Available online at www.sciencedirect.com
ScienceDirect
Journal of Sound and Vibration 311 (2008) 608–622
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Anomaly detection in flexible mechanical couplings via symbolic time series analysis

Amol Khatkhate[a], Shalabh Gupta[a], Asok Ray[a,*], Ravi Patankar[b]

[a]*The Pennsylvania State University, University Park, PA 16802, USA*
[b]*Intelligent Automation, Inc., 15400 Calhourn Drive, Suite 400, Rockville, MD 20855, USA*

Received 8 January 2006; received in revised form 11 September 2006; accepted 20 September 2007
Available online 26 November 2007

Abstract

Critical components of a rotating machinery such as bearings and couplings are often subjected to unbalanced axial and radial loads due to excessive machine vibrations arising from shaft misalignment(s). The paper presents Symbolic Time Series Analysis (STSA) of bearing acceleration data for detection and estimation of gradually developing parametric changes in flexible disc/diaphragm couplings. The analytical method is built upon the principles of Symbolic Dynamics, Automata Theory, and Statistical Pattern Recognition. The anomaly detection methodology is validated on a real-time simulation test bed, where the dynamic model of a flexible mechanical coupling is subjected to angular misalignment(s) leading to coupling failure. Damage patterns are identified from STSA of multiple data sets generated for different input torques. Statistical estimates are obtained for small changes in the coupling stiffness based on the information derived from the ensemble of damage patterns.

1. Introduction

The problem of shaft misalignment is encountered in diverse industrial applications, such as those involving turbo-machinery. Recent advances in the laser alignment technology [1] have made it possible to align a prime mover with the driven machine within a very fine tolerance. Despite these efforts, extremely precise alignment may not be possible due to rotor imbalance and lateral deflections of the rotating shafts. Consequently, harmonic forces are generated, which cause large stresses and premature failures in bearings, seals, and couplings and, in some cases, may lead to distortion of the rotor shaft(s). Shaft misalignment is categorized into two classes [2]:

(i) *Offset misalignment* that occurs if the centerlines of two shafts are parallel but do not meet at the power transfer point;
(ii) *Angular misalignment* that occurs if the non-parallel centerlines of two shafts intersect at the power transfer point.

*Corresponding author. Tel.: +1 814 865 6377; fax: +1 814 863 4848.
E-mail address: axr2@psu.edu (A. Ray).

0022-460X/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.jsv.2007.09.046

Nomenclature

$\mathscr{A}$	alphabet size
C_b	bearing rotation damping coefficient
D	window length on a symbolic sequence
F_b	force vector applied on bearing
g	acceleration due to gravity
H	entropy of the symbol sequence
I_2	inertia of the diaphragm along the diameter
J_z	shaft rotation inertia
k_b	bearing spring stiffness in x and y direction
k_ε	elastic constant corresponding to translation in the x direction
k_η	elastic constant corresponding to translation in the y direction
k_β	elastic constant corresponding to rotation about the x-axis
k_γ	elastic constant corresponding to rotation about the y-axis
K	diaphragm stiffness coefficient
m	mass of diaphragm
$\mathscr{M}_k$	anomaly measure computed at time epoch t_k
$\mathbf{p}_k$	state probability vector at time epoch t_k
$\mathscr{P}$	pattern matrix generated in the analysis part
r_c	radius of diaphragm
r_b	radius of bearing
T	torque applied on the shaft
x_b	vertical displacement of the bearing
x_c	vertical displacement of diaphragm center
y_b	horizontal displacement of the bearing
y_c	horizontal displacement of diaphragm center
β	rotation of diaphragm about x-axis
γ	rotation of diaphragm about y-axis
ε	displacement of diaphragm center in x direction
η	displacement of diaphragm center in y direction
Ω	rotation rate of the shaft about its bearing
Π_k	state transition matrix at time epoch t_k

Often misalignment in actual machinery exhibits a combination of both. Coupling failure, bearing failures, distorted rotors, and bearing housing damage are the commonly encountered effects of misalignment [3]. Proper shaft alignment reduces axial and radial loads with the associated reduction of noise, vibration, and local temperature rise. It also decreases wear of mechanical components and practically eliminates the possibility of shaft failure due to cyclic fatigue. Thus, availability, reliability, and maintainability are enhanced with the attendant reduction of life cycle cost [4].

Angular misalignment produces oscillatory forces that, in turn, generate cyclic stresses in the coupling materials. The stress amplitude becomes larger with increased angular misalignment and thereby accelerate fatigue failure of the coupling components. However, recent research [5] reveals that angular misalignment has a much smaller impact on bearing life than offset misalignment. The objective here is to investigate the effects of angular misalignment on the life of flexible couplings and infer the changes in coupling stiffness for prognosis of fatigue failure after prolonged usage under the following assumptions:

(1) The rotor shaft is sufficiently rigid so that the offset misalignment at steady state can be neglected.
(2) The initial angular misalignment under the nominal healthy condition is bounded within tolerance limits of 0.05° ($\sim$0.8 mills/in) [4].
(3) The load torque is statistically stationary with a small coefficient of variation (i.e., ratio of standard deviation to the mean).
(4) The shaft speed is proportional to the driving torque (e.g., positive displacement pumps, mixers, and extruders).

This paper presents a novel symbolic time series analysis (STSA)-based method [6,7] for early detection of fatigue damage in flexible couplings, primarily due to angular misalignment. The results of analysis are validated on a simulation test bed that is built upon a nonlinear dynamic model of a shaft undergoing the whirling phenomena [2]. Whirling is inevitable when the driver and driven shafts are coupled to each other and

are in motion. Although shafts are designed to avoid whirling (within specified tolerances), misalignments do occur over a period of time due to diverse reasons, such as thermal expansion/contraction and small differences in surface levels of the driver and driven test beds. The test-bed encompasses the effects of angular misalignment in the coupling by translating the exerted forces to the accelerometers mounted on the bearings. The information, needed for anomaly detection and damage analysis, is derived from the accelerometer readings. Statistical patterns of the accelerometer readings are primarily dependent on variations in effective stiffness of the coupling and the changes due to small torque fluctuations are relatively insignificant; the rationale is that loss of coupling stiffness induces much larger vibrations at the bearings. Model-based analysis provides ample insight into the physical phenomena; future work will involve fabrication of a laboratory apparatus for real-time experimental verification of the concept.

The paper is organized in six sections including the present one. Section 2 outlines the procedure for identification of statistical patterns from the time series data and formulates the problem of anomaly detection. Section 3 describes the simulation model of the coupling based on the whirling phenomena used for generation of the sensor time series. Section 4 introduces the underlying concepts of STSA for anomaly detection [6]. Section 5 discusses the solution procedure for anomaly detection and estimation problem and also presents the results of the estimation of the stiffness parameter of the coupling. Finally, the paper is concluded in Section 6 along with recommendations for future research.

2. Problem formulation for anomaly detection

Anomaly is defined as deviation from the nominal behavior of a dynamical system and is often associated with parametric and non-parametric changes that may gradually evolve in time. Anomaly detection is based on the analysis of observed time series data under known external stimuli (i.e., input excitation) or under self-excitation of the dynamical system [9]. The goal is to make inferences on occurrence of *slow-time-scale* anomalies based on observed changes in behaviorial pattern of the *fast-time-scale* process dynamics. From the above perspectives, anomaly detection in dynamical systems is formulated as a two-time-scale problem, in which the phase trajectories evolve in the fast time scale and anomalies, if any, progress in the slow-time scale.

Statistical pattern identification is categorized by two inter-related problems:

(i) the forward (or analysis) problem and the inverse
(ii) (or synthesis) problem.

The primary objective of the forward problem is to identify the patterns in the process dynamics and to track statistical changes occurring over the entire span of slow time. Specifically, the forward problem aims at detecting the deviations in the statistical patterns in the time series data, generated at slow-time epochs, from the nominal behavior pattern.

Solutions of the forward problem require the following steps:

(a) generation of multiple sensor time series data sets under self-excitation, or under external stimuli or using simulations, spanning the system behavior under different operational conditions;
(b) analysis of the data sets to characterize the system's behavioral pattern based on certain features at different slow-time epochs as the process evolves.

The inverse problem infers the anomalies based on the observed time series data and the information generated in the forward problem. The major role of the inverse problem is to provide information for monitoring the system behavior.

The entire framework for the problem is depicted in Fig. 1. The inputs to the model are: (i) the instantaneous torque that is assumed to be constant under steady-state operation; and (ii) range of the bending stiffness coefficients, k_β and k_γ, of the diaphragm. The outputs of the model are the bearing accelerometer readings that, in turn, are provided to the STSA subsystem. The output of the STSA subsystem is anomaly measure (see Section 4.3) curve which indicates the progress of anomaly (i.e., reduction in stiffness) in the disc/diaphragm coupling. Under multiple torque inputs in the range of 3000–6500 N m, a family of anomaly

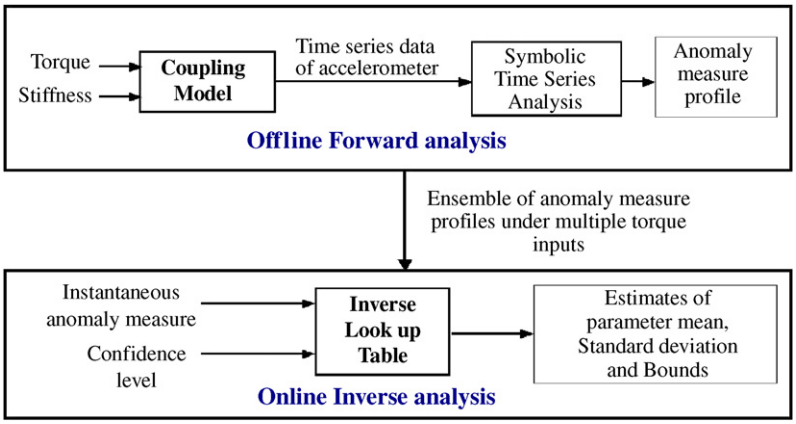

Fig. 1. Overall architecture of the STSA method for anomaly detection in flexible mechanical couplings.

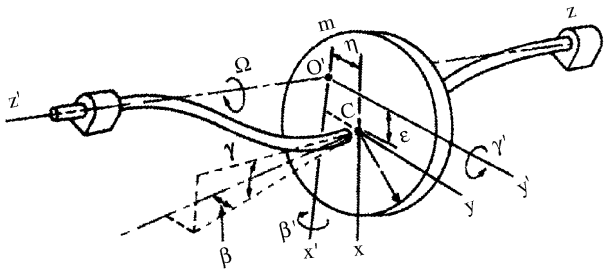

Fig. 2. Schematic of the disc/diaphragm coupling.

measure curves is generated that provides a stochastic pattern of anomaly growth (see Section 5.1). This pattern set is utilized by the inverse analysis section to generate an inverse look up table (see Section 5.2). This table provides the estimates of mean, std. deviation and confidence intervals of diaphragm stiffness for any arbitrary value of anomaly measure obtained in real time. The entire architecture has been built in Simulink within the MATLAB environment. The proposed architecture in Fig. 1 is application-independent and is portable to other applications.

3. Description of coupling simulation

This section presents nonlinear modelling of a flexible disc/diaphragm coupling, based on the whirling phenomena [2], to generate simulated data of accelerometer readings under different load excitation. The flexible coupling under consideration connects two shafts, end-to-end in the same line. The main objective is to transmit power/torque from one shaft to the other, causing both to rotate in unison, at the same rotational speed. Another objective is to compensate for minor amounts of misalignment and random movement between the two shafts. The misalignment is primarily caused due to dynamic movement (i.e. whirling) of shafts at high speeds.

Fig. 2 elucidates the effects of whirling phenomena in turbo machinery in which a rotating shaft undergoes deformation like a beam. The shaft supports the diaphragm that is assumed to be of negligible mass relative to the shaft and is flexible in bending but rigid in extension and torsion. The diaphragm is welded to the shaft, such that its center C coincides with the centerline of the shaft. The rotational speed of the shaft about its bearing is constant at Ω.

Let x–y–z be a reference frame that rotates at the constant angular speed Ω, with its z-axis concurrent with the line connecting the bearings. Let the deflected position of the center of the diaphragm, relative to the x, y, z

axes, be denoted as $\vec{r} = \varepsilon\vec{i} + \eta\vec{j}$. Furthermore, let x', y', and z' be the principal centroidal axes for the diaphragm, and let β and γ be the angular rotations about the x and y axes, respectively, of the x–y–z frame. For a shaft of symmetric cross section and a rotor concentrically situated at the midpoint of the shaft, each deformation is resisted solely by a corresponding proportional elastic force or moment; the elastic constants are k_ε, k_η, k_β, and k_γ.

The main assumptions in the development of the disc/diaphragm model are delineated below:

(1) The shaft is rigid (i.e., $k_\varepsilon = \infty$ and $k_\eta = \infty$), implying no translation or offset misalignment under steady state (i.e., $\varepsilon = 0$ and $\eta = 0$).
(2) The initial angular misalignment is non-zero within the tolerance limits of 0.05° (approx. 0.8 mills/in) [4] for the steady-state rotational speed range of 300–1000 rev/min.
(3) Changes in stiffness due to angular misalignment have the same effects along both the rotation axes (i.e., $k_\beta = k_\gamma = K$).
(4) The bearing stiffness, k_b, and bearing rotation damping coefficient, C_b, remain constant.
(5) Changes in the statistical patterns of the accelerometer readings are solely due to structural degradation of the coupling.
(6) Bearing life remains unaffected due to the misalignments under consideration.

The nonlinear equations of translational and rotational motion for the coupling dynamics are derived as follows:

$$\ddot{\varepsilon} + 2\Omega\dot{\eta} + k_\varepsilon/(m - \Omega^2)\varepsilon = g\cos\Omega t, \tag{1}$$

$$\ddot{\eta} - 2\Omega\dot{\varepsilon} + k_\eta/(m - \Omega^2)\eta = g\sin\Omega t, \tag{2}$$

$$\ddot{\beta} - 2\Omega\dot{\gamma} + k_\beta/(I_2 - \Omega^2)\beta = 0, \tag{3}$$

$$\ddot{\gamma} + 2\Omega\dot{\beta} + k_\gamma/(I_2 - \Omega^2)\gamma = 0, \tag{4}$$

where $I_2 = mr_c^2/4$ is the polar inertia of the disc/diaphragm. The governing equation for deflection disc/diaphragm is given as

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \begin{pmatrix} \cos\Omega t & -\sin\Omega t \\ \sin\Omega t & \cos\Omega t \end{pmatrix} \begin{pmatrix} \eta \\ \varepsilon \end{pmatrix}. \tag{5}$$

The force transmitted to the bearing due to translation and rotation (i.e., offset and angular misalignment) is given as

$$F_b = m\begin{pmatrix} \ddot{x}_c \\ \ddot{y}_c \end{pmatrix} + \frac{1}{r_b}\begin{pmatrix} \cos\Omega t & -\sin\Omega t \\ \sin\Omega t & \cos\Omega t \end{pmatrix} \begin{pmatrix} I_2\ddot{\beta}\sin\beta \\ I_2\ddot{\gamma}\sin\gamma \end{pmatrix}. \tag{6}$$

The torque is obtained as

$$T = J_z\dot{\Omega} + B\Omega, \tag{7}$$

where $J_z = (mr_c^2/2) + m(\eta^2 + \varepsilon^2)$ is the effective inertia of the diaphragm coupling. The governing equation for deflection of the bearing is given as

$$C_b\begin{pmatrix} \dot{x}_b \\ \dot{y}_b \end{pmatrix} + k_b\begin{pmatrix} x_b \\ y_b \end{pmatrix} = F_b. \tag{8}$$

The above set of equations is expressed in the state-space form as follows:

$$\begin{aligned}
&\mathbf{S}_1 = \varepsilon, \quad \mathbf{S}_2 = \dot{\varepsilon}, \quad \mathbf{S}_3 = \eta, \quad \mathbf{S}_4 = \dot{\eta}, \quad \mathbf{S}_5 = \beta, \quad \mathbf{S}_6 = \dot{\beta},\\
&\mathbf{S}_7 = \gamma, \quad \mathbf{S}_8 = \dot{\gamma}, \quad \mathbf{S}_9 = \Omega, \quad \mathbf{S}_{10} = x_b, \quad \mathbf{S}_{11} = y_b, \quad \dot{\mathbf{S}}_1 = \mathbf{S}_2, \quad \dot{\mathbf{S}}_3 = \mathbf{S}_4,\\
&\dot{\mathbf{S}}_5 = \mathbf{S}_6, \quad \dot{\mathbf{S}}_7 = \mathbf{S}_8,\\
&\dot{\mathbf{S}}_2 = -2\mathbf{S}_9\mathbf{S}_4 - k_\varepsilon/(m - \mathbf{S}_9^2)\mathbf{S}_1 + g\cos\mathbf{S}_9 t,\\
&\dot{\mathbf{S}}_4 = 2\mathbf{S}_9\mathbf{S}_2 - k_\eta/(m - \mathbf{S}_9^2)\mathbf{S}_3 + g\sin\mathbf{S}_9 t,\\
&\dot{\mathbf{S}}_6 = 2\mathbf{S}_9\mathbf{S}_8 - k_\beta/(I_2 - \mathbf{S}_9^2)\mathbf{S}_5,\\
&\dot{\mathbf{S}}_8 = -2\mathbf{S}_9\mathbf{S}_6 - k_\gamma/(I_2 - \mathbf{S}_9^2)\mathbf{S}_7,\\
&\dot{\mathbf{S}}_9 = \frac{T - B\mathbf{S}_9}{(mr_c^2/2) + m(\mathbf{S}_3^2 + \mathbf{S}_1^2)}
\end{aligned}$$

and

$$\begin{aligned}
\begin{bmatrix}\dot{\mathbf{S}}_{10}\\ \dot{\mathbf{S}}_{11}\end{bmatrix} = {}& \frac{k_b}{C_b}\begin{bmatrix}\mathbf{S}_{10}\\ \mathbf{S}_{11}\end{bmatrix} + \frac{m}{C_b}\begin{bmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{bmatrix}\begin{bmatrix}\mathbf{S}_3\\ \mathbf{S}_1\end{bmatrix} + \frac{2m}{C_b}\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}\begin{bmatrix}\mathbf{S}_4\\ \mathbf{S}_2\end{bmatrix}\\
&+ \frac{m}{C_b}\begin{bmatrix}\cos\mathbf{S}_9 t & -\sin\mathbf{S}_9 t\\ \sin\mathbf{S}_9 t & \cos\mathbf{S}_9 t\end{bmatrix}\begin{bmatrix}\dot{\mathbf{S}}_4\\ \dot{\mathbf{S}}_2\end{bmatrix} + \frac{I_2}{r_b}\begin{bmatrix}\cos\mathbf{S}_9 t & -\sin\mathbf{S}_9 t\\ \sin\mathbf{S}_9 t & \cos\mathbf{S}_9 t\end{bmatrix}\begin{bmatrix}\dot{\mathbf{S}}_6\\ \dot{\mathbf{S}}_8\end{bmatrix},
\end{aligned}$$

where

$$\begin{aligned}
T_{11} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)^2\cos\mathbf{S}_9 t - 2\dot{\mathbf{S}}_9\sin\mathbf{S}_9 t,\\
T_{12} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)^2\sin\mathbf{S}_9 t - 2\dot{\mathbf{S}}_9\cos\mathbf{S}_9 t,\\
T_{21} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)^2\sin\mathbf{S}_9 t - 2\dot{\mathbf{S}}_9\cos\mathbf{S}_9 t,\\
T_{22} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)^2\cos\mathbf{S}_9 t - 2\dot{\mathbf{S}}_9\sin\mathbf{S}_9 t,\\
U_{11} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)\sin\mathbf{S}_9 t,\\
U_{12} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)\cos\mathbf{S}_9 t,\\
U_{21} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)\cos\mathbf{S}_9 t \quad \text{and}\\
U_{22} &= -(\mathbf{S}_9 + \dot{\mathbf{S}}_9 t)\sin\mathbf{S}_9 t.
\end{aligned}$$

The bearing acceleration readings were generated using the following equations:

$$\ddot{x}_b = \ddot{S}_{10}, \tag{9}$$

$$\ddot{y}_b = \ddot{S}_{11}. \tag{10}$$

4. Symbolic time series analysis

This section briefly describes the core concept of STSA, which has been recently reported in literature [6]. Symbolic time series analysis takes advantage of the information generated by partitioning the time series data in its wavelet domain. The details of STSA method are presented in Ref. [6] while the concepts are briefly described here for the sake of completeness of the paper.

A data sequence (e.g., time series data) is converted to a symbol sequence by partitioning a compact region Ω of the phase space, over which the trajectory evolves, into finitely many discrete blocks. Let $\{\Phi_1, \Phi_2, \ldots, \Phi_m\}$ be a partitioning of Ω, such that it is exhaustive and mutually exclusive set, i.e.,

$$\bigcup_{j=1}^{m}\Phi_j = \Omega \quad \text{and} \quad \Phi_j \cap \Phi_k = \emptyset \ \ \forall j \neq k. \tag{11}$$

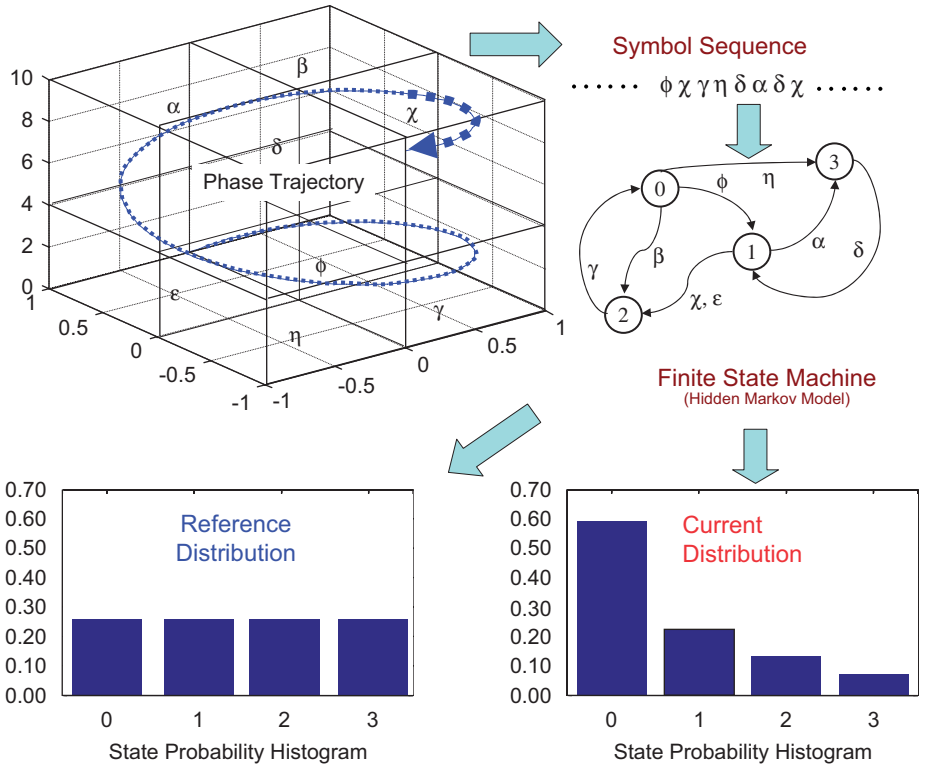

Fig. 3. Conceptual view of STSA.

Each block Φ_j is labelled as the symbol $\sigma_j \varepsilon \mathscr{A}$, where the symbol set $\mathscr{A}$ is called the *alphabet set* consisting of m different symbols (i.e., $m = |\mathscr{A}|$). As the system evolves in time, it travels through various blocks in its phase space and the corresponding symbol $\sigma_j \varepsilon \mathscr{A}$ is assigned to it, thus converting a data sequence to a symbol sequence $\ldots \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_k} \ldots$.

Fig. 3 elucidates partitioning of a compact (i.e., closed and bounded) region of the phase space and a mapping from the partitioned space into the symbol alphabet, which becomes a representation of the system dynamics defined by the trajectories. It also shows the construction of a hidden Markov model [8] from the symbol sequence as a finite state machine. The quasi-stationary probability histograms of the states represent patterns that are indicatives of the nominal (or reference) and anomalous behavior of the dynamical system, as explained in subsections below.

4.1. Wavelet space partitioning

Several partitioning techniques have been reported in literature for symbol generation [10,11]. Recent literature has reported wavelet-based partitioning approach [6,12], where wavelet transform [13] was shown to be particularly effective for noisy data from high-dimensional dynamical systems. In this method, the time series data is first converted to the wavelet transform data, where wavelet coefficients are generated at different scales and time shifts. The graphs of wavelet coefficients versus scale, at selected time shifts, are stacked starting with the smallest value of scale and ending with its largest value and then back from the largest value to the smallest value of the scale at the next instant of time shift. The arrangement of the resulting *scale series* data in the wavelet space is then partitioned with an alphabet size $|\mathscr{A}|$ by horizontal lines such that the regions with more information are partitioned finer and those with sparse information are partitioned coarser. In this approach, the Shannon entropy is maximized by the partition that induces uniform probability distribution of

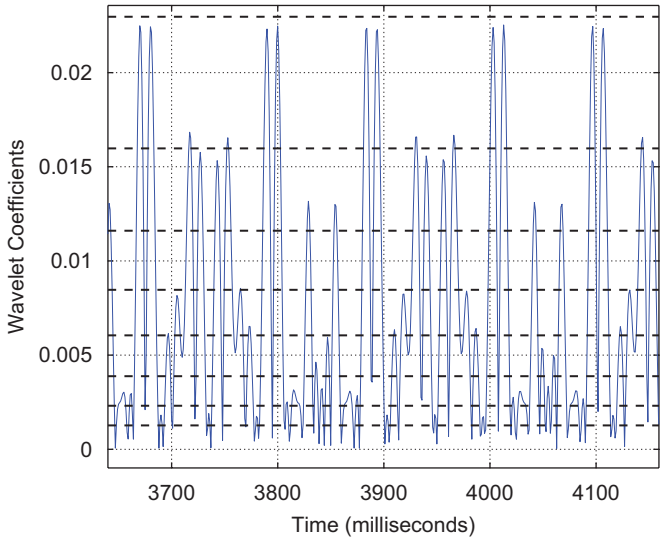

Fig. 4. Maximum entropy partitioning with alphabet size $\mathscr{A} = 8$ in the wavelet domain.

the symbols in the symbol alphabet [12]. Shannon entropy is defined as

$$H = -\sum_{i=1}^{|\mathscr{A}|} p_i \log(p_i), \tag{12}$$

where p_i is the probability of the ith state and summation is taken over all possible states. Uniform probability distribution of states is a consequence of the maximum entropy partitioning. Also, this paper has utilized the absolute value of the scale series data points for partitioning because of the symmetry of data points about the mean position. Fig. 4 shows an example of the maximum entropy partitioning in the wavelet space for alphabet size $|\mathscr{A}| = 8$, where the partitioned regions are marked by symbols ranging from 0 to 7. As such, if a data point in the wavelet domain falls in a particular partitioning region it is assigned the corresponding symbol of that region and in this manner the data in the wavelet domain is transformed to a symbol sequence. The combination of symbolic dynamics and wavelet space partitioning enables compression of the salient information from multivariate statistics in the form of univariate statistics. Once the symbol sequence is obtained the next step is to construct a finite state machine as described in the following section.

4.2. D-Markov machine construction

The partitioning as described in the previous subsection is performed at time epoch t_0 of the nominal condition that is chosen to be a healthy condition having zero anomaly measure. A finite state machine is then constructed, where the states of the machine are defined corresponding to a given *alphabet* $\mathscr{A}$ and window length D. The alphabet size $|\mathscr{A}|$ is the total number of partitions while the window length D is the length of consecutive symbol words forming the states of the machine [6]. The states of the machine are chosen as all possible words of length D from the symbol sequence, thereby making the number n of states to be equal to the total permutations of the alphabet symbols within word of length D, (i.e., $n \leqslant |\mathscr{A}|^D$). The choice of $|\mathscr{A}|$ and D depends on specific experiments, noise level and also the available computation power. A large *alphabet* may be noise-sensitive while a small alphabet could miss the details of signal dynamics. Similarly, a high value of D is extremely sensitive to small signal distortions but would lead to larger number of states requiring more computation power. For machine construction, the window of length D on the symbol sequence $\ldots \sigma_{i_1}\ \sigma_{i_2} \ldots \sigma_{i_k} \ldots$ is shifted to the right by one symbol, such that it retains the last $(D-1)$ symbols of the previous state and appends it with the new symbol σ_{i_ℓ} at the end. The symbolic permutation in the current window gives rise to a new state. The machine constructed in this fashion is called D-Markov machine [6] because of its Markov properties.

Definition 1. A symbolic stationary process is called D-Markov if the probability of the next symbol depends only on the previous D symbols, i.e., $P(\sigma_{i_0}/\sigma_{i_{-1}}\ldots.\sigma_{i_{-D}}\sigma_{i_{-D-1}}\ldots.) = P(\sigma_{i_0}/\sigma_{i_{-1}}\ldots.\sigma_{i_{-D}})$.

The finite state machine constructed above has D-Markov properties because the probability of occurrence of symbol σ_{i_ℓ} on a particular state depends only on the configuration of that state, i.e., previous D symbols. For example, if $\mathscr{A} = \{0, 1\}$, i.e., $|\mathscr{A}| = 2$ and $D = 2$, then the number of states is $n \leqslant |\mathscr{A}|^D = 4$; and the possible states are Q = {00, 01, 10, 11}, some of which may be forbidden.

Once the partitioning alphabet $\mathscr{A}$ and word length D are determined at the nominal condition (time epoch t_0), they are kept constant for all (slow time) epochs $\{t_1, t_2, \ldots t_k \ldots.\}$, i.e. the structure of the machine is fixed at the nominal condition. The states of the machine are marked with the corresponding symbolic word permutation and the edges joining the states indicate the occurrence of an event σ_{i_ℓ}.

Definition 2. The probability of transitions from state q_j to state q_k belonging to the set Q of states under a transition $\delta : Q \times \mathscr{A} \to Q$ is defined as

$$\pi_{jk} = P(\sigma \in \mathscr{A} | \delta(q_j, \sigma) \to q_k); \sum_k \pi_{jk} = 1. \tag{13}$$

Thus, for a D-Markov machine, the irreducible stochastic matrix $\Pi \equiv [\pi_{ij}]$ describes all transition probabilities between states such that it has at most $|\mathscr{A}|^{D+1}$ non-zero entries. The left eigenvector p corresponding to the unit eigenvalue of Π is the state probability vector under the (fast time scale) stationary condition of the dynamical system [6]. On a given symbol sequence $\ldots.\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_l}\ldots.$ generated from the time series data collected at slow-time epoch t_k, a window of length (D) is moved by keeping a count of occurrences of word sequences $\sigma_{i_1}\cdots\sigma_{i_D}\sigma_{i_{D+1}}$ and $\sigma_{i_1}\cdots\sigma_{i_D}$ which are, respectively, denoted by $N(\sigma_{i_1}\cdots\sigma_{i_D}\sigma_{i_{D+1}})$ and $N(\sigma_{i_1}\cdots\sigma_{i_D})$. Note that if $N(\sigma_{i_1}\cdots\sigma_{i_D}) = 0$, then the state $q \equiv \sigma_{i_1}\cdots\sigma_{i_D} \in Q$ has zero probability of occurrence. For $N(\sigma_{i_1}\cdots\sigma_{i_D}) \neq 0$, the transition probabilities are then obtained by these frequency counts as follows:

$$\pi_{jk} \equiv P[q_k|q_j] = \frac{P(\sigma_{i_1}\cdots\sigma_{i_D}\sigma)}{P(\sigma_{i_1}\cdots\sigma_{i_D})} \approx \frac{N(\sigma_{i_1}\cdots\sigma_{i_D}\sigma)}{N(\sigma_{i_1}\cdots\sigma_{i_D})}, \tag{14}$$

where the corresponding states are denoted by $q_j \equiv \sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_D}$ and $q_k \equiv \sigma_{i_2}\cdots\sigma_{i_D}\sigma$.

The time series data under the nominal condition (set as a benchmark) generates the *state transition matrix* Π_0 that, in turn, is used to obtain the *state probability vector* p_0 whose elements are the stationary probabilities of the state vector, where p_0 is the left eigenvector of Π_0 corresponding to the (unique) unit eigenvalue. A stopping rule is implemented to determine the minimum length of the symbol sequence that is required to generate the probability vector at a particular slow-time epoch [14].

4.3. Computation of anomaly measure

Subsequently, state probability vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_k \ldots.$ are obtained at slow-time epochs $t_1, t_2, \ldots t_k \ldots.$ based on the respective time series data. The behavioral changes from nominal condition are described as anomalies which are characterized by a scalar called *Anomaly Measure* ($\mathscr{M}$). The anomaly measure at slow-time epoch t_k is obtained as

$$\mathscr{M}_k \equiv d(\mathbf{p}_k, \mathbf{p}_0), \tag{15}$$

where $d(\bullet, \bullet)$ is an appropriately defined distance function. In this case, the distance function is chosen to be the standard Euclidean norm of the difference between the probability vectors. As such, anomaly measure is given by

$$\mathscr{M}_k \equiv \left[\sum_{j=1}^{n} (|\mathbf{p}_k(j) - \mathbf{p}_0(j)|)^2\right]^{1/2}. \tag{16}$$

4.4. Advantages of STSA

After having discussed the underlying principles and essential features of STSA-based anomaly detection, the major advantages of the technique for anomaly detection are listed below:

(1) The procedure is robust to measurement noise and spurious disturbances [15] and it filters out the noise at different steps. First of all, coarse graining of the continuous data (i.e., partitioning into finite blocks) and generation of a symbol sequence eliminate small measurement noise [6]. Secondly, the wavelet transform also contributes in signal–noise separation of the raw time series data by proper choice of scales [15]. Finally, the state probabilities are generated by passing a long symbol sequence over the finite state machine, which further eliminates small (zero-mean) measurement noise.
(2) Adaptability to low-resolution sensing due to coarse graining in space partitions [6].
(3) Capability for early detection of anomalies because of sensitivity to signal distortion and real-time execution on commercially available inexpensive platforms [7].
(4) Applicability to networked communication systems due to the capability of data compression into low-dimensional pattern vectors.

5. Data acquisition and results

This section presents the data acquisition and the solution procedures for both the forward and the inverse problems using the STSA method.

5.1. Solution procedure of the forward problem

The objective of the forward problem is to generate a pattern set which is representative of damage evolution under different excitation levels. The requirement of the forward problem is to create time series data of system responses under both healthy and anomalous conditions and to define an appropriate anomaly measure which can identify subtle changes in data sets and is capable of recognizing small deviations from the nominal behavior. To solve the forward problem, time series data sets are generated for constant torque (or speed) levels under both nominal condition (i.e., when the stiffness parameter has the nominal value and it is not degraded due to structural damage) and anomalous conditions (i.e., when the stiffness of the coupling has degraded from the nominal value due to structural damage). One of the major features of the STSA technique is that the anomaly is quantified relative to the nominal condition, which is assigned zero anomaly measure. Furthermore, uncertainties due to small (zero-mean) torque fluctuations and exogenous disturbances are largely attenuated due to the filtering effects of wavelet-based symbolic dynamic analysis [15].

In this paper, the bending stiffness parameters along both the x and y axes are assumed to be affected to an equal extent. Initially, the disc is allowed to rotate freely at an angular speed of 40 rad/s ($\sim$380 rev/min). The steady-state response of the accelerometer data is recorded for analysis when the angular speed reaches a constant value after $\sim$0.5 s. In each run, 8000 data points are generated at the sampling frequency of 1.0 kHz. For a single excitation level, system responses are measured under different values of the stiffness parameter that is reduced by 5% for each run until 75% of the nominal value is reached (see Fig. 5), where the system is considered to be no longer usable and hence out of service.

Anomaly measures are then calculated using the procedure described in Section 4 with respect to the nominal condition. Similarly, time series data are generated under both nominal and anomalous conditions for multiple torque levels in the operating range of 3000–6500 N m, separated at an interval of 100 N m, thereby generating an ensemble of data sets for multiple excitations. Anomaly measures for different stiffness values are then calculated for each excitation level as mentioned above. In practice, accelerometers are required to be routinely calibrated before usage for data acquisition. Furthermore, variations in the parameters (e.g., frequency bandwidth) between successive maintenance actions are usually insignificant. Therefore, having known that there is no significant change in the load torque, any abrupt changes in the anomaly measure can be attributed to large failure(s) (e.g., due to sensor malfunction). The effects of bias errors (i.e., low-frequency disturbances) and unbiased (i.e., zero-mean) random noise can be detected and

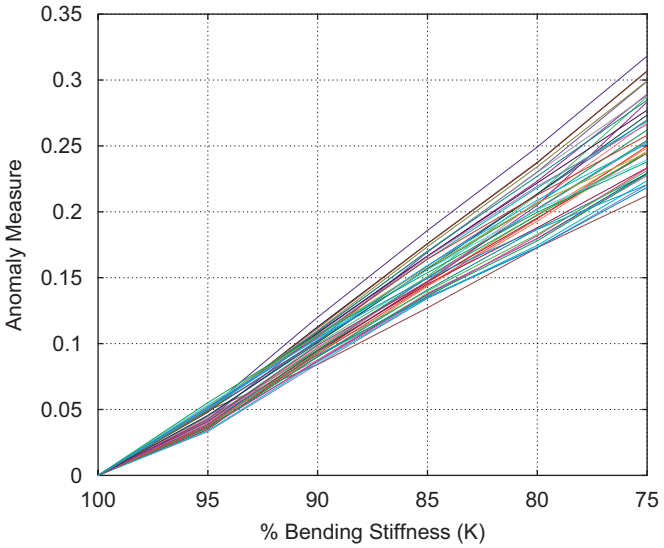

Fig. 5. Family of anomaly measure plots for different levels of constant torque inputs in the range of 3000–6500 N m vs. diaphragm stiffness K.

compensated by the STSA technique. For example, the system dynamics are likely to change due to bias errors and therefore will call for calibration, which is done in the forward problem. The unbiased random noise is practically eliminated in the STSA algorithm [15].

The resultant $\ddot{z}_b$ of the bearing accelerometer measurements $\ddot{x}_b$ and $\ddot{y}_b$ in the x and y directions, respectively, is used for generating the anomaly measures using the STSA algorithm, where

$$\ddot{z}_b = \sqrt{\ddot{x}_b^2 + \ddot{y}_b^2}. \tag{17}$$

The alphabet size, window length, and the wavelet basis function were chosen as $A = 8$, $D = 1$ and "*gaus*2" [16], respectively (see Section 4.1), for all time series data sets collected for different stiffness values and torque levels. Increasing the alphabet size further did not show any appreciable improvement in the results and increasing the value of depth D created a much larger number of states of the finite state machine, many of them having very small or zero probabilities. The finite state machine constructed with the above choice of parameters has 8 states; the algorithm is computationally fast and can be used on inexpensive processors for real-time execution. The wavelet basis "*gaus*2" provided better results than many other wavelets of the Daubechie's family. The reason for this choice is attributed to close matching of the shapes of the signal and the wavelet basis function.

Fig. 5 exhibits the profiles of anomaly measures for varying stiffness under different excitations, separated by 100 N m in the range of 3000–6500 N m. This ensemble of the anomaly measure plots as seen in Fig. 5 forms a pattern set for anomaly detection under different excitations in the operating range of 3000–6500 N m. The ensemble consists of $\ell = 42$ plots derived from the simulation data generated in the operating range. This pattern set has been used to solve the inverse problem for estimating the stiffness value at any particular instant and torque level in the operating range, as presented in the next section.

5.2. Solution procedure of the inverse problem

The objective of the inverse problem is identification of anomalies and estimation of the fault parameters based on the family of anomaly measure profiles generated in the forward problem. It is essential to detect changes in the stiffness of the coupling, due to evolving fatigue damage, so that appropriate remedial action(s) can be taken before the onset of widespread damage. It is also necessary to relate changes in stiffness to the magnitude and location of incipient damage. The rationale is that loss of stiffness of the coupling may lead to

further misalignment, excessive noise and vibrations causing rapid failure of other rotor components. Hence, the estimated change in stiffness during the operating period is crucial to allow for scheduled operation and maintenance. The following procedure for stiffness estimation is based on the statistical pattern changes in the observed time series data.

Statistical techniques have been utilized for structural health monitoring and failure prognosis of mechanical systems in the past few decades [17]. For example, a commonly used tool is principal component analysis that relies on the estimation of eigenvalues and eigenvectors [18,19] of the dynamical system. These parameters are dependant on the mass, stiffness and damping properties and as well as on the structural flaws in the system. They provide critical information about the characteristics and behavior of dynamical systems. In the approach described in this section, the objective is to estimate the critical parameter(s) of the system, based on the observed time series data response. Statistical analysis of the pattern set of the anomaly measure profiles (see procedure outlined in Section 5.1) provides the estimate of the coupling stiffness.

This section describes the procedure for parameter (i.e., stiffness) estimation of flexible mechanical couplings subjected to exogenous or self excitation in the operating range. For solution of the inverse problem (i.e., estimation of stiffness at any instant under arbitrary load), time series data is generated under both nominal condition and arbitrarily chosen anomalous conditions (e.g., between 100% and 75% of the nominal value of stiffness). Input excitation can be arbitrarily chosen in the range of 3000–6500 N m. Anomaly measure is then calculated with respect to the nominal condition based on the observed time series data as described in Section 5.1. It is considered here that the exact determination of the *stiffness* parameter based on the derived value of anomaly measure at any particular instant under arbitrary load is not possible due to the variation observed in the forward problem. Therefore, the *stiffness* parameter is treated as a random variable because of the uncertainty in determining its exact value. It is assumed here that the only known quantity is the anomaly measure value which is derived from the observed time series data.

The range of anomaly measure $\mathscr{M}$ (i.e. the ordinate in Fig. 5) is discretized into n uniformly spaced levels. A pattern matrix $\mathscr{P}$ of dimension $\ell \times n$ is then derived from the anomaly measure plots as shown in Fig. 5. Each column of the pattern matrix $\mathscr{P}$ corresponds to the stiffness values calculated for ℓ torque inputs at a particular anomaly measure. The elements belonging to a particular column of the pattern matrix $\mathscr{P}$ are distributed over a certain range of stiffness values. For the elements of each column, a two-parameter *lognormal* distribution [20,21,17] is hypothesized, and its goodness of fit is examined by both χ^2 and Kolmogorov–Smirnov tests [22]. Fig. 6 compares the lognormal-distributed probability density functions (pdf's) of the diaphragm stiffness K with the corresponding histograms generated at four different anomaly measure values in the range. The number of bins were taken to be $r = 8$. With $f = r - 2 - 1 = 5$ degrees of freedom for the two-parameter lognormal distribution, the χ^2-test shows that, for each of the n data sets, the hypothesis passed the 10% significance level [22] which suffices the conventional standard of 5% significance level. Also, for each of the n data sets, the hypothesis passed the 30% significance level of the Kolmogorov–Smirnov test which again suffices the conventional standard of 5% significance level. A good discussion of these statistical tests is provided in Ref. [22]. The rationale for the choice of the lognormal distribution is presented below:

- The fact that the lognormal distribution is one directional on the positive axis is consistent with the useful life of a structure.
- The probability of failure is generally high for a certain range of life and gradually decreases thereafter. This behavior is easily modelled by the lognormal distribution, i.e., the failure rate shapes and the pdf of lognormal distribution are physically similar.
- Since the logarithm of the stiffness parameter is Gaussian, many standard statistical tools are available for further analysis. This makes the lognormal distribution a very natural choice for failure analysis.

Once the fits of lognormal distribution are obtained, the bounds for different confidence intervals are computed [22,23]. Confidence intervals (CI) are obtained at a certain confidence level such as 95%. This implies that at a particular value of the (real time) anomaly measure, the confidence interval is a measure of the spread of the stiffness with the specified level of confidence. As such, for a confidence level of 95%, the

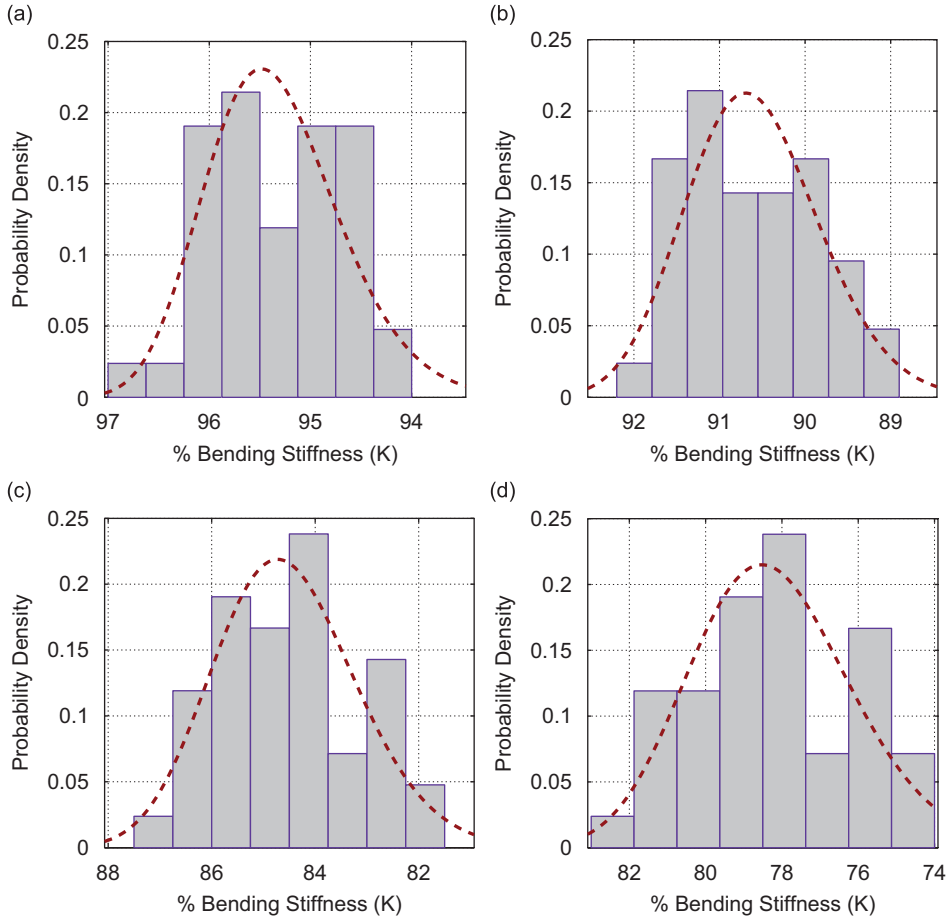

Fig. 6. Plots of actual probability distribution of diaphragm stiffness K and corresponding log-normal fit for four levels of anomaly measure ($\mathscr{M}$): (a) 0.0445, (b) 0.0954, (c) 0.1591, (d) 0.2227.

probability that the actual stiffness will lie between the obtained confidence level is 95%. Fig. 7 shows the bounds for 99%, 95% and 90% confidence levels.

For validation of the proposed methodology, time series data have been generated at three different arbitrary torque levels in the range of 3000–6500 N m for the nominal value of the stiffness parameter. Also to compute the anomaly measure using STSA (see Section 5.1), time series data is generated at three randomly selected stiffness values between 75% and 100% of the nominal value, for all three torque levels, as seen in Table 1.The chosen torque levels in the inverse problem are deliberately made to be different from those used in the forward problem (see Section 5.1) but each of them lies within the operating range of 3000–6500 N m. For the generated anomaly measure values, the mean diaphragm stiffness $\hat{K}$, its standard deviation $\hat{\sigma}$ and confidence intervals are estimated at three confidence levels of 99%, 95%, 90% from the properties of the lognormal distribution [22] and using the solution procedure of the inverse problem (see Section 5.2). Table 1 provides the actual stiffness values used for generation of anomaly measures, and the estimated parameters obtained using the inverse problem solution. As seen in Table 1, the confidence intervals become tighter as the confidence level reduces.

6. Summary, conclusions, and future work

This paper presents the concept of a novel method for identification of statistical patterns for damage detection in rotating machinery and its validation on a simulation test bed. The underlying principle of the

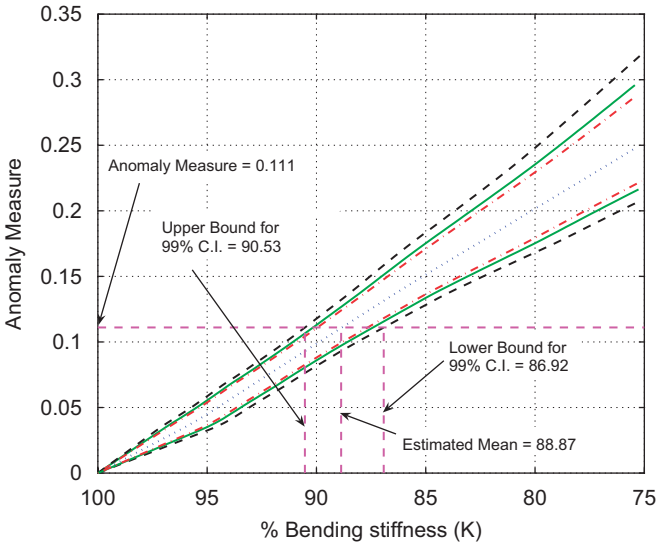

Fig. 7. Plots of confidence interval bounds of the log-normal distribution are shown at 99% (outermost bounds), 95% (middle bounds) and 90% (innermost bounds) confidence levels. Estimated stiffness is plotted as the mean (centerline) of the distribution. Bounds of confidence level 99% and estimated mean are shown at an arbitrary value of anomaly measure ($\mathscr{M} = 0.111$).

Table 1
Estimated values of diaphragm stiffness (K) for three different confidence levels for three different Torque loads

Torque (N m)	$\mathscr{M}$	K (%)	$\hat{K}$ (%)	$\hat{\sigma}$	99% C.I.		95% C.I.		90% C.I.	
					Lower	Upper	Lower	Upper	Lower	Upper
3800	0.0599	93	93.64	0.539	92.1	94.89	92.5	94.62	92.69	94.47
	0.1167	87	88.33	0.736	86.29	90.08	86.8	89.69	87.07	89.48
	0.1863	78	81.59	1.363	77.78	84.77	78.73	84.07	79.23	83.69
4800	0.0328	96	96.30	0.454	94.95	97.29	95.31	97.09	95.49	96.97
	0.111	88	88.87	0.700	86.92	90.52	87.41	90.16	87.66	89.96
	0.2038	79	79.88	1.498	75.65	83.37	76.73	82.59	77.27	82.18
5800	0.0181	98	97.96	0.251	97.21	98.51	97.41	98.39	97.51	98.33
	0.1013	90	89.78	0.645	87.98	91.3	88.44	90.96	88.67	90.78
	0.1882	82	81.41	1.380	77.53	84.62	78.52	83.91	79.02	83.52

proposed method is based on STSA [24,6,15,7] of the observed process variable(s). The *STSA* algorithm is built upon the concepts of Automata theory, information theory, and statistical pattern recognition.

Efficacy of the proposed STSA method for identification of behavior patterns is demonstrated on a real-time simulation test bed of mechanical structures, where the source of possible anomalies is the change in stiffness of the disc/diaphragm coupling due to angular misalignment. The damage information is extracted from accelerometer sensor signals mounted on bearings. Both forward and inverse problems in STSA-based pattern identification have been tested. The histograms on state probability distribution are generated from the observed time series data to serve as patterns of the evolving behavior change, resulting from stiffness reduction due to fatigue damage in flexible couplings.

The STSA method has been experimentally validated in real time on electronic system [15] and electromechanical system [7] test beds. The superiority of STSA over other pattern recognition techniques has been effectively demonstrated in a previous publication [25]. The reported work on STSA pattern identification is a step toward building a reliable instrumentation system for early detection of structural damage in human-engineered systems such as rotating machinery; further research and experimental

validation is necessary before its usage in industry. Future work is planned for fabrication of a disc/diaphragm coupling apparatus and installation of accelerometers at the site(s) of interest. Implementation and experimental validation of the STSA algorithm on such an experimental test bed is a topic of future research. Further theoretical research on choice of STSA parameters such as depth D and alphabet size $|\mathscr{A}|$ is suggested as an area of future work.

Acknowledgements

This work has been supported in part by the US Army Research Laboratory and the US Army Office under Grant No. DAAD19-01-1-0646.

References

[1] A. Luedeking, Getting the best alignment possible, *Pumps and Processes Magazine* (2002) pp. 28–30.
[2] J.H. Ginsberg, *Advanced Engineering Dynamics*, second ed., Cambridge University Press, New York, 1998.
[3] R.C. Eisenmann Sr., R.C. Eisenmann Jr., *Machinery Malfunction Diagnosis and Correction: Vibration Analysis and Troubleshooting for the Process Industries*, Prentice-Hall PTR, Upper Saddle River, NJ, 1997.
[4] J. Piotrowski, *Shaft Alignment Handbook*, second ed., Marcel Dekker Inc., New York, NY, 1995.
[5] J.W. Hines, S. Jesse, A. Edmondson, D. Nower, Motor shaft misalignment bearing load analysis, *Proceedings of the Maintenance and Reliability Conference (MARCON 99)*, Gatlinburg, TN, May 10–12, 1999.
[6] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, *Signal Processing* 84 (7) (2004) 1115–1130.
[7] S. Gupta, A. Ray, E. Keller, Symbolic time series analysis of ultrasonic data for early detection of fatigue damage, *Mechanical Systems and Signal Processing* 21 (2) (2007) 866–884.
[8] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech processing, *Proceedings of the IEEE* 77 (2) (1989) 257–286.
[9] H.D.I. Abarbanel, *The Analysis of Observed Chaotic Data*, Springer, New York, NY, 1996.
[10] R. Davidchack, Y. Lai, E. Bolt, H. Dhamala, Estimating generating partitions of chaotic systems by unstable periodic orbits, *Physical Review E* 61 (2000) 1353–1356.
[11] M.B. Kennel, M. Buhl, Estimating good discrete partitions form observed data: symbolic false nearest neighbors, *Physical Review E* 91 (8) (2003) 84–102.
[12] V. Rajagopalan, A. Ray, Wavelet-based space partitioning for symbolic time series analysis, *Proceedings of 44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, 2005.
[13] S. Mallat, *A Wavelet Tour of Signal Processing*, second ed., Academic Press, San Diego, CA, 1999.
[14] S. Gupta, A. Ray, A. Mukhopadhyay, Anomaly detection in thermal pulse combustors, *Proceedings of the I Mech E Part I Journal of Systems and Control Engineering* 220 (5) (2006) 339–351.
[15] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, *Signal Processing* 86 (11) (2006) 3309–3320.
[16] *Matlab Wavelet Toolbox*, Mathworks Inc., 2006.
[17] A. Ray, Stochastic measure of fatigue crack damage for health monitoring of ductile alloy structures, *Structural Health Monitoring* 3 (3) (2004) 245–263.
[18] G.M.L. Gladwell, Inverse problems in vibration, *Applied Mechanics Reviews* 39 (7) (1986) 1013–1018.
[19] R.A. Ibrahim, Structural dynamics with parameter uncertainties, *Applied Mechanics Reviews* 40 (3) (1987) 309–328.
[20] J.L. Bogdanoff, F. Kozin, *Probabilistic Models of Cumulative Damage*, Wiley, New York, 1985.
[21] NIST/SEMATECH e-Handbook of Statistical Methods ⟨http://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm⟩.
[22] H.D. Brunk, *An Introduction to Mathematical Statistics*, third ed., Xerox Publishing, Lexington, MA, 1975.
[23] G.W. Snedecor, W.G. Cochran, *Statistical Methods*, eighth ed., Iowa State University Press, Ames, Iowa, 1989.
[24] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, *Review of Scientific Instruments* 74 (2) (2004) 915–930.
[25] S. Chin, A. Ray, V. Rajagopalan, Symbolic time series analysis for anomaly detection: a comparative evaluation, *Signal Processing* 85 (9) (2005) 1859–1868.