Symbolic Time Series Analysis for Anomaly Detection in Measure-Invariant Ergodic Systems

Najah F. Ghalyan
Department of Mechanical Engineering,
The Pennsylvania State University,
University Park, PA 16802;
Department of Mechanical Engineering,
The University of Kerbala,
Kerbala 56001, Iraq
e-mail: nfjasim76@gmail.com

Asok Ray[1]
Fellow ASME
Department of Mechanical Engineering;
Department of Mathematics,
The Pennsylvania State University,
University Park, PA 16802
e-mail: axr2@psu.edu

This paper presents a novel framework of symbolic time series analysis (STSA) for anomaly detection in dynamical systems. The core concept is built upon a property of measure-preserving transformation (MPT) sequence, acting on a probability space with ergodic measure, that the eigenfunctions of these transformations would be time-invariant. As a result, unlike a standard STSA that is required to generate time-homogeneous Markov chains, the proposed MPT-based STSA is allowed to have time-inhomogeneous Markov chains, where the (possibly time-varying) state transition probability matrices have time-invariant eigenvectors. Such a time-invariance facilitates analysis of the dynamical system by using short-length time series of measurements. This is particularly important in applications, where the underlying dynamics and process anomalies need fast monitoring and control actions in order to mitigate any potential structural damage and/or to avoid catastrophic failures. The MPT-based STSA has been applied for low-delay detection of fatigue damage, which is a common source of failures in mechanical structures and which is known to have uncertain dynamical characteristics. The underlying algorithm has been validated with experimental data generated from a laboratory apparatus that uses ultrasonic sensors to detect fatigue damage in polycrystalline–alloy specimens. The performance of the proposed MPT-based STSA is evaluated by comparison with those of a standard STSA and a hidden Markov model (HMM) on the same experimental data. The results consistently show superior performance of the MPT-based STSA. [DOI: 10.1115/1.4046156]

Keywords: anomaly detection, ergodicity, fatigue damage, measure invariance, spectral analysis, symbolic dynamics

1 Introduction

Data-driven detection of anomalous behavior [1] is important in diverse engineering applications (e.g., prediction of fatigue failures, condition-based maintenance, and radar systems) as well as for malware detection (e.g., Ref. [2] and references therein). One of the commonly used methods for anomaly detection is the *cumulative sum* technique, developed by Page [3], which has been widely used for the detection of change points [4]. This tool is very efficient for detecting changes in the time series, which may occur as a result of abrupt variations in the underlying model structure. However, changes in the time series may often happen due to gradual degradation in the underlying dynamical system. Detection of such changes is more challenging than those due to abrupt variations.

Hidden Markov models (HMMs) [5] have been widely used for both change point and anomaly detection in diverse applications such as electronic systems [6], bio-informatics [7], target detection [8], brain imaging [9], early detection of thermo-acoustic instabilities in combustion systems [10,11], and detection of fatigue damage in structural materials [12]. In this setting, the (nominal) HMM is trained by using observed time series that is known to represent the nominal behavior. Then, if a change occurs in the time series, the likelihood of the new observed subsequence under the nominal HMM is expected to deviate significantly from the nominal likelihood [11].

In a similar context, the concept of *symbolic time series analysis* (STSA) has been used by many researchers (e.g., Refs. [13–15]) for constructing Markov chain models from the observed time series. In the STSA framework, a (finite length) time series is partitioned for conversion into a string of symbols from a (finite cardinality) alphabet of symbols (e.g., Refs. [16–20]). Subsequently, a PFSA (is constructed from the symbol string (e.g., Refs. [21–23]), in which the probability distribution of the emitted symbols depends upon the immediately preceding D symbols, where the Markov depth D is a positive integer. Such a PFSA is called a *D-Markov machine*, which has found diverse applications in pattern recognition and anomaly detection (e.g., Refs. [15,16,23–25]). The main distinction between HMMs and D-Markov machines is that the state transition in HMMs could have a nondeterministic algebraic structure [22], which requires an iterative method (e.g., the Baum-Welch algorithm [5]) for training the HMM parameters; in this scenario, the algorithm might lead to a poor local optimum. In contrast, D-Markov machines have a deterministic algebraic structure [23], which makes computation much simpler and less prone to local optimality issues.

In the STSA setting, the selection of the window length of time series to construct the PFSA largely depends on several parameters (e.g., Markov depth and alphabet size) and the nature of the particular underlying process that generates the time series [23]. 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 window lengths. Under the assumption of statistical stationarity [15], the (possibly time-varying) state transition probability matrix converges to a constant matrix when the window length may become arbitrarily large [15]. Thus, one may choose a minimum window length at which this matrix tends to

[1]Corresponding author.

Contributed by the Dynamic Systems Division of ASME for publication in the Journal of Dynamic Systems, Measurement, and Control. Manuscript received July 30, 2019; final manuscript received January 26, 2020; published online March 3, 2020. Assoc. Editor: Suman Chakravorty.

be approximately time-invariant. The resulting model in this case would be a time-homogeneous Markov chain [26]. However, this scenario would typically require a large window length, which could be infeasible in many applications where a decision needs to be made with low delay tolerance.

The notion of measure-preserving transformation (MPT) has been widely used to represent Hamiltonian (i.e., conservative) dynamical systems evolving on a probability space such that the total energy of the dynamical system is invariant [27,28]. A key concept in this regard is that even though a measure-preserving dynamical system could be described by a sequence of transformations with time-varying eigenvalues, the eigenfunctions may remain unchanged with time under the ergodicity assumption. Based on this rationale, this paper presents a methodology for constructing PFSAs, from short-length time series, which may generate a nonhomogeneous Markov chain model for adequately describing the underlying stochastic process. As a result, the time-invariance of eigenvectors, which reflects measure-invariance of the underlying ergodic dynamical system, can be used to decide the window length of the time series required to construct the PFSA. Unlike a standard STSA [15,23] that requires increasing the window length until the resulting PFSA is no longer significantly changing, the proposed MPT-based STSA increases the window length until the eigenvectors are nearly constant. 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 are no longer time-invariant. Along this line, a metric of variability in the eigenvectors is proposed as a measure of anomaly. The application example in this paper focuses on fatigue damage in polycrystalline alloys (e.g., Ref. [29]), which is a common source of failure in mechanical structures. The performance of the proposed MPT-based STSA is evaluated by comparison with those of a standard STSA and an HMM on the same experimental data sets. These data sets have been generated from a laboratory apparatus that uses ultrasonic sensor measurements to detect fatigue damage in polycrystalline–alloy specimens.

Major contributions of the paper:

(1) *Construction of an MPT-based framework of STSA for nonhomogeneous Markov chain modeling*: The underlying algorithms are built upon the concept of measure-invariance in dynamical systems [28,30–32], which facilitates ergodic measure-preserving modeling of dynamical systems from short-length time series.
(2) *Identification of a metric for machine/process anomaly*: Evolving anomalies are quantified as a norm of deviations in the eigenvectors of the constructed sequence of stochastic matrices by utilizing the invariant property of eigenvectors. This norm tends to be small if the dynamical system remains in the nominal state. As the system starts deviating from the nominal state, the eigenvectors no longer remain constant, and hence the quantified anomaly tends to increase. Anomalous patterns are detected as the metric exceeds a user-selected threshold.
(3) *Validation with experimental data*: The proposed anomaly detection methodology is validated on a laboratory-scale experimental apparatus for detection of fatigue damage in polycrystalline–alloy structures.

Organization of the paper: The paper is organized in six sections, including the present one. Section 2 provides background information on measure-preservation and ergodicity in dynamical systems. Section 3 briefly describes the principle of STSA for anomaly detection. Section 4 presents the technical approach for developing STSA-based anomaly detection algorithms that rely on spectral properties of sequences of MPTs. Section 5 presents the results of validation with experimental data. Section 6 summarizes and concludes the paper along with recommendations for future research.

2 Measure-Preserving Transformation and Ergodicity

This section provides a brief introduction to the notion of MPT that forms the backbone of the methodology presented in this paper; the details are extensively reported in literature (e.g., Refs. [12] and [33]). The following definitions are presented below for completeness of this paper and ease of readability.

DEFINITION 1. *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* $\Omega \backslash E \in \mathcal{E}$
- *A countable union of measurable sets is measurable, i.e., if* $\{E_k\}$ *is a countable collection of members of* $\mathcal{E}$, *then* $\cup_k E_k \in \mathcal{E}$.

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

DEFINITION 2. *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* $P[\cup_k E_k] \leq \sum_k P[E_k]$; *and the equality holds if the members of* $\{E_k\}$ *are pairwise disjoint, i.e.,* $E_i \cap E_j = \varnothing \ \forall i \neq j$.

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

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) or for P-almost all $x \in \Omega$. Therefore, all measurable sets that are equal P-ae form an equivalence class; members of this equivalence class are P-almost equal sets. Note: The symmetric difference $(E \Delta F) \triangleq (E \backslash F) \cup (F \backslash E)$.

DEFINITION 3. *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 $T^{-1}E \in \mathcal{E} \ \forall E \in \mathcal{E}$.*

A measurable set $E \in \mathcal{E}$ is called T-invariant if $P[E \Delta T^{-1}E] = 0$, which implies that $\mathrm{Tx} \in E$ for P-almost all $x \in E$. Furthermore, a function $f : \Omega \to \mathbb{C}$ is called T-invariant if $f(\mathrm{Tx}) = f(x)$ for P-almost all $x \in \Omega$, where $\mathbb{C}$ is the field of complex numbers.

A measurable transformation T is called a MPT if $P[T^{-1}E] = P[E] \ \forall E \in \mathcal{E}$.

A measure-preserving transformation T is called an endomorphism if T is surjective (i.e., onto).

Remark 1. The concept of MPT has been widely used to investigate the asymptotic properties of random sequences in statistical mechanics [33]. For an MPT T on a (finite) measure space $(\Omega, \mathcal{E}, P)$, every set $E \in \mathcal{E}$ has a recurrence property in the sense that once E is visited, it will be revisited infinitely many times; that is, if $x \in E$, then there are (countably) infinitely many values of n such that $T^n x \in E$ [33].

DEFINITION 4. [34] *Let $\{T^n\}$ be a one-parameter semigroup of MPTs on a probability space $(\Omega, \mathcal{E}, P)$. Note: An algebraic system $(S, \circ)$ is called a semigroup if the following two conditions hold: (i) closure,* i.e., $\circ : S \times S \to S$ *and (ii) associativity,* i.e., $\circ(x, \circ(y, z)) = \circ(\circ(x, y), z) \ \forall x, y, z \in S$.

A function $f \in L_1(P)$ is said to be an eigenfunction of a sequence $\{T^n\}$ of transformations with the corresponding sequence $\{\lambda_n\}$ of (scalar) eigenvalues if f is a nonzero function such that $f(T^n) = \lambda_n f$ P-ae and $\forall n \in \mathbb{N}$.

The sequence $\{T^n\}$ is said to be ergodic and P is called an ergodic measure if each T^n-invariant set $E \in \mathcal{E}$ is trivial, i.e., *either $P[E] = 0$ or $P[E] = 1$.*

The following theorem provides a property of MPTs on a probability space $(\Omega, \mathcal{E}, P)$ with an ergodic measure P.

THEOREM 2.1. [33] *Let $(\Omega, \mathcal{E}, P)$ be a probability space, and let $\{T^t\}$ be a semigroup of MPTs, where $t \in [0, \infty)$. Then, $\{T^t\}$ 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 MPT sequence $\{T^t\}$*, then* $|f(x)|$ *is a constant for P-almost all* $x \in \Omega$.

Proof. The proof of the theorem is given in Ref. [33]. ■

3 Symbolic Time Series Analysis

Before embarking on a description of the technical approach and the algorithms therein, it is necessary to provide the background for construction of a probabilistic finite state automaton (PFSA) (see Sec. 3.1) [15] and *D*-Markov machines (see Sec. 3.2) [23]. Section 3.1 refers to Sec. 2 that addresses the measure-preservation property of ergodic transformations.

3.1 Probabilistic Finite State Automata. A (finite length) time series is converted into a symbol string by partitioning the signal space into a finite number of cells, where the number of cells is identically equal to the cardinality $|\mathcal{A}|$ of the (symbol) alphabet $\mathcal{A}$, and each cell is assigned exactly one of the symbols in $\mathcal{A}$. At a given instant of time, a data point is assigned the symbol corresponding to the cell within which the data point is located; details are reported by Mukherjee and Ray [23]. The resulting symbol string is used to construct a *D*-Markov model, defined in Sec. 3.2, which models the statistics of the underlying stochastic process. The following definitions, which are available in standard literature (e.g., Refs. [15] and [23]), are recalled here for completeness of the paper and ease of readability.

DEFINITION 5. *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., its cardinality* $|\mathcal{A}| \in \mathbb{N}$.
- $\mathcal{Q}$ *is a (nonempty) finite set of states, i.e., its cardinality* $|\mathcal{Q}| \in \mathbb{N}$.
- $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ *is a (deterministic) state transition map.*

DEFINITION 6. *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* $w \triangleq s_1 s_2 \cdots s_\ell$ *with* $s_i \in \mathcal{A}$ *is* $|w| = \ell$*, and the length of the empty word* ε *is* $|\varepsilon| = 0$*. 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* $\mathcal{A}^\ell$.

A symbol string (or word) is generated from a (finite length) time series by symbolization.

DEFINITION 7. *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* $\pi : \mathcal{Q} \times \mathcal{A} \to [0, 1]$ *is also known as the symbol generation probability function that satisfies the condition:* $\sum_{\sigma \in \mathcal{A}} \pi(q, \sigma) = 1$ *for all* $q \in \mathcal{Q}$.

Often 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* $\mathcal{K} = (\mathcal{A}, \mathcal{Q}, \mathcal{T})$.

Remark 2. It is noted that the $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition probability matrix $\mathcal{T}$ is stochastic [35] (i.e., each element of $\mathcal{T}$ is nonnegative and each row sum is unity). Ergodicity of the underlying process, from which $\mathcal{T}$ is constructed, is equivalent to irreducibility of $\mathcal{T}$ [35], 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., $|\lambda| \leq 1$). The (sum-normalized) left eigenvector v corresponding to the unity eigenvalue (i.e., $\lambda = 1$) represents the stationary state probability vector of the Markov chain [35].

The mathematical concept of ergodic semigroup of endomorphisms and some of the relevant results have been presented in Sec. 2. Following Theorem 2.1, the absolute values of the eigenfunctions of individual transformations T^n at a time epoch n in the one-parameter ergodic semigroup of endomorphisms do not change with n although the respective eigenvalues may vary with n. These results are explained below in the context of STSA.

In the probability space $(\Omega, \mathcal{E}, P)$ for STSA, the sample space Ω is the (finite) state space $\mathcal{Q}$ of the PFSA under consideration, the associated σ-algebra $\mathcal{E}$ is the power set $2^{\mathcal{Q}}$, and P is the probability measure (see Sec. 2). The objective here is to model the system dynamics from a time series of measurements for anomaly detection in the STSA setting. With symbols $s \in \mathcal{A}$ occurring randomly, the state transition map $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ in Definition 5 becomes a random mapping $T : (\mathcal{Q}, \mathcal{E}, P) \to (\mathcal{Q}, \mathcal{E}, P)$ 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:

$$P[\{T(q)\} \in 2^{\mathcal{Q}}] \triangleq \sum_{\tilde{q} \in \mathcal{Q}} \kappa(q, \tilde{q}) \ \forall q \in \mathcal{Q} \tag{1}$$

where the state transition probability $\kappa(q, \tilde{q})$ is computed 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 $\kappa(q, \cdot)$. Note: The $|\mathcal{Q}| \times |\mathcal{Q}|$ state transition probability matrix $\mathcal{T}$ in Definition 7 is a stochastic representation of the random mapping T (see Remark 2).

Example 1. *In the probability space* $(\mathcal{Q}, \mathcal{E}, P)$*, let* $\mathcal{Q} = \{q_1, q_2\}$ *with the* σ*-algebra* $\mathcal{E} = 2^{\mathcal{Q}} \triangleq \{\varnothing, \{q_1\}, \{q_2\}, \mathcal{Q}\}$*, and* $P : \mathcal{E} \to [0, 1]$*. Let* $\{T^k\}$ *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 7) 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 matrix* $\mathcal{T}^k$ *are:* $\lambda_1^k = 1$ *and* $\lambda_2^k = (p^k + \tilde{p}^k - 1)$*; and the corresponding absolute-sum-normalized left eigenvectors are:* $v_1^k = [(1-\tilde{p}^k)/(2-p^k-\tilde{p}^k)) \ (1-p^k/(2-p^k-\tilde{p}^k))]$ *that are k-variant, in general, and* $v_2^k = [0.5 - 0.5]$ *that are k-invariant. Hence,* $P[(T^k)^{-1}\mathcal{Q}] = P[\mathcal{Q}] = 1$ *and* $P[(T^k)^{-1}\varnothing] = P[\varnothing] = 0$*. Furthermore, the pre-image of* $\{q_1\}$ (i.e., $(T^k)^{-1}\{q_1\}$) *is either* $\{q_1\}$ *itself or* $\{q_2\}$*; similar results hold for* $\{q_2\}$*. Then, it follows that* $P[(T^k)^{-1}E] = P[E] \forall E \in \mathcal{E}$ (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* $\{T^k\}$ *is ergodic, because the only* T^k*-invariant measurable sets (*i.e., *members of* $\mathcal{E}$*) are* $\varnothing$ *and* $\mathcal{Q}$*. Then, it follows from Definitions 3 and 4 that, in this example, the sequence* $\{T^k\}$ *is measure-preserving and ergodic if and only if* $\tilde{p}^k = p^k$ *and* $p^k \in [0, 1) \forall k$.

Following Definitions 4 and 7, a sequence $\{\mathcal{K}^n\} \triangleq \{(\mathcal{A}, \mathcal{Q}, \mathcal{T}^n)\}$ 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 this paper, follows from Theorem 2.1 for PFSAs and is presented as the following corollary.

COROLLARY 1. *Let* $\{\mathcal{K}^n : n \in \mathbb{N}\}$*, be a sequence of measure-preserving and ergodic PFSAs on a probability space* $(\mathcal{Q}, \mathcal{E}, P)$*. Then, the (sum-normalized) left eigenvector* v_1^n*, corresponding to the eigenvalue* $\lambda_1 = 1$*, is uniformly distributed,* i.e.,

$$v_1^n = \left[\frac{1}{|\mathcal{Q}|}, \ldots, \frac{1}{|\mathcal{Q}|}\right] \ \forall n$$

3.2 D-Markov Machines. In the construction of a *D*-Markov machine, it is assumed that the generation of the next symbol

depends 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 [23] is defined as follows.

DEFINITION 8. *A D-Markov machine is a PFSA in the sense of Definition 7 and it 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* $S = \cdots s_{-1}s_0s_1\cdots$, *where the probability of occurrence of a new symbol depends only on the last consecutive (at most) D symbols, i.e.,*

$$P[s_n|\cdots s_{n-D}\cdots s_{n-1}] = P[s_n|s_{n-D}\cdots s_{n-1}] \tag{2}$$

Consequently, for $w \in \mathcal{A}^D$ *(see Definition 6), the equivalence class* $\mathcal{A}^*w$ *of all (finite length) words of suffix w, is qualified to be a D-Markov state that is denoted as w.*

Remark 3. While the algebraic structure of the PFSA in D-Markov machines is deterministic [15,23], a HMM may have a nondeterministic algebraic structure [21,22]; this difference yields significant computational advantages of D-Markov machines over HMMs at the expense of limited loss of modeling flexibility. Moreover, since HMMs are typically trained by expectation maximization [5], the underlying algorithms might suffer from having a poor local optimum. In addition to its iterative computation cost, HMMs may not be sufficiently robust in terms of convergence even to a locally optimum point. In contrast, the deterministic algebraic structure of D-Markov machines makes the modeling process much simpler and less prone to the local optimum issue, where the model can be trained by frequency counting [23], for example.

The PFSA of a D-Markov machine is capable of generating symbol strings. That is, such a generated symbol string has the form $\{s_1s_2\cdots s_\ell\}$, where $s_j \in \mathcal{A}$ and ℓ is a positive integer. Both morph function π and state probability transition 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 [23].

For construction of the proposed D-Markov-based PFSA, there are four primary choices as enumerated below:

- *Alphabet size* ($|\mathcal{A}|$): Larger is the alphabet size, the dynamical system is more discretely represented as different symbols, which will require more training data. Therefore, selection of the alphabet $\mathcal{A}$ is a critical step of PFSA construction; while there are several techniques for selection of $\mathcal{A}$ (e.g., Refs. [19] and [23]), the alphabet size $\mathcal{A}$ has been chosen to be very small (e.g., $|\mathcal{A}|= 2$ or 3) in this paper so that the number of states $|\mathcal{Q}|$ is also small to limit the required window length L.
- *Partitioning method*: While there are many data partitioning techniques, maximum entropy partitioning (MEP) [16,18] and K-means partitioning [36] have been chosen as they are commonly used for symbolization of time series.
- *Depth (D) in the D-Markov machine*: In this paper, $D = 1$ has been primarily chosen for PFSA construction to limit the window length L. Higher values of the positive integer D may lead to better results at the expense of increased number of states $|\mathcal{Q}|$ and central processing unit (CPU) time due to larger dimension of the feature space and may need more training data.
- *Choice of feature*: The feature needs to be the one that best captures the physical nature (e.g., texture) of the signal and also that is not computationally too expensive. In this paper, (absolute sum-normalized) left eigenvectors of the $\mathcal{T}$-matrix are selected as features.

3.3 Anomaly Detection in a Standard Symbolic Time Series Analysis Setting. From the perspectives of discrete-time measurements and their discrete-space symbolization, usage of D-Markov machines is an efficient and convenient way of modeling a dynamical system. In this setting, the time series $\{x_n\}$ is converted into a symbol string $\{s_n\}$, $s_n \in \mathcal{A}$, where $\mathcal{A}$ is a finite cardinality alphabet. Then, PFSAs are constructed from the symbol strings, which in turn generate low-dimensional feature vectors that are used for detection of anomalous patterns. The procedure for anomaly detection using a standard STSA [23], is executed in the following steps:

(1) *Select* a block of a time series, called the nominal block, for which the system is in a normal operating condition.
(2) *Construct* a partitioning for the nominal block and convert it into a symbol string to construct the nominal PFSA. The emission matrix (and hence the state transition probability matrix) of the PFSA are constructed by frequency counting [23]. This *learned* nominal model generates a probability vector v^0 that represents the nominal pattern.
(3) *Select* a new block of the time series up to the current time n and convert it into a symbol string using the learned nominal partition. This yields a new PFSA with a new (quasi-) stationary probability vector v^n that represents the feature vector at the time epoch n.
(4) *Compute* the (scalar-valued) anomaly metric θ^n at the time epoch n from a string of divergences between the nominal feature vector v^0 and the current feature vector v^m and by sliding the block of data N times as:

$$\theta^n \triangleq \frac{1}{N}\sum_{m=n}^{n+N-1} d_{KL}\left(v^0, v^m\right) \tag{3}$$

where $d_{KL}(\bullet,\bullet)$ is the Kullback–Leibler divergence [37], and the number N of sliding windows serves the sole purpose of data smoothing. The anomaly metric θ^n in Eq. (3) is used in this paper for the standard STSA, which is more robust to outliers and fast fluctuations in the time series than if an individual term is used for computation of $d_{KL}(v^0, v^m)$.

4 Technical Approach

A major issue in a standard STSA-based anomaly detection is that it may require a long time series to construct a stationary state transition probability matrix $\mathcal{T}$, from which the stationary state probability vector v is generated to serve as a feature vector. However, many applications do require low-delay detection of anomalous events, for which short-length windows of time series must be used. For example, fatigue damage in critical mechanical structures must be detected as early as possible to avoid a catastrophic failure. This requirement mandates early detection of the damage by using short-length windows for PFSA construction [11].

To address the above-mentioned issue, the dynamical system is described by an ergodic sequence of MPTs $\{\mathcal{T}^n\}$, $n \in \mathbb{N}$, acting on a probability space $(\Omega, \mathcal{E}, P)$. It is stated in Sec. 3.1 and is reiterated here that the sample space Ω is the (finite) state space $\mathcal{Q}$ of the PFSA of a D-Markov machine, the associated σ-algebra $\mathcal{E}$ is the power set $2^{\mathcal{Q}}$, and P is the probability measure. Based on the concept of MPT-based STSA, if a time-varying PFSA is constructed using short-length windows of time series, then the resulting sequence of state transition probability matrices $\{\mathcal{T}^n\}$ would describe nonhomogeneous Markov chain models. Although the underlying stochastic process could be stationary, short-length windows of time series may produce time-varying $\mathcal{T}$-matrices; hence, in general, the eigenvalues and eigenvectors of $\mathcal{T}$ may be time-varying as well. However, in light of Corollary 1 in Sec. 3.1, the left eigenvectors of $\{\mathcal{T}^n\}$ should be n-invariant to reflect measure-invariance and ergodicity of the dynamical system model. Therefore, rather than increasing the length of windows until a constant $\mathcal{T}$ is obtained, as required by a standard STSA, the proposed MPT-based STSA requires that $\mathcal{T}$ may only satisfy the property of time-invariant left eigenvectors given by Corollary

1. Moreover, a metric of variability of left eigenvectors is proposed as a measure of anomaly in the dynamical system.[2]

A modification of a standard STSA is now proposed based on the variations of eigenvectors. Table 1 lists the major differences, demonstrated in the sequel, between a standard STSA and the proposed MPT-based STSA.

Algorithm 1 MPT-based selection of window length

INPUT: Alphabet $\mathcal{A}$, Markov depth D, a time series $\{x_n\}$ generated by an ergodic measure-preserving stochastic process, number of sliding windows $N \in \mathbb{N}$, increment of the window length ΔL in each iteration, and a threshold $\tau_1 > 0$.
OUTPUT: Window length L of the data blocks $\{x_{n:n+L}\}$ for construction of a D-Markov model of the stochastic process.
1: Choose an initial window size L.
2: **do**
3: Convert time series blocks $\{x_{n+1:n+L}\}$, $n = 0, 1, \ldots, N-1$, into symbol strings $\{s_{n+1:n+L}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
4: Using frequency counting, construct a D-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{\mathcal{T}^n\}$.
5: Find the (sum-normalized) left eigenvectors $\{v_1^n\}$, corresponding to $\lambda_1 = 1$, for each one of the state transition probability matrices.
6: $\bar{v}_1 \leftarrow \frac{1}{N}\sum_{n=0}^{N-1} v_1^n$.
7: $\theta_1 \leftarrow \frac{1}{N}\sum_{n=0}^{N-1} |(v_1^n - \bar{v}_1)|$.
8: $\rho \leftarrow ||\theta_1||_{l_1}$, where $||\theta_1||_{l_1} \triangleq \sum_{j=1}^{|\mathcal{Q}|} |\theta_{1,j}|$.
9: $L \leftarrow L + \Delta L$.
10 **while** $\rho > \tau_1$

Algorithm 2 MPT-based anomaly detection

INPUT: Alphabet $\mathcal{A}$, Markov depth D, window length L, time series $\{x_n\}$, number of sliding windows $N \in \mathbb{N}$, a threshold $\tau_2 > 0$.
OUTPUT: The decision on whether the system is nominal or anomalous.
1: Convert time series blocks $\{x_{n+1:n+L}\}$, $n = 0, 1, \ldots, N-1$, into symbol strings $\{s_{n+1:n+L}\}$, $s_i \in \mathcal{A}$, using one of the STSA partitioning methods.
2: Using frequency counting, construct a D-Markov machine based on each $s_{n+1:n+L}$ to obtain state transition probability matrices $\{\mathcal{T}^n\}$.
3: Find the (sum-normalized) left eigenvectors $\{v_1^n\}$, corresponding to $\lambda_1 = 1$, for each one of the state transition probability matrices.
4: $\bar{v}_1 \leftarrow \frac{1}{N}\sum_{n=0}^{N-1} v_1^n$.
5: $\theta_1 \leftarrow \frac{1}{N}\sum_{n=0}^{N-1} |(v_1^n - \bar{v}_1)|$.
6: $\rho \leftarrow ||\theta_1||_{l_1}$, where $||\theta_1||_{l_1} \triangleq \sum_{j=1}^{|\mathcal{Q}|} |\theta_{1,j}|$.
7: **if** $\rho > \tau_2$ **then**
8: declare the system as anomalous
9: **else**
10: declare the system as nominal
11: **end if**

Algorithms 1 and 2 implement the theory of proposed MPT-based STSA. Algorithm 1 presents a procedure for identification of the window length L when the measure-preserving and ergodicity conditions are satisfied. In Algorithm 1, the parameter L is kept on increasing until a metric of variations of the left eigenvector, hereafter called ergodicity metric ρ, which should be a constant for measure-preserving and ergodic processes, is less than the (user-specified) threshold parameter τ_1. Algorithm 2 presents the steps for detecting anomalous patterns using short-length windows by constructing a D-Markov machine for the underlying stochastic process, given the alphabet size $|\mathcal{A}|$, the Markov depth D, and window length L. In this situation, an anomaly is detected online when the aforementioned ergodicity metric ρ exceeds the (user-specified) threshold parameter τ_2. It is demonstrated in Sec. 5 that this criterion can be used to achieve class separability in the feature space for pattern classification. Figures 1 and 2 show the flowcharts of Algorithms 1 and 2, respectively.

The threshold parameters τ_1 and τ_2 in Algorithms 1 and 2, respectively, are interdependent in the sense that a smaller τ_1 would result in a larger window length L (i.e., increased delay), which would require a smaller τ_2 under a fixed maximum allowable false positive rate (FPR) for similar qualities of decisions on anomaly detection. Both parameters, tolerated delay and maximum allowable FPR, are generally application-dependent. In this regard, Pareto optimization for selection of the threshold parameters τ_1 and τ_2 is recommended as a topic of future research in Sec. 6.

5 Experimental Validation

This section evaluates the performance of the MPT-based STSA for early prediction of fatigue damage in a polycrystalline–alloy material. In this regard, an ensemble of time series has been generated for both nominal (e.g., undamaged) and anomalous (e.g., damaged) conditions; the objective here is to evaluate the performance of the MPT-based STSA, presented in Algorithms 1 and 2, for online anomaly detection using short-length windows and low-dimensional feature vectors. The performance of the MPT-based STSA is compared with those of a standard STSA (e.g., Ref. [23]) and an HMM [5][3] on the same experimental data. In all three methods (i.e., standard STSA, MPT-based STSA, and HMM), the number of sliding windows is taken as $N = 30$ (see Eq. (3), Algorithms 1 and 2).

Seventeen experiments have been conducted on the test apparatus, shown in Fig. 3(*a*), which is built upon a computer-instrumented and computer-controlled fatigue testing machine equipped with ultrasonic sensing and optical microscopy. The test specimens (see Fig. 3(*b*)) are made of polycrystalline aluminum 7075-T6 alloy, where each specimen is 3 mm thick and 50 mm wide with a notch of 1.58 mm × 4.57 mm on one side of the specimen. The notch is made to increase the (local) stress concentration factor that ensures crack initiation and propagation at the notch end.

The test specimens have been subjected to sinusoidal loading on the apparatus under tension–tension stress at a frequency of ∼50 Hz. The ultrasonic time series collected during each experiment contains approximately 1,000,000 data points. Since it is not possible by inspection to determine the exact point at which a change in the time series has occurred, an interval is identified (within which this change point is located) to serve as the ground truth, against which efficacy of the test methods are evaluated. A part of the time series before that interval and another part of the time series after that interval are selected such that these two parts can be concatenated to form a single time series with a clearly defined instant of change time; the total length of these two parts of the selected time series is ∼10, 000.

Figure 4 shows the reconstructed ultrasonic signals for sixteen (16) test specimens, and Fig. 5 shows a typical sample of the reconstructed ultrasonic signal with a fatigue failure onset at the approximate stress cycle 4525. The goal here is to promptly detect such an onset point using a short-length window of the ultrasonic signal. The main challenges in doing so are:

(1) Fatigue failure phenomena in polycrystalline alloys are stochastic, which increases the probabilities of missed detections and false alarms due to large uncertainties of (randomly changing) failure onset points in individual specimens, even under identical loading conditions.
(2) Typically the change in a signal pattern is very small, as seen in the expanded view of Fig. 5 over a span of 1000

[2]One may check the eigenvectors, which bear pertinent information on occurrence of anomalies. Apparently, it suffices to look at the left eigenvector v_1, corresponding to the eigenvalue $\lambda_1 = 1$, to validate the anomaly detection algorithm in Sec. 5. However, depending on the particular application, examination of some of the other eigenvectors could be effective in quantifying anomalies of the system.

[3]Gaussian mixtures have been used for state-conditional density functions in the HMM implemented here, where the number of mixture components is determined by Bayesian information criterion (BIC) [36]. Details on the HMM-based anomaly detection are given in Ref. [11].

Table 1 Comparison of standard STSA and MPT-based STSA

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

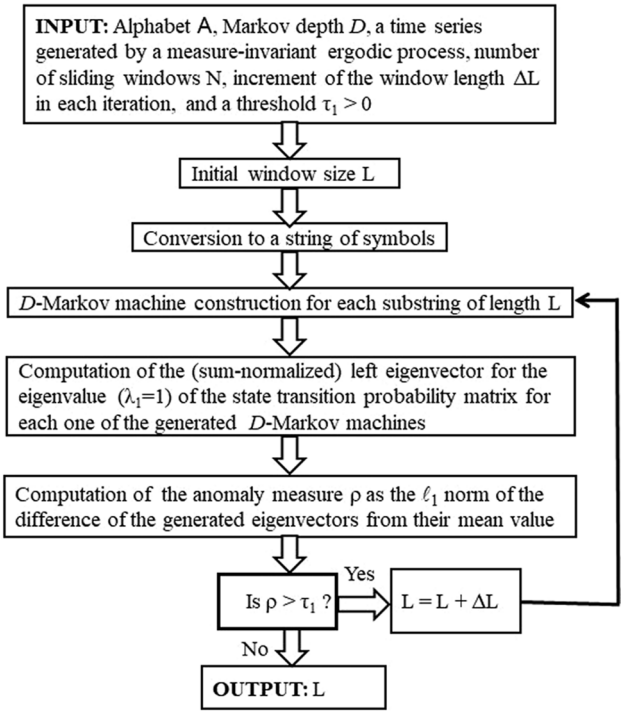

Fig. 1 Flowchart for Algorithm 1

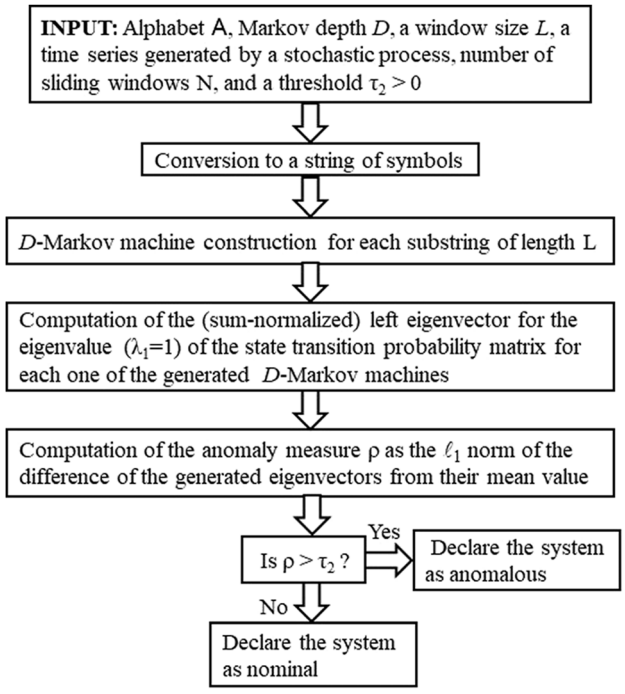

Fig. 2 Flowchart for Algorithm 2

cycles at the top right-hand corner. The challenge is to detect such pattern changes and to identify the onset point in near-real-time (e.g., within a short window of length, say, ranging from $L = 50$ to $L = 200$).

(3) Low-cost ultrasonic sensors have been used for generating and measuring the signals. These sensors are noisy and are very sensitive to the location where they are fixed on the specimen. Therefore, a slight change in their relative locations may produce a significant change in the measured signals. These phenomena would increase the uncertainty of the failure onset point and make its detection even more challenging.

Considering the healthy part of an ultrasonic signal in Fig. 5, nearly all the signal components are available at the receiver sensor. The signal energy is *nearly* constant and the process can be considered as *nearly* measure-preserving. Moreover, if the signal space is partitioned and the signal time series is converted to a symbol sequence, then every cell is visited many times if the healthy state prevails for a long time period. Therefore, the resulting symbolic dynamics can be considered as measure-preserving and ergodic. In contrast, the transient part of the signal cannot be considered as measure-preserving because some of the signal components are reflected back rather than being available at the receiver, and the received signals are attenuated as a result of the decrease in energy. Consequently, by following Corollary 1 in Sec. 3.1, it is expected that eigenvectors would tend to be time-invariant in the nominal (healthy) phase and time-varying in the transient (anomalous) phase.

The above-mentioned facts are demonstrated in Fig. 6 by analysis of experimental data, where the behavior of the eigenvectors is displayed over time for different window lengths L of time-series blocks. The state transition probability matrix $\mathcal{T}$ is constructed for an alphabet size of $|\mathcal{A}| = 2$ and Markov depth $D = 1$, which leads to the number of states $|\mathcal{Q}| = 2$ (see Sec. 3.1). Since $|\mathcal{Q}| = 2$ in this case, there are only two (two-dimensional) eigenvectors and only one element of the eigenvector v_1 corresponding to the eigenvalue λ_1 is needed.[4]

Figure 6 shows that the eigenvectors (corresponding to the eigenvalue $\lambda_1 = 1$) tend to become time-invariant in the nominal (healthy) phase as the window length L is increased, while they are grossly time-varying in the transient (damaged) phase. The results displayed in Fig. 6 are, only to some extent, consistent with Corollary 1 in the sense that the (sum-normalized) eigenvectors, corresponding to $\lambda_1 = 1$, should converge to the uniform

[4]In this paper, eigenvectors are always absolute-sum-normalized; therefore, given an $M = 2$-dimensional eigenvector, only $(M - 1 = 1)$ element(s) of the eigenvector need to be specified because the remaining Mth component will be a linear combination of the rest.

distribution provided that the underlying stochastic process is truly ergodic and measure-preserving. However, the left eigenvector, corresponding the eigenvalue $\lambda_1 = 1$ in the nominal phase, converges to a constant vector that deviates from being uniformly distributed (i.e., $v_{1,1} = 0.3$ instead of 0.5), as seen in Fig. 6. A possible reason for this deviation is that the dynamical system is not perfectly measure-preserving, because the ultrasonic signal is received with some loss or dissipation. An explanation following Example 1 is that $\tilde{p}_k \neq p_k$ in the (2×2) state transition probability matrix $\mathcal{T}^k = \begin{bmatrix} p^k & 1-p^k \\ 1-\tilde{p}^k & \tilde{p}^k \end{bmatrix}$ where $p^k, \tilde{p}^k \in [0, 1)$. For a large window length (e.g., $L = 1000$), a typical value of the state transition probability matrix in the nominal phase is: $\mathcal{T} \approx \begin{bmatrix} 0.65 & 0.35 \\ 0.15 & 0.85 \end{bmatrix}$, which yields a nonuniform left eigenvector $v_1 = [0.3\ 0.7]$. In this case, the stochastic model of the dynamical system is ergodic but *not exactly* measure-preserving. In this context, the system in the nominal phase is considered to be ergodic and *approximately* measure-preserving. This property is largely violated in the transient phase, where energy dissipation and loss in the ultrasonic signals are much more significant. It is seen in Fig. 6 that once the damage evolution starts, the eigenvectors lose the time-invariance property and become time-varying.

Figures 7 and 8 compare the receiver operating characteristic (ROC)[5] performance of MPT-based STSA, described in Algorithm 2, with those of a standard STSA and HMM in terms of the area under the curve of ROC plots, for different parameters (i.e., window length L and alphabet size $\mathcal{A}$) and partitioning methods. While the results of MPT-based STSA and standard STSA are different for MEP and K-means partitioning in the ROC plots of Figs. 7 and 8, respectively, the results for HMM in these two figures are largely similar because data partitioning is not relevant in the (stochastic) algorithm of HMM. All eight plates in Figs. 7 and 8 exhibit that the performance of MPT-based STSA is consistently superior.[6]

Since the (two-dimensional) eigenvectors are sum-normalized, a change in one component implies a change in the other one. Therefore, only one component of the eigenvector, corresponding to the eigenvalue $\lambda_1 = 1$, has been used. Different values of window lengths are considered: $L = 50$, 100, and 200. As shown in Figs. 7 and 8, a significant improvement over the standard STSA and some improvement over HMM are obtained by MPT-based STSA for different values of L, with an excellent performance at $L = 200$, where area under the curve ~ 1, using K-means for partitioning. It is also noticed from the figures that increasing the number of hidden states may degrade the detection performance of HMM. This behavior is due to proneness of HMMs with Gaussian mixtures to singularity, especially when a short-length time series is used for training. This proneness is increased when the number of hidden states is increased [39], in which case the maximization of the observed data likelihood may fail or end up with a poor local minimum.

Table 2 lists the statistics of CPU execution time[7] of standard STSA, MPT-based STSA, and HMM for fatigue damage detection using an ensemble of ultrasonic signals. As explained in Ref. [11], the decision of the HMM involves only the computation of the joint likelihood of the observed time series conditioned on the nominal HMM. This is efficiently computed by using the forward procedure only, which does not require iterations. Table 2 shows that HMM typically takes more CPU execution time than both STSA methods.

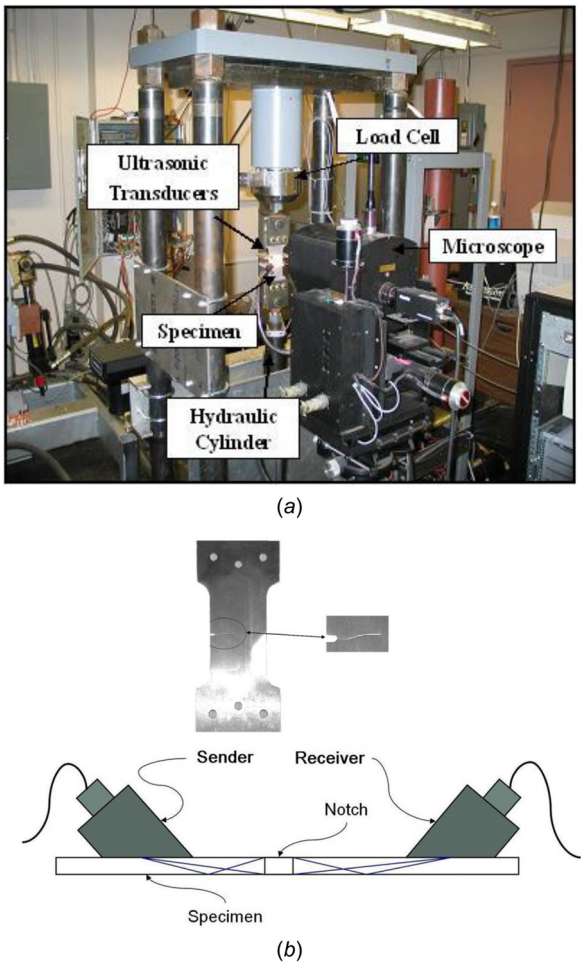

Fig. 3 (*a*) Fatigue testing apparatus and (*b*) details of a test specimen

6 Summary, Conclusion, and Future Work

This paper has presented an alternative framework for STSA by constructing time-inhomogeneous Markov models of dynamical systems, based on the theory of ergodic sequences of MPTs. This concept of MPT is used to develop an STSA-based anomaly detection method. The underlying algorithm has been validated on experimental data for fatigue damage detection in polycrystalline alloys by using short-length time series of ultrasonic measurements. The experimental validation has been conducted with different detection parameters, data partitioning methods, and window lengths of the observed time series. The performance of the proposed MPT-based STSA has been compared with those of a standard STSA [23] and a HMM [5,10,11] for anomaly detection; the results show consistent superiority of MPT-based STSA in terms of detection accuracy. An important property of MPT-based STSA is that it enhances robustness of decisions on anomaly detection to changes in the nominal phase. It is concluded that the MPT-based STSA is suitable for near-real-time detection of anomalies and prediction of (possibly) forthcoming failures.

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 ROC curves have been used here for assessing the detection performance by varying the parameter τ_2 [38], where each point in a ROC curve corresponds to a specific value of the threshold τ_2. It is noted that τ_2 can be determined from the ROC curve by specifying a maximum allowable FPR, which may depend on the application. If the cost of a positive false alarm is low, the maximum FPR could be increased. On the other hand, for applications where the cost for a positive false alarm is high, a small value for the maximum FPR should be selected.

[6]For fair comparison, the number of states for both STSA methods and HMM in each experiment is the same. That is, if an alphabet $\mathcal{A}$ is used with a Markov depth D in STSA, then $|\mathcal{A}|^D$ hidden states are used in HMM.

[7]The results have been computed on a DELL PRECISION T3400 with an Intel(R) Core(TM)2 Quad CPU Q9550 at 2.83 GHz and 8 GB RAM, running under Windows 7.

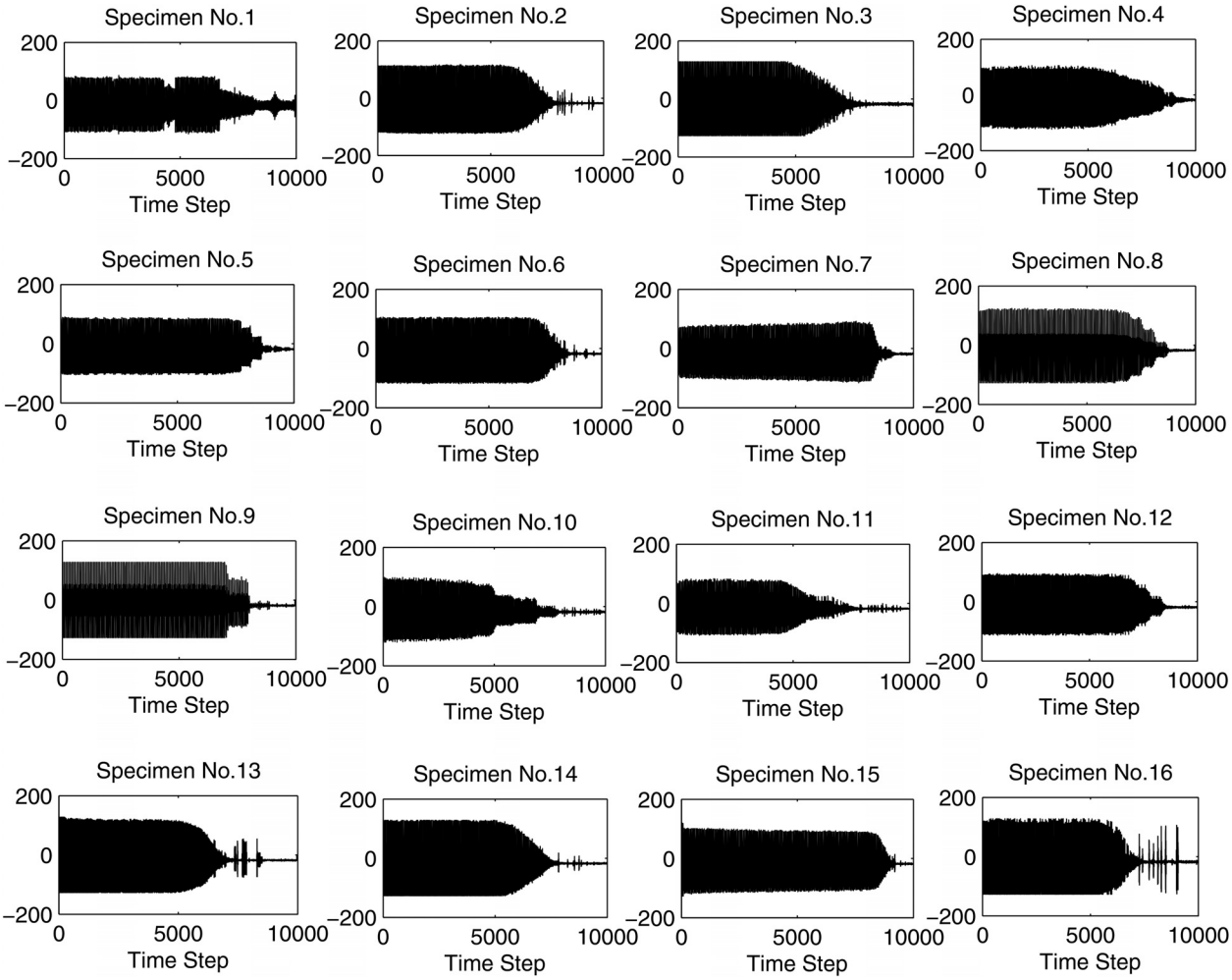

Fig. 4 Ultrasonic signals for several typical specimens

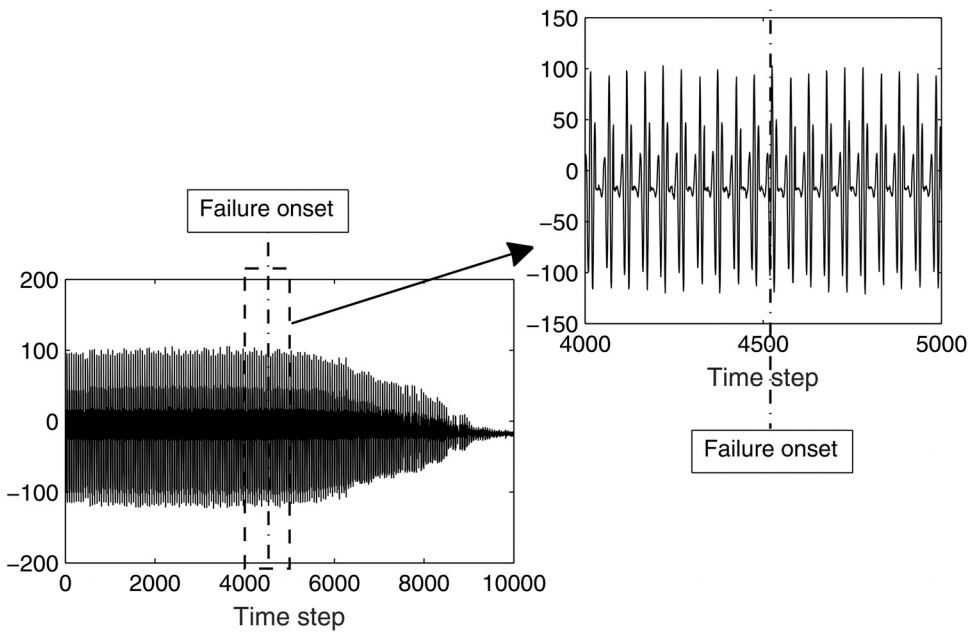

Fig. 5 Ultrasonic signal with fatigue onset at 4525 cycles

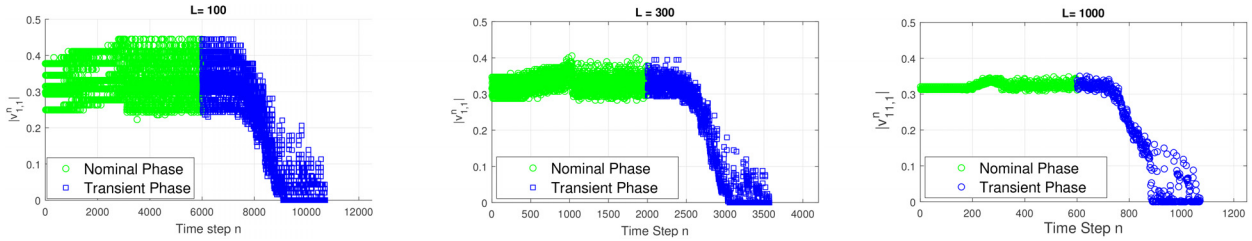

Fig. 6 Convergence of eigenvectors, corresponding to the eigenvalue $\lambda_1 = 1$ to constants in the nominal phase, with a time-varying behavior in the transient phase, for $|\mathcal{A}| = 2$, $D = 1$, and three different window lengths L

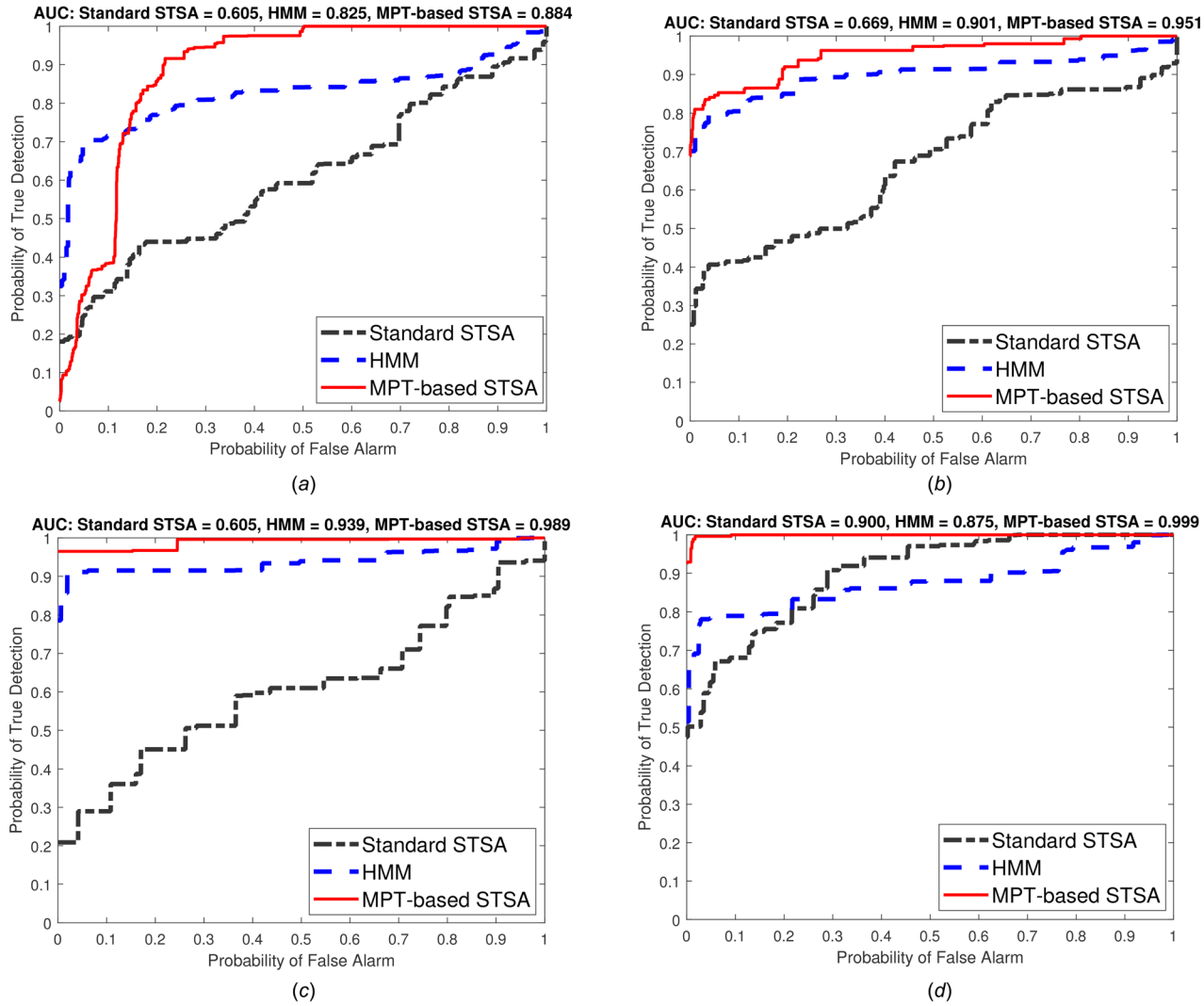

Fig. 7 ROC performance of the MPT-based STSA and the standard STSA, with MEP partitioning, and HMM for fatigue damage detection: (*a*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 50$ with MEP, (*b*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 100$ with MEP, (*c*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 200$ with MEP, and (*d*) $|\mathcal{A}| = 3$, $D = 1$, and $L = 200$ with MEP

(1) *Pareto optimization for selection of the threshold parameters τ_1 and τ_2 in Algorithms 1 and 2, respectively*: This will also involve rigorous statistical analysis (e.g., Ref. [40]).
(2) *Comparative evaluation of MPT-based STSA with other techniques of anomaly detection*: Examples of potential candidates are neural network-based forecasting [41], and dynamic mode decomposition [42].
(3) *Investigation of sufficient conditions for commutativity of MPTs*: In this case, the commutator norm can be used as a measure of evolving anomalies, which is expected to yield a significant reduction in CPU execution time.
(4) *Experimental validation of MPT-based STSA in diverse applications*: These applications should demonstrate high performance and robustness of MPT-based STSA for real-time monitoring and active control.

Acknowledgment

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

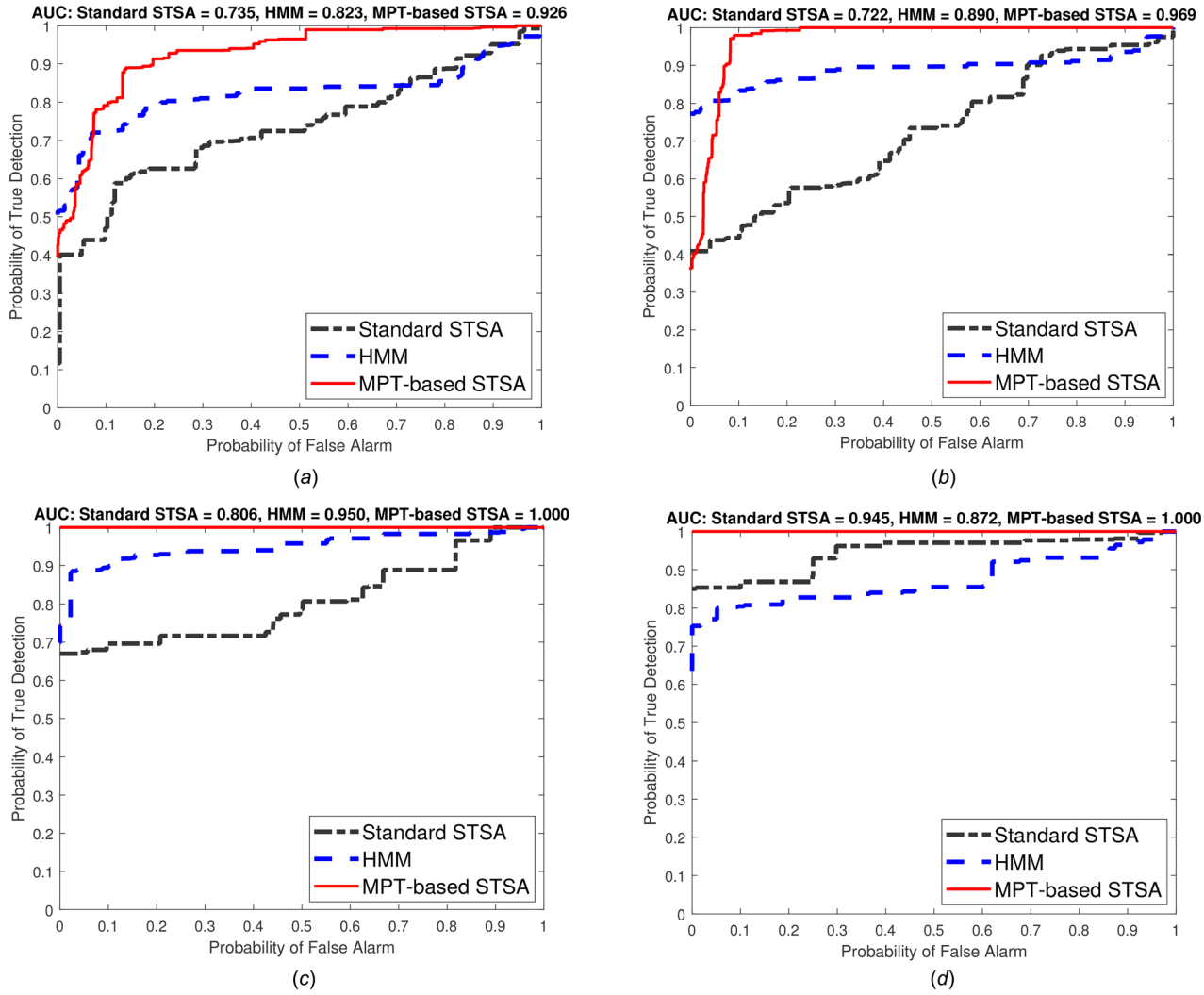

Fig. 8 ROC performance of the MPT-based STSA and the standard STSA, with K-means partitioning, and HMM for fatigue damage detection: (*a*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 50$ with *K*-means, (*b*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 100$ with *K*-means, (*c*) $|\mathcal{A}| = 2$, $D = 1$, and $L = 200$ with *K*-means, and (*d*) $|\mathcal{A}| = 3$, $D = 1$, and $L = 200$ with *K*-means

Table 2 Statistics of CPU execution time (sec) per window for online fatigue damage detection

	Standard STSA		MPT-based STSA		HMM	
Parameters and partitioning method	Mean	Std dev	Mean	Std dev	Mean	Std dev
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 50$, MEP	2.133×10^{-5}	9.609×10^{-6}	2.133×10^{-5}	9.610×10^{-6}	2.754×10^{-4}	2.372×10^{-5}
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 100$, MEP	3.866×10^{-5}	9.110×10^{-6}	3.866×10^{-5}	9.110×10^{-6}	3.956×10^{-4}	3.422×10^{-5}
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 200$, MEP	9.031×10^{-5}	7.117×10^{-5}	9.031×10^{-5}	7.117×10^{-5}	7.693×10^{-4}	4.282×10^{-5}
$\|\mathcal{A}\| = 3$, $D = 1$, $L = 200$, MEP	1.733×10^{-4}	9.122×10^{-5}	1.733×10^{-4}	9.119×10^{-5}	8.831×10^{-4}	6.494×10^{-5}
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 50$, *K*-means	1.053×10^{-4}	5.696×10^{-5}	1.054×10^{-4}	5.696×10^{-5}	2.783×10^{-4}	3.226×10^{-5}
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 100$, *K*-means	1.339×10^{-4}	1.616×10^{-5}	1.339×10^{-4}	1.616×10^{-5}	4.109×10^{-4}	3.606×10^{-5}
$\|\mathcal{A}\| = 2$, $D = 1$, $L = 200$, *K*-means	2.177×10^{-4}	8.809×10^{-5}	2.177×10^{-4}	8.809×10^{-5}	7.816×10^{-4}	4.791×10^{-5}
$\|\mathcal{A}\| = 3$, $D = 1$, $L = 200$, *K*-means	3.048×10^{-4}	1.272×10^{-4}	3.048×10^{-4}	1.272×10^{-4}	9.063×10^{-4}	6.529×10^{-5}

by U.S. Air Force Office of Scientific Research (AFOSR) under Grant No. FA9550-15-1-0400 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.

Funding Data

- Higher Committee for Education Development (HCED) (Funder ID: 10.13039/501100009928).
- U.S. Air Force Office of Scientific Research (AFOSR) (Grant No. FA9550-15-1-0400; Funder ID: 10.13039/100000181).

References

[1] Chamroukhi, F., Same, A., Govaert, G., and Aknin, P., 2009, "Time Series Modeling by a Regression Approach Based on a Latent Process," Neural Networks, **22**(5–6), pp. 593–602.

[2] Frick, K., Munk, A., and Sieling, H., 2014, "Multiscale Change Point Inference," J R. Stat Soc. Ser. B: Stat. Methol., **76**(3), pp. 495–580.

[3] Page, E. S., 1954, "Continuous Inspection Schemes," Biometrika, **41**(1–2), pp. 100–115.

[4] Basseville, M., and Nikiforov, I. V., 1993, *Detection of Abrupt Changes—Theory and Application*, Prentice Hall, Englewood Cliffs, NJ.
[5] Rabiner, L., 1989, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, **77**(2), pp. 257–286.
[6] Dorj, E., and Chen, C., 2013, "Anomaly Detection Approach Using Hidden Markov Model," IEEE Aerospace Conference, Ulaanbaatar, Mongolia, June 28–July 1, pp. 1–10.
[7] Fridlyand, J., Snijders, A. M., Pinkel, D., Albertson, D. G., and Jain, A. N., 2004, "Hidden Markov Models Approach to the Analysis of Array CGH Data," J. Mult. Anal., **90**(1), pp. 132–153.
[8] Blanding, W. R., Willett, P. K., Bar-Shalom, Y., and Coraluppi, S., 2009, "Multisensor Track Management for Targets With Fluctuating Snr," IEEE Trans. Aero Elect. Sys., **45**(4), pp. 1275–1292.
[9] Zhang, Y., Brady, M., and Smith, S., 2001, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," IEEE Trans. Med. Imaging, **20**(1), pp. 45–57.
[10] Mondal, S., Ghalyan, N. F., Ray, A., and Mukhopadhyay, A., 2019, "Early Detection of Thermoacoustic Instabilities Using Hidden Markov Models," Combust. Sci. Technol., **191**(8), pp. 1309–1336.
[11] Ghalyan, N. F., Mondal, S., Miller, D. J., and Ray, A., 2019, "Hidden Markov Modeling-Based Decision-Making Using Short-Length Sensor Time Series," ASME J. Dyn. Sys. Meas. Control, **141**(10), p. 104502.
[12] Ghalyan, N., 2019, "Sequential Machine Learning for Decision Making in Mechanical Systems," Ph.D. thesis, Pennsylvania State University, University Park, PA.
[13] Beim Graben, P., 2001, "Estimating and Improving the Signal-to-Noise Ratio of Time Series by Symbolic Dynamics," Phys. Rev. E, **64**(5), p. 051104.
[14] Daw, C. S., Finney, C. E. A., and Tracy, E. R., 2003, "A Review of Symbolic Analysis of Experimental Data," Rev. Sci. Instrum., **74**(2), pp. 915–930.
[15] Ray, A., 2004, "Symbolic Dynamic Analysis of Complex Systems for Anomaly Detection," Signal Process., **84**(7), pp. 1115–1130.
[16] Rajagopalan, V., and Ray, A., 2006, "Symbolic Time Series Analysis Via Wavelet-Based Partitioning," Signal Process., **86**(11), pp. 3309–3320.
[17] Gupta, S., and Ray, A., 2007, "Symbolic Dynamic Filtering for Data-Driven Pattern Recognition," *Pattern Recognition: Theory and Application*, Chap. 2, E. A. Zoeller, ed., Nova Science Publishers, Hauppage, NY.
[18] Subbu, A., and Ray, A., 2008, "Space Partitioning Via Hilbert Transform for Symbolic Time Series Analysis," Appl. Phys. Lett., **92**(8), p. 084107.
[19] Sarkar, S., Chattopdhyay, P., and Ray, A., 2016, "Symbolization of Dynamic Data-Driven Systems for Signal Representation," Signal, Image, Video Process., **10**(8), pp. 1535–1542.
[20] Ghalyan, N. F., Miller, D. J., and Ray, A., 2018, "A Locally Optimal Algorithm for Estimating a Generating Partition From an Observed Time Series and Its Application to Anomaly Detection," Neural Comput., **30**(9), pp. 2500–2529.
[21] Dupont, P., Denis, F., and Esposito, Y., 2005, "Links Between Probabilistic Automata and Hidden Markov Models: Probability Distributions, Learning Models and Induction Algorithms," Pattern Recognit., **38**(9), pp. 1349 –1371.
[22] Vidal, E., Thollard, F., de la Higuera, C., Casacuberta, F., and Carrasco, R., 2005, "Probabilistic Finite-State machines—Part I and Part II," IEEE Trans. Pattern Anal. Mach. Intel., **27**(7), pp. 1013–1039.
[23] Mukherjee, K., and Ray, A., 2014, "State Splitting and Merging in Probabilistic Finite State Automata for Signal Representation and Analysis," Signal Process., **104**, pp. 105–119.
[24] Rao, C., Ray, A., Sarkar, S., and Yasar, M., 2009, "Review and Comparative Evaluation of Symbolic Dynamic Filtering for Detection of Anomaly Patterns," Signal, Image, Video Process., **3**(2), pp. 101–114.
[25] Bahrampour, S., Ray, A., Sarkar, S., Damarla, T., and Nasrabadi, N., 2013, "Performance Comparison of Feature Extraction Algorithms for Target Detection and Classification," Pattern Recognit. Lett., **34**(16), pp. 2126–2134.
[26] Hajek, B., 2015, *Random Processes for Engineers*, Cambridge University Press, Cambridge, UK.
[27] Beck, C., and Schlögl, F., 1993, *Thermodynamics of Chaotic Systems: An Introduction*, Cambridge University Press, Cambridge, UK.
[28] Halmos, P., 2017, *Lectures on Ergodic Theory* (Dover Books on Mathematics), Dover Publications, New York.
[29] Aeran, A., Siriwardane, S., Mikkelsen, O., and Langen, I., 2017, "A New Nonlinear Fatigue Damage Model Based Only on *s-n* Curve Parameters," Int. J. Fatigue, **103**, pp. 327–341.
[30] Walters, P., 2000, *An Introduction to Ergodic Theory* (Graduate Texts in Mathematics), Springer, New York.
[31] Billingsley, P., 1965, *Ergodic Theory and Information*, Wiley, New York.
[32] Lasota, A., and Mackey, M., 1994, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed., Springer-Verlag, New York.
[33] Cornfeld, I., Fomin, S., and Sinai, Y., 1982, *Ergodic Theory*, Springer-Verlag, New York.
[34] Keynes, H. B., and Robertson, J. B., 1969, "Eigenvalue Theorems in Topological Transformation Groups," Trans. Am. Math. Soc., **139**(1969), pp. 359–369.
[35] Berman, A., and Plemmons, R., 1994, *Nonnegative Matrices in the Mathematical Sciences*, SIAM Press, Philadelphia, PA.
[36] Murphy, K., 2012, *Machine Learning: A Probabilistic Perspective*, 1st ed., The MIT Press, Cambridge, MA.
[37] Cover, T., and Thomas, J., 2006, *Elements of Information Theory*, Wiley-Interscience, Hoboken, NJ.
[38] McDonough, R. N., and Whalen, A. D., 1995, *Detection of Signals in Noise*, 2nd ed., Academic Press, Boston, MA.
[39] Bishop, C. M., 2006, *Pattern Recognition and Machine Learning*, Springer, New York.
[40] Pastor, D., and Nguyen, Q.-T., 2013, "Random Distortion Testing and Optimality of Thresholding Tests," IEEE Trans. Signal Process., **61**(16), pp. 4161–4171.
[41] Hauser, M., Fu, Y., Phoha, S., and Ray, A., 2018, "Neural Probabilistic Forecasting of Symbolic Sequences With Long Short-Term Memory," ASME J. Dyn. Sys. Meas. Control, **140**(3), p. 084502.
[42] Schmid, P., August 2010, "Dynamic Mode Decomposition of Numerical and Experimental Data," J. Fluid Mech., **656**(2), pp. 5–28.