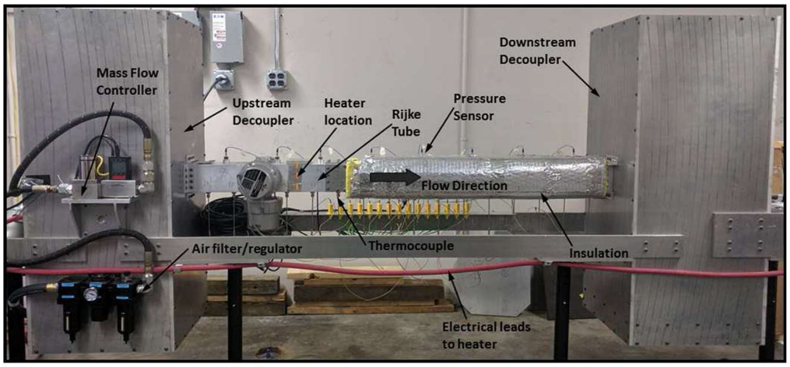


Fig. 2. The electrically heated Rijke tube apparatus.

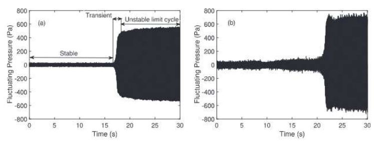


Fig. 3. Unsteady pressure signals showing the transition from stable (nominal) to unstable limit cycle (anomalous) behavior: (a) input power abruptly increased to 1800 W with air flow rate 210 LPM; (b) input power abruptly increased to 2000 W with air flow rate 250 LPM.

state is expected to be revisited infinitely many times, if the system (hypothetically) stays ergodic (i.e., stable) forever. Once the transient phase starts, more energy is added to the system, and some of the PFSA states are no longer revisited many times. Therefore, in the transient phase, both measure-preservation and ergodicity properties are presumed to be lost.

5.2. Results of experimental data analysis

An ensemble of pressure time series have been generated for the nominal (stable) operation and for the anomalous (transient to unstable) operation. The left eigenvector, v_1, of $\mathcal{T}$, corresponding to the eigenvalue $\lambda_0 = 1$, serves as the feature vector. The effects of different values of window length L are investigated in the plates of Fig. 4, while the other two parameters are set at $|\mathcal{A}| = 2$ and $D = 1$. Therefore, the number of D-Markov states is $|\mathcal{Q}| = 2$ (see Section 3); consequently, the (sum-normalized) eigenvector is 2-dimensional in the form of $v_1(n) = [v_{1,1}(n) \quad (1 - v_{1,1}(n))]$ and only the (scalar) values of $v_{1,1}(n)$ are plotted.[3] Fig. 4 shows how the absolute value of the left eigenvector (corresponding to $\lambda_0 = 1$) converges to a nearly constant vector [0.5 0.5] for the stable phase as the window size L is increased, while characteristics of the transient zone remain time-varying. In view of Corollary 24 in Section 3.2, this observation supports the postulation that the stable phase is considered to be ergodic and measure-preserving while the transient phase is not. The unstable zones are ignored in Fig. 4, because the main objective here is to detect the onset of transience from the stable condition.

Fig. 5 shows the convergent behavior of the absolute value of the left eigenvector (corresponding to $\lambda_0 = 1$) in the stable phase with $|\mathcal{A}| = 3, D = 1$ (i.e., $|\mathcal{Q}| = 3$), and different window length L's. It follows from Fig. 5 that the absolute value of the left eigenvectors converge to a nearly constant vector $[1/3 \;\; 1/3 \;\; 1/3]$ as the window length L is increased, which is supported by Corollary 24 in Section 3.2.

[3] The eigenvectors are absolute-sum-normalized, i.e., for a $|\mathcal{Q}|$-dimensional eigenvector, only ($|\mathcal{Q}|$-1) elements need to be specified.

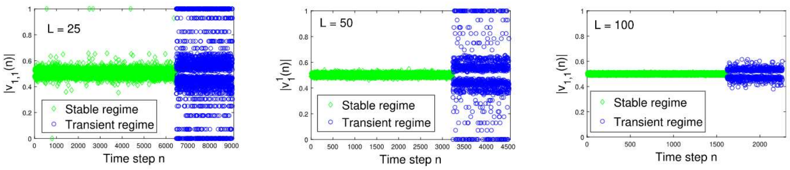


Fig. 4. For the MPT-based STSA, convergence of the left eigenvector $v_1 = [v_{1,1}\ (1 - v_{1,1})]$, corresponding to the eigenvalue $(\lambda_0 = 1)$, to an approximately constant function in the stable combustion phase, and to an oscillating function for transience to the unstable phase $(|\mathcal{A}| = 2, D = 1)$.

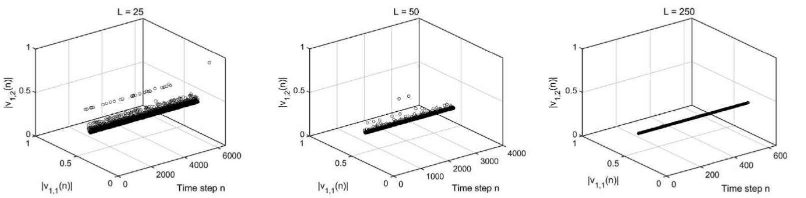


Fig. 5. For MPT-based STSA, convergence of the left eigenvector $v_1 = [v_{1,1}\ v_{1,2}\ (1 - v_{1,1} - v_{1,2})]$, corresponding to the eigenvalue $\lambda_0 = 1$, to the uniform distribution $[1/3\ 1/3\ 1/3]$ in the stable combustion phase $(|\mathcal{A}| = 3, D = 1)$.

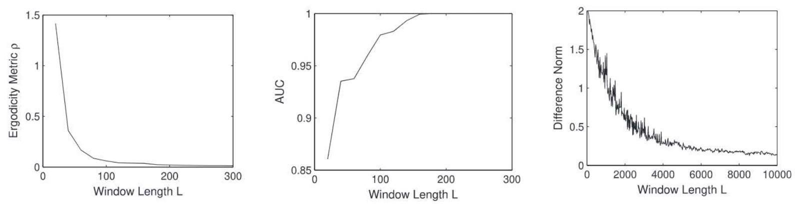


Fig. 6. Comparison of convergence of MPT-based STSA (using left eigenvectors corresponding to the eigenvalue $\lambda_0 = 1$) to that of standard STSA (*K*-means has been used for partitioning in both methods with $|\mathcal{A}| = 2$ and $D = 1$): (a) convergence of the ergodicity metric ρ as *L* increases for a typical time series; (b) increase of area under curve (AUC) of the ROC, using Algorithm 1 for the same time series, as *L* increases; and (c) convergence of the difference norm of consecutive state transition probability matrices for the same time series.

Algorithm 1 is validated in Fig. 6, where the ergodicity metric ρ is evaluated for different values of *L* for the pressure time series from a typical experiment, and the area under the curve (AUC) of the ROC curve is obtained by using Algorithm 1 for the same experiment and the same values of *L*, with parameters $|\mathcal{A}| = 2, D = 1$, and *K*-means partitioning.[4]

As seen in plates (a) and (b) of Fig. 6, the profile of ROC performance (e.g., AUC) is strongly correlated to that of the ergodicity metric ρ in Algorithm 1; it is concluded that smaller is the value of ρ, more accurate is the *D*-Markov model of the stochastic process and hence better ROC performance is achieved by the *D*-Markov model. This conclusion demonstrates the efficacy of Algorithm 1 for selection of the window length *L* based on the value of the ergodicity metric ρ. In practice, one may use a number of time series from several experiments to construct a family of plots. Then, the threshold parameter τ_1 in Algorithm 1 can be identified by specifying the desired performance in terms of the areas under the ROC curves. That is, τ_1 can be chosen such that the corresponding area under the ROC curve for each experiment (alternatively, the average area of the ROC curves) is equal to or greater than the desired AUC.

[4] It is noted that *K*-means is an iterative clustering technique that relies on minimization of the clusters' distortion [33]. There are two main hyperparameters in *K*-means, namely, number of clusters and initial locations of the clusters' centroids. In this paper, number of clusters = $|\mathcal{A}|$, which is specified in the sequel for each one of the plots. The other hyperparameter is initial locations of the clusters centroids that are determined based on *K*-means++ algorithm [39], which is a randomized algorithm for improving the performance and computation time of the conventional *K*-means [33].

A comparison of the two plates (a) and (c) in Fig. 6 reveals the rate of convergence of the ergodicity metric ρ, proposed in Algorithm 1, relative to that of the state transition probability matrices $\left\{\mathcal{T}^{(n)}\right\}$, when the window length L is monotonically increased. This comparison shows that the convergence of ρ is much faster than that of the norm of differences in consecutive state transition probability matrices $\left\{\mathcal{T}^{(n)}\right\}$, and the profile of ρ is also more smooth. In particular, plates (a) and (b) suggest that if we want to use MPT-STSA for this typical experiment, we need a time series observation window of length $L \sim 160$ in order to achieve $AUC \sim 1$. However, plate (c) suggests that we need an observation window of length $L \sim 10,000$ in order to achieve a comparable performance, i.e. $AUC \sim 1$, if we want to use the standard STSA. Notice the large delay reduction we gain by using MPT-based STSA over the standard STSA. In summary, Fig. 6 suggests the following:

> The window length L required for constructing a time-inhomogeneous Markov model by relying on MPT-based STSA is much smaller than that required for a comparably accurate time-homogeneous Markov model by using a standard STSA with a stationary state transition probability matrix [14].

Figs. 7 and 8 exhibit a family of ROC curves generated by Algorithm 1 with K-means and MEP, respectively. The individual plates in these two figures show performance comparison of MPT-based STSA, HMM, and CUSUM for different combinations of alphabet size $|\mathcal{A}|$, Markov depth D, and window length L. Since the results of standard STSA are significantly inferior to those of both HMM and CUSUM due to short window lengths L (and small values of $|\mathcal{A}|$), the plots of standard STSA are not included in Figs. 7 and 8.

Before delving into the details of comparison in Figs. 7 and 8, we very briefly present how the two well-known detection methods, HMM and CUSUM, have been used to generate the results that serve as benchmarks.

- *Hidden Markov Model (HMM)*: A standard version of HMM [33] has been used in this paper by computing the joint likelihood of L observations, where L is the window length, conditioned on a *null* HMM trained on the stable phase. Then, N such joint likelihoods are computed by sliding the window N times, whose average serves as the anomaly measure, as detailed by Miller et al. [6].
- *Cumulative Sum (CUSUM)*: A standardized version of CUSUM [7] returns the first index of the time series $\{x_k\}$ that has drifted k standard deviations away from the nominal mean; this detection approach is suitable for a deviation in the mean only. As seen in Fig. 3, since the data under consideration show changes in variance, a slightly modified version [6] of CUSUM has been adopted here to account for deviations in variance. The *stopping rule* of the modified CUSUM algorithm follows:

$$\sigma_0^2 = \frac{1}{N_0 - 1}\sum_{k=1}^{N_0} x_k^2; \text{ and } \rho(n) = \sum_{k=1}^{n} x_k^2 - n\sigma_0^2 \tag{7}$$

where N_0 is the number of samples taken from the nominal state with the data shifted to have zero mean so that σ_0^2 is an unbiased variance estimate. Given the window length L, the measure $\rho(n)$ is computed upon a data window $x_{n:n+L-1}$; this window is then slid N times to generate N values for ρ. Based on the averaged measure ρ_{av}, anomaly decisions are made by using the following stopping rule:

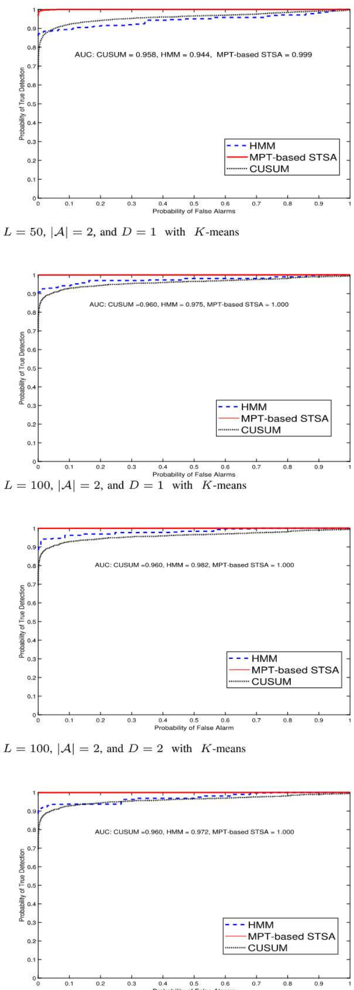
$$N_{stop} = \inf\{n : |\rho_{av}(n)| \geqslant \delta\} \tag{8}$$

where N_{stop} is the *stopping time* at which a change detection is declared, and δ is a threshold that is set to achieve a specified false positive rate (FPR). Since we use ROC for computation of detection performance, the threshold δ is adjusted [27] in the range $(0, \infty)$ to generate the ROC curves for CUSUM. Further details on how the ROC curves are constructed and used to choose detection thresholds are provided later in Section 5.3.

It is noted that each of the three methods, HMM, CUSUM, and MPT-Based STSA, is a function of both L and N, whose values influence the delay tolerated at the given application. In Figs. 7 and 8, we consider $L = 50$ and $L = 100$, while $N = 30$ is kept fixed for all cases therein.

Referring back to the subject of performance comparison, it is seen in Figs. 7 and 8 that the performance of all three detection methods generally improve as the window length L is increased; CUSUM has slightly better performance compared to HMM for $L = 50$, while HMM achieves slightly better performance for $L = 100$. In summary, a comparison of Figs. 7 and 8 consistently shows that:

1. Both HMM and CUSUM are outperformed by MPT-based STSA in all cases.
2. MPT-based STSA achieves excellent performance for both K-means and MEP, which demonstrates that the algorithm of MPT-based STSA is robust to an appropriate choice of time series partitioning for symbolization.



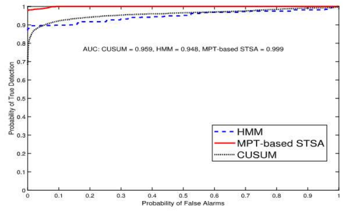
(a) $L = 50$, $|\mathcal{A}| = 2$, and $D = 1$ with K-means

(b) $L = 100$, $|\mathcal{A}| = 2$, and $D = 1$ with K-means

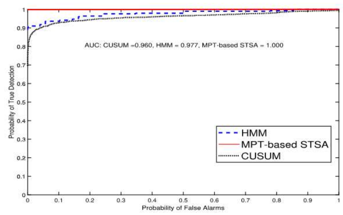
(c) $L = 100$, $|\mathcal{A}| = 2$, and $D = 2$ with K-means

(d) $L = 100$, $|\mathcal{A}| = 3$, and $D = 1$ with K-means

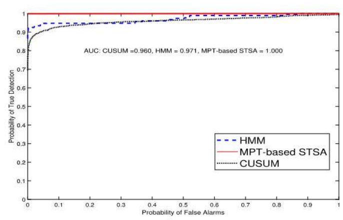
Fig. 7. Comparison of ROC performance of HMM and CUSUM with that of MPT-based STSA (having *K*-means partitioning) by using eigenvectors corresponding to the eigenvalue $\lambda_0 = 1$.



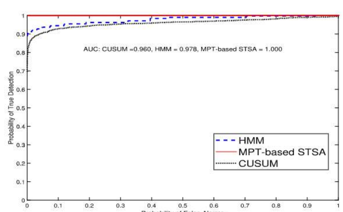
(a) $L = 50$, $|\mathcal{A}| = 2$, and $D = 1$ with MEP



(b) $L = 100$, $|\mathcal{A}| = 2$, and $D = 1$ with MEP



(c) $L = 100$, $|\mathcal{A}| = 2$, and $D = 2$ with MEP



(d) $L = 100$, $|\mathcal{A}| = 3$, and $D = 1$ with MEP

Fig. 8. Comparison of ROC performance of HMM and CUSUM with that of MPT-based STSA (having maximum entropy partitioning (MEP)) by using eigenvectors corresponding to the eigenvalue $\lambda_0 = 1$.

Table 2
Statistics of CPU execution time (sec) per window for online anomaly detection.

Parameters & partitioning method	CUSUM Mean	CUSUM Std dev	MPT-based STSA Mean	MPT-based STSA Std dev	HMM Mean	HMM Std dev
$\lvert\mathcal{A}\rvert = 2,\ D = 1,\ L = 50$, MEP	6.127e−06	7.834e−07	1.737e−04	1.316e−05	2.625e−03	2.104e−05
$\lvert\mathcal{A}\rvert = 2,\ D = 1,\ L = 100$, MEP	9.883e−06	6.615e−06	3.001e−04	2.221e−05	4.266e−03	3.391e−05
$\lvert\mathcal{A}\rvert = 2,\ D = 2,\ L = 100$, MEP	9.883e−06	6.615e−06	6.441e−04	4.727e−05	4.778e−03	3.803e−05
$\lvert\mathcal{A}\rvert = 3,\ D = 1,\ L = 100$, MEP	9.883e−06	6.615e−06	5.723e−04	4.119e−05	4.570e−03	3.487e−05
$\lvert\mathcal{A}\rvert = 2,\ D = 1,\ L = 50, K$-means	6.127e−07	7.834e−07	4.379e−04	3.240e−05	2.666e−03	2.138e−05
$\lvert\mathcal{A}\rvert = 2,\ D = 1,\ L = 100, K$-means	9.883e−06	6.615e−06	5.662e−04	4.174e−05	4.275e−03	3.378e−05
$\lvert\mathcal{A}\rvert = 2,\ D = 2,\ L = 100, K$-means	9.883e−06	6.615e−06	9.152e−04	6.744e−05	4.689e−03	4.013e−05
$\lvert\mathcal{A}\rvert = 3,\ D = 1,\ L = 100, K$-means	9.883e−06	6.615e−06	8.300e−04	6.004e−05	4.471e−03	3.168e−05

5.3. Threshold parameters τ_1 and τ_2 in Algorithm 1

This subsection suggests a procedure for selecting the threshold parameters τ_1 and τ_2 in Algorithm 1 from the receiver operator characteristic (ROC) curves. Rigorous statistical analysis in this direction is suggested as a topic of future research in Section 6.

The ROC curves in Figs. 7 and 8 have been generated via Algorithm 1 by varying the threshold parameter τ_2 [27]. Each point in the ROC curve corresponds to a specific value of τ_2. As a result, the design parameter τ_2 can be chosen from the ROC curve by fixing a maximum false positive rate (FPR); however, this assertion is typically application-dependent. For example, in applications where the cost of a positive false alarm is relatively low, the maximum FPR could be increased. In contrast, if the cost of a positive false alarm is high, the maximum allowable FPR should be decreased. The value of the threshold τ_1 in Algorithm 1 can be decided by specifying a maximum delay determined by the window length L.

5.4. Computational time for algorithm execution

Table 2 lists the second-order statistics of CPU execution time[5] of CUSUM, MPT-based STSA, and HMM for anomaly detection using the same ensemble of time series data. It is seen that, on the average, the execution time of CUSUM is approximately two orders of magnitude less than that of MPT-based STSA. In contrast, MPT-based STSA is about one order of magnitude faster than HMM especially if MEP is used for partitioning instead of *K*-means in the MPT-based STSA; the rationale is that K-means is iterative while MEP is not.

Table 2 also shows that as the window length L is increased, standard deviation of the CPU execution time tends to increase initially and eventually stabilizes. The rationale is that, as L is increased, the number of total windows for the same time series would decrease resulting in a higher standard deviation. Referring to the detection performance in the ROC curves of Figs. 7 and 8, it s seen that each of HMM and CUSUM would require much longer window lengths (L) of time series to achieve similar performance in comparison to MPT-based STSA. This implies that MPT-based STSA would be subjected to much less time delay, compared to HMM and CUSUM, to acquire the requisite data for online anomaly detection.

6. Summary, conclusion and future work

This paper has presented and validated a symbolic time series analysis (STSA) methodology for pattern classification and online anomaly detection in uncertain dynamical systems, where the underlying concept is built upon spectral analysis of ergodic sequence of measure-preserving transformations. Rather than constructing a homogeneous (i.e., time-invariant) Markov-chain model, which may require a longer window of time series, a non–homogeneous Markov-chain model is constructed, for which consecutive stochastic matrices might be different but they share nearly the same (absolute) values of the eigenvectors' components. Based on this information, Markov-chain models are constructed using short windows of time series to facilitate online anomaly detection. The proposed methodology of pattern classification and anomaly detection, called MPT-based STSA, has been validated by comparison with well-known methods of hidden Markov model (HMM) [5] and cumulative sum (CUSUM) [7,40] on experimental data generated from a laboratory apparatus for detection of thermoacoustic instabilities (TAI) in combustion systems. In this application, MPT-based STSA is consistently seen to be superior to both HMM and CUSUM.

While there are many areas of theoretical and experimental research to enhance the work reported in this paper, the authors suggest the following topics for future research:

[5] The experimental results were generated on a DELL PRECISION T3400 with an Intel (R) Core (TM)2 Quad CPU Q9550 at 2.83 GHz, with 8 GB RAM, running under Windows 10.

1. Rigorous statistical analysis [41] and Pareto optimization [42] for selection of the threshold parameters τ_1 and τ_2 in Algorithm 1.
2. Investigation on usage of a commutator operator on ergodic matrices instead of the norm of eigenvector difference for enhancement of computational efficiency.
3. Experimental validation of MPT-based STSA in diverse industrial applications for pattern classification and anomaly detection.

CRediT authorship contribution statement

Najah F. Ghalyan: Conceptualization, Methodology, Software, Validation, Writing - original draft. **Asok Ray:** Supervision, Formal analysis, Writing - review & editing.

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

The authors are grateful to Dr. Sudeepta Mondal who has kindly provided the experimental data on the Rijke tube apparatus. The first author would like to thank the Higher Committee for Education Development (HCED) in Iraq for their financial support. The work reported in this paper has been supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grant Nos. FA9550-15-1-0400 and FA9550-18-1-0135 in the area of dynamic data-driven application systems (DDDAS). Any opinions, findings, and conclusions in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

References

[1] A. Shiryaev, On optimum methods in quickest detection problems, Theory Probab. Appl. 8 (1) (1963) 22–46.
[2] A. Tartakovsky, V. Veeravalli, General asymptotic Bayesian theory of quickest change detection, Theory Probab. Appl. 49 (3) (2005) 458–497.
[3] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes – Theory and Application, Prentice Hall Inc., 1993.
[4] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[5] B. Hajek, Random Processes for Engineers, Cambridge University Press, Cambridge, UK, 2015.
[6] D.J. Miller, N.F. Ghalyan, S. Mondal, A. Ray, HMM conditional-likelihood based change detection with strict delay tolerance, Mech. Syst. Sig. Process. 147 (2021) 107109 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327020304957.
[7] E.S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.
[8] C.-M. Kuan, K. Hornik, The generalized fluctuation test: a unifying view, Econom. Rev. 14 (2) (1995) 135–161.
[9] S. Mishra, O.A. Vanli, C. Park, A multivariate cumulative sum method for continuous damage monitoring with lamb-wave sensors, Int. J. Progn. Health Manage. 6 (2015) 1–11.
[10] J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, Econometrica 66 (1) (1998) 47–78.
[11] J. Bai, P. Perron, Computation and analysis of multiple structural change models, J. Appl. Econom. 18 (1) (2003) 1–22.
[12] C.S. Daw, C.E.A. Finney, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. 74 (2003) 915–930.
[13] P. Beim Graben, Estimating and improving the signal-to-noise ratio of time series by symbolic dynamics, Phys. Rev. E 64 (5) (2001) 051104.
[14] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Sig. Process. 84 (7) (2004) 1115–1130.
[15] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Sig. Process. 86 (11) (2006) 3309–3320.
[16] A. Subbu, A. Ray, Space partitioning via hilbert transform for symbolic time series analysis, Appl. Phys. Lett. 92 (8) (2008) 084107.
[17] S. Sarkar, P. Chattopadhyay, A. Ray, Symbolization of dynamic data-driven systems for signal representation, Sig. Image Video Process. 10 (8) (2016) 1535–1542.
[18] N.F. Ghalyan, D.J. Miller, A. Ray, A locally optimal algorithm for estimating a generating partition from an observed time series and its application to anomaly detection, Neural Comput. 30 (9) (2018) 2500–2529.
[19] P. Dupont, F. Denis, Y. Esposito, Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms, Pattern Recogn. 38 (9) (2005) 1349–1371.
[20] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, R. Carrasco, Probabilistic finite-state machines – Part I and Part II, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1013–1039.
[21] K. Mukherjee, A. Ray, State splitting and merging in probabilistic finite state automata for signal representation and analysis, Sig. Process. 104 (2014) 105–119.
[22] S. Gupta, A. Ray, Symbolic dynamics filtering for data-driven pattern recognition, Pattern Recogn. Theory Appl. (2007) 17–71.
[23] P. Halmos, Lectures on Ergodic Theory, Dover Books on Mathematics, Dover Publications, 2017 [Online]. Available: https://books.google.com/books?id=E3BADwAAQBAJ.
[24] I. Cornfeld, S. Fomin, Y. Sinai, Ergodic Theory, Springer-Verlag, New York, NY, USA, 1982.
[25] N.F. Ghalyan, A. Ray, Symbolic time series analysis for anomaly detection in measure-invariant ergodic systems, J. Dyn. Syst., Measure., Control 142 (2020) 061003.
[26] T. Lieuwen, V. Yang, Combustion instabilities in gas turbine engines: operational experience, fundamental mechanisms, and modeling, Am. Inst. Aeronaut. Astronaut. ch. 1 (2005) 3–26.
[27] H. Poor, An Introduction to Signal Detection and Estimation, Spriger-Verlag, New York, NY, USA, 1994.
[28] H.B. Keynes, J.B. Robertson, Eigenvalue theorems in topological transformation groups, Trans. Am. Math. Soc. 139 (1969) 359–369.
[29] R. Durrett, Probability: Theory and Examples, fourth ed., Cambridge Series in Statistical and Probabilistic Mathematics, 2010. [Online]. Available: https://books.google.com/books?id=evbGTPhuvSoC.
[30] J. Munkres, Topology, 2nd ed., Prentice-Hall, Upper Saddle River, NJ, USA, 2000.
[31] C. Beck, F. Schlögl, Thermodynamics of Chaotic Systems: An Introduction, Cambridge University Press, Cambridge, U.K., 1993.
[32] A. Berman, R. Plemmons, Nonnegative Matrices in the Mathematical Sciences, SIAM Press, Philadelphia, PA, USA, 1994.

[33] K. Murphy, Machine Learning: A Probabilistic Perspective, 1st ed., The MIT Press, 2012.
[34] T. Cover, J. Thomas, Elements of Information Theory, Wiley-Interscience, Hoboken, NJ, USA, 2006.
[35] N.F. Ghalyan, S. Mondal, D.J. Miller, A. Ray, Hidden Markov modeling-based decision-making using short-length sensor time series, ASME J. Dyn. Syst. Meas. Control 141 (10) (2019) 104502.
[36] D. Jha, N. Virani, J. Reimann, A. Srivastav, A. Ray, Symbolic analysis-based reduced order markov modeling of time series data, Sig. Process. 149 (2018) 68–81.
[37] S. Mondal, N.F. Ghalyan, A. Ray, A. Mukhopadhyay, Early detection of thermoacoustic instabilities using hidden markov models, Combust. Sci. Technol. 191 (8) (August 2019) 1309–1336.
[38] N. Jamieson, G. Rigas, M. Juniper, Experimental sensitivity analysis via a secondary heat source in an oscillating thermoacoustic system, Int. J. Spray Combust. Dyn. 9 (4) (2017) 230–240.
[39] D. Arthur, S. Vassilvitskii, K-means++: the advantages of careful seeding, in: Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, 2007.
[40] V. Kireichikov, V. Mangushev, I. Nikiforov, Investigation and application of cusum algorithms to monitoring of sensors, in: Statistical Problems of Control (in Russian), 1990, pp. 124–130..
[41] D. Pastor, Q.-T. Nguyen, Random distortion testing and optimality of thresholding tests, IEEE Trans. Sig. Process. 61 (16) (2013) 4161–4171.
[42] K. Miettinen, Nonlinear Multiobjective Optimization, Kluwer Academic, Boston, MA, USA, 1998.