Heterogeneous Recurrence Analysis

Yun Chen and Hui Yang
Department of Industrial and Management Systems Engineering
University of South Florida, Tampa

Abstract: Process monitoring of dynamic transitions in complex systems is more concerned with aperiodic recurrences and heterogeneous recurrence variations. However, traditional recurrence methods treat recurrence states homogeneously. This paper presents a new approach of heterogeneous recurrence analysis for process monitoring and control. Experimental results show that the proposed methodology effectively monitors the changes in the dynamics of complex systems.

Key words: Heterogeneous recurrence, dynamical transition, fractal representation, multivariate analysis.

I. Introduction

Recurrence (i.e., approximate repetitions of a certain event) is one of the most common phenomena in natural and engineering systems. For examples, stamping machines are cyclically forming sheet metals during production [1]; human heart is near-periodically beating to maintain vital living organs. Technological advances bring the proliferation of sensor signals gathered from these complex processes. This offers an unprecedented opportunity to exploit recurrence dynamics for process monitoring and anomaly detection. However, most of existing approaches adopt linear methodologies for the analysis of recurrence behaviors. They have encountered certain difficulties to capture nonlinearity, nonstationarity and high-order variations in complex systems.

Process monitoring of disease conditions or manufacturing quality is more concerned with aperiodic recurrences and recurrence variations in nonlinear and nonstationary systems. The approach of nonlinear recurrence analysis characterizes the proximity of two state vectors \( \mathbf{s}(i) \) and \( \mathbf{s}(j) \) in the high-dimensional state space, i.e., \( R(i,j) = \Theta(\varepsilon - \| \mathbf{s}(i) - \mathbf{s}(j) \|) \), where \( \Theta \) is the Heaviside function, \( \varepsilon \) is the threshold and \( \| \cdot \| \) is a distance measure. The recurrence plot \( R(i,j) \), introduced by Eckmann et al. in the late 1980’s [2], captures topological relationships in the state space as a 2D image (see Fig. 1). The structure of recurrence plot has distinct topology and texture patterns. The ridges locate the nonstationarity and/or the switching between local behaviors. The parallel diagonal lines indicate the near-periodicity of system behaviors.

Recurrence quantification analysis (RQA) measures small-structures, chaos-order transitions, as well as chaos-chaos transitions, in the recurrence plot. Examples of recurrence quantifiers include recurrence rate (RR), determinism (DET), entropy (ENT), and laminarity (LAM) [3, 4].

However, very little research has been done to characterize and quantify heterogeneous recurrences. Although the theory of recurrence analysis has been significantly advanced in the past few decades, most of existing works are based on homogeneous recurrence (i.e., treating all recurring states in the same way
as black dots in the recurrence plot). Fig. 1 illustrates the basic concept of heterogeneity in the recurrence plot. For a given time series with embedding dimension \( m = 3 \) and time delay \( \tau = 2 \), there are two pairs of recurrence states, i.e., \( \vec{s}(15) = (x_{15}, x_{17}, x_{19}) \) and \( \vec{s}(1) = (x_1, x_3, x_5) \). \( \vec{s}(32) = (x_{32}, x_{34}, x_{36}) \) and \( \vec{s}(30) = (x_{30}, x_{32}, x_{34}) \). In the traditional recurrence analysis, they are treated in the same way as black dots in the recurrence plot. Nonetheless, their recurrence behaviors are heterogeneous. Notably, heterogeneous recurrence patterns are not specifically considered in the state of the art. This present paper is the first of its kind that aims to investigate heterogeneous recurrence behaviors for complex systems informatics and monitoring.

In addition, significant level of detected variations is not well established for RQA measures. Recently, Marwan et al. introduced a novel bootstrap-based approach to establish confidence intervals of the variations of RQA measures for the purpose of process monitoring (i.e., detecting the changes and transitions in the dynamics of a complex system) [5]. It may be noted that this approach is analogous to the concept of statistical process control (SPC) in quality engineering (e.g., hypothesis testing and control limits) [6], but SPC methods focus on the variation of quality characteristics in production systems and overlook the changes in nonlinear dynamical systems. Few, if any, previous works integrated nonlinear recurrence analysis with SPC methods for the change detection in the dynamics of a complex system.

This paper presents a new approach of heterogeneous recurrence analysis for process monitoring and anomaly detection in complex systems. We designed a new fractal representation of state space that efficiently delineates heterogeneous recurrence states for both single state and multi-state sequence. Further, we extracted a new set of heterogeneous recurrence quantifiers from fractal representation in the transformed space. Experimental results on stochastic Markov processes and distribution-based processes show that the proposed approach not only captures heterogeneous recurrence patterns in the fractal representation, but also effectively monitors the changes in the dynamics of a complex system.

This paper is organized as follows: Section II presents the state of the art in recurrence analysis and process monitoring. Section III introduces the research methodology. Section IV presents the materials and experimental design. Section V presents the experimental results, and Section VI includes the discussion and conclusions arising out of this investigation.

II. Research Background

In the past few decades, significant advancements have been made on the theory of recurrence analysis. Nonlinear recurrence methods have found successful applications in various disciplines, e.g., biology [7], manufacturing [8], and neuroscience [9]. Recurrence plots and related quantification methods provide an effective means to characterize and measure nonlinear dynamical properties of complex systems from time series, which are usually indiscernible with traditional statistical analysis and linear models. Notably, recurrence characteristics will be perturbed by the changes in the dynamics of a complex system. Hence, many previous works have recognized the value of time-varying recurrence patterns in the monitoring of complex systems. For example, Marwan et al. extracted recurrence quantifiers from the moving windows of time series, and then employed the bootstrap resampling to establish confidence intervals of RQA measures for the detection of dynamic transitions [5]. This approach is the first of its kind that brings the confidence interval and hypothesis testing from the domain of statistical process control to gaining confidence in recurrence-based monitoring of nonlinear system dynamics.

However, most of existing recurrence methods and tools are based on homogeneous recurrences. In other words, all recurring states are treated in the same way as black dots and non-recurrence states are white in the recurrence plot. Few, if any, previous works considered the differentiation and quantification of various types of recurrence behaviors (i.e., heterogeneous recurrences) for process monitoring. For example, a Markov chain is a stochastic process with a finite set of states. The recurrence of state \( i \) is different from the recurrence of state \( j \) in the stochastic process. Heterogeneous recurrences of states are
pertinent to the transition probabilities represented in the transition matrix. The recurrences of a single state, as well as multi-state sequences characterize dynamical properties of the stochastic process. It may be noted that two states with a large transition probability are more likely to recur in the future than two states with a small transition probability. Nonetheless, if the transition probability from one state to the other is zero, the pattern of the two-state sequence will not appear in the stochastic process. Therefore, this present investigation aims to develop an effective representation of heterogeneous recurrences and further quantify heterogeneous recurrence patterns for process monitoring.

III. Heterogeneous Recurrence Analysis

This present paper delineates heterogeneous recurrence characteristics in complex systems through a new fractal representation of nonlinear time series, as opposed to the conventional state space reconstruction. As shown in Fig. 2, this present investigation is embodied by three core components focusing on the development of heterogeneous recurrence methodology for complex systems monitoring and control. (1) The first component is aimed at the design of a new fractal representation (rather than traditional state space reconstruction) of nonlinear time series that effectively present heterogeneous recurrence information. (2) The second component will develop new statistical quantifiers (rather than traditional fractal dimensions) that measure heterogeneous recurrence patterns in the fractal domain. The extracted features should be sensitive to the changes in the dynamics of complex systems. (3) The third component aims to design and develop multivariate control charts with confidence intervals (rather than univariate change-point detection without significance measures) for process monitoring. All three components are eventually integrated together in the framework of heterogeneous recurrence analysis to make complex systems monitoring more effective and efficient.

![Figure 2: The methodology of heterogeneous recurrence analysis.](image)

A. Fractal Representation

Fractals are typically self-similar patterns regardless of the magnification. Fractal geometry often refers to self-similar patterns across geometric scales, but the attractor reconstructed from time series in recurrence analysis is seldom fractal. This poses a significant challenge to estimate fractal dimensions of the geometric attractor reconstructed from time series. Therefore, most of existing works resort to self-similar structures of time series across temporal scales, instead of geometric scales. Notably, Webber [10] implemented recurrence analysis on systems of common fractals and demonstrated the recurrence plots of real and imaginary parts in the dynamics of Mandelbrot set. However, limited work has been done to tightly melt recurrence with fractal concepts for nonlinear time series analysis. In this present investigation, we propose the use of iterative function system (IFS) [11] to represent heterogeneous recurrences in time series. For a discrete process with a finite set of states, the IFS sequentially maps each state \( \vec{s}(n) \) to an address (i.e., a point) \( [c_x(n), c_y(n)] \) in the 2D coordinate system as:

\[
\vec{s}(n) \rightarrow k \in \mathcal{K} = \{1, 2, \cdots, K\}
\]
\[
\begin{bmatrix}
    c_x(n) \\
    c_y(n)
\end{bmatrix} = \varphi(k, \begin{bmatrix}
    c_x(n-1) \\
    c_y(n-1)
\end{bmatrix}) = \begin{bmatrix}
    \alpha & 0 \\
    0 & \alpha
\end{bmatrix} \begin{bmatrix}
    c_x(n-1) \\
    c_y(n-1)
\end{bmatrix} + \begin{bmatrix}
    \cos(k \times \frac{2\pi}{K}) \\
    \sin(k \times \frac{2\pi}{K})
\end{bmatrix}
\]

where \( \begin{bmatrix}
    c_x(0) \\
    c_y(0)
\end{bmatrix} = \begin{bmatrix}
    0 \\
    0
\end{bmatrix} \). Each state \( s(n) \) is assigned to a categorical variable \( k \) that belongs to a finite set \( \mathcal{K} \) of positive integers. For a discrete process with a finite set of states, the state space \( S \) can be directly mapped to the set of categorical variables \( \mathcal{K} = \{1,2,\cdots,K\} \).

The IFS of a circle transformation (Eq. 1) is an iterative contractive mapping \( \varphi(k, c) \) that represents the states as vectors in \( \mathbb{R}^2 \). The iterative contractive mapping is composed of two parts. The first part accounts for the previous state's address with a weight \( \alpha \). The second part uniformly distributes \( K \) addresses on the unit circle. Therefore, the IFS transforms each state \( s(n) \) into an address that depends on all of its previous states, as well as the corresponding categorical variable. The IFS transformation ensures a unique address for every state in the 2D fractal graph, because two states have the same address if and only if they share the same category variable and all of their previous states are the same.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{Fractal representation of a Markov process with 8 categories (Fig. 2) via the IFS of a circle transformation: (a) addresses of individual states; (b) addresses of two-state sequence; (c) addresses of three-state sequence.}
\end{figure}

Fig. 3 shows the fractal representation of a Markov process with 8 categories via the IFS of a circle transformation. Each circle in Fig. 3a shows the recurrence of one out of 8 possible categories. These categories are centered on 8 addresses that are uniformly distributed on the unit circle. Zooming into circle 8 (i.e., marked by the blue rectangle in Fig. 3a) leads to Fig. 3b. Every circle in Fig. 3b represents one of eight two-state sequence: 18,28, ...,88 in the time series (i.e., heterogeneous recurrences in the level of two-state sequence). Notably, there are three two-state sequences (i.e., 48, 78 and 88) missing in Fig. 3b, which indicates zero transition probability between these two states. Further, Fig. 3c is the zoom into category 28. Each circle represents the recurrence of one of three-state sequences. Notably, the density and distribution of points in each circle characterize heterogeneous recurrence variations.

To this end, the IFS of a circle transformation effectively reveals heterogeneous recurrence characteristics in time series. Each type of recurrences is tagged by an address in the 2D fractal graph. By zooming into local regions, we can highlight various types of recurrences that are elegantly grouped and centered at a unique address in the graph. Further, we will focus on the quantification of heterogeneous recurrence patterns for describing dynamical properties of complex systems as detailed in Section III.B.

B. Heterogeneous Recurrence Quantification

Traditional recurrence plots treat all recurrence states in the same way as black dots. The proposed fractal representation tags heterogeneous recurrences at different addresses in the 2D graph. This facilitates the delineation of heterogeneous recurrences in recurrence plots. However, such salient patterns of
heterogeneous recurrences tend to be overlooked by most of existing dynamic quantifiers. Conventional RQA statistics focus on the structures of diagonal and vertical lines in recurrence plots. However, diagonal and vertical lines have faded in heterogeneous recurrence plots, due to the discrete process and the separation of recurrences. This makes the use of conventional RQA statistics impractical for quantifying the recurrence heterogeneity. We have also attempted to quantify the fractal dimension from the 2D graph. Notably, fractal dimension focuses on the quantification of self-similar structures across geometric or temporal scales, but overlooks heterogeneous recurrence patterns in this investigation.

Hence, we propose the development of new quantifiers that describe heterogeneous recurrence patterns in the fractal representation. Because the contractive mapping \( \varphi(k, c): \mathbf{s}(n) \to k \in \mathcal{K} = \{1, 2, ..., K\} \) effectively clusters all the states with the same categorical variable \( k \) at local regions in the 2D graph, we denote these clustered states as heterogeneous recurrence sets, i.e.,

\[
W_{k_1, k_2, ..., k_L} = \{ \varphi(k_1|k_2, ..., k_L): \mathbf{s}(n) \to k_1, \mathbf{s}(n-1) \to k_2, ..., \mathbf{s}(n-L+1) \to k_L \} \quad \text{and} \quad k_1, k_2, ..., k_L \in \mathcal{K}.
\]

Here, \( k_1, k_2, ..., k_L \) denotes the \( L \)-state sequence. On the basis of the set \( W_{k_1, k_2, ..., k_L} \), we develop 3 new quantifiers of heterogeneous recurrences, namely heterogeneous recurrence rate (HRR), heterogeneous mean (HMean) and heterogeneous entropy (HENT).

**HRR** measures the percentage of recurrences in the heterogeneous recurrence plot from the set \( W_{k_1, k_2, ..., k_L} \), which is analogous to recurrence rate in conventional RQA measures.

\[
\text{HRR} = \left( \frac{\overline{W}_{k_1, k_2, ..., k_L}}{N} \right)^2
\]

where \( \overline{W}_{k_1, k_2, ..., k_L} \) denotes the cardinality of set \( W_{k_1, k_2, ..., k_L} \). The square of \( \overline{W}_{k_1, k_2, ..., k_L} \) corresponds to the number of colored dots in the heterogeneous recurrence plot. To be consistent with the definition of recurrence rate, HRR is defined as the square of the ratio between the cardinality \( \overline{W}_{k_1, k_2, ..., k_L} \) and total number of states in the stochastic process.

In addition, it should be noted that the set of \( W_{k_1, k_2, ..., k_L} \) denotes the recurrences of the same \( L \)-state sequence \([\mathbf{s}(n) \to k_1, \mathbf{s}(n-1) \to k_2, ..., \mathbf{s}(n-L+1) \to k_L]\) that are clustered at local regions in the 2D fractal graph (see Fig. 3). However, the addresses of \( L \)-state sequences in the set \( W_{k_1, k_2, ..., k_L} \) are not exactly the same and are distributed in the local region. Hence, we computed the distance matrix in the heterogeneous recurrence set \( W_{k_1, k_2, ..., k_L} \) as:

\[
\mathcal{D}_{k_1, k_2, ..., k_L}(i, j) = \| \varphi^i - \varphi^j \|, \quad \varphi^i, \varphi^j \in W_{k_1, k_2, ..., k_L}; \quad i, j = 1, 2, ..., \overline{W}; \quad i < j
\]

where \( \varphi^i \) and \( \varphi^j \) are the \( i \)th and \( j \)th elements in the set \( W_{k_1, k_2, ..., k_L} \).** Heterogeneous mean (HMean)** is defined as the average distance of \( \mathcal{D}_{k_1, k_2, ..., k_L} \):

\[
\text{HMean} = \frac{2}{\overline{W}(\overline{W} - 1)} \sum_{i=1}^{\overline{W}} \sum_{j=i+1}^{\overline{W}} \mathcal{D}_{k_1, k_2, ..., k_L}(i, j)
\]

HMean provides general information about the average distance among elements in the set \( W_{k_1, k_2, ..., k_L} \). As shown in Fig. 3, some sets (e.g., \( W_{40}, W_{79}, W_{89} \)) are empty and others (e.g., \( W_{318}, W_{418}, W_{518} \)) have an irregular distribution of elements in the space. Therefore, HMean describes the dynamic property of heterogeneous sets of recurrences.

**Heterogeneous entropy (HENT)** is based on the measure of Shannon entropy of the probability distribution of \( \mathcal{D}_{k_1, k_2, ..., k_L}(i, j) \). It should be noted that we divide the distance matrix \( \mathcal{D}_{k_1, k_2, ..., k_L} \) into \( B \) equally bins from 0 to \( \text{max}(\mathcal{D}) \) and compute the probability as

\[
p(b) = \frac{1}{\overline{W}(\overline{W} - 1)} \# \left\{ b - 1 \leq \text{max}(\mathcal{D}) < b \right\}
\]

where \( b = 1, 2, ..., B \). Hence, the HENT is defined as

\[
\text{HENT} = - \sum_{p=1}^{B} p(b) \ln p(b)
\]
HENT provides general information on the uncertainty in the recurrence of an L-state sequence. Note that the recurrences of an L-state sequence do not coincide in the same address but scatter in local regions in the fractal graph. Very little work has discerned the uncertainty for the same type of recurrence. This investigation developed three new measures (i.e., HRR, HMean and HENT) that quantify heterogeneous recurrence patterns hidden in the time series.

C. Multivariate Process Monitoring

Heterogeneous recurrence quantification leads to multiple features pertinent to the dynamics of a complex system. The hypothesis test for the significance of dynamic transitions is to determine whether there is a significant mean shift in the feature vector \( \mathbf{y} = [y_1, y_2, ..., y_p]^T \), where \( p \) is the dimensionality of feature vector. Suppose that \( \mathbf{y} \) is a multivariate random variable with population mean \( \mu \) and covariance matrix \( \Sigma \). Under the null hypothesis \( H_0 \), the dynamics of complex systems do not change over time. However, population mean \( \mu \) and covariance matrix \( \Sigma \) need to be estimated from the data. If we replace \( \mu \) with the sample mean \( \bar{y} \) and \( \Sigma \) with sample covariance matrix \( \mathbf{S} \), the test statistic becomes \( T^2 = (\mathbf{y} - \bar{y})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{y}) \), named Hotelling’s \( T^2 \) statistic. The upper control limit for the Hotelling’s \( T^2 \) statistic is: \( UCL = \frac{p(M+1)(M-1)}{M^2-Mp}F_{a,p,M-p} \), where \( M \) is the number of samples, \( F_{a,p,M-p} \) is the upper 100\( \alpha \)% critical point of \( F \) distribution with \( p \) and \( M - p \) degrees of freedom.

However, a significant challenge resides in the inversion of sample covariance matrix \( \mathbf{S} \). Because the transition probability between two states may be zero in the Markov Chain, this leads to an empty set in the fractal representation (e.g., 48, 78 and 88 in Fig. 3b). As a result, the covariance matrix is singular due to zeros in heterogeneous recurrence quantifiers, thereby making the computation of Hotelling’s \( T^2 \) statistic impractical. Therefore, we transformed the feature matrix \( \mathbf{Y}_{M \times p} \) into a set of principal components (PCs) that are linearly uncorrelated. First, the feature matrix \( \mathbf{Y}_{M \times p} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_M]^T \) is centered by subtracting off column means, i.e., \( \mathbf{Y} = [\mathbf{y}_1 - \bar{y}, \mathbf{y}_2 - \bar{y}, ..., \mathbf{y}_M - \bar{y}]^T \). Then, the singular value decomposition (SVD) of \( \mathbf{Y} \) will be

\[
\mathbf{Y} = \mathbf{U}\Psi \mathbf{V}^T
\]

where \( \mathbf{U} \) and \( \mathbf{V} \) are \( M \times M \) and \( p \times p \) orthogonal matrices, \( \Psi \) is a \( M \times p \) diagonal matrix, with diagonal entries \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \) (i.e., singular values of \( \mathbf{Y} \)). Hence, the covariance matrix \( \mathbf{Y}^T \mathbf{Y} \) is

\[
\mathbf{Y}^T \mathbf{Y} = \mathbf{U} \Psi \Psi^T \mathbf{V}^T = \mathbf{V} \Psi^2 \mathbf{V}^T
\]

which is the eigen decomposition of \( \mathbf{Y}^T \mathbf{Y} \). The eigenvectors \( \mathbf{v}_i \) are the principal component directions of \( \mathbf{Y} \). As a result, principal components are given as

\[
\mathbf{Z} = \mathbf{Y} \Psi = \mathbf{U} \Psi \mathbf{V}^T \mathbf{V} = \mathbf{U} \Psi
\]

The first principal component \( \mathbf{Z}(i,:) = \mathbf{Y} \mathbf{v}_1 \) has the largest sample variance among all the principal components. The sample covariance matrix \( \mathbf{S} \) is easily seen to be:

\[
\mathbf{S} = \frac{1}{M-1} \sum_{i=1}^{M} (\mathbf{y}_i - \bar{y}) (\mathbf{y}_i - \bar{y})^T = \frac{1}{M-1} \mathbf{V} \mathbf{Z}^T \mathbf{Z} \mathbf{V}^T = \mathbf{V} \Sigma \mathbf{V}^T
\]

Notably, principal components \( \mathbf{Z}(i,:) \) and \( \mathbf{Z}(j,:) \) are orthogonal with each other. Hence, the covariance matrix of principal components \( \mathbf{S}_z \) is a diagonal matrix with diagonal entries \( \lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_p^2 \). The Hotelling’s \( T^2 \) statistic becomes

\[
T^2(i) = (\mathbf{y}_i - \bar{y})^T \mathbf{S}^{-1} (\mathbf{y}_i - \bar{y}) = (\mathbf{y}_i - \bar{y})^T \mathbf{V} \mathbf{S}_z^{-1} \mathbf{V}^T = \sum_{k=1}^{p} \frac{Z(i,k)^2}{\lambda_k^2}
\]

where \( \mathbf{Z}(i,:) \) is the projection of the \( i \)th sample on principal component directions, that is, \( (\mathbf{y}_i - \bar{y})^T \mathbf{V} \). The eigenvectors \( \mathbf{v}_i \) and eigenvalues \( \lambda_i \) of \( \mathbf{S} \) can be obtained by considering only the first \( q \) eigen values and eigenvectors \( \mathbf{v}_i \). Hence, the Hotelling’s \( T^2 \) statistic in the reduced dimension \( q \) is:

\[
\tilde{T}^2(i) = \sum_{k=1}^{q} \frac{Z(i,k)^2}{\lambda_k^2}
\]

Because it is difficult to
graphically construct the control ellipse for multivariate recurrence features (i.e., >3), the Hotelling $T^2$ statistic can be plotted for each sample on a control chart with the upper control limit. This control chart characterizes the distribution of multivariate recurrence features by a single number (i.e., Hotelling $T^2$ statistic), and effectively detects the mean shift in any dimension of the feature vector (i.e., multivariate process monitoring). The design of experiments to evaluate the methodology of heterogeneous recurrence analysis is described in Section IV.

IV. Materials and Experimental Design

In this present investigation, we illustrated and evaluated the proposed methodology of heterogeneous recurrence analysis on three experimental scenarios. (1) in-control vs. out-of-control Markov processes: Suppose the Markov process which generated the fractal representation in Fig. 3 is in control. The out-of-control Markov process was generated from another transition matrix, which is significantly different from the in-control one. (2) in-control vs. slightly-changed Markov processes: Furthermore, we slightly perturb the transition matrix for the in-control Markov process and generate a new Markov process. Deviation from the in-control process is obtained by slightly changing one row in the transition matrix and relatively preserving the summation in the row to be 1. (3) distribution-based processes (uniform vs. normal): We have generated the in-control process randomly from a discrete uniform distribution. The out-of-control process is simulated when the underlying distribution is deviated to the normal distribution. Each process is represented by a time series that are random realizations of this process. In all three scenarios, each point in the control chart is the Hotelling $T^2$ statistic calculated from multivariate recurrence features. These features are extracted from a sample of time series with 10,000 data points. For each experimental scenario, the first 100 samples are random realizations from the in-control process and the last 50 samples are from the out-of-control process. The experimental results are detailed in Sec. V.

V. Results

This present investigation made an attempt to delineate heterogeneous recurrences hidden in the time series and thereby detect the changes in the dynamics of a complex system. As discussed in the experimental design, we studied the performance of the proposed methodology of heterogeneous recurrence analysis on three experimental scenarios. The results are as follows:

A. In-Control vs. Out-of-Control Markov Processes

As aforementioned, the out-of-control Markov process was generated from the transition matrix that is different from the in-control one. Multivariate features of heterogeneous recurrences were extracted from 100 samples of in-control process and 50 samples of out-of-control process. Further, Hotelling $T^2$ statistics are derived from heterogeneous recurrence quantifiers to detect the changes in the Markov process. In Phase I, we used the in-control Markov process to establish the control limits. In Phase II monitoring, if a new sample comes, a new point (i.e., Hotelling $T^2$ statistic) will be added in the control chart. The upper control limit of control charts for HRR, ENT, and HMean features in the level of individual states (see Fig. 4) is $UCL = 23.57$, where $M = 100, p = 8, \alpha = 0.01$. In addition, we developed a Hotelling $T^2$ control chart for the combination of HRR, ENT, and HMean features in the level of individual states. The upper control limit for overall features is $UCL = 64.57$, where $M = 100, p = 24, \alpha = 0.01$.

Fig. 4 shows Hotelling $T^2$ control charts with the feature set of HRR, HENT, HMean and overall quantifiers, respectively for in-control and out-of-control Markov processes. It can be seen from Fig. 4 that all three quantifiers of heterogeneous recurrences show significant differences between in-control and out-of-control Markov processes, thereby leading to effective monitoring schemes. In addition, the discriminatory power of HRR is better than HMean and HENT (see Fig. 4a, 4b and 4c). In other words, the percentage of heterogeneous recurrences is more effective in differentiating in-control from out-of-control Markov processes than statistical measures of the distance distribution in one specific category of
recurrences. Moreover, overall quantifiers are the combination of HRR, HENT, HMean features, which also significantly increase the discriminatory power of Hotelling $T^2$ control charts (see Fig. 4d).

**Figure 4:** Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean; (d) Overall quantifiers.

**B. In-Control vs. Slightly-Changed Markov Processes**

As aforementioned, the slightly-changed Markov process is obtained by slightly changing one row in the transition matrix of in-control Markov process and relatively preserving the summation in the row to be 1, while others remain the same. This slight change makes it difficult to visually differentiate fractal patterns between in-control and slightly-changed Markov processes. Nonetheless, heterogeneous recurrence analysis discloses hidden information on various types of recurrences from the level of individual states to multi-state sequences. Therefore, it facilitates the detection of slight changes in the dynamics of a complex system. Now, the question becomes “Can heterogeneous recurrence quantifiers effectively identify slight changes in the Markov Process?”

**Figure 5:** Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean; (d) Overall quantifiers. (Note: The green dash line represents the upper control limit.)

Further, we derived Hotelling $T^2$ statistics from heterogeneous recurrence quantifiers for in-control and slightly-changed Markov processes. Fig. 5 shows Hotelling $T^2$ control charts for the feature set of HRR, HENT, HMean and overall quantifiers. Notably, HRR is still capable of identifying slight changes in the Markov process. However, the discriminatory power of HRR is as not significant as in the out-of-control cases (see Fig. 4). Because of the iterative nature of fractal representation, the address of each state depends on not only its categorical variable but also all of its previous states. Therefore, a slight change in the transition probability affects the distance distribution in one specific category of recurrences. Consequently, Fig. 5 shows that HENT and HMean quantifiers show comparable performances as HRR between in-control and slightly-changed Markov processes. Moreover, overall quantifiers are the integration of HRR, HENT, HMean features, which also increase the discriminatory power of Hotelling $T^2$ control charts (see Fig. 5d).

**C. Distribution-Based Processes (Uniform vs. Normal):**

In this experimental scenario, we simulated the in-control process randomly from a discrete uniform distribution. Deviation from the in-control process is obtained by changing the underlying distribution to the normal distribution. The proposed heterogeneous recurrence method is appealing for the monitoring
of distribution-based processes. Fig. 6 presents Hotelling $T^2$ control charts with control limits for the feature set of HRR, HENT, HMean and overall quantifiers. As demonstrated, four control charts clearly distinguish between uniformly-distributed (samples 1 to 100) and normally-distributed processes (samples 101-150). Notably, HRR yields better discriminatory power than HENT and HMean due to significant differences in the distribution and density of point addresses between two processes. In addition, overall quantifiers combine HRR, HENT, HMean features and yield the best discriminatory power among these four control charts (see Fig. 6d).

![Figure 6: Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean; (d) Overall quantifiers. (Note: The green dash line represents the upper control limit.)](image)

### D. Multivariate Process Monitoring in the Reduced-Dimension Space

It may be noted that we only used the set of heterogeneous recurrence features (i.e., HRR, HENT, and HMean) in the first level of individual states (see Fig. 3a) in Sections A-C. If zooming into each state, there are 8 clusters of points characterizing heterogeneous recurrences of two-state sequences (see Fig. 3b). The next question is “Will heterogeneous recurrence analysis in multiple scales improve the monitoring performance?” If we consider the levels of both single states and two-state sequences, there will be 72 features (i.e., $8 \times 8 = 72$) for each of three heterogeneous recurrence quantifiers (i.e., HRR, HENT, or HMean). In total, there will be 216 features for each sample of 10,000 data points in the stochastic process. This gives rise to the problem of “curse of dimensionality”. Hence, we conducted further experiments to investigate the performances of multivariate process monitoring in the reduced-dimension space (Eq. 11). Here, we adopted the average run length (ARL) to evaluate the performance of monitoring schemes. If the process is out of control, the ARL is the average period at which a process-monitoring scheme first signals. In this present paper, the ARL is calculated as: $ARL = 1/(1 - \beta)$, where $\beta$ is the probability of samples falling within control limits after the process is out of control.

#### Table I: Comparison of process-monitoring performances in the reduced-dimension space

<table>
<thead>
<tr>
<th>Percentage of variance explained</th>
<th>In-control vs. Slight change</th>
<th>In-control vs. Out-of-control</th>
<th>Uniform vs. Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARL</td>
<td>ARL</td>
<td>ARL</td>
</tr>
<tr>
<td>50%</td>
<td>N/A</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>60%</td>
<td>100</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>70%</td>
<td>100</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>80%</td>
<td>50</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>90%</td>
<td>33.3</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>95%</td>
<td>8.3</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>98%</td>
<td>1</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

As shown in Table I, we considered the percentage of variance explained from 50% to 98% and calculated the ARL for all three experimental scenarios. Notably, we kept a reduced set of principal components (PCs) with large eigenvalues that explains a specific percentage of data variance. For 50% of variance explained, 8 PCs were retained and the ARL is not applicable in the scenario of in-control vs. slight change. When the percentage of variance explained increases from 60% to 98%, the ARL decreases from 100 to 1. For in-control vs. out-of-control cases, the ARL is 1 when the percentage of variance explained increases from 50% to 98%. For uniform vs. normal cases, the number of PCs retained is from
15 to 54 and yields the ARL from 1.01 to 1 when the percentage of variance explained increases from 50% to 98%. Therefore, experimental results show that multivariate monitoring performance is impacted by the number of PCs retained in the reduced-dimension space. Nonetheless, the percentage of variance explained is suggested to be greater than 98% to reduce the dimension and establish an effective and efficient monitoring scheme for a high-dimensional set of features.

VI. Conclusions

This paper presents a new approach of heterogeneous recurrence analysis for complex systems informatics, process monitoring and anomaly detection. In the state of the art, nonlinear recurrence methods are mainly based on homogeneous recurrences. In other words, all recurring states are treated in the same way in the recurrence plot. Very little work has been done to differentiate and quantify various types of recurrence behaviors (i.e., heterogeneous recurrences) in the dynamics of complex systems. However, process monitoring of dynamic transitions in complex systems is more concerned with aperiodic recurrences and heterogeneous recurrence variations in nonlinear and nonstationary systems. The present paper is the first of its kind to not only exploit heterogeneous recurrence dynamics but also design multivariate control charts for effectively monitoring complex systems. Experimental results on stochastic Markov processes and distribution-based processes show that the proposed methodology not only captures heterogeneous recurrence patterns, but also effectively monitors the changes in the dynamics of complex systems. Notably, multi-scale analysis of heterogeneous recurrences is shown to disclose more hidden information that is usually buried in the single-scale recurrence analysis.

As a final remark, this paper is presented in the context of nonlinear stochastic processes with a discrete and finite state space, which has many applications in various disciplines, e.g., sleep apnea study, queueing theory, quality control in the semiconductor industry. However, the proposed methodology of heterogeneous recurrence analysis is extensible to continuous state space that can be discretized into a finite set of ranges of interests. Our future research will focus on the investigation of heterogeneous recurrences in the continuous state space for process monitoring and control.

References