Computers in Industry 64 (2013) 625–633

Contents lists available at SciVerse ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Analytical approach to similarity-based prediction of manufacturing system performance

Alexander Bleakie, Dragan Djurdjanovic *

University of Texas at Austin, Department of Mechanical Engineering, Austin, TX, United States

ARTICLE INFO

Article history:
Received 23 April 2012
Received in revised form 27 February 2013
Accepted 28 February 2013
Available online 18 April 2013

Keywords:
Dynamic system performance prediction
Time series modeling
Gaussian mixture model
Weighted likelihood estimation

ABSTRACT

In this paper, a new method is proposed that is capable of predicting system condition by comparing the similarity of the most recent performance signatures with the known degradation patterns available in the historical records. For predicting the future performance, the similarities of the current performance signatures to each known degradation pattern are utilized in an analytically tractable manner to slant the prediction distributions toward most similar past degradation patterns. The newly proposed method was applied to prediction of sensor signatures coming from an industrial plasma enhanced chemical vapor deposition (PECVD) tool operating in a major semiconductor manufacturing fab. Results showed that the proposed method significantly improves the long-term time series prediction accuracy in terms of mean squared errors over the traditional autoregressive moving average (ARMA) model and additionally showed comparable mean squared prediction errors to another recently introduced similarity-based algorithm for long-term prediction of non-linear and non-stationary time series. However, the analytical structure of the method proposed in this paper enables computation of the prediction distributions an order of magnitude faster.

1. Introduction

Degradation and failure of machines and products occur in many settings every day, leading to significant cost, wasted material, and hazardous environments to employees and consumers. In 1981, maintenance costs in the United States were estimated at $600 billion, a figure that doubled in the subsequent 20 years, with an estimated one-third of these costs wasted through ineffective maintenance and unexpected failures [1]. These losses are one of the key motivating factors driving condition-based and predictive maintenance research. These efforts pursue maintenance policies based on the current and predicted system conditions, as assessed based on the readings from sensors mounted on the monitored system.

The prediction accuracy of degrading performance, particularly in the long-term, is an essential part of having cost effective maintenance policies [2–4]. Predictive condition information about the monitored system allows one to predict its future degradation state, failure modes, and possibly remaining useful life (RUL), thus possibly enabling one to avoid sudden unexpected failures and allow optimal maintenance strategies [5,6]. In particular, accurate predictions of the long-term performance of degrading equipment are necessary in order to have sufficient time to prepare maintenance operations [7,8]. Hence, long-term prediction accuracy of the time-series' of sensory signatures indicative of the condition of the monitoring process plays an important role for implementing predictive maintenance in many manufacturing industries.

The condition of the monitored system is usually not directly observable and has to be inferred from the available sensor readings [8]. For example, small cracks in pipes and vessels of a chemical plant cannot be directly measured and their presence and severity need to be inferred from the strategically distributed sensing of flows and pressures across the system [9]. Generally, the degradation process in a dynamic system is associated with a set of features that can be extracted from the available sensor readings. In that context, degradation prediction boils down to prediction of the time-series of sensory features indicative of the system condition, such as vibration levels, forces, thermodynamic states, etc. [9]. Prediction of the time-series of features then allows one to predict future degradation patterns, fault modes, RUL, or probabilities of unacceptable behavior of the system over time.

* Corresponding author. Tel.: +1 512 232 0815.
E-mail addresses: a_bleakie@mail.utexas.edu (A. Bleakie), dragand@me.utexas.edu (D. Djurdjanovic).

0166-3615/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compind.2013.02.013

Numerous techniques for time-series prediction exist and great importance is placed on the accuracy and computation complexity of the method in general situations. One can separate these techniques into three general types: physics-based models, linear data-driven models and non-linear data-driven models.

Physical models predict the future behavior of a system by utilizing first-principle theories tested through years of application [9,10]. Identification of constituent equations and model parameters is not a trivial task for any real world system. In addition, for manufacturing equipment and other complex systems, there are sometimes thousands of interacting components with highly variable geometries and uncertain forces affecting the system [9]. These conditions and requirements bring inherent problems to physical models, such as long computation due to large number of states, numerical instability, and inaccuracies when compared to the real world system due to large number of estimated parameters and geometries. One can however, still use a detailed physical model to see how certain parameters shift over time or how disturbances affect the system, and then simplify the model as needed for predictive maintenance purposes [10].

Even though there is no replacement for an accurate underlying physical model capturing relevant degradation phenomena, various linear and non-linear data-driven models can be used in cases where the underlying physics is uncertain and/or requires too much computation to obtain accurate predictions.

Linear prediction techniques include multiple-linear regression [11], least-squares regression [11], auto-regressive moving average (ARMA) models [12], and Kalman filters [13] in the presence of stationary noise processes. These methods work well for short-term predictions due to the assumptions that the data is generated from a linear system. Unfortunately, any real process with non-stationarities or non-linearities in the time-series will cause the prediction errors to increase drastically.

Non-linear parametric techniques also include numerous methods, such as non-linear regression [14], fuzzy ARMA modeling [15,16], Bayesian curve fitting [17], wavelet methods [18], support vector machine [19], probability density functions [17,20] and neural networks based techniques [22–24]. These methods have been shown to give better results in terms of modeling and predicting non-stationary and non-linear time series. However, compared to linear methods, they usually require much more computational effort in order to identify the time-series model, as well as to obtain prediction results from that model.

For both linear and non-linear data-driven approaches, selection of an appropriate model form is an important and difficult task. As an example, predictions of probability density functions (pdf) of future signatures can be accomplished by parameterizing them using Gaussian mixture models [21] or kernel density estimation [17]. Zivkovic follows that approach in [20] offering a recursive finite mixture model algorithm for predicting multidimensional features that automatically selects the number of components in the mixture model. Bayesian curve fitting gets around this by allowing a distribution of solutions to be generated around one model structure [17].

Neural network based prediction methods represent a rich family of non-linear data-driven time-series prediction approaches, with tremendous applicability to modeling and prediction of non-linear and non-stationary time-series [22]. The topology of interconnections among neurons, number of layers in the network, the number of neurons in each layer, and the type of transfer functions between neurons define the structure of these essentially non-linear parametric models, and commonly we see this structure determined via ad hoc selection by the user. Each neuron has a transfer function (linear, Gaussian, sigmoidal, etc. [22]), with parameters of the neural function and inter-neural weights determined via a gradient descent based training algorithm [23]. The most common type of neural network used for time-series prediction is the so-called Recurrent Neural Network (RNN) [23,24]. RNNs use internal delayed feedback links to take into account temporal dependencies in the data, which enables them to approximate a wide class of non-linear dynamic systems [24]. The gradient descent training algorithms of these networks however have certain problems when dealing with long-term time dependencies, which can limit their accuracy. Also, choosing the number of layers, number of neurons, and feedback structure is still an active area of research [22].

The similarity-based methods for time-series prediction represent another powerful family of non-linear data-driven time-series prediction methods. These methods make forecasts using a similarity metric of the currently observed time-series realizations with historical records of prior realizations of those time-series. In terms of performance prediction in condition-based maintenance, these approaches predict the future degradation states of the monitored system using time-series corresponding to past degradation patterns and fault modes of the monitored system. Mahalanobis and Euclidean distances [11] have been used extensively for finding the similarity between vector trajectories in the multidimensional space in which the time-series resides [8,25–28]. Similarity based approaches are of particular importance in manufacturing applications. Namely, in manufacturing environments, systems experience multiple degradation cycles, failures, and maintenance events in their lifetimes [8,31]. Signatures extracted from sensor readings over the life of a monitored system yield vital historical databases that can be used to predict the performance of the upcoming operations [8,31].

Liu et al. [8] used similarity measures of past time-series trajectories and heuristically inspired methods to weight the forecasts toward the most similar time-series trajectories observed in the past. This method was shown to have much lower long-term mean prediction errors than ARMA and RNN models. Wang et al. [28] offer another example of similarity-based time-series prediction methods. The authors accomplish time-series prediction using similarity measures and optimized alignment of the currently observed signature patterns with the past patterns of the time-series. The method was demonstrated in RUL prediction based on signatures emitted by turbo fans and was shown to improve the long-term predictions over an existing exponential curve fitting procedure. While these methods achieve long-term accuracy, it is at the cost of long computation times, which hampers prediction in very highly dimensional spaces, and the prediction of fast degradation processes.

The review of available literature implies that long-term prediction of the most complicated time-series can be accomplished utilizing some similarity measure between the current and past time-series behavior, even though the benefits of such approaches come with a prohibitively high computation cost. The goal of the method developed in this paper is to achieve long-term prediction accuracy of performance signatures by incorporating the similarity of historical degradation processes, while at the same time achieving a level of analytical tractability that accelerates the process of postulating and updating such similarity based predictions.

The rest of the paper is organized as follows. In Section 2, the new similarity-based prediction method is introduced. In Section 3, we offer results of the newly proposed approach applied to the prediction of signatures from an industrial semiconductor plasma enhanced chemical vapor deposition (PECVD) tool. Section 4 provides conclusions and possible avenues of future work.

2. Methodology

2.1. Method overview

Let us first introduce the following terminology that will be used throughout the paper:

- The term *feature vector* is used to indicate the signatures extracted from the raw sensors mounted on the monitored system that are known to characterize the condition of that system. Evolution of these signatures is then indicative of system degradation and their behavior needs to be predicted.
- The term *cycle* is used to indicate a single operation by the system, emitting a single feature vector. This can be any manufacturing operation or single use of a product. Repeated cycles degrade the system, causing the signatures to evolve and ultimately lead to maintenance events.
- The term *run* is used to indicate the time interval between two consecutive maintenance events. These maintenance events can be component replacements, repairs, cleaning, etc. Thus, a time-series of feature vectors in a past run represents a particular degradation trajectory known from historical data.

The method described in this paper is relevant for any dynamic system that has existing historical records of signal features from either past runs containing degradation patterns/failure modes or physical models containing degradation dynamics. It is assumed that the degradation process is described by a time-series of feature vectors extracted from the sensors relevant to the behavior of the dynamic system. The complexity of the degradation and failure modes of a system, such as wear, aging, pitting, cracking, corrosion and other mechanisms [9] often makes the resulting time-series of features indicative of system degradation non-stationary and non-linear in their dynamics. For example, Fig. 1 shows the possible complexity of six degradation trajectories using features from a simulation of a hypothetical non-linear two-output system [9].

As mentioned in the introduction, similarity-based methods with an underlying non-linear parametric model can be effectively utilized to account for non-linear, non-stationary degradation dynamics and thus result in high long-term prediction accuracy. This involves the comparison of the large number of historical runs and combining that information into a dynamic model that can predict the current run's degradation process.

The method described in this paper follows this paradigm and uses pdf-s representing feature vectors from the previous runs as the models for evaluating similarities between the newly observed

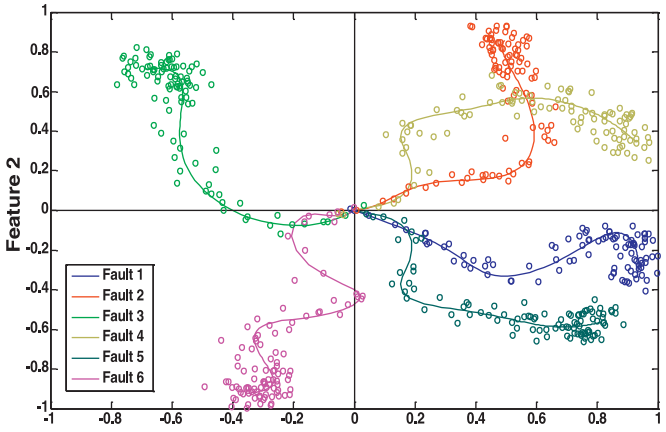

Fig. 1. Conceptual illustration of degradation dynamics for a system with six different fault modes, monitored using two features.

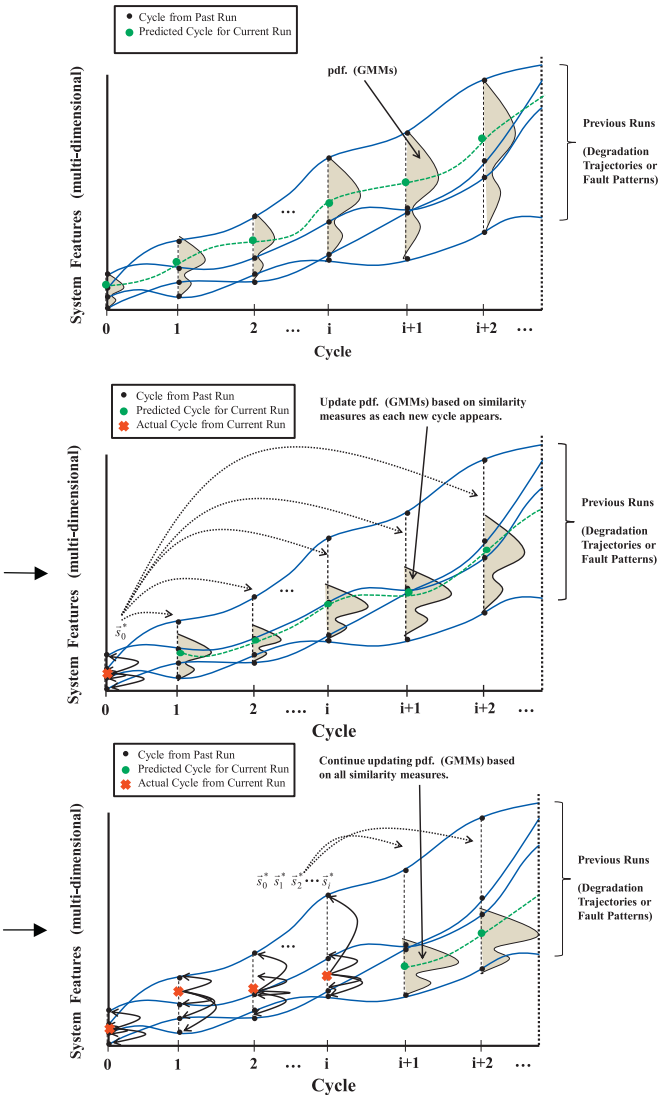

Fig. 2. Steps for the similarity-based prediction algorithm described in this paper.

trajectory of signatures and those observed in the past. The pdf-s are approximated using Gaussian mixture models (GMM) due to their ability to model any distribution within a desired accuracy, given enough Gaussian components [21]. Fig. 2 conceptually describes the newly proposed time-series prediction algorithm. Right after a maintenance operation (i.e. just before a new run starts), the only information that we have about the run that is about to start are feature vector realizations observed during the previous runs of the monitored system. At each cycle, GMMs of feature vectors corresponding to that cycle in the previous runs can be formulated. As the current run progresses, feature vectors from more and more cycles are observed and similarity measures between those feature vectors and all feature vectors corresponding to that cycle in past runs can be evaluated. Consequently, at cycle number i, we can observe the set of vectors $\vec{s}_c^*$; $c \in \{0, 1, 2, \ldots, i\}$, composed of similarities between feature vectors observed at cycle number c, with feature vectors observed at cycle c in all previous runs.[1] These similarity measures can then be used to skew the GMMs of feature vectors corresponding to future cycles of the current run (cycles $i+1$, $i+2$,...) towards feature

[1] Each vector $\vec{s}_c^*$ will have as many elements as there are past runs that performed c cycles.

vectors from previous runs that in the past cycles showed more similarity with the current run. Thus, as time progresses and more and more signatures are collected during the current run, the feature models shift toward the most similar runs observed in the past. When the current run is completed, it can be incorporated into the library of previous runs, thus enabling continuous learning as the system progresses through its lifetime.

Details of the prediction method conceptually described above will be given in the remainder of this section, which is organized as follows. Section 2.2 discusses the calculation of the similarity measures used in updating of the GMMs. Sections 2.3 and 2.4 review the method of modifying the log-likelihood function of a pdf in order to handle weighted data points and also discusses the modifications to the well-known expectation–maximization (EM) algorithm [36] used to fit a GMM to the library of past feature vectors, while incorporating the similarity measures.

2.2. Concept of similarity vectors

Let us assume that the monitored system has just performed cycle i of the current run and let $\vec{x}_{i,\text{current}}$ denote the corresponding feature vector consisting of D components (i.e., feature vectors are of dimensionality D). Given K previous runs in the historical database, one can compute distance measures, d_{ik}, between the current feature vector and the ith feature vectors $\vec{x}_{ik}$ from all previous runs k, $k \in \{1,2,3,\ldots,K\}$. Following Liu et al. [8], in this paper we use the Mahalanobis distance

$$d_{ik} = \sqrt{(\vec{x}_{ik} - \vec{x}_{i,\text{current}})^{\mathrm{T}} \bar{\Sigma}^{-1} (\vec{x}_{ik} - \vec{x}_{i,\text{current}})} \tag{1}$$

where the scaling covariance matrix $\bar{\Sigma}$ is characteristic of the normal operation of the monitored system, i.e. estimated from feature vectors observed soon after maintenance operations [4]. Depending on the application and topology of the feature space, other distance measures could be used [11]. The similarity s^*_{ik} of the ith cycle of the current run with the kth previous run can be expressed as

$$s^*_{ik} = \exp(-d_{ik}) \tag{2}$$

which gives values close to one when the Mahalanobis distance between the two feature vectors is small, and approaches zero as the distance grows large.

The K similarities can be combined into a vector $\vec{s}^*_i = \left[s^*_{i1} \quad s^*_{i2} \quad \cdots \quad s^*_{ik} \right]^{\mathrm{T}}$, containing similarities for cycle i in the current run. Observations from Liu et al. [8] and our own experiments showed that of the similarity vectors $\vec{s}^*_i$ tend to be rather noisy and a filtering step is needed before those similarities can be used for updating of predictions. In this paper, an exponentially weighted moving average (EWMA) [32] with smoothing factor λ was used as a simple way to combine all of the similarity vectors $\vec{s}^*_0 \ldots \vec{s}^*_i$ into an *overall similarity vector*, $\vec{s}_i$, that can be used in predicting future cycles of the current run (i.e. feature vectors for cycles $i+1$, $i+2$, …). The overall similarity vector $\vec{s}_i$ contains filtered information about similarities between all the past cycles of the current run and corresponding cycles in the past runs of the monitored system, with similarities corresponding to the most recent cycles being more emphasized by the EWMA filter. Based on these concepts of feature comparison between the past and current degradation trajectories (runs) of the monitored system, we can now describe the use of the similarity vectors $\vec{s}_i$ for prediction of future feature vectors for the current run.

2.3. Weighted likelihood estimation of mixture Gaussians with similarity vectors

Once again, let us assume that the monitored system just completed cycle i of the current run and that a filtered vector of similarities, $\vec{s}_i$, has been obtained using methods described in Section 2.2. At any future cycle (i.e. cycle, $i+1$, $i+2$, …), a multi-dimensional GMM, $f(\vec{x})$, of feature vectors will be pursued in the form [21]

$$f(\vec{x}) = \sum_{m=1}^{M} w_m N(\vec{x} | \vec{\mu}_m, \bar{\Sigma}_m) \tag{3}$$

where M is the total number of Gaussian components in the GMM, w_m is the mixture weight corresponding to the mth component of the GMM and

$$N(\vec{x} | \vec{\mu}_m, \bar{\Sigma}_m) = \left(\frac{1}{(2\pi)^{D/2} |\bar{\Sigma}_m|^{1/2}} \right) \exp\left[-\frac{1}{2} (\vec{x} - \vec{\mu}_m)^{\mathrm{T}} \bar{\Sigma}_m^{-1} (\vec{x} - \vec{\mu}_m) \right] \tag{4}$$

denotes the mth Gaussian component with mean $\vec{\mu}_m$ and covariance matrix $\bar{\Sigma}_m$. The reader should be reminded that D is the dimensionality of the feature vectors and thus represents the dimensionality of the GMMs describing feature distributions at each cycle.

Let us assume we need to fit a GMM to the observed feature vectors $\vec{x}_k$; $k \in \{1,2,\ldots,K\}$. It is easy to show that the log-likelihood $\ln Pr(X|\theta)$ of $X = \{\vec{x}_1 \ldots \vec{x}_K\}$ is

$$\ln Pr(X|\theta) = \sum_{k=1}^{K} \ln[f(\vec{x}_k)] \tag{5}$$

In order to fit a GMM, this function is commonly maximized by a gradient ascent algorithm that seeks parameters $\theta = \{ \vec{\mu}_m \quad \bar{\Sigma}_m \quad w_m \}$, $m = 1, \ldots, M$, that maximize the likelihood (5) for the given data X [21].

The log-likelihood function of a pdf can always be modified with a prior distribution if one has information about the system behavior. Since, we have prior information about system behavior from the previous runs, we will use this opportunity to skew the GMMs of future cycles of the current run towards corresponding cycles of the previous runs that resemble the current run the most.

Weighted Likelihood Estimation (WLE) is a Bayesian estimation methodology based on modifying the likelihood function of a pdf with some weighting distribution, vector, or observation-dependent function that is a priori known [33,34]. Various applications in statistical analysis, machine learning, medicine, and gambling have used these techniques to obtain better prediction accuracy from many probabilistic model forms [34,35]. One of the most common formulations of WLE is to assign each feature vector a unique weight α_k based on the prior information about the data (higher α_k meaning higher prior probability for a given observation). The log-likelihood function of the data X for this type of a WLE takes the following form

$$\begin{aligned} \ln Pr(X|\theta)_{\text{WLE}} &= \sum_{k=1}^{K} \ln\left[(f(\vec{x}_k))^{\alpha_k} \right] \\ &= \sum_{k=1}^{K} \alpha_k \ln\left[\left(\sum_{m=1}^{M} w_m N(\vec{x}_k | \vec{\mu}_m, \bar{\Sigma}_m) \right) \right] \end{aligned} \tag{6}$$

In our context, the weight α_k is the kth component of the overall similarity vector $\vec{s}_i$ at cycle i. In such a way, the similarities with previous runs can be used to modify the log-likelihood function in a way that emphasizes the most similar previous runs and vice versa, thus increasing the prediction accuracy. Based on the definitions of the WLE of a GMM, one can now state the necessary

modifications for the gradient-based EM algorithm used for fitting the GMM parameters.

2.4. Expectation–maximization algorithm modifications for the weighted likelihood estimation of Gaussian mixtures with similarity vectors

The so-called expectation–maximization (EM) algorithm is commonly utilized for fitting GMM parameters [36]. The EM algorithm is an iterative method for estimating the maximum likelihood solution for a set of parameters in a statistical model. In this subsection, we will introduce a modification of this algorithm, which will enable the inclusion of similarities of the current run with the pervious runs into the estimation process.

Let us assume that a feature distribution for any given cycle in the current run is composed of M Gaussians[2] and that we have the corresponding initial guesses for the M mean vectors $\vec{\mu}_m$, and M covariance matrices $\bar{\Sigma}_m$ [21]. The procedure for finding the GMM components $\theta = \{\, \vec{\mu}_m \quad \bar{\Sigma}_m \quad w_m \,\}$, $m = 1, \ldots, M$ consists of two steps. In the first step, referred to as the Expectation calculation, the likelihood that a given feature vector in the dataset X belongs to each GMM component is evaluated. The next step, referred to as Maximization, updates the GMM parameters in order to maximize the likelihood function of the data, given the probabilities from the Expectation step [36]. The algorithm repeats the Expectation and Maximization steps until convergence, with possible convergence criteria being the number of Expectation-Maximization iterations, bounds on the change in the log-likelihood function, or the change in GMM parameters [21].

In our work, the WLE log-likelihood function (6) replaces the standard formulation and the EM process will be consequently modified by maximizing (6) with respect to $\theta = \{\, \vec{\mu}_m \quad \bar{\Sigma}_m \quad w_m \,\}$, $m = 1, \ldots, M$. For a given set of GMM parameters, the Expectation step can be performed for a GMM in standard fashion, by taking the weighted sum of the different Gaussian component contributions to each feature vector

$$Pr(m|\vec{x}_k, \theta) = \frac{w_m N(\vec{x}_k|\vec{\mu}_m, \bar{\Sigma}_m)}{\sum_{m=1}^{M} w_m N(\vec{x}_k|\vec{\mu}_m, \bar{\Sigma}_m)} \tag{7}$$

where $Pr(m|\vec{x}_k, \theta)$ is the probability that feature vector k is produced by the Gaussian component m. In order to obtain the maximum likelihood parameters, $\hat{\theta}_{ML}$, for the model, one must maximize the log-likelihood function with respect to the model parameters:

$$\begin{aligned}\hat{\theta}_{ML} &= \arg\max_{\theta}\{\ln Pr(X|\theta)_{\mathrm{WLE}}\} \\ &= \arg\max_{\theta}\left\{\sum_{k=1}^{K}\alpha_k \ln\left[\left(\sum_{m=1}^{M} w_m N(\vec{x}_k|\vec{\mu}_m, \bar{\Sigma}_m)\right)\right]\right\}\end{aligned} \tag{8}$$

Based on the iterative procedure proposed in [21] for the EM based identification of GMM parameters, the EM algorithm for the modified likelihood function (6) yields the following iterative procedure for finding GMM parameters:

At each iteration l, the current solution for the GMM parameters $\theta^l = \left\{\, \vec{\mu}_m^{\,l} \quad \bar{\Sigma}_m^{\,l} \quad w_m^l \,\right\}$, $m = 1, 2, \ldots, M$ is transformed into the next iterative approximation of the solution $\theta^{l+1} = \left\{\, \vec{\mu}_m^{\,l+1} \quad \bar{\Sigma}_m^{\,l+1} \quad w_m^{l+1} \,\right\}$, $m = 1, 2, \ldots, M$ according to formulae

$$\vec{\mu}_m^{\,l+1} = \frac{\sum_{k=1}^{K} \vec{x}_k \alpha_k Pr(m|\vec{x}_k, \theta^l)}{\sum_{k=1}^{K} \alpha_k Pr(m|\vec{x}_k, \theta^l)} \tag{10}$$

$$\begin{aligned}\bar{\Sigma}_m^{\,l+1} = &\left\{\frac{\sum_{k=1}^{K} \alpha_k Pr(m|\vec{x}_k, \theta^l)}{\left[\sum_{k=1}^{K} \alpha_k Pr(m|\vec{x}_k, \theta^l)\right]^2 - \sum_{k=1}^{K}\left[\alpha_k Pr(m|\vec{x}_k, \theta^l)\right]^2}\right\}\sum_{k=1}^{K}(\vec{x}_k \\ &- \vec{\mu}_m^{\,l+1})(\vec{x}_k - \vec{\mu}_m^{\,l+1})^T \alpha_k Pr(m|\vec{x}_k, \theta^l)\end{aligned} \tag{11}$$

$$w_m^{l+1} = \frac{\sum_{k=1}^{K} \alpha_k Pr(m|\vec{x}_k, \theta^l)}{\sum_{m=1}^{M}\sum_{k=1}^{K} \alpha_k Pr(m|\vec{x}_k, \theta^l)} \tag{12}$$

where

$$Pr(m|\vec{x}_k, \theta^l) = \frac{w_m^l N(\vec{x}_k^{\,l}|\vec{\mu}_m^{\,l}, \bar{\Sigma}_m^{\,l})}{\sum_{m=1}^{M} w_m^l N(\vec{x}_k^{\,l}|\vec{\mu}_m^{\,l}, \bar{\Sigma}_m^{\,l})}$$

The expectation and maximization steps are repeated until one of the convergence criteria mentioned previously is met [21]. The modified EM algorithm for estimating GMM parameters based on the WLE formulation (6) is summarized in the pseudo code shown in Fig. 3.

In this paper, at any cycle i, the WLE-modified EM algorithm is used to rapidly estimate GMM parameters for feature vectors corresponding to cycles $i+1, i+2, \ldots$. When the feature for cycle i in the current run is observed, the similarity vector $\vec{s}_i$ is updated using the procedure described in Section 2.2 and the modified EM algorithm described above is invoked to rapidly update the GMMs for the cycles $i+1, i+2, \ldots$.

3. Application of the methods on an industrial tool

3.1. Description of the tool and process under study

The data set used to evaluate the performance of the newly proposed similarity-based time-series prediction method was gathered from a standard 300 mm plasma enhanced chemical vapor deposition (PECVD) tool operating in the facilities of a major domestic manufacturer of integrated circuits. PECVD tools are used for depositing thin films onto silicon wafer substrates, which is one of the crucial steps in manufacturing of microelectronic circuits and solar cells. It is the most common method for producing conductors and dielectrics with excellent film growth properties necessary for small chip components [37]. Inside a PECVD tool chamber, reactive gases pass over silicon wafers and are absorbed onto the surfaces to form a thin layer. The gases are excited into the state of energetic plasma through radio frequency (RF) electrical power, which eases the process of film deposition. More specifically, the plasma state allows the reactions to take place at lower temperatures, which is more suitable for integrated circuit fabrication on large silicon wafers.

A general PECVD tool is composed of a reaction chamber, RF plasma generation system, gas delivery system, wafer load locks, and a robotic arm to carry wafers to and from the tool. Fig. 4 shows a diagram of the main components of a PECVD tool. High frequency energy is generated by the RF generation system and is delivered to gasses in the chamber via two matching capacitors (so-called load and tune capacitors) that adapt their capacitances to maximize the power delivered to the chamber. The RF energy excites the flowing gas into the plasma state necessary for lower temperature depositions. The gas delivery system consists of mass flow controllers (MFCs) that control the delivery of each gas used in

[2] Literature has shown works that find optimal numbers of components in mixture models in an unsupervised manner, and therefore this assumption is not particularly restrictive [17,20].

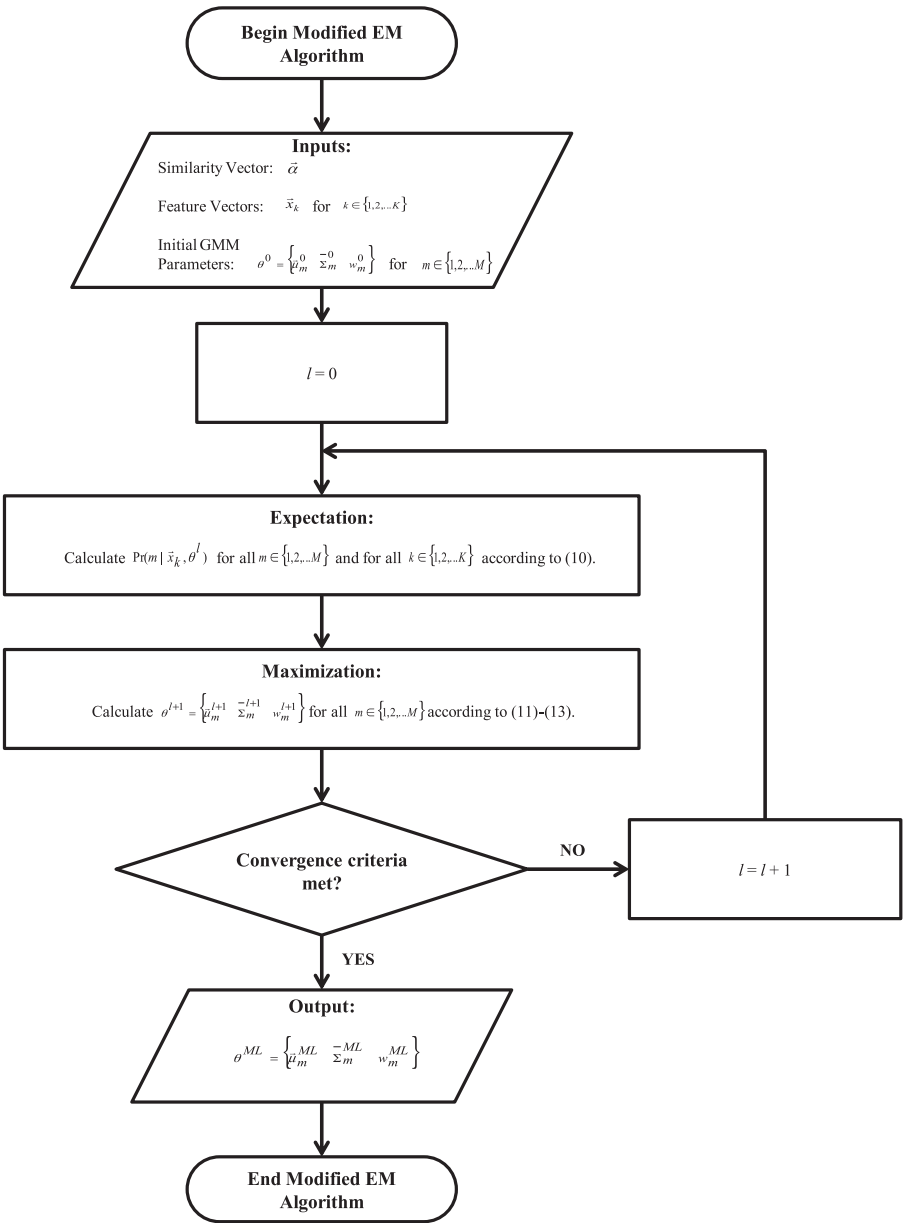

Fig. 3. Pseudo-code for the modified EM algorithm utilized in this work for GMM fitting.

various depositions. A control valve, often referred to as "pendulum valve," controls the chamber pressure and evacuates deposition gases from the chamber, while temperature controlled top and lower chamber plates enclose the chamber.

Various chemical compounds can be deposited using PECVD tools. Silicon nitride (Si–N) and silicon dioxide ($Si–O_2$) are some of the most common thin films deposited in these tools, even though other compounds can be used, depending on the conductivity, as well as mechanical and reliability requirements on the film [37].

In addition to the deposition cycles that contribute to the process of chip-making, PECVD tools in semiconductor manufacturing facilities also perform automatic in situ cleaning programs after a predetermined total film accumulation limit (corresponding to approximately 25–100 wafers, depending on the film chemistry and thickness). A common way of performing the in situ cleans is by flowing plasma-excited fluorine (F^-) into the chamber to eat away films deposited on the tool surfaces. These cleans are performed periodically in order to bring the tool back into a lower state of degradation. Unfortunately, this tool-cleaning procedure is imperfect, since residual films can be left in parts of the chamber, while at the same time, some tool surfaces can also be etched away during the process. This results in a long-term degradation of the tool, which over time leads to the production of wafers with noticeable defects, unless preventive maintenance (PM) actions are undertaken. Thus, besides the short-term accumulation drifts caused by successive wafer depositions and remedied via in situ cleans, one can also observe a long-term degradation of the tool condition as numerous in situ clean cycles are executed.

The next section details the dataset and features that were extracted from sensor readings to yield the multivariate time-series on which the newly introduced time-series prediction methodology was applied.

3.2. Description of the dataset and features extracted from sensor readings

In this study, we focused on roughly 8 months of production data and on the recipe (film thickness) that significantly dominated

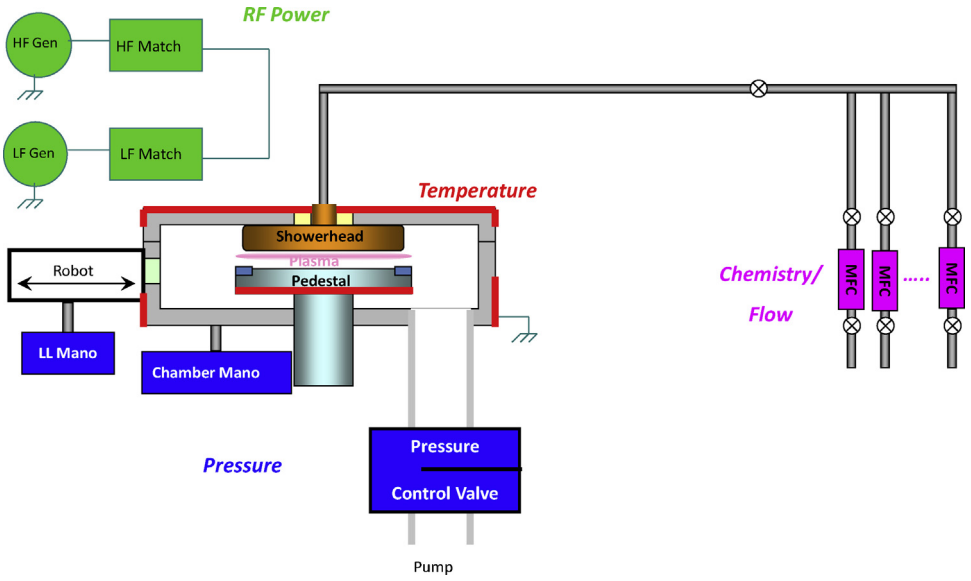

Fig. 4. Diagram representing the various components of the PECVD tool.

Table 1
Features extracted from the data analyzed in this study.

Signal	Signal features			
Top plate temperature	Mean	Minimum	Amplitude	
Chamber temperature	Mean	Minimum	Amplitude	
Pedestal 1 temperature	Mean minimum	Amplitude		
LF forward power	Steady state error	Tune time		
LF load power	Steady state error	Tune time		
LF reflected power	Steady state error	Tune time	Maximum	
HF forward power	Steady state error	Tune time		
HF load power	Steady state error	Tune time		
HF reflected power	Steady state error	Tune time	Maximum	
Load capacitor voltage	Steady state	Tune time	Overshoot high	Overshoot low
Tune capacitor Voltage	Steady state	Tune time	Overshoot high	Overshoot low
Pendulum valve angle	Steady state	Maximum		
Process chamber pressure	Steady state error	Rise time	Overshoot	Minimum
Liquid flow rate TEOS	Steady state error	Rise time	Overshoot	

the operations (about 80% of operations). The approximate size of the entire data set for the recipe considered here corresponds to about 40,000 wafers, with 1500 in situ cleans. Thus we have roughly 1500 runs and 40,000 cycles.

The signals used in this study were the RF power characteristics (forward, load, and reflected power), voltages of RF matching network capacitors, MFC flow rates, top plate temperature, chamber temperature, pedestal temperature, chamber pressure and the pendulum valve angle. All sensor readings were concurrently collected using a 10 Hz sampling rate, which is an order of magnitude higher than the prevalent 300 mm fab standards. The signals were obtained from a single tool continuously depositing recipes of $Si–O_2$ films with various thicknesses, while tetraethylorthosilicate (TEOS) was used as the main reactant.

Table 1 summarizes the 40 features that were extracted from PECVD tool sensor readings. Signal features that are the most descriptive of tool condition and degradation were then sought within this comprehensive feature set using the linear discriminant analysis (LDA) [14]. More specifically, LDA was utilized to extract 10 features showing the most significant statistical change before and after the in situ cleans. These features are listed in Table 2 and their evolution trajectories over multiple in situ cleans form the dataset on which the similarity-based time-series prediction method proposed in this work is applied. The weights of the principal discriminant vector corresponding to the top 10 most sensitive features are also listed in Table 2, indicating relative sensitivities of the 10 features listed in Table 2. All other features had significantly smaller weighting in the LDA principal discriminant vector [14] and showed no visible changes as depositions progressed. Our prior publication [38] discusses the physical reasoning of why these features appear to be most sensitive.

Fig. 5 shows three distinct evolution trajectories (runs) of two of the top sensitive features listed in Table 2 (the load capacitor steady state voltage and the top plate temperature mean). From

Table 2
Features showing the highest sensitivity to degradation between in situ cleans. They formulate the feature set used in validating the prediction methodology introduced in this paper.

Principal vector (w)	Feature name
0.60	Load capacitor overshoot low
0.41	Load capacitor overshoot high
0.40	Load capacitor steady state
0.27	HF load power steady state error
0.26	Top plate temperature mean
0.25	Top plate temperature minimum
0.22	Pedestal 1 temperature minimum
0.21	Pedestal 1 temperature mean
0.13	LF reflected power tune time
0.06	Process pressure minimum

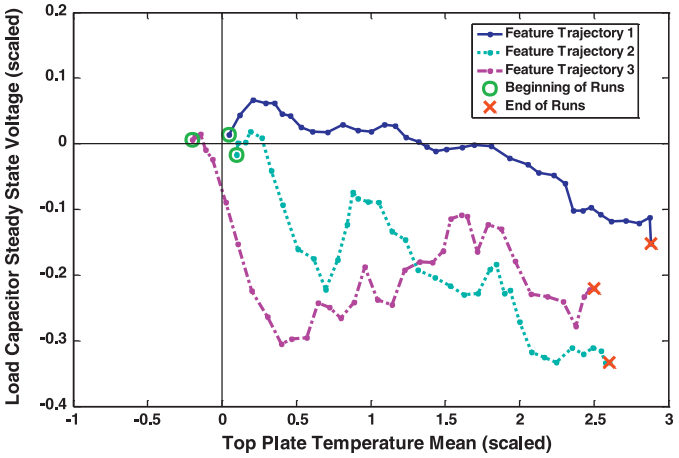

Fig. 5. Various degradation paths for three select runs of the PECVD tool using two sensitive dynamic features.

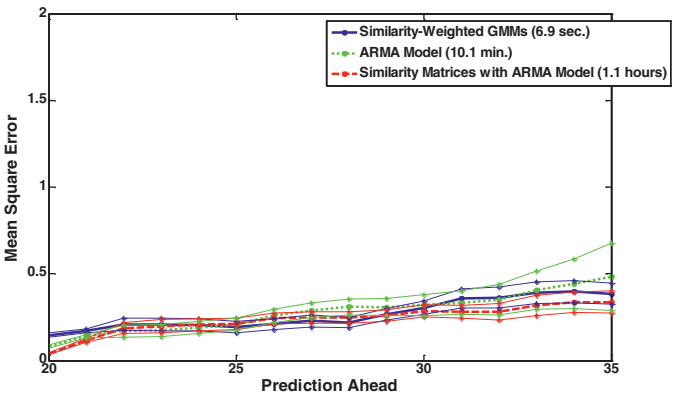

Fig. 7. Mean squared errors and computation time for the three prediction algorithms, starting at cycle 20 in each of the 40 test runs and predicting up to 35 cycles (wafers) ahead. Results were averaged over 40 test runs.

the figure, one can see that the trajectories through feature space can show vastly different behavior as the runs progress. In addition, Fig. 5 also shows obvious recoveries at the beginning of the runs, which all begin in similar areas of the feature space before progressing into different directions, indicating that various degradation patterns occurred during the course of many inter in situ clean runs observed in this data set. Thus, Fig. 5 illustrates that this dataset is a typical situation that necessitates the use of similarity based prediction methods.

3.3. Prediction results

The newly introduced similarity-based time-series prediction methodology was used to predict the behavior of the features listed in Table 2. The new method is compared against two other time-series prediction models – the standard auto-regressive moving-average model [12] and a recently introduced prediction method also based on the use of similarity matrices [8]. Mean squared errors and computation times associated with each method are used as comparison metrics. Signatures corresponding to wafer depositions from initial 30 runs were used to form the historical database of past runs, while signatures from the another 40 runs were used for testing of the prediction methods. Within each test run, predictions of feature vectors were made up to 35 cycles (wafers) ahead, after which the squared prediction errors and times associated with computation of all the predictions were averaged over the 40 test runs. Fig. 6 shows the results for mean

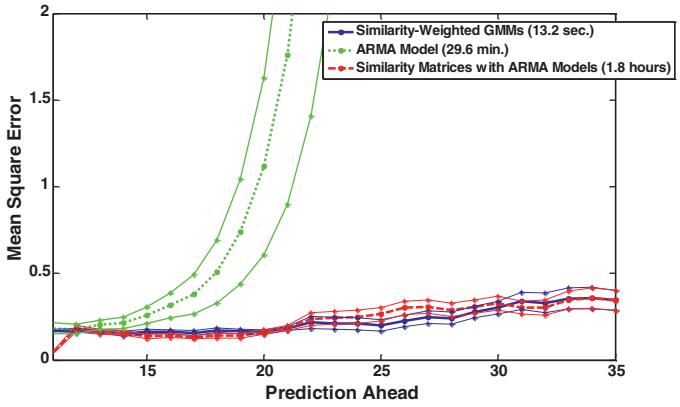

Fig. 6. Mean squared errors and computation time for the three prediction algorithms, starting at cycle 11 in each of the 40 test runs and predicting up to 35 cycles (wafers) ahead. Results were averaged over 40 test runs.

square errors and computation times for each of the three algorithms for the case when predictions were made starting at cycle (wafer) 11 in each of the test runs. Fig. 7 presents the same metrics for the case when predictions were made starting at cycle 20 in each of the test runs.

One can clearly see that ARMA prediction errors grow rapidly as the prediction horizon increases in Fig. 7. The reason is that ARMA prediction uses only dynamics of the features observed in the run up to that cycle and does not fully exploit the rich information available in the previous run history. The newly proposed method and the similarity matrix based approach from [8] both have much smaller errors and are comparable with each other in terms of prediction accuracy and confidence. However, the weighted GMM based algorithm introduced in this paper takes an order of magnitude less time to compute the results than the other two methods (minutes instead of hours necessary to obtain prediction results for all 40 test runs). When predictions start from further ahead in the run, the prediction errors are comparable for all three methods, even though one can observe that the ARMA prediction errors start drifting after predicting more than 10 cycles ahead. Once again, the ARMA and similarity matrix methods take an order of magnitude longer to compute the prediction results for all 40 test runs than the newly proposed algorithm. The computational advantage comes from the analytical character of the GMM based estimation of future feature distributions used in the newly proposed time-series prediction method, which bypasses the need for Monte-Carlo sampling utilized in [8]. Thus, the new method incorporates the rich information from the past runs, but, unlike what we see in [8], it does not do that at the expense of tremendous computational effort and sacrificing analytical tractability.

4. Conclusions

In this paper, we present a novel time-series prediction algorithm capable of dealing with a long-term prediction of non-stationary multivariate time-series. The method is based on the concept of similarity-weighted Gaussian mixture models (GMMs) obtained via comparisons of signatures describing the current degradation process with those observed on the same machine/process in the past. It provides one with a natural way to derive the predicted feature distributions over time, which could be used to obtain information about the remaining useful life, predicted probabilities of failure, or unacceptable behavior. The new method was tested in predicting signatures extracted during the operation of an industrial plasma enhanced chemical vapor deposition (PECVD) tool. The results showed that the newly

proposed prediction method yields noticeably smaller mean squared errors, compared with ARMA based prediction and comparable mean squared errors to another recently introduced similarity-based time-series prediction model. However, the analytical structure of the method proposed in this paper computes the prediction distributions an order of magnitude faster.

An avenue for possible future work is potential grouping of similar degradation trajectories (runs) having similar evolutions of the time-series of sensory features, which will enable one to reduce the number of degradation trajectories that need to be kept in the historical database and used for predictions. In that context, a new run would be added to the library of past runs only after a degradation trajectory is observed that is "sufficiently different" from the ones seen in the past. This problem is outside the scope of this paper, but carries significant potential benefits for eventual users of the new time-series prediction method.

Acknowledgment

This work was supported in part by the International SEMATECH Manufacturing Initiative (ISMI).

References

[1] A. Heng, S. Zhang, A.C.C. Tan, J. Mathew, Rotating machinery prognostics: state of the art, challenges and opportunities, Mechanical Systems and Signal Processing 3 (23) (2009) 724–739.
[2] D. Djurdjanovic, J. Lee, J. Ni, Watchdog agent – an infotronics based prognostics approach for product performance assessment and prediction, International Journal of Advanced Engineering Informatics, Special Issue on Intelligent Maintenance Systems 17 (3–4) (2003) 109–125.
[3] A. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition based maintenance, Mechanical Systems and Signal Processing 20 (2006) 1483–1510.
[4] Djurdjanovic, D., Liu, Y., 2006. Survey of predictive maintenance research and industry best practice. University of Michigan, Ann Arbor, MI.
[5] J. Hu, L. Zhang, L. Ma, W. Liang, An integrated method for safety pre-warning of complex system, Safety Science 48 (2010) 580–597.
[6] J. Hu, L. Zhang, W. Liang, An adaptive online safety assessment method for mechanical system with pre-warning function, Safety Science 3 (50) (2012) 385–399.
[7] M. You, L. Li, G. Meng, J. Ni, Cost-effective updated sequential predictive maintenance policy for continuously monitored degrading systems, IEEE Transactions on Automation Science and Engineering 2 (7) (2010) 257–265.
[8] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, J. Lee, Similarity based method for manufacturing process performance prediction and diagnosis, Computers in Industry 6 (58) (2007) 558–566.
[9] J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, New York, 1998, Ch. 1.
[10] R. Serway, J. Jewett, Physics for Scientists and Engineers, eighth ed., Brooks/Cole Publishers, 2010.
[11] R. Harris, A Primer of Multivariate Statistics, Academic Press, New York, 1975.
[12] S.M. Pandit, S.M. Wu, Time Series and System Analysis with Applications, John Wiley, 1983.
[13] R. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME Journal of Basic Engineering 82 (1960) 34–45.
[14] R. Duda, P. Hart, D. Stork, Pattern Classification, second ed., Wiley & Sons, NY, 2001117–124.
[15] A. Geva, Non-stationary time-series prediction using fuzzy clustering, in: Presented at Proceedings of NAFIPS-99: 18th International Conference of the North American Fuzzy Information Processing Society, New York, 1999.
[16] C. Lee, A. Liu, W. Chen, Pattern discovery of fuzzy time series for financial prediction, IEEE Transactions on Knowledge and Data Engineering 18 (2006) 613–625.
[17] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006 Text (1).
[18] O. Renaud, J. Starck, F. Murtagh, Wavelet-based combined signal filtering and prediction, IEEE Transactions on Systems, Man, Cybernetics, Part B: Cybernetics 35 (2005) 1241–1251.
[19] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, Cambridge, 2000.
[20] Z. Zivkovic, F. Heijden, Recursive unsupervised learning of finite mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (26) (2004) 651–656.
[21] B. Lindsay, Mixture Models: Theory, Geometry and Applications, Institute of Mathematical Statistics, CA, 19951–6.
[22] S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice Hall, Upper Saddle River, NJ, 1999.
[23] M. Han, J. Xi, S. Xu, F. Yin, Prediction of chaotic time series based on the recurrent predictor neural network, IEEE Transactions on Signal Processing 52 (2004) 3409–3416.
[24] J. Connor, R. Martin, L. Atlas, Recurrent neural networks and robust time series prediction, IEEE Transactions on Neural Networks 5 (1994) 240–253.
[25] J. Yu, Bearing performance degradation assessment using locality preserving projections, Expert Systems with Applications 38 (2011) 7440–7450.
[26] H.E. Kim, A.C.C. Tan, J. Mathew, New machine prognostics approach base on health state probability estimation, Australian Journal of Mechanical Engineering 2 (8) (2011) 79–89.
[27] Y. Yang, Y. Liao, G. Meng, J. Lee, A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis, Expert Systems with Applications 38 (2011) 11311–11320.
[28] T. Wang, J. Yu, D. Siegel, J. Lee, A similarity-based prognostics approach for remaining useful life estimation of engineered systems, International Conference on Prognostics and Health Management (2008) 1–6.
[31] J.C. Li, Signal processing in manufacturing monitoring, Condition Monitoring and Control for Intelligent Manufacturing (2006) 245–265.
[32] D.C. Montgomery, Introduction to Statistical Quality Control, fourth ed., John Wiley, New York, 2001.
[33] F. Hu, J. Zidek, The weighted likelihood, The Canadian Journal of Statistics 30 (3) (2002) 347–371.
[34] F. Hu, W. Rosenberger, J. Zidek, Relevance weighted likelihood for dependent data, Journal of Metrika 51 (3) (2000) 223–243.
[35] F. Hu, J. Zidek, Forecasting NBA Basketball Playoff Outcomes Using the Weighted Likelihood, Institute of Mathematical Statistics, Lecture Notes-Monograph Series 45 (2004) 385–395.
[36] Chen, Y., Gupta, M., 2010. EM Demystified: An Expectation Maximization Tutorial, UWEE Technical Report, University of Washington Department of EE.
[37] H. Geng, Semiconductor Manufacturing Handbook, McGraw-Hill Handbooks, NY, 2005, Ch. 14.
[38] A. Bleakie, D. Djurdjanovic, Dynamic Feature Monitoring Technique Applied to Thin Film Deposition Processes in an Industrial PECVD Tool, in: Proceedings of the ASME 2011 International Manufacturing Science and Engineering Conference, June 13–17, 2011.

Alexander Q. Bleakie was born in Houston, Texas, USA. He received a B.S. degree in Mechanical Engineering from Texas A&M University in 2008 with Summa Cum Laude Honors, and a M.S. degree in Mechanical Engineering from the University of Texas at Austin in 2010. He is currently pursuing a Ph.D. degree in Mechanical Engineering at The University of Texas at Austin with emphasis on Dynamic Systems and Control. His current research interests include dynamic modeling, control theory, fault diagnosis, and estimation.

Dragan Djurdjanovic received the B.S. degree in mechanical engineering and applied mathematics from the University of Nis, Nis, Serbia, in 1997, the M.Eng. degree in mechanical engineering from Nanyang Technological University, Singapore, in 1999, the M.S. degree in electrical engineering (systems), and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, in 2002.

He is currently an Assistant Professor of mechanical engineering, and operations research and industrial engineering, with the Department of Mechanical Engineering, University of Texas, Austin. He has co-authored 45 published or accepted journal publications, 3 book chapters, and more than 30 conference publications. His current research interests include advanced quality control in multistage manufacturing systems, intelligent proactive maintenance techniques, and applications of advanced signal processing in biomedical engineering.

Dr. Djurdjanovic was the recipient of several prizes and awards, including the 2006 Outstanding Young Manufacturing Engineer Award from the Society of Manufacturing Engineers, the 2005 Teaching Incentive Award from the Department of Mechanical Engineering, University of Michigan, Nomination for the Distinguished Ph.D. Thesis from the Department of Mechanical Engineering, University of Michigan in 2003, and the Outstanding Paper Award at the 2001 SME North American Manufacturing Research Conference.