Heterogeneous recurrence monitoring and control of nonlinear stochastic processes

Hui Yang and Yun Chen

Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science 24, 013138 (2014); doi: 10.1063/1.4869306
View online: http://dx.doi.org/10.1063/1.4869306
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/24/1?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in
Laser process monitoring and automatic control at kHz rates through inline coherent imaging

A System Oriented Approach for the Optimal Control of Process Chains under Stochastic Influences
AIP Conf. Proc. 1389, 419 (2011); 10.1063/1.3636752

On Stochastic Processes with Constant Valuation

Relativistic Stochastic Processes: A review

Geometrical stochastic control and quantization
Heterogeneous recurrence monitoring and control of nonlinear stochastic processes

Hui Yang and Yun Chen
Complex Systems Monitoring, Modeling and Analysis Laboratory, University of South Florida, Tampa, Florida 33620, USA

(Received 10 December 2013; accepted 11 March 2014; published online 26 March 2014)

Recurrence is one of the most common phenomena in natural and engineering systems. Process monitoring of dynamic transitions in nonlinear and nonstationary systems is more concerned with aperiodic recurrences and recurrence variations. However, little has been done to investigate the heterogeneous recurrence variations and link with the objectives of process monitoring and anomaly detection. Notably, nonlinear recurrence methodologies are based on homogeneous recurrences, which treat all recurrence states in the same way as black dots, and non-recurrence is white in recurrence plots. Heterogeneous recurrences are more concerned about the variations of recurrence states in terms of state properties (e.g., values and relative locations) and the evolving dynamics (e.g., sequential state transitions). This paper presents a novel approach of heterogeneous recurrence analysis that utilizes a new fractal representation to delineate heterogeneous recurrence states in multiple scales, including the recurrences of both single states and multi-state sequences. Further, we developed a new set of heterogeneous recurrence quantifiers that are extracted from fractal representation in the transformed space. To that end, we integrated multivariate statistical control charts with heterogeneous recurrence analysis to simultaneously monitor two or more related quantities. Experimental results on nonlinear stochastic 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. © 2014 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4869306]

Rapid advancements of sensing technology have improved our ability to monitor complex systems in real time. Process monitoring of dynamic transitions in complex systems (e.g., disease conditions or manufacturing quality) is more concerned about aperiodic recurrences and heterogeneous types of recurrence variations. However, traditional recurrence plots treat all recurrence states homogeneously using the Heaviside function (i.e., the color code of all recurrence states is black, and non-recurrence is white). Very little work has been done to delineate heterogeneous recurrences (i.e., different kinds of recurrence behaviors). Notably, recurrence states can be different in kind because of state properties (e.g., state values and relative locations in the time series) and the evolving system dynamics (e.g., sequential state transitions before and after). Characterizing heterogeneous recurrences will bring the unique advantage to detect the variations in the nonlinear stochastic dynamics of complex systems, thereby improving the quality and integrity of operating processes. This paper presents a new approach of heterogeneous recurrence analysis for complex systems informatics, process monitoring and anomaly detection. Experimental results on stochastic Markov processes and distribution-based processes show that the proposed methodology not only delineates heterogeneous recurrence patterns in the fractal representation but also effectively monitors the changes in the dynamics of complex systems. The proposed methodology of heterogeneous recurrence analysis is presented in the context of nonlinear stochastic processes with a discrete and finite state space that will find many real-world applications in different disciplines, for examples, sleep apnea study, queueing theory, quality control in the semiconductor industry. Further, this present work is extensible to continuous state space that can be discretized into a finite set of ranges of interests.

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, the human heart is near-periodically beating to maintain vital living organs, and stamping machines are cyclically forming sheet metals during production. 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. Traditional linear methods interpret the regular structure, e.g., dominant frequencies in the signals. They have encountered certain difficulties to capture nonlinearity, nonstationarity and high-order variations in complex systems. For example, Fourier analysis does not provide the temporal localization of frequency components and assume spectral components exist at all times (i.e., stationarity).
Process monitoring of disease conditions or manufacturing quality is more concerned with aperiodic recurrences and recurrence variations in nonlinear and nonstationary systems. As shown in Fig. 1, vectorcardiogram (VCG) signal waveform at different segments changes significantly within one cycle, corresponding to different stages of cardiac operations. In addition, the waveform in one cycle is similar to others but with variations between cycles (Fig. 1(a)). The approach of nonlinear recurrence analysis characterizes recurrence behaviors in the high-dimensional state space. The recurrence plot was introduced by Eckmann et al. in the late 1980’s. As shown in Fig. 1(b), it captures topological relationships in the state space as a 2D image. If two states are located close to each other in the m-dimensional state space (e.g., 3D space in Fig. 1(a)), the color code is black (Fig. 1(b)). If they are located farther apart, the color is white. The structure of recurrence plot has distinct topology and texture patterns (Fig. 1(b)). 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 (e.g., small dots, vertical, and diagonal lines), 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).

However, very little work 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 using the Heaviside function in recurrence plots). As shown in Fig. 1, the color code of all recurrence states is black, and non-recurrence is white. Different kinds of recurrence behaviors, i.e., heterogeneous recurrences, are not distinguished in the high-dimensional state space. Notably, recurrence states can be different in kind because of state properties (e.g., state values and relative locations in the time series) and the evolving system dynamics (e.g., state transitions before and after). Fig. 2 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. However, their recurrence behaviors are heterogeneous, including state properties and pertinent system dynamics. Notably, the values and relative locations of $\vec{s}(30)$ and $\vec{s}(32)$ are different from $\vec{s}(1)$ and $\vec{s}(15)$ in the time series. States $\vec{s}(1)$ and $\vec{s}(15)$ have wider oscillations and their relative positions have a bigger span in the time series that spreads them further from the diagonal line in the recurrence plot (see Fig. 1). Furthermore, nonlinear stochastic dynamics differentiate them by state transitions that happen before and after. Nonetheless, heterogeneous recurrence patterns are not specifically considered in the state of the art. Characterizing heterogeneous recurrences will bring the unique advantage to detect the variations in the nonlinear stochastic dynamics of complex systems, thereby improving the quality and integrity of operating processes. This present paper is the first of its kind that aims to investigate heterogeneous recurrence behaviors for complex systems informatics and monitoring.

Process monitoring of dynamic transitions in nonlinear stochastic systems remains a challenging problem. There is an urgent need to develop new recurrence methods and tools that delineate heterogeneous recurrence variations for process monitoring and anomaly detection. Rapid advancements of sensing technology have improved our ability to monitor complex systems in real time. For examples, telemedicine remotely gathers ECG signals from cardiac patients in non-clinical settings. Manufacturing machines are installed a
variety of sensors to monitor process operations. However, current practices of process monitoring heavily rely on the human visual inspection (i.e., physicians or quality technicians) in both healthcare and manufacturing industries. It is prevalent in most cases to wait until the end of data acquisition to perform an inspection and assess process anomaly and operational quality. Such process monitoring practices inevitably lead to long time delays. This will not only cause a significant increase of defective products in manufacturing but also life-threatening events in cardiac care.

However, most of existing works focus on the characterization and quantification of recurrence dynamics in complex systems, without fully exploiting heterogeneous recurrence variations and linking with the objectives of process monitoring and anomaly detection. Recurrence quantification analysis (RQA) provides an effective means to quantify laminar, divergent and nonlinear transition dynamics in complex systems. To some limited extent, RQA measures appear to be promising features for the applications in real-time process monitoring. Nonetheless, traditional RQA measures are based on the recurrence plot that treats all recurrence states homogeneously. As a result, existing RQA measures cannot capture the patterns of heterogeneous recurrences. A primary reason is that conventional recurrence methods focus on continuous state space that is lag-reconstructed from time series. This poses significant challenges to delineate various types of recurrence states. Hence, there is an urgent need to design a new representation that present salient patterns of heterogeneous recurrences in the transformed space, as well as develop heterogeneous recurrence measures for the detection of dynamic transition in complex systems.

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). 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), 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, including the recurrences of both single state and multi-state sequence. Further, we developed a new set of heterogeneous recurrence quantifiers that are extracted from fractal representation in the transformed space. To that end, we integrated multivariate SPC methods with heterogeneous recurrence analysis to simultaneously monitor two or more related quantifiers. 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 experimental results, and Secs. VI and VII include 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., physiology, biology, economy, manufacturing, geophysics and neuroscience. In particular, various modifications and extensions of recurrence plot (RP) have been developed, including corridor thresholded recurrence plot, perpendicular recurrence plot, meta recurrence plot, order patterns recurrence plot to name a few. Further, Marwan et al. developed the bivariate cross recurrence plot to compare simultaneous recurrences of states in two different systems. Romano et al. developed a multivariate extension of RP, i.e., joint recurrence plot, for the detection of phase synchronization. These variations of RPs facilitate the prominence of hidden recurrence dynamics and provide graphical insights into dynamics of complex systems. However, very little work on the graphical representation of heterogeneous recurrences has been reported in the literature.

Recurrence quantification analysis was developed by Zbilut and Webber and extended by Marwan et al. to provide statistical measures of recurrence plots. It was shown that recurrence plots contain sufficient information pertinent to dynamical invariants of a complex system. Further recurrence networks were introduced to derive network measures (or graph theoretical properties) from the recurrence-based adjacency matrix for quantifying dynamical properties of complex systems. Network-theoretic measures (e.g., average path length, clustering coefficient, and degree centrality) provide new means to quantify recurrence dynamics of a complex system. In addition, we developed a novel multiscale framework to quantify recurrence dynamics in complex systems and resolve computational issues for big data. As opposed to traditional single-scale recurrence analysis, recurrence dynamics are characterized and quantified in multiple wavelet scales. Wavelet transform was shown to effectively separate system transient, intermittent and steady behaviors, thereby facilitating multi-resolution recurrence analysis. Furthermore, we have developed a local recurrence modeling approach that delineates time-varying recurrence characteristics with piece-wise models. Specifically, local recurrence patterns are exploited to predict the evolution of nonlinear systems under highly nonstationary conditions.

Recurrence plots and related quantification methods provide an effective means to characterize and measure 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 examples, Gao et al. and Donges et al. utilized recurrence network analysis to detect the transition in the dynamics of nonlinear time series and multivariate signals. Becker et al. proposed the use of recurrence determinism extracted from a single-channel EEG (Electroencephalography) activity to detect the consciousness of surgical patients. 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. This approach is the first of its kind to bring the confidence interval and hypothesis testing to the domain of statistical process control to gaining confidence in recurrence-based monitoring of nonlinear system dynamics.

Statistical control charts have been widely used in the monitoring and quality control of engineering processes since early 1920s. Note that control charts equally perform well for a complex process with measurable quality statistics. Some successful applications can be found in many disciplines, e.g., manufacturing processes, business marketing, surgical outcomes, customer service quality. However, nonlinear time series are usually gathered in real-time monitoring of complex systems. In this case, nonlinear features that represent hidden dynamical properties (or quality characteristics) need to be extracted for process monitoring. In-control status refers to the fact that the dynamics of complex systems do not change over time (i.e., null hypothesis $H_0$), while out-of-control is the transition of system dynamics (i.e., alternative hypothesis $H_1$). Notably, hypothesis testing to establish the confidence limits usually requires the knowledge of probability distribution of quality features in statistical process control. For example, Shewhart chart assumed the normal distribution of quality features for the detection of mean shifts in the process. Hence, there is a need to investigate the distribution of nonlinear features (e.g., recurrence quantifiers) extracted from time series. In addition, multiple nonlinear features are often extracted simultaneously, as opposed to a single feature. It is a challenging task to establish multivariate control charts for monitoring the changes in the dynamics of a complex system. However, very little work has been done to exploit nonlinear recurrence dynamics and establish statistical control charts of multivariate recurrence quantifiers for the change detection in the dynamics of a complex system.

Furthermore, 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 (see Fig. 3). The recurrence of state $i$ is different from the recurrence of state $j$ in the stochastic process. Fig. 3(a) shows the transition matrix $P$ of a Markov chain, where the $(i, j)$th element of $P$ is $p_{ij} = Pr(S_{n+1} = j|S_n = i)$. As shown in Fig. 3(b), each state transits to other states or itself with different probabilities, and the total transition probability equals one. Hence, each row of $P$ sums to one and all elements are non-negative. The probability of the first recurrence to $i$ at the $n$th step is $f(n) = Pr\{S_n = i, S_1 \neq i, \cdots, S_{n-1} \neq i|S_0 = i\}$. Heterogeneous recurrences of states are pertinent to the transition probabilities represented in the

\[
P = \begin{pmatrix}
0 & 0.0455 & 0.0999 & 0.1363 & 0.1818 & 0.2273 & 0.2273 & 0.0999 \\
0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \\
0.3846 & 0.077 & 0.2308 & 0 & 0 & 0.1538 & 0.1538 & 0.3846 \\
0.3846 & 0 & 0 & 0.3846 & 0 & 0.2308 & 0 & 0 \\
0.0968 & 0.129 & 0.0908 & 0.1613 & 0.2581 & 0 & 0 & 0.258 \\
0.0588 & 0.1176 & 0.0588 & 0.1176 & 0.0588 & 0.1176 & 0.0588 & 0.412 \\
0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 \\
0.3333 & 0 & 0 & 0.3333 & 0 & 0 & 0.3334 & 0
\end{pmatrix}
\]
transition matrix. Fig. 3(b) shows the state diagram that is a directed graph describing the state transitions. 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. This paper is presented in the context of nonlinear stochastic processes with a discrete and finite state space. 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.

Nonetheless, discrete processes with a finite state space have many applications in various disciplines. For examples, the study of sleep apnea often involves a discrete time series with a finite set of sleep stages (i.e., 0: Wake, 1: REM, 2: Stage1, 3: Stage2, 4: Stage3, 5: Stage4, 6: Artifact, 7: Indeterminate). In queueing theory, the analytical treatment of queues is also based on the discrete process with a finite state space. In the semiconductor industry, process monitoring examines the wafer and counts the number of defects on the wafer. Quality-improvement efforts adjust the process to keep the variations of the number of defects in control (i.e., below a control limit). Hence, the proposed methodology in the context of discrete processes with a finite state space will find many real-world applications in different disciplines.

### 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 time-delayed reconstruction. Few, if any, previous recurrence methods have focused on the characterization and quantification of heterogeneous recurrence behaviors and their applications in the change detection of dynamics of complex systems. The present investigation is the first of its kind to not only exploit heterogeneous recurrence dynamics but also design multivariate control charts for effectively monitoring the dynamics of complex systems.

As shown in Fig. 4, 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 time-delayed reconstruction) of nonlinear time series that effectively present heterogeneous recurrence information. As such, mathematical descriptions of salient patterns pertinent to heterogeneous recurrences as well as the procedures for feature extraction are much simpler and efficient in the transformed domain. (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.

#### 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. Ivanov et al. investigated the multifractality in the time series of human heartbeats and showed that multifractal spectrum for healthy controls is significantly different from congestive heart failures. Here, fractals refer to self-similar structures across temporal scales in the heartbeat time series. The method of wavelet transform modulus maxima was utilized to characterize self-similar structures across temporal scales and quantify the multifractal spectrum in nonlinear time series. It should be noted that the computation of fractal dimensions prefers a large amount of data because of the need of multiscale characterization. In addition, Webber 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, very little 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)\(^{43-46}\) to represent heterogeneous recurrences in time series. The basic motivation is to transform nonlinear time series from temporal domain to the 2-dimensional fractal domain and thereby extract the salient patterns of heterogeneous recurrences within the fractal space. As such, heterogeneous recurrences are effectively separated out of 8 possible categories. These categories are centered on the IFS mapping graph.\(^{43-46}\) Nonetheless, few previous works explored heterogeneous recurrences embedded in time series with the use of the fractal representation (e.g., IFS of a circle transformation).

For a discrete process with a finite set of states, the IFS sequentially maps each state \(\bar{s}(n)\) to an address (i.e., a point) \([c_x(n), c_y(n)]\) in the 2D coordinate system as:

\[
\bar{s}(n) \to k \in \mathcal{K} = \{1, 2, \ldots, K\}, \quad \begin{bmatrix} c_x(n) \\ c_y(n) \end{bmatrix} = \varphi(k, \begin{bmatrix} c_x(n-1) \\ c_y(n-1) \end{bmatrix}) = \begin{bmatrix} \cos(k \times \frac{2\pi}{K}) \\ \sin(k \times \frac{2\pi}{K}) \end{bmatrix}, \quad (1)
\]

where \(\begin{bmatrix} c_x(0) \\ c_y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\). Each state \(\bar{s}(n)\) 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 \(\mathcal{S}\) can be directly mapped to the set of categorical variables \(\mathcal{K} = \{1, 2, \ldots, K\}\). It should also be noted that the state space of a continuous process can be discretized into a finite set of categorical variables.

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 \([c_x(n-1), c_y(n-1)]\) with a weight \(x\) for both dimensions. Hence, the current state’s address \([c_x(n), c_y(n)]\) depends on all of its previous states because of the iterative function. The second part (i.e., \(\cos(k \times \frac{2\pi}{K}), \sin(k \times \frac{2\pi}{K})\)) uniformly distribute \(K\) addresses on the unit circle. For the current state \(\bar{s}(n)\), its address will be close to one of \(K\) addresses pertinent to its categorical variable \(k\). Therefore, the IFS transforms each state \(\bar{s}(n)\) 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 categorical variable and all of their previous states are the same. However, this is rarely true for time series from a nonlinear stochastic process.

Notably, the iterative mapping \(\varphi(k, c)\) is only contractive in the metric space \((X, d)\) if \(d(\varphi(k, v), \varphi(k, \omega)) \leq \vartheta_k \cdot d(v, \omega)\), \(\forall v, \omega \in X\) for the contractive factor \(\vartheta_k < 1\), where \(d\) is the distance measure and \(v, \omega\) are vectors in \(\mathbb{R}^2\).\(^{43-46}\) In addition, \(K\) addresses are uniformly distributed on the unit circle. The parameter \(x\) is important to avoid potential overlaps of two of \(K\) categories in the graph. Hence, the inequality of \(\vartheta_k \triangleq \sin(\frac{k}{K})\) needs to be satisfied so as to ensure that \(K\) categories are disconnected on the IFS mapping graph.\(^{43-46}\)

Fig. 5 shows the fractal representation of a Markov process with 8 categories (Fig. 3) via the IFS of a circle transformation. Each circle in Fig. 5(a) 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. As such, heterogeneous recurrences are effectively separated in the 2D graph in the level of individual states. Zooming
into circle 8 (i.e., marked by the blue rectangle in Fig. 5(a)) leads to Fig. 5(b). Every circle in Fig. 5(b) 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). The density of points in each circle is pertinent to the frequency of a two-state sequence. Notably, there are three two-state sequences (i.e., 48, 78, and 88) missing in Fig. 5(b), which indicates zero transition probability between these two states (also see the transition matrix in Fig. 3(a)). Further, Fig. 5(c) is the zoom into category 28. Each circle represents the recurrence of one of three-state sequences (i.e., 128, 228, 328, 428, 528, 628, 728, 828). 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 (from the individual state to multi-state sequences) 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 Sec. III B.

B. Heterogeneous recurrence quantification

Traditional recurrence plots treat all recurrence states in the same way as black dots. The proposed fractal representation (see Fig. 5) tags heterogeneous recurrences at different addresses in the 2D graph. This facilitates the delineation of heterogeneous recurrences in recurrence plots. In the level of individual states, the IFS of a circle transformation (see Eq. 1) centered the recurrences of single states on addresses that are uniformly distributed along the unit circle (see Fig. 5(a)). As such, heterogeneous recurrences of single states are effectively separated by their locations in the 2D graph. Traditionally, the recurrences of one single state (e.g., state 7) are treated in the same way as the recurrences of other states (e.g., state 2) in recurrence plots, because the distance is 0 for the recurrences of all the individual states.

However, the fractal representation provides an important opportunity to delineate various types of recurrences. Based on the addresses of single-state recurrences, Fig. 6(a) coded heterogeneous types of recurrences of 8 individual states with different colors. In contrast with black-white recurrence plots, a great deal of information is disclosed by separating various types of recurrences in heterogeneous recurrence plots. Furthermore, if we isolate (or zooming into) the recurrences of state 8 from Fig. 6(a) (or Fig. 5(a)), Figs. 6(b) and 5(b) represent only the recurrences of state 8. However, it should be noted that there are 8 different states that precede the state 8, giving rise to eight 2-state sequences (i.e., 18, 28, ..., 88). As a result, the fractal representation provides heterogeneous types of recurrences in the level of two-state sequence. Fig. 6(b) shows only the recurrences of state 8 from Fig. 6(a), but coded with 8 different colors based on the addresses of two-state sequences in Fig. 5(b). In other words, Fig. 6(b) further coded the recurrences of state 8 with 8 different colors based on the preceding state (see Fig. 6(b)). Similarly, Fig. 6(c) shows the recurrence patterns of 3-state sequences in Fig. 5(c). Notably, this process iteratively reveals salient information of heterogeneous recurrences in the level of individual states, 2-state sequences and 3-state sequences that are usually hidden in traditional recurrence plots.

Notably, 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 (see Fig. 6), 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 \( \varphi(k, c) : \tilde{s}(n) \rightarrow K = \{1, 2, \ldots, K\} \) 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., \( W_{k_1, k_2, \ldots, k_L} = \{\varphi(k_1, k_2, \ldots, k_L) : \tilde{s}(n) \rightarrow k_1, \tilde{s}(n-1) \rightarrow k_2, \ldots, \tilde{s}(n-L+1) \rightarrow k_L\} \) and \( k_1, k_2, \ldots, k_L \in K \). Here, \( k_1, k_2, \ldots, k_L \) denotes the \( L \)-state sequence. In other words, \( W_{k_1, k_2, \ldots, k_L} \) denotes the recurrence set of an individual state \( k_1, W_{k_1, k_2} \) denotes the recurrence set of 2-state sequence \( \{\tilde{s}(n) \rightarrow k_1, \tilde{s}(n-1) \rightarrow k_2\} \), and \( W_{k_1, k_2, \ldots, k_L} \) denotes the recurrence set of \( L \)-state sequence \( \{\tilde{s}(n) \rightarrow k_1, \tilde{s}(n-1) \rightarrow k_2, \ldots, \tilde{s}(n-L+1) \rightarrow k_L\} \). On the basis of the set \( W_{k_1, k_2, \ldots, k_L} \), 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 \( W_{k_1, k_2, \ldots, k_L} \), which is analogous to recurrence rate in conventional RQA measures. However, it may be noted that HRR is a fine-grained decomposition of recurrence rate.

\[
HRR = \left( \frac{\bar{w}_{k_1, k_2, \ldots, k_L}}{N} \right)^2,
\]

(2)
where \( \bar{w}_{k_1, k_2, \ldots, k_L} \) denotes the cardinality of set \( W_{k_1, k_2, \ldots, k_L} \). The square of \( \bar{w}_{k_1, k_2, \ldots, k_L} \) 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 \( \bar{w}_{k_1, k_2, \ldots, k_L} \) and total number of states in the stochastic process.

In addition, it should be noted that the set of \( W_{k_1, k_2, \ldots, k_L} \) denotes the recurrences of the same \( L \)-state sequence \( \{\tilde{s}(n) \rightarrow k_1, \tilde{s}(n-1) \rightarrow k_2, \ldots, \tilde{s}(n-L+1) \rightarrow k_L\} \) that are clustered at local regions in the 2D fractal graph (see Fig. 5). However, the addresses of \( L \)-state sequences in the set \( W_{k_1, k_2, \ldots, k_L} \) are not exactly the same and are distributed in the local region. Hence, we computed the distance matrix in the heterogeneous recurrence set \( W_{k_1, k_2, \ldots, k_L} \) as:

\[
D_{k_1, k_2, \ldots, k_L}(i, j) = \|\varphi^j - \varphi^i\|, \\
\varphi^j, \varphi^i \in W_{k_1, k_2, \ldots, k_L}; \quad i, j = 1, 2, \ldots, \bar{w}; \quad i < j,
\]

(3)
where \( \varphi^j \) and \( \varphi^i \) are the \( i \)th and \( j \)th elements in the set \( W_{k_1, k_2, \ldots, k_L} \). Heterogeneous mean (HMean) is defined as the average distance of \( D_{k_1, k_2, \ldots, k_L} \):

\[
HMean = \frac{2}{W(W-1)} \sum_{i=1}^{W} \sum_{j=i+1}^{W} D_{k_1, k_2, \ldots, k_L}(i, j).
\]

(4)
HMean provides general information about the average distance among elements in the set \( W_{k_1, k_2, \ldots, k_L} \). As shown in Fig. 5, some sets (e.g., \( W_{48}, W_{78}, W_{888} \)) are empty and others (e.g., \( W_{518}, W_{418}, W_{518} \)) 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 \( D_{k_1, k_2, \ldots, k_L}(i, j) \). It should be noted that we divide the distance matrix \( D_{k_1, k_2, \ldots, k_L} \) into \( B \) equally bins from 0 to \( max(D) \) and compute the probability as

\[
p(b) = \frac{1}{W(W-1)} \# \left\{ \frac{b - 1}{B} max(D) < D_{k_1, k_2, \ldots, k_L}(i, j) \leq \frac{b}{B} max(D) \right\},
\]

(5)
where \( b = 1, 2, \ldots, B \). Hence, the HENT is defined as

\[
HENT = -\sum_{b=1}^{B} p(b) \ln p(b).
\]

(6)
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 test for significance of dynamic transition is a test to determine whether there is a significant mean shift in the feature vector \( y = [y_1, y_2, \ldots, y_p]^T \), where \( p \) is the dimensionality of feature vector. Suppose that \( y \) 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. Hence, we assume that \( y \) follows a multivariate normal distribution, i.e.,

\[
f(y) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp \left[ -\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu) \right].
\]

(7)
If the transition of a dynamical system occurs, the joint distribution of multivariate features will vary. The hypothesis test is to identify data-driven evidences that reject the null hypothesis \( H_0 \) at a specific level of significance. Notably, normality assumption is required to formally establish confidence limits in the statistical test, but a slight deviation from normality will not seriously impact the results.\(^{13}\) In order to validate the normality assumption, we will use multivariate normal probability plotting to assess whether these heterogeneous recurrence features are approximately normally distributed, which will be detailed in the section of results.

Fig. 7 shows two possible schemes for monitoring the multivariate vector of heterogeneous recurrence features, and
thereby detecting the transition in the dynamics of a complex system. The first scheme to monitor each feature independently is a common approach in the literature\(^\text{13}\) (see Fig. 7(a)). The null hypothesis \(H_0\) holds only if all sample means \(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p\) fall within their perspective control limits. In the bivariate case, the pair of sample means \(\bar{y}_1, \bar{y}_2\) resides in the rectangular region in Fig. 7(a). However, this is very misleading because if the probability of Type I error is \(\alpha\) for each feature, then the true probability of Type I error for monitoring \(p\) features independently is \(1 - (1-\alpha)^p\). The probability that all \(p\) sample means are within the confidence limits is \((1-\alpha)^p\) if system dynamics do not change over time.\(^\text{13}\) Hence, the distortion is severe as the dimensionality of feature vector increases. In addition, the independent assumption among \(p\) features is often not valid because they are extracted from the dynamics of the same complex system. As a result, the first scheme to monitor each feature independently is inefficient and can potentially lead to erroneous conclusions.

Hence, there is a dire need to consider multivariate methods that jointly monitor these features. Fig. 7(b) shows the second scheme that assumes multivariate normal distribution for the \(p\) recurrence features under the null hypothesis \(H_0\) (i.e., the dynamics of complex systems do not change over time). The objective is to test the hypothesis that whether process means of multivariate random variables shift to some new values. Hence, the test statistic \(\chi^2 = (y - \mu)'\Sigma^{-1}(y - \mu)\) follows a chi-square distribution with \(p\) degrees of freedom. If there are not shifts in the means of multivariate recurrence features, then \(\chi^2\) values should be less than the upper control limit \(UCL = \chi^2_{\alpha,p}\), where \(\alpha\) is the significance level. If there are shifts in at least one of the means in multivariate recurrence features, then \(\chi^2\) values will be above the upper control limit. Fig. 7(b) shows the control ellipse \(\chi^2_{\alpha,p}\) graphically for two dependent features. Notably, the principal axes of the ellipse are no longer parallel to the \(y_1, y_2\) axes.\(^\text{13}\)

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 with sample covariance matrix \(S\), the test statistic becomes \(T^2 = (y - \bar{y})'S^{-1}(y - \bar{y})\), named Hotelling \(T^2\) statistic.\(^\text{17}\) The upper control limit for the Hotelling \(T^2\) statistic is:

\[
UCL = \frac{p(M + 1)(M - 1)}{M^2 - Mp} F_{x,p,M-p},
\]

where \(M\) is the number of samples, \(F_{x,p,M-p}\) is the upper 100% 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 \(S\). Because the transition probability between two states may be zero in the Markov Chain, this leads to an empty set for two-state sequences in the fractal representation (e.g., 48, 78, and 88 in Fig. 5(b)). As a result, the covariance matrix is singular due to zeros in heterogeneous recurrence quantifiers, thereby making the computation of Hotelling \(T^2\) statistic impractical.

In order to address this challenge, we transformed the feature matrix \(Y_{M \times p}\) into a set of principal components (PCs) that are linearly uncorrelated. First, the feature matrix \(Y_{M \times p} = [y_1, y_2, \ldots, y_M]'\) is centered by subtracting off column means, i.e., \(Y = [y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_M - \bar{y}]'\). Then, the singular value decomposition (SVD) of \(Y^*\) will be:

\[
Y^* = U \Psi V^T,
\]

where \(U\) and \(V\) 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 \(Y^*\)). Hence, the covariance matrix \(Y^*Y^*\) is decomposed as:

\[
Y^*Y^* = U \Psi^2 V^T U^T = \Psi^2 V^T,
\]

which is the eigen decomposition of \(Y^*Y^*\). The eigenvectors \(v_i\) are the principal component directions of \(Y^*\). As a result, principal components are given as:

\[
Z = Y^*V = U \Psi V^T V = U \Psi.
\]

The first principal component \(Z(:,1) = Y^*v_1\) has the largest sample variance among all the principal components. The sample covariance matrix \(S\) is easily seen to be:

\[
S = \frac{1}{M-1} \sum_{i=1}^{M} (y_i - \bar{y})(y_i - \bar{y})^T = \frac{1}{M-1} VZ^T Z V^T = V S_z V^T.
\]

Notably, principal components \(Z(:,i)\) and \(Z(:,j)\) are orthogonal with each other. Hence, the covariance matrix of principal components \(S_z\) is a diagonal matrix with diagonal entries \(\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_p^2\). The Hotelling \(T^2\) statistic becomes:

\[
T^2(i) = (y_i - \bar{y})'S^{-1}(y_i - \bar{y}) = Z(i,:)V^T (V S_z V^T)^{-1} VZ(i,:)^T,
\]

\[
= Z(i,:)S_z^{-1}Z(i,:)^T = \sum_{k=1}^{p} \frac{Z(i,k)^2}{\lambda_k^2},
\]
where \( Z(i,:) \) is the projection of the \( ith \) sample on principal component directions, that is, \( (y_i - \bar{y})'V \). In addition, a truncated \( M \)-by-\( q \) score matrix \( Z_q \) can be obtained by considering only the first \( q \) eigen values and eigenvectors \( (q < p) \), which explains the majority of variations in the feature matrix \( Y_{M\times p} \). Hence, the Hotelling \( T^2 \) statistic in the reduced dimension \( q \) is:

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

(14)

Because it is difficult to graphically construct the control ellipse for multivariate recurrence features (i.e., multivariate process monitoring). Therefore, 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 Sec. 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, i.e., in-control vs. out-of-control Markov processes, in-control vs. slightly changed Markov processes, and distribution-based processes (uniform vs. normal). Each process is represented by a time series (i.e., discrete and finite state space) that are random realizations from the Markov process. As discussed in the experimental design, we studied the performance of the proposed methodology of heterogeneous recurrence analysis in characterizing and monitoring distribution-based processes. 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

Suppose the Markov process in Fig. 3 is in control. The out-of-control Markov process was generated from another transition matrix (i.e., as shown in the \( P_{OC} \) below), which is significantly different from the in-control one (see Fig. 3(a)). In total, 100 samples are randomly generated for the in-control case and 50 samples for the out-of-control. As a result, heterogeneous recurrence patterns are significantly varied from in-control to out-of-control Markov processes.

B. In-control vs. slightly changed Markov processes:

Furthermore, we slightly perturb the transition matrix for the in-control Markov process (see Fig. 3) and generate a new Markov process. Such slight variations in the transition matrix are aimed at testing the sensitivity of heterogeneous recurrence analysis. Deviation from the in-control process is obtained by changing one row in the transition matrix and relatively preserving the summation in the row to be 1. Here, we changed the \( 7th \) row of in-control transition matrix (see \( p_{ij} \) in Fig. 3(a)) from \([0, 0, 0.5, 0, 0, 0.5, 0, 0] \) to \([0, 0, 0.6, 0, 0, 0.4, 0, 0] \), which is the transition probability from state 7 to other states. A total of 100 samples were generated from the in-control Markov process, and 50 samples from the slightly-changed Markov process. Each sample is a time series with 10 000 data points (i.e., random realizations from the Markov process).

C. Distribution-based processes (uniform vs. normal):

In addition to Markov processes, 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. Notably, distribution-based processes are random in terms of the underlying distribution and do not rely on the transition probability as in the markov process. Hence, this experimental scenario is to test the capability of heterogeneous recurrence analysis in characterizing and monitoring distribution-based processes. The experimental results are detailed in Sec. V.
missing, namely 128, 528, and 728. This indicates that transition probabilities of 1 → 2 → 8, 5 → 2 → 8 and 7 → 2 → 8 are zeroes in the out-of-control Markov process. By comparing fractal representations in Figs. 5 and 8, it is evident that heterogeneous recurrence patterns are effectively represented in the 2D fractal graph. Also, heterogeneous recurrence patterns are significantly different between in-control and out-of-control Markov processes from the level of individual states to multi-state sequences.

Next, multivariate features of heterogeneous recurrences were extracted from 100 samples of in-control process and 50 samples of out-of-control process. Each sample is a time series with 10,000 data points that are random realizations of this process. Before the implementation of multivariate Hotelling $T^2$ control chart, the normality assumption of heterogeneous recurrence features needs to be assessed. As aforementioned, we assumed that heterogeneous recurrence features $y$ follow a multivariate normal distribution if the dynamics of complex systems do not change over time. Here, we used the graphical tool of multivariate normal probability plotting. Suppose $y_i = [y_{i1}, y_{i2}, ..., y_{ip}]^T$ is a random sample from a $p$-dimensional normal distribution with population mean $\mu$ and covariance matrix $\Sigma$. Then, the quantity $r^2 = (y - \mu)^T \Sigma^{-1} (y - \mu)$ has the $\chi^2_p$ distribution with $p$-degrees of freedom. Let $r^2_j, j = 1, ..., M$ be the set of $M$ quantities derived from the feature matrix $Y_{M \times p}$. Multivariate normal probability plotting consists of two steps: (1) Order the derived quantities $r^2_1, ..., r^2_M$ in an increasing order $r^2_1 < r^2_2 < ... < r^2_M$, where parenthesized subscripts denote the ascending order of $r^2_j$ quantities. (2) Plot $r^2_j$ against the chi-squared probability points $\chi^2_p(j/p)$, $j = 1, 2, ..., M$, where $p_j$’s are equally spaced probabilities between 0 to 1, i.e., $p_j = (j - 0.5)/M$, and $\chi^2_p(\alpha)$ is the upper $\alpha$ probability point of chi-squared distribution. If heterogeneous recurrence features $y$ are multivariate normal, then $r^2_j$ is approximately $\chi^2_p$ distributed and the plot will be a roughly straight line. Fig. 9 shows the multivariate normal probability plot for 100 samples from in-control Markov processes. Notably, it is an approximately straight line, which indicates the validity of multivariate normal assumption in heterogeneous recurrence features extracted from in-control Markov processes.

Further, Hotelling $T^2$ statistics are derived from heterogeneous recurrence quantifiers to detect the changes in the Markov process. Notably, we transformed the data set of heterogeneous recurrence quantifiers into principal components to avoid the singular issue in the computation of Hotelling $T^2$ statistics. In Phase I, we used the in-control Markov process to establish the control limits. There are 100 in-control samples in Phase I to estimate eigenvectors $V$ and principal components $Z$. In Phase II monitoring, if a new sample comes, a new row will be added in the matrix of principal components and 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. 10) is:

$$UCL = \frac{p(M + 1)(M - 1)}{M^2 - Mp} F_{x,p,M-p} = 23.57,$$

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:

$$UCL = \frac{p(M + 1)(M - 1)}{M^2 - Mp} F_{x,p,M-p} = 64.57,$$

where $M = 100$, $p = 24$, $\alpha = 0.01$. 

![FIG. 8. Fractal representation of out-of-control Markov process via the IFS of a circle transformation: (a) addresses of individual states; (b) addresses of two-state sequence; (c) addresses of three-state sequence.](image-url)

![FIG. 9. Multivariate normal probability plot of in-control Markov processes.](image-url)
Fig. 10 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. 10 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 Figs. 10(a)–10(c)). 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. 10(d)).

### B. In-control vs. slightly changed Markov processes

Figs. 11 and 12 show fractal representations of in-control vs. slightly changed Markov processes, which were constructed with a circle transformation of a sample of 10000 data points that are random realizations from each process. As aforementioned, $p_{ij}$ in the transition matrix is changed from $[0, 0, 0, 5, 0, 0, 0, 5, 0, 0]$ to $[0, 0, 0, 4, 0, 0, 4, 0, 0, 0]$, while others remain the same. Such a slight change makes it difficult to visually differentiate fractal patterns between in-control and slightly changed Markov processes.

As shown in Figs. 11(a) and 12(a), we can hardly distinguish recurrence patterns of 8 possible categories in the large scale. Nonetheless, the IFS transforms each state $s(n)$ into an address that depends on all of its previous states, as well as its corresponding categorical variable. Although fractal patterns are not visually discernible in Figs. 11(a) and 12(a), the address for every state is different in the 2D fractal graph from in-control to slightly changed Markov processes. Figs. 11(b) and 12(b) are the zoom into category 3 in Figs. 11(a) and 12(a). It can be seen that some slight changes in the small-scale density of two-state sequences emerge. However, fractal patterns are still not visually distinguishable in Figs. 11(b) and 12(b). If we continue zooming into category 73, Figs. 11(c) and 12(c) show that three-state sequences have different distributions of point addresses between in-control and slightly changed Markov processes. For example, category 673 has one more cluster of points in Fig. 12(c) but not in Fig. 11(c). In addition, the density and distributions of point addresses in category 673 are also different (see red rectangular areas in Figs. 11(c) and 12(c)).

Fractal representations in Figs. 11 and 12 demonstrate the need to delineate heterogeneous recurrence in fine-grained resolutions. In contrast with homogeneous recurrence treatments in traditional recurrence plots, 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. In addition,
Fig. 9 shows that heterogeneous recurrence features extracted from in-control Markov processes are approximately multivariate normal. Now, the question becomes “Can heterogeneous recurrence quantifiers effectively identify slight changes in the Markov Process?”

Further, we derived Hotelling $T^2$ statistics from heterogeneous recurrence quantifiers for in-control and slightly-changed Markov processes. Fig. 13 shows Hotelling $T^2$ control charts for the feature set of HRR, HENT, HMean, and overall quantifiers. Due to the slight change in $p_{7j}$ in the transition matrix, the density of point addresses in multi-state sequences involving state 7 also changes. Hence, 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 Markov process (see Fig. 10). Because of the iterative nature of fractal representation, the address of each state depends not only on its categorical variable but also on 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. 13 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. 13(d)).

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. Figs. 14 and 15 show fractal representations of the uniformly distributed and normally distributed processes, respectively. Notably, the differences in the density and distribution of point addresses are visually appealing between two processes in Figs. 14 and 15. Zooming into local regions in the 2D fractal graph leads to fine-grained plots of multi-state recurrences (see Figs. 14(b)–15(c)). In contrast with the fractal representation of in-control Markov process (see Fig. 5), we found that point addresses are approximately uniformly distributed in Fig. 14 and normally distributed in Fig. 15. It can be seen from Fig. 15 that point addresses are dense when close to state 4 but sparse around states 1 and 8. In addition, few missing categories were found in the first three levels of fractal representations for the uniformly-distributed process (see Fig. 14). However, normally distributed process tends to have more missing categories starting from the third level of fractal representation (see Fig. 15). Such significant changes in fractal representations make the proposed heterogeneous recurrence method appealing for the monitoring of distribution-based processes.

Further, Fig. 16 shows the multivariable normal probability plot for heterogeneous recurrence features extracted from the uniformly distributed process. Note that this plot used a total of 100 samples and each has 24 features, including HRR, HENT and HMean for 8 individual states as shown in Fig. 14(a). The feature set from 100 samples plots roughly a straight line, indicating the validity of the assumption of multivariate normal distribution.
Fig. 17 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 (see Figs. 14 and 15). In addition, overall quantifiers combine HRR, HENT, HMean features and yield the best discriminatory power among these four control charts (see Fig. 17(d)).

VI. DISCUSSIONS

In this present investigation, multivariate control chart uses a single number (i.e., $T^2$ statistic) to characterize the feature vector of the process. Hence, process performance can be visually monitored by plotting the time sequence of $T^2$ statistics with control limits. This mitigates the difficulty of constructing the control ellipse for multivariate features when monitoring each feature independently. Notably, Hotelling $T^2$ statistic is a statistical measure of the multivariate distance of each feature from the center of the feature set. This is an analytical way to detect the out-of-control point in process dynamics. However, after the out-of-control point is detected, an immediate question is “which dimension in the feature vector is responsible for it?” Mason et al. proposed the decomposition of $T^2$ statistic to identify the responsible feature within the $p$-dimensional feature vector. This decomposition reflects the contribution of each individual feature by decomposing the $T^2$ statistic into $p$ orthogonal components as $T^2 = T^2_1 + T^2_2 + \ldots + T^2_p$. The first term $T^2_1$ is an unconditional Hotelling $T^2$ for the first feature in the feature vector. Conditional terms $T^2_{j,1,2,\ldots,j-1}$, $j = 1,\ldots,p$ represent conditional distribution of $y_j$ given all
other features. As such, the trigger of an out-of-control point is identified. For more details, see Ref. 47, for example.

In addition, 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 in the section of results (see Section V). If zooming into each state, there are 8 clusters of points characterizing heterogeneous recurrences of two-state sequences. The next question is “Does 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., $8 + 8 \times 8 = 72$) 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. As shown in Figs. 18–20, Hotelling $T^2$ control charts with heterogeneous recurrence features from both single states and two-state sequences clearly show much better discriminatory power for all three experimental scenarios than those only from the level of single states (see Figs. 10–17). These results show that multi-scale analysis of heterogeneous recurrences discloses more hidden information that is usually buried in the single-scale recurrence analysis.

It is worth mentioning that we used a reduced set of 50 principal components for the in-control Markov process in Figs. 18 and 19. This is because 50 out of 72 principal

FIG. 17. Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean; (d) Overall quantifiers, which are derived from the fractal representation in the first level of individual states. (Note: The first 100 samples are from the uniformly-distributed process and the last 50 samples are from the normally distributed process. The green dashed line represents the upper control limit.)

FIG. 18. Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean extracted from fractal representation in the levels of single states and two-state sequences for the in-control process (samples 1 to 100) and the out-of-control process (samples 101 to 150).

FIG. 19. Multivariate Hotelling $T^2$ control charts of (a) HRR; (b) HENT; (c) HMean extracted from fractal representation in the levels of single states and two-state sequences for the in-control process (samples 1 to 100) and the slightly-changed process (samples 101 to 150).
components have relatively large eigenvalues and explain nearly 100% variance in the feature data set. Thus, the upper control limit for in-control cases is $UCL = 194.88$, where $M = 100$, $p = 50$, $z = 0.01$. For the uniform case, we used all the 72 PCs with positive eigenvalues and the upper control limit is $UCL = 573.79$, where $M = 100$, $p = 72$, $z = 0.01$. Furthermore, HENT and HMean yield better discriminatory power than HRR in Fig. 18 for the in-control vs. out-of-control processes, while all three features yield comparable performances in Fig. 19 for the in-control vs. slightly changed processes. However, HRR is the most significant feature in Fig. 20 for uniformly distributed vs. normally distributed processes. These patterns are similar to those in the first level of individual states (see Figs. 10–17) but show much better discriminatory power in multi-scale analysis of heterogeneous recurrences.

Furthermore, if we combine all three heterogeneous recurrence quantifiers, there will be 216 features for each sample. This gives rise to the problem of “curse of dimensionality.” In other words, the number of features is greater than the dimensionality of available samples, thereby failing to establish an effective multivariate monitoring scheme. In order to address this problem, we reduced the dimensionality of feature set by only keeping those principal components with relatively large eigenvalues that explains the majority of data variance. Fig. 21 shows the Pareto chart of the percentage of variance explained by principal components that are sorted with respect to eigenvalues. As shown in Fig. 21(a), the principal component with the largest eigenvalue explains around 10% of total variance, and the first 30 principal components explain 95% of total variance in the feature data set from in-control Markov processes. However, only 5% of total variance is explained by the principal component with the largest eigenvalue in the uniform case, and the first 47 principal components explain 95% of total variance.

Hence, we conducted further experiments to investigate the performances of multivariate process monitoring in the reduced-dimension space. Here, we adopted the average run length (ARL) as a metric 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

$$ARL = \frac{1}{1 - \beta}, \quad (17)$$

where $\beta$ is the probability of type II error. In other words, $\beta$ is the probability of samples falling within control limits after the process is out of control.

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. It should be noted that 35 PCs (reduced from 216 features) explained 98% of variance and yielded the ARL of 1 for both in-control vs. slight change cases. This indicates that 1 sample is needed in average for the monitoring scheme with 35 PCs to first signal the out of control. In contrast, the ARL is 1 for in-control vs. out-of-control cases when the percentage of variance explained increases from 50% to 98%. Moreover, 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% for uniform vs. normal cases. 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.

<table>
<thead>
<tr>
<th>Percentage of variance explained</th>
<th>In-control vs. Slight change</th>
<th>ARL</th>
<th># of PCs</th>
<th>In-control vs. Out-of-control</th>
<th>ARL</th>
<th># of PCs</th>
<th>Uniform vs. Normal</th>
<th>ARL</th>
<th># of PCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>N/A</td>
<td>8</td>
<td>1</td>
<td>8.01</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td>100</td>
<td>11</td>
<td>1</td>
<td>11.01</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>100</td>
<td>14</td>
<td>1</td>
<td>14.01</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>50</td>
<td>18</td>
<td>1</td>
<td>18.01</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>33.3</td>
<td>25</td>
<td>1</td>
<td>25.01</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>8.3</td>
<td>30</td>
<td>1</td>
<td>30.01</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98%</td>
<td>1</td>
<td>35</td>
<td>1</td>
<td>35.01</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE I. Comparison of process-monitoring performances in the reduced-dimension space.

As a final remark, this paper is presented in the context of nonlinear stochastic processes and distribution-based processes show that the proposed methodology not only captures heterogeneous recurrence patterns in the fractal representation, 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. Furthermore, we reduced the dimensionality of feature set by only keeping those principal components with relatively large eigenvalues that explains the majority of data variance. This effectively addresses the singular problem of covariance matrix in the computation of Hotelling $T^2$ statistics, as well as the problem of “curse of dimensionality.” It should be noted that 35 PCs (reduced from 216 features) yielded the ARL of 1 for both Markov processes and distribution-based processes. Experimental results suggested the percentage of variance explained to be greater than 98% so as to reduce the dimension and establish an effective and efficient monitoring scheme for a high-dimensional set of features.

As a final remark, this paper is presented in the context of nonlinear stochastic processes with a discrete and finite state space. Notably, discrete processes with a finite state space have many applications in various disciplines, for examples, sleep apnea study, queueing theory, quality control in the semiconductor industry. It is expected that the proposed methodology in the context of discrete processes with recurrence behaviors (i.e., heterogeneous recurrences) in the dynamics of complex systems. Notably, most of existing works focus on the characterization and quantification of homogeneous recurrence dynamics in complex systems, without fully exploiting heterogeneous recurrence variations and linking with the objectives of process monitoring and anomaly detection. Traditional RQA measures are based on the recurrence plot that treats all recurrence states homogeneously, thereby failing to delineate heterogeneous recurrence patterns.

However, process monitoring of dynamic transitions in complex systems (e.g., disease conditions or manufacturing quality) is more concerned with aperiodic recurrences and heterogeneous types of 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 the dynamics of complex systems. First, we proposed a new fractal representation (rather than traditional time-delayed reconstruction) of nonlinear time series that effectively represent salient patterns of heterogeneous recurrences. As such, mathematical descriptions of heterogeneous recurrence patterns as well as the procedures for feature extraction are much simpler and efficient. Second, we developed a new set of statistical quantifiers that characterize and measure heterogeneous recurrence patterns from fractal representation in the transformed space. Experimental results showed that these extracted features are sensitive to the changes in recurrence dynamics of complex systems, instead of other extraneous factors. Third, we developed multivariate control charts with confidence intervals (rather than univariate change-point detection without significance measures) to simultaneously monitor two or more related quantifiers.
a finite state space will find many real-world applications in different disciplines. In addition, 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 complex systems informatics, process monitoring and anomaly detection.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (CMMI-1266331 and IOS-1146882) for the support of this research.