ELSEVIER

Symbolic time series analysis via wavelet-based partitioning ☆

Venkatesh Rajagopalan, Asok Ray*

The Pennsylvania State University, University Park, PA 16802, USA

Abstract

Symbolic time series analysis (*STSA*) of complex systems for anomaly detection has been recently introduced in literature. An important feature of the *STSA* method is extraction of relevant information, imbedded in the measured time series data, to generate symbol sequences. This paper presents a wavelet-based partitioning approach for symbol generation, instead of the currently practiced method of phase-space partitioning. Various aspects of the proposed technique, such as wavelet selection, noise mitigation, and robustness to spurious disturbances, are discussed. The wavelet-based partitioning in *STSA* is experimentally validated on laboratory apparatuses for anomaly/damage detection. Its efficacy is investigated by comparison with phase-space partitioning.
© 2006 Elsevier B.V. All rights reserved.

1. Introduction

The concept of symbolic time series analysis (*STSA*) has been recently proposed for anomaly detection in complex systems [1]. Several case studies [2–5] in anomaly detection show that *STSA* can be more effective than existing pattern recognition techniques (e.g., principal component analysis and neural networks). The *STSA* method has also been demonstrated for fault detection in electromechanical systems, such as three-phase induction motors [6] and helical gearbox in rotorcraft [7].

A crucial step in *STSA* is extraction of relevant information, imbedded in the measured time series data, to generate symbol sequences. Symbol generation requires partitioning of the data space to obtain the symbol sequences [8,9]. Various partitioning techniques have been suggested in literature for symbol generation, which include variance-based [10], entropy-based [11], and hierarchical clustering [12] methods. A survey of clustering techniques is provided in [13]. In addition to these methods, another scheme of partitioning, based on *symbolic false nearest neighbors* (*SFNN*), was reported by Kennel and Buhl [14]. The objective of *SFNN* partitioning is to ensure that points that are close to each other in the symbol space are also close to each other in the phase space. Partitions that yield a smaller proportion of *SFNN* are considered optimal. However, this partitioning method may become computationally

very expensive if the dimension of the phase space is large or if the data set is contaminated by noise, since noise induces false symbols.

Partitioning of wavelet-transformed data potentially alleviates the above-mentioned shortcomings and is particularly effective with noisy data from high-dimensional dynamical systems. Usage of wavelet transform for symbolization has been recently introduced by the authors [1,15]. This paper elaborates the concept of wavelet-based partitioning for *STSA* and its major features are delineated below.

- Selection of the wavelet basis and scale range.
- Noise mitigation in the measured time series data prior to symbol generation.
- Robustness of extracted information from symbol sequences.
- Enhancement of computational efficiency for anomaly detection.
- Validation of the *STSA* method on laboratory apparatuses for anomaly/damage detection.

The paper is organized into six sections including the present section. Section 2 focuses on key technical aspects of wavelet analysis such as selection of wavelet basis and scales. Section 3 presents the maximum entropy method of partitioning the space of wavelet coefficients. Section 4 elucidates the underlying principles of wavelet-based partitioning via simulation examples. Section 5 presents experimental results on laboratory apparatuses to demonstrate efficacy of wavelet-based *STSA* for anomaly detection. Section 6 summarizes the paper and makes major conclusions along with recommendations for future research.

2. Wavelet analysis of time series data

This section presents generation of wavelet coefficients from measured time series data, and their arrangement for symbol generation. Specifically, issues of wavelet basis and scale range selection are addressed.

Preprocessing of time series data is often necessary for extraction of pertinent information. Fourier analysis is sufficient if the signal to be analyzed is stationary and if the time period is accurately known. However, Fourier analysis may not be appropriate if the signal has non-stationary characteristics such as drifts, abrupt asynchronous changes, and frequency trends. Wavelet analysis alleviates these difficulties via adaptive usage of long windows for retrieving low-frequency information and short windows for high-frequency information [16,17]. The ability to perform flexible localized analysis is one of the striking features of wavelet transform.

In multi-resolution analysis (*MRA*) of wavelet transform, a continuous signal $f \in \mathbb{H}$, where $\mathbb{H}$ is a Hilbert space, is decomposed as a linear combination of time translations of scaled versions of a suitably chosen scaling function $\phi(t)$ and the derived wavelet function $\psi(t)$. Let the sequence $\{\phi_{j,k}\}$ belong to another Hilbert space $\mathbb{M}$ with a countable measure, where the scale $s = 2^j$ and time translation $\tau = 2^{-j}k$. If the sequence $\{\phi_{j,k}\}$ is a frame for the Hilbert space $\mathbb{H}$ with a frame representation operator $\mathbb{L}$, then there are positive real scalars A and B such that

$$A\|f\|_{\mathbb{H}}^2 \leqslant \|\mathbb{L}f\|_{\mathbb{M}}^2 \leqslant B\|f\|_{\mathbb{H}}^2 \quad \forall f \in \mathbb{H}, \tag{1}$$

where $\mathbb{L}f = \{\langle f, \phi_{j,k}\rangle\}$ and $\|\mathbb{L}f\|_{\mathbb{M}}$ is an appropriate norm, e.g., $\|\mathbb{L}f\|_{\mathbb{M}} = \sqrt{\sum_j \sum_k |\langle f, \phi_{j,k}\rangle|^2}$ is a candidate norm; and $\langle x, y\rangle$ is the inner product of x and y, both belonging to $\mathbb{H}$.

The above relationship is a norm equivalence and represents the degree of coherence of the signal f with respect to the frame set of scaling functions; it may be interpreted as enforcing an approximate energy transfer between the domains $\mathbb{H}$ and $\mathbb{L}(\mathbb{H})$. In other words, for all signals $f \in \mathbb{H}$, a scaled amount of energy is distributed in the coefficient domain where the scale factor lies between A and B [16]. However, the energy distribution is dependent on the signal's degree of coherence with the underlying frame $\{\phi_{j,k}\}$. For a signal f, which is coherent with respect to the frame $\{\phi_{j,k}\}$, norm equivalence in the frame representation necessarily implies that a few coefficients contain most of the signal energy and hence have relatively large magnitudes. Similarly, pure noise signal w being incoherent with respect to the set $\{\phi_{j,k}\}$, must have a frame representation in which the noise energy is spread out over a large number of coefficients. Consequently, these coefficients have a relatively small magnitude [17].

Let $\tilde{f}$ be a noise corrupted version of the original signal f expressed as

$$\tilde{f} = f + \sigma w, \tag{2}$$

where w is additive white gaussian noise with zero mean and unit variance and σ is the noise level.

Then, the inner product of $\tilde{f}$ and $\phi_{j,k}$ is obtained as

$$\langle \tilde{f}, \phi_{j,k} \rangle = \underbrace{\langle f, \phi_{j,k} \rangle}_{\text{signal part}} + \sigma \underbrace{\langle w, \phi_{j,k} \rangle}_{\text{noise part}}. \qquad (3)$$

The noise part in Eq. (3) may further be reduced if the scales over which coefficients are obtained are properly chosen.

For every wavelet, there exists a certain frequency called the center frequency F_c that has the maximum modulus in the Fourier transform of the wavelet [18]. The pseudo-frequency f_p of the wavelet at a particular scale α is given by the following formula [18,19]:

$$f_p = \frac{F_c}{\alpha \Delta t}, \qquad (4)$$

where Δt is the sampling interval. Fig. 1 depicts the center frequency associated with the Daubechies 4 wavelet 'db4' [16,20].

The power spectral density (PSD) of the signal provides the information about the frequency content of the signal. This information along with Eq. (4) can be used for scale selection. The procedure of selecting the scales is summarized below:

• Identification of the frequencies of interest through PSD analysis of time series data.
• Substitution of the above frequencies in place of f_p in Eq. (4) to obtain the respective scales in terms of the known parameters F_c and Δt.

The wavelet coefficients of the signal are significantly large when the pseudo-frequency f_p of the wavelet corresponds to the locally dominant frequencies in the underlying signal. Example 1 in Section 4 illustrates how the choice of wavelet and scale affect the coefficients. Examples 2 and 3
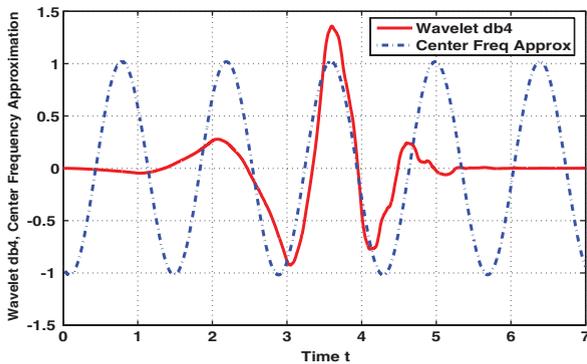
Fig. 1. Center frequency approximation for wavelet db4.

illustrate noise suppression and robustness. Three experimental studies in Section 5 illustrate enhancement of anomaly/damage detection using STSA and make comparisons of wavelet-based partitioning and SFNN phase-space partitioning in this regard.

Upon selection of the wavelet basis and scale range, the wavelet coefficients are obtained. These coefficients are stacked at selected time-shift positions, 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. In the sequel, this one-dimensional array of arranged wavelet coefficients is called the *scale series* data, which is structurally similar to time series data in the phase space. For symbol generation, the scale series data can be handled in a similar way as time series data.

3. Symbolization of scale series data

This section presents symbolization of wavelet coefficients by maximum entropy partitioning. The scale series data are partitioned to construct the symbol alphabet and to generate symbol sequences.

In the wavelet-based partitioning scheme, as introduced in an earlier publication [1], the maximum and minimum of the scale series are evaluated and the ordinates between the maximum and minimum are divided into equal-sized regions. These regions are mutually disjoint and thus form a partition. Each region is then labelled with one symbol from the alphabet. If the data point lies in a particular region, it is coded with the symbol associated with that region. Thus, a sequence of symbols is created from a given sequence of scale series data. This type of partitioning is called uniform partitioning in the sequel. The left-hand plot in Fig. 2 depicts uniform partitioning of noise contaminated signal $\sin(2\pi t)$ with alphabet size $|\Sigma| = 4$. Note that the partition segments are of equal size.

Intuitively, it is more reasonable if the information-rich regions of the data set are partitioned finer and those with sparse information are partitioned coarser. To achieve this objective, a partitioning method is adopted such that the entropy of the generated symbol sequence is maximized [15]. Maximum entropy partitioning is abbreviated as ME partitioning in the sequel. The procedure for obtaining an ME partition is described below.

Let N be the length of the data set and $|\Sigma|$ be the size of the symbol alphabet (i.e., the number of
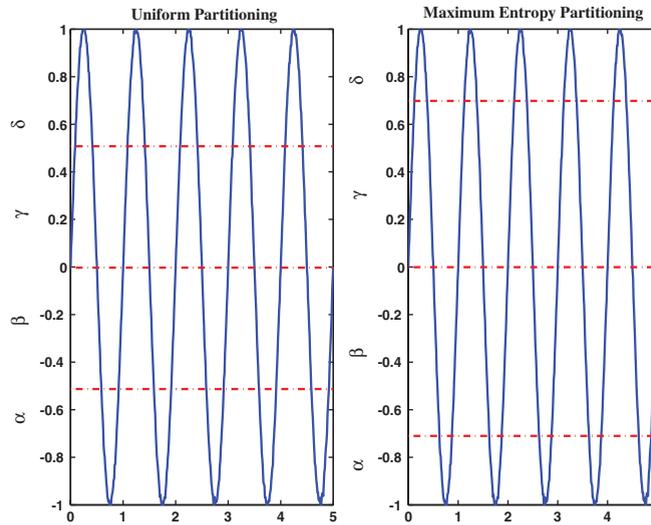
Fig. 2. Examples of uniform and ME partitioning with $|\Sigma| = 4$.

disjoint elements in the partition). The data set is sorted in ascending order. Starting from the first point in the sorted data, every consecutive data segment of length $\lfloor N/|\Sigma| \rfloor$ forms a distinct element of the partition, where $\lfloor x \rfloor$ represents the greatest integer less than or equal to x.

The right-hand plot in Fig. 2 shows ME partitioning for the noise-contaminated signal $\sin(2\pi t)$ with $|\Sigma| = 4$. As expected, the size of the partitions are not equal, but the probabilities of the symbols are equal. Variations in data patterns are more likely to be reflected in the symbol sequence obtained under ME partitioning than other partitioning.

The choice of the alphabet size $|\Sigma|$ plays a crucial role in STSA. For example, a small value of $|\Sigma|$ may prove inadequate for capturing the characteristics of the scale series data. On the other hand, a large value of $|\Sigma|$ may lead to redundancy and waste of computational resources.

Selection of the alphabet size $|\Sigma|$ is an area of active research; an entropy-based approach has been adopted for selecting $|\Sigma|$ in this paper. Let $H(k)$ denote the Shannon entropy of the symbol sequence obtained by partitioning the data set with k symbols:

$$H(k) = -\sum_{i=1}^{i=k} p_i \log_2 p_i, \qquad (5)$$

where p_i represents the probability of occurrence of the symbol σ_i. Note that $H(1) = 0$ because $p_i = 0$ or 1 with $i = 1$. If the underlying data set has sufficient information content, then the entropy achieved

under ME partitioning would be $\log_2(k)$, which corresponds to the uniform distribution. We define a quantity $h(\cdot)$ to represent the change in entropy with respect to the number $|\Sigma|$ of symbols as

$$h(k) \triangleq H(k) - H(k-1) \quad \forall k \geqslant 2. \qquad (6)$$

The algorithm for alphabet size selection is given below.

Step 1: Set $k = 2$. Choose a threshold ε_h, where $0 < \varepsilon_h \ll 1$.

Step 2: Sort the data set (of length N) in the ascending order.

Step 3: Every consecutive segment of length $\lfloor N/k \rfloor$ in the sorted data set (of length N) forms a distinct element of the partition.

Step 4: Convert the raw data into a symbol sequence with the partitions obtained in Step 3. If the data point lies within or on the lower bound of a partition, it is coded with the symbol associated with that partition.

Step 5: Compute the symbol probabilities p_i, $i = 1, 2, \ldots, k$.

Step 6: Compute $H(k) = -\sum_{i=1}^{i=k} p_i \log_2 p_i$ and $h(k) = H(k) - H(k-1)$.

Step 7: If $h(k) < \varepsilon_h$, then exit; else increment k by 1 and go to Step 3.

In general, a small ε_h leads to a large size of the symbol alphabet, resulting in increased computation. Also, a larger alphabet makes the partitioning finer. This might increase the probability of false symbols being induced by noise. On the other hand, a large ε_h leads to a small alphabet size that may prove inadequate for capturing the pertinent

information. Hence, there is a trade-off between accuracy and computational speed when ε_h is chosen. The variance of the noise process associated with the signal may serve as a guideline for selection of ε_h.

4. Validation of wavelet-based partitioning

This section presents simulation cases to validate symbolization of measured time series data via partitioning of the wavelet coefficients. The underlying concepts are illustrated by three examples.

4.1. Example 1: choice of wavelet parameters

This example illustrates how the choice of wavelet basis and scale range affects the coefficients that, in turn, determine symbol generation for anomaly detection [1]. Let us consider the following sinusoidal signal:

$$y(t) = \cos(2\pi t) \quad \forall t \in [-5, +5]. \tag{7}$$

The frequency of $y(t)$ in Eq. (7) is 1.00 Hz. The Gaussian wavelet 9 ('gaus9') (see [20, p. 74]) matches the shape of $y(t)$ well, as seen in Fig. 3 that compares an appropriately scaled and translated version of the 'gaus9' wavelet with the signal $y(t)$.

To demonstrate the impact of the chosen wavelet parameters on the coefficients, the wavelet basis 'db1' is also considered for comparison purposes. The signal $y(t)$ is sampled at 100 Hz (i.e., the sampling interval $\Delta t = 0.01$ s). The wavelet coefficients of the signal $y(t)$ are obtained for various scales with both wavelets, 'gaus9' and 'db1'. The

norm of the coefficients corresponding to each scale and the pseudo-frequencies of the wavelet corresponding to the chosen scales are calculated. Fig. 4 shows the plot of the norm of coefficients and the pseudo-frequencies of the wavelet.

It is observed in Fig. 4 that, for both wavelets 'gaus9' and 'db1', the maximum of the norm is obtained at $f_p \approx 1.00$ Hz. In fact, it is exactly at 1.00 Hz for 'gaus9'. Furthermore, the value of the peak norm achieved with 'gaus9' is appreciably greater than that with wavelet 'db1'. In other words, the coefficients obtained with 'gaus9' are more significant than those obtained with 'db1'. Another observation is that the norm curve for 'gaus9' shows a greater rate of decay across pseudo-frequencies than that of 'db1'. More energy is concentrated in a narrow band frequencies around 1.00 Hz in the case of 'gaus9'. These observations imply that high energy compaction can be achieved with fewer coefficients if the wavelet and the scales are chosen as stated in Section 2. A favorable implication of fewer coefficients is fewer number of symbols for analysis and hence an improvement in computational efficiency.

4.2. Example 2: noise suppression

This example demonstrates how noise suppression is achieved with wavelets. Let the signal $y(t)$ in Eq. (7) be corrupted with additive zero-mean white Gaussian noise $w(t)$,

$$\tilde{y}(t) = y(t) + \sigma w(t). \tag{8}$$

A common measure of noise in a noise-corrupted signal is the signal-to-noise ratio (SNR) that is
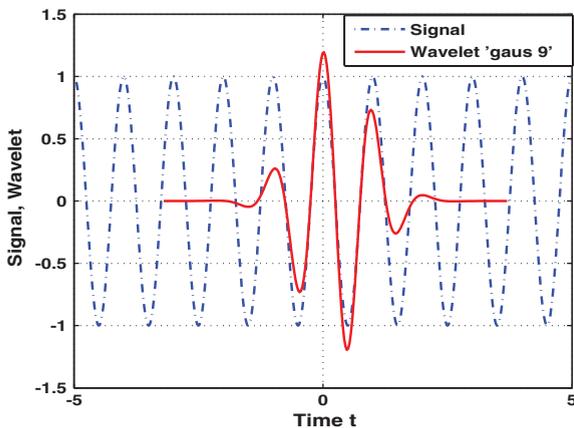
Fig. 3. Comparison of the wavelet basis 'gaus9' and the signal.
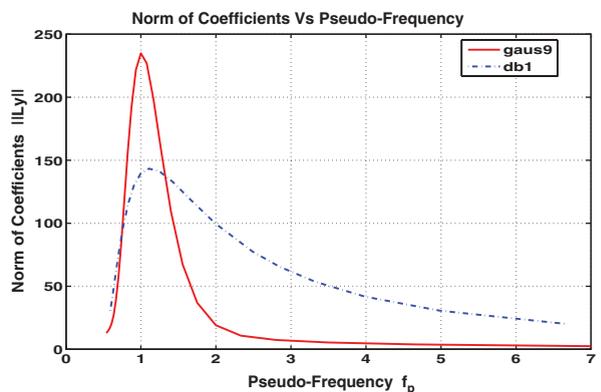
Fig. 4. Coefficient norm and pseudo-frequency for different wavelets.
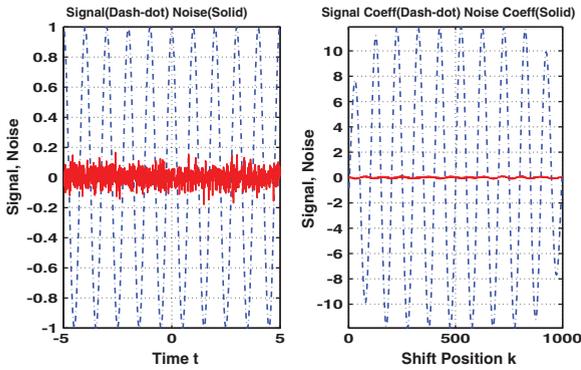
Fig. 5. Signal and noise profiles at $\sigma = 0.05$.
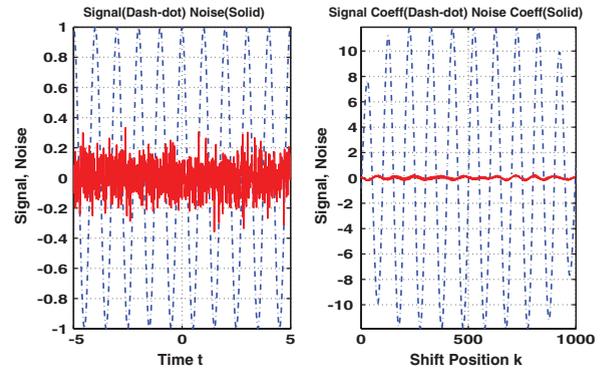
Fig. 6. Signal and noise profiles at $\sigma = 0.10$.

defined as:

$$\text{SNR} \triangleq \frac{\|y\|_{\mathbb{H}}^2}{\|\sigma\, w\|_{\mathbb{H}}^2}, \tag{9}$$

where y and w are functions of time. Similar to Eq. (9), the SNR in the wavelet domain is defined as

$$\widetilde{\text{SNR}} \triangleq \frac{\|\mathbb{L}y\|_{\mathbb{M}}^2}{\|\sigma\, \mathbb{L}w\|_{\mathbb{M}}^2}, \tag{10}$$

where $\mathbb{L}y$ and $\mathbb{L}w$, which belong to the Hilbert space $\mathbb{M}$ (see Section 2), represent the wavelet coefficients of the signal y and the noise w.

Numerical experiments have been performed with $\sigma \in \{0.05, 0.1\}$. The signal is sampled at 100 Hz (i.e., $\Delta t = 0.01$ s). The scales are determined following Eq. (4), such that the pseudo-frequency of the wavelet matches the frequency of the signal. Fig. 5 depicts the time domain plot (left plate) and coefficient plot (right plate) of the signal y and white Gaussian noise having standard deviation $\sigma = 0.05$. Similarly, Fig. 6 depicts the time domain plot (left plate) and coefficient plot (right plate) of the signal y and white Gaussian noise having standard deviation $\sigma = 0.10$. Table 1 lists the values of SNR and $\widetilde{\text{SNR}}$, averaged over 20 simulation runs.

Both Figs. 5 and 6 show that $\widetilde{\text{SNR}}$ is significantly larger than SNR. That is, the wavelet-transformed signal is significantly de-noised relative to the time domain signal. This is expected because the noise is incoherent with the wavelet while the signal enjoys a great degree of coherence with the wavelet. Thus, symbols generated from wavelet coefficients would reflect the characteristics of the signal with more fidelity than those obtained with time domain signals.

Table 1
SNR values

	$\sigma = 0.05$	$\sigma = 0.1$
SNR	191.55	50.89
$\widetilde{\text{SNR}}$	25195	4281.5

4.3. Example 3: robustness of symbol probability vector

The symbolization scheme is developed to enhance real-time detection of slowly varying anomalies in dynamical systems [1]. Of critical importance is the symbol probability vector **p** whose elements denote the probability of occurrence of individual symbols in the symbol sequence. The vector **p** must be robust relative to measurement noise and spurious disturbances and, at the same time, be sensitive enough to detect small slowly varying anomalies from the observed data set. A distortion measure for the symbol probability vector is introduced below

$$\delta_t \triangleq \|p_t - \tilde{p}_t\|_1, \tag{11}$$

where the subscript t denotes that the probability vectors correspond to symbols generated from time domain signals; and $\|\bullet\|_1$ is the sum of the absolute values of the elements of the vector $\bullet$. The vector p_t, with $\|p_t\|_1 = 1$, corresponds to the uncorrupted signal and $\tilde{p}_t$ corresponds to the corrupted signal. Similar to Eq. (11), distortion measure in the wavelet scale domain is defined as

$$\delta_s \triangleq \|p_s - \tilde{p}_s\|_1, \tag{12}$$

where the subscript s denotes that the probability vectors correspond to symbols generated from

wavelet scale domain signals (i.e., scale series data). Therefore, lower is the distortion ratio, closer is the probabilistic representation of the corrupted signal to that of the uncorrupted signal, i.e., greater is the robustness to noise and spurious disturbances.

The partitions are obtained, in case of time domain, by employing the maximum entropy criterion on the time series data of the signal. In the wavelet domain, the partitions are obtained with the corresponding scale series data, as defined in Section 2. In both time domain and wavelet domain, the probability vectors p and $\tilde{p}$ are computed with the same partitions for the uncorrupted and corrupted signals.

The symbol alphabet size $|\Sigma|$ is chosen to be 4 in this example. The partitions are obtained as mentioned before for the signal y and its transform, i.e., the coefficient vector $\mathbb{L}y$. Table 2 lists the values of distortion ratios δ_t and δ_s, averaged over 20 simulation runs.

It is seen that distortion measures are far smaller in the wavelet scale domain than those in the time domain. This observation implies that the symbol probabilities are significantly more robust to measurement noise and spurious disturbances in the wavelet domain than in the time domain. Hence, it may be inferred that symbols generated from the wavelet coefficients would be better for anomaly detection as the effects of noise to induce errors in the symbol probabilities are significantly mitigated.

5. Experimental results on anomaly detection

This section evaluates the performance of wavelet-based partitioning in *STSA* for anomaly/damage detection. This is demonstrated via experimentation on the following laboratory apparatuses:

- Anomaly detection on a nonlinear electronic system apparatus [1].
- Structural damage detection on a mechanical vibration system apparatus [3].
- Damage detection in polycrystalline alloys on a fatigue testing apparatus [5].

Table 2
Distortion ratios

	$\sigma = 0.05$	$\sigma = 0.1$
δ_t	0.040	0.054
δ_s	0.006	0.010

5.1. Anomaly detection in nonlinear systems

This example demonstrates efficacy of the *STSA* method for anomaly detection in nonlinear systems. Experiments have been conducted on a laboratory apparatus [1] that emulates the forced Duffing equation [21], modelled as

$$\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + y(t) + y^3(t) = A\cos(\Omega t), \qquad (13)$$

where the dissipation parameter β varies slowly with respect to the response of the dynamical system; $\beta = 0.1$ represents the nominal condition; and a change in the value of β is considered as an anomaly. With amplitude $A = 22.0$ and $\Omega = 5.0$, a sharp change in the behavior is noticed around $\beta = 0.29$, possibly due to bifurcation. The phase plots and time-response plots, depicting this drastic change behavior, are not presented here as they are provided in an earlier publication [1].

The objective of anomaly detection is to identify small changes in the parameter β as early as possible and well before it manifests a drastic change in the system dynamics. The details of the experimental apparatus are provided in [22]. Time series data of the signal $y(t)$ from the experimental apparatus is used for symbolic analysis.

The first step in the analysis is selection of the wavelet basis. The time series data of the signal and a scaled and translated version of the wavelet 'gaus1' are shown in the left-hand plate of Fig. 7. For the purpose of comparison, the right-hand plate of Fig. 7 shows the same time series data of the signal and a scaled and translated version of the wavelet 'db1' that was used in [1] for wavelet analysis. Since 'gaus1' matches the shape of the signal more closely than 'db1', the wavelet 'gaus1' is
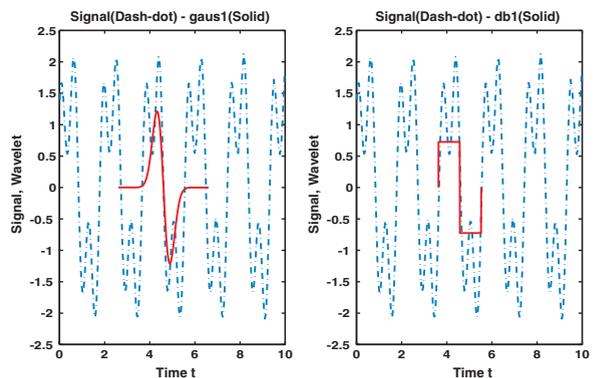
Fig. 7. Profiles of signal and wavelet.

better suited for STSA of this data sequence than 'db1'. Once the wavelet is chosen, the next step is identification of the frequencies of interest. The top and bottom plates of Fig. 8 depict the power spectra of the system for $\beta = 0.1$ and 0.29, respectively.

Fig. 8 shows that frequencies in the neighborhood of 0.54 Hz, though present at the nominal condition $\beta = 0.10$, are absent at $\beta = 0.29$. The wavelet coefficients at scales, corresponding to the pseudo-frequency of 0.54 Hz, would be smaller in magnitude in the anomalous condition(s) when compared with those in the nominal condition. Hence, by choosing scales that correspond to pseudo-frequencies around 0.54 Hz, early detection can be achieved more effectively.

The next aspect of anomaly detection via STSA is the choice of number of symbols, i.e., cardinality $|\Sigma|$ of the symbol alphabet Σ. The scale series data, at the nominal condition, is partitioned into a symbol sequence starting with $|\Sigma| = 2$ and the threshold parameter ε_h is chosen to be 0.2. Fig. 9 depicts the change in entropy h versus the number of symbols $|\Sigma|$. It is seen that h monotonically decreases with $|\Sigma|$ and becomes less than ε_h when $|\Sigma| = 8$. Accordingly, the number of symbols $|\Sigma|$ is chosen to be eight. A smaller value of ε_h results in increased number of symbols, which would increase computation with (possibly) no significant gain in accuracy of anomaly detection.

The partition is obtained using data obtained under the nominal ($\beta = 0.1$) condition. Once the partition is generated, it remains invariant. As the dynamical behavior of the system changes due to variations in β, the statistical characteristics of the
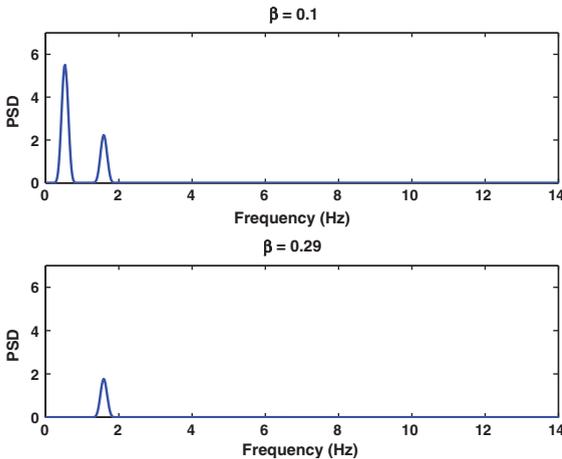
Fig. 9. Selection of number of symbols from change in entropy.

symbol sequences are also altered and so do the symbol probabilities. A measure could be induced on the symbol probability vectors obtained under different anomalous conditions, to quantify these changes. Such a measure is called the anomaly measure $\mathbf{M}$. The metric $\mathbf{M_k} = d(\mathbf{p_0}, \mathbf{p_k})$ is an anomaly measure, where $\mathbf{p_0}$ and $\mathbf{p_k}$ represent the symbol probability vectors under nominal and anomalous conditions, respectively. A candidate anomaly measure is the angle between the symbol probability vectors under nominal and anomalous conditions. This measure is defined as

$$\mathbf{M_k} = \arccos\left(\frac{\langle \mathbf{p_0}, \mathbf{p_k} \rangle}{\|\mathbf{p_0}\|_2 \|\mathbf{p_k}\|_2}\right), \qquad (14)$$

where $\langle x, y \rangle$ is the inner product of the vectors x and y; and $\|x\|_2$ is the Euclidean norm of x.

Fig. 10 compares the profiles of the anomaly measures $\mathbf{M_k}$ in Eq. (14) obtained with wavelet 'gaus1' under ME partitioning and uniform partitioning. With β increasing from 0.1, there is a gradual increase in the anomaly measure much before the abrupt change in the vicinity of $\beta = 0.29$ takes place. This indicates growth of the anomaly even before any notable change in the dynamical behavior takes place. At this point, the anomaly measure starts increasing relatively more rapidly suggesting the onset of a forthcoming catastrophic failure. Under ME partitioning, the larger values of the anomaly measure at smaller values of β and gradual increase in both slope and curvature of the anomaly measure curve would facilitate anomaly detection significantly before it is possible to do so under uniform partitioning. Therefore, with regard to early detection of anomalies, ME partitioning appears to be more effective than uniform partitioning.

Fig. 8. Power spectrum plots at nominal and anomalous conditions.
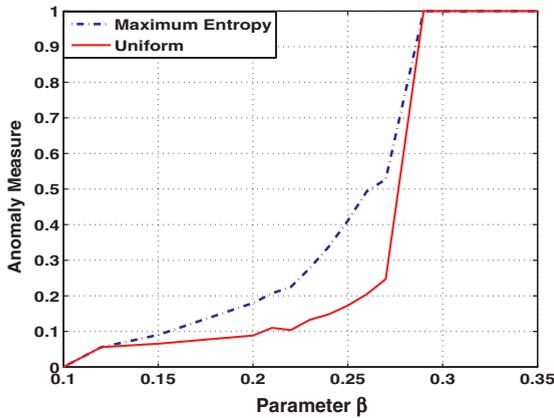
Fig. 10. Maximum entropy partitioning and uniform partitioning.
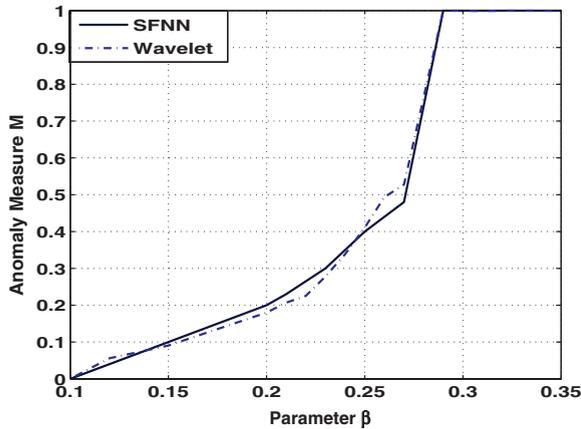
Fig. 11. Anomaly detection on the electronic system apparatus.

5.2. Structural damage detection in mechanical systems

This example demonstrates efficacy of the *STSA* method for early detection of damage in mechanical structures. Laboratory experiments have been conducted on a multi-degree-of-freedom mechanical vibration apparatus [3]. The mechanical system in the apparatus is persistently excited at a frequency of 10.4 Hz, which is a close approximation of one of resonance frequencies of the mechanical structure. During the experiments, time series data have been collected from the measurements of displacement sensors, and each set contains 30 s of data under persistent vibratory motion of the mass-beam system. The resulting cyclic stresses induce (irreversible) fatigue crack damage in the critical structures, which cause gradual reduction in stiffness. Consequently, the statistics of time series data undergo changes. The objective here is to detect these changes as early as possible in real time.

The first data set, which is dominated by a sinusoid of frequency ~10.4 Hz and represents the nominal behavior of the mechanical vibration system, is considered to be the reference point. The wavelet 'gaus9' is chosen for analysis because this wavelet closely matches the shape of the signal. A small set of scales around the frequency of 10.4 Hz are chosen as per procedure outlined in Section 2. The alphabet size is chosen, based on the entropy rate, as eight, i.e., $|\Sigma| = 8$.

Fig. 12 depicts two plots of anomaly measure, which are obtained from the same data set with

Fig. 11 depicts the anomaly measure plots obtained with wavelet partitioning ('gaus1') and phase-space partitioning (*SFNN*). These two profiles of anomaly measure are generated from the same time series data, where the number of symbols $|\Sigma| = 8$ in both cases. It is observed that *ME* partitioning with 'gaus1' wavelet is comparable to *SFNN* partitioning. However, in this problem, the execution time for *SFNN* to generate the partition is found to be ≈ 4 h, while that for *ME* partitioning is ≈ 100 ms on the same computer. Therefore, it may be inferred from this experiment that *ME* partitioning is computationally several orders of magnitude less intensive than *SFNN* partitioning while they yield similar performance from the perspectives of anomaly detection.

Fig. 12. Anomaly detection on the mechanical vibration apparatus.

different partitioning methods.

- *SFNN* partitioning [14].
- *ME* partitioning with wavelet 'gaus9'.

It is observed in Fig. 12 that *ME* partitioning with 'gaus9' wavelet is comparable to *SFNN* partitioning. However, in this problem, the execution time for *ME* partitioning is about five orders of magnitude less than that for *SFNN* partitioning. Hence, it may be inferred from this experiment that *ME* partitioning, is better suited for real-time structural damage detection in mechanical systems.

5.3. Fatigue damage detection in polycrystalline alloys

This example presents a comparison between wavelet space and phase-space partitioning in the context of fatigue crack damage detection in polycrystalline alloys. The details of the experimental apparatus, equipped with an ultrasonic flaw detector and an optical microscope, on which experiments were conducted are reported in [5] and references therein. Low-cycle fatigue tests have been conducted with specimens, made of aluminum alloy 7075-T6. A sinusoidal load with maximum and minimum loads of 87 and 4.85 MPa, respectively, is applied. A significant amount of internal damage occurs before the crack appears on the surface of the specimen, where it is observed by the optical microscope. This internal damage caused by multiple small cracks and dislocations affect the ultrasonic waves as they pass through the regions where these faults have developed. This phenomenon causes signal distortion and attenuation at the receiver end of the ultrasonic flaw detector.

The crack propagation stage starts when the internal damage eventually develops into a single large crack. Subsequently, the crack growth rate increases rapidly and when the crack is sufficiently large, the transmitted ultrasonic signal is almost completely attenuated. The process of obtaining the ultrasonic time series data is described in [5]. The results of *STSA*-based anomaly/damage detection, obtained with *ME* and *SFNN* partitioning methods, are presented below.

Wavelet 'gaus2' is used for obtaining the coefficients that are stacked to form the *scale series* data set. The alphabet size $|\Sigma|$ is chosen to be eight. The *scale series* data set is partitioned by imposing the ME criterion to generate the symbols. Fig. 13
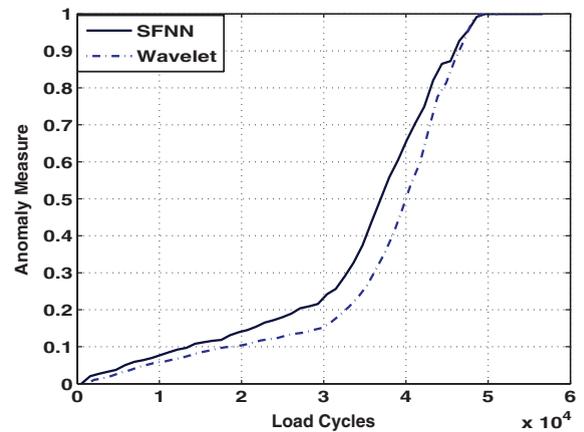
Fig. 13. Anomaly measure plots for fatigue crack detection.

depicts two anomaly measure plots that are obtained from the same data set by using *ME* and *SFNN* partitioning.

While the *SFNN* partitioning yields slightly higher values of the anomaly measure and comparable slope, the profile of wavelet partitioning is more smooth and yields a larger change in the curvature around 32 kilocycles, which is an early warning for the forthcoming failure. Simultaneous consideration of the anomaly measure, slope, and curvature provides a robust method of failure prediction and reduces the probability of false alarms.

Similar to the previous experiments, it is observed that the execution time for *ME* partitioning is approximately five orders of magnitude less than that for *SFNN* partitioning. Hence, it may be inferred that *ME* partitioning, is better suited for real-time fatigue damage detection in polycrystalline alloys.

6. Summary and conclusions

This paper presents a novel method of partitioning the data space for symbolic time series analysis (*STSA*). In this approach, symbols are generated from the wavelet coefficients of the time series data, instead of obtaining them directly from the time series data. Various aspects of this method, such as selection of wavelet basis and scale range, are systematically investigated for: (i) suppression of noise and spurious disturbances and (ii) enhancement of sensitivity to changes in signal dynamics. The advantages of using wavelet-based partitioning over phase-space partitioning are demonstrated by

numerical simulation and laboratory experimentation. It has been shown that measurement noise suppression results in robust symbol generation. The symbol sequences, generated from the wavelet coefficients of a noisy signal, are able to capture the changes in signal information better than those obtained directly from the time series data of the signal. It is also shown that the choice of an appropriate wavelet and scales significantly improves computational efficiency and thereby enhances implementation of the anomaly detection technique for real-time applications. An entropy-based algorithm is introduced for selection of the symbol alphabet size.

A partitioning method, based on maximum entropy, is introduced and is compared with the uniform partitioning method. The performance of wavelet-based partitioning has been tested via experimentation for: (i) anomaly detection on a nonlinear electronic system apparatus [1]; (ii) structural damage detection on a mechanical vibration system apparatus [3]; and (iii) damage detection in polycrystalline alloys on a Fatigue Testing apparatus [5].

Wavelet-based maximum entropy (*ME*) partitioning has been compared with symbolic false nearest neighbor (*SFNN*) partitioning [14] with regard to anomaly detection. Based on the time series data from three experimental apparatuses, it is observed that the aforementioned partitioning methods yield comparable results while the computation time for wavelet-based *ME* partitioning is observed to be several orders of magnitude smaller than that for *SFNN* partitioning.

A major conclusion based on this investigation is that wavelet-based *ME* partitioning, combined with an appropriate choice of wavelet and scales, significantly enhances computational efficiency and anomaly detection capabilities beyond what has been reported in literature. The field of STSA is relatively new and its application to anomaly detection is very recent. Therefore, the proposed method of symbol generation for anomaly detection requires continued theoretical and experimental research. In this context, future research is recommended in the following areas:

- Exploration of lifting techniques [23] for wavelet customization;
- Extension of *ME* partitioning to multi-dimensional time series;
- Noise reduction in time series for robust anomaly/damage detection.

Acknowledgements

References

[1] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.

[2] S. Chin, A. Ray, V. Rajagopalan, Symbolic time series analysis for anomaly detection: a comparative evaluation, Signal Processing 85 (9) (2005) 1859–1868.

[3] A. Khatkhate, A. Ray, S. Chin, V. Rajagopalan, E. Keller, Detection of fatigue crack anomaly: a symbolic dynamic approach, Proceedings of American Control Conference, Boston, MA, June–July 2004, pp. 3741–3746.

[4] D. Tolani, M. Yasar, A. Ray, V. Yang, Anomaly detection in aircraft gas turbine engines, AIAA J. Aerospace Comput., Inform. Comm. 3 (2006) 44–51.

[5] S. Gupta, A. Ray, E. Keller, Symbolic time series analysis of ultrasonic data for early detection of fatigue damage, in press, Elsevier. Corrected proof available online.

[6] S. Bhatnagar, V. Rajagopalan, A. Ray, Incipient fault detection in mechanical power transmission systems, Proceedings of American Control Conference, Portland, OR, 2005, pp. 472–477.

[7] R. Samsi, V. Rajagopalan, J. Mayer, A. Ray, Early detection of voltage imbalances in induction machines, Proceedings of American Control Conference, Portland, OR, June 2005, pp. 478–483.

[8] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. 74 (2) (2003) 915–930.

[9] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, second ed., Cambridge University Press, Cambridge, UK, 2004.

[10] C.J. Veenman, M.J.T. Reinders, E.M. Bolt, E. Baker, A maximum variance cluster algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9) (2002) 1273–1280.

[11] T. Chau, A.K.C. Wong, Pattern discovery by residual analysis and recursive partitioning, IEEE Trans. Knowledge Data Eng. 11 (6) (1999) 833–852.

[12] Y. Kakizawa, R.H. Shumway, N. Taniguchi, Discrimination and clustering for multivariate time series, J. Amer. Statist. Assoc. 93 (441) (1999) 328–340.

[13] T.W. Liao, Clustering of time series data—a survey, Pattern Recognition 38 (2005) 1857–1874.

[14] M.B. Kennel, M. Buhl, Estimating good discrete partitions form observed data: symbolic false nearest neighbors, Phys. Rev. E 91 (8) (2003) 084102.

[15] V. Rajagopalan, A. Ray, Wavelet-based space partitioning for symbolic time series analysis, Proceedings of IEEE Conference on Decision and Control (CDC) and European Control Conference (ECC), Seville, Spain, 2005, pp. 5245–5250.

[16] S. Mallat, A Wavelet Tour of Signal Processing, second ed., Academic Press, San Diego, CA, 1998.

[17] A. Teolis, Computational Signal Processing with Wavelets, Birkhäuser, Boston, MA, 1998.

[18] Wavelet Toolbox, MATLAB. Mathworks Inc, Natick, MA.

[19] P. Abry, Ondelettes et turbulence, multirésolutions, algorithmes de décomposition, invariance d'échelles, Diderot Editeur, Paris, 1997.

[20] G. Kaiser, A Friendly Guide to Wavelets, Birkhäuser, Boston, MA, 2004.

[21] J.M.T. Thompson, H.B. Stewart, Nonlinear Dynamics and Chaos, Wiley, Chichester, UK, 1986.

[22] V. Rajagopalan, R. Samsi, A. Ray, J. Mayer, C. Lagoa, A symbolic dynamics approach for early detection of slowly evolving faults in nonlinear systems, Proceedings of IASTED-CSS, Clearwater, FL, Paper No. 449–082, 2004.

[23] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM J. Math. Anal. 29 (2) (1998) 511–546.