Mechanical Systems and Signal Processing 147 (2021) 107109

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

HMM conditional-likelihood based change detection with strict delay tolerance

David J. Miller [a,*], Najah F. Ghalyan [b,c], Sudeepta Mondal [c], Asok Ray [d]

[a] School of Electrical Engineering and Computer Science, Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Mechanical Engineering, The University of Kerbala, Kerbala 56001, Iraq
[c] Department of Mechanical Engineering, Pennsylvania State University, University Park, PA 16802 USA
[d] Departments of Mechanical Engineering and Mathematics, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 13 July 2019
Received in revised form 1 June 2020
Accepted 28 June 2020

Keywords:
Anomaly detection
Change point detection
Hidden Markov modeling
Conditional likelihood
Forward recursion
Delay tolerance
Fatigue failure
Combustion instability

ABSTRACT

Hidden Markov models (HMMs) have been widely used for anomaly and change point detection due to their representation power and computational efficiency in capturing statistical dependencies in time series. However, often information is integrated over relatively long observation windows, with detections made when the observed sequence's likelihood under the (null) HMM deviates significantly from its typical range. Three related limitations are: i) use of long windows entails large decision delay, which may e.g. fail to prevent machine failure/damage; ii) typical approaches do not narrowly identify an interval within which the change point occurred. Such information could be useful e.g. for process control, where one wants to know how long it takes for control inputs to induce desired change points; iii) The decision statistic is usually the likelihood of the data in the current window, without consideration of past observations. This is suboptimal – this likelihood should be conditioned on past observations to optimally account for statistical dependency in the time series. In this paper, we propose a framework for change point detection which overcomes all of these limitations: i) it applies a standard HMM Forward recursion, but used to evaluate the likelihood of an observation subsequence conditioned on the subsequence's entire past. This approach is used to efficiently evaluate the conditional likelihoods of all intervals of fixed length (hence with fixed delay, d), until a change point is first detected. Here d is a design parameter whose proper value (needed to have a quick response/mitigate damage) may be known for a given application domain; ii) the algorithm narrowly estimates the interval within which a detected change point lies; iii) we propose a novel performance criterion well-matched to low-delay, narrowly localized change point detection – the true detection interval rate (TDIR) – and also evaluate the false positive rate (FPR) and the bias and variance of the estimated change point, all as a function of d. The proposed method is shown to outperform a CUSUM algorithm, symbolic time series analysis (STSA) methods, and a standard HMM method (evaluating the unconditioned likelihood) for instability onset in combustion systems and fatigue failure initiation in a material.

* Corresponding author.
E-mail address: djmiller@engr.psu.edu (D.J. Miller).

https://doi.org/10.1016/j.ymssp.2020.107109
0888-3270/

1. Introduction

1.1. Detection literature review

Change point detection [1] has important applications e.g. to speech recognition, radar systems, condition-based maintenance, fatigue failure detection, geology signals, and even to DNA analysis and malware detection (see, e.g., [2,3] and references therein). If the distribution of observations before a change point occurs (the null distribution) and the distribution after a change point are both known, the problem can be treated e.g. as a standard (two-class) hypothesis testing problem. However, as is more realistic in practice, the distribution after change is typically unknown. In this case, change point detection involves choosing a cost function that measures goodness-of-fit of the generated time series to the null model, and identifying signal segments for which the goodness of fit falls below a threshold. One of the standard methods is the cumulative sum (CUSUM) technique, developed by Page [4], which has been widely used [5–7]. CUSUM is a sequential method which involves the calculation of a cumulative statistic, thresholded to decide whether a change has occurred or not. The decision threshold can be chosen to fix the false alarm rate based on a collection of time series for which it is known that no change point has occurred. In another approach, Bai and Perron [8] considered a time series generated by a linear regression model that undergoes multiple structural changes (breaks) at unknown times with the goodness of fit the sum of squared residuals between the regression model outputs and the observed time series. Later, in [9] they addressed the problem of estimating the change points, introducing a computationally-efficient algorithm that finds the global minimum of the sum of squared residuals using dynamic programming. Bai [10] also proposed a likelihood-ratio model that can detect multiple structural changes in (generally non-stationary) regression models based on hypothesis testing. The null hypothesis is that there are l (known or estimated) breaks in the time series, while the alternative hypothesis is that there are $l+1$ breaks. When $l=0$ a single break/change point is sought.

The approach in [9] considers the entire time series in making change point detections and thus imposes no constraint on the amount of allowed delay in making detections. In many applications, detection needs to be achieved with low delay and/or using short observation windows, e.g. in order to trigger machine damage mitigation. If the true change point occurs at time τ, detection should occur using observations only up to time $\tau+d$, where d, a delay tolerance design parameter, is chosen sufficiently small that detection can result in e.g. expedient damage mitigation. The maximum tolerable value of d may in fact be known for a given application based on past experience [11]. While some detection algorithms such as CUSUM possess optimality properties in minimizing the expected detection delay (given a fixed false positive detection rate) [6,12], this is not the same as the strict delay requirement (d) considered here – achieving an average detection delay of d, or even a value significantly below d, does not ensure that the probability of detection delay exceeding d is small. Moreover, such optimality results assume something is known about the distribution after change, e.g. that the distribution parametric form is the same, post-change, but with a change in the mean parameter, or the variance parameter. More generally, nothing is known about the post-change distribution. Moreover, since d may be very small, a generalized likelihood ratio detection framework, where one estimates the putative distribution post-change and uses it in a likelihood ratio test, may not be feasible, since there will in general be too few observations (d) post-change to accurately estimate the post-change distribution.

Various scenarios are considered in the change point detection literature. First, as aforementioned, one may have knowledge of the observed data distribution both before and after a change has occurred [6]. We do not assume such information here. Second, there may be at most one change point or possibly multiple change points (each possibly to different states, with different data distributions). In the latter case, if there is no delay requirement, one can view the problem as signal segmentation given the whole observed time series [6], as is done in [9]. In [11], the authors distinguish a quickest change detection (QCD) problem and a transient change detection (TCD) problem. In the former, the duration of the change state is infinite, whereas in the latter, the duration of the change state is finite, before a return to the null/normal state or perhaps to a failure state. In this latter (TCD) problem setting, for particular applications, strict detection delay tolerance is needed, while strict delay tolerance is not usually considered in the QCD setting. Some works that establish theoretically optimal detectors, minimizing average detection delay, again based on knowledge of what statistics are changing, include [13,14]. In this work, we address detection problems where the change state duration may be finite or infinite; regardless, we impose strict delay tolerance in order to ensure that true detections will allow damage mitigation. More specifically, we consider a non-Bayesian scenario[1] that well-captures certain machine/process failure for condition-based maintenance applications: the machine is first operating in a normal state and may continue to do so for its entire operating cycle. Alternatively, at some point, there may be a change to an abnormal state. This abnormal state may itself be a system fault/failure state or it may be a transient state prior to entering fault/failure. Either way, there is no transition back to the normal state. The former case is a QCD problem and the latter essentially a TCD one because the duration of the (transient) change state is finite.[2] Irrespective, strict detection delay tolerance may be needed to ensure that damage mitigation (or recovery of normal state) is possible.

[1] The change point treated as deterministic but unknown, i.e. even if the change point is random we have no knowledge of its distribution.

[2] It may be possible to force the system back to the normal state if a proper control action is made in a timely fashion (although in this paper we are not addressing such control actions, and therefore in our case the final transition is to a failure state, not back to the normal state).

1.2. STSA-based detection

Symbolic time series analysis (STSA) has been efficiently used for anomaly detection [15–19]. In STSA, the time series is converted to a symbol sequence from a finite-cardinality alphabet $\mathscr{A}$. The resulting discrete-valued sequence is modeled as a Probabilistic Finite State Automata [20] (PFSA) that mirrors (tracks) the original discrete-time dynamical system [21]. The PFSA states are concatenations of symbols from $\mathscr{A}$. Anomaly detection using STSA is done by first training a null PFSA on data generated from the nominal phase. The resulting states' stationary probability vector is used as a null phase "reference". Once new data arrives, a new PFSA is constructed, whose states' stationary probability vector is computed and compared with that of the null. Based on the Kullback–Leibler "distance" between these two vectors, an anomaly is declared if this distance exceeds a user-defined threshold. One main drawback of STSA anomaly detection is that it discards information in the initial discretization (quantization) step that yields a (finite cardinality) symbol sequence.

1.3. HMM-based detection

Alternatively, Hidden Markov models (HMMs) have been widely used for both change point and anomaly detection (AD) in diverse applications (see e.g. [22–35]). In contrast to STSA, HMMs exploit a latent discrete symbol sequence representation that does not require (hard, information-lossy) quantization of the time series. Moreover, in HMMs one sums over all possible discrete state sequences, in evaluating the log-likelihood fit of the observed data to the (null) model, unlike in STSA – an HMM's joint probability (or joint likelihood) for an observation sequence $x_{n:n+M} = (x_n, x_{n+1}, \ldots, x_{n+M})$, starting at time n, is found by marginalizing over the set of all possible hidden state sequences. This computation is made efficiently via the iterative Forward algorithm [36], which can be used to assign likelihoods to observation sequences in an on-line manner, as each observation arrives. In an AD setting the (null) HMM model is trained using observed time series known to represent normal behavior (using the same (null phase) data used to estimate a null STSA model). Then, when the HMM is applied operationally, if a change point occurs within an observation sequence, the likelihood under the HMM of an observed subsequence that contains the change point is expected to deviate significantly from the typical likelihood [37], once enough observations following the change point are included in the subsequence.[3] Based on a properly chosen threshold (fixed e.g. to a specified false detection rate), one can decide whether change has occurred within a given subsequence. Using HMM inference to narrow down the change point in a long time series to within a small interval requires computing the joint-likelihoods of many possible observation subsequences. However, these many evaluations can be efficiently made by exploiting the standard HMM Forward recursion. Therefore, HMMs have been efficiently used for anomaly detection with the observation subsequence likelihood the decision statistic [22,38].

1.4. Low-delay detection

In this paper, we introduce a detection algorithm specifically designed to make strictly low-delay detections and to accurately identify a narrow interval within which the true change point is estimated to lie. The algorithm exploits the standard Forward recursion for HMM inference to successively evaluate the likelihoods of the observation subsequences $x_{n:n+d-1} = (x_n, \ldots, x_{n+d-1})$, conditioned on $(x_1, x_2, \ldots, x_{n-1}), n = 1, 2, \ldots$, until a change point is detected, where d is a detection delay parameter that controls the subsequence size and which is user-chosen, based on the delay tolerance of the application. Note that there is a tradeoff in the choice of d – if d is very small, only very low delay detections are tolerated, but the true detection rate may be too low (due to an insufficient number of post-change samples in the window). On the other hand, choosing larger d tends to increase the true detection rate, but may introduce an unacceptable delay in making detections. Note that past works such as [25–28,34,35] use HHMs with a sliding window for detection. However they do not condition on the past, in evaluating an observation subsequence's likelihood, for detection/inference purposes. In Appendix A, we theoretically support our use of the conditional likelihood for detection by proving that use of the true Bayes class posterior, conditioning on all available observations, is optimal not only in the well-known sense of minimum expected classification error but also in the sense of maximum true detection rate given a fixed false positive rate. Moreover, we prove that the true detection rate for such a detector, given fixed false positive rate, is non-decreasing as d is increased. Note also that evaluating likelihoods for a fixed window size, d, rather than for a growing window starting at the beginning of the time series, is expected to yield both quicker detection (many "normal" samples will "contaminate" a growing window's anomaly decision statistic, which should lead to delayed detections) and narrow identification of the change point interval.

1.5. Novel performance criteria

Furthermore, for our strictly low-delay setting, care is needed in defining detection events. Specifically, suppose the true change point occurs at time t_c. Then, we define a true detection interval event as one wherein the change is detected at a time t (based on observations up to t) such that $t \geqslant t_c$ and $t \leqslant t_c + d$, i.e. detections that are either too premature or too late

[3] Note that this use of a time-domain window of observations is commonly applied, both for detection as well as for estimation, i.e. Wiener filter smoothing, to estimate one random process given a window of observations from a statistically related process.

are not considered to be true. Also, consider a time series (corresponding to a machine cycle) that does not contain a change point. If a change point is detected anywhere in this time series (which is finite-length for a given machine's cycle, but may still be quite long), this we consider to be a false positive detection event. This is in contrast to a definition of false positives that includes events involving time series that do contain a change point, but where the detection is made too prematurely, or too late. Our false positive rate is meaningful in real applications because detection of even a single change may lead to a (costly) interruption of the machine's cycle, which is warranted only if the detection is likely to be that of a true change point.

Accordingly, we propose to evaluate several criteria that give a rich characterization of low-delay tolerant change point detection performance. Specifically, as a function of delay tolerance, d, we propose to evaluate: 1) bias of the estimated change point; 2) variance of the estimated change point; 3) We define the event that an estimated interval contains the true change point. Accordingly, we then evaluate a novel criterion proposed here, the true detection interval rate (TDIR), the probability that the true change point lies in the estimated interval. TDIR is a more demanding measure of performance than the true detection rate (TDR), which for successful detection would simply require a detection to be made after the change point has occurred; 4) Finally, we evaluate the false positive rate (FPR), the probability that a detection is made when in fact there is no change point in the given time series.

Contributions of the paper: The main contributions of the paper are:

- Detection algorithm: A novel HMM-based detection algorithm is developed that makes strictly low-delay change point detections and identifies a narrow interval within which the true change point is estimated to lie. The algorithm applies the standard HMM Forward recursion, but to evaluate the conditional likelihoods of all successive observation subsequences of length d, given the entire past of the time series relative to the subsequence, with the algorithm terminated either when a detection is first made or when the observation sequence has been exhausted with no detection made.
- Detection evaluation criteria: Criteria for evaluating performance in our strict delay setting, with at most one change point, are proposed, including the novel TDIR, FPR, and bias and variance in estimation of the change point, all as a function of delay tolerance (d) and the associated observation window length (also d).
- Evaluation on real-life application domains: Performance of the proposed algorithm is experimentally validated and compared with a CUSUM, STSA, and standard HMM detection techniques [38,22] on two different applications: detection of combustion instability onset and detection of fatigue failure start in polycrystalline alloys. The results for both applications show excellent performance of the proposed scheme to detect and estimate the change point and to define a narrow interval within which the change point lies. Substantial performance gains, with respect to all four criteria, are achieved compared with the abovementioned detection methods.

Organization of the paper: Section 2 introduces an algorithm for change point detection and performance criteria suitable for low-delay detection. Section 3 experimentally validates the proposed scheme in comparison with several popular methods. Section 4 summarizes and concludes the paper along with recommendations for future research. Appendix A presents a proof that theoretically supports our use of the conditional likelihood for detection and that detection performance improves with increasing d.

2. A conditional likelihood based HMM change point detection algorithm

2.1. Conditional likelihood calculation

Consider a null first-order HMM (learned using the standard Baum-Welch re-estimation algorithm for HMMs [39] based on a null time series, containing "normal" observations with no change point), described by $\Lambda = \{\{\pi_i\}, \{a_{j/i}\}, \{b_j(x)\}\}$, where $\{\pi_i\}$ is the initial state probability vector, $\{a_{j/i}\}$ is the state-transition probability matrix, and $\{b_j(x)\}$ are the state-conditional observation densities. Consider a new time series $x_{1:N} = (x_1, \ldots, x_N)$ being operationally interrogated for possible detection of a change point. Let $s_{1:N} = \{s_1, s_2, \ldots, s_N\}$ be a particular hidden state sequence realization that could have been used to generate the given observed time series, under the null HMM model. Recall the standard Forward variable for HMMs [39], i.e. $\alpha_n[i] = P[x_{1:n}, S_n = i; \Lambda], i = 1, \ldots, L, L$ the number of HMM states. Here, S_n is a random variable representing the state at time n and i is its (state) realization. These Forward variables can be recursively computed and used to evaluate the joint likelihood of an observation sequence, as follows [39][4]:

- Initialization:

$$\alpha_1[i] = \pi_i b_i(x_1), \ i = 1, \ldots, L \tag{1}$$

- Recursion:

[4] To avoid numerical underflow, we apply the scaling procedure for the Forward algorithm suggested in [39], which does not introduce any loss in precision, in computing the log-likelihood.

$$\alpha_{n+1}[j] = \left[\sum_{i=1}^{L} \alpha_n[i] a_{j/i}\right] b_j(x_{n+1}) \quad (2)$$
$$j = 1, \ldots, L, \; n = 1, \ldots, N-1$$

- Termination:

$$P[x_1, x_2, \ldots, x_N; \Lambda] = \sum_{i=1}^{L} \alpha_N(i) \quad (3)$$

Now consider the potential application of this Forward algorithm to detection of a change point within a given subsequence $x_{n:n+d-1}$. An approach that has been commonly applied in prior works [38,22] is to make the approximation that the time series in fact begins at time n. Thus, the Forward recursion above is run, but with the initialization step replaced by: $\hat{\alpha}_n[i] = \pi_i b_i(x_n), i = 1, \ldots, L$. Here, we use $\hat{}$ to denote the fact that an approximate Forward variable is being computed. The result of running the Forward recursion starting from this approximate initialization at time n is an approximation of the joint likelihood: $\widehat{P}[x_{n:n+d-1}; \Lambda] = \sum_{i=1}^{L} \hat{\alpha}_{n+d-1}[i]$. Use of this in making detections will be referred to as standard HMM inference for detection.

There are in fact two limitations of this standard approach. First, $\widehat{P}[x_{n:n+d-1}; \Lambda]$ is clearly only an approximation of the joint likelihood – the true joint likelihood can only be obtained by marginalizing out the random variables $X_1, \ldots, X_{n-1}$ from $P[X_1, \ldots, X_{n-1}, x_n, x_{n+1}, \ldots, x_{n+d-1}; \Lambda]$. In general, this is analytically intractable. Second, it would in fact be suboptimal to perform such marginalization (see Appendix A). Rather, it is optimal to account for the observed past by conditioning on it, i.e. evaluating the conditional likelihood $P[x_{n:n+d-1} | x_{1:n-1}; \Lambda]$. This conditional likelihood can in fact be efficiently evaluated. Note in particular that

$$P[x_{n:n+d-1} | x_{1:n-1}; \Lambda] = \frac{P[x_{1:n+d-1}; \Lambda]}{P[x_{1:n-1}; \Lambda]} = \frac{\sum_{i=1}^{L} \alpha_{n+d-1}[i]}{\sum_{i=1}^{L} \alpha_{n-1}[i]}, \quad (4)$$

The latter equality is achieved simply by realizing that the joint likelihoods in the numerator and denominator are efficiently computed by running the Forward HMM recursion described above in (1)–(3), i.e. starting from the beginning of the time series, and terminating at the appropriate times for the numerator and denominator likelihood calculations, respectively. We have seen little prior work using the conditional likelihood for HMM-based inference and HMM-based anomaly detection.

2.2. Detection algorithm based on the conditional likelihood

Consider a time series $x_{1:N}$, and assume there is at most one change point. Such an assumption is reasonable in many real-life applications, such as fault and machine failure detection. In these examples, it is valuable to be able, with low delay, to detect the beginning of the change from nominal behavior so as to make a proper action to avoid further consequences of such change. Therefore, we want to detect the start of the change, which corresponds to a single time point within the time series.

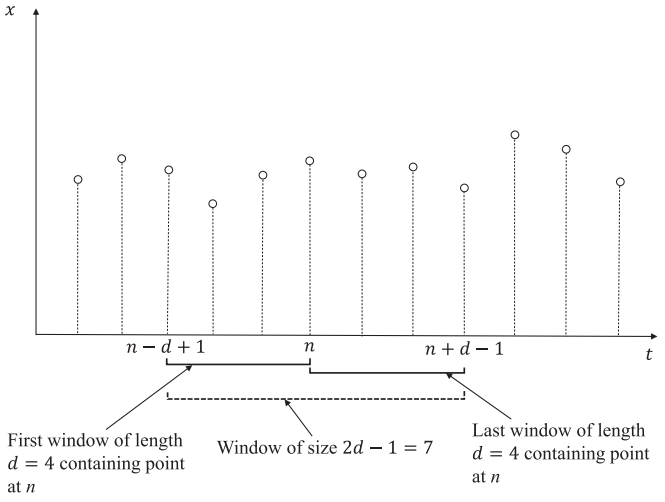

Fig. 1. Schematic representation of the sliding window in Algorithm 1, illustrating d subintervals of length d containing a given candidate change point, n.

To do so for allowed delay tolerance d, we consider intervals of length d that contain a candidate change time point, n. There are d such intervals, as indicated in Fig. 1. We evaluate the conditional log-likelihood of each such interval, using (4), and make a detection at time point n if the minimum absolute deviation from a (null) reference log-likelihood value, over all the d intervals, exceeds a threshold. Thresholding the minimum deviation is a conservative detection rule, with the resulting algorithm called HMM$^{(2)}$-min. At the other extreme, an aggressive detection rule would be to threshold the maximum of the d deviations, with the resulting algorithm called HMM$^{(2)}$-max. Both such detection rules are in fact evaluated in the sequel. If a detection is made, the detection interval is chosen as $[n-d+1, n+d-1]$ and the change point itself is estimated as n. If no detection is made, the window (in Fig. 1), centered on n, is slid one position to the right, with the new candidate change point as $n+1$. Pseudocode for our detection procedure is given below in Algorithm 1.

Algorithm 1 Proposed Detection Approach

Input: Time series $x_{1:N}$, a null HMM Λ, delay tolerance d, null (reference) normalized (by d) log-likelihood $\bar{L}$, and threshold δ.

Output: Either a "no change point" decision ($z_d = 0$) or a "change point detected" decision ($z_d = 1$) and the estimated change point $\hat{n}$, which determines the estimated interval $[\hat{n}-d+1, \hat{n}+d-1]$.

1: $z_d \leftarrow 0$, $n \leftarrow d$;
2: Run the Forward recursion to compute $\alpha_1[i], \ldots, \alpha_{2d-2}[i], i = 1, \ldots, L$.
3: Set $\sum_{i=1}^{L} \alpha_0[i] = 1$.
4: **do**
5: Take one Forward recursion step to compute $\alpha_{n+d-1}[i],\ i = 1, \ldots, L$.
6: Compute $\log P\left[x_{n-d+1+j:n+j} | x_{1:n-d+j}; \Lambda\right] = \log\left(\frac{\sum_{i=1}^{L} \alpha_{n+j}[i]}{\sum_{i=1}^{L} \alpha_{n-d+j}[i]}\right),\ j = 0, 1, \ldots, d-1$.
7: $D \leftarrow \min_{j \in \{0,1,\ldots,d-1\}} \left| \frac{1}{d} \log P\left[x_{n-d+1+j:n+j} | x_{1:n-d+j}; \Lambda\right] - \bar{L} \right|$.
8: **if** $D > \delta$ **then**
9: $\hat{n} \leftarrow n$
10: Estimated interval is: $[\hat{n}-d+1, \hat{n}+d-1]$.
11: $z_d \leftarrow 1$
12: **end if**
13: $n \leftarrow n+1$
14: **while** $((n < N)$ AND $(z_d == 0))$

Comments:

1. The algorithm starts by considering the candidate change point $n = d$, with associated observation window $[1, \ldots, 2d-1]$. The first interval evaluated for this window is $[1, \ldots, d]$. For this interval, there are no past observations. Thus, the "conditional" likelihood is actually the unconditioned likelihood $P[x_{1:d}; \Lambda]$. The algorithm computes this likelihood by using $\sum_{i=1}^{L} \alpha_0[i]$ in the denominator term in (4), which in fact equals 1.
2. For each candidate change point, just one new forward recursion step is needed. Thus, the detection algorithm complexity scales with the complexity of the Forward algorithm which, for a sequence of length M, is $O\left(L^2 M\right)$.
3. The reference normalized log-likelihood $\bar{L}$ can be computed in different ways, customized to the application domain of interest: i) We can simply set $\bar{L} = 0$ if we expect the conditional likelihood of the sliding interval to significantly change once the change point is involved in the interval, ii) We can compute an ensemble average $\bar{L}$ from multiple "normal" time series that are known to contain no change point. Moreover, since this reference average log-likelihood may be a function of both n and d, we can compute an ensemble average value for each (n, d) pair; iii) $\bar{L}$ can be time series realization dependent and can again be a function of n and d (for example, it can be evaluated by locally, temporally averaging the log-likelihoods for each of the d time intervals of length d in the window $[n-d+1, n+d-1]$), for all values n that are "early enough" such that it is known that $[1, \ldots, n]$ does not contain a change point; iv) It can again be time series dependent and a function of d but not a function of n. In this case, we assume that $\bar{L}$ does not change much, as the past context (as n) increases; v) For each time series and at any given time n, we can compute $\bar{L}$ as the time-average of the conditional likelihoods of all intervals of length d up to (but not including) time n (i.e., including all times up until the present that have been rejected as change points). In the sequel we will exposit how we chose $\bar{L}$ in our experiments.
4. We use the absolute deviation from a null reference log-likelihood because for some applications anomalies will be associated with lower likelihoods than typical under the null whereas, for others, anomalies will actually yield higher likelihoods than are null-typical. This will be seen in our fatigue failure experiments.
5. The detection threshold δ can be set in one of several ways, e.g. to control the FPR for a given d.

3. Experimental results

In this section we validate the proposed HMM-based detection method using experiments from two different domains: instabilities in combustion systems and fatigue failure in a polycrystalline alloy material. In both cases we have time series generated from experiments with failures (instabilities) and with no failures (no instabilities). For simplicity, we assume there are two hidden (null) HMM states and we use Gaussian mixtures for the state-conditional density functions. The number of mixture components for each state is selected using the Bayesian Information Criterion (BIC) [40]. We also apply a CUSUM algorithm and the symbolic time series analysis technique (STSA) [41] with K-means and maximum entropy partitioning (MEP) [42] used to perform the symbolization (quantization) required by STSA. Finally, we compare with standard use of HMMs for detection, evaluating the joint likelihood of observations in the window [38,22], as described in 2.1, rather than the conditional likelihood. We will compare the results for these methods with our proposed method.

In this paper, we evaluate an STSA change point algorithm that mimics the structure of Algorithm 1. Specifically, a null time series (known to contain exclusively "normal" observations) is quantized into a discrete symbol sequence (e.g., using the K-means algorithm), which is then used to estimate a null PFSA. Then, operationally, for any new (test) window of observations, a new PFSA is estimated using the same quantizer partition as was used to construct the null PFSA. Therefore, we have a null PFSA (and its steady state probability vector), which was estimated based on an observation window for which it is known no anomaly is present, and a new PFSA, which is estimated based on a test observation window. The states' stationary probability vector at a given time n is obtained from the new PFSA at time n, and a surrogate for the log-likelihood deviation defined in Algorithm 1 is obtained by computing the Kullback–Leibler (KL) divergence between the new probability vector at time n and the probability vector obtained from the null PFSA. Then, mirroring the HMM-based algorithm, we find the minimum KL value over all d windows of length d that contain the point n and compare this to a detection threshold. Hence, for STSA, we apply the proposed Algorithm 1 structure (unlike previous change point detection algorithms based on STSA), but based on the KL measure rather than the absolute log-likelihood deviation. The KL measure is expected to increase when an anomaly occurs.

3.1. Detection of thermoacoustic instability onset in combustion systems

Thermoacoustic instabilities (TAI) in combustion systems are usually caused by spontaneous excitation of one or more natural modes of acoustic waves [43]. TAI are typically manifested by large-amplitude self-sustained chaotic pressure oscillations in the combustion chamber [44], which may lead to damage in mechanical structures if the pressure oscillations match one of the natural frequencies of the system. Traditional techniques of pressure-oscillation measurement for TAI detection, reported in the literature [45,46], attempt to extract the growth-rate information using the entire envelope of pressure oscillations. These techniques thus observe the entire acoustic time series and may not be suitable for online estimation of transient growth of oscillations and early detection of instability. Moreover, the time scales of TAI are on the order of milliseconds, which, given practical sampling rates, mandates an algorithm that can accurately detect an onset of TAI based on short-length sensor data.

The detection approach given by Algorithm 1 is our proposed candidate for such detection. The performance of this method is compared here with CUSUM and STSA techniques based on K-means and MEP [42] partitions. We also evaluate the standard HMM method, in which the conditional likelihood $P\left[x_{n-d+1+j:n+j}|x_{1:n-d+j};\Lambda\right]$ used in Algorithm 1 is replaced by the unconditioned likelihood $\widehat{P}\left[x_{n-d+1+j:n+j}|\Lambda\right]$. This approach is consistent with the HMM detection approaches in [38,22]. In the sequel, HMM$^{(1)}$ stands for the conventional HMM-based method, and HMM$^{(2)}$ stands for our conditional-likelihood HMM-based method.

For each experiment, the null HMM Λ is estimated by the Baum-Welch algorithm [36] using the first 1/10-th of the observations generated in the stable phase. Since the observations' variance increases as the system becomes unstable, the log conditional likelihood is expected to significantly decrease. Therefore, we set $\bar{L}=0$. The threshold δ is computed as

$$\delta = \max_{n\in\{1,\ldots,d\}} D(n) + \epsilon \tag{5}$$

where $D(n)$ is the anomaly measure, at time n, used in Algorithm 1, and ϵ is a hyperparameter chosen by splitting the experiments into two equal-sized sets; training and test set experiments. We evaluate a grid of candidate ϵ values and pick the one which achieves the best TDIR performance over the training experiments. Then, the performance of the proposed algorithm is evaluated (for this ϵ choice) using only the time series in the test set. For fair comparison, the same method is used for computing the threshold for the STSA techniques. Moreover, for each specimen, the same observations used to train the null HMM are also used to train the STSA null model.

3.1.1. Description of the experimental apparatus

Experiments in this subsection are obtained by using an electrically heated Rijke tube apparatus shown in Fig. 2 (see [34] for more details), where the process starts with a stable combustion that gradually becomes unstable. Experiments have been conducted by varying the air flow rate (Q) and the power input to the heater (E_{in}). A time series of pressure oscillations is collected over 30 s for each experiment, sampled at 8192 Hz and high-pass-filtered to attenuate the effects of low-

frequency environmental acoustics. For each experiment, the setup is heated to steady state with a power input of 200 W to the heater. Then the power input is abruptly raised to a high value, which eventually showed a limit cycle behavior in each experiment. A sample pressure signal is shown in Fig. 3, where instability is indicated by a large amplitude limit cycle. As shown in the figure, the pressure signal starts with a relatively small amplitude, and once the process starts becoming unstable the amplitude starts growing due to the onset of a Hopf bifurcation in the system [47,45]. This is a TCD problem (as indicated from the figure), with a short transient stage between the stable and unstable stages. Thermoacoustic instability is triggered by changing Q (ranging between 130 LPM and 250 LPM at increments of 20 LPM) and E_{in} (ranging between 800 W and 2000 W at increments of 200 W). For some values of Q and E_{in}, instability does not set in, which is depicted in the stability map of the system, as described in the work by Mondal et al. [34].

The onset time of Hopf bifurcation (m) changes from one experiment to another. To visually identify a unique change time as ground truth, against which we compare our method's estimate, we remove a small block of the time series that includes the point x_m. By doing so, we remove visual uncertainty, creating certitude that all the points in the time series up to the beginning of the removed block are from the stable phase, with points after the removed block from the unstable phase. However, the block removed is on the order of 100 points, while the total transient block is on the order of $10,000$ points; thus the removed block does not significantly change the gradual transition from stable to unstable combustion. Therefore we consider the time point immediately after the removed block as the true change time. Since each original pressure signal has millions of sample points, we consider the last $1/10$-th of the signal before the transition from stable to unstable combustion, the whole transient zone, and the first $1/10$-th of the signal after the end of the transient zone. The resulting signal is further downsampled by 100 so that the signal length is reduced while maintaining the main shape and features of the original signal. This is for signals that do contain a transition from stable to unstable combustion. For signals with purely stable combustion, we consider a $1/10$-th segment of the signal and downsample by 100. We use the same process for the ultrasonic signals considered in the next subsection. Fig. 4 shows samples of the resulting pressure signals after downsampling. Fig. 5 shows a sample pressure signal after downsampling, and a magnification of the transient part, which shows that the transition from stable to unstable combustion is gradual, not abrupt.

3.1.2. Experimental validation

We conducted 145 experiments. In each, there is a single change from stable combustion to the unstable phase, with the change time not fixed across the experiments. Then for each experiment we extract two pressure time series. One is safely confined to cover a sub-interval of the stable phase, with the second one consisting of the full time series. Therefore, we have 290 time series: 145 with transitions from stable to unstable combustion (72 used in the training set for setting ϵ and 73 in the test set), and 145 with purely stable combustion (all in the test set). For each time series, we attempt to make a single change point detection. By applying a detection algorithm to each of the time series, we can measure the TDIR, FPR, and bias and variance of the estimator, for increasing values of the sliding window length d.

We start with the CUSUM method. One standardized version of the CUSUM algorithm returns the first index of the time series $x_{1:N}$ that has drifted k standard deviations outside the nominal mean. This detection approach is suitable for a deviation in mean. However, as shown in Fig. 3, our data shows a change in variance. Therefore, we use a CUSUM variant suitable for change due to variance deviation. Inspired by [48] we introduce the modified CUSUM algorithm described by the following stopping rule:

$$T_s = \inf\{n : |\rho(n)| \geqslant \delta\} \tag{6}$$

$$\rho(n) = \sum_{k=1}^{n} x_k^2 - n\sigma_0^2 \tag{7}$$

$$\sigma_0^2 = \frac{1}{N_0}\sum_{k=1}^{N_0} x_k^2 \tag{8}$$

where T_s is the stopping time at which change detection is declared, N_0 is the number of samples taken from the nominal state to compute σ_0^2, and with the data shifted to have zero mean (so that σ_0^2 is a variance estimate). We tried different values of the threshold δ and picked the one with best TDIR performance. The results are shown in Fig. 6. At each value of the window size, d, the modified CUSUM is applied to the window $x_{n:n+d-1}$ to make a detection, where the window $x_{n:n+d-1}$ is slid from the beginning until the end of the time series $x_{1:N}$, until a detection is made. As shown in Fig. 6, the detection performance generally improves with window size. Furthermore, TDIR achieves its maximum attainable value, 1.0, at $d \sim 425$, and FPR achieves zero for the same value of d. These delay values are much larger than for the comparison methods, as will be seen shortly.

Let us now apply STSA techniques. We consider two STSA methods; the first uses K-means clustering for partitioning and the second uses the maximum entropy partition (MEP) [40,42]. The results for K-means are given in Fig. 7, and for MEP in Fig. 8, both evaluated for different values of the symbol alphabet size (number of clusters, K). As shown in these two figures, STSA performance is much better than modified CUSUM. We notice for both STSA methods how the TDIR converges close to one with window size ~ 120, much less than ~ 425 required for the modified CUSUM. A comparison between the two figures for K-means and MEP shows a little improvement for MEP in TDIR performance. However, the opposite is true for FPR, where

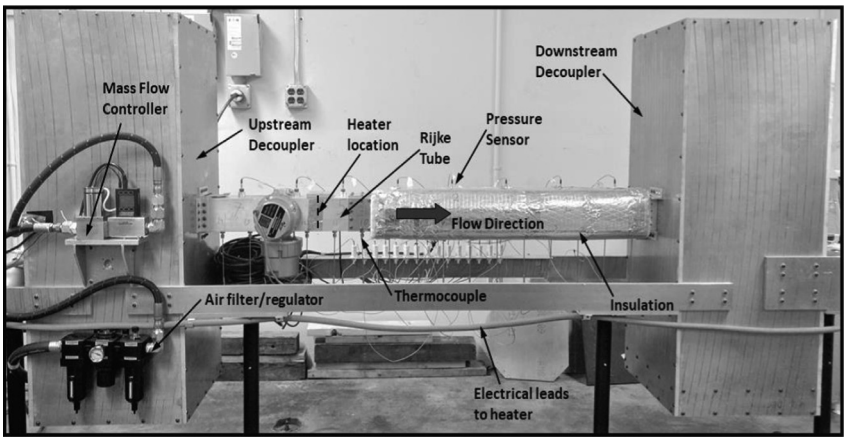

Fig. 2. Rijke Tube combustion apparatus.

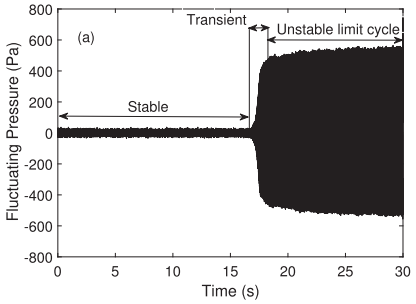

Fig. 3. A pressure signal showing the transition from a stable combustion to unstable limit cycle.

MEP converges to zero at window size $\geqslant 80$ with alphabet sizes 3 and 4, while K-means achieves zero FPR at window size $\geqslant 60$ (with alphabet sizes 3, 4, and 5), both having much better FPR performance compared to the modified CUSUM.

Next, we pick the K-means and MEP solutions with best TDIR performance (i.e., alphabet size = 5 for K-means and 3 for MEP), and compare them with the results obtained for the proposed HMM-based detection method (Algorithm 1), denoted by $\text{HMM}^{(2)}$. The results are given in Fig. 9, which shows excellent results of the proposed HMM scheme for all performance measures: TDIR, FPR, bias and variance of the estimated change time, with consistent improvement achieved over the STSA techniques. Note also that the STSA results in Figs. 7 and 8 are much better than CUSUM in Fig. 6. Thus, the HMM method, which dominates the STSA methods, is also substantially better than CUSUM, with respect to all four performance measures

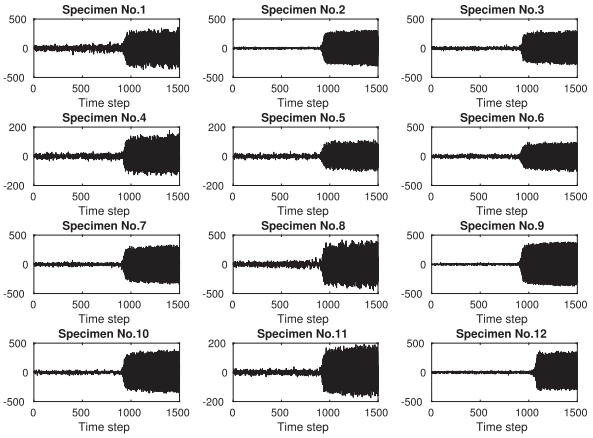

Fig. 4. Samples of pressure signals after downsampling.

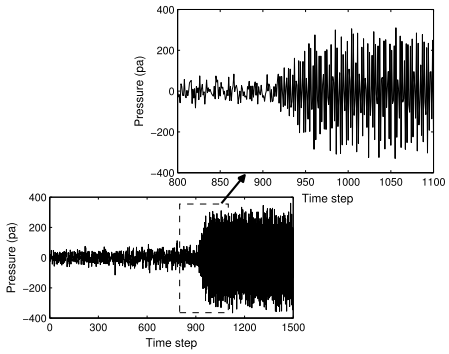

Fig. 5. A sample pressure signal after downsampling, showing the gradual transition from a stable combustion to unstable limit cycle.

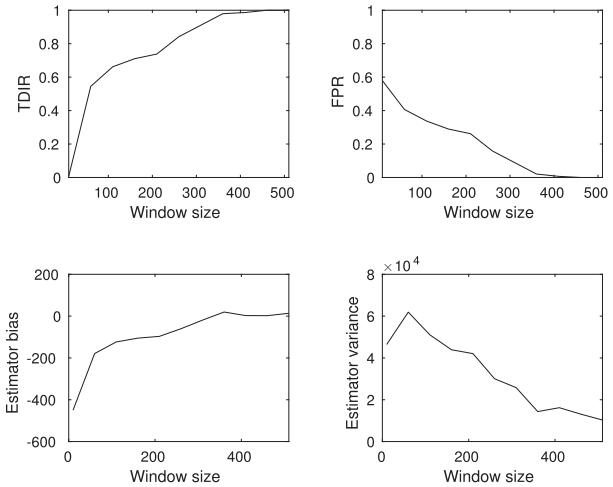

Fig. 6. Detection of thermoacoustic instabilities using modified CUSUM.

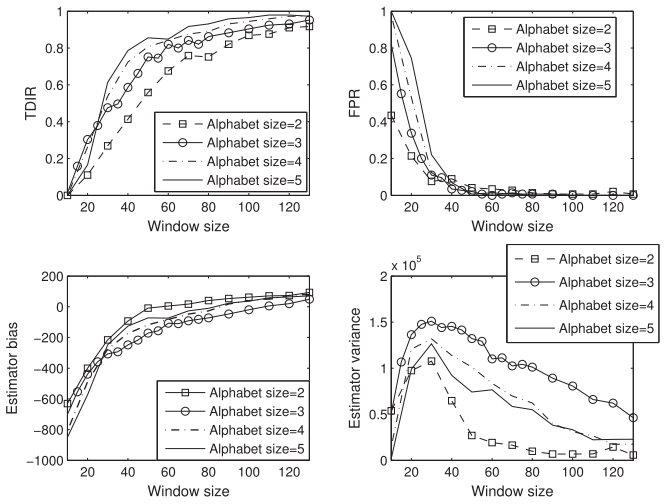

Fig. 7. Detection of thermoacoustic instabilities using STSA with K-means.

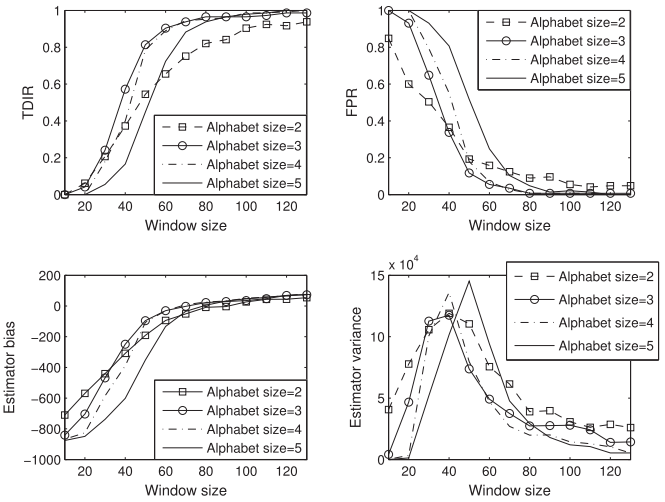

Fig. 8. Detection of thermoacoustic instabilities using STSA with MEP.

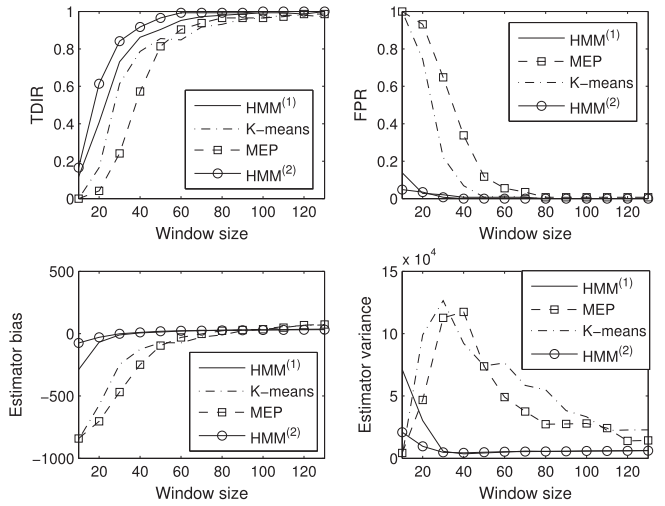

Fig. 9. Detection of thermoacoustic instabilities using K-means (alphabet size = 5), MEP (alphabet size = 3), $HMM^{(1)}$ (number of states = 2), and $HMM^{(2)}$ (number of states = 2).

evaluated. The HMM method's FPR goes to zero much faster than the STSA methods; the HMM method dominates at all window sizes. But most notable is the much better variance characteristic of the HMM method. Also, the figure shows the performance of the standard HMM-based method, $HMM^{(1)}$, which uses the approximate joint subsequence likelihood, i.e. which ignores the past when evaluating the likelihood of a subsequence. Note that the conditional likelihood based approach outperforms the joint likelihood based variant for all window sizes, d, with respect to all four performance measures, but especially with respect to TDIR.

Fig. 10 shows the performance of the proposed HMM-based method with the anomaly measure obtained by using the min (denoted as $HMM^{(2)}$-min) and max (denoted as $HMM^{(2)}$-max) operators (over all subsequences of length d for the current candidate change point). The figure shows generally comparable results, with $HMM^{(2)}$-max giving better performance for TDIR and $HMM^{(2)}$-min giving better performance for FPR, the bias, and variance of the estimated change point. These results are not surprising since 'max' is expected to be a more sensitive (but less specific) statistic than 'min'. Note also in Figs. 6–10 that, for all the methods, TDIR increases, and FPR decreases, with increasing window size, d. This experimental observation is theoretically supported in Appendix A. This suggests one might want to choose d quite large. However, for a given application, d must be restricted in order to achieve low-latency detection and the potential for damage mitigation/response.

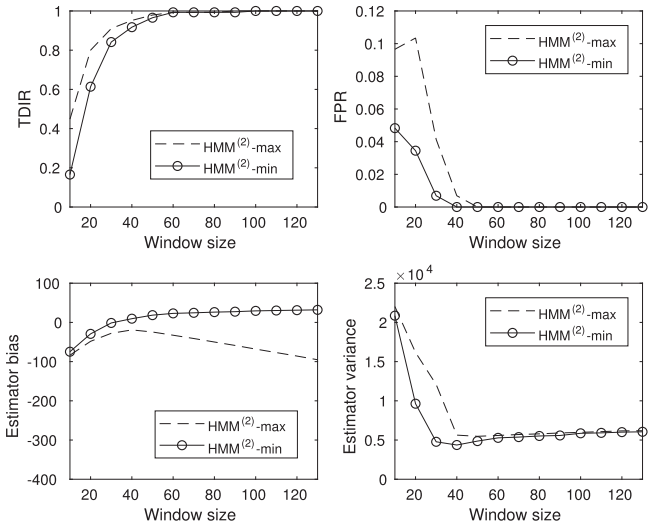

Fig. 10. HMM-based method performance with min and max operators used to compute the anomaly measure.

3.2. Detection of fatigue crack initiation in a polycrystalline alloy material

Modeling fatigue failure in mechanical structures has received great attention by many researchers, e.g. [49,50]. In any mechanical structure, millions of initial materials' defects (such as dislocations, voids, inclusions and slip bands) exist inside the microstructure even before the structure is used. In general, fatigue damage is critically dependent on these initial defects, from which cracks start to nucleate and merge together, generating bigger cracks, leading to the failure of the structure [51]. These initial defects are usually distributed in a highly random fashion, producing large uncertainties in the crack initiation and propagation process even under identical loading. Therefore, fatigue failure is considered an unpredictable and highly stochastic process. In the following, we give a brief description of the apparatus we use to get ultrasonic signals correlated with fatigue damage.

3.2.1. Description of the experimental apparatus

Although structural fatigue damage is not easily measured directly, damage may be correlated with signals that can be measured and used for fatigue damage detection. In this work we use ultrasonic signals, which have been commonly used for real time damage sensing in the aerospace and nuclear power industries to detect flaws in structures [52]. Fig. 11 shows the experimental apparatus, built upon an MTS 831.1 fatigue testing machine, which can be used to apply external load to test specimens with the desired cyclic properties: amplitude, frequency, and the shape of the force function; it is also capable of applying random loading. The other component is the ultrasonic part, which functions by emitting high frequency ultrasonic pulses by using a piezoelectric transducer. A Matec TB1000 Gated Amplifier PC add-in card drives the piezoelectric transducer with a gated sine wave with amplitude of 300 V. The generated signal consists of short bursts of a sine wave of constant amplitude interrupted by relatively long periods of 0 V. This signal propagates through the specimen and is captured by transducers arrayed on the other side of the notch. The received signal is routed through a high frequency selector switch to a National Instruments NI5911 Oscilloscope card in the PC. The acquisition of the ultrasonic signal is synchronized with the load applied to the specimen so that data is acquired at the low load and at the high load (see [53] for more details). When cracks occur, part of the ultrasonic signal will be reflected, and hence the received signal is attenuated. This attenuation increases as more cracks initiate and propagate, until the specimen breaks and only noise signal is left. In this paper, data are collected at a sampling rate $\sim 21,800$ Hz.

As we did for the combustion signals, we remove a small block from the signal at the visually assessed change point. We then consider the last $1/10$-th of the signal just before the transition from healthy to damaged state, the whole transient zone (except for the small removed block), and the first $1/10$-th segment of the signal just after the end of the transient zone. Then the resulting signal is further downsampled by 100 so that the signal length is reduced while maintaining its main shape and features. This is for signals that contain a transition from a healthy to a damaged state. For signals with purely healthy state, we only consider a $1/10$-th segment of the signal, and then downsample by 100. Thus the data are collected at an effective sampling rate ~ 218 Hz.

Fig. 12 shows such ultrasonic signals, after this segmentation and downsampling, for 12 sample specimens made of steel-aluminum alloy. As shown in the figure, the signal begins to significantly attenuate at a certain time instant. Such time instants roughly estimate phase transitions in fatigue damage, where cracks reach critical lengths and the damage process starts growing aggressively (until the specimen breaks and only a noise signal remains). As the figure indicates, this is a QCD

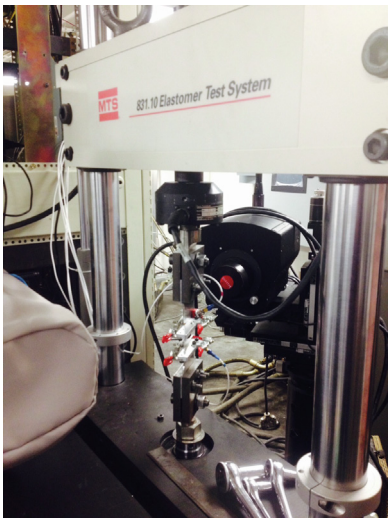

Fig. 11. Apparatus for fatigue failure experiments.

problem, with the system remaining in the transient (increasing damage, indicated by increasing signal attenuation) phase until the specimen breaks (with the signal then flat). Although all the specimens used in this work have the same dimensions and are made from the same material, the plots in Fig. 12 show that the change time is different from one specimen to another. This difference is due to initial microstructural defects which are specimen-dependent.

3.2.2. Experimental validation

We conducted 17 experiments and proceeded in a similar way as for the combustion experiments to generate 34 ultrasonic time series – 17 with transitions from healthy to a gradually damaged state until the specimen breaks (8 used in the training set for setting ϵ and 9 in the test set); another 17 with healthy condition, for which there is no detectable fatigue failure. As we mentioned, the ultrasonic time series obtained here are highly stochastic and highly noisy. Moreover, unlike the pressure signals in the combustion experiments, the ultrasonic signals at the receiver tend to decrease in variance (approaching the null density mean value – zero) as damage accumulates in the specimens. In this situation, unlike for the combustion experiment, the likelihood under the null is expected to (atypically) increase as the data variance decreases, because after the change point the observations tend to hover close to the null mean – this is why we use the absolute deviation in Algorithm 1. It covers both atypically small and atypically large log-likelihoods compared to the null (reference) log-likelihood.

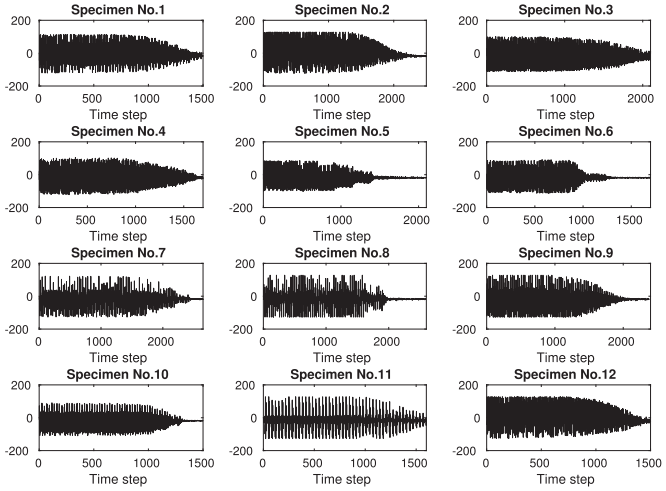

Fig. 12. Ultrasonic signals for sample specimens, with signal strength value on the y-axis and time on the x-axis.

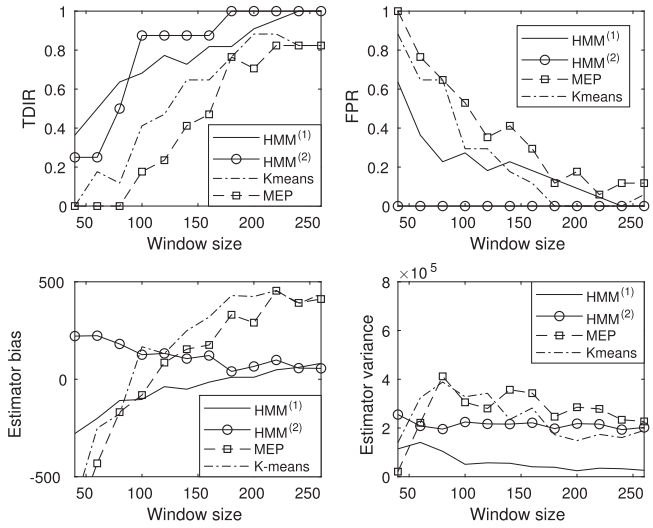

Fig. 13. Detection of fatigue failure onset using K-means (alphabet size = 3), MEP (alphabet size = 4), HMM$^{(1)}$ (number of states = 2), and HMM$^{(2)}$ (number of states = 2).

We applied STSA methods using K-means and MEP for partitioning, and HMM$^{(2)}$ inference with the proposed forward recursion, both combined with the proposed detection approach in Algorithm 1. We also applied the standard HMM method, HMM$^{(1)}$, which uses the approximate joint likelihood instead of the conditional likelihood. Again, for STSA, the surrogate quantity for the deviation in log-likelihood in Algorithm 1 is the Kullback divergence between the current PFSA states' stationary probability vector and the null PFSA states' stationary probability vector [41,42].

We follow the same methods we used in the combustion experiments to compute the null models and the thresholds for both HMM and STSA techniques. However, since the likelihood under the null is expected to increase as the data variance decreases, it is not reasonable to set $\overline{L} = 0$, as we did in the combustion experiments. Alternatively, we compute $\overline{L}$ as follows. For HMM$^{(1)}$, $\overline{L}$ is the mean of the log-likelihoods of the first d intervals of length d, i.e.,

$$\overline{L} = \frac{1}{d}\sum_{n=1}^{d} L(n) \tag{9}$$

However, this choice of $\overline{L}$ is not plausible for HMM$^{(2)}$. In this case, $\overline{L}$ should be time-varying, consistent with the conditional log likelihood depending on the growing past. Therefore, we compute a time-varying $\overline{L}$ as

$$\overline{L}(n) = \begin{cases} \frac{1}{n}\sum_{m=1}^{n} L(m), & \text{if } n \leqslant T_0 \\ \frac{1}{T_0}\sum_{m=1}^{T_0} L(m), & \textit{otherwise} \end{cases} \tag{10}$$

for all $n > 0$, where T_0 is chosen to be well within the normal phase (much less than the true change instant). We used $T_0 = \frac{\tau}{10}$, where τ is the true change instant.

Fig. 13 shows the results for STSA and HMM methods. For K-means and MEP, we used different values of alphabet size, ranging from 2 to 5, and picked the one with the best TDIR performance. Specifically, we used alphabet size = 4 for MEP and 3 for K-means. The results show excellent performance of the HMM algorithms (HMM$^{(1)}$ and HMM$^{(2)}$) in terms of TDIR, FPR, and estimator bias and variance, with consistent improvement achieved over the STSA techniques. Again, it is seen that the conditional likelihood based (proposed) method (HMM$^{(2)}$) overall outperforms the joint likelihood based variant of our approach (HMM$^{(1)}$) with respect to TDIR and FPR. The bias of the two methods is comparable and the variance of HMM$^{(1)}$ is smaller.

4. Conclusion

The proposed strictly low-delay, localizing, HMM conditional-likelihood based transient change detection algorithm has been shown to achieve substantially improved performance over comparison methods with respect to appropriate performance criteria proposed here: a novel TDIR, FPR, the bias in estimating the change point, and the variance of the change point

estimate, all as a function of the delay tolerance parameter d. We also proved, within a supervised setting, that conditioning (as we do) is required for optimal detection and that detection performance is monotonically non-decreasing with increasing window size, d, consistent with our experimental results. Since many systems that involve change point detection involve control inputs, in our future work we may consider the framework of input–output hidden Markov models (IOHMMs) [54] in devising a scheme that chooses control inputs to e.g. maximize TDIR under strict detection delay given a fixed FPR. Such control inputs could also be used to induce desired change points (associated with improved performance, rather than with system failure).

CRediT authorship contribution statement

David J. Miller: Conceptualization, Methodology, Writing - original draft, Supervision. **Najah F. Ghalyan:** Methodology, Software, Validation, Writing - review & editing. **Sudeepta Mondal:** Validation, Writing - review & editing. **Asok Ray:** Supervision, 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.

Acknowledgment

The work reported in this paper has been supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grant Nos. FA9550-17–1-0070, 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. The second author would like to thank the Higher Committee for Education Development (HCED) in Iraq for their financial support.

Appendix A. Proof that i) conditioning on past observations is required for optimality in a (supervised) detection setting and ii) detection performance is monotonically non-decreasing with increasing delay tolerance *d*

In this appendix, to simplify theoretical analysis, we consider detection based on a window of observations $\underline{x}_n = \{x_n, \ldots, x_{n+d-1}\}$. Further, solely for clarity's sake, suppose that, if there is a change point in the time series, it is at x_n, i.e., in this case all the observations in $\{x_n, \ldots, x_{n+d-1}\}$ were generated according to the new (change state) random process. Also define the growing window vector $\underline{x}_n = \{x_0, \ldots, x_{n+d-1}\}$. Denote the joint density of $\underline{X}_n$ under the null hypothesis by $f_{\underline{X}_{n/0}}(.;\Lambda_0)$. The alternative hypothesis is that $\{x_0, \ldots, x_{n-1}\}$ was generated according to the nominal random process and $\{x_n, \ldots, x_{n+d-1}\}$ was generated according to the new process (starting with change point x_n), $f_{\underline{X}_{n/1}}(.;\Lambda_1)$, independent of past observations. Suppose we wish to make optimal (minimum error rate) detection, given knowledge of $f_{\underline{X}_{n/0}}(.)$ and $f_{\underline{X}_{n/1}}(.)$. The error rate can be expressed as:

$$P_e[n] = P_1 \int_{R_0} f_{\underline{X}_{n-d/0}}\left(\underline{z}_{n-d};\Lambda_0\right) f_{\underline{X}_{n/1}}(\underline{z}_n;\Lambda_1) d\underline{z}_n + P_0 \int_{R_1} f_{\underline{X}_{n/0}}\left(\underline{z}_n;\Lambda_0\right) d\underline{z}_n, \tag{11}$$

where $\underline{z}_n$ is a vector of dummy variables of integration and where R_0 is the decision region assigned to the no-change state and R_1 is the decision region assigned to the change state. If the normal and changed state class priors P_1 and P_0 are unknown, they can be set to $[1/2,\ 1/2]$. Using a standard simplification, this is re-expressed as:

$$P_e[n] = P_0 + \int_{R_0} \left(P_1 f_{\underline{X}_{n-d/0}}\left(\underline{z}_{n-d};\Lambda_0\right) f_{\underline{X}_{n/1}}(\underline{z}_n;\Lambda_1) - P_0 f_{\underline{X}_{n/0}}\left(\underline{z}_n;\Lambda_0\right)\right) d\underline{z}_n,$$

All the values for which the integrand is negative should be assigned to R_0, to minimize $P_e[n]$. All the values for which the integrand is positive should be assigned to R_1 (so that they do not make positive contribution to the integral). Thus, we have the Bayes-optimal rule:

$$R_0 = \left[\underline{x} : P_0 f_{\underline{X}_{n/0}}\left(\underline{x};\Lambda_0\right) \geqslant P_1 f_{\underline{X}_{n-d/0}}\left(\underline{x}_{n-d}\right) f_{\underline{X}_{n/1}}(\underline{x}_n;\Lambda_1)\right],$$

with R_1 the complement of this region in $\mathbb{R}^d$. This can be rewritten, cancelling the common term $f_{\underline{X}_{n-d/0}}(.;\Lambda_0)$, as

$$R_0 = \left[\underline{x} : P_0 f_{\underline{X}_n/\underline{X}_{n-d}.0}\left(\underline{x}_n/\underline{x}_{n-d};\Lambda_0\right) \geqslant P_1 f_{\underline{X}_{n/1}}(\underline{x}_n;\Lambda_1)\right].$$

Note that the minimum error rate rule is based on the (conditional) density, considering observations (from 0 to $n-d$) outside the current window.

Now, consider the more relevant (detection) problem, with the goal to minimize the missed detection rate given a fixed false positive rate. That is, the problem:

$$\min_{R_0} P_1 \int_{R_0} f_{\underline{\underline{X}}_{n-d/0}}\left(\underline{\underline{z}}_{n-d};\Lambda_0\right) f_{\underline{X}_{n/1}}(\underline{z}_n;\Lambda_1) d\underline{\underline{z}}_n$$

subject to

$$P_0 \int_{\overline{R}_0} f_{\underline{\underline{X}}_{n/0}}\left(\underline{\underline{z}}_n;\Lambda_0\right) d\underline{\underline{z}}_n = \gamma$$

The associated Lagrangian objective is:

$$L = P_1 \int_{R_0} f_{\underline{\underline{X}}_{n-d/0}}\left(\underline{\underline{z}}_{n-d};\Lambda\right) f_{\underline{X}_{n/1}}(\underline{z}_n;\Lambda_1) d\underline{\underline{z}}_n + \lambda\left(P_0 \int_{\overline{R}_0} f_{\underline{\underline{X}}_{n/0}}\left(\underline{\underline{z}}_n;\Lambda_0\right) d\underline{\underline{z}}_n - \gamma\right)$$

Now, dividing by $P_1 + \lambda P_0$, note that minimizing L with respect to R_0 is equivalent to minimizing

$$\frac{P_1}{P_1+\lambda P_0} \int_{R_0} f_{\underline{\underline{X}}_{n-d/0}}\left(\underline{\underline{z}}_{n-d};\Lambda\right) f_{\underline{X}_{n/1}}(\underline{z}_n;\Lambda_1) d\underline{\underline{z}}_n + \frac{\lambda P_0}{P_1+\lambda P_0}\left(\int_{\overline{R}_0} f_{\underline{\underline{X}}_{n/0}}\left(\underline{\underline{z}}_n;\Lambda_0\right) d\underline{\underline{z}}_n - \gamma'\right) \tag{12}$$

where $\gamma' = \gamma/P_0$. Now simply recognize, ignoring the γ' term which does not affect the solution, Eq. (12) can be reinterpreted as a probability of error expansion, just as Eq. (11), but with class priors $\left[\frac{P_1}{P_1+\lambda P_0} \;\; \frac{\lambda P_0}{P_1+\lambda P_0}\right]$.

Thus, our previous probability of error analysis also holds for Eq. (12), i.e. the optimal decision rule, minimizing Eq. (12) and thus achieving least missed detection rate given fixed false positive rate, must use the (conditional) density, conditioning on $\underline{\underline{X}}_{n-d} = \underline{x}_{n-d} = \{x_0,\ldots,x_{n-1}\}$, just as one must do to achieve the Bayes-optimal minimum error rate classifier.

Moreover, consider two window sizes d and $d', d' < d$. Note that making decisions based on the window of observations $\{x_n,\ldots,x_{n+d'-1}\}$ is equivalent to observing $\{x_n,\ldots,x_{n+d-1}\}$ but only using the observations up to time $n+d'-1$ to make decisions. But the above formulation shows that the optimal detection rule, determining $\overline{R}_0$, makes use of all the observations in the window $\{x_n,\ldots,x_{n+d-1}\}$. Thus, given a fixed false positive rate γ, the true detection rate of any detector using just the window $\{x_n,\ldots,x_{n+d'-1}\}$ cannot be larger than the true detection rate of the optimal detector making use of $\{x_n,\ldots,x_{n+d-1}\}$. This proves monotonically non-decreasing detection performance, with increasing window size, d.

References

[1] F. Chamroukhi, A. Same, G. Govaert, P. Aknin, Time series modeling by a regression approach based on a latent process, Neural Networks 22 (2009) 593–602.
[2] K. Frick, A. Munk, H. Sieling, Multiscale change point inference, Journal of the Royal Statistical Society. Series B: Statistical Methology 76 (3) (2014) 495–580.
[3] J. Chen, A. Gupta, Parametric Statistical Change Point Analysis: with Applications to Genetics, Medicine, and Finance, second ed., Springer, 2012.
[4] E.S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.
[5] C.-M. Kuan, K. Hornik, The generalized fluctuation test: A unifying view, Econometric Reviews 14 (2) (1995) 135–161.
[6] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes - Theory and Application, Prentice Hall Inc, 1993.
[7] S. Mishra, O.A. Vanli, C. Park, A multivariate cumulative sum method for continuous damage monitoring with lamb-wave sensors, International Journal of Prognostics and Health Management 6 (2015) 1–11.
[8] J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, Econometrica 66 (1) (1998) 47–78.
[9] B. Jushan, P. Pierre, Computation and analysis of multiple structural change models, Journal of Applied Econometrics, 18 (1), 1–22.
[10] J. Bai, Likelihood ratio tests for multiple structural changes, Journal of Econometrics 91 (2) (1999) 299–323.
[11] D. Egea-Roca, J.A. Lopez-Salcedo, G. Seco-Granados, H. Vincent Poor, Performance bounds for finite moving average tests in transient change detection, IEEE Transactions on Signal Processing 66 (2018).
[12] G.V. Moustakides, Multiple optimality properties of the shewhart test, Sequential Analysis 33 (3) (2014) 318–344.
[13] A. Shiryaev, On optimum methods in quickest detection problems, heory of Probability & Its Applications 8 (1) (1963) 22–46.
[14] A. Tartakovsky, V. Veeravalli, General asymptotic Bayesian theory of quickest change detection, Theory of Probability & Its Applications 49 (3) (2005) 458–497.
[15] S.C. Chin, A. Ray, V. Rajagopalan, Symbolic time series analysis for anomaly detection: A comparative evaluation, Siginal Processing 85 (9) (2005) 1859–1868.
[16] X. Jin, Y. Guo, S. Sarkar, A. Ray, R.M. Edwards, Anomaly detection in nuclear power plants via symbolic dynamic filtering, IEEE Transactions on Nuclear Science 58 (1) (Feb 2011) 277–288.
[17] Y. Li, D.K. Jha, A. Ray, T.A. Wettergren, Information fusion of passive sensors for detection of moving targets in dynamic environments, IEEE Transactions on Cybernetics 47 (1) (Jan 2017) 93–104.
[18] Y. Li, D.K. Jha, A. Ray, T.A. Wettergren, Information-theoretic performance analysis of sensor networks via markov modeling of time series data, IEEE Transactions on Cybernetics 48 (6) (June 2018) 1898–1909.
[19] D. Jha, N. Virani, J. Reimann, A. Srivastav, A. Ray, Symbolic analysis-based reduced order Markov modeling of time series data, Signal Processing 149 (2018) 68–81.
[20] J. Hopcroft, R. Motwani, J. Ullman, Introduction to Automata Theory, Languages, and Computation, 2nd ed., Addison-Wesley, 2001.
[21] B.P. Kitchens, Symbolic Dynamics: One-sided, Two-sided and Countable State Markov Shifts, Springer-Verlag, Berlin Heidelberg GmbH, 1998.
[22] T.M. Luong, V. Perduca, G. Nuel, Hidden Markov model applications in change-point analysis, arXiv preprint arXiv:1212.1778, 2012.
[23] C.-D. Fuh, Sprt and cusum in hidden markov models, The Annals of Statistics, 31 (3) (2003) 942–977. [Online]. Available: doi: 10.1214/aos/1056562468.
[24] A.G. Tartakovsky, On asymptotic optimality in sequential changepoint detection: Non-iid case, IEEE Transactions on Information Theory 63 (2017) 3433–3450.

[25] W. Khreich, E. Granger, R. Sabourin, A. Miri, Combining hidden markov models for improved anomaly detection, in: 2009 IEEE Int Conf on Communications, June 2009, pp. 1–6.
[26] Y. Yuan, Y. Meng, L. Lin, H. Sahli, Y. Anzhi, J. Chen, Z. Zhao, Y. Kong, D. He, Continuous change detection and classification using hidden markov model: A case study for monitoring urban encroachment onto farmland in beijing, Remote Sensing, 7 (11) (2015) 15-318–15-339.
[27] Z. Bouyahia, L. Benyoussef, S. Derrode, Unsupervised sar images change detection with hidden markov chains on a sliding window, Proceedings of SPIE – The International Society for Optical Engineering 6748 (2007) 10.
[28] S.D. Zied Bouyahia, Lamia Ben Youssef, Change detection in synthetic aperture radar images with a sliding hidden markov chain model, Journal of Applied Remote Sensing, 2 (1) (2008) 1–13 – 13 [Online]. Available: doi: 10.1117/1.2957968.
[29] E. Dorj, C. Chen, Anomaly detection approach using hidden Markov model, 2013 IEEE Aerospace Conference (2013) 1–10.
[30] W.R. Blanding, P.K. Willett, Y. Bar-Shalom, S. Coraluppi, Multisensor track management for targets with fluctuating snr, IEEE Transactions on Aerospace and Electronic Systems 45 (4) (2009) 1275–1292.
[31] T.M. Luong, Y. Rozenholc, G. Nuel, Fast estimation of posterior probabilities in change-point analysis through a constrained hidden Markov model, Computational Statistics & Data Analysis 68 (2013) 129–140.
[32] A. Sultana, A. Hamou-Lhadj, M. Couture, An improved hidden markov model for anomaly detection using frequent common patterns, in: 2012 IEEE International Conference on Communications, June 2012, pp. 1113–1117.
[33] H.O. Omoregbee, P.S. Heyns, Fault detection in roller bearing operating at low speed and varying loads using Bayesian robust new hidden Markov model, Journal of Mechanical Science and Technology, 32 (9) (2018) 4025–4036. [Online]. Available: doi: 10.1007/s12206-018-0802-8.
[34] S. Mondal, N.F. Ghalyan, A. Ray, A. Mukhopadhyay, Early detection of thermoacoustic instabilities using hidden markov models, Combustion Science and Technology 191 (8) (2019) 1309–1336.
[35] N.F. Ghalyan, S. Mondal, D.J. Miller, A. Ray, Hidden Markov modeling-based decision-making using short-length sensor time series, The ASME Journal of Dynamic Systems, Measurement, and Control, 141 (10) (2019) 104 502–104 502–6.
[36] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, 1st ed., Prentice Hall, 1993.
[37] T. Cover, J. Thomas, Elements of Information Theory, Wiley-Interscience Publication, 2006.
[38] H. Nguyen, G. McLachlan, P. Orban, P. Bellec, A. Janke, Maximum pseudolikelihood estimation for model-based clustering of time series data, Neural Computation 29 (2017) 990–1020.
[39] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings on IEEE 77 (2) (1989) 257–286.
[40] K. Murphy, Machine Learning: A Probabilistic Perspective, first ed., The MIT Press, 2012.
[41] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[42] S. Bahrampour, A. Ray, S. Sarkar, T. Damarla, N. Nasrabadi, Performance comparison of feature extraction algorithms for target detection and classification, Pattern Recognition Letters 34 (2013) 2126–2134.
[43] T. Lieuwen, V. Yang, Combustion Instabilities in gas turbine engines: operational experience, fundamental mechanisms, and modeling, American Institute of Aeronautics and Astronautics (2005) 3–26, ch. 1.
[44] K. Matveev, Thermoacoustic instabilities in the Rijke tube: experiments and modeling, Ph.D. dissertation, California Institute of Technology, 2003.
[45] G. Rigas, N. Jamieson, L. Li, M. Juniper, Experimental sensitivity analysis and control of thermoacoustic systems, Journal of Fluid Mechanics 787 (2016).
[46] N. Jamieson, G. Rigas, M. Juniper, Experimental sensitivity analysis via a secondary heat source in an oscillating thermoacoustic system, International Journal of Spray and Combustion Dynamics 9 (4) (2017) 230–240.
[47] D. Zhao, Transient growth of flow disturbances in triggering a Rijke tube combustion instability, Combustion and Flame 159 (2012) 2126–2137.
[48] 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.
[49] S. Kwofie, N. Rahbar, A fatigue driving stress approach to damage and life prediction under variable amplitude loading, International Journal of Damage Mechanics 22 (2012) 393–404.
[50] A. Aeran, S. Siriwardane, O. Mikkelsen, I. Langen, A new nonlinear fatigue damage model based only on s-n curve parameters, International Journal of Fatigue 103 (2017) 327–341.
[51] S. Suresh, Fatigue of Materials, second ed., Cambridge University Press, 2004.
[52] S. Gupta, A. Ray, Symbolic dynamics filtering for data-driven pattern recognition, Pattern Recognition: Theory and Applications, 2007, pp. 17–71.
[53] E. Keller, Real-time sensing of fatigue crack damage for information-based decision and control, Ph.D. dissertation, Pennsylvania State University, 2001. [Online]. Available: https://books.google.com/books?id=FTH6jwEACAAJ.
[54] Y. Bengio, P. Frasconi, Input-output HMMs for sequence processing, IEEE Transactions on Neural Networks 7 (1996) 1231–1249.