TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2023

 | ISBN 979-8-9882856-1-8

Mining Nonlinear Dynamics in Operational Data for Process Improvement

Hui Yang
Complex Systems Monitoring, Modeling and Control Laboratory, Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA 16802, USA, huiyang@psu.edu

Abstract The emergence of nonlinear and nonstationary dynamics is common when multiple entities collaborate, compete, or interfere in manufacturing and service operations. Operational management calls upon effective monitoring, modeling and control of in-process nonlinear dynamics. This, in turn, can result in significant economic and societal benefits. Nevertheless, traditional reductionist approaches often fall short in comprehending nonlinear dynamical systems. Also, the theory of nonlinear dynamics is mainly studied in mathematics and physics. A critical gap remains in the knowledge base that pertains to integrating nonlinear dynamics research with operations engineering. The need to leverage nonlinear dynamics has become increasingly urgent for the development of high-quality products and services. This tutorial presents a review of nonlinear dynamics methods and tools for real-time system informatics, monitoring and control. Specifically, we discuss the characterization and modeling of recurrence dynamics, network dynamics, and self-organizing dynamics hidden in operational data for process improvement. Further, we contextualize the theory of nonlinear dynamics with real-world case studies and discuss future opportunities to improve the monitoring and control of manufacturing and service operations. We posit this work will help catalyze more in-depth investigations and multi-disciplinary research efforts at the intersection of nonlinear dynamics and data mining for operational excellence.

Keywords nonlinear dynamics; operations research; data mining; system informatics

1. Introduction

Manufacturing and service operations often involve multifarious entities that cooperate, compete, or interfere with each other. As a result, sensor observations of these operations exhibit nonlinear and nonstationary dynamics. For example, Figure 1(a) shows nonlinear waveforms of power consumption signals when a lathe machine is removing metal from a workpiece [62, 75]. Figure 1(b) shows nonlinear variations of electrocardiogram (ECG) signals when the human heart maintains blood circulation through orchestrated depolarization and repolarization of cardiac cells [69, 72]. As complex systems evolve over time, sensor signals exhibit dynamic behaviors. Whether the system settles down to the steady state, undergoes incipient transitions, or deviates into high-order variations, sensor signals help analyze system dynamics. Operational management calls upon effective monitoring, modeling and control of in-process nonlinear dynamics which can result in significant economic and societal benefits. For example, manufacturing processes can make products achieving better quality and higher throughput. Heart disease is one leading cause of death in the world, amounting to an annual loss of $448.5 billion in the US [61]. Improving the quality of cardiac operations will reduce healthcare costs and increase the health of our society.

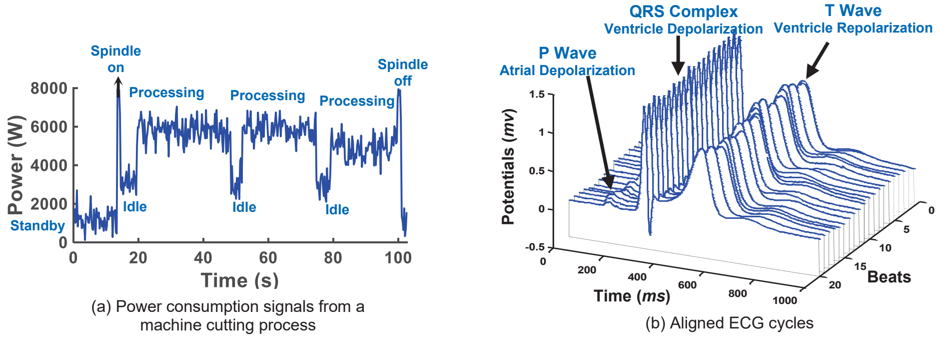


FIGURE 1. (a) Sample power consumption signals from a machining process. (b) Ensemble visualization of multiple ECG cycles from a human heart. Each ECG cycle consists of a PQRST complex, which is segmented into the P wave, QRS complex, and T wave to delineate atrial depolarization (and systole), ventricular depolarization (and systole), and ventricular repolarization (and diastole), respectively [25, 69]. A software package for ECG ensemble visualization is available via `https://www.mathworks.com/matlabcentral/fileexchange/77993-ensembles-and-waterfall-visualization-of-ecg-heart-beats`.

Nonlinear dynamics, however, pose significant challenges to data mining for operations management (e.g., the extraction of useful information from sensing data, and the development of data-driven intelligence). The traditional reductionist approach, which involves analyzing individual components separately and then combining them to understand a system's behavior, is limited in its ability to handle and comprehend nonlinear dynamical systems [50]. For example, it is often assumed that drugs which significantly reduce arrhythmic behaviors in isolated cardiac cells would have a similar effect in the heart, but later this concept was shown inadequate in two large clinical trials [30, 31]. In order to cope with system complexity, modern industries are investing in advanced sensing modalities (e.g., distributed sensor networks and imaging technology) for process monitoring [22, 23]. Real-time sensing gives rise to large amounts of data. Realizing the full potential of big data for quality control requires fundamentally new methodologies to harness and exploit complexity. However, there is a critical gap in the knowledge base that pertains to integrating nonlinear dynamics research with operations engineering. Specifically, the theory of nonlinear dynamics has been primarily studied within the domain of mathematics and physics [41, 58]. Available nonlinear dynamics techniques are either not concerned with specific objectives in operations engineering or fail to effectively analyze big data to extract useful information for process control. There is an urgent need to mine nonlinear dynamics in operational data for process monitoring and control.

This tutorial presents a review of nonlinear dynamics methods and tools for real-time system informatics, monitoring and control. To demonstrate general applicability of these methods, we use examples from two disparate disciplines, manufacturing and healthcare. Specifically, this tutorial consists of the following three sections:

(1) *Mining nonlinear recurrences for process modeling, monitoring and control*: This section reviews a suite of nonlinear methods to characterize and quantify multi-scale, heterogeneous recurrence dynamics in high-dimensional sensor signals. Specifically, statistical process control is realized by real-time computing and modeling of latent states hidden in the variations of process recurrences.

(2) *Network modeling and analysis of industrial imaging data*: This section focuses on the analysis of 3D images over time, also called 4D image profiles. Each 3D image is represented as a network and then structured into local communities. The evolving dynamics in community patterns of networks, rather than in the stream of raw images, are investigated for process monitoring and quality control.

(3) *Self-organizing dynamics for data visualization, topology learning and variable clustering*: High-dimensional variables (e.g., variation patterns of predictor variables or multi-lead sensor signals) are embedded as network nodes in the space. These nodes are then driven by nonlinear-coupling forces (i.e., measured between sensing variables) to form a self-organizing network. The resulting network topology helps the learning of geometrical structures, the clustering of sensing variables, the identification of anomaly patterns, and also the analysis of root causes.

Finally, we contextualize the theory of nonlinear dynamics with prior works and real-world case studies to improve the monitoring and control of manufacturing and service operations. Specifically, as shown in Figure 1, we will use sensor signals collected from machining operations and human heart studies to demonstrate the important concepts in nonlinear dynamics. Software packages and YouTube videos are also included for public dissemination. Sensor-based modeling in the contexts of manufacturing and healthcare will enrich the theory of nonlinear dynamics. We posit that this work will help bring nonlinear thinking into advanced sensing, system informatics and control, and further catalyze more in-depth investigations and multi-disciplinary research efforts at the intersection of nonlinear dynamics and data mining for operational excellence.

2. Background

Advanced sensing provides in-process observations of system dynamics and results in large amounts of data. In general, "informatics" refers to the extraction of useful information about system dynamics from the data [83]. We naturally embrace and accept data that are linear, stationary, or clean. However, real-world data are often nonlinear, nonstationary, and noisy. A linear dynamical system is often defined in the form of differential equations as:

$$\begin{cases} \dot{\mathbf{X}} = \mathbf{A}\mathbf{X} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases} ; \mathbf{X} \in R^n \tag{1}$$

Note that the system's evolution is linear, and the general solution can be obtained analytically. In other words, the solution is explicitly known for any time t:

$$\mathbf{X}(t) = e^{At}\mathbf{X}_0 \tag{2}$$

The stability of linear systems is determined by the eigenvalues of matrix $\mathbf{A}$. If $Re(\lambda) < 0$, the system is stable; but if $Re(\lambda) > 0$, it is unstable [19, 20].

On the other hand, nonlinear systems are difficult or impossible to solve analytically. The system's evolving dynamics form an attractor. Therefore, the attractor is also called the basin of attraction, which is defined as a set of states toward which nonlinear systems are dynamically evolving under a variety of initial conditions. A nonlinear dynamical system is often modeled in the form of differential equations as:

$$\dot{\mathbf{X}}(t) = \frac{d\mathbf{X}}{dt} = F(\mathbf{X}), F \in R^n \to R^n \tag{3}$$

where $F(\cdot)$ is a nonlinear function. The states of a nonlinear system $\dot{\mathbf{X}}$ consist of interdependent components. The space in which "states" are evolving is called a phase space (or state space) in the dynamical systems theory. For example, Eq. 4 shows the Lorenz model of atmospheric convection [3, 56]. Each state includes the x, y and z components that are are mixed and interdependent. Although the Lorenz model is deterministic, it is nonlinear. Figure 2 shows that the system's evolving dynamics form the Lorenz attractor. Time series observations of x, y and z components exhibit chaotic behaviors

$$\begin{cases} \dot{x} = 10(y - x) \\ \dot{y} = x(28 - z) - y \\ \dot{z} = xy - \frac{3}{8}z \end{cases} \tag{4}$$

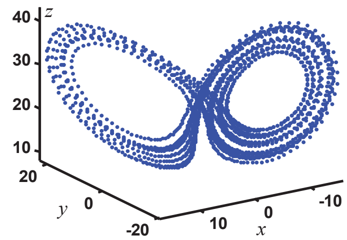


Figure 2. An example of the Lorenz attrator.

If stochastic components (e.g., random noise) are added into the Lorenz model (e.g., in any of the x, y or z dimensions), then it will become a nonlinear stochastic system. An attractor is formed when the dynamical system is evolving over time in the phase space. In a d-dimensional space, the state of a dynamical system is defined as a d-dimensional vector at time t

$$\vec{x} = [x_1(t), x_2(t), \ldots, x_d(t)]^T \tag{5}$$

The dynamics or equations of motion provide the causal relation between the present state and the next state, as shown in the Lorenz model (also see Eq. 4). This gives a rule for time evolution in the state space and helps predict what happens in the next step. Sensor observations of system dynamics often show complexity, e.g., randomness and irregularity. As a result, complex-structured models (e.g., high-order statistical models or stochastic processes) are generally developed to better fit the data and cope with the complexity.

An example of logistic map: However, it is not uncommon that complex behaviors (e.g., randomness and nonstationarity) can be exhibited from simple nonlinear systems. The *logistic difference equation* (or logistic map) is represented as $x_{t+1} = rx_t(1-x_t)$, which is widely used in ecology to describe the behavior of a population under the constraint of limited resources (e.g., food, water, space). The growth rate is not constant but rather proportional to the remaining capacity $(1-x_t)$. As shown in Figure 3, the observation x_t can reach steady state, exhibit periodic or chaotic behavior when its parameter r is varied. Figure 3(a) shows the observation x_t evolves into a steady state when $r = 2.8$, while Figure 3(b-c) shows x_t reaches period 2 when $r = 3.14$, and period 4 when $r = 3.5$. Figure 3(d) shows that logistic map exhibits chaotic behaviors when $r = 3.8$. See more details about chaotic behaviors of the logistic map in [27, 28].

Clearly, the logistic map shows more complex behaviors than the traditional linear growth model, $x_{t+1} = rx_t$, where the growth rate is constant. The next generation is linearly proportional to the population in the present generation. For example, if the population of fish in a lake grows by 10% in every generation, then the linear growth is modeled as $x_{t+1} = 1.1x_t$, where $r = 1.1$ is the constant growth rate. If the initial population is $x_0 = 100$, then the population sequence will be $100, 110, 121, 133$, and so on. Further, it is worth noting that logistic map possesses different characteristics from the exponential growth model, which uses the linear differential equation to describe that the rate of population change is proportional to the present population size as:

$$\dot{x} = \frac{dx}{dt} = rx, \quad \ln\frac{x(t)}{x_0} = rt, \quad \text{and} \quad x(t) = x_0 e^{rt} \tag{6}$$

Both linear growth and exponential growth models are predictable with analytical solutions. However, for logistic growth, a small difference in the value of r or x_0 can make a huge difference in the system behavior at time t. As shown in Figure 3(d), when $r = 3.8$, the simplest form of nonlinear system exhibits chaotic behaviors. The system has sensitive dependence on initial conditions (i.e., $x_0 = 0.2, 0.3$). It becomes difficult to predict the value of x_t when there is a small perturbation in the initial conditions.

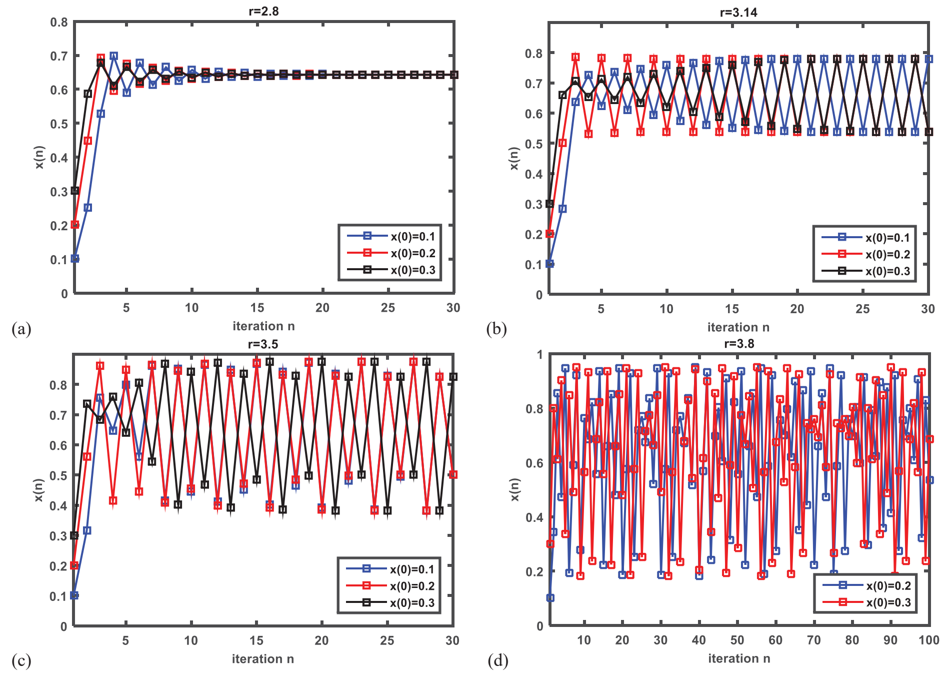


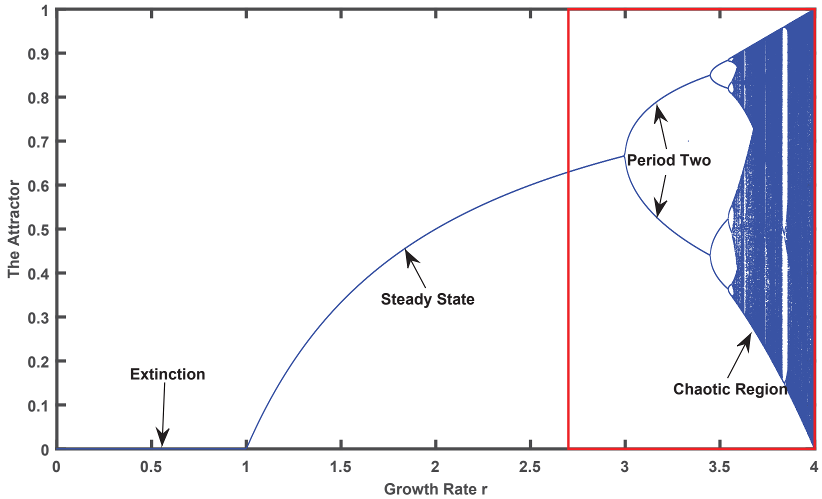
Figure 3. Examples of time series from the logistic difference equation (or logistic map) $x_{t+1} = rx_t(1-x_t)$: (a) r=2.8, (b) r=3.14, (c) r=3.5, (d) r=3.8. A small difference in the value of r or x_0 can make a huge difference in the system behavior at time t.

Figure 4 shows the bifurcation diagram of logistic map, which gives long-term system behaviors as a function of the parameter r. The x-axis represents different values of the parameter r, while the y-axis represents a set of states (or possible x_t values) towards which the logistic map evolves into. Self-similarity can be observed when zooming into the bifurcation diagram at different scales; also see the successive magnified views of the red rectangle regions in Figures 4a to 4c. In mathematics, self-similarity, also called fractal, is defined as a geometric object looking similar or approximately similar to a part of itself when you zoom in using different scales [51, 81]. Interestingly, there are narrow windows of order in the chaotic region (i.e., white vertical strips when $3.5 < r < 4$ in Figures 4a to 4c).

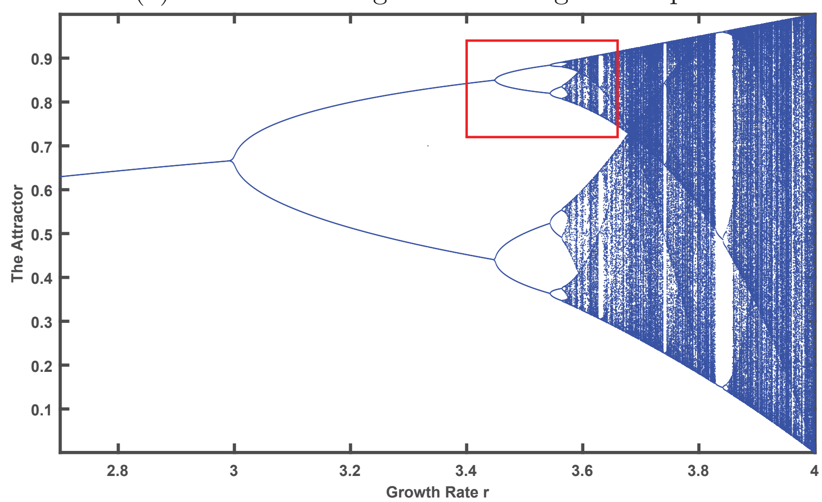
Furthermore, nonlinearity is not uncommon in real-world dynamical systems, e.g., the human heart [65] or the lathe machine [7]. Near-periodical beatings of the human heart provide nourishment to all parts of the body. Lathe machines cyclically process workpieces during production. Advanced sensing brings the proliferation of in-situ measurements of process dynamics (e.g., sensor signals, or functional profiles). There is an urgent need to harness and exploit nonlinear complexity underlying sensing signals for quality improvements in both cardiac and machining operations. Although the lathe machine and human heart are disparate in disciplines, both systems exhibit nonlinear dynamics phenomena as follows:

- **Nonlinear Dynamics**: Figure 1 shows nonlinear waveform signals of a single sensor from machining and cardiac operations. Distributed sensing allows for the deployment of multiple sensors at different locations and thereby offers multi-directional measurements of nonlinear dynamics present in the underlying processes [54]. Traditional linear methodologies mainly focus on analyzing individual components separately and then combining them to understand a system's behavior. Principal component analysis (PCA), Fourier analysis, and factor analysis are examples of such methods. These methods tend to be limited in their ability to capture nonlinear, nonstationary and high-order variations.

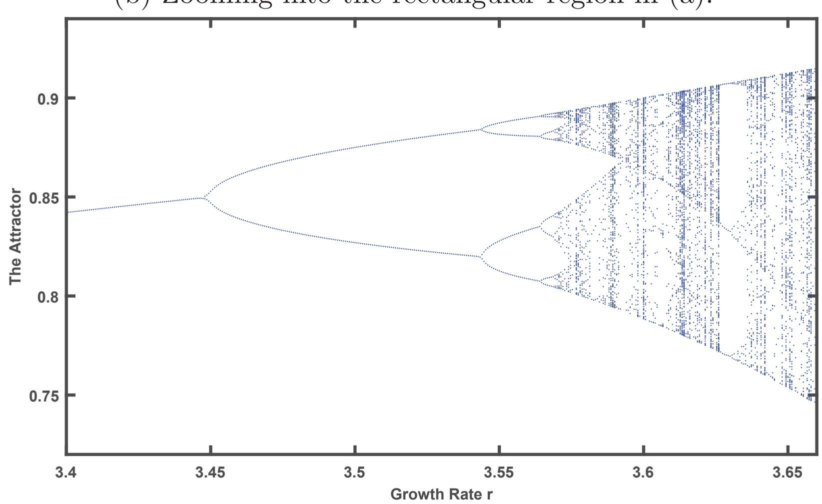
  In contrast, Poincaré proposed geometric analysis of nonlinear trajectories in the phase space (see Figure 5). This is known as geometric thinking of nonlinear dynamical systems in nonlinear theory. In other words, when analytical solutions are difficult or impossible



(a) Bifurcation diagram of the logistic map.



(b) Zooming into the rectangular region in (a).



(c) Zooming into the rectangular region in (b).

FIGURE 4. Bifurcation diagram of the logistic difference equation $x_{t+1} = rx_t(1-x_t)$. This diagram gives long-term system behaviors as a function of the parameter r.

to reach for nonlinear systems, Poincaré's way of nonlinear thinking focuses more on invariant properties of nonlinear attractors in the phase space.

- **Recurrence**: Nonlinear dynamical systems exhibit recurrence characteristics in the phase space. Figure 1 shows that the time-domain waveforms change significantly within one cycle. However, they are similar to each other between cycles, albeit with variations. Figure 5 shows an example of the ECG phase space constructed from multi-lead ECG signals using the Takens' embedding theorem [64]. The Poincaré recurrence theorem shows that if a dynamical system has the measure-preserving transformation, its trajectories eventually reappear in the ϵ-neighborhood of former states (see Figure 5) [42]. In other words, let T be a measure-preserving transformation of a probability space $(\mathfrak{X}, P)$ and let $\mathcal{A} \subset \mathfrak{X}$ be a measurable set. Then, for any natural number $N \in \mathbb{N}$, the trajectory will eventually reappear at neighborhood $\mathcal{A}$ of former states:

$$\Pr(\{x \in \mathcal{A} | \{T^n(x)\}_{n \geq N} \subset \mathfrak{X} \backslash \mathcal{A}\}) = 0 \tag{7}$$

  Heterogeneous recurrence variations are closely pertinent to dynamic transitions in complex systems (e.g., disease conditions, machine faults, product defects). There is an urgent need to investigate the variations of phase-space recurrences and link this with quality control objectives such as anomaly detection, process monitoring and fault diagnostics.
- **Self-organization**: As shown in Figure 5, the evolution of a dynamical system is self-organizing to form an attractor. The attractor is a set of states defining the basin of attraction in the phase space when complex systems are dynamically evolving in the long term. Self-organizing dynamical systems have the ability to adapt their behaviors to random perturbations and automatically evolve in the vicinity of the attractor. Examples of self-organizing processes include self-assembly of nanoparticles – a "bottom-up" nanomanufacturing technique to fabricate nanostructure materials [29], and electrical conduction in the heart – an orchestrated function of cardiac cells to generate heartbeats [24]. Notably, a simple interrelationship between entities (e.g., electrostatic force between nanoparticles, reaction/diffusion between cells) generates systems of enormous complexity. However, little has been done to exploit the knowledge of bio-inspired self-organization for the purpose of quality monitoring and diagnosis of root causes.

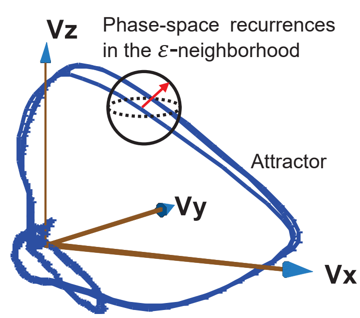


Figure 5. An example of nonlinear and nonstationary ECG trajectories in the 3-dimensional phase space. Here, the three dimensions are the Frank XYZ orthogonal ECG signals that capture geometric characteristics of the heart dynamics, as opposed to traditional time series observations. (Software package is available via https://www.mathworks.com/matlabcentral/fileexchange/58238-electrocardiogram-animation-and-state-state-representation).

Increasing levels of system complexity calls upon the development of new sensor-based models for better predictability and control. Nonetheless, there is limited work on studying phase-space nonlinear dynamics underlying sensing data for quality control. A review of literature pertinent to sensor-based modeling and analysis is detailed in the next section.

2.1. Manufacturing

Complex data structures (e.g., time series, image profiles) are commonly encountered in real-time sensing. Traditional statistical process control (SPC) methodologies primarily focus on monitoring key product or process characteristics and are not specifically designed to handle complex data structures. Therefore, in-process quality improvement is increasingly concerned with complex data structures (e.g., linear or nonlinear profiles, sensor signals, imaging data) over the past two decades [57]. For example, statistical monitoring of linear profiles was addressed in [40, 43]. In the case of nonlinear profiles, quality features are typically extracted for the purpose of monitoring and diagnosis. Koh and Shi et al. [44] leveraged wavelet transformation to analyze tonnage signals and detect faults in stamping processes. Zhou et al. [84] designed an SPC monitoring system specifically for near-periodic waveform signals. This system not only has the capability to detect a process change but also estimates the location and magnitude of mean shifts within the signal. Jin et al. [32] also developed new wavelet methods for "feature-preserving" compression of tonnage signals, and further decomposed tonnage signals into constituent parts at individual stations [33]. Paynabar et al. [55] developed wavelet mixed-effect models to characterize both within-profile and between-profile variations. Also, Gebraeel et al. [26] investigated sensor-based modeling and analysis of system degradation for reliability improvements and maintenance logistics. Bukkapatnam et al. [5, 6, 7] showed the existence of low-dimensional chaos in the machining process and further studied chatter control and fault diagnosis.

Despite these important advances, little has been done in the literature to investigate nonlinear dynamics in sensor signals for on-line quality monitoring. In general, there is little adoption of nonlinear thinking in quality engineering. Most existing works are more concerned with time-domain signals from a single sensor to extract quality characteristics. Although underutilized, multi-sensing capabilities provide a higher level of flexibility to reconstruct the phase-space attractor for the analysis of nonlinear dynamics. One advantageous feature of nonlinear thinking is that geometric segmentation is adaptive to system configuration, which circumvents the need to detect and select windows in segmenting time-domain signals. In other words, time windows are often required to segment time-domain signals, but are not necessary for geometric segmentation in the multi-dimensional space.

2.2. Healthcare

Quality is also the lifeline for electro-mechanical operations of the heart. However, there exists a substantial gap between quality engineering and cardiovascular science due to the isolation of individual disciplines as well as a lack of understanding of cardiovascular biophysics. ECG signals are rich in dynamic information that is highly relevant to cardiac operations and is thereby crucial for improving cardiac care. For instance, real-time ECG monitoring is conducive to early diagnosis and prognosis of cardiac ailments, thereby providing a new sensor-based framework for smart health management. However, a 1-lead ECG only captures 1-dimensional temporal view of space-time cardiac electrical activity. Multi-lead ECG systems (e.g., 12-lead ECG and vectorcardiographic (VCG) lead systems) deploy a distributed network of sensors (or electrodes) on the body surface. This helps to provide multi-directional views of such space-time dynamics [21, 69].

In the time domain, a normal ECG tracing is often segmented into the P wave, QRS complex, and T wave (see Figure 1b) to delineate atrial depolarization (and systole), ventricular depolarization (and systole), and ventricular repolarization (and diastole), respectively [74]. Most existing work has focused on the analysis of time-domain ECG signals from a single sensor. Time-domain algorithms were usually developed to quantify the characteristics of ECG wave deflections (i.e., P, QRS and T waves) [4]. Examples of ECG features include PR interval, RR interval, ST elevation/depression, QT interval, R amplitude. Also, Fourier analysis was utilized to transform time-domain ECGs to extract hidden features in the frequency domain [2, 60]. However, Fourier analysis does not provide temporal location of frequency components, and assumes spectral components exist at all times (i.e., stationarity). Nonstationarity in cardiovascular systems fueled increasing interests in wavelet analysis of ECG signals to delineate local time and frequency information for applications such as adaptive representation [64], PQRST segmentation [45], noise cancellation [1], and arrhythmia recognition [46]. Further, nonlinear methods were developed to reconstruct the phase space from 1-lead ECG and then characterize the dynamics of cardiovascular systems [82]. Notably, statistical control charts were also designed to monitor time-domain features (e.g., ST interval) extracted from 1-lead ECG signals [8, 59].

However, one limitation of existing methodologies is that they underutilize multi-lead ECG signals and overlook spatio-temporal dynamics in the heart. Multiple sensors at various locations on the human body respond to process changes differently. Time-domain ECG, i.e., a projected view of space-time cardiac electrical activity, diminishes important spatial information pertinent to faults in cardiovascular systems (e.g., myocardial infarction). Most existing methods are influenced by such an information loss, thereby failing to extract effective ECG biomarkers sensitive to cardiac malfunctions. Another issue is pertinent to root cause diagnosis in electromechanical operations of cardiac processes. Heart disease is complex and heterogeneous. Most existing studies focused on the binary interpretation of cardiac conditions, e.g., asking whether the heart is healthy or not. Nonetheless, heart disease can take place at different locations of the heart and with different temporal severity. Disease complexity results in variable patterns in multi-lead ECGs and poses significant challenges to root cause diagnosis. As a result, existing methodologies fall short of effectively addressing fundamental issues important to cardiac process monitoring and fault diagnostics. The existence of such methodological barriers hampers the transformation from clinic-centered care to patient-centered care that extends to the home, workplace, and community. Although an ECG is always obtainable in real time, cardiologists are not always available or able to analyze big data for warning signs. The challenges call upon the development of sensor-based models to address the complexity in cardiovascular systems.

3. Mining nonlinear recurrences in operational data for process modeling, monitoring and control

Recurrence (i.e., approximate repetitions of a certain event) is one of the most common phenomena in natural and engineering systems. The Complex Systems Lab at Penn State has done extensive research to design and develop recurrence methods for modeling and analysis of nonlinear dynamics in automotive assembly lines [66, 68] and cardiovascular systems [12, 65, 67]. The recurrence plot (RP) is a graphical tool to characterize the recurrence patterns of states in the phase space. As shown in Figure 5, a 3D ball that defines the ϵ-neighborhood is moving along the trajectory to capture phase-space recurrences. In other words, a recurrence plot characterizes the proximity of two states $\vec{x}(i)$ and $\vec{x}(j)$, i.e.,

$$R(i,j) := \Theta(\epsilon - \|\vec{x}(i) - \vec{x}(j)\|) \tag{8}$$

where $\Theta(\cdot)$ is the Heaviside function*, and $\|\cdot\|$ is a distance measure. The "rules of thumb" for choosing the ϵ-neighborhood include: (i) a fixed percentage of the maximum diameter of the state space; (ii) a fixed recurrence rate (i.e., the proportion of black dots in an RP); (iii) the number of neighbors for every state; (iv) the standard deviation of observational data. It is worth mentioning that if ϵ is too small, there will be few recurrence dots in the RP. As a result, RP patterns can not be observed and learned. If ϵ is too large, then almost every state will be a recurrence of every other state. There will be too many recurrence dots in

*The Heaviside function $\Theta(\nu) = 1$, if $\nu > 0$, and $\Theta(\nu) = 0$, if $\nu < 0$

the RP. In a nutshell, the choice of ϵ depends on the geometric shape and diameters of the attractor from a nonlinear system [52].

As shown in Figure 6, recurrence plot is an effective graphical tool that can reflect distinct structures of time series observations from dynamical systems. Here, we show RP examples for three types of different systems, including sine-wave observations from a linear system, Lorenz time series from a nonlinear system, and random noise from a randomized system. The pattern differences can be captured via both small-scale and large-scale structures. In small-scale structures, we can see the difference from single dots, diagonal and vertical lines. Single dots are signs of random fluctuation in the process. The RP of random noise shows a scattering of isolated dots, which indicates an uncorrelated random process. Diagonal lines are observed when the system is periodic and the system's states are recurring at different times. For example, the RP of sine waves shows regular lines along the diagonal direction, which indicate periodicity in the system dynamics. On the other hand, the RP of a Lorenz system only shows intermittent diagonal lines along with the scatter of isolated dots. Also, the RP of a Lorenz system shows intermittent vertical lines. This indicates that, for some time, the system's states do not change or change slowly. In large-scale structures, there are different homogeneous, periodic and disruptive patterns. If large-scale homogeneity is observed in RPs, then the system is stationary (e.g., diagonal lines in sine waves, and the homogeneous scatter of single dots in random noises). The period patterns correspond to the cyclicity. The period is correlated to the distance between diagonal lines. The disruptive patterns (e.g., vertical disruptions in the Lorenz RP) often indicate nonstationarity due to switching or transitional behaviors in the system's evolving dynamics.

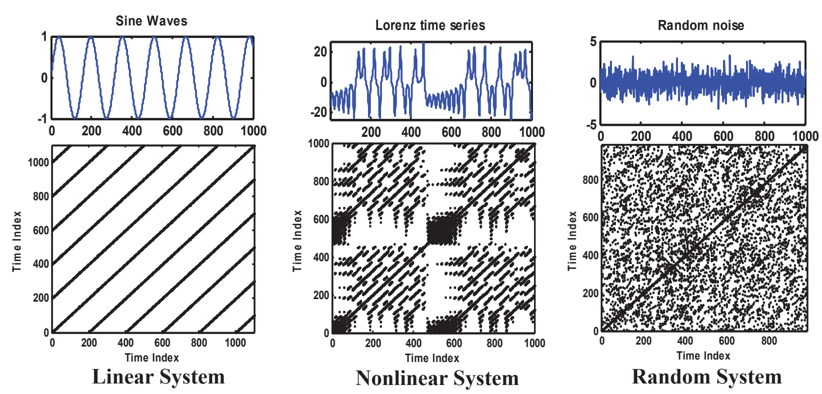


FIGURE 6. Illustrations of distinct RP patterns (e.g., either small-scale or large-scale structures) for sine waves, Lorenz time series, and random noise.

In addition to simulation datasets, Figure 7 is an example of recurrence plot that captures geometric recurrences in the ECG phase space of Figure 5. Recurrence quantification analysis (RQA) measures intriguing structures and patterns in the recurrence plot [63]. Examples of RQA measures include recurrence rate, determinism, entropy, and laminarity; see the detailed definitions in [53, 79]. Chen and Yang et al. [12, 65, 67] also developed a multi-scale framework to characterize and quantify the dynamics of transient, intermittent and steady recurrences within wavelet scales. Multi-scale analysis facilitates the prominence of hidden recurrence properties that are usually buried in a single scale. Shorter wavelet subseries also make expensive recurrence computations not only plausible but also more effective within wavelet scales. Moreover, Yang et al. [66, 68] developed a local recurrence modeling approach to delineate time-varying recurrence characteristics in multi-stage manufacturing systems. Specifically, local recurrence patterns were investigated for real-time performance prediction under highly nonstationary conditions.

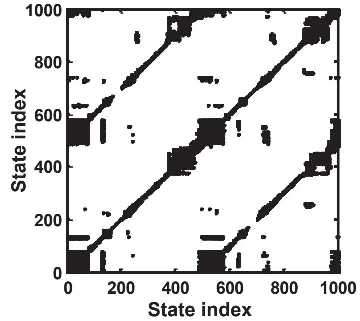


FIGURE 7. The recurrence plot characterizes the proximity of two states $\vec{x}(i)$ and $\vec{x}(j)$, i.e., $R(i,j) := \Theta(\epsilon - \|\vec{x}(i) - \vec{x}(j)\|)$, where Θ is the Heaviside function and $\|\cdot\|$ is a distance measure. Software package is available via https://www.mathworks.com/matlabcentral/fileexchange/58246-tool-box-of-recurrence-plot-and-recurrence-quantification-analysis.

Further, our prior work has developed a new approach to investigate heterogeneous recurrence variations and link with quality control objectives, namely "**heterogeneous recurrence analysis** (HRA)"[15, 71]. Cardiovascular and manufacturing operations are more concerned with heterogeneous recurrences and their variations hidden in nonlinear and nonstationary sensor signals. Traditional recurrence plots treat all recurrence states homogeneously using the Heaviside function. As shown in Figure 7, the color code of all recurrence states is black and non-recurrence is white. In other words, if there is a spatial partition of the phase space (e.g., Figure 8), recurrences in one box are treated the same as in other boxes. In the state of the art, little has been done to delineate and characterize heterogeneous types of recurrence behaviors (e.g., different kinds of recurrence patterns in vicinity regions in the phase space). As shown in Figure 8, recurrence patterns can be different in kind because of state properties (e.g., state values and relative locations due to the spatial partition of the phase space) and the evolving system dynamics (e.g., sequential state transitions before and after a present state) [71].

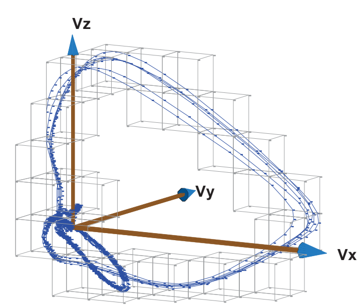


FIGURE 8. An example of spatial partition of the ECG phase space.

If a categorical index k is assigned to each box in Figure 8, then the states $\vec{x}(n)$ tend to make transitions between these boxes when the system is dynamically evolving. The HRA framework starts with the design of a fractal representation of Markov states via the iterative function system (IFS), which was originally developed for fractal analysis of categorical time series. For a finite state space, the IFS sequentially maps each Markov state $\vec{x}(n)$ to an address (i.e., a point $[c_x(n), c_y(n)]$) in the 2D coordinate system using

$$\vec{x}(n) \longrightarrow k \in \mathcal{K} = \{1, 2, \ldots, K\} \tag{9}$$

$$\begin{bmatrix} c_x(n) \\ c_y(n) \end{bmatrix} = \varphi\left(k, \begin{bmatrix} c_x(n-1) \\ c_y(n-1) \end{bmatrix}\right) = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} c_x(n-1) \\ c_y(n-1) \end{bmatrix} + \begin{bmatrix} \cos(k \times \frac{2\pi}{K}) \\ \sin(k \times \frac{2\pi}{K}) \end{bmatrix} \tag{10}$$

where the initial address is $\begin{bmatrix} c_x(0) \\ c_y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. See more mathematical details in [71]. Figure 9 illustrates the representation of a stochastic Markov process with 8 states (i.e., $K = 8$). Each circle in Figure 9a shows the recurrence of one of the 8 states. Zooming into circle 1 leads to Figure 9b. Every circle in Figure 9b provides information on heterogeneous recurrences in state 1 (i.e., all two-state sequences that end in state 1). Figure 9c shows a magnified view of category 21. Each circle represents the recurrence of one of the three-state sequences (i.e., 121, 221, 321, 421, 521, 621, 721, 821). Notably, the density and distribution of points in each circle characterize heterogeneous recurrence variations, which have a direct association with conditions and settings of process operations.

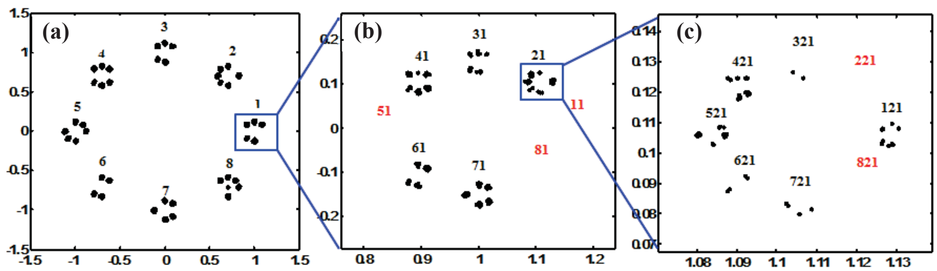


Figure 9. Multi-scale fractal representation of a Markov process with 8 discrete states via the iterative function system: (a) addresses of mono-state; (b) addresses of two-state sequences; (c) addresses of three-state sequences, where the x-axis represents $c_x(n)$ and the y-axis is $c_y(n)$. Software package is available via `https://www.mathworks.com/matlabcentral/fileexchange/58250-heterogeneous-recurrence-analysis-toolbox`.

The novel HRA approach has found diverse applications in disparate disciplines of manufacturing and healthcare. For example, our prior work investigated heterogeneous recurrence variations in the nonlinear state space from a time series of heart rate variability (i.e., extracted from single-channel ECGs) [18]. Experimental results showed that heterogeneous recurrence analysis is an effective tool to detect obstructive sleep apnea using one-lead ECG signals. Also, we reconstructed the state space of in-situ vibration signals that are collected during the ultraprecision machining (UPM) process. Heterogeneous recurrence variations were then studied and demonstrated to effectively capture dynamic transients for monitoring the quality of UPM surface finishes [35]. Further, Chen et al. [10] leveraged the heterogeneous recurrence analysis to delineate disease-altered spatiotemporal patterns in multi-channel cardiac signals. In addition, Yang et al. [70] extended the heterogeneous recurrence approach for spatial data analysis. Spatial data were traversed with the Hilbert Space-Filling Curve to transform the variations of recurrence patterns from the spatial domain to the state-space domain. The spatial HRA approach was evaluated and demonstrated in a real-world case study to analyze microscopic images for manufacturing quality control.

Finally, our prior work has developed a multivariate strategy of sparse particle filtering (SPF) for on-line monitoring and predictive control of process recurrences [11, 14]. This is realized by real-time computing, recursive predicting and updating of latent states hidden in the variations of process recurrences. Specifically, SPF tackles two main challenges: (i) *Nonlinear and non-Gaussian properties*: To effectively model the dynamics of process recurrences, one has to consider measurement and process functions in nonlinear forms and model the hidden states in a non-Gaussian distribution. However, most existing methods are only effective when the state space model is linear and the posterior density follows a Gaussian distribution. For example, the Kalman filter assumes a linear state-space model. Dynamic factor analysis requires the assumption of a Gaussian distribution. The SPF can effectively model the evolving dynamics of latent states, as well as represent complex distributions (e.g., non-Gaussian, multi-modal) with particles generated by sequential Monte Carlo sampling.

(ii) *Curse of dimensionality*: It is also worth noting that the dimensionality of recurrence variables is very large. This poses a serious "curse of dimensionality" problem for estimation and prediction steps. However, traditional latent-variable methodologies (e.g., PCA, ICA, nonlinear factor analysis) do not account for the dynamics of hidden variables even though complex systems evolve dynamically. The SPF approach identifies a compact set of latent variables sufficient to model the nonlinear dynamic process.

4. Dynamic network modeling of industrial imaging data

Rapid advances in 3D metrology now bring the increasing availability of imaging data for quality inspection and process improvement. For example, the quality of a 3D-printed part is pertinent to melt pool geometry variations (also see Figure 10a) in the laser powder bed fusion (LPBF) process [49, 80]. Ultra-precision machining (UPM) is equipped with air-bearing spindles and diamond tools for high-end manufacturing applications such as aerospace and semiconductor manufacturing. UPM is capable of generating optical surface finishes with an average roughness $\leq 10nm$ (see Figure 10b). Time-varying 3D images can be captured through optical mapping techniques to observe the propagation and conduction of electrical waves in the whole heart (Figure 10c). Cardiac imaging is critical to studying the mechanisms of heart diseases and observing real-time feedback in the course of cardiac surgeries [17, 73]. The laser-scanning video microscopy captures in-vivo 4D images of living cells (Figure 10d) during the synthesis of biopharmaceutical products [34, 39]. Dealing with 3D and time-varying 3D (4D) imaging data is a general problem facing both traditional and next-generation innovation practices in manufacturing and biotechnology.

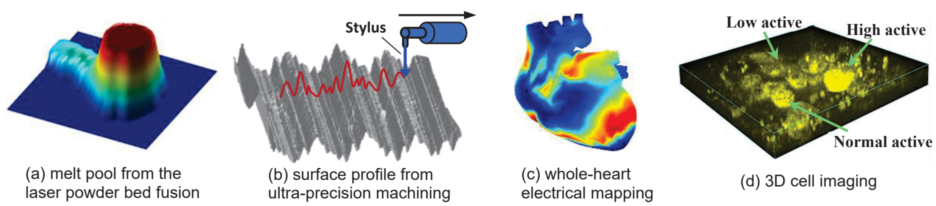


Figure 10. Examples of industrial imaging data from manufacturing and healthcare domains. See our software package for whole-heart simulation via https://www.mathworks.com/matlabcentral/fileexchange/78194-simulating-spatiotemporal-dynamics-in-the-whole-heart.

However, traditional studies tend to focus more on statistical modeling and analysis of 2D surface profiles, and are limited in the ability to model and analyze complex spatiotemporal patterns in 3D (4D) imaging data. When there is a stream of 3D images dynamically evolving over time, we obtain a new 4D "image" – a sequence of 3D images for every point in time (e.g., whole-heart electrical mapping and 3D cell imaging in Figures 10c and d). This section will present our previous research works in the design and development of dynamic network methods for monitoring and control of high-dimensional imaging streams [9, 37, 38, 70]. Notably, dynamic networks have the following advantageous features: (i) Network nodes can effectively represent high-dimensional data (e.g., either 3D or 4D imaging). (ii) Network edges preserve complex spatial structures and patterns in the time-varying stream of images. The dynamic network methodology is evaluated and validated with real-world case studies for quality monitoring of cardiac operations and surface finishes in manufacturing processes. Our prior work has developed a novel two-phase greedy procedure for automatic detection and characterization of calcium sparks in living cells [39]. Time-varying imaging data are taken at different times in a biomanufacturing process that produces genetically-engineered cells. Each image contains complex patterns of disease-altered cellular behaviors (e.g., density, activity).

4.1. Dynamic network representation and community detection

Our prior work [37, 38] proposed to represent industrial imaging data as dynamic networks and then detect subnetwork structures in the network. As shown in Figure 11, each image will be cast as a weighted and undirected network. Each node represents a pixel or a group of pixels in the image. The edge $E_{i,j}=1$ if node i is linked to node j and $E_{i,j}=0$ otherwise, i.e., $E_{i,j}=\Theta(\epsilon-\|v_i-v_j\|)-\delta_{i,j}$, where v_i is the intensity/color of node i, and $\delta_{i,j}$ is the Kronecker delta used to avoid self-loops. Each edge is assigned a weight W_{ij}, which depends on both spatial and intensity factors, i.e., $W_{ij}=\alpha\exp(\|s_i-s_j\|/s_{max})+(1-\alpha)\exp(\|v_i-v_j\|/v_{max})$. The regularization term α controls the balance between spatial and intensity similarities, s_i is the spatial location of pixel i, v_{max} is the maximal intensity and s_{max} is the maximal spatial distance. Thus, two connected nodes in an image have a smaller weight if their intensities and geometric locations are closer to each other.

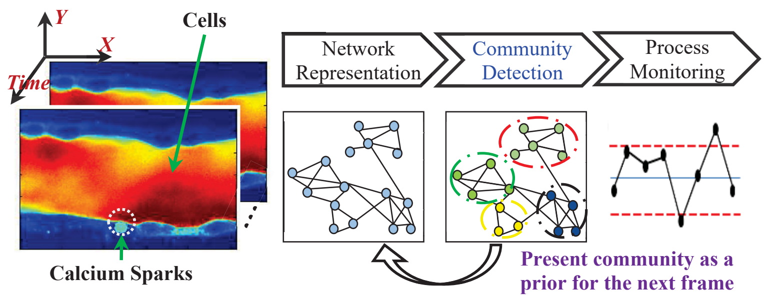


Figure 11. Dynamic network modeling and monitoring of biomanufacturing imaging data.

However, the challenge lies in computing $E_{i,j}$ and $W_{i,j}$ for a large image or improving the computational speed for a moderate-size image. Traditionally, an image is resized (or compressed) by imposing a coarse grid to compute local averages. This unavoidably leads to information loss and influences the performance of fault detection. Our prior work [70] investigated a wavelet network casting approach to improve computational efficiency, and extract hidden properties that are usually buried in the single-scale image. Notably, wavelet transforms model irregular image patterns such as abruptions and intermittences better than the Fourier transform and standard statistical methods (e.g., splines or polynomials). Multi-scale decomposition and subsampling reduce the image size significantly, but do not lose any information. The original image can be perfectly reconstructed from the set of wavelet sub-images.

After the image is represented as a network, there are sub-groups or sub-communities to be identified in the next step. Each community is characterized as a group of tightly connected nodes within the network. As such, these sub-communities provide pixel clusters that correspond to different types of inherent patterns in the image. We use a "Potts type" variable g_i to denote the community label of node i. If there are K communities in the network, then $1\le g_i\le K$. The vector of $\{g_i\}_{i=1,...,N}$ denotes the community labels for all the N nodes. A subgroup of nodes with the same value, e.g., $g_i=k$ and $1\le k\le K$, belong to the same community C_k. For a specific partition of the network into K communities, the Hamiltonian function is introduced to define the total energy function of the network as:

$$\mathcal{H}(\{g_s\}_{s=1}^N)=\frac{1}{2}\sum_{k=1}^{K}\sum_{i,j\in C_k}(W_{ij}-\overline{W})[\Theta(\overline{W}-W_{ij})+\gamma\Theta(W_{ij}-\overline{W})]\delta(g_i,g_j) \tag{11}$$

where $\overline{W}$ is the average of all weights, γ is a regularization parameter, $\delta(g_i,g_j)$ is the Kronecker delta, and $i,j\in C_k$ denote nodes i and j belonging to the same community C_k. If the Hamiltonian function is minimized, $min_{g_i,g_j}\mathcal{H}(\{g_s\}_{s=1}^N)$, then the total energy of the network is reduced to a minimal state. The network tends to be more stable. As a result, the Hamiltonian minimization process favors strongly-connected nodes and penalizes loosely-related nodes appearing in the same community. In other words, if strongly-connected nodes are in the same community, network energy will be reduced. If loosely-related nodes are in the same community, network energy will be increased. Similarly, the Hamiltonian minimization process penalizes strongly connected nodes and favors loosely related nodes being in different communities. See mathematical details in [38]. This optimization problem is analogous to community detection for social or metabolic networks. However, traditional studies have been limited to networks and did not specifically consider industrial imaging data. This section focuses on community detection in a stream of dynamic images for process monitoring and quality control, which is even more challenging.

4.2. Dynamic network process monitoring and quality control

Solving this optimization model requires considerable time for an image, which can be longer than the sampling rate of dynamic images. In this case, it will be computationally infeasible for on-line process monitoring. Thus, we propose an adaptive learning approach that recursively updates the network communities, as shown in Figure 11. In other words, community detection does not start from scratch for every frame of image, but utilizes a priori information in previous frames. Notably, the difference between successive frames is small in functional image data. This is because the functional image tracks fine-grained details of the initiation and progression of events. As a result, adaptive learning will make on-line monitoring computationally efficient and feasible.

Further, we propose to develop community statistics on the basis of the state of subgraph structures and further design multivariate network control charts for process monitoring. The variations of community densities distinctly mark the emergent properties of anomalies in the functional image. Most importantly, each pixel (or node) has a "Potts type" variable that represents the community label. This offers a great advantage for fault pattern detection, including the time, location, and size factors. Our prior work [38] evaluates the performance of the dynamic network methodology under various uncertainty factors (e.g., fault types, nonstationary movements of subjects, and noisy background) in the functional image and how they influence process monitoring and fault pattern detection. Experimental results show that new dynamic network monitoring schemes are adaptive to nonlinear and nonstationary variations, and are generally applicable to research problems in diverse fields such as biomanufacturing quality control, whole-heart imaging analysis, surface profile monitoring, as well as laser processing of metal powders.

5. Self-organizing dynamics for data visualization, topology learning and variable clustering

This section discusses the exploitation of self-organizing dynamics for the purpose of data visualization, topology learning and variable clustering. Self-organizing dynamical systems possess the capability to adjust their actions in response to unexpected disturbances and naturally evolve in the proximity of a nonlinear attractor. The attractor is a set of states defining the basin of attraction in the phase space when complex systems are dynamically evolving in the long term. For example, the human heart is an autonomous and self-organizing dynamical system. The orchestrated function of cardiac cells leads to the propagation and conduction of electrical waves and generates near-periodic heartbeats [24]. Also, the self-assembly of nanoparticles is used as a "bottom-up" nanomanufacturing technique to fabricate nanostructure materials [29]. A simple interrelationship between entities (e.g., an electrostatic force between nanoparticles, or a reaction/diffusion between cells) often generates system dynamics of enormous complexity.

Our prior work [77] leveraged the self-organizing principles to derive and learn the unique topology of a recurrence network. See an example of the self-organizing process for Lorenz nonlinear systems at this site `https://youtu.be/bhcj4ljKjYs`. As shown in Figure 7, the recurrence plot characterizes the state space of a nonlinear dynamical system via a binary matrix with "zeros" and "ones" as

$$R_{ij} := \Theta(\epsilon - \|\vec{\boldsymbol{x}}(i) - \vec{\boldsymbol{x}}(j)\|) \tag{12}$$

This binary matrix is further leveraged to derive the adjacency matrix A_{ij} of a complex network, as shown in Figure 12. A unifying framework to define the recurrence network is given by $A_{ij} = R_{ij} - \delta_{ij}$, where the Kronecker delta δ_{ij} is used to avoid self loops.

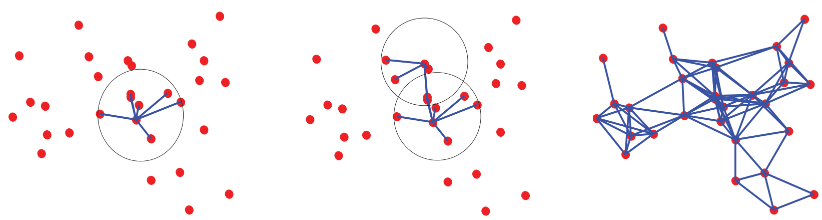


Figure 12. The construction of $\epsilon-$neighborhood recurrence networks.

However, the adjacency matrix only provides connectivity information for nodes and cannot determine the relevant spatial locations among nodes. From a recurrence adjacency matrix, we have a variety of possible topological structures for the network; also see Figure 13. A common question is "how to learn a unique and stable topological structure for a recurrence network?" To derive the geometric structure, we simulate the recurrence network as a physical system by treating edges as springs with the attractive force, $f_a(i,j) = |\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j|^2/K$, when nodes i and j are connected by an edge $i \leftrightarrow j$, and treating nodes as electrically-charged particles with the repulsive force $f_r(i,j) = -CK^3/|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j|^2, i \neq j$.

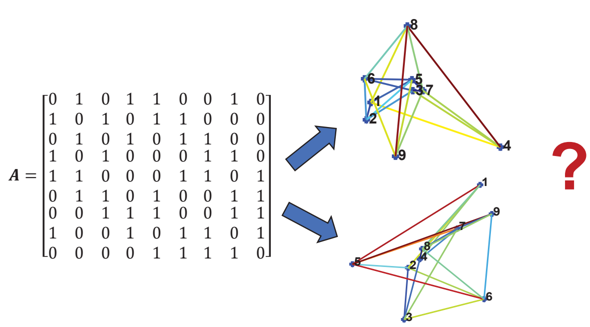


Figure 13. Two different topological structures from the same network adjacency matrix.

Therefore, the combined force on a node i from all other nodes is

$$f(i, \vec{\boldsymbol{x}}, K, C) = \sum_{i \neq j} -\frac{CK^3}{|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j|^3}(\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j) + \sum_{i \leftrightarrow j} \frac{|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j|}{K}(\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j) \tag{13}$$

where K is the natural spring length, C regulates the relative strength of repulsive and attractive forces, and $(\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j)$ is the force-directional vector, which is separated from $f_r(i,j)$ and $f_a(i,j)$ to define the direction of the combined force. The objective function is to minimize the network Hamiltonian (or energy): $\arg\min_x \sum_i f^2(i,x,K,C)$. See the details of optimization algorithms in [77]. Our prior work also carried out the experiments to investigate the impact of parameters K, C, and self-organized network topology of nonlinear dynamical systems such as Lorenz and Rossler systems [77]. As an example, Figure 14 illustrates the self-organization of a network with 292 nodes, which are first randomly distributed in space from two different initial conditions. Although the initial layouts are different, the network topology is self-organizing and converges to a unique and stable topological structure when the network Hamiltonian is iteratively minimized.

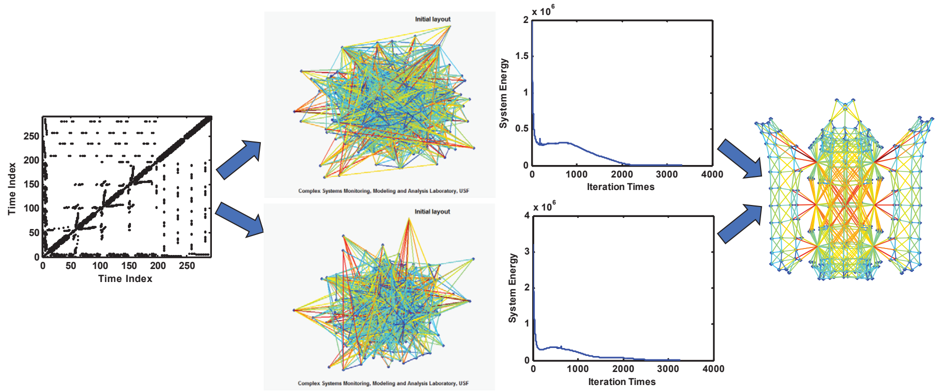


FIGURE 14. Self-organizing processes until convergence from two random initial conditions, i.e., layout A https://youtu.be/AF-WyYJeIbU vs. layout B https://youtu.be/utKWbof6yRw.

Further, our prior work developed self-organizing methods for variable clustering [16, 47, 48] that are aimed at detecting the subsets of homogeneous variables (i.e., time-series variables or predictor variables). In other words, variable clustering identifies the subgroups of variables that have stronger interrelations with each other. If two variables are weakly interrelated, they are categorized into different groups or clusters. It is important to note that variable clustering is different from data clustering. In the literature, "clustering" is generally referred to as "data clustering". Figure 15a illustrates that data clustering involves grouping or partitioning samples into clusters based on similarities in their attributes or characteristics. Each point represents a data sample with two coordinates, e.g., (0.36, 0.39) in the 2-dimensional space. To cluster data samples, a distance measure such as Euclidean distance is often used to determine the similarities and dissimilarities between samples. Data clustering is an unsupervised technique that involves partitioning data samples into clusters that are internally homogeneous and externally heterogeneous. However, Figure 15b illustrates the clustering of 20 variables. Each point in Figure 15b is a variable (e.g., $\mathbf{x}_{11}$ is a predictor variable with the dimensionality 1000×1) instead of a data sample (e.g., 2 dimensions in the example of Figure 15a). As shown in Figure 15c, for the table-form data, data clustering is more concerned about the samples in rows (i.e., $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$), but variable clustering focuses more on the variables in columns (i.e., $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$, where $K \ll N$).

Variable clustering takes into account the interdependence structure among variables, rather than just their individual attributes, when grouping them into clusters. Traditionally, such interrelationships are estimated with methods such as correlation, mutual information, and dynamic time warping. Yet, correlation is a second-order quantity evaluating merely linear dependency among data. Mutual information quantifies both linear and nonlinear dependency between variables but requires stationarity in the computation [16]. Dynamic time warping measures the pattern similarity between two univariate time series [36, 74].

Our prior work [53, 76] exploited cross recurrences between two nonlinear trajectories (i.e., $\overrightarrow{\boldsymbol{x}}$ and $\overrightarrow{\boldsymbol{y}}$) in the state space for nonlinear coupling analysis. See the relevant literature and software package for Cross Recurrence Quantification Analysis (CRQA) in https://www.mathworks.com/matlabcentral/fileexchange/

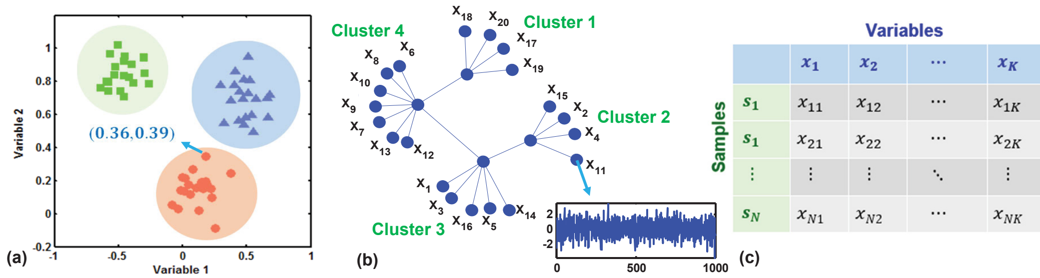


Figure 15. (a) Data clustering of samples; (b) variable clustering; (c) table-form data with variables arranged in columns and samples in rows.

100164-cross-recurrence-plot-and-quantification-analysis-crp-crqa. As such, there will be a $K \times K$ nonlinear coupling matrix for K geometric trajectories (i.e., constructed from multi-sensor signals or predictor variables such as $x_1, x_2, \ldots, x_K$ in Figure 15c). A new idea is to embed nonlinear trajectories as nodes in a complex network, and move the nodes with nonlinear-coupling forces to derive a self-organizing topology of the network. As such, the geometric map of networks pinpoints each symptom to root causes (or heart diseases) in subnetwork communities and also gradually learns the emerging causes. Note that this is different from the Kohonen map in neural networks, which learns self-organizing positions of neurons based on distance measures in the data space [13]. In contrast, our prior work [53, 76] seeks to self-organize the multi-lead sensor signals (e.g., high-dimensional nonlinear trajectories) for signal processing and fault identification. Furthermore, a "self-organizing network" provides an effective variable-grouping strategy that maximizes redundancy within a group and minimizes redundancy between groups, thereby greatly improving the performance of group variable selection [16, 48].

In addition, our prior work [78] developed a new self-organizing network representation of 3D objects (e.g., computer-aided designs (CAD) or 3D scanned objects), also called an "intra-design network" that is shown to effectively characterize both topological and geometrical information of 3D models in a sparse form. Industrial scanners can easily convert real-world 3D objects into digital CAD models, which can then be fabricated using optimized process plans for subtractive manufacturing or newer additive manufacturing (AM) technologies that build the object layer by layer. In order to create a searchable database for engineering designs that allows for effective matching and retrieval, there is a dire need to develop new analytical methods that can represent and reconstruct real-world 3D objects. Nonetheless, 3D models are often stored in different formats (or in different representations) on the internet. Thus, it is impractical to directly input heterogeneous formats of these models for retrieval, matching and search. This self-organizing "intra-design network" is shown to effectively unify the input format of each 3D design (or CAD model). As shown in Figure 16, when the self-organizing process converges and reaches a stable period, the resulting recurrence network closely resembles the geometric structure of a chair. See also the examples of a self-organizing process for a 3D drone model https://youtu.be/ac2bLZEmUco and a 3D ant model https://youtu.be/X1Knc70vbBQ.

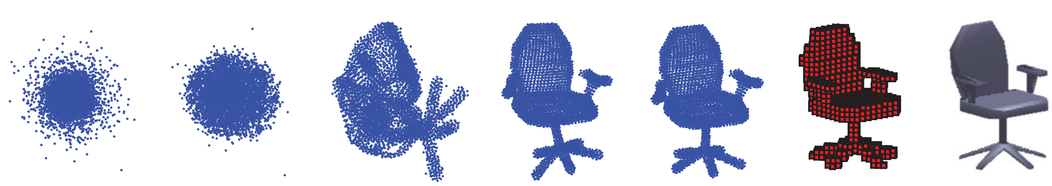


Figure 16. The self-organizing process of a recurrence network for the 3D design of a chair (YouTube video demo: https://youtu.be/FHI3eN5X6po).

6. Conclusions

Modern industries are investing in advanced sensing modalities such as wearable body area sensor networks, internet-of-things sensors, 3D(4D) imaging and scanning technologies. The emergence of extraordinary sensing capabilities has resulted in the accumulation of vast amounts of data, commonly referred to as “big data”. To fully exploit the potential of ”big data” for operations engineering, it is imperative to develop fundamentally new methodologies that can effectively harness and leverage the inherent complexity of high-dimensional sensor signals. Therefore, this tutorial presents a review of nonlinear dynamics methods and tools for real-time system informatics, monitoring and control.

Specifically, this tutorial consists of a series of research outcomes by students, collaborators and faculty members from the Complex Systems Lab at the Pennsylvania State University, which include: (1) A novel nonlinear recurrence methodology, which effectively characterizes and models multi-scale, heterogeneous recurrences in the phase-space domain of multi-sensor signals (rather than single-scale recurrences in time-domain signals); (2) A new sparse particle filtering method that compactly models nonlinear dynamic processes in a data-rich environment, and provides on-line quality monitoring and predictive control for process improvement (rather than off-line simulation and quality inspection); (3) An innovative dynamic network methodology to tackle the emerging challenges in real-time monitoring and control of time-varying 3D (4D) imaging profiles (rather than low-dimensional quality variables or profiles); and (4) A new bio-inspired self-organizing methodology for data visualization, topology learning and variable clustering (rather than only change detection). This tutorial reflects an effective integration of recurrence dynamics, network dynamics and self-organizing dynamics with operations engineering.

Our prior work and continued research aim to (1) advance the state of the art in modeling and control of complex systems by contributing new “nonlinear dynamics” concepts, methods and algorithms; (2) bridge the gap between cardiovascular research which is in dire need of enabling quality technologies for smart health, and quality engineering that tends to overlook electro-mechanical operations in the heart; and (3) enrich the theory of nonlinear dynamics and expand its research domain to interdisciplinary applications in both manufacturing and healthcare. In this tutorial, we contextualize the theory of nonlinear dynamics with prior work and real-world case studies on a wide variety of complex systems exhibiting nonlinear dynamics, e.g., heart, brain, sleep apnea, precision machining, multi-stage assembly line, biomanufacturing, and additive manufacturing.

Acknowledgements

The author gratefully acknowledges the generous support of NSF CAREER award CMMI-1617148. Also, the author would like to sincerely thank editors, anonymous reviewers, and colleagues who provided constructive comments to improve this work. In addition, the author thank Penn State and the Fulbright Scholar program for financial support when writing this tutorial during his sabbatical leave. Any opinions, findings, or conclusions found in this tutorial are those of the authors and do not necessarily reflect the views of sponsors.

References

[1] Paul S Addison. Wavelet transforms and the ECG: A Review. *Physiological Measurement*, 26(5):R155, 2005.

[2] Valtino X Afonso, Willis J Tompkins, Truong Q Nguyen, and Shen Luo. ECG beat detection using filter banks. *IEEE Transactions on Biomedical Engineering*, 46(2):192–202, 1999.

[3] Eduardo L Brugnago, Jason AC Gallas, and Marcus W Beims. Predicting regime changes and durations in Lorenz's atmospheric convection model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103109, 2020.

[4] Satish Bukkapatnam, Ranga Komanduri, Hui Yang, Prahalad Rao, Wen-Chen Lih, Milind Malshe, Lionel M Raff, Bruce Benjamin, and Mark Rockley. Classification of atrial fibrillation episodes from sparse electrocardiogram data. *Journal of Electrocardiology*, 41(4):292–299, 2008.

[5] Satish Bukkapatnam, Soundar Kumara, and Akhlesh Lakhtakia. Analysis of acoustic emission signals in machining. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 121(4):568–576, 1999.

[6] Satish Bukkapatnam, Soundar Kumara, and Akhlesh Lakhtakia. Real-time chatter control in machining using chaos theory. *CIRP Journal of Manufacturing Systems*, 29:321–326, 1999.

[7] Satish Bukkapatnam, Akhlesh Lakhtakia, and Soundar Kumara. Analysis of sensor signals shows turning on a lathe exhibits low-dimensional chaos. *Physical Review E*, 52(3):2375–2387, 1995.

[8] Charles D Callahan and David L Griffen. Advanced statistics: Applying statistical process control techniques to emergency medicine: A primer for providers. *Academic Emergency Medicine*, 10(8):883–890, 2003.

[9] Cheng-Bang Chen, Hui Yang, and Soundar Kumara. Recurrence network modeling and analysis of spatial data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(8):085714, 2018.

[10] Ruimin Chen, Farhad Imani, and Hui Yang. Heterogeneous recurrence analysis of disease-altered spatiotemporal patterns in multi-channel cardiac signals. *IEEE Journal of Biomedical and Health Informatics*, 24(6):1619–1631, 2019.

[11] Yun Chen, Gang Liu, and Hui Yang. Sparse particle filtering for modeling space–time dynamics in distributed sensor networks. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 626–631. IEEE, 2014.

[12] Yun Chen and Hui Yang. Multiscale recurrence analysis of long-term nonlinear and nonstationary time series. *Chaos, Solitons & Fractals*, 45(7):978–987, 2012.

[13] Yun Chen and Hui Yang. Self-organized neural network for the quality control of 12-lead ECG signals. *Physiological Measurement*, 33(9):1399, 2012.

[14] Yun Chen and Hui Yang. Sparse modeling and recursive prediction of space–time dynamics in stochastic sensor networks. *IEEE Transactions on Automation Science and Engineering*, 13(1):215–226, 2015.

[15] Yun Chen and Hui Yang. Heterogeneous recurrence representation and quantification of dynamic transitions in continuous nonlinear processes. *The European Physical Journal B*, 89:1–11, 2016.

[16] Yun Chen and Hui Yang. A novel information-theoretic approach for variable clustering and predictive modeling using Dirichlet process mixtures. *Scientific Reports*, 6(1):38913, 2016.

[17] Yun Chen and Hui Yang. Numerical simulation and pattern characterization of nonlinear spatiotemporal dynamics on fractal surfaces for the whole-heart modeling applications. *The European Physical Journal B*, 89:1–16, 2016.

[18] Changqing Cheng, Chen Kan, and Hui Yang. Heterogeneous recurrence analysis of heartbeat dynamics for the identification of sleep apnea events. *Computers in Biology and Medicine*, 75:10–18, 2016.

[19] Charles M Close, Dean K Frederick, and Jonathan C Newell. *Modeling and Analysis of Dynamic Systems*. John Wiley & Sons, 2001.

[20] Edward J. Davison. A method for simplifying linear dynamic systems. *IEEE Transactions on Automatic Control*, 11(1):93–101, 1966.

[21] Drew Dawson, Hui Yang, Milind Malshe, Satish TS Bukkapatnam, Bruce Benjamin, and Ranga Komanduri. Linear affine transformations between 3-lead (Frank XYZ leads) vectorcardiogram and 12-lead electrocardiogram signals. *Journal of Electrocardiology*, 42(6):622–630, 2009.

[22] Yu Ding, Elsayed A Elsayed, Soundar Kumara, J-C Lu, Feng Niu, and Jianjun Shi. Distributed sensing for quality and productivity improvements. *IEEE Transactions on Automation Science and Engineering*, 3(4):344–359, 2006.

[23] Yu Ding, Pansoo Kim, Dariusz Ceglarek, and Jionghua Jin. Optimal sensor distribution for variation diagnosis in multistation assembly processes. *IEEE Transactions on Robotics and Automation*, 19(4):543–556, 2003.

[24] Dongping Du, Hui Yang, Sarah A Norring, and Eric S Bennett. In-silico modeling of glycosylation modulation dynamics in hERG ion channels and cardiac electrical signals. *IEEE Journal of Biomedical and Health Informatics*, 18(1):205–214, 2013.

[25] Dale Dubin. *Rapid Interpretation of EKG's: An Interactive Course*. Cover Publishing Company, 2000.

[26] Nagi Gebraeel. Sensory-updated residual life distributions for components with exponential degradation patterns. *IEEE Transactions on Automation Science and Engineering*, 3(4):382–393, 2006.

[27] James Gleick. *Chaos: Making a New Science*. Penguin, 2008.

[28] Michael E. Goggin, Bala Sundaram, and Peter W. Milonni. Quantum logistic map. *Physical Review A*, 41(10):5705, 1990.

[29] Marek Grzelczak, Jan Vermant, Eric M Furst, and Luis M Liz-Marzán. Directed self-assembly of nanoparticles. *ACS Nano*, 4(7):3591–3605, 2010.

[30] Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *New England Journal of Medicine*, 321(6):406–412, 1989.

[31] Cardiac Arrhythmia Suppression Trial II Investigators*. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *New England Journal of Medicine*, 327(4):227–233, 1992.

[32] Jionghua Jin and Jianjun Shi. Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, 41(4):327–339, 1999.

[33] Jionghua Jin and Jianjun Shi. Press tonnage signal decomposition and validation analysis for transfer or progressive die processes. *J. Manuf. Sci. Eng.*, 127(1):231–235, 2005.

[34] Chen Kan, Ruimin Chen, and Hui Yang. Image-guided quality control of biomanufacturing process. *Procedia CIRP*, 65:168–174, 2017.

[35] Chen Kan, Changqing Cheng, and Hui Yang. Heterogeneous recurrence monitoring of dynamic transients in ultraprecision machining processes. *Journal of Manufacturing Systems*, 41:178–187, 2016.

[36] Chen Kan and Hui Yang. Dynamic spatiotemporal warping for the detection and location of myocardial infarctions. In *2012 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1046–1051. IEEE, 2012.

[37] Chen Kan and Hui Yang. Network models for monitoring high-dimensional image profiles. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1078–1083. IEEE, 2015.

[38] Chen Kan and Hui Yang. Dynamic network monitoring and control of in situ image profiles from ultraprecision machining and biomanufacturing processes. *Quality and Reliability Engineering International*, 33(8):2003–2022, 2017.

[39] Chen Kan, Kay-Pong Yip, and Hui Yang. Two-phase greedy pursuit algorithm for automatic detection and characterization of transient calcium signaling. *IEEE Journal of Biomedical and Health Informatics*, 19(2):687–697, 2014.

[40] Lan Kang and Susan L Albin. On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, 32(4):418–426, 2000.

[41] Holger Kantz and Thomas Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2004.

[42] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*. Cambridge university press, 1995.

[43] Keunpyo Kim, Mahmoud A Mahmoud, and William H Woodall. On the monitoring of linear profiles. *Journal of Quality Technology*, 35(3):317–328, 2003.

[44] Christopher KH Koh, Jianjun Shi, and William J Williams. Tonnage signature analysis using the Haar transform. *Transactions-North American Manufacturing Research Institution of SME*, pages 229–234, 1995.

[45] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1):21–28, 1995.

[46] Chia-Hung Lin, Yi-Chun Du, and Tainsong Chen. Adaptive wavelet network for multiple cardiac arrhythmias recognition. *Expert Systems with Applications*, 34(4):2601–2611, 2008.

[47] Gang Liu and Hui Yang. A self-organizing method for predictive modeling with highly-redundant variables. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1084–1089. IEEE, 2015.

[48] Gang Liu and Hui Yang. Self-organizing network for variable clustering. *Annals of Operations Research*, 263:119–140, 2018.

[49] Runsang Liu and Hui Yang. Multimodal probabilistic modeling of melt pool geometry variations in additive manufacturing. *Additive Manufacturing*, 61:103375, 2023.

[50] Klaus Mainzer. *Thinking in Complexity: The Computational Dynamics of Matter, Mind, and Mankind*. Springer, 2004.

[51] Benoit B Mandelbrot. Self-affine fractal sets, I: The basic fractal dimensions. In *Fractals in Physics*, pages 3–15. Elsevier, 1986.

[52] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.

[53] Adam Meyers, Mohammed Buqammaz, and Hui Yang. Cross-recurrence analysis for pattern matching of multidimensional physiological signals. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(12):123125, 2020.

[54] Adam Meyers and Hui Yang. Markov chains for fault-tolerance modeling of stochastic networks. *IEEE Transactions on Automation Science and Engineering*, 19(3):2591–2606, 2021.

[55] Kamran Paynabar and Jionghua Jin. Characterization of non-linear profiles variations using mixed-effect models and wavelets. *IISE Transactions*, 43(4):275–290, 2011.

[56] Bo-Wen Shen, Roger A Pielke Sr, Xubin Zeng, Jong-Jin Baik, Sara Faghih-Naini, Jialin Cui, and Robert Atlas. Is weather chaotic? Coexistence of chaos and order within a generalized Lorenz model. *Bulletin of the American Meteorological Society*, 102(1):E148–E158, 2021.

[57] Jianjun Shi. In-process quality improvement: Concepts, methodologies, and applications. *IISE Transactions*, 55(1):2–21, 2023.

[58] Steven H Strogatz. *Nonlinear Dynamics and Chaos With Student Solutions Manual: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC press, 2018.

[59] Eric Tatara and Ali Cinar. Interpreting ECG data by integrating statistical and artificial intelligence tools. *IEEE Engineering in Medicine and Biology Magazine*, 21(1):36–41, 2002.

[60] Nitish V Thakor and Y-S Zhu. Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection. *IEEE Transactions on Biomedical Engineering*, 38(8):785–794, 1991.

[61] Connie W Tsao, Aaron W Aday, Zaid I Almarzooq, Alvaro Alonso, Andrea Z Beaton, Marcio S Bittencourt, Amelia K Boehme, Alfred E Buxton, April P Carson, Yvonne Commodore-Mensah, et al. Heart disease and stroke statistics—2022 update: A report from the American Heart Association. *Circulation*, 145(8):e153–e639, 2022.

[62] Qiulian Wang and Hui Yang. Sensor-based recurrence analysis of energy efficiency in machining processes. *IEEE Access*, 8:18326–18336, 2020.

[63] Charles L Webber Jr and Joseph P Zbilut. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, 94(2005):26–94, 2005.

[64] H Yang, ST Bukkapatnam, and R Komanduri. Nonlinear adaptive wavelet analysis of electrocardiogram signals. *Physical Review E*, 76(2):026214, 2007.

[65] Hui Yang. Multiscale recurrence quantification analysis of spatial cardiac vectorcardiogram signals. *IEEE Transactions on Biomedical Engineering*, 58(2):339–347, 2010.

[66] Hui Yang and Satish TS Bukkapatnam. Recurrence based performance prediction and prognostics in complex manufacturing systems. In *Proceedings of IIE Annual Conference*, pages 1453–1458. Institute of Industrial and Systems Engineers (IISE), 2009.

[67] Hui Yang and Satish TS Bukkapatnam. Multiscale recurrence analysis of complex physiological rhythmic dynamics. In *Proceedings of IIE Annual Conference*, pages 1218–1224. Institute of Industrial and Systems Engineers (IISE), 2010.

[68] Hui Yang, Satish TS Bukkapatnam, and Leandro G Barajas. Local recurrence based performance prediction and prognostics in the nonlinear and nonstationary systems. *Pattern Recognition*, 44(8):1834–1840, 2011.

[69] Hui Yang, Satish TS Bukkapatnam, and Ranga Komanduri. Spatiotemporal representation of cardiac vectorcardiogram (VCG) signals. *Biomedical Engineering Online*, 11:1–15, 2012.

[70] Hui Yang, Cheng-Bang Chen, and Soundar Kumara. Heterogeneous recurrence analysis of spatial data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(1):013119, 2020.

[71] Hui Yang and Yun Chen. Heterogeneous recurrence monitoring and control of nonlinear stochastic processes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(1):013138, 2014.

[72] Hui Yang, Yun Chen, and Fabio Leonelli. Characterization and monitoring of nonlinear dynamics and chaos in complex physiological systems. *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement*, pages 59–93, 2016.

[73] Hui Yang, Yun Chen, and Fabio M Leonelli. Whole heart modeling—spatiotemporal dynamics of electrical wave conduction and propagation. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5575–5578. IEEE, 2016.

[74] Hui Yang, Chen Kan, Gang Liu, and Yun Chen. Spatiotemporal differentiation of myocardial infarctions. *IEEE Transactions on Automation Science and Engineering*, 10(4):938–947, 2013.

[75] Hui Yang, Soundar Kumara, Satish TS Bukkapatnam, and Fugee Tsung. The internet of things for smart manufacturing: A review. *IISE Transactions*, 51(11):1190–1216, 2019.

[76] Hui Yang and Fabio Leonelli. Self-organizing visualization and pattern matching of vectorcardiographic QRS waveforms. *Computers in Biology and Medicine*, 79:1–9, 2016.

[77] Hui Yang and Gang Liu. Self-organized topology of recurrence-based complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(4):043116, 2013.

[78] Hui Yang, Runsang Liu, and Soundar Kumara. Self-organizing network modelling of 3D objects. *CIRP Annals*, 69(1):409–412, 2020.

[79] Hui Yang, Milind Malshe, Satish TS Bukkapatnam, and Ranga Komanduri. Recurrence quantification analysis and principal components in the detection of myocardial infarction from vectorcardiogram signals. In *Proc. of the 3rd INFORMS Workshop on Data Mining and Health Informatics*, 2008.

[80] Hui Yang, Siqi Zhang, Yan Lu, Paul Witherell, and Soundar Kumara. Spatiotemporal monitoring of melt-pool variations in metal-based additive manufacturing. *IEEE Robotics and Automation Letters*, 7(3):8249–8256, 2022.

[81] Bing Yao, Farhad Imani, Aniket S Sakpal, Edward Reutzel, and Hui Yang. Multifractal analysis of image profiles for the characterization and detection of defects in additive manufacturing. *Journal of Manufacturing Science and Engineering*, 140(3), 2018.

[82] Joseph P Zbilut, Nitza Thomasson, and Charles L Webber. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Medical Engineering & Physics*, 24(1):53–60, 2002.

[83] Kang Zhao, Yao Xie, Kwok-Leung Tsui, Qingming Wei, Wenpo Huang, Wei Jiang, Yanting Li, Sugon Cho, Seoung Bum Kim, Kaibo Liu, et al. System informatics: From methodology to applications. *IEEE Intelligent Systems*, 30(6):12–29, 2015.

[84] Shiyu Zhou, Baocheng Sun, and Jianjun Shi. An SPC monitoring system for cycle-based waveform signals using Haar transform. *IEEE Transactions on Automation Science and Engineering*, 3(1):60–72, 2006.