Time Series Forecasting for Nonlinear and Nonstationary Processes: A Review and Comparative Study

7 authors, including:

Akkarapol Sa-Ngasoongsong
Oklahoma State University - Stillwater
13 PUBLICATIONS   **286** CITATIONS

Omer Faruk Beyca
Istanbul Technical University
21 PUBLICATIONS   **444** CITATIONS

Trung Le
North Dakota State University
34 PUBLICATIONS   **315** CITATIONS

Some of the authors of this publication are also working on these related projects:

Internet of Hearts View project

Multiscale modeling and optimization of cardiovascular systems View project

Time series forecasting for nonlinear and non-stationary processes: a review and comparative study

Changqing Cheng, Akkarapol Sa-Ngasoongsong, Omer Beyca, Trung Le, Hui Yang, Zhenyu (James) Kong & Satish T.S. Bukkapatnam

Time series forecasting for nonlinear and non-stationary processes: a review and comparative study

CHANGQING CHENG[1], AKKARAPOL SA-NGASOONGSONG[2], OMER BEYCA[3], TRUNG LE[4], HUI YANG[1], ZHENYU (JAMES) KONG[5] and SATISH T.S. BUKKAPATNAM[6*]

[1]*Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, FL 33620, USA*
[2]*School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK 74078, USA*
[3]*Industrial Engineering Department, Fatih University, Istanbul, Turkey 34500*
[4]*Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843, USA*
[5]*Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[6]*Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA*
E-mail: satish@tamu.edu

Forecasting the evolution of complex systems is noted as one of the 10 grand challenges of modern science. Time series data from complex systems capture the dynamic behaviors and causalities of the underlying processes and provide a tractable means to predict and monitor system state evolution. However, the nonlinear and non-stationary dynamics of the underlying processes pose a major challenge for accurate forecasting. For most real-world systems, the vector field of state dynamics is a nonlinear function of the state variables; i.e., the relationship connecting intrinsic state variables with their autoregressive terms and exogenous variables is nonlinear. Time series emerging from such complex systems exhibit aperiodic (chaotic) patterns even under steady state. Also, since real-world systems often evolve under transient conditions, the signals obtained therefrom tend to exhibit myriad forms of non-stationarity. Nonetheless, methods reported in the literature focus mostly on forecasting linear and stationary processes. This article presents a review of these advancements in nonlinear and non-stationary time series forecasting models and a comparison of their performances in certain real-world manufacturing and health informatics applications. Conventional approaches do not adequately capture the system evolution (from the standpoint of forecasting accuracy, computational effort, and sensitivity to quantity and quality of *a priori* information) in these applications.

Keywords: Nonlinear and non-stationary processes, complex systems, time series, forecasting

1. Introduction

Recent advancements in sensor, computing, and communication technologies and the consequent availability of abundant data sources in the form of time series can transform the way real-world complex systems are monitored and controlled. For instance, the production processes in semiconductor, automotive, and aerospace manufacturing industries are moving toward expensive, large part sizes (e.g., > 300 mm wafer size, integrally machined, as opposed to welded aerospace and automotive panels) with stringent quality requirements (e.g., defect-free sub-20-nm feature sizes, with surface finish $Ra \sim 1$ nm). Consequently, the cost of quality loss is becoming increasingly prohibitive,

and industries are beginning to upgrade the design and control of their manufacturing systems to *prevent*, as opposed to *mitigate*, quality and integrity losses (see Fig. 1). Similarly, with the nation's annual healthcare costs exceeding $2 trillion, the U.S. health industry is moving away from reactive disease control in favor of *preventative medicine* (e.g., P4—Predictive, Preventive, Personalized, and Participatory Medicine paradigm; Tian *et al.* (2012)). Such a shift from *detection–diagnosis–mitigation* of anomalies to *prediction–prognosis–prevention* is seen across various engineering domains beyond manufacturing and healthcare, including telecommunication and utility networks, infrastructure, and lifeline systems. In particular, the rich causal and dynamic information discernible from time series data has made forecasting of the evolution of complex biological, physical, and engineering system dynamics (Lang *et al.*, 2007; Duy and Peters, 2008; Wang *et al.*, 2008)—crucial for their preventative control—possible, and
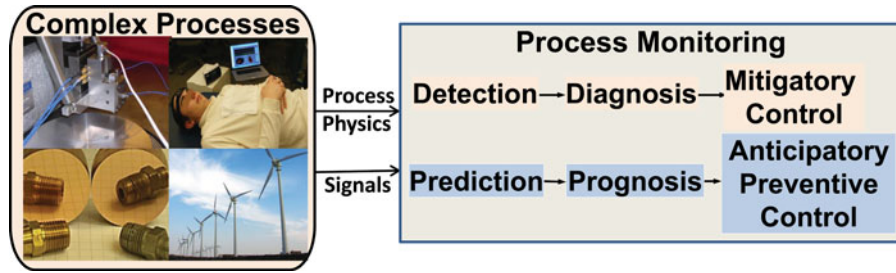
Fig. 1. Time series forecasting application for complex nonlinear processes.

predictive analytics is emerging as one of the most important branches of modern "big data" analytics and informatics. In fact, time series forecasting of complex systems evolution is considered one of the emerging challenges of modern science (Kumar *et al.*, 2003).

Effective prediction of future states of a complex system from time series remains a challenge, mainly due to diverse combinations of the nonlinear and non-stationary dynamic behaviors exhibited by these systems. Statistically speaking, time series may be considered as emerging from a nonlinear dynamic system if it is marked by characteristics such as non-normality, aperiodicity, asymmetric cycles, multi-modality, nonlinear causal relationships among the lagged variables, variation of forecasting performance over the state space, and time irreversibility, among others (Fan and Yao, 2003). For instance, the time series $y(t)$ shown in Fig. 2(a) exhibits a stationary, aperiodic pattern. Figure 2(b) shows a two-dimensional (2D) projection of the state space $x(t)$ reconstructed from the delays (time shifts) of $y(t)$ (Marwan *et al.*, 2007), explained in Section 2. The aperiodic pattern and stationary (time-invariant) non-ellipsoidal distribution of points in the state space are typical characteristics of a nonlinear time series. Non-stationary behavior refers to the time-varying nature of the underlying distributions and is marked by variation in the first, second, and/or higher moments of the underlying stochastic processes or transient regime-change behavior of the system over time. Prominent examples of such nonlinear and non-stationary time series include acoustic emission from manufacturing machinery and structural systems, customer demand and production rates in a manufacturing enterprise, electricity load in a smart grid, traffic at a node (in-

tersection) across transportation and telecommunication networks, stock prices and other market indices, EEG and ECG signals, and temperature and other weather signals (Bukkapatnam and Cheng, 2010; Yang *et al.*, 2011; An *et al.*, 2012; Sa-ngasoongsong, 2012; Le *et al.*, 2013).

In addition to the complexity of the dynamics, the structure of the nonlinear relationships among the measured signals and states remains unknown if not indeterminable in most cases. It is also customary to assume that a linear or a simple nonlinear relationship exists between past and future observations. However, this assumption does not hold for most real-world time series, which clearly indicate nonlinear patterns and can also be classified as non-stationary. Consequently, classic time series models, such as autoregressive and moving average with or without co-integrated terms and external inputs (ARMA, ARMAX) tend to have serious shortcomings for forecasting the evolution of complex systems.

In the past 10–15 years, several advancements have taken place toward capturing the *transient* events and *nonlinear* structure of the underlying systems from time series data. A recent review of time series forecasting models (De Gooijer and Hyndman, 2006) provides extensive coverage of the classic forecasting models (e.g., exponential smoothing) and some nonlinear models, such as regime-switching (Tong, 1990) and neural network models. Nevertheless, a comprehensive review of recent developments in nonlinear and non-stationary complex system forecasting models is still lacking. The primary purpose of this article is to summarize the state-of-the-art of forecasting models for complex systems and present a comparative study of the alternative models, including the contributions from our research group in local recurrence models (Yang *et al.*, 2011), local Markov model (Bukkapatnam *et al.*, 2009), local Gaussian process (Bukkapatnam and Cheng, 2010), Dirichlet process-based mixture Gaussian process models (Le *et al.*, 2013), recurrent neural network-based particle filtering (Rao *et al.*, 2013), and empirical mode decomposition (Sa-ngasoongsong and Bukkapatnam, 2011) models for forecasting of time series in real-world complex systems. The remainder of this article is organized as follows: Section 2 presents an overview of nonlinearity and non-stationary behavior in time series; a concise classification and review of nonlinear and non-stationary
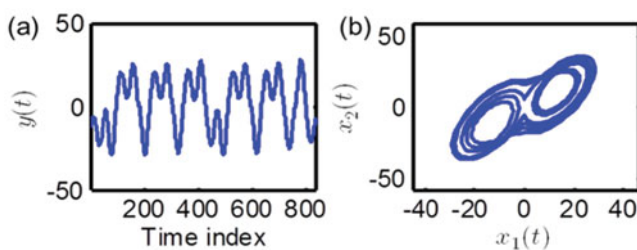

Fig. 2. An illustration of (a) a nonlinear and non-stationary time series with (b) its portrait in a 2D projection of its state space.

time series forecasting models is provided in Section 3; Section 4 contains case studies for state and performance forecasting in real-world nonlinear and non-stationary systems; and conclusions are presented in Section 5.

2. Overview of nonlinear and nonstationary dynamics of complex system time series

This section presents a concise exposition of nonlinear dynamics and non-stationary process principles, purely toward providing the relevant background on nonlinear and non-stationary time series. For further details, the readers may consult Kantz and Schreiber (1997), Casdagli (1989), and Fan and Yao (2003). The nonlinear dynamic process is governed by a nonlinear stochastic differential (or difference) equation of the following form:

$$dx/dt = F(x, \theta, \varepsilon), \qquad (1)$$

which specifies the evolution of process states $x(t)$ from an initial condition $x(0)$, where θ is the process parameter vector, and ε is the system noise. The solution of (1)

is given by $x(t) = \phi(x(0), t)$. Here, $\phi(\cdot, \cdot)$ is called the flow, or the state transition function. The measured signals $y(t)$ from these processes are a function of the state $x(t)$; i.e., $y = h(x)$. Many real-world systems exhibit nonlinear stochastic dynamics, and the measured signals from such systems are referred to as nonlinear time series, which exhibit non-Gaussian, multi-modality, time irreversibility, and other properties (Fan and Yao, 2003).

From an initial condition $x(0)$, the system state evolution $x(t)$ oftentimes settles down asymptotically to executing steady-state (stationary) dynamics, after prosecuting transients. The set of points constituting an invariant set in the state space that describes this (stationary) steady-state behavior is called an attractor **A**. Although the concept of an attractor is well understood in the context of deterministic nonlinear systems, characterization of attractors in stochastic systems remains a subject of active research (Craul *et al.*, 1997). The attractor **A** of a deterministic system can be (i) static (no drift); (ii) periodic; (iii) quasi-periodic; or (iv) chaotic (see Fig. 3). For a static attractor, the steady state is simply a fixed point, and the system is at an equilibrium state. An attractor is periodic
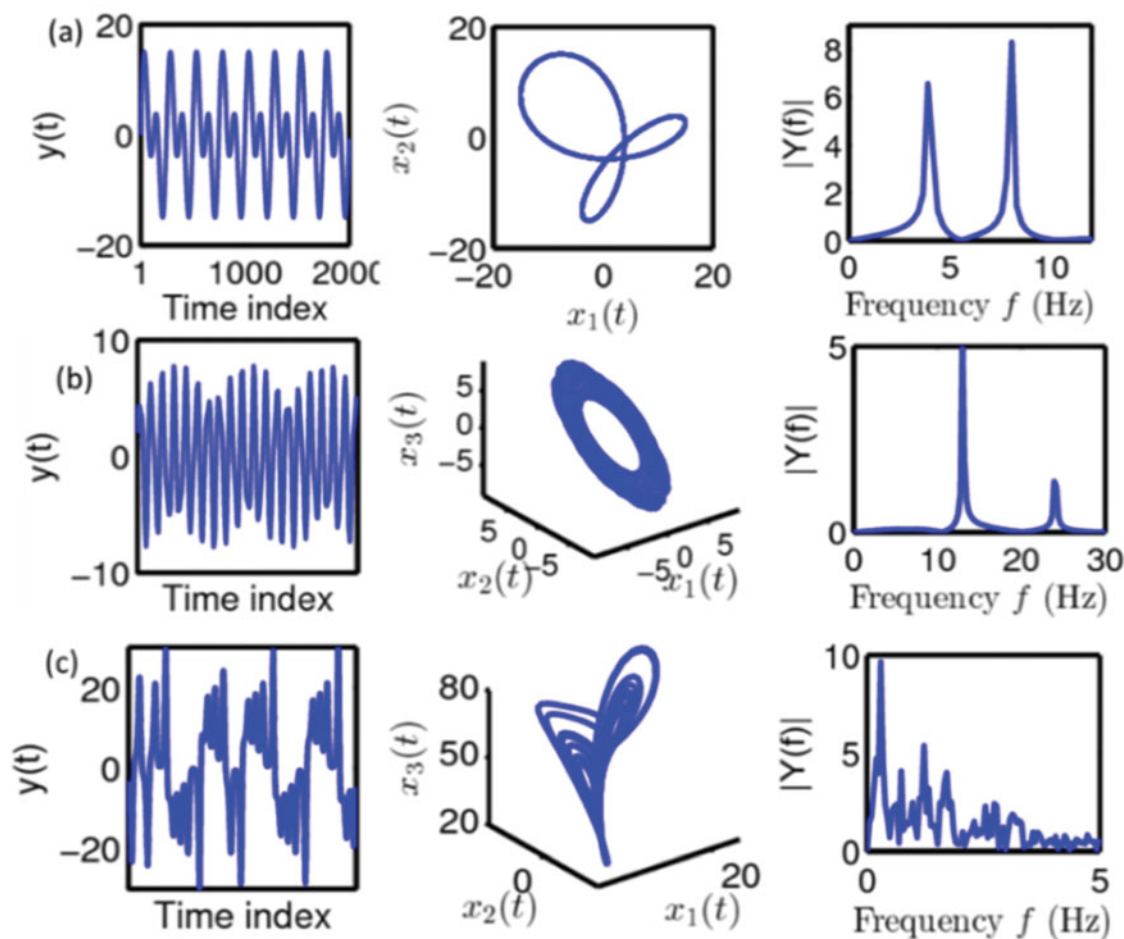

Fig. 3. Time series, reconstructed state portrait, and the frequency spectrum of (a) periodic, (b) quasi-periodic, and (c) chaotic processes.

(also referred to as a limit cycle) if the system trajectories evolve into a closed cyclic path in the state space (Kantz and Schreiber, 1997). For example, Fig. 3(a) shows the time, state, and frequency portraits of a periodic signal composed of two frequencies, namely, 4 and 8 Hz. The state portrait shows a closed path along which the system evolves under steady state. A quasi-periodic near-toroidal attractor will be formed if the frequencies of the system dynamics $\omega_1, \omega_2, \ldots, \omega_n$ are rationally independent (e.g., the ratio of at least two ω_i s is an irrational number); i.e., for all $k \in Z^n \backslash 0$ one has $\sum_{i=1}^{n} \omega_i k_i \neq 0$ (Broer *et al.*, 1996). For example, Fig. 3(b) shows the portraits of a quasi-periodic signal that consists of two frequencies, 13 and $17\sqrt{2}$ Hz. The figure suggests that although the signal is composed of just two frequency components, the time portrait exhibits a somewhat irregular pattern, and the evolution in the state space does not constitute a closed loop. A chaotic process displays irregular ("noisy") oscillations with a continuous band of frequency components (Kantz and Schreiber, 1997), as indicated in the frequency plot in Fig. 3(c). The chaotic attractor is aperiodic, as the system dynamics never repeats. The evolution of the chaotic process is so irregular that the system response is highly sensitive to initial conditions. Initially, small deviations $\Delta x(0)$ between two system trajectories under steady state grow exponentially over time (although the state of the system remains on the attractor), impairing long-term predictability; i.e., the deviation at time t may be expressed as $\Delta x(t) \approx e^{\lambda t} \Delta x(0)$, where λ, the largest Lyapunov component, limits the av-

erage exponential rate of separation. Also, $\lambda > 0$ indicates that the system is chaotic. Time series from chaotic systems are known to present significant challenge for forecasting even under steady state (Muckerjee *et al.*, 1997; Menezes and Barreto, 2008; Hamid and Noorani, 2013) due to the sensitive dependence on initial conditions. Techniques collectively termed nonlinear time series analysis (Kantz and Schreiber, 1997) are employed toward characterizing the local dynamics underlying the measured time series to facilitate forecasting of chaotic time series (Casdagli, 1989).

Whenever the governing equations of complex system dynamics are unknown or indeterminable, the state space (for both deterministic as well as stochastic processes) may be reconstructed from the delay coordinates of the measured time series $y(t)$ (Kantz and Schreiber, 1997) as

$$
\begin{aligned}
x_i &= x(t_i) \\
&= [y(t_i), y(t_i + \tau), y(t_i + 2\tau), \ldots, y(t_i + (m-1)\tau)],
\end{aligned}
\tag{2}
$$

where m is the embedding dimension and τ is the delay time. The minimum sufficient embedding dimension m is usually estimated using a false nearest neighbor method (Takens, 1981; Kennel *et al.*, 1992) and the optimal τ by minimizing the mutual information function (Fraser and Swinney, 1986). See Fig. 4(a) and (b) for the reconstruction of Lorenz state space from time series. It may be noted that the reconstructed state portrait holds strong similarity to (strictly speaking, diffeomorphic [in the sense that there is
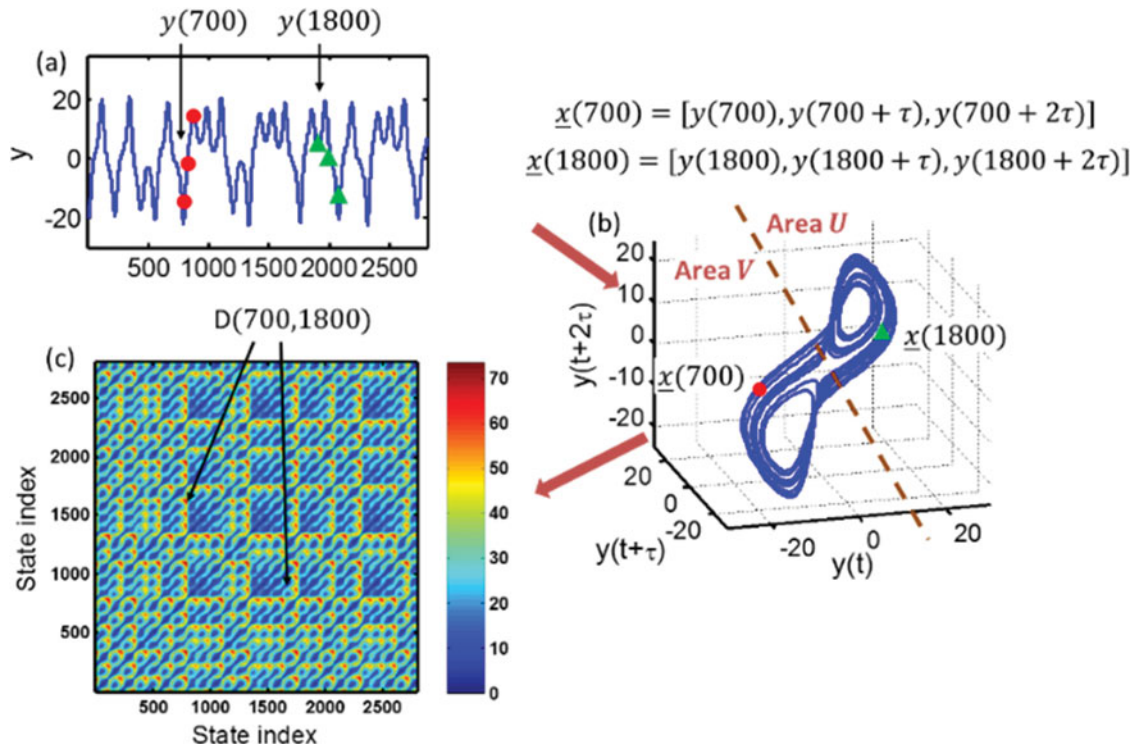

Fig. 4. Graphical illustration of the process of reconstructing an attractor and a recurrence plot from a time series (Yang *et al.*, 2011).

a smooth function that maps all points in the underlying state space to reconstructed state space, and *vice versa*] to the observable subset of) the original state space (Kantz and Schreiber, 1997).

Characterization of attractor topology can be helpful in investigating system dynamics and predicting its evolution. Recurrence property is one of the most important characteristics of Attractor **A**. Poincare recurrence theorem (Katok and Hasselblatt, 1995) suggests that the trajectories of nonlinear dynamic systems eventually return to the neighborhood of the state space of every initial condition after a finite time; i.e., for every $\varepsilon > 0$, and almost every $x(0) \in A$, $\exists t > 0$, such that $|x(0) - x(t)| \leq \varepsilon$, where ε is a distance measure (e.g., Euclidean norm). In effect, all trajectories within an attractor remain bounded (Kantz and Schreiber, 1997). Recurrence analysis invokes a recurrence plot (Marwan *et al.*, 2007) to visualize the recurrence property of the state space. Recurrence plots delineate the distance of every point x_i, the state vector realized at time t_i, to all the others in the state space; i.e., $D(i, j) = C(x_i - x_j)$, where $C(\cdot)$ is the color code that maps the distance to a color scale (see Fig. 4(c)). Furthermore, the patterns discerned from a recurrence plot can be used to quantify the underlying nonlinear dynamics and non-stationary behaviors (Eckmann *et al.*, 1987; Casdagli, 1997). For instance, the recurrence plot can be used to estimate the boundary between the two local near-stationary dynamic behaviors in a state space (e.g., areas U and V shown in Figs. 4(b) and 4(c)). Most real-world nonlinear dynamic systems operate under transient (i.e., non-stationary) conditions (Schreiber, 1999). From the statistical perspective, the stationarity of time series $y(t)$ requires the joint distribution of every collection $[y(t + \tau_1), y(t + \tau_2), \ldots, y(t + \tau_k)]$ to be invariant to τ_i ($i = 1, 2, \ldots, k$) for any k. Even under non-stationary conditions, complex system dynamics may be treated as a concatenation of much simpler piecewise transient or near-stationary behaviors, each reflecting a finite time sojourn of system dynamics in the vicinity of an attractor of

a nonlinear dynamic process (Amaral *et al.*, 2006; Bukkapatnam and Cheng, 2010). Figure 5 shows a summary of common non-stationary behaviors. Most commonly, non-stationary behavior is attributed to specific deterministic and stochastic trends in the moments (Kirchgassner *et al.*, 2013). For example, a first-order non-stationary time series may be expressed as $y(t) = \mu_t + w(t)$, where μ_t is a deterministic function of time, and $w(t) \sim N\left(\mu_w, \sigma_w^2\right)$ is a stationary disturbance process. The stationarity is violated, as $E[y(t)] = \mu_t$ depends on time t. According to the explicit form of the drift μ_t, the time series may exhibit growth (upward trend) and/or decay (downward trend). If a repetitive up-and-down trend is present, it is called a cycle. When a cyclic trend occurs at fixed and known periods, it is called seasonality. Additionally, processes embodying a stochastic trend such as the first-order random walk $y(t) = \alpha + y(t - 1) + w(t)$ have time-varying conditional means and variances with stationary increments ($\alpha \neq 0$). Autoregressive integrated moving average (ARIMA; Fan and Yao (2003)) models are often used to represent these patterns. The evolution of asset prices and other derivatives of financial markets is often treated as a process with time-varying and auto-correlated σ_w^2, also known as varying volatility or heteroskedastic processes (Engel, 1982). Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models are often employed to capture these behaviors (Marchesi and Lux, 2000). Additional second-order non-stationary stochastic processes include *periodic time series* where there is a $T > 0$ such that the expected values and covariance satisfy the following conditions: $\mu_t = \mu_{t+T}$ for all t and $\mathrm{cov}\,(y_{t_1}, y_{t_2}) = \mathrm{cov}\,(y_{t_1+T}, y_{t_2}) = \mathrm{cov}\,(y_{t_1}, y_{t_2+T})$ for all t_1, t_2 almost surely, and *cyclo-stationary time series* where there is a $T > 0$ such that $\mu_t = \mu_{t+T}$ for all t and $\mathrm{cov}\,(y_{t_1}, y_{t_2}) = \mathrm{cov}\,(y_{t_1+T}, y_{t_2+T})$ for all t_1, t_2 almost surely (Stark and Woods, 2011). Cyclo-stationary time series analysis is often seen in the modeling of weather parameters, such as temperatures that can be highly non-stationary during a 12-month period in a
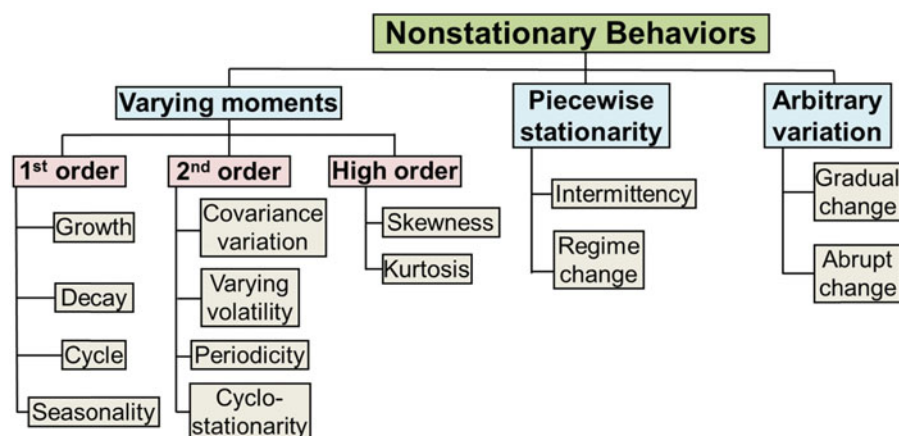

Fig. 5. Classification of salient non-stationary behaviors.

certain region but variation in each winter for the past years can be regarded as stationary (Koutsoyiannis and Montanari, 2007). Apart from these, forecasting of time series with varying higher-order moments (e.g., kurtosis and skewness) has also been investigated (Genton and Thompson, 2004).

Time series from a plethora of real-world systems, including human physiological processes, financial and social networks, and various natural phenomena can be perceived as piecewise stationary; i.e., weak stationarity rules are largely satisfied in some local regimes although the time series as a whole exhibit complex behavioral patterns. Most financial time series are non-stationary, in that they can be modeled by a multi-regime stochastic process, and the time series throughout each regime are generated by a stationary process. The transition between those regimes is instantaneous. In contrast, irregular, yet a more gradual alternation of different regimes connotes a behavior known as dynamic *intermittency*. Intermittency is marked by random switching of system trajectories between relatively regular laminar phase(s) and irregular turbulent phase(s) (Heagy *et al.*, 1994; Wang *et al.*, 2014). Figure 6 shows a representative time series from a dynamic system with a simple on–off intermittency. Here, the dynamic system stays in the vicinity of an attractor of a state for long periods of time ("on"), and then transitions out of the attractor occasionally ("off"; Hamilton (1989); Battaglia and Protopapas (2011)).

Additionally, a highly complex non-stationary time series, whose patterns cannot be simply represented in terms of varying moments or piecewise stationarity, can be classified to exhibit arbitrary variation. Here, the relationships among the intrinsic and exogenous variables (and their autoregressive terms) can change either abruptly or gradually over time due to the stationary or non-stationary changes in the parameters θ (Last and Shumway, 2008; Rosen *et al.*, 2012). A non-stationary evolution of the parameters can follow any of the aforementioned behaviors. For example, Tay and Cao (2002) developed a modified Support Vector Machine (SVM) model for non-stationary financial time series forecasting by giving higher weight to recently realized data compared with distant ones. Their approach is based on the premise that the dependency between input and output variables gradually change over time.
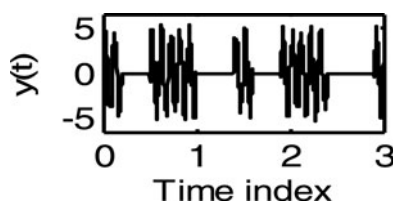
3. Classification and review of nonlinear and non-stationary time series forecasting models

Although real-world systems exhibit mostly nonlinear and non-stationary behaviors, the majority of forecasting methods reported in the literature assume linearity and/or stationarity of the underlying dynamics (De Gooijerand Hyndman, 2006) or consider simple forms of non-stationarity, such as well-defined trends and variations in the first two moments, simple forms of piecewise stationary regimes (Battaglia and Protopapas, 2011). Although significant advancements have been reported for forecasting such complex time series, as detailed in the remainder of this section, accurate forecasting of the evolution of time series generated from real-world complex systems that exhibit nonlinear and non-stationary dynamics remains a challenge.

In this section, various methods reported in the literature are reviewed and categorized based on how they have been applied for forecasting real-world time series data (see Fig. 7). Particular emphasis has been placed on nonlinear and non-stationary time series forecasting in systems and processes of interest to IE researchers. Broadly speaking, these forecasting methods may be classified based on the premises or the approaches to treating non-stationarity under nonlinear conditions, in that they assume (i) a known form of the trend in the first few moments; (ii) piecewise stationarity of the signals (mostly, models developed in our research group); (iii) progressively varying parameters (hidden Markov model-based models); or (iv) decomposability of the signal into stationary segments in a transformed domain, and they are either parametric or nonparametric depending on whether the predictor takes a predetermined form or is constructed purely according to the data (e.g., the number of latent variables is allowed to vary).

3.1. *Parametric models*

A parametric forecasting model (Fan and Yao, 2003) specifies an explicit function form with a finite number of pa-
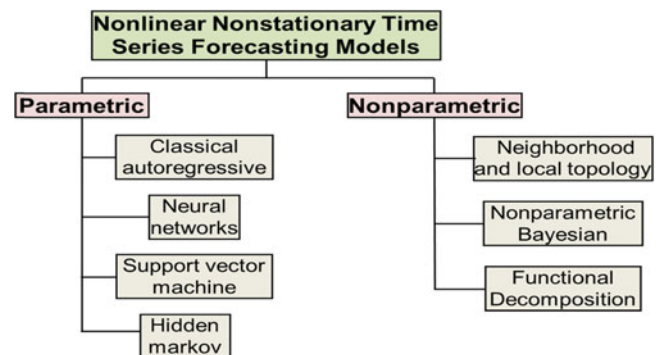

Fig. 6. On–off intermittent time series from a nonlinear dynamical process.

Fig. 7. Classification of nonlinear and non-stationary time series forecasting models.

rameters θ to describe the relationship between the input, consisting of the intrinsic and exogenous variables and their autoregressive (lag) terms, and output, consisting of the future values of the intrinsic variable $y(t+1)$. The model parameters are estimated from the time series realizations.

3.1.1. *Classic autoregressive models*

AutoRegressive (AR) models are most widely studied because of their flexibility to model many stationary processes. These include the well-known ARMA model (Fan and Yao, 2003), which assumes a linear relationship between the lagged variables as $y_{t+1} = \sum_{i=0}^{p-1} \phi_i y_{t-i} + \sum_{i=0}^{q-1} \theta_i \varepsilon_{t-i} + \varepsilon_t$. Weron and Misiorek (2008) reviewed AR models and their extensions for short-term time series forecasting. However, these linear models are only a coarse approximation to real-world complex systems and generally fail to accurately predict the evolution of nonlinear and non-stationary processes. ARMA model performance frequently degrades considerably whenever time trends and seasonality features are present. Methods such as ARIMA (Fan and Yao, 2003), which are based on the evolution of the increment $\Delta y_t = y_{t+1} - y_t$ or $\Delta_{y_t}^2$, are used at times to remove/reduce first-order (moment) non-stationarity. However, differencing generally amplifies high-frequency noise in the time series, and great effort is required to determine the order of an ARIMA model. Also, ARIMA models are largely limited to capturing the first-order non-stationarity. Engle (1982) introduced the ARCH model to capture the second-order (moment) non-stationarity; i.e., time-varying conditional variance or volatility. The GARCH model (Bollerslev, 1986), developed later, represents the variance of the error term as a function of its AR terms, thereby allowing a more parsimonious representation.

Toward incorporating nonlinearity as part of an ARMA structure, Tong (1990) proposed Threshold AR models (TAR) that assume piecewise linearity (i.e., different linear forms in different regions of the input space). Non-overlapped partition of the input space can be specified in terms of a threshold variable. TAR models have been successfully applied for time series forecasting in economics and neuroscience among other fields (Yadav *et al.*, 1994). Also, the threshold itself is assumed to be time-varying and specified as an AR process in Self-Excited Threshold AR (SETAR) models (Tiao and Tsay, 1994; Goldberger, 1999). These models can be regarded as an extension of AR models that allow a higher degree of flexibility in model parameters through a regime-switching behavior. Smooth Transition AR (STAR) models (Terasvirta, 1994; Schreiber, 1999) were introduced to capture smooth transitions between regimes, as opposed to the abrupt switch used in SETAR models. McAleer and Medeiros (2008) investigated heterogeneous AR models with multi-regime smooth transition for financial time series forecasting. This model can approximate volatility and long-term memory behavior as

well as sign and size asymmetries inherent to financial time series. Later, TAR models that switch both according to time and lagged-value state space were designed. Lundbergh *et al.* (2003) proposed a model to consider simultaneously regime alternating and time-varying coefficients; i.e., AR coefficients change both according to time and to an input variable that switches smoothly between two regimes. However, in practice, many parameters have to be optimized to build a consistent time-varying STAR model, and no efficient analytical optimization is available.

Recently, Battaglia and Protopapas (2011) investigated time-varying multi-regime threshold models to consider regime transitions according to both time and time series values. They employed a genetic algorithm to optimize parameters, such as threshold and orders. Hamilton (1989) used a discrete state Markov process to represent the switch between different regimes of the threshold models to tackle highly nonlinear problems. Lineesh and John (2010) studied trend and threshold AR models for nonlinear and non-stationary time series. They employed wavelet bases to decompose time series into orthogonal trend series and detail series, and then used ARMA and TAR models to forecast each decomposed series, respectively. Krishnamurthy and Yin (2002) combined a hidden Markov model and AR models under a Markov regime, where AR parameters switch in time according to the realization of a finite-state Markov chain, for nonlinear time series forecasting. However, most of these methods tend to be limited for nonlinear and stationary time series forecasting by the local linearity assumption implicit with an AR-type structure. Multivariate time series–based classic regression forecasting models, such as Vector AR (VAR), have also been studied to consider the effects of exogenous variables. However, the non-stationary evolution of such variables, which can deviate permanently from previous states, presents additional challenges to the long-term forecasting of the target variables. A vector Error Correction Model (ECM) can be useful in capturing such behaviors, especially when exogenous variables are co-integrated; i.e., their linear combination is stationary. ECMs are a category of multivariate time series models that directly estimate the speed at which the target variable returns to equilibrium after a change in an exogenous variable and can be used to estimate both short-term and long-term effects of one time series on another. For example, Sa-ngasoongsong *et al.* (2012) combined VAR and ECM to forecast the long-term variation of automobile sales, by estimating the co-integration vectors of the intrinsic variables.

3.1.2. *Neural networks and neuro-fuzzy-type models*

Neural Networks (NNs) and other nonlinear operator-based models have been used for nonlinear time series forecasting in manufacturing systems, finance, health informatics, and energy grids (Lowe and Webb, 1991; Kohzadi *et al.*, 1996; Hippert *et al.*, 2001), among other fields. These models do not need prior assumptions on the form

of nonlinearity and are universal approximators (Park and Sandberg, 1991); i.e., they can approximate any continuous function to an arbitrary precision. A recent review of NN models for time series forecasting has been provided by Zhang (2012).

Feed-forward Neural Network models (FNNs) parameterized with a back-propagation algorithm have been employed for nonlinear time series forecasting (Lapedes and Farber, 1987; French *et al.*, 1992). They are known to outperform traditional statistical methods such as regression and Box–Jenkins approaches in functional approximation, but they assume the dynamics underlying time series are time-invariant. FNNs with recurrent feedback connections have also been attempted for time series forecasting (De Groot and Wuertz, 1991). Such dynamic Recurrent NN (RNN) models allow forecasting of nonlinear time series occurring in various fields (Grudnitski and Osburn, 1993; Kuan and Liu, 1995). Rao *et al.* (203) studied Recurrent Predictor NN (RPNN) model for one-step-ahead prediction of the nonlinear signal patterns during ultra-precision machining processes and combined particle filtering models for detection of changes in NN weights. Menezes and Barreto (2008) built a recurrent network structure of nonlinear AR models with exogenous input for multi-step forecasting of chaotic time series. Various types of Radial Basis Function (RBF) NN models, such as those employing dynamic regularization (Yee and Haykin, 1999), orthogonal least squares learning (Chen *et al.*, 1991), and recursion (Chen *et al.*, 1992) have been investigated to capture different forms of trends and volatility in the time series.

Barreto *et al.* (2007) reviewed time series forecasting approaches using Self-Organizing Map (SOM) NN models. The local approximation property inherent in these models can improve the forecasting accuracy of nonlinear time series compared to global models such as FNN. Additionally, they obviate the need to specify the number of neurons in advance by allowing the network architecture to grow based on the data.

Ensemble (Zhang and Berardi, 2001; Lai *et al.*, 2006) or hybrid NN models (Kodogiannis and Lolis, 2002; Zhou *et al.*, 2004), such as wavelet NN models, have also been attempted for nonlinear time series forecasting. For example, Lai *et al.* (2006) studied ensemble nonlinear NN models for financial time series forecasting. NN models with different initial weights were generated first, and then Principal Component Analysis was adopted to select appropriate NN ensembles. A wavelet NN combines the functional approximation superiority of FNNs with the strength of wavelet analysis to express non-stationary behaviors of time series in terms of both time and frequency characteristics (Zhou *et al.*, 2004). Here, the signal components from each scale of wavelet decomposition are modeled using an FNN. This simplifies the learning of the new time series, especially those with slow frequencies, and is also useful in separating noise from relevant information (Soltani, 2002). Zhang (2003) used an FNN to model the residual from an ARIMA

model for nonlinear time series forecasting. His investigations suggested that such an approach can reduce model uncertainty and ease over-fitting problems in NN models. There has also been a growing interest in neuro-fuzzy models, where the best kernels for capturing linguistic variables are learned from data, for time series forecasting applications (Kodogiannis and Lolis, 2002; Nayak *et al.*, 2004; Aznarte *et al.*, 2007; Khashei *et al.*, 2008). Neuro-fuzzy models can handle large nonlinear time series where the underlying physical relationships are unknown (Nayak *et al.*, 2004).

3.1.3. *SVM models*

SVM-based forecasting methods use a class of generalized regression models, such as Support Vector Regression (SVR) and Least-Squares Support Vector Machines (LS-SVMs; Smola and Scholkopf (2004)), that are parameterized using convex quadratic programming methods (Balabin and Lomakina, 2011). An SVM maps the inputs x_i, which may consist of AR terms of intrinsic and exogenous variables, into a higher-dimensional feature space $\varphi(x_i)$. We need not compute the transform $\varphi(x_i)$ explicitly; instead, we only need to estimate the inner product of the mapped patterns $k(x_i, x_j) = |\varphi(x_i), \varphi(x_j)|$, where $|\langle\rangle$ denotes the inner product. The inner product is expressed as a linear combination of specified kernel functions, based on which SVMs are categorized into linear, Gaussian or RBF, polynomial, and multilayer perceptron classifiers. A linear regressor is then constructed by minimizing the structural risk minimization (the upper bound of the generalization error), leading to better forecasting performance than conventional techniques (Cao, 2003).

Mukherjee *et al.* (1977) investigated the application of SVMs for chaotic time series forecasting. They showed that SVMs have higher forecasting accuracy than NN models and employ fewer parameters. Lau and Wu (2008) reviewed Least Squares- (LS) and RBF-based predictors and designed a local SVM (defined in the reconstructed state space) for chaotic time series forecasting. Their investigations suggested that local SVM models can provide higher accuracy for long-term forecasting compared with local LS- and RBF-based polynomial predictors. Gestel *et al.* (2001) investigated a Bayesian method to parameterize LS-SVMs for financial time series forecasting especially in scenarios where the noise levels are comparable to the underlying signal energy and forecasting of the second moment (volatility) is needed. Cao and Gu (2002) developed a dynamic SVM model that uses an exponentially increasing regularization constant and an exponentially decreasing tube size to deal with structural changes in the data.

Hybrid SVM models have also been considered for time series forecasting applications. Zhang *et al.* (2004) developed a wavelet SVM that employs a wavelet kernel to approximate arbitrary nonlinear functions. Cao's SVM expert forecasting model (Cao, 2003) uses a two-stage architecture: an SOM to partition the input space into several

disjointed regions and multiple SVMs (SVM experts) to provide forecasts in each of the partitioned regions. Khemchandani *et al.* (2009) investigated fuzzy SVR models for non-stationary time series forecasting, with the basic assumption that the most recent data samples provide more relevant information for forecasting. This approach reduces over-fitting and computational costs inherent in traditional SVR.

3.1.4. *Hidden or state-observer Markov models*

Most of the models reviewed above involve batch processing, where the model is fit and updated intermittently using batches of historic data. However, the curse of dimensionality due to the prohibitive computational effort, memory requirements, and large data sizes hampers their applicability to many real-world problems, especially for online process monitoring. A variety of sequential (also known as online or recursive) forecasting models, such as Hidden Markov Models (HMMs; Rabiner (1989)), are investigated to surmount this limitation. An HMM is a special class of mixture models, where the observed time series $y(t)$ is treated as a function of the underlying, unobserved states vector x_t (see Fig. 8). A state vector may be reconstructed from autoregressive terms of $y(t)$. The state x_t is a Markov process, whose evolution is given by $x_t = f_t(x_{t-1}, v_{t-1})$. Here, $f_t : R^{n_x} \times R^{n_v} \to R^{n_x}$ is the Markov transition function, v_{t-1} is an independent and identically distributed process noise, and n_x and n_v are the dimensions of the state and process noise, respectively. The observed time series is given by $y(t) = h_t(x_t, w_t)$, where $h_t : R^{n_x} \times R^{n_w} \to R^{n_y}$ is a function of x_t, w_t is the measurement noise, and n_y and n_w are the dimensions of y and w, respectively. One limitation of HMMs is that their forecasting performance tends to be particularly sensitive to the order of the Markovian employed to represent the state. Additionally, the parametric form of HMM models can be unwieldy if not inadequate to capture the evolution of highly non-stationary time series. Consequently, HMMs are attractive to model the evolution of time series generated from a discrete or discretized state space (in contrast with a continuous state space) where the switching between the finite state follows a definite pattern (e.g., electricity price fluctuation due to competitor strategies; Gonzalez *et al.* (2005)).

Generally speaking, state space models such as Kalman Filter (KF) and Particle Filter (PF; Arulampalam *et al.* (2002)) can be classified as HMMs. In order to relax restrictive Gaussian and linearity assumptions in KF, Extended Kalman filters (EKFs; Wang *et al.* (2009)) have

been attempted for nonlinear time series forecasting. An EKF model still assumes a Gaussian posterior and uses a first-order Taylor series expansion to approximate state dynamics. Therefore, Unscented KFs (UKFs; Wan and van der Merwe (2000)) have been introduced to overcome the limitations of EKFs. Instead of local linearization and to avoid the Jacobian matrix calculation inherent in EKFs, UKFs choose a small sample of points to achieve a more accurate estimate of local dynamics, and the evolution of these sample points is propagated at each estimation step. A mixture of KFs (Cen and Liu, 2000) was developed to approximate the system dynamics as a mixture of Gaussian distributions. Dual KF methods (Moradkhani *et al.*, 2005) and EM-algorithms for nonlinear state space models (Ghahramani and Roweis, 1999) have also been used for nonlinear and non-stationary time series forecasting.

PF models (Gordon *et al.*, 1993; Kong *et al.*, 2010), which allow structured approximation via Bayesian estimation, have become one of the most effective means for nonlinear time series forecasting with satisfactory computational efficiency whenever the estimation of a posterior distribution is not analytically tractable. PF models approximate the posterior density by discretizing the continuous state variables into samples. A large number of samples (particles) are generated using sequential Monte Carlo sampling methods. Kong *et al.* (2010) assumed a polynomial form of the state space model in the PF structure and predicted the material removal rate in a Chemical-Mechanical Planarization (CMP) process using nonlinear and non-Gaussian time series data from online vibration sensors. Their model outperformed ARMA, KF, and other commonly used forecasting models. They further improved the forecasting performance by considering a logistic model (Kong *et al.*, 2011) to represent the state space structure for PF estimation, and formulated a utility function to establish the stopping criterion for the CMP process based on predicted signal features.

Significant efforts have also been made towards reducing computational overhead arising from sequential Markov Chain Monte Carlo (MCMC) calculations during the fitting of PF and other HMM models. For example, Van der Merwe and Wan (2003) investigated a recursive Bayesian estimation algorithm that combines an importance sampling-based measurement update step with sigma-point KFs for the time update and distribution generation. The posterior state density was represented by a Gaussian mixture model. Kotecha and Djuric (2003) developed a Gaussian PF model that approximated the posterior distributions by single Gaussian. It was shown that under Gaussian assumptions, the PF is asymptotically optimal in the number of particles and hence improves performance and versatility over other Gaussian filters, especially for nonlinear time series. Ko and Fox (2009) combined nonparametric Gaussian Process (GP) models with different versions of Bayesian filters, including PF, EKF, and UKF, in which both forecasting and observation models are learned
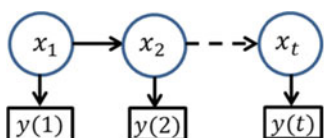

Fig. 8. Graphical representation of an HMM.

from training data using a GP model. The nonparametric GP-based filtering offers several advantages over standard parametric filter models, as it does not require accurate parametric models of the system and degrades gracefully under increasing model uncertainty.

3.2. *Nonparametric models*

Although parametric models can provide accurate forecasts when the models are correctly specified, they tend to become highly suboptimal whenever the underlying dynamics are unknown or indeterminable; additionally, the issue of model biases persists as the dynamics of most real-world complex systems are inherently nonlinear and non-stationary (Fan and Yao, 2003). In contrast, nonparametric models can provide a complete representation of the dynamics based on the observed data, do not impose any structural assumption, and simplify modeling efforts. Consequently, the modeling and forecasting accuracy for nonlinear and non-stationary time series is improved. However, compared with parametric models, nonparametric models usually require larger datasets wherefrom information on the underlying relationships may be effectively gleaned. In this section, we review recent developments in nonparametric models, including state space neighborhood, Bayesian mixture, and decomposition models, for time series forecasting.

3.2.1. *State space neighborhood and local topology-based models*

Nearest neighborhood approaches predict future values by selectively resampling historical observations, with the basic assumption that future behavior varies smoothly; i.e., observations similar to the target one are likely to have similar outcomes. Those models are attractive for complex system dynamics forecasting because of their simplicity and accuracy (Yankov *et al.*, 2006).

In the k nearest neighbor resampling approach with multiple predictor variables of Mehrotra and Sharma (2006), an influence weight was assigned to each predictor to identify nearest neighbors. Hamid *et al.* (2013) investigated a variety of neighbor-based methods for the forecasting of chaotic time series; e.g., zeroth-order approximation (one nearest neighbor), k nearest neighbors (multiple neighbors), and weighted distance approximation (distance weighted average of multiple neighbors) models.

For most complex dynamic systems, it is not possible to observe all relevant variables. Oftentimes, the time evolution of only certain variables is observed, but the relationship to the state variables is unknown or indeterminable. The state space reconstructed from time delay embedding (see Equation (2) and Fig. 4 for details) holds strong similarities (i.e., a diffeomorphic image) to the underlying state space (as $x(t)$ in Equation (1); Kantz and Schreiber (1997)), and offers a unique way for nonlinear time series forecasting

(Finkenstadt and Kuhbier, 1995). Casdagli (1989, 1997) developed a local linear model of the reconstructed state space for chaotic time series forecasting. The predicted value of the current observation was derived from the most recent w embedding vectors. Then k nearest neighbors were identified within the window width w based on the recurrence property of the reconstructed state space. Yang *et al.* (2011) partitioned the reconstructed state space into various near-stationary segments based on local recurrence properties. The evolutionary trajectories were further decomposed using a principal component approach to identify the principal evolution directions and thus predict future states. Similarly, Bukkapatnam *et al.* (2009) developed an LMM based on the recurrence properties. A Markov transition matrix derived from the reconstructed state space was applied for system pattern analysis and to partition the trajectories into piecewise stationary segments. An LMM was employed to predict future states in each near-stationary segment based on the obtained transition matrix. McNames (1998) studied time series forecasting with little or no noise. Neighbors in the reconstructed state space were identified by an optimized weighted Euclidean metric, and a novel ρ-step-ahead cross-validation error was used to assess model accuracy. Regonda *et al.* (2005) investigated a local polynomial regression model using neighbors and future evolutions thereof in the reconstructed state space. An ensemble nearest-neighbor model was then implemented via selecting a suite of parameter combinations for the local regression model. This ensemble approach was able to capture the effects of parameter uncertainty adequately.

3.2.2. *Bayesian nonparametric models*

Bayesian modeling is basically a process of incorporating prior information to render posterior inference; i.e., estimating the conditional distribution $p(\theta|y)$ of the hidden model or parameters θ given an observed time series $y(t)$ (Gershman and Blei, 2012). Different from other Bayesian methods, Bayesian nonparametric models assume the hidden structure here grows with the data. In other words, Bayesian nonparametric models seek a single model from infinitely many possibilities (i.e., θ may be infinite-dimensional) whose complexity (e.g., the number of parameters to estimate) is adapted according to the data.

Among Bayesian nonparametric models, GP models have been the most widely studied for time series forecasting (Rasmussen and Williams, 2006). A GP model seeks to establish a mapping f of the form $y = f(x) + \varepsilon$, where x can be constructed from historical realizations as indicated in Section 2, as well as exogenous variables, and assumes $y \sim N(0, K)$. Here, K is the covariance matrix, which depends on the covariance function. Although an explicit form of f is not specified, the covariance function, which varies with the sample size, is used to capture the underlying relationships. A GP model provides not just a point estimate but a complete distribution of the forecast-

ing. GP models have two major limitations, namely, the computational expense to perform a matrix inversion and the stationary covariance function assumption. Many attempts have been investigated in the literature to address these issues. For instance, non-stationary covariance functions introduced to overcome the stationarity assumptions (Brahim-Belhouari and Bermak, 2004; Plagemann *et al.*, 2008; Wang *et al*, 2008) are suited only for simple nonlinear and non-stationary forms, such as linear trends, and introduce additional parameters to fit. The GP model of Adams and Stegle (2008) treats data as a point-wise product of two GPs to handle non-stationarity. Zhou *et al.* (2006) investigated a Multi-scale representation of GP (MGP) for nonlinear and non-stationary time series forecasting. In their model, the GP was represented as a linear combination of a series of basis functions, which were defined as the dilation and translation of a scaling function. The MGP model can consider the non-stationarity of the time series, guarantee the optimal value of hyper-parameters, and improve the forecasting accuracy. The recurrence-based local GP model of Bukkapatnam and Cheng (2008) is based on partitioning the time series into near-stationary segments based on the recurrence of the state space reconstructed from the signals. A local GP model was built in each segment to forecast the evolution of carbon nanotube length increments in a chemical vapor deposition process (Cheng *et al.*, 2012). Duy and Peters (2008) presented recursive local GP models via partitioning the training data into local regions according to a distance metric. A weighted sum of the local model estimates was used to provide the forecast.

Recently, multi-step forecasting of time series with GP models has also been explored. Palm (2007) combined GP with the Takagi–Sugeno fuzzy model for multi-step forecasting. Girard *et al.* (2003) studied a GP model for multi-step time series forecasting with input uncertainty, where the input for next-step forecasting was derived from sampling the output distribution, specified in terms of the mean and variance estimated at the previous time step of the Gaussian process. Hachino and Kadirkamanathan (2007) presented a direct approach for multi-step forecasting using GP models. Time series forecasting was carried out directly by each step-ahead predictor to avoid accumulated prediction error, and genetic algorithm was used to optimize hyper-parameters. Mixture GP models have also been explored to combine local GP models for nonlinear and non-stationary time series forecasting (Shi *et al.*, 2005). Additionally, Dirichlet Process (DP) nonparametric models (Blei and Jordan, 2005) are being investigated for adaptively determining the number of clusters based on the complexity of the data. More pertinently, DP-based infinite mixture GP models (Blei and Jordan, 2005; Meeds and Osindero, 2006; Chatzis and Demiris, 2012) have been studied to handle various forms of non-stationarity in nonlinear time series forecasting.

3.2.3. *Functional decomposition models*

Functional decomposition models (Frei and Osorio, 2007) for time series forecasting have received increased attention in recent years. These models rely on the adaptive basis, which does not require *a priori* specification of a parametric functional (e.g., sine wave basis in Fourier analysis and Haar wave basis in wavelet analysis). Instead, local characteristics of the time series, including the spacing between successive local extrema (maxima and minima) over different timescales, are used to construct the decomposition basis. These models can be used to capture the drifts and nonlinear modes of any nonlinear and non-stationary processes. Most of the forecasting methods in this category follow a hybrid approach in that they use the decomposition model to represent the time series and employ another parametric or nonparametric technique to forecast the evolution of the individual components. Although we reviewed these models under nonparametric approaches, some functional decomposition models presented hereunder, such as Karhunen Loeve (Yang *et al.*, 2007), employ parametric forms.

Functional decomposition often suppresses or disperses over different functional components the effects of non-stationarity (e.g., trend and seasonality) in the time series. Multiplicative decompositions may also be used to tackle trends in seasonal variations; i.e., the time series can be considered as the multiplication of seasonal and non-seasonal components. A multiplicative decomposition technique-based ARIMA model was reported to address seasonality and other non-stationarity issues in monthly energy consumption forecasting (Damrongkulkamjorn and Churueang, 2005). The energy consumption time series were decomposed into trend cycle and seasonality components and a seasonal ARIMA model was applied to the trend-cycle part. If the seasonal variation is relatively constant with the trend, additive decomposition is usually preferred (Zhang and Qi, 2005).

Among nonparametric decomposition models, Empirical Mode Decomposition (EMD; Huang *et al.* (1998)) can decompose non-stationary time series into a finite number of components called Intrinsic Mode Functions (IMFs), such that evolutions of each IMF can be explored individually at different timescales via classic time series forecasting techniques, such as AR or ARMA models (Wang *et al.*, 2007; Xu *et al.*, 2010). Because EMD allows perfect reconstruction of the original time series with IMFs and isolation of trend and noise components from a non-stationary process (Stroeer *et al.*, 2009), it can improve long-term forecasting accuracy. Sa-ngasoongsong and Bukkapatnam (2011) developed a two-step EMD model and applied it to long-term customer willing-to-pay forecasting. The wind farm power forecasting model of An *et al.* (2012) was based on estimating the largest Lyapunov exponents for each IMF. They applied the largest Lyapunov exponent forecasting method (Wolf *et al.*, 1985) to provide higher forecasting

accuracy than other direct EMD-based forecasting models. SVR models built on each IMF were also fused in different ways (Zhu *et al.*, 2007; Yu *et al.*, 2008; Li *et al.*, 2012) to forecast non-stationary time series. EMD has also been applied for weather time series forecasting (Yang *et al.*, 2010; Lee and Ouarda, 2011). Although attractive for nonlinear and non-stationary forecasting, EMD poses some mathematical challenges due to the end effects (Huang, 2005). Several possible fixes such as adaptive decompositions have been attempted, especially in the context of financial time series forecasting (Fu, 2010; Hong, 2011).

Recently, Intrinsic Time-scale Decomposition (ITD) has been investigated for precise time-frequency-energy analysis of time series (Frei and Osorio, 2007). ITD overcomes the limitations of classic Fourier, wavelet, and EMD approaches for nonlinear and non-stationary time series modeling and decomposes time series into proper rotation components (a signal with strictly positive values at all local maxima and strictly negative values at all local minima; Frei and Osorio (2007)) with defined frequency and amplitude and a monotonic trend. This decomposition preserves precise temporal information regarding critical points (e.g., local extrema) in the time series, with a temporal resolution equal to the timescale of extrema occurrence. The ITD model has been used for nonlinear biomedical signal characterization (Martis *et al.*, 2013), and ITD-based ECM was attempted to forecast long-term automobile demand (Sa-ngasoongsong, 2012).

3.3. *Summary*

Table 1 presents a summary of the comparison of different forecasting models, including ARCH/GARCH, MR-TAR (Multi-regime threshold AR), WNN (wavelet neural network), NNM (nearest neighborhood model), PF, ITD, LGP (local recurrence-based GP), and DPMG, in terms of model capacity, such as nonlinearity form (e.g., linear trend), non-stationary form (e.g., changed variance), computational (training) complexity, and noise effect, based on our own investigation. The decomposition models can accommodate large noise variances in the underlying system, and MRTAR and PF models can incorporate exogenous variables as the unobserved state. We use tc to represent the computational complexity. It is a function of various factors, including model order, number of inputs, input dimension, and so on. Here, n is the number of inputs, p and q are the model orders, n_c is the number of regimes or clusters, n_n is the number of neurons in NN, and n_x is the state dimension. Compared with classic GP models, LGP and DPMG models can be applied to non-stationary time series and reduce the computational overhead, as the number of training points in each segment or cluster is reduced.

4. Case studies

During the past 7 years, the authors have investigated parametric and nonparametric nonlinear models that assume

Table 1. Summary of the performance of different prediction models in prior empirical studies

Prediction models	Nonlinearity	Non-stationarity	Training complexity	Remarks
ARCH/GARCH	Linear	Time-varying volatility	$tc(n, p, q)$	Linear structures; can model varying second moments; sensitive to noise
MRTAR	Piecewise linear	Switch according to time and indicator	$tc(n_c, p, q, n)$	Exogenous variables used to indicate switch
Wavelet-NN	Between lagged variables	Non-stationarity dispersed over scales	$tc(n_n, n)$	Can separate relatively large noise
PF	Between lagged variables	Adaptive parameter update	$tc(n_x, n)$	Posterior approximation; MCMC computations are cumbersome
NNM	Local linearity/nonlinearity	Local stationarity	$tc(n)$	Sensitive to neighborhood threshold; can accommodate large noise level
ITD	Decomposed into monotonic trend	Decomposed into rotation components	$tc(n)$	Construct piecewise linear baseline signal; reduce boundary effect in EMD; reduced sensitivity to noise
LGP	Between lagged variables	Piecewise stationary	$tc(n)$	Recurrence based; sensitive to threshold; GP requires relatively low noise level
DPMG	Between lagged variables	Piecewise stationary	$tc(n_c, n)$	DP clustering; GP accuracy sensitive to noise

Table 2. Comparison of the accuracy of sleep apnea forecasts

	ARMA	*LGP*	*LRM*	*PF*	*RPNN*	*EMD*	*DPMG*
Forecasting horizon	3	5	5	4	1	3	3
R^2 (first step)	0.37	0.53	0.51	0.42	0.32	0.45	0.92

piecewise stationarity to forecast the evolution of complex systems. The research has resulted in models based on local GP (Bukkapatnam and Cheng, 2010), local recurrence characteristics (Yang *et al*., 2011), local Markovian properties (Bukkapatnam *et al*., 2009), particle filtering with RNNs (Rao *et al*., 2013), EMD (Sa-ngasoongsong and Bukkaptnam, 2011), and DPMG (Le *et al*., 2013) for complex systems forecasting. In this section we present three case studies to compare the performance of various forecasting methods in applications ranging from biomedical and manufacturing shop-floor to precision manufacturing processes.

4.1. *Sleep apnea episode prediction*

Obstructive Sleep Apnea (OSA) is a common breathing disorder that affects a tenth of the world's adult population (Young *et al*., 1993). An OSA episode is marked by the interruption of the airflow during sleep. It is known to reduce sleep quality and increase the risk for hypertension, coronary artery diseases, and stroke. Current widespread therapies, such as Continuous Positive Airway Pressure (CPAP), are not compatible with an estimated 46–83% of patients with OSA (Le *et al*., 2013).

Forecasting of OSA events from tracing the evolution of biorhythm signatures is vital toward calibrating airflow and allowing proactive adjustment of airflow and body positions to mitigate OSA, thus improving the adherence of patients to the CPAP therapy. A nonparametric DPMG model has been developed to predict the complex evolution of OSA signatures, such as power spectral density in a 0.04–0.12 Hz frequency range, and the longest vertical length (Le *et al*., 2013). These signatures were extracted from heart rate signals measured from a unique wireless biometric (multisensory) sleepwear. The forecasting approaches were extensively tested using 20 recordings from the Physionet database and 10 recordings from eight subjects wearing the sleepwear. A comparison of the performance of different forecasting models is summarized in Table 2. Here, we performed recursive multi-step forecasting for each model where the forecasts at a particular look-ahead were used as the input for the subsequent forecasting step. The forecasting horizon here refers to the number of forecasting steps (minutes) ahead before the R^2 value degrades to below zero. Among the methods tested, the DPMG model effectively captures the nonlinear and non-stationary evolution of the signatures and provides OSA prediction accuracy of 83% for 1-min-ahead and 77% for 3-min-ahead. This level of accuracy amounts to some

20–40% improvement compared with other methods. The improved accuracy from DPMG compared with LGP and LRM results from its use of a DP to identify the local clusters. LGP and LRM are sensitive to the threshold used to define the local clusters, and the highly non-stationary time series here make it unwieldy to select a suitable threshold. The early prediction from DPMG can spur the development of adaptive flow control in CPAP devices and lead to the advent of devices to induce minor adjustments to body positions to mitigate OSA.

4.2. *Real-time throughput forecasting in an automotive assembly line*

The time series shown in Fig. 9 represents the number of parts produced (also called the throughput) in an 8-hour shift from a station of an automotive assembly line (Yang *et al*., 2011). As shown in the figure, the values fluctuate rapidly and erratically, thereby making conventional forecasting approaches unwieldy. A comparison of forecasting accuracies of different models we investigated is summarized in Table 3. Here, the PF/RPNN model did not converge because of high non-stationarity and data sparsity. All methods tested except ARMA provide comparable one-step forecasting accuracy (R^2 in the range of 0.5–0.6), as they are designed to capture the nonlinear and non-stationary evolution. The linear structure of ARMA model is evidently inadequate to capture the complex throughput variation. Among the nonlinear and non-stationary forecasting methods, EMD provides the highest long term (six-step) forecasting accuracy as some of the IMFs exhibit a near-stationary evolution. LGP has the best prediction accuracy in terms of the second moment, and over 85% of the realized values are within two-sigma (estimated) limits of the predictions likely due to effective partitioning of the state space into local near-stationary segments.
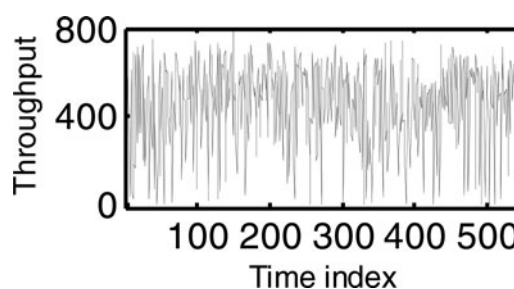

Fig. 9. Automotive assembly line throughput time series.

Table 3. Comparison of forecasting horizon and accuracy of automotive assembly line throughput

	ARMA	LGP	LRM	LMM	EMD	DPMG
Forecasting horizon	2	3	3	5	6	5
R^2 (first step)	0.10	0.61	0.53	0.60	0.60	0.59

4.3. *Endpoint detection in CMP processes*

The semiconductor industry relies on CMP processes to polish wafer surfaces to meet the strongest surface roughness, flatness, and defect-control regularizations. Various sensors can be used during the polishing process for *in situ* monitoring and control of Material Removal Rate (MRR) and surface quality (Kong *et al.*, 2010; Kong *et al.*, 2011). In a recent study, we instrumented a LapMaster 12 benchtop lapping machine with wireless vibration sensors for polishing small-diameter (ϕ 100 mm) copper wafers (Kong *et al.*, 2011). The sensor signals were gathered from the process under various conditions, and the dynamics underlying the measured signals were found to be nonlinear and non-stationary (Kong *et al.*, 2010; Kong *et al.*, 2011). Fourteen different features capturing the various complex patterns of the signals were extracted, and nine principal components of these features were used to predict MRR, necessary for timely control of wafer height. An NN model was used to predict MRR from the forecasts of the first nine principal features. An evolution of the first principal feature (see Fig. 10) suggests significant non-stationarity and aperiodicity of the dynamics. Table 4 summarizes the accuracy of forecasting first principal feature and MRR using alternative methods. Although PF can adaptively adjust to the nonlinear dynamics and provide the highest accuracy for forecasting the feature evolution as well as MRR, EMD offers higher accuracy for relatively long-term forecasting, as the decomposed IMFs represent the long-term
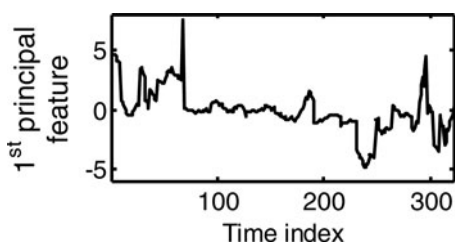
Fig. 10. Evolution of the first principal feature of the vibration signal over time.

trend in the time series. Also, 91% of the realized feature values were within two-sigma (estimated) limits of DPMG forecasts, thus facilitating accurate endpoint detection to mitigate over- and under-polishing defects in semiconductor wafers (Ghahramani and Roweis, 1999).

5. Conclusions

Time series forecasting has become essential toward advancing the way we manage and control complex real-world systems—from a conventional *detection–diagnosis–mitigation* to a more proactive *prediction–prognosis–prevention* paradigm. Time series from real-world complex systems exhibit multifarious nonlinear and non-stationary dynamic behaviors including aperiodic patterns, intermittencies, and other transient behaviors. The genesis of such behaviors is mainly attributed to the nature of the relationships among the intrinsic and exogenous state variables and their AR terms. Additionally, the dynamics are often unknown or indeterminable, rendering tractable forecasting of the evolution of complex system dynamics a daunting challenge.

In this article, we have reviewed the recent developments in nonlinear and non-stationary time series forecasting and have provided a comparative evaluation of the performance of alternative models in real-world application case studies, especially where traditional models (ARMA and KF) fail to capture the system evolutions. The following implications may be elicited based on this review.

1. A vast majority of nonlinear and non-stationary process forecasting methods consider only a narrow set of non-stationary behaviors, namely, linear or nonlinear (seasonal) drifts in the first moment. A few models reported in the literature consider variations in the second moment and covariance characteristics, including periodic and cyclo-stationary behaviors. However, very little has been reported on the forecasting of piecewise stationary nonlinear processes, as well as those with continuous,

Table 4. Comparison of forecasting horizon and accuracies for the first principal feature and MRR

	ARMA	LGP	LRM	PF	RPNN	EMD	DPMG
Forecasting horizon	1	3	2	3	1	4	3
R^2 (first step)	0.07	0.53	0.56	0.80	0.10	0.75	0.61
MRR estimation R^2 (first step)	0.01	0.63	0.67	0.89	0.05	0.84	0.71

arbitrary drifts in higher-order moments. With an increasing abundance of data, Bayesian sequential learning approaches—that can relax many of the assumptions commonly made on the nature of time-varying characteristics and nonlinearity—provide an attractive means to incorporate prior information for nonlinear and non-stationary time series forecasting.

2. Although the majority of previous forecasting methods use parametric models, recently there has been an increased interest in nonparametric models for forecasting nonlinear and non-stationary time series. The use of nonparametric models is more convenient because the function form of underlying relationships is unknown or indeterminable for the vast majority of complex systems. Among these, mixture models (e.g., DMGP) are being considered for capturing various non-Gaussian and non-stationary behaviors.

3. Adapting the principles of nonlinear dynamic systems into nonparametric modeling approaches can provide an attractive means to further advance forecasting under intermittencies and other piecewise stationary conditions. Further investigations into nonlinear transforms and multi-scale models such as EMD may offer new opportunities for long-term forecasting of process states. The robustness of these methods to noise needs to be enhanced to make them attractive for real-world applications.

With the increasing availability of large streams of high-dimensional data, there is an urgent need to develop tractable models, which are collectively identified under "big data" analytics, to effectively extract the hidden patterns in complex systems, reduce the dimensionality of data, determine dynamic causality, and leverage a compact set of exogenous factors for accurate forecasting. These advanced techniques for "big data" forecasting are being increasingly sought in various engineering domains, including manufacturing (quality monitoring and prognostics), civil infrastructure and mechanical systems (design and planning), healthcare delivery (point of care), and energy (smart grid control), and they are essential for ushering in the P4 paradigm to improve the overall quality and integrity of complex nonlinear and non-stationary systems.

Acknowledgments

References

Adams, R.P. and Stegle, O. (2008) Gaussian process product models for nonparametric nonstationarity, in *Proceedings of the 25th International Conference on Machine Learning*, ACM, New York, NY, pp. 1–8.

Amaral, G.F.V., Letellier, C. and Aguirre, L.A. (2006) Piecewise affine models of chaotic attractors: the Rössler and Lorenz systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **16**(1), 013115–013315-4.

An, X., Jiang, D., Zhao, M. and Liu, C. (2012) Short-term prediction of wind power using EMD and chaotic theory. *Communications in Nonlinear Science and Numerical Simulation*, **17**(2), 1036–1042.

Arulampalam, M.S., Maskell, S., Gordon, N. and Clapp, T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50**(2), 174–188.

Aznarte, J.L., Benítez Sánchez, M.J.M., Lugilde, D.N., de Linares Fernández, C., de la Guardia, C.D. and Sánchez, F.A. (2007) Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. *Expert Systems with Applications*, **32**(4), 1218–1225.

Balabin, R.M. and Lomakina, E.I. (2011) Support vector machine regression (SVR/LS-SVM)–an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst*, **136**(8), 1703–1712.

Barreto, G., Hammer, B. and Hitzler, P. (2007) Time Series prediction with the self-organizing map: a review perspectives of neural-symbolic integration, in *Perspectives of Neural-Symbolic Integration*, eds. B. Hammer and P. Hitzler, Springer, Berlin, pp. 135–158.

Battaglia, F. and Protopapas, M.K. (2011) Time-varying multi-regime models fitting by genetic algorithms. *Journal of Time Series Analysis*, **32**(3), 237–252.

Blei, D.M. and Jordan, M.I. (2005) Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, **1**, 121–144.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307–327.

Brahim-Belhouari, S. and Bermak, A. (2004) Gaussian process for non-stationary time series prediction. *Computational Statistics & Data Analysis*, **47**(4), 705–712.

Broer, H.W., Huitema, G.B. and Sevryuk, M.B. (1996) Quasi-periodic motions in families of dynamical systems: order amidst chaos. *Lecture Notes in Mathematics*, **1645**.

Bukkapatnam, S.T.S. and Cheng, C. (2010) Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian process models. *Physical Review E*, **82**(5), 056206–056206-12.

Bukkapatnam, S.T.S., Yang, H. and Madhavi, F. (2009) Towards prediction of nonlinear and nonstationary evolution of customer preferences using local Markov models, in *Product Research*, Raghavan, N.R.S. and Cafeo, J.A. (eds), Springer, Dordrecht, The Netherlands, pp. 271–287.

Cao, L. (2003) Support vector machines experts for time series forecasting. *Neurocomputing*, **51**, 321–339.

Cao, L. and Gu, Q. (2002) Dynamic support vector machines for non-stationary time series forecasting. *Intelligent Data Analysis*, **6**(1), 67–83.

Casdagli, M. (1989) Nonlinear prediction of chaotic time series. *Physica D: Nonlinear Phenomena*, **35**(3), 335–356.

Casdagli, M.C. (1997) Recurrence plots revisited. *Physica D: Nonlinear Phenomena*, **108**(1–2), 12–44.

Chatzis, S.P. and Demiris, Y. (2012) Nonparametric mixtures of Gaussian processes with power-law behavior. *IEEE Transactions on Neural Networks and Learning Systems*, **23**(12), 1862–1871.

Chen, R. and Liu, J.S. (2000) Mixture Kalman filters. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **62**, 493–508.

Chen, S., Billings, S.A. and Grant, P.M. (1992) Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *International Journal of Control*, **55**(5), 1051–1070.

Chen, S., Cowan, C.F.N. and Grant, P.M. (1991) Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, **2**(2), 302–309.

Cheng, C., Bukkapatnam, S.T.S., Raff, L.M., Hagan, M. and Komanduri, R. (2012) Monte Carlo simulation of carbon nanotube nucleation and growth using nonlinear dynamic predictions. *Chemical Physics Letters*, **530**, 81–85.

Crauel, H., Debussche, A. and Flandoli, F. (1997) Random attractors. *Journal of Dynamics and Differential Equations*, **9**(2), 307–341.

Damrongkulkamjorn, P. and Churueang, P. (2005) Monthly energy forecasting using decomposition method with application of seasonal ARIMA, in *Procedings of the Seventh International Power Engineering Conference*, IEEE, Singapore, pp. 1–229.

De Gooijer, J.G. and Hyndman, R.J. (2006) 25 years of time series forecasting. *International Journal of Forecasting*, **22**, 443–473.

De Groot, C. and Wuertz, D. (1991) Analysis of univariate time series with connectionist nets: a case study of two classical examples. *Neurocomputing*, **3**(4), 177–192.

Duy, N.-T. and Peters, J. (2008) Local Gaussian process regression for real-time model-based robot control, in *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Nice, France, pp. 380–385.

Eckmann, J.P., Kamphorst, S.O. and Ruelle, D. (1987) Recurrence plots of dynamical systems. *Europhysics Letters*, **4**(9), 973–977.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**(4), 987–1007.

Fan, J. and Yao, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag, New York, NY.

Finkenstädt, B. and Kuhbier, P. (1995) Forecasting nonlinear economic time series: a simple test to accompany the nearest neighbor approach. *Empirical Economics*, **20**(2), 243–263.

Fraser, A.M. and Swinney, H.L. (1986) Independent coordinates for strange attractors from mutual information. *Physical Review A*, **33**(2), 1134–1140.

Frei, M.G. and Osorio, I. (2007) Intrinsic time-scale decomposition: time–frequency–energy analysis and real-time filtering of nonstationary signals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, **463**(2078), 321–342.

French, M.N., Krajewski, W.F. and Cuykendall, R.R. (1992) Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, **137**(1), 31–31.

Fu, C. (2010) Forecasting exchange rate with EMD-based support vector regression, in *Proceedings of the International Conference on Management and Service Science*, IEEE, Wuhan, China, pp. 1–4.

Genton, M. and Thompson, K. Skew-elliptical time series with application to flooding risk, in *Time Series Analysis and Applications to Geophysical Systems*, Brillinger, D., Robinson, E. and Schoenberg, F. (eds), Springer, New York, pp. 169–185.

Gershman, S.J. and Blei, D.M. (2012) A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, **56**(1), 1–12.

Gestel, T.V., Suykens, J.A.K., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., Moor, B.D. and Vandewalle, J. (2001) Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks*, **12**, 809–821.

Ghahramani, Z. and Roweis, S. (1999) Learning nonlinear dynamical systems using an EM algorithm. *Advances in Neural Information Processing Systems*, **11**, 431–437.

Girard, A., Rasmussen, C.E., Quinonero-Candela, J. and Murray-Smith, R. (2003) Gaussian process priors with uncertain inputs? Application to multiple-step ahead time series forecasting, in *Proceedings*

of the Conference on Neural Information Processing Systems, MIT Press, Cambridge, Massachusetts, pp. 9–14.

Goldberger, A.L. (1999) Implications for neuroautonomic heart rate control in health and disease, in *The Autonomic Nervous System*, Bolis, C. and Licinio, J. (eds), World Health Organization, Geneva.

Gonzalez, A.M., Roque, A.M.S. and Garcia-Gonzalez, J. (2005) Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on Power Systems*, **20**(1), 13–24.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F, Radar and Signal Processing*, **140**(2), 107–113.

Grudnitski, G. and Osburn, L. (1993) Forecasting S&P and gold futures prices: an application of neural networks. *Journal of Futures Markets*, **13**(6), 631–643.

Hachino, T. and Kadirkamanathan, V. (2007) Time series forecasting using multiple Gaussian process prior model, in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, IEEE, Honolulu, HI, pp. 604–609.

Hamid, N.Z.A. and Noorani, M.S.M. (2013) Modeling of prediction system: an application of nearest neighbor approach to chaotic data. *Applied Mathematics and Computational Intelligence*, **2**, 137–148.

Hamilton, J.D. (1989) A new approach to the economic analysis of time series. *Econometrica*, **57**(2), 357–384.

Heagy, J.F., Platt, N. and Hammel, S.M. (1994) Characterization of on-off intermittency. *Physical Review E*, **49**(2), 1140–1150.

Hippert, H.S., Pedreira, C.E. and Souza, R.C. (2001) Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, **16**(1), 44–55.

Hong, L. (2011) Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition. *Energy Procedia*, **5**, 1333–1340.

Huang, N.E. (2005) Introduction to the Hilbert-Huang transform and its related mathematical problems. *Interdisciplinary Mathematics*, **5**, 1–26.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **454**(1971), 903–995.

Kantz, H. and Schreiber, T. (1997) *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK.

Katok, A. and Hasselblatt, B. (1995) *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, Cambridge, UK.

Kennel, M.B., Brown, R. and Abarbanel, H.D. (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, **45**(6), 3403–3411.

Khashei, M., Reza Hejazi, S. and Bijari, M. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*, **159**(7), 769–786.

Khemchandani, R., Jayadeva and Chandra, S. (2009) Regularized least squares fuzzy support vector regression for financial time series forecasting. *Expert Systems with Applications*, **36**(1), 132–138.

Kirchgässner, G., Wolters, J. and Hassler, U. (2013) *Introduction to Modern Time Series Analysis*, Springer, Berlin, Germany.

Ko, J. and Fox, D. (2009) GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models. *Autonomous Robots*, **27**(1), 75–90.

Kodogiannis, V. and Lolis, A. (2002) Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural Computing & Applications*, **11**(2), 90–102.

Kohzadi, N., Boyd, M.S., Kermanshahi, B. and Kaastra, I. (1996) A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, **10**(2), 169–181.

Kong, Z., Beyca, O., Bukkapatnam, S.T.S. and Komanduri, R. (2011) Nonlinear sequential Bayesian analysis-based decision making for

end-point detection of chemical mechanical planarization (CMP) processes. *IEEE Transactions on Semiconductor Manufacturing*, **24**(4), 523–532.

Kong, Z., Oztekin, A., Beyca, O.F., Phatak, U., Bukkapatnam, S.T.S. and Komanduri, R. (2010) Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear Bayesian analysis and statistical modeling. *IEEE Transactions on Semiconductor Manufacturing*, **23**(2), 316–327.

Kotecha, J.H. and Djuric, P.M. (2003) Gaussian particle filtering. *IEEE Transactions on Signal Processing*, **51**(10), 2592–2601.

Koutsoyiannis, D. and Montanari, A. (2007) Statistical analysis of hydroclimatic time series: uncertainty and insights. *Water Resources Research*, **43**(5), W05429:1–9.

Krishnamurthy, V. and Yin, G.G. (2002) Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime. *IEEE Transactions on Information Theory*, **48**(2), 458–476.

Kuan, C.M. and Liu, T. (1995) Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, **10**(4), 347–364.

Kumar, C., Patel, N. and Jarudi, I. (2003) 21st century physics: grand challenges. *Journal of the Federation of American Scientists*, **56**(2), 9–11.

Lai, K., Yu, L., Wang, S. and Wei, H. (2006) A novel nonlinear neural network ensemble model for financial time series forecasting. *Lecture Notes in Computer Science*, **3991**, 790–793.

Lang, T., Plagemann, C. and Burgard, W. (2007) Adaptive non-stationary kernel regression for terrain modeling, in *Proceedings of the Robotics: Science and Systems Conference*, MIT Press, Cambridge, MA, pp. 81–88.

Lapedes, A. and Farber, R. (1987) Nonlinear signal processing using neural networks: prediction and system modeling. Report, Los Alamos Nationary Laboratory, Los Alamos, NM.

Last, M. and Shumway, R. (2008) Detecting abrupt changes in a piecewise locally stationary time series. *Journal of Multivariate Analysis*, **99**(2), 191–214.

Lau, K.W. and Wu, Q.H. (2008) Local prediction of non-linear time series using support vector regression. *Pattern Recognition*, **41**(5), 1539–1547.

Le, T., Cheng, C., Sangasoongsong, A., Wongdhamma, W. and Bukkapatnam, S.T.S. (2013) Wireless wearable multisensory suite and real-time prediction of sleep apnea episodes. *IEEE Journal of Translational Engineering in Health and Medicine*, **1**(1), 2700109.

Lee, T. and Ouarda, T.B.M.J. (2011) Prediction of climate nonstationary oscillation processes with empirical mode decomposition. *Journal of Geophysical Research: Atmospheres*, **116**(D6), D06107, 1–15.

Li, J.P., Tang, L., Sun, X.L., Yu, L.A., He, W. and Yang, Y.Y. (2012) Country risk forecasting for major oil exporting countries: a decomposition hybrid approach. *Computers and Industrial Engineering*, **63**(3), 641–651.

Lineesh, M.C. and John, C.J. (2010) Analysis of non-stationary time series using wavelet decomposition. *Nature and Science*, **8**(1), 53–59.

Lowe, D. and Webb, A.R. (1991) Time series prediction by adaptive networks: a dynamical systems perspective. *IEE Proceedings of Radar and Signal Processing*, **138**(1), 17–24.

Lundbergh, S., Teräsvirta, T. and van Dijk, D. (2003) Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics*, **21**(1), 104–121.

Martis, R.J., Acharya, U.R., Tan, J.H., Petznick, A., Tong, L., Chua, C.K. and Ng, E.Y.K. (2013) Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction. *International Journal of Neural Systems*, **23**(05), 1350023.

Marwan, N., Carmen Romano, M., Thiel, M. and Kurths, J. (2007) Recurrence plots for the analysis of complex systems. *Physics Reports*, **438**(5–6), 237–329.

Meeds, E. and Osindero, S. (2006) An alternative infinite mixture of Gaussian process experts, in *Advances in Neural Information Processing Systems*, Weiss, Y., Scholkopf, B. and Platt, J. (eds), MIT Press, Cambridge, MA, pp. 883–890.

Menezes, J.M.P. and Barreto, G.A. (2008) Long-term time series prediction with the NARX network: an empirical evaluation. *Neurocomputing*, **71**(16-18), 3335–3343.

Marchesi, M. and Lux, T. (2000) Volatility clustering in financial markets: a microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, **3**(04), 675–702.

McAleer, M. and Medeiros, M.C. (2008) A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, **147**(1), 104–119.

McNames, J. (1998) A nearest trajectory strategy for time series prediction, in *Proceedings of the International Workshop on Advanced BlackBox Techniques for Nonlinear Modeling*, Katholieke Universiteit, Leuven Belgium, pp. 112–128.

Mehrotra, R. and Sharma, A. (2006) Conditional resampling of hydrologic time series using multiple predictor variables: a K-nearest neighbour approach. *Advances in Water Resources*, **29**(7), 987–999.

Moradkhani, H., Sorooshian, S., Gupta, H.V. and Houser, P.R. (2005) Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, **28**(2), 135–147.

Mukherjee, S., Osuna, E. and Girosi, F. (1997) Nonlinear prediction of chaotic time series using support vector machines, in *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing*, IEEE, Amelia Island, FL, pp. 511–520.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. (2004) A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, **291**(1–2), 52–66.

Palm, R. (2007) Multiple-step-ahead prediction in control systems with Gaussian process models and TS-fuzzy models. *Engineering Applications of Artificial Intelligence*, **20**(8), 1023–1035.

Park, J. and Sandberg, I.W. (1991) Universal approximation using radial-basis-function networks. *Neural Computation*, **3**(2), 246–257.

Plagemann, C., Kersting, K. and Burgard, W. (2008) Nonstationary Gaussian process regression using point estimates of local smoothness. *Lecture Notes in Computer Science*, **5212**, 204–219.

Rabiner, L. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

Rao, K.P., Bukkapatnam, S.T.S., Beyca, O.F., Kong, Z. and Komanduri, R. (2013) Real-time identification of incipient surface morphology variations in ultra-precision machining process. *ASME Journal of Manufacturing Science and Engineering*, **136**(2), 021008 1–11.

Rasmussen, C.E. and Williams, C.K.I. (2006) *Gaussian Process for Machine Learning*, MIT Press, Cambridge, MA.

Regonda, S., Rajagopalan, B., Lall, U., Clark, M. and Moon, Y.-I. (2005) Local polynomial method for ensemble forecast of time series. *Nonlinear Processes in Geophysics*, **12**, 397–406.

Rosen, O., Wood, S. and Stoffer, D.S. (2012) AdaptSPEC: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association*, **107**(500), 1575–1589.

Sa-ngasoongsong, A. (2012) Long-term demand prediction using long-run equilibrium relationship of intrinsic time-scale decomposition components, in *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, FL.

Sa-ngasoongsong, A. and Bukkapatnam, S.T.S. (2011) Willingness-to-pay prediction based on empirical mode decomposition, in *Proceedings of the Industrial Engineering Research Conference*, Reno, NV, USA.

Sa-ngasoongsong, A., Bukkapatnam, S.T.S., Kim, J., Iyer, P.S. and Suresh, R.P. (2012) Multi-step sales forecasting in automotive industry based on structural relationship identification. *International Journal of Production Economics*, **140**(2), 875–887.

Schreiber, T. (1999) Interdisciplinary application of nonlinear time series methods. *Physics Reports*, **308**(1), 1–64.

Shi, J.Q., Murray-Smith, R. and Titterington, D.M. (2005) Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, **15**(1), 31–41.

Smola, A. and Schölkopf, B. (2004) A tutorial on support vector regression. *Statistics and Computing*, **14**(3), 199–222.

Soltani, S. (2002) On the use of the wavelet decomposition for time series prediction. *Neurocomputing*, **48**(1-4), 267–277.

Stark, H. and Woods, J. (2011) *Probability, Statistics, and Random Processes for Engineers*, Prentice Hall, Englewood Cliffs, NJ.

Stroeer, A., Cannizzo, J.K., Camp, J.B. and Gagarin, N. (2009) Methods for detection and characterization of signals in noisy data with the Hilbert-Huang transform. *Physical Review D*, **79**(12), 124022 1–10.

Takens, F. (1981) Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, **898**, 366–381.

Tay, F.E.H. and Cao, L.J. (2002) Modified support vector machines in financial time series forecasting. *Neurocomputing*, **48**(1), 847–861.

Teräsvirta, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**(425), 208–218.

Tian, Q., Price, N.D. and Hood, L. (2012) Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *Journal of Internal Medicine*, **271**(2), 111–121.

Tiao, G.C. and Tsay, R.S. (1994) Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, **13**(2), 109–131.

Tong, H. (1990) *Non-Linear Time Series: A Dynamical System Approach*, Clarendon Press, Oxford, UK.

Van der Merwe, R. and Wan, E. (2003) Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models, in *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal*, IEEE, Vancouver, BC, Canada, pp. VI-701-4.

Wan, E.A. and van der Merwe, R. (2000) The unscented Kalman filter for nonlinear estimation, in *The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, IEEE, Lake Louise, Alberta, Canada, pp. 153–158.

Wang, J.M. Fleet, D.J. and Hertzmann, A. (2008) Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, 283–298.

Wang, X., Ma, J.J., Wang, S. and Bi, D.W. (2007) Time series forecasting energy-efficient organization of wireless sensor networks. *Sensors*, **7**(9), 1766–1792.

Wang, Z., Bukkapatnam, S.T.S., Kumara, S.R.T., Kong, Z. and Katz, Z. (2014) Change detection in precision manufacturing processes under transient conditions. *CIRP Annals - Manufacturing Technology*, **63**, 449–452.

Wang, Z., Liu, X., Liu, Y., Liang, J. and Vinciotti, V. (2009) An extended Kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series. *IEEE Transactions on Computational Biology Bioinformatics*, **6**(3), 410–419.

Weron, R. and Misiorek, A. (2008) Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, **24**(4), 744–763.

Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. (1985) Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, **16**(3), 285–317.

Xu, X., Qi, Y. and Hua, Z. (2010) Forecasting demand of commodities after natural disasters. *Expert Systems with Applications*, **37**(6), 4313–4317.

Yadav, P.K., Pope, P.F. and Paudyal, K. (1994) Threshold autoregressive modeling in finance: the price differences of equivalent assets. *Mathematical Finance*, **4**(2), 205–221.

Yang, H., Bukkapatnam, S.T.S. and Barajas, L.G. (2011) Local recurrence based performance prediction and prognostics in the nonlinear and nonstationary systems. *Pattern Recognition*, **44**(8), 1834–1840.

Yang, H., Bukkapatnam, S.T.S. and Komanduri, R. (2007) Nonlinear adaptive wavelet analysis of electrocardiogram signals. *Physical Review E*, **76**(2), 026214: 1–8.

Yang, P., Wang, G., Bian, J. and Zhou, X. (2010) The prediction of nonstationary climate series based on empirical mode decomposition. *Advances in Atmospheric Sciences*, **27**(4), 845–854.

Yankov, D., DeCoste, D. and Keogh, E. (2006) Ensembles of nearest neighbor forecasts. *Lecture Notes in Computer Science*, **4212**, 545–556.

Yee, P. and Haykin, S. (1999) A dynamic regularized radial basis function network for nonlinear, nonstationary time series prediction. *IEEE Transactions on Signal Processing*, **47**(9), 2503–2521.

Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S. and Badr, S. (1993) The occurrence of sleep-disordered breathing among middle-aged adults. *New England Journal of Medicine*, **328**(17), 1230–1235.

Yu, C., He, Y. and Quan, T. (2008) Frequency spectrum prediction method based on EMD and SVR, in *Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications*, IEEE, Kaohsiung, Tawain, pp. 39–44.

Zhang, G.P. (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159–175.

Zhang, G.P. (2012) Neural networks for time-series forecasting, in *Handbook of Natural Computing*, Rozenberg, G., Bäck, T. and Kok, J. (eds), Springer, Berlin, Germany, pp. 461–477.

Zhang, G.P. and Berardi, V.L. (2001) Time series forecasting with neural network ensembles: an application for exchange rate prediction. *Journal of the Operational Research Society*, **52**, 652–664.

Zhang, G.P. and Qi, M. (2005) Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, **160**(2), 501–514.

Zhang, L., Zhou, W. and Jiao, L. (2004) Wavelet support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **34**(1), 34–39.

Zhou, B., Shi, A., Cai, F. and Zhang, Y. (2004) Wavelet neural networks for nonlinear time series analysis, in *Advances in Neural Networks*, Yin, F.L., Wang, J. and Guo, C. (eds), Springer, Berlin/Heidelberg, pp. 430–435.

Zhou, Y., Zhang, T. and Li, X. (2006) Predicting nonstationary time series with multi-scale Gaussian processes model. *Lecture Notes in Computer Science*, **4232**, 513–520.

Zhu, Z., Sun, Y. and Li, H. (2007) Hybrid of EMD and SVMs for short-term load forecasting, in *Proceedings of the IEEE International Conference on Control and Automation*, IEEE, Guangzhou, China, pp. 1044–1047.

Biographies

Changqing Cheng is a visiting professor at Department of Industrial and Management Systems Engineering at the University of South Florida. He earned his Ph.D. degree in Industrial Engineering and Management from Oklahoma State University in 2013. His research mainly focuses on advanced analytics and quality monitoring of complex systems, with applications in ultra-precision nanomanufacturing processes and healthcare systems. He was the recipient of Oklahoma State University Outstanding Doctoral Student Award, IIE Research Conference Council of Industrial Engineering Academic Department Heads (CIEADH) Best Poster Award, and NSF-ASME Poster Symposium Scholarship.

Akkarapol Sa-ngasoongsong was born in Bangkok, Thailand. He received a B.Eng. degree in Electrical Engineering from Mahidol University, Thailand, and M.S. and Ph.D. degrees in Industrial Engineering and Management from Oklahoma State University, Stillwater, Oklahoma. His current research interests include advanced time series forecasting, data mining, and analytics. He was a recipient of two best paper awards, Gilbreth Memorial Fellowship, and John S.W. Fargher Scholarship from

the Institute of Industrial Engineers. He was also a recipient of IEM doctoral student award, Distinguished Graduate Fellowship, and a finalist of Phoenix Award from Oklahoma State University.

Omer Beyca is an Assistant Professor in the Industrial Engineering Department at Fatih University, Turkey. He received his B.S. degree from Fatih University and Ph.D. degree from Oklahoma State University in Industrial Engineering and Management. His research interests include dynamic modeling of nonlinear systems, quality control in micromachining, and sensor-based modeling.

Trung Q. Le is an Assistant Research Scientist at the Center for Remote Healthcare Technologies of Texas A&M University. He is also a fellow of the Vietnam Education Foundation. His research interests are wireless wearable sensors and systems, biomedical signal processing, and nonstationary and nonlinear dynamic analysis with a focus on improving the monitoring, diagnostics, and prognostics of cardiovascular and neurological disorders. He has published his research in *IEEE Transactions on Biomedical Engineering, IEEE Journal of Translational Engineering in Health and Medicine*, and *Medical Engineering & Physics Mechanical*. He is a member of IIE, INFORMS, and IEEE.

Hui Yang is an Assistant Professor of Industrial and Management Systems Engineering and the Director of Complex Systems Monitoring, Modeling and Analysis Laboratory at the University of South Florida. He received his B.S. and M.S. degrees in Electrical and Computer Engineering with honors from China University of Mining and Technology (Beijing) in 2002 and 2005, respectively, and his Ph.D. degree in Industrial Engineering and Management from Oklahoma State University in 2009. His research focuses on sensor-based modeling and analysis of complex systems for process monitoring/control, system diagnostics/prognostics, quality improvement, and performance optimization, with special focus on nonlinear stochastic dynamics and the resulting chaotic, recurrence, multifractal, self-organizing behaviors. His research received several best paper awards such as the IBM Best Paper Award in 2011 IEEE Annual Conference of Engineering in Medicine and Biology Society and the IERC Conference Best Paper Awards (2009, 2010, 2014). Currently, he is a board member of the Data Mining section and Quality, Statistics and Reliability section of INFORMS, Also, he is the guest editor for *IEEE Intelligent Systems and Information Systems* and *e-Business Management* (Springer) and a professional member of IEEE, IIE, AHA, ASEE, and INFORMS.

Zhenyu "James" Kong earned his Ph.D. in Industrial and Systems Engineering from the University of Wisconsin–Madison in 2004 and his bachelor's and master's degrees in Mechanical Engineering from the

Harbin Institute of Technology in China in 1993 and in 1995, respectively. He joined the Department of Industrial and System Engineering at Virginia Tech as an Associate Professor with tenure in August 2013. He started his academic career at Oklahoma State University's School of Industrial Engineering and Management in August 2006. He was promoted to Associate Professor with tenure in 2012. Prior to his stint at Oklahoma, he spent 2 years as a senior research engineer at Dimensional Control Systems, Inc., of Troy, Michigan. His research interests are in the monitoring and control of micro/nano manufacturing operations via MEMS sensor-based predictive modeling. He also focuses on compressive sensing-based modeling, synthesis, and diagnosis for large and complex manufacturing systems. His research has been sponsored by the National Science Foundation, the Department of Transportation, and the National Institute of Standards and Technology. He was awarded the Halliburton Outstanding Faculty Award from the College of Engineering, Architecture and Technology at Oklahoma State University in 2013.

Satish Bukkapatnam serves as a Rockwell International Professor of Industrial and Systems Engineering at Texas A&M University with a joint (courtesy) appointment in the Department of Biomedical Engineering and the director of Texas Engineering Experimentation Station Institute for Manufacturing Systems. He has previously served as an AT&T Professor at the Oklahoma State University and as an Assistant Professor at the University of Southern California. His research addresses the harnessing of high-resolution nonlinear dynamic information, particularly from wireless MEMS sensors, to improve the monitoring and prognostics, mainly of ultraprecision and nanomanufacturing processes and machines and cardiorespiratory processes. His research has led to 125 peer-reviewed publications (73 published/accepted in journals and 52 in conference proceedings) and five pending patents and has been the basis for 10 Ph.D. dissertations. His research has received support from federal agencies including the National Science Foundation, Department of Energy, and Department of Defense and the private sector including General Motors, Ford, National Instruments, and the Central Rural Electric Cooperative. His work with his students has received over a dozen best paper/poster/innovation awards, and his contributions have been recognized with Oklahoma State University Regents Distinguished Researcher (2011), Halliburton Outstanding College of Engineering Faculty (2011 and 2012), the Institute of Industrial Engineers (IIE) Eldin Outstanding Young Industrial Engineer (2012), IIE Boeing Technical Innovation (2014), and the Society of Manufacturing Engineers (SME) Dougherty Outstanding Young Manufacturing Engineer (2005) awards. He received his master's degree and Ph.D. from the Pennsylvania State University and undergraduate degree from S.V. University, Tirupati, India.