Journal of Civil Structural Health Monitoring manuscript No.
(will be inserted by the editor)

Hidden Markov Models for Pipeline Damage Detection Using Piezoelectric Transducers

Mingchi Zhang · Xuemin Chen · Wei Li

Received: March 15, 2021/ Accepted: date

Abstract Oil and gas pipeline leakages lead to not only enormous economic loss but also environmental disasters. How to detect the pipeline damages including leakages and cracks has attracted much research attention. One of the promising leakage detection method is to use lead zirconate titanate (PZT) transducers to detect the negative pressure wave when leakage occurs. PZT transducers can generate and detect guided waves for crack detection also. However, the negative pressure waves or guided stress waves may not be easily detected with environmental interference, e.g., the oil and gas pipelines in offshore environment. In this paper, a Gaussian mixture model - hidden Markov model (GMM-HMM) method is proposed to process PZT transducers' outputs for detecting the pipeline leakage and crack depth in changing environment and time-varying operational conditions. Leakages in different sections or crack depths are considered as different states in hidden Markov models (HMMs). One time domain damage index and one frequency domain damage index are extracted from signals collected from PZT transducers, then extracted indices are formed as observation emissions in the HMM. The observation probability distribution matrix in HMM is initialized by a Gaussian mixture model (GMM) to address signal uncertainties. After the HMM parameter initialization, an iterative training process through the Baum-Welch algorithm is applied to get the optimized parameters of the GMM-HMM. Leakage location or crack depth is decided by the maximum posterior probability from the trained model. Two different experimental settings and results show that the GMM-HMM method can recognize the crack depth and leakage of pipeline such as whether there is a leakage, where the leakage is.

Keywords Gaussian mixture model · Hidden Markove model · Lead zirconate titanate (PZT) · Leakage detection · Pipeline

This material is based upon work supported by the National Science Foundation under Grants No. 1801811. Thanks to Smart Materials and Structures Laboratory at the University of Houston for providing the pipeline leakage experimental data.

Mingchi Zhang
Department of Computer Science, Texas Southern University

Xuemin Chen (corresponding author)
Department of Engineering, Texas Southern University
Tel.: +1-713-3137285
Fax: +1-713-3134486
E-mail: xuemin.chen@tsu.edu

Wei Li
Department of Computer Science, Texas Southern University

1 Introduction

Thousands of miles of pipelines crisscrossed on the Gulf of Mexico seafloor are the veins for offshore oil and gas industry of the U.S. or even the whole world, while the leaks and ruptures of those pipelines lead to not only enormous economic loss but also environmental disasters [1]. In the last few decades, many pipeline structural health monitoring techniques have been used to monitor damages [1, 2]. One of the promising methods for pipeline leakage detection is based on the negative pressure wave (NPW) [3–7]. NPW is generated at the leak point when the fluid or gas escapes in the form of a high velocity jet [8]. Then the NPW propagates along pipeline in both directions, i.e., the upstream and downstream of leakage point. The NPW can be detected by lead zirconate titanate (PZT) transducers. PZT transducers are made of piezoelectric materials which can convert mechanical energy to electrical energy and vice versa. This piezoelectric effect leads to PZT transducers work as passive sensors or active actuators. PZT transducers can be effectively used as passive sensors to catch the acoustic signals propagating along the pipeline. On the other hand, the pipeline health condition needs to be evaluated periodically to provide early warning. To address this demand, PZT transducers have been used in active sensing mode to detect crack [9]. PZT transducers have found lots of applications for structural health monitoring. Samantaray et al. [10] used piezoelectric transducer to monitor looseness in bolted joint structure based on the electro-mechanical impedance (EMI) method. Wang et al. [11] invented a wearable PZT transducers which can be easily and noninvasively "worn" onto the flanged valve for bolted joint in real-time. Zhang et al. [12] proposed an active piezoelectric sensing system for concrete crack detection based on the energy diffusivity method. Gulizzi et al. [13] proposed to use piezoelectric transducers bonded or embedded to the structure for SHM based on the simultaneous use of ultrasounds and EMI methods. Gong et al. [14] developed an algorithm to process signals collected by PZT sensor for automatic extraction of the stress wave reflection period. In these applications, the structural health condition is evaluated according to extracted features. However, the changing environment and time-varying operational conditions make the reliability of damage evaluation facing the challenge in a real-world application, especially for the offshore pipelines damage detection in the submarine environment. In recent years, many researchers set their sights on probabilistic and statistical models to improve the damage evaluation reliability under uncertainties . For example, cointegration method was proposed to deal with the interference [15, 16]; a Dirichlet process Gaussian mixture model (DP-GMM) is proposed to adaptively and unsupervised learn structural data collected guided wave sensor measurements [17]; a generalized hidden Markov model (HMM) was used to classify the fault types and fault severity levels of rolling bearings [18].

Among the existing probabilistic and statistical models, HMM has a strong capability in pattern classification, especially for signals with non-stationary natures and poor repeatability and reproducibility [19]. HMM and its variants have been extensively used for speech recognition [20], hand gesture recognition [21], handwritten word recognition [22], and newly applied to spam SMS detection [23]. In recent years, this technology has been gradually applied to structural health monitoring (SHM) [24–28]. In the field of SHM, several researchers have also applied the HMM to damage evaluation. Rammohan and Taha [29] using a standard HMM to model the simulated data of a pre-stressed concrete bridge. Tschöpe and Wolff [30] studied the HMM for damage degree classification on plate-like structures. Mei et al. [31] proposed an HMM based unweighted moving average trend estimation (HMM-UMATE) method to improve the damage evaluation reliability of real aircraft structures under time-varying conditions. A Gaussian mixture model-hidden Markov model (GMM-HMM) was proposed to evaluate damage severity using guided wave. These studies indicate that the HMM is robust to uncertainties. The HMM and its variants have been used for detect pipeline health monitoring also. Ai et al. [32] designed a pipeline health monitoring and leak detecting system based on extracted linear prediction cepstrum coefficient from acoustic sensors. HMM was used to recognize damage types. Qiu et al. [33] proposed an early-warning model of compressor in gas pipeline. A deep belief network (DNN) was used to extract signal features which formed the observation sequence for HMM to estimate the operating status of compressor unit. Tejedor et al. [34] proposed a smart surveillance system for pipeline integrity detection using fiber optic sensors. A GMM-HMM based pattern classification combining with contextual feature information for decision making was presented.

In this paper, a GMM-HMM method is proposed to detect pipeline leakage and crack depths using piezoelectric transducers. One time domain damage index and one frequency domain damage index are extracted from signals collected from PZT transducers. Different from the existing work, our work aims to find the leakage location of a long pipeline based on the stress wave collected by PZT sensors and detect the crack depth on pipeline based on guided wave which is generated and collected by PZT transducers.

2 GMM-HMM based leakage detection method

In this section, method for damage indices extraction from the acoustic signals detected by PZT sensors is discussed. Then the design and training of GMM-HMM are presented.

2.1 Damage indices extraction

To establish the relationship between characteristics change of the sampled waveform and damage parameters, two damage indices are adopted to indicate the signal variations and serve as observations of the HMM model. The first damage index (DI_1) is a time-domain damage index [35], defined in the Eq. (1):

$$DI_1 = 1 - \sqrt{\frac{\int (s_1(t) - \bar{s}_1)(s_2(t) - \bar{s}_2)\, dt}{\int (s_1(t) - \bar{s}_1)^2\, dt \int (s_2(t) - \bar{s}_2)^2\, dt}}, \qquad (1)$$

where $s_1(t)$ is the baseline waveform and $s_2(t)$ is the comparison waveform at time t. The $\bar{s}_1$ and $\bar{s}_2$ is average value of $s_1(t)$ and $s_2(t)$. The baseline waveform represent the incident waveform, and the comparison waveform denote the captured waves detected by sensors. Unity minus the absolute value of the Pearson correlation coefficient is used as the time domain damage index which can identify the signal difference. The time domain damage index is different from the one used by Mei et al. [31]. In [31], a signal energy was calculated as the time domain damage index.

The second one (DI_2) is a frequency-domain damage index [31], the amplitude of peak frequency, as defined in the Eq. (2):

$$DI_2 = \max_{f_1 \leq f \leq f_2}(|X(f)|), \tag{2}$$

where $X(f) = \int_{t_1}^{t_2} X(t)e^{-2\pi ift}\,dt$, $i = \sqrt{-1}$, f_1 and f_2 are the start and stop frequency corresponding to the selected frequency spectrum window.

2.2 HMM and the Baum–Welch algorithm

HMM is a powerful probabilistic and statistical modeling system and defines a probability distribution over sequences of observations by invoking another sequence of hidden states which has Markov dynamics. Each hidden state has a probability distribution over the possible leakage states. The posterior probabilities of leakage states obtained by HMM can be used to obtain a probabilistic evaluation of the pipeline damage.

To reduce the calculation complexity and keep the model works efficiently, only first-order HMM with continuous observation is adopted in this study. A typical HMM model can be defined by a three-tuple:

$$\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}, \tag{3}$$

where $\boldsymbol{\pi}$ is the initial probability distribution, $\boldsymbol{A}$ is the state transition probability matrix and $\boldsymbol{B}$ is the observation probability distribution matrix or the emission matrix. Common notations are used in this paper as follows.

a) $\boldsymbol{S}$: A set of N hidden states is denoted as $\boldsymbol{S} = \{s_1, s_2, s_3, \ldots, s_N\}$. The state of model at time t is denoted by $q_t \in \boldsymbol{S}$, which denotes the current state.

b) $\boldsymbol{V}$: A set of observed emissions in a specifically range or a combination of time intervals. The observed state at time t is denoted by $O_t \in \boldsymbol{V}$.

c) $\boldsymbol{\pi}$: an $N \times 1$ initial probability distribution over the state. π_i is the probability that the Markov chain will start in the state s_i. In general, we will denote it by

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N). \tag{4}$$

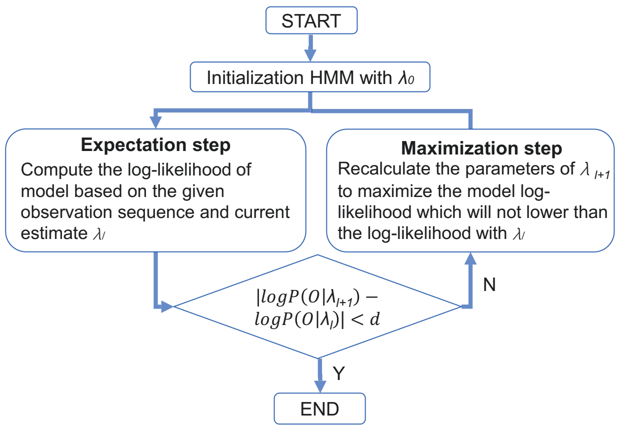

Fig. 1 Baum-Welch algorithm (EM algorithm)

d) $\boldsymbol{A}$: $N \times N$ state transition probability matrix, N is the number of hidden states. $\boldsymbol{A} = [a_{ij}], 1 \leq i, j \leq N$.

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i), \tag{5}$$

where a_{ij} is the probability of transition from structural damage state s_i at time t to structural damage state s_j at time $t+1$ and $\sum_{j=1}^{N} a_{ij} = 1$.

e) $\boldsymbol{B}$: a state-dependent observation density column vector $\boldsymbol{B} = (b_1(O_t), b_2(O_t), ..., b_N(O_t))^\tau$, where $b_j(O_t)$ is the probability density function of observation O_t at time t since GMM is used to model the distribution of two damage indices in this paper, given that it is in state s_j.

There are three basic problems in HMM: evaluation, decoding, and learning [20]. Given a training set of observations, HMM is trained to find the optimized parameters $\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}$ by Baum-Welch expectation-maximization algorithm (EM algorithm), which is a learning problem of HMM. After the HMM trained, this model can be applied to evaluate the damage states of a new defective pipeline in various conditional environments.

The Baum-Welch algorithm is employed in HMM to perform the training as shown in Fig. 1, which is an iterative process to adjust the parameters of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\pi}$.

It is an EM algorithm which uses dynamic programming (forward and backward algorithm) to simplify the calculation by terminating at iteration step j whenever for a given small positive value d that

$$|logP(O|\lambda_{j+1}) - logP(O|\lambda_j)| < d, \tag{6}$$

where O is the observation sequence and λ_j represents the system after j-th step iteration.

2.3 Gaussian mixture model

In this paper, the Gaussian mixture model[31, 36] is chosen to model the distribution of observed damage indices to obtain B in the hidden Markov model, which elements are given by Eq. (7).

$$b_j(O_t) = \sum_{k=1}^{K} c_{jk} N(O_t, \mu_{jk}, \Sigma_{jk}), \tag{7}$$

where c_{jk}, μ_{jk} and Σ_{jk} are the mixture weight, mean vector, the covariance matrix of the K-component Gaussian mixture model. The EM algorithm is implied in the GMM model, which contains the 4 steps, i.e., initialization, expectation, maximization and iteration as described below.
a) Initialization: Setting the number of mixture component K, for each component initializing it with c_{jk}, μ_{jk} and Σ_{jk}.
b) Expectation step (E step): Calculate the posterior probability based on current c_{jk}, μ_{jk} and Σ_{jk}.
c) Maximization step (M step): Calculate new c_{jk}, μ_{jk}, Σ_{jk} based on the $\gamma(z_{tk})$. $\gamma(z_{tk})$ represents the probability of occupying mixture Gaussian component K of structural damage state S at time t, which is used to update the state-dependent observation density $\boldsymbol{B}$.
d) Iteration: Iterate these steps until the GMM model converges.

2.4 Flow chart of GMM-HMM method

Based on the proposed method in sections 2.2 and 2.3, GMM-HMM method is shown in Fig. 2. There are five major steps included in the damage detection process.

1) States of the HMM are designed based on the placement of pipelines and sensors.
2) Stress waves are obtained from the sensors. Damage indices are extracted from the stress waves. The GMM model calculates the parameter of each Gaussian model which formed matrix $\boldsymbol{B}$ in HMM.
3) A typical HMM model defined by a three-tuple: $\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}$. To initialize the HMM, $\boldsymbol{\pi}$ and $\boldsymbol{A}$ is estimate by the prior probability.
4) After the parameter initialization of the HMM, the training is implemented by an iterative process through the Baum-Welch algorithm.
5) Given the trained HMM and the new observations to find out the current pipeline status.

3 Experimental Results

3.1 Leakage detection

3.1.1 Experiment setup

The purpose of the experiment is to detect pipeline leakage utilizing PZT transducers. The negative pressure wave is generated by leakage in the pipeline and propagates along the pipeline from leakage point to both ends.

The experimental pipeline that was built at the University of Houston is shown in Fig. 3. It consists of a series of PVC plain-end pipe sections connected together to form a pipeline with a total length of 55.78 m. Six PZT sensors (P_1 to P_6, size is 15 $mm \times 10$ mm) are directly mounted on the pipeline to detect NPW signal arrival. A NI PXI-5105 digitizer is used as a data acquisition system. The digitizer is triggered

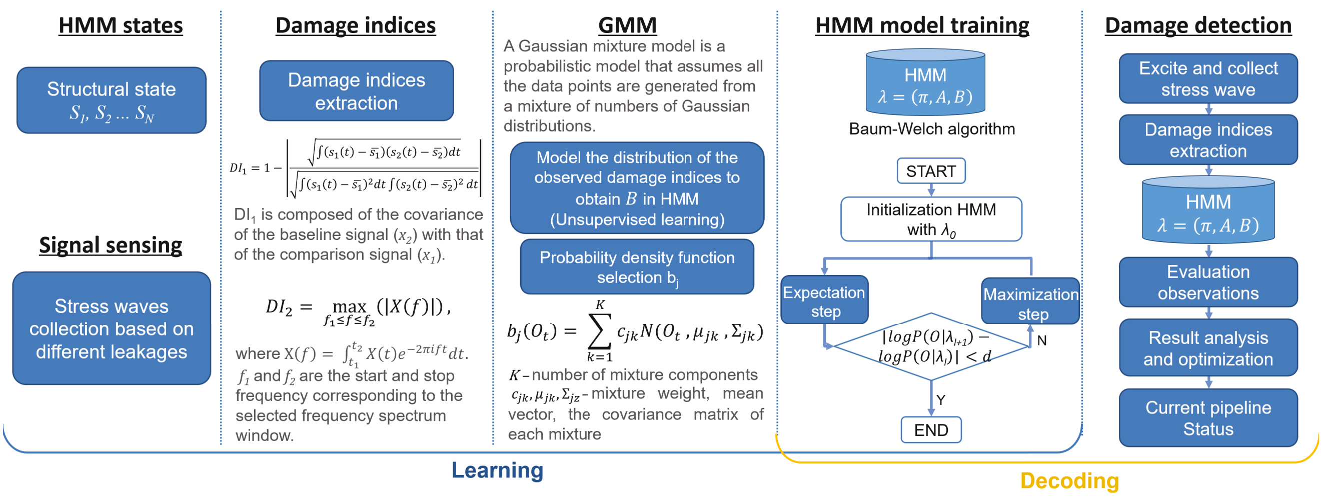

Fig. 2 Flow chart of GMM-HMM model

by the voltage signal of PZT sensor (p_1) with the trigger level at $-0.02V$ and all the signals from six PZT sensors are recorded simultaneously at a sampling rate of 100 KS/s. The model is implemented based on the hmmlearn Python package [37] and seqHMM R package [38], ran on a desktop with 64 bit Windows 10 and Intel core i7-8700 CPU. The seqHMM R package is used for Gaussian mixture model component number K estimation. The same computational environment is used for leakage location detection and crack depth inspection presented in Section 3.2. and 3.3. respectively.

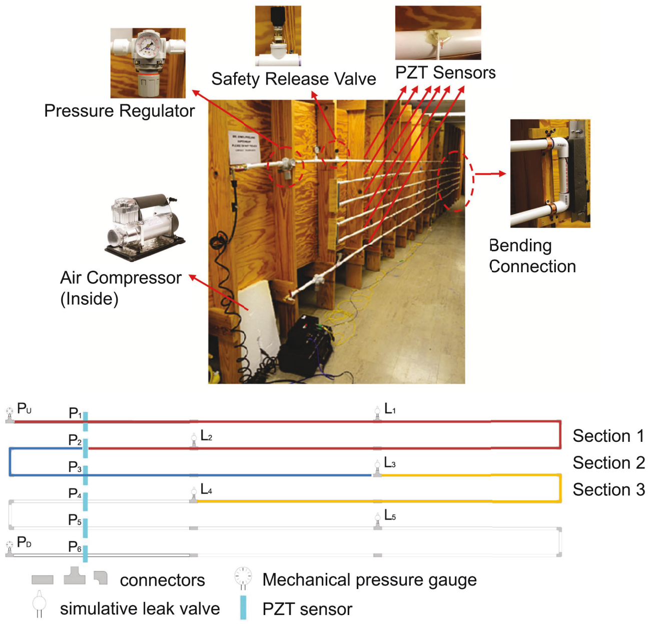

Fig. 3 Settlement of the pipeline and sensors [4]

This experiment has served a published paper using the latency of signal to locate leakage [4]. Although it is not targeted to solve the problem of submarine pipeline leakage detection, the experimental data are used to validate the proposed GMM-HMM method.

Two states, leaking or not leaking, are chosen as the states in left to right HMM model, where the transitions only go from one state to itself or the unique follower. Leakage signal collected by sensor P_1 and baseline signals of three different sections used as original data of the model.

3.1.2 Results of leakage detection

Signals of the different states are shown in Fig. 4. The blue line represents leakage state signal and red solid line represents no leakage state signal. For each test, 90000 samples have been used in calculation.

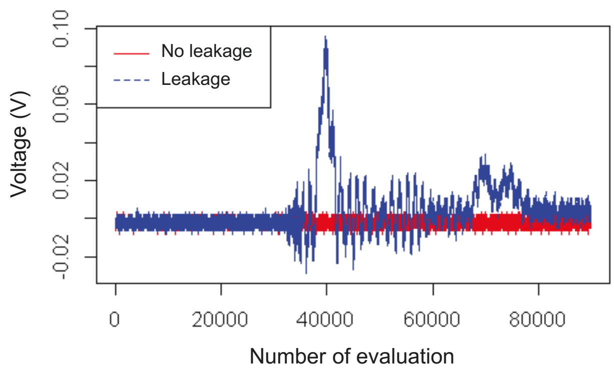

Fig. 4 Signal of different states

The experiment of collecting signals for each leakage state is repeated by 20 times. Randomly selected experiment data from the same sensor are separated into two groups, 70% of training data and 30% of testing data. For training data, a total of 70 groups, 100000 samples per group are collected.

Two damage indices, i.e, DI_1 and DI_2, in two states are obtained by using Eq. 1 and Eq. 2 with 45 data points for each state as shown in Fig. 5. 45 data points

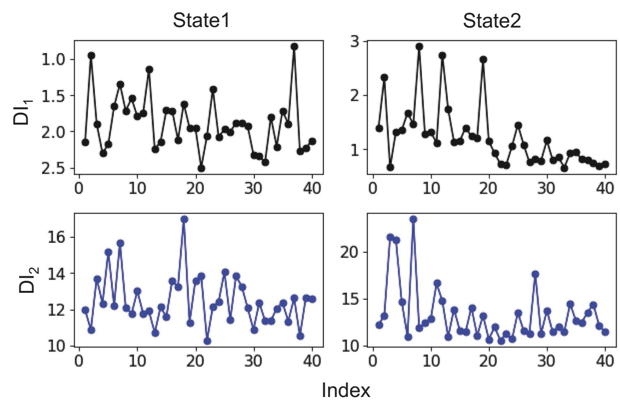

Fig. 5 Damage indices

for each state the time domain damage index DI_1 and frequency domain damage index DI_2 in two states as shown in Fig. 5. In the DI_1, measurement scales from -2.5 to -1.0 in state1 and scales from 1 to 3 in state2. Form the distribution of data, the difference between each state can be distinguished. The combined consideration of two different dimensions DIs improves the richness of the data and makes the results more reliable. The parameter of GMM can be calculated by EM algorithm. After the calculation, damage indices are gathered into two distributions which can be predicted and measured by the negative log-likelihood as shown in Fig. 6.

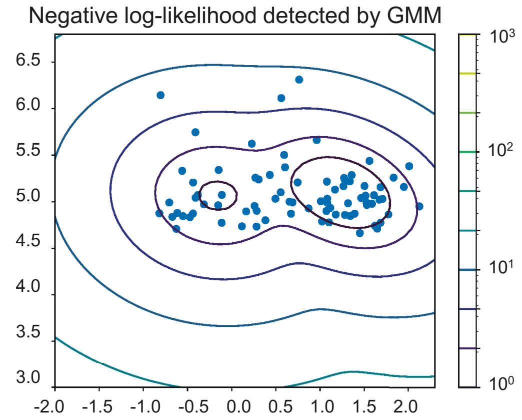

Fig. 6 Negative log-likelihood predicted by GMM

As for this experiment, since the leaking and not leaking are under investigation, the HMM state number has been set to 2. Also, by noting that two damage indices are considered, the dimension of the corresponding GMM is set by 2. By using the R seqHMM package, the number of component number K is set as 2 similar as that in [38]. The parameters of HMM model are initialized as:

$$\boldsymbol{\pi} = [1, 0], \boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix}.$$

The parameters of two states HMM model are reassessed by the Baum-Welch algorithm based on Eq. 6. The maximum number of iteration is set to 10 times and the tolerance is 0.01 where EM will stop if the gain in log-likelihood is below.

After training, without lose of generality we use the same notations for state stationary probability and state transmission matrix in the paper, the updated parameters are obtained as follows:

$$\boldsymbol{\pi} = [1, 0], \boldsymbol{A} = \begin{bmatrix} 0.983 & 0.017 \\ 0 & 1 \end{bmatrix}.$$

Based on the two damage indices input, mixture weight c_{jk}, mean vector μ_{jk} and covariance matrix Σ_{jk} of the 2-component Gaussian mixture model obtained are as follows:

As for the state 1, the corresponding two mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{1,1:2} = [0.794, 0.206],$$

$$\mu_{1,1:2} = \{\begin{pmatrix} 1.380 \\ 5.059 \end{pmatrix}; \begin{pmatrix} 6.283 \\ 4.954 \end{pmatrix}\},$$

$$\Sigma_{11} = \begin{pmatrix} 0.0744 & -0.0239 \\ -0.0239 & 0.0501 \end{pmatrix};$$

$$\Sigma_{12} = \begin{pmatrix} 0.3572 & 0.0255 \\ 0.0255 & 0.0079 \end{pmatrix}.$$

As for the state 2, the corresponding two mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{2,1:2} = [0.634, 0.366],$$

$$\mu_{2,1:2} = \{\begin{pmatrix} 1.380 \\ 5.059 \end{pmatrix}; \begin{pmatrix} 6.283 \\ 4.954 \end{pmatrix}\},$$

$$\Sigma_{21} = \begin{pmatrix} 0.7378 & 0.0614 \\ 0.0614 & 0.0320 \end{pmatrix};$$

$$\Sigma_{22} = \begin{pmatrix} 0.2644 & 0.0496 \\ 0.0496 & 0.2192 \end{pmatrix}.$$

The testing is performed after the training process of the HMM model which parameters have been optimized. The testing is used to validate the detection capability of the HMM model and the result is showed in Fig. 7.

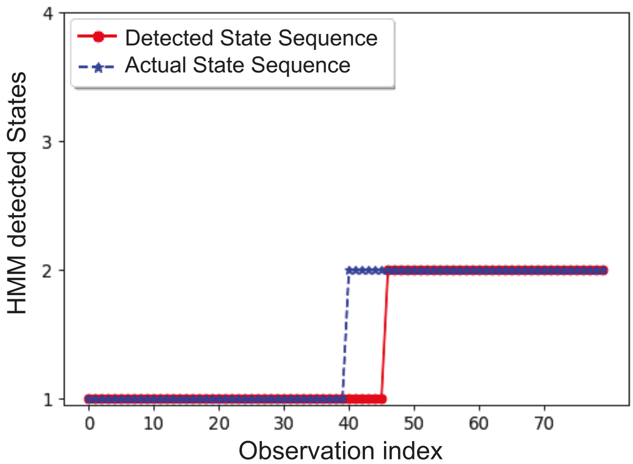

Fig. 7 Pipeline status output

The blue dashed line represents actual state sequence and red solid line represents detected state sequence. The accuracy of testing performance between detected state sequence and actual state sequence are approaching 92.51%. The accuracy is defined as the ratio of correctly detected states to the total number of state sequence.

3.2 Leakage location detection

3.2.1 Experiment setup

In this experiment, all the data are based on the previous experimental setup. By changing the leakage state to different leakage locations, the output of HMM model will be changed to leakage location detection. As shown in Fig. 8, there are three leakage locations corresponding to three different states in HMM. State 1 denotes leakage occurred at section 1 of pipeline, state 2 and 3 denote leakages occur at section 2 and 3 of pipeline.

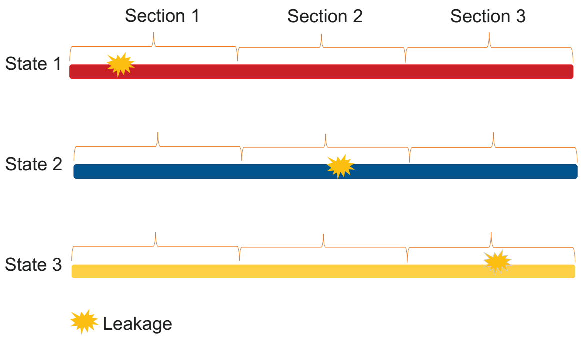

Fig. 8 Schematic diagram of pipeline statues

Therefore, there are three states in this Markov process which allowing for transitions from any emitting state to any other emitting state.

Leakage signals are collected by sensor P_1. As shown in Fig. 9. The red solid line represents state 1, where the leakage happened in section 1 of pipeline. The blue dashed line denote state 2 where leakage happened in section 2. The yellow solid line indicate state 3. A total of 100 groups, 200000 samples per group are collected. To reduce the computational volume, data for this experience are cropped from the original data by 90000 samples to each state.

Two damage indices are obtained by using Eq. 1 and Eq. 2 with 45 data points for each state in HMM as shown in Fig. 10. These three states represent different leakage location on the pipeline. In the time domain DI_1, there are clear difference between each states which is not obvious in the raw data. The above processing makes the size of the data suitable for HMM to operate.By extracting DI the changes of states are presented more obvious without losing too many features.

3.2.2 Results of leakage location detection

The 100 groups of measurements from three different leakage locations are separated into a training group

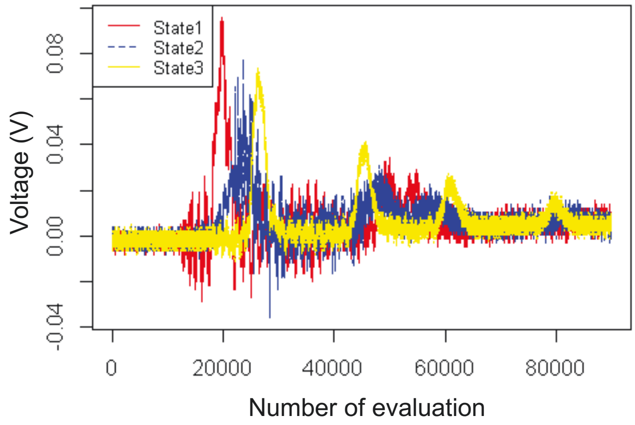

Fig. 9 Signal of different states

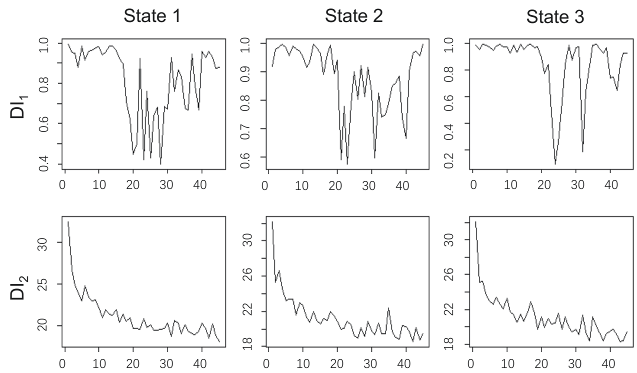

Fig. 10 Damage indices

and a testing group. In this experiment, leakage section1, leakage section2, and leakage section3 are the three hidden states of HMM. The dimension of GMM is two because two damage indices are considered. By using the R seqHMM package, the number of component number K is set to 3 similar as that in [38]. According to the observation density estimation, the parameters of GMM for each leakage state can be calculated by EM algorithm. The parameters are initialized as:

$$\boldsymbol{\pi} = [1, 0, 0], \boldsymbol{A} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

The parameters of three states HMM model are reassessed by the EM algorithm based on Eq. 6. The maximum number of iteration is set to 10 times and the tolerance is 0.01 where EM will stop if the gain in log-likelihood is below. After training, the updated parameters are obtained as follows:

$$\boldsymbol{\pi} = [0, 0, 1], \boldsymbol{A} = \begin{bmatrix} 0.232 & 0.357 & 0.411 \\ 0 & 0.525 & 0.475 \\ 0.539 & 0.102 & 0.359 \end{bmatrix}.$$

Based on the two damage indices input, mixture weight c_{jk}, mean vector μ_{jk} and covariance matrix Σ_{jk} of 3-component Gaussian mixture model obtained are as follows:

As for the state 1, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{1,1:3} = [0.298, 0.106, 0.596],$$

$$\mu_{1,1:3} = \{\begin{pmatrix} 8.640 \\ 20.344 \end{pmatrix}; \begin{pmatrix} 9.835 \\ 22.195 \end{pmatrix}; \begin{pmatrix} 1.785 \\ 21.097 \end{pmatrix}\},$$

$$\Sigma_{11} = \begin{pmatrix} 0.216 & -0.622 \\ -0.622 & 1.799 \end{pmatrix};$$

$$\Sigma_{12} = \begin{pmatrix} 0.062 & -0.202 \\ -0.202 & 1.301 \end{pmatrix};$$

$$\Sigma_{13} = \begin{pmatrix} 0.003 & -0.057 \\ -0.057 & 4.048 \end{pmatrix}.$$

As for the state 2, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{2,1:3} = [0.532, 0.390, 0.078],$$

$$\mu_{2,1:3} = \{\begin{pmatrix} 9.275 \\ 20.596 \end{pmatrix}; \begin{pmatrix} 9.681 \\ 32.290 \end{pmatrix}; \begin{pmatrix} 7.957 \\ 19.559 \end{pmatrix}\},$$

$$\Sigma_{21} = \begin{pmatrix} 0.009 & 0.053 \\ 0.053 & 2.838 \end{pmatrix};$$

$$\Sigma_{22} = \begin{pmatrix} 0.127 & 0.028 \\ 0.028 & 0.036 \end{pmatrix};$$

$$\Sigma_{23} = \begin{pmatrix} 0.098 & -0.652 \\ -0.652 & 4.367 \end{pmatrix}.$$

As for the state 3, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{3,1:3} = [0.389, 0.331, 0.280],$$

$$\mu_{3,1:3} = \{\begin{pmatrix} 9.572 \\ 22.944 \end{pmatrix}; \begin{pmatrix} 9.474 \\ 22.675 \end{pmatrix}; \begin{pmatrix} 5.619 \\ 19.705 \end{pmatrix}\},$$

$$\Sigma_{31} = \begin{pmatrix} 0.431 & 0.310 \\ 0.310 & 0.236 \end{pmatrix};$$

$$\Sigma_{32} = \begin{pmatrix} 1.009 & -0.432 \\ -0.432 & 0.5338 \end{pmatrix};$$

$$\Sigma_{33} = \begin{pmatrix} 1.262 & -0.320 \\ -0.320 & 0.219 \end{pmatrix}.$$

Then, the testing is performed after the parameter optimization of HMM model. The model calculates posterior probability when new testing data are fed. By maximizing the posterior probability, the detected

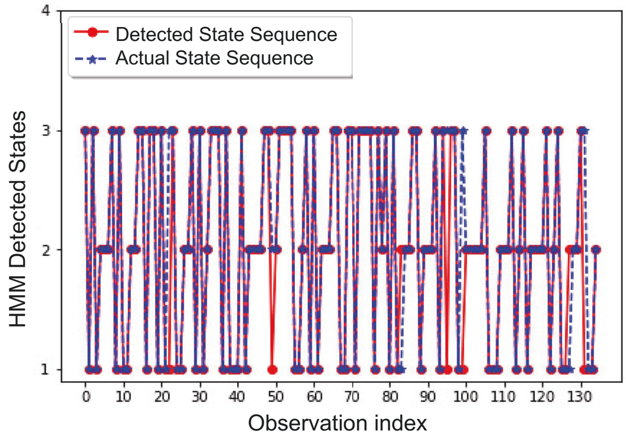

Fig. 11 Actual state vs. HMM detected state

state is obtained. The leakage location evaluation result is shown in Fig. 11. The blue dashed line denote actual state sequence. The red solid line indicate detected state sequence giving by HMM. The accuracy of testing performance between detected state sequence and actual state sequence are approaching 94.81%, which shows the high performance of this model.

3.3 Crack depth inspection

3.3.1 Experiment setup

The purpose of the experiment is to detect the depth of cracks utilizing PZT sensors. The experimental pipeline was built at the University of Houston is shown in Fig. 12. It consists of a section of galvanized steel pipe with a total length of 3 meters. A PZT array with sixteen PZT transducers are directly mounted on the pipeline to detect defection signal. Longitudinal axially symmetric mode, L(0,2), is utilized in this experiment.

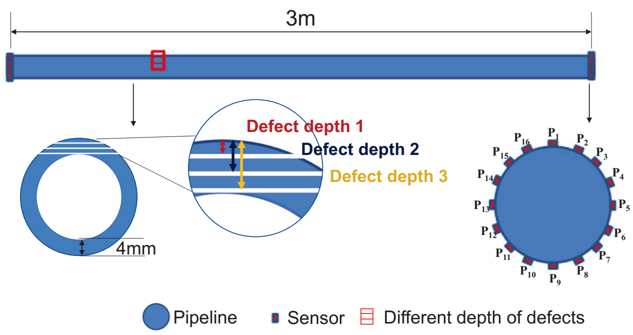

Fig. 12 Setup of the defect detection experiment

The guided wave is generated from sensors at the left side of pipe at the frequency of 50 kHz, propagates along the pipeline through the defect point, and received by the sensors at other side. This PZT array can detect the pipeline defect location based on time-reversal method and matching pursuit de-noising [39]. In addition, this PZT array has been used for underwater communication by using stress wave propagation along pipelines [40].

Three states, defect depth 1, defect depth 2, and defect depth 3, are chosen as the states in the left to right HMM model, which means the depth of defect is increased unidirectionally. Reflect signal is collected from the right-hand side sensors. Raw data collected with noise in three different depths will be used as original data for the GMM-HMM model as shown in Fig. 13.

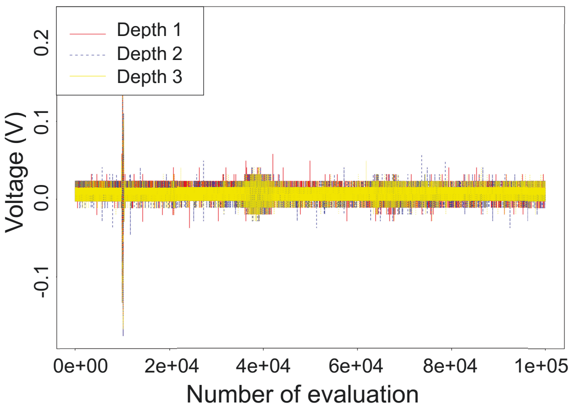

Fig. 13 Signal of different states

Damage indices extracted from defection signals are shown in Fig. 14. Two damage indices calculation are based on the Eqs. (1) and (2). These three states represent different depth of cracks on the pipeline. The original data is noisy, without denoising procedure the damage index extract the features and distinguish each state efficiently.

3.3.2 Results of depth inspection

The 154 group samples from three defect depth inspection are separated into two groups. 70% measurements is used as a training and the rest is used for testing. In this experiment, there are three crack depth, in an other word, the state number of HMM is 3. Since two damage indices are considered, two dimension GMM is applied. The parameters of the Gaussian mixture model are also calculated based on the observation density estimation method and the number of Gaussian Mixture K is calculated by the R seqHMM package, and is set

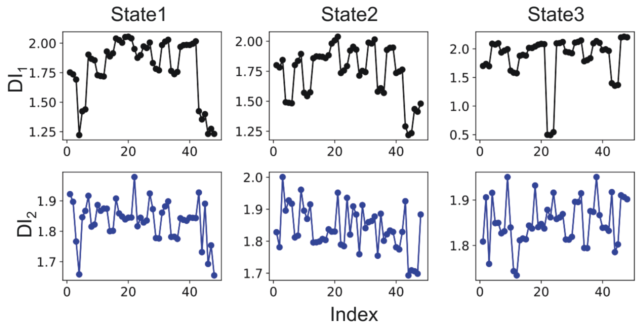

Fig. 14 Damage indices

to 3 similar as in [38]. The parameters of the left to right HMM model are initialized as:

$$\boldsymbol{\pi} = [1, 0, 0], \boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The parameters of the three states HMM model are also reassessed by the Baum-Welch algorithm based on Eq. 6. The maximum number of iteration is set to 10 times and the tolerance is 0.01 where EM will stop if the gain in log-likelihood is below. After training, the updated parameters are obtained as follows:

$$\boldsymbol{\pi} = [1, 0, 0], \boldsymbol{A} = \begin{bmatrix} 0.974 & 0.026 & 0 \\ 0 & 0.976 & 0.024 \\ 0 & 0 & 1 \end{bmatrix}.$$

Based on the two damage indices input, mixture weight c_{jk}, mean vector μ_{jk} and covariance matrix Σ_{jk} of 3-component Gaussian mixture model obtained are as follows: As for the state 1, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{1,1:3} = [0.422, 0.522, 0.054],$$

$$\mu_{1,1:3} = \{\begin{pmatrix} -0.432 \\ 1.624 \end{pmatrix}; \begin{pmatrix} 0.439 \\ 1.666 \end{pmatrix}; \begin{pmatrix} -0.542 \\ 1.572 \end{pmatrix}\},$$

$$\Sigma_{11} = \begin{pmatrix} 4.888 & 4.998 \\ 4.998 & 8.517 \end{pmatrix};$$

$$\Sigma_{12} = \begin{pmatrix} 3.605 & 2.828 \\ 2.828 & 2.516 \end{pmatrix};$$

$$\Sigma_{13} = \begin{pmatrix} 6.182 & 3.852 \\ 3.852 & 2.545 \end{pmatrix}.$$

As for the state 2, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{2,1:3} = [0.081, 0.767, 0.152],$$

$$\mu_{2,1:3} = \{\begin{pmatrix}0.739\\1.629\end{pmatrix}; \begin{pmatrix}0.542\\1.020\end{pmatrix}; \begin{pmatrix}1.075\\1.733\end{pmatrix}\},$$

$$\Sigma_{21} = \begin{pmatrix}1.597 & 2.437\\2.437 & 4.448\end{pmatrix};$$

$$\Sigma_{22} = \begin{pmatrix}2.514 & -2.392\\-2.392 & 5.933\end{pmatrix};$$

$$\Sigma_{23} = \begin{pmatrix}7.849 & 7.833\\7.833 & 9.896\end{pmatrix}.$$

As for the state 3, the corresponding three mixture 2-dimensional Gaussian models have the following parameters, respectively.

$$c_{3,1:3} = [0.124, 0.541, 0.335],$$

$$\mu_{3,1:3} = \{\begin{pmatrix}-0.889\\1.670\end{pmatrix}; \begin{pmatrix}-0.313\\1.690\end{pmatrix}; \begin{pmatrix}-1.485\\1.651\end{pmatrix}\},$$

$$\Sigma_{31} = \begin{pmatrix}2.026 & 8.372\\8.372 & 4.173\end{pmatrix};$$

$$\Sigma_{32} = \begin{pmatrix}1.064 & 1.426\\1.426 & 7.605\end{pmatrix};$$

$$\Sigma_{33} = \begin{pmatrix}2.170 & 1.153\\1.153 & 6.182\end{pmatrix}.$$

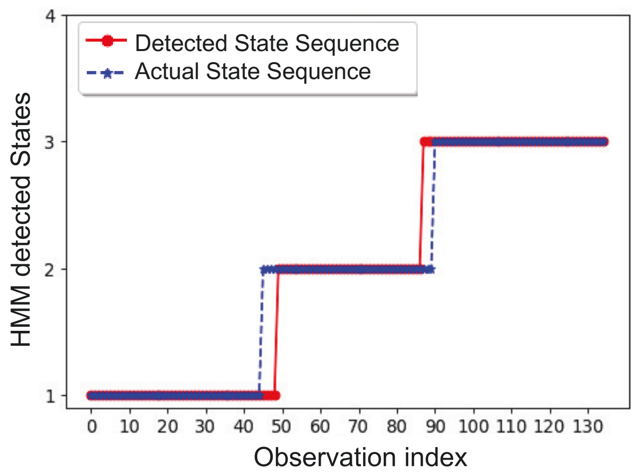

Fig. 15 HMM output

By maximizing the posterior probability, the detected state will be obtained. The crack depth evaluation results are shown in Fig. 15. The blue dashed line represent actual state sequence and the red solid line indicate detected state sequence giving by HMM. The accuracy of cracked depth inspection performance is approaching 93.23%.

4 Conclusion

Pipeline leakage detection and crack depth identification are difficult especially in changing environment and time-varying operational conditions. This study applied a GMM-HMM method on pipeline damage detection to answer the research questions, i.e., whether the pipeline has a leak, where the leakage location is and how deep the crack is. One time domain damage index based on the Pearson correlation was used to identify the signal difference and one frequency domain damage index based on Fourier transform to find the peak frequency. The experimental results illustrate the effectiveness of proposed GMM-HMM method. In the future work, we will use the signal energy as time domain damage index to see if it can improve the detection accuracy. The hybrid damage indices will be explored also. In addition, a better way to extract the signal features will be investigated.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Ho M, El-Borgi S, Patil D, Song G (2020) Inspection and monitoring systems subsea pipelines: A review paper. Structural Health Monitoring 19(2):606–645, DOI 10.1177/1475921719837718
2. Adegboye MA, Fung WK, Karnik A (2019) Recent advances in pipeline monitoring and oil leakage detection technologies: principles and approaches. Sensors 19(11):2548, DOI 10.3390/s19112548
3. Zhang G, Ho SCM, Huo L, Zhu J (2019) Negative pressure waves based high resolution leakage localization method using piezoceramic transducers and multiple temporal convolutions. Sensors 19(9):1990, DOI 10.3390/s19091990
4. Zhu J, Ren L, Ho SC, Jia Z, Song G (2017) Gas pipeline leakage detection based on PZT sensors. Smart Materials and Structures 26(2):025022, DOI 10.1088/1361-665x/26/2/025022
5. Hou Q, Ren L, Jiao W, Zou P, Song G (2013) An improved negative pressure wave method for natural gas pipeline leak location using FBG based strain sensor and wavelet transform. Mathematical Problems in Engineering 2013:1–8, DOI 10.1155/2013/278794
6. Liang W, Zhang L, Xu Q, Yan C (2013) Gas pipeline leakage detection based on acoustic tech-

nology. Engineering Failure Analysis 31:1–7, DOI 10.1016/j.engfailanal.2012.10.020

7. Li Y, Sun L (2009) Leakage detection and location for long range oil pipeline using negative pressure wave technique. In: 2009 4th IEEE Conference on Industrial Electronics and Applications, IEEE, DOI 10.1109/ICIEA.2009.5138796
8. Wikipedia contributors (2020) Leak detection — Wikipedia, the free encyclopedia. URL https://en.wikipedia.org/wiki/Leak_detection, [Online; accessed 13-August-2020]
9. Li N, Chen X, Song G (2019) Fatigue crack width detection based on the active sensing method: A feasibility study. In: 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), IEEE, IEEE, pp 186–189, DOI 10.1109/icnsc.2019.8743163
10. Samantaray SK, Mittal SK, Mahapatra P, Kumar S (2018) An impedance-based structural health monitoring approach for looseness identification in bolted joint structure. Journal of Civil Structural Health Monitoring 8(5):809–822, DOI 10.1007/s13349-018-0307-2
11. Wang C, Wang N, Ho SC, Chen X, Pan M, Song G (2018) Design of a novel wearable sensor device for real-time bolted joints health monitoring. IEEE Internet of Things Journal 5(6):5307–5316, DOI 10.1109/jiot.2018.2852653
12. Zhang C, Yu X, Alexander L, Zhang Y, Rajamani R, Garg N (2016) Piezoelectric active sensing system for crack detection in concrete structure. Journal of Civil Structural Health Monitoring 6(1):129–139, DOI 10.1007/s13349-015-0143-6
13. Gulizzi V, Rizzo P, Milazzo A, Ribolla ELM (2015) An integrated structural health monitoring system based on electromechanical impedance and guided ultrasonic waves. Journal of Civil Structural Health Monitoring 5(3):337–352, DOI 10.1007/s13349-015-0112-0
14. Gong P, Luo M, Zhou L, Jiang L, Chen X (2020) An image processing method for extraction of the stress wave reflection period. Applied Sciences 10(10):3486, DOI 10.3390/app10103486
15. Worden K, Cross E, Antoniadou I, Kyprianou A (2014) A multiresolution approach to cointegration for enhanced SHM of structures under varying conditions–an exploratory study. Mechanical Systems and Signal Processing 47(1-2):243–262, DOI 10.1016/j.ymssp.2013.10.012
16. Dao PB, Staszewski WJ (2014) Lamb wave based structural damage detection using cointegration and fractal signal processing. Mechanical Systems and Signal Processing 49(1-2):285–301, DOI 10.1016/j.ymssp.2014.04.011
17. Chakraborty D, Kovvali N, Papandreou-Suppappola A, Chattopadhyay A (2015) An adaptive learning damage estimation method for structural health monitoring. Journal of Intelligent Material Systems and Structures 26(2):125–143, DOI 10.1177/1045389X14522531
18. Liu J, Hu Y, Wu B, Wang Y, Xie F (2017) A hybrid generalized hidden Markov model-based condition monitoring approach for rolling bearings. Sensors 17(5):1143
19. Li Z, Han J, Sun J, He Y, Chu F (2007) Fault recognition method based on independent component analysis and hidden Markov model. Journal of Vibration and Control 13(2):125–137, DOI 10.1177/1077546307066964
20. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2):257–286, DOI 10.1016/b978-0-08-051584-7.50027-9
21. Wilson AD, Bobick AF (1999) Parametric hidden markov models for gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(9):884–900, DOI 10.1109/34.790429
22. Mohamed MA, Gader P (2000) Generalized hidden Markov models. II. Application to handwritten word recognition. IEEE Transactions on Fuzzy Systems 8(1):82–94, DOI 10.1109/91.824774
23. Xia T, Chen X (2020) A discrete hidden Markov model for SMS spam detection. Applied Sciences 10(14):5011, DOI 10.3390/app10145011
24. Bunks C, McCarthy D, Al-Ani T (2000) Condition-based maintenance of machines using hidden Markov models. Mechanical Systems and Signal Processing 14(4):597–612, DOI 10.1006/mssp.2000.1309
25. Baruah P, Chinnam RB (2005) HMMs for diagnostics and prognostics in machining processes. International Journal of Production Research 43(6):1275–1293, DOI 10.1080/00207540412331327727
26. Zaidi SSH, Aviyente S, Salman M, Shin KK, Strangas EG (2010) Prognosis of gear failures in dc starter motors using hidden Markov models. IEEE Transactions on Industrial Electronics 58(5):1695–1706, DOI 10.1109/tie.2010.2052540
27. Geramifard O, Xu JX, Zhou JH, Li X (2012) A physically segmented hidden Markov model approach for continuous tool condition monitoring: Diagnostics and prognostics. IEEE Transactions on Industrial Informatics 8(4):964–973, DOI 10.1109/tii.2012.2205583

28. Soualhi A, Razik H, Clerc G, Doan DD (2013) Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system. IEEE Transactions on Industrial Electronics 61(6):2864–2874, DOI 10.1109/tie.2013.2274415
29. Rammohan R, Taha MR (2005) Exploratory investigations for intelligent damage prognosis using hidden Markov models. In: 2005 IEEE International Conference on Systems, Man and Cybernetics, IEEE, vol 2, pp 1524–1529, DOI 10.1109/ICSMC.2005.1571363
30. Tschope C, Wolff M (2009) Statistical classifiers for structural health monitoring. IEEE Sensors Journal 9(11):1567–1576, DOI 10.1109/jsen.2009.2019330
31. Mei H, Yuan S, Qiu L, Zhang J (2016) Damage evaluation by a guided wave-hidden Markov model based method. Smart Materials and Structures 25(2):025021, DOI 10.1088/0964-1726/25/2/025021
32. Ai C, Zhao H, Ma R, Dong X (2006) Pipeline damage and leak detection based on sound spectrum LPCC and HMM. In: Sixth International Conference on Intelligent Systems Design and Applications, IEEE, vol 1, pp 829–833, DOI 10.1109/isda.2006.215
33. Qiu J, Liang W, Zhang L, Yu X, Zhang M (2015) The early-warning model of equipment chain in gas pipeline based on DNN-HMM. Journal of Natural Gas Science and Engineering 27:1710–1722, DOI 10.1016/j.jngse.2015.10.036
34. Tejedor J, Macias-Guarasa J, Martins HF, Martin-Lopez S, Gonzalez-Herraez M (2019) A contextual GMM-HMM smart fiber optic surveillance system for pipeline integrity threat detection. Journal of Lightwave Technology 37(18):4514–4522, DOI 10.1109/jlt.2019.2908816
35. Torkamani S, Roy S, Barkey ME, Sazonov E, Burkett S, Kotru S (2014) A novel damage index for damage identification using guided waves with application in laminated composites. Smart Materials and Structures 23(9):095015, DOI 10.1088/0964-1726/23/9/095015
36. Bishop CM (2006) Pattern recognition and machine learning. Springer, DOI 10.1198/tech.2007.s518
37. Lebedev S (2016) hmmlearn/hmmlearn: Hidden Markov models in Python, with scikit-learn like API. `https://github.com/hmmlearn/hmmlearn`
38. Helske S, Helske J (2017) Mixture hidden Markov models for sequence data: The seqHMM package in R. arXiv preprint arXiv:170400543 pp 24–25
39. Xu Y, Luo M, Liu Q, Du G, Song G (2019) PZT transducer array enabled pipeline defect locating based on time-reversal method and matching pursuit de-noising. Smart Materials and Structures 28(7):075019, DOI 10.1088/1361-665X/ab1cc9
40. He S, Wang N, Ho M, Zhu J, Song G (2020) Design of a new stress wave communication method for underwater communication. IEEE Transactions on Industrial Electronics DOI 10.1109/tie.2020.3003634