Computation Intelligence in Online Reliability Monitoring

Ratna Babu Chinnam

Department of Industrial and Manufacturing Engineering, Wayne State University, Detroit

Bharatendra Rai

Ford Motor Company, Dearborn

9.1 Introduction

9.1.1 Individual Component versus Population Characteristics

The lifecycle of a product is generally associated with two key players viz., a producer and an end-user. Although producers and end-users depend on each other, they have their own priorities. For a producer, the characteristics of a population of units are more important than the individual units, whereas for an end-user the opposite is true. The government makes laws for the population of a country, but a parent may be more concerned about its impact on the future of their individual children. A car manufacturer targets consistency in fuel efficiency for a population of cars, whereas a car owner has concerns about the fuel efficiency of his/her car. Similar differences in concerns also apply to tennis racquet manufacturer versus a tennis player or a cutting tool manufacturer versus cutting tool user.

The producer commonly uses time-to-failure data to assess and predict reliability of a population of products. For highly reliable products, when time-to-failure data are difficult to obtain, degradation data are also used for such an analysis. On the other hand, for an end-user assessing and predicting reliability of an individual part or component often assumes more importance. For example, a producer involved in voluminous production of drill-bits needs to assess and monitor reliability on a regular basis to ensure consistent population characteristics for end users. However, the end-

R.B. Chinnam and B. Rai: *Computation Intelligence in Online Reliability Monitoring*, Computational Intelligence in Reliability Engineering (SCI) **40**, 223–260 (2007)
www.springerlink.com

user of such drill-bits generally has more interest in the reliability of the individual drill-bits. Voluminous amount of work has been published for reliability modeling and analysis related to population characteristics (Kapur and Lamberson 1977; Lawless 1982; Nelson 1982; Lewis 1987; Elsayed 1996; Meeker and Escobar 1998). A good review of literature on degradation signals can be obtained from Tomsky (1982), Lu and Pantula (1989), Nelson (1990), Lu and Meeker (1993), Tseng et al. (1995), Tang et al. (1995), Chinnam et al. (1996), Lu et al. (1997), Meeker et al. (1998), Wu and Shao (1999), Wu and Tsai (2000), and Gebraeel et al. (2004).

This chapter focuses on monitoring of reliability from end-users viewpoint. For monitoring reliability of individual unit, time-to-failure data are not of much use. For example, Fig. 1 gives a plot of the life of 16 M-1 grade quarter-inch high speed twist drills, measured in number of holes successfully drilled in quarter-inch steel plates, when operated with no coolant at a speed 2000 rpm and a feed 20 inches/min (Chinnam 1999). Even though the drill-bits came from the same manufacturer in the same box, it is obvious from the figure that the dispersion in life (ranging from 17 holes to 58 holes) is far too large with respect to the mean time-to-failure (around 28 holes), and hence, information about the population would be of little value to the end user. In contrast, the end user would greatly benefit from an on-line estimate of the reliability of the drill-bit, to make effective decisions regarding optimal drill bit replacement strategies, essentially lowering production costs by fully utilizing the drill-bit.

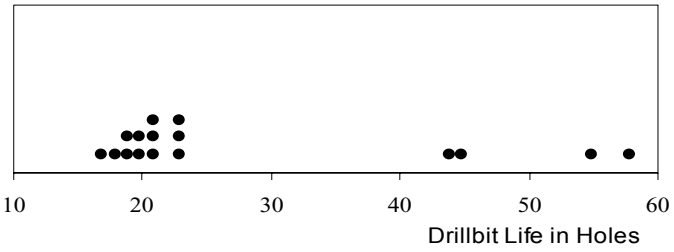


Fig. 1. A plot of life of 16 drill-bits, measured in number of holes successfully drilled

For individual units, condition at different points of time or degradation levels are more useful to arrive at optimal component replacement or maintenance strategies leading to improved system utilization, while reducing the risk and maintenance costs. While gathering data based on direct measurements of the condition or degradation level of a unit is not impossible, such methods are not practical due to the intrusive nature for

applications requiring online monitoring of reliability. For example, measuring the amount of wear on a milling insert or a drill bit after every operation would slow down the production drastically. Thus, applications requiring online reliability monitoring of individual units more often use indirect and non-intrusive measurements. The nature and type of such indirect measurements may vary from application to application, however, their selection is critical for an effective decision making process. A basic criterion they need to fulfill, apart from being non-intrusive, is to have a good correlation with the degradation level of the part/component of interest.

9.1.2 Diagnostics and Prognostics for Condition-Based Maintenance

In many physical and electro-mechanical systems, the system or unit under consideration generates degradation signals that contain valuable information about the health/well-being of the system. These degradation signals, such as power consumption of a metal cutting machine tool, error rate of a computer hard disk, temperature of a drill-bit, vibration in machinery, color spectrum from an arc welder, loads acting on a structure, tend to be non-stationary or transitory signals having drifts, trends, abrupt changes, and beginnings and ends of events. Despite considerable advances in intelligent degradation monitoring for the last several decades, on-line condition monitoring and diagnostics are still largely reserved for only the most critical system components and have not found their place in mainstream machinery and equipment health management (Kacprzynski and Roemer, 2000). If one were to talk about predictive maintenance technologies, in particular prognostics and on-line reliability assessment, there exist no robust methods for even the most critical system components. *Diagnostics has traditionally been defined as the ability to detect and classify fault conditions*. Literature is extremely vast in this area. *Prognostics on the contrary is defined here as the capability to provide early detection of the precursor to a failure condition and to manage and predict the progression of this fault condition to component failure*. Recognizing the inability to prevent costly unscheduled equipment breakdowns through Preventive Maintenance (PM) activities and basic diagnostic condition monitoring methods, there seem to be consensus among industry and federal agencies that one of the next great opportunities for product differentiation and successful competition in the world markets lies in true prognostics based condition-based maintenance (CBM).

Condition-Based Maintenance (*CBM*) *is a philosophy of performing maintenance on a machine or system only when there is objective evidence of need or impending failure.* CBM typically involves mounting non-intrusive sensors on the component to capture signals of interest and subsequent interpretation of these signals for the purpose of developing a customized maintenance policy. Given recent advances in the areas of non-intrusive sensors, data acquisition hardware, and signal processing algorithms, combined with drastic reductions in computing and networking costs and proliferation of information technology products that integrate factory information systems and industrial networks with web-based visual plant front-ends, it is now possible to realize systems that can deliver cost effective diagnostics, prognostics, and CBM for a variety of industrial systems. The basic elements necessary for successful diagnostics and prognostics for CBM are illustrated in Fig. 2 (Chinnam and Baruah 2004).

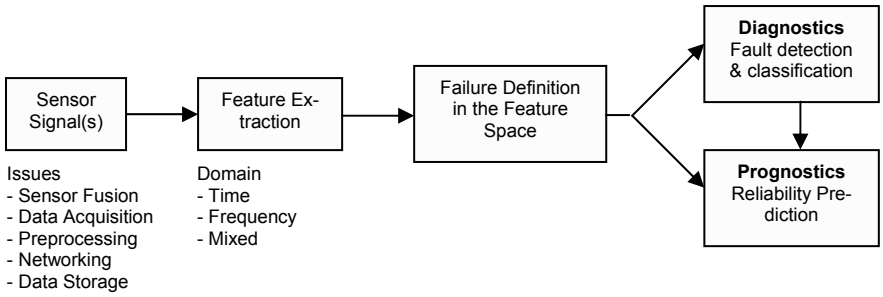


Fig. 2. Basic elements of diagnostics and prognostics for CBM.

Sensor signal(s): Sensing techniques commonly used include touch sensors, temperature, thrust force, vibration, torque, acoustic emission, voltage, noise, vision systems, etc. The data obtained from sensor signals contain useful information about the condition or degradation levels of a unit. As data recording is automated, it is not uncommon to see several thousands of measurements recorded per incremental usage condition. Extraction of appropriate and useful features from such data is critical before models for online monitoring and prediction of individual component reliability can be developed.

Feature extraction: Extraction of useful features typically involves analysis of data in several different domains. Basic time domain signal parameters utilized in conventional diagnostics include amplitude, crest factor, kurtosis, RMS values, and various measures of instantaneous and cumulative energy (Zhou et al. 1995; Quan et al. 1998; Kuo 2000; Dzenis and

Qian 2001; Sung et al. 2002). Frequency domain spectral parameters include the Fourier transform and linear spectral density, while advanced spectral measures include higher-order (or displaced) power spectral density. Extensive research over the last two decades has resulted in a long list of promising features or *Figures-of-Merit* (FOMs) for different applications. For example, Lebold et al. (2000) discussed 14 FOMs for gearbox diagnostics and prognostics, employing a vibration signal alone. The literature offers FOMs for monitoring and diagnosis of mechanical systems such as gearboxes, pumps, motors, engines, and metal cutting tools. In recent years, mixed-domain analysis methods such as Wavelets are gaining popularity for their ability to offer a shorter yet accurate description of a signal by employing scale-based basis functions. Wavelet analysis, while still being researched for machine diagnostics and prognostics (Chinnam and Mohan 2002; Vachtsevanos *et al.* 1999), is well established for such applications as image processing. In the last few years, the *Empirical Mode Decomposition* method has received much attention for its ability to analyze non-stationary and nonlinear time series, something not possible with methods such as Wavelet analysis (Huang et al. 1998).

Failure Definition in the Feature Space: A specified level of degradation in feature space is generally used to define failure. Such a threshold limit is required to assess and predict reliability of a unit. Sometimes the features of interest may consistently show significantly different degradation levels before the physical failure occurs. In such situations, failure definition in the feature space may be easier to determine. In situations where this is not the case, arriving at a failure definition may be more involved. We later discuss a fuzzy inference model to arrive at failure definition in the feature space in Section 4.

Diagnostics: During the diagnostics process, specific FOMs are typically compared to threshold limits (Begg et al. 1999). Additional processing may determine a signature pattern in one, or multiple, fault measure(s). Automated reasoning is often used to identify the faulty type (cracked gear tooth, bearing spall, imbalance etc.), location, and severity. The core problem of diagnostics is essentially a problem of classification (Elverson 1997). Discriminant transformations are often used to map the data characteristic of different failure mode effects into distinct regions in the feature subspace (Byington and Garga 2001). The task is relatively straightforward in the presence of robust FOMs. The literature is vast in this area and commercial technologies are well established. Depending on the application, these systems employ model-based methods, any number of statistical methods, and a variety of computational intelligence methods.

Dimla et al. (1997) provide a critical review of the neural network methods used for tool condition monitoring.

Prognostics: Contrary to diagnostics, the literature on prognostics, called the Achilles' heel of the CBM architecture (Vachtsevanos et al. 1999), is extremely sparse. Unfortunately, the large majority of the literature that employs the term 'prognostics' in the title ends up discussing diagnostics. To achieve prognostics, there need to be features that are suitable for tracking and prediction (Begg et al. 1999). It is for this reason that prognostics is receiving the most attention for systems consisting of mechanical and structural components, for unlike electronic or electrical systems, mechanical systems typically fail slowly as structural faults progress to a critical level (Mathur et al., 2001).

9.2 Performance Reliability Theory

Let $\{y(s)\}$ represent a scalar time-series generated by sampling the performance degradation signal (or a transformation thereof). Suppose that $\{y(s)\}$ can be described by a *nonlinear regressive model* of order p as follows:

$$y(s) = f(y(s-1), y(s-2), \cdots, y(s-p)) + \varepsilon(s) \tag{1}$$

where f is a nonlinear function and $\varepsilon(s)$ is a residual drawn from a *white Gaussian noise process*. In general, the nonlinear function f is unknown, and the only information we have available to us is a set of observables: $y(1), y(2), \cdots, y(S)$, where S is the total length of the time-series. Given the data set, the requirement is to construct a physical model of the time-series. To do so, we can use any number of statistical (for example, Box and Jenkin's auto-regressive integrated moving average (ARIMA) models) or computational intelligence based (for example, feed-forward neural networks such as multi-layer perceptron (MLP)) forecasting techniques as a one-step predictor of order p. Specifically, the model is estimated to make a prediction of the sample $y(s)$, given the immediate past p samples $y(s-1), y(s-2), \cdots, y(s-p)$, as shown by

$$\hat{y}(s) = \tilde{f}(y(s-1), y(s-2), \cdots, y(s-p)) + e(s). \tag{2}$$

The nonlinear function $\tilde{f}$ is the approximation of the unknown function f, built in general to minimize some cost function (J) of the prediction error

$$e(s) = y(s) - \hat{y}(s), \quad p+1 \le s \le S \tag{3}$$

Note that a single model can be potentially built to simultaneously work with many degradation signals and also allow different prediction orders for different signals. Of course, the structure and complexity of the model will increase with an increase in the number of degradation signals jointly modeled.

Now, let $F(t)$ denote the probability that failure of a component takes place at a time or usage less than or equal to t (i.e., $F(t) = P(T \le t)$), where the random variable T denotes the time to failure. From the definition of conditional probability, the conditional reliability that the component will fail at some time or usage $T > t + \Delta t$, given that it has not yet failed at time $T = t$ will be:

$$R\left((t+\Delta t) \mid t\right) = \left[1 - P\left(T \le t + \Delta t\right)\right] \Big/ P\left(T > t\right). \tag{4}$$

Let $\mathbf{y} = [y_1, y_2, ..., y_m]$ denote the vector of m degradation signals (or a transformation thereof) being monitored from the system under evaluation. Let $\mathbf{y}^{pcl} = [y_1^{pcl}, y_2^{pcl}, ..., y_m^{pcl}]$ denote the vector of deterministic *performance critical limits* (PCLs), which represent an appropriate definition of failure in terms of the amplitude of the m degradation signals. For any given operating/environmental conditions, *performance reliability* can be defined as "the conditional probability that $\mathbf{y}$ does not exceed $\mathbf{y}^{pcl}$, for a specified period of time or usage." Obviously, the above definition directly applies to the case where the amplitudes of the degradation signals are preferred to be low (lower-the-better signals with higher critical limits), and can be easily extended to deal with higher-the-better signals (with lower critical limits) and nominal-value-is-best signals (with two-sided critical limits), and any combinations in between. Without loss of generality, for illustrative purposes, let us make the assumption here that all the m degradation signals are of lower-is-better type signals.

Since a model estimated using past degradation signals collected from other similar components keeps providing us with an estimate of $\mathbf{y}$ into the future, denoted by $\hat{\mathbf{y}}(t_f)$, under the assumption that the change in $\mathbf{y}$

from the current time point (t_c) to the predicted time point (t_f) is either monotonically decreasing or increasing, the reliability that the component or system will operate without failure until t_f is given by:

$$R\left(\left(T \geq t_f\right) | t_c\right) = \int_{y_1=-\infty}^{y_1^{pcl}} \int_{y_2=-\infty}^{y_2^{pcl}} .. \int_{y_m=-\infty}^{y_m^{pcl}} g\left(\hat{\mathbf{y}}\left(t_f\right)\right) dy_1 dy_2 .. dy_m \tag{5}$$

where $g\left(\hat{\mathbf{y}}\left(t_f\right)\right)$ denotes the probability density function of $\hat{\mathbf{y}}(t_f)$. The assumption here is that y_i^{pcl} is a constant for any given i and is independent of $\mathbf{y}(t_f)$. Under these conditions, the failure space is bounded by orthogonal hyper planes. If the independence assumption is not justified, one could use a hyper-surface to define the failure boundary (Lu et al. 2001). If need be, one could even relax the assumption of a deterministic boundary and replace it with a stochastic boundary model. However, such an extension is non-trivial.

For the special case where there exists just one lower-the-better degradation signal, this process is illustrated in Fig. 3 (Chinnam and Baruah 2004). The shaded area of $g\left(\hat{\mathbf{y}}\left(t_f\right)\right)$ at any t_f denotes the conditional-unreliability of the unit. That is, given that the unit has survived until t_c, the shaded area denotes the probability that the unit will fail by t_f. To obtain mean residual life (MRL), using r_{MRL}, the least acceptable reliability, one can estimate t_{MRL}, the time instant/usage at which the reliability of the unit reaches r_{MRL}. Thus, one can calculate the MRL to be the time difference between t_c and t_{MRL}.

9.3 Feature Extraction from Degradation Signals

Feature extraction is an important step in developing effective procedures for online reliability monitoring using degradation signals. In this section we discuss time, frequency and mixed-domain analysis techniques for preprocessing degradation signals and feature extraction.

9.3.1 Time, Frequency, and Mixed-Domain Analysis

As is pointed out in the introduction, the most important and fundamental variables in degradation signal processing are time and frequency. In addition, the degradation signals often tend to be stochastically non-stationary, rendering the fast Fourier transform (FFT) spectrum (a transform that is quite popular for frequency analysis) inadequate, for it can only evaluate an average spectrum over a definite time period and loses the non-stationary characteristics of the signals (Yen and Lin 2000). Given this, in many real world applications, it is far more useful to characterize the signal in both the time- and frequency- domains, simultaneously.

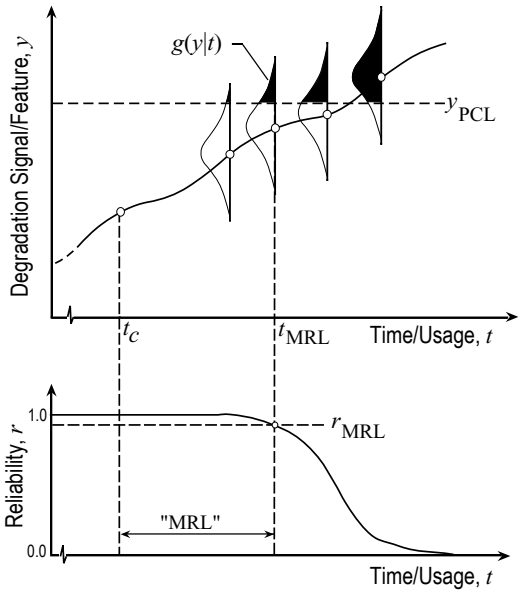


Fig. 3. Degradation signal forecasting model coupled with a failure definition PCL to estimate MRL

Several joint time-frequency (or mixed-domain) alternatives have been proposed in the literature. Some of the alternatives include the short term Fourier transform (STFT) and the Wavelet transform (WT), to name a few. Joint time-frequency methods are conventionally classified into two categories: linear and quadratic (Qian and Chen, 1996). The principle of linear time-frequency representation involves decomposing any signal into a linear expansion of functions that belong to a set of redundant elementary functions. All linear transformations are achieved by comparing the ana-

lyzed signal with a set of prudently selected elementary functions. While the functions for STFT are obtained by frequency modulation of sine and cosine waves in STFT, in the WT, the functions are obtained by scaling and shifting the mother wavelet.

The most important relationship in terms of joint time-frequency analysis is the relation between signal's time window duration and frequency bandwidth. Several different definitions are offered in the literature for specifying the time window duration and frequency bandwidth (Qian and Chen, 1996; Akay and Mello, 1998). In general, for mixed-domain methods, there is a tradeoff between time resolution and frequency resolution for there is an upper bound on the product of the two resolutions. In other words, an increase in the time resolution results in a loss of frequency resolution, and vice versa. In STFT, since the elementary function is the same for all the frequency components, time and frequency resolutions are fixed on the time-frequency plane once the elementary function has been chosen. Hence, the choice of time window duration is the key for any good STFT representation. In WT, time and frequency resolutions are not fixed over the entire time-frequency plane.

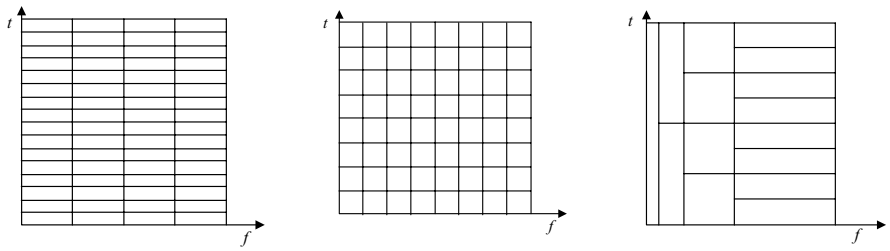


Fig. 4. Comparison of the STFT and the wavelet transform in terms of time and frequency resolution.

The tiling of the windows in the joint time-frequency plane is illustrated for STFT and WT in Fig. 4. While the STFT tilling is linear, the WT tilling is logarithmic. In Fig. 4(a) and 4(b), when the length of window is specified, the time and frequency resolution remains constant throughout the plane. In Fig. 4(c), time and frequency resolution is not fixed over the entire time-frequency domain: time resolution becomes good at higher frequencies whereas frequency resolution becomes good at lower frequency.

9.3.2 Wavelet Preprocessing of Degradation Signals

As discussed in the introduction, degradation signals often tend to be rich in time and frequency components, and hence, lend themselves for better representation in mixed-domain analysis. The treatment on wavelets that follows is borrowed heavily from DeVore and Lucier (1992). The term wavelet denotes a univariate function ψ (multivariate wavelets exist as well), defined on $\mathbf{R}$, which, when subjected to the fundamental operations of shifts (i.e., translation by integers) and dyadic dilation, yields an orthogonal basis of $L_2(\mathbf{R})$. That is, the functions $\psi_{j,k} := 2^{k/2}\psi(2^k \cdot - j)$, $j,k \in \mathbf{Z}$, form a complete orthonormal system for $L_2(\mathbf{R})$. Such functions are generally called orthogonal wavelets, since there are many generalizations of wavelets that drop the requirement of orthogonality.

One can view a wavelet ψ as a "bump" and think of it as having compact support, though it need not. Dilation squeezes or expands the bump and translation shifts it. Thus, $\psi_{j,k}$ is a scaled version of ψ centered at the dyadic integer $j2^{-k}$. If k is large positive, then $\psi_{j,k}$ is a bump with small support; if k is large negative, the support $\psi_{j,k}$ is large. The requirement that the set $\{\psi_{j,k}\}_{j,k\in\mathbf{Z}}$ forms an orthonormal system means that any function $f \in L_2(\mathbf{R})$ can be represented as a series

$$f = \sum_{j,k\in\mathbf{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} \tag{6}$$

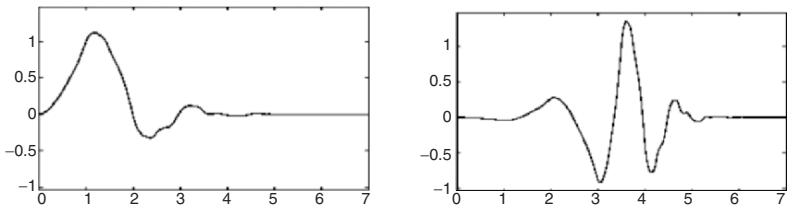
with $\langle f, g \rangle := \int_{\mathbf{R}} f\bar{g}dx$ the usual inner product of two $L_2(\mathbf{R})$ functions. One can view Eq. (6) as building up the function f from the bumps. Bumps corresponding to small values of k contribute to the broad resolution of f; those corresponding to large values of k give finer detail.

The decomposition of Eq. (6) is analogous to the Fourier decomposition of a function $f \in L_2(\mathbf{R})$ in terms of the exponential functions $e_k := e^{ik\cdot}$, but there are important differences. The exponential functions e_k have global support. Thus, all terms in the Fourier decomposition contribute to the value of f at a point x. On the other hand, wavelets are usually either of compact support or fall off exponentially at infinity. Thus, only the

terms in Eq. (6) corresponding to $\psi_{j,k}$ with $j2^{-k}$ near x make a large contribution to x. The representation Eq. (6) is in this sense local.

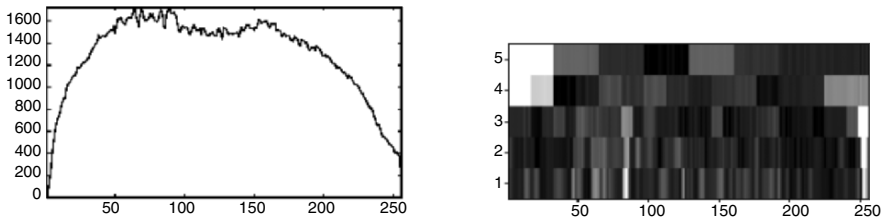
All this would be of little more than theoretical interest if it were not for the fact that one can efficiently compute wavelet coefficients and reconstruct functions from these coefficients. These algorithms, known as "fast wavelet transforms" are the analogue of the Fast Fourier Transform and follow simply from the refinement of the dilation and shift equation mentioned above.

In summary, the wavelet transform results in many wavelet coefficients, which are a function of scale (or level or frequency) and position. Hence, a wavelet plot is a plot of coefficients on time-scale axis. The higher the scale, the more stretched the wavelet. The more stretched the wavelet, the longer the portion of the signal with which it is compared, and thus the coarser the signal features being measured by the wavelet coefficients. Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal. The coefficients constitute the results of a regression of the original signal performed on the wavelets.



(a) Scaling function for DB4 transform. (b) Wavelet function for DB4 transform.

Fig. 5. Daubechies DB4 wavelet tranform.



(a) Thrust-force singnal from drill-bit #8 hole #1. (b) DB4 wavelt coefficients at different scales.

Fig. 6. Thrust-force degradation signal from drill-bit #8 hole #1 and their transformed DB4 wavelet coefficients.

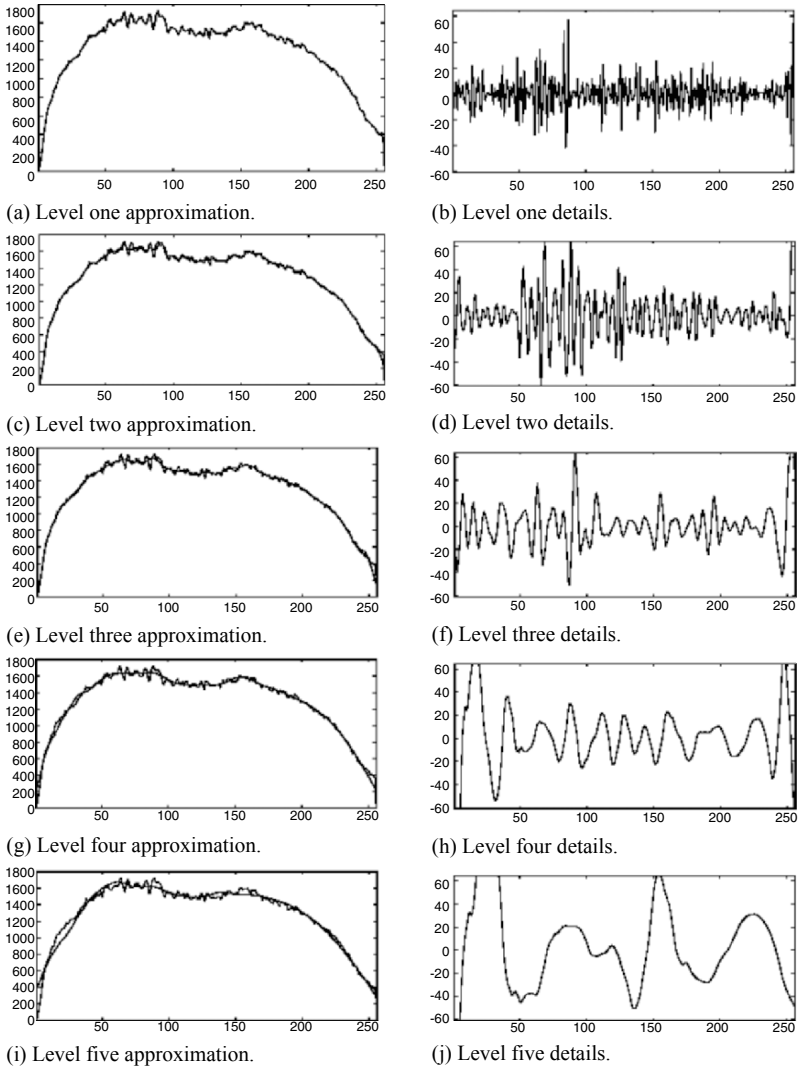


Fig. 7. Discrete wavelet analysis of thrust-force signal from drill-bit #8 hole #1

The particular wavelet transform considered in this paper is the compactly supported Daubechies' wavelet transform. The transform is compactly supported with extreme phase and highest number of vanishing moments for a given support width. One particular Daubechies' wavelet transform, the DB4 discrete wavelet transform function and its associated scaling function, is shown in Fig. 5. For illustrative purposes, the thrust-

force degradation signal measured from drill-bit #8 while drilling hole #4 is shown in Fig. 6(a). The wavelet plot resulting from the DB4 transform of this degradation signal is shown in Fig. 6(b). In this plot, the x-axis represents time while the y-axis represents scales, and the darker cells in the plot represent coefficients that are low in amplitude. The reconstruction of this signal at different levels (or scales) is shown in Fig. 7 along with the detail counterparts. The transform is performed using MatLab's Wavelet Toolbox. We request the reader to see Daubechies (1990) for a detailed treatment of this transform.

9.3.3 Multivariate Methods for Feature Extraction

When multiple features are used to represent degradation signals, multivariate methods can also be used for extracting useful features. Rai, Chinnam, and Singh (2004) used Mahalanobis-Taguchi System (MTS) analysis for predicting drill-bit breakage from degradation signals. A MTS analysis consists of four stages (Taguchi and Jugulum 2002). In the first stage of analysis a measurement scale is constructed from a standardized (by subtracting the mean and dividing by the standard deviation) 'normal' group of features using Mahalanobis distances (MDs) given by,

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^{'} C^{-1} Z_{ij} \tag{7}$$

where,

j = Observation number in the normal group (1 to m)
i = Feature number (1 to k)
$Z_{ij} = (z_{1j}, z_{2j}, ..., z_{kj})$ = Standardized vector
C^{-1} = Inverse of the correlation matrix.

In the second stage, larger values of MDs obtained from an abnormal group are used for validating the measurement scale developed in the first stage. In the third stage, useful features are extracted from those under study using signal-to-noise ratio values. A S/N ratio for say qth trial with 't' features present in the combination can be obtained as,

$$\eta_q = -10 \log \left[\frac{1}{t} \sum_{j=1}^{t} \frac{1}{MD_j} \right] \tag{8}$$

For a given feature, an average value of the S/N ratio is determined separately at level-1 indicating presence and at level-2 indicating absence

of a feature. Subsequently, gain in S/N ratio values is obtained by taking difference of the two average values as,

$$\text{Gain} = (\text{Avg. S/N Ratio})_{\text{Level-1}} - (\text{Avg. S/N Ratio})_{\text{Level-2}} \quad (9)$$

A positive gain for a feature indicates its usefulness and vice-versa. And finally, in the fourth stage of analysis a threshold value for the MDs is developed from the normal group to enable degradation level prediction using the useful features.

9.4 Fuzzy Inference Models for Failure Definition

Most prognostics methods in the literature for on-line estimation of MRL utilize trending or forecasting models in combination with mechanistic or empirical failure definition models. However, in spite of significant advances made throughout the last century, our understanding of the physics of failure is not quite complete for many electro-mechanical systems. In the absence of sound knowledge for the mechanics of degradation and/or adequate failure data, it is not possible to establish practical failure definition models in the degradation signal space. Under these circumstances, the sort of procedures illustrated in Section 2 is not feasible. However, if there exist domain experts with strong experiential knowledge, one can potentially establish fuzzy inference models for failure definition. In this section, we suggest the incorporation of fuzzy inference models to introduce the definition of failure in the degradation signal space using domain experts with strong experiential knowledge. While the trending or forecasting subcomponent will predict the future states of the system in the degradation signal space, it is now the task of the fuzzy inference model to estimate the reliability associated with that forecast state. If one were to compare this procedure with that discussed in Section 2, it is equivalent to replacing the right hand side of Eq. (5) with a fuzzy inference model.

One might argue that probabilistic models could be potentially used for modeling experiential knowledge of domain experts. However, it is widely accepted that classical probability theory has some fundamental shortcomings when it comes to modeling the nature of human concepts and thoughts, which tend to be abstract and imprecise. While probability theory is developed to model and explain randomness, fuzzy arithmetic and logic is developed to model and explain the uncertain and imprecise nature of abstract thoughts and concepts. Over the last three decades, since Lofti Zadeh authored his seminal paper in 1965 on fuzzy set theory (Zadeh 1965), the scientific community had made major strides in extending the

set theory to address applications in areas such as automatic control, data classification, decision analysis, and time series prediction (Jang et al. 1997).

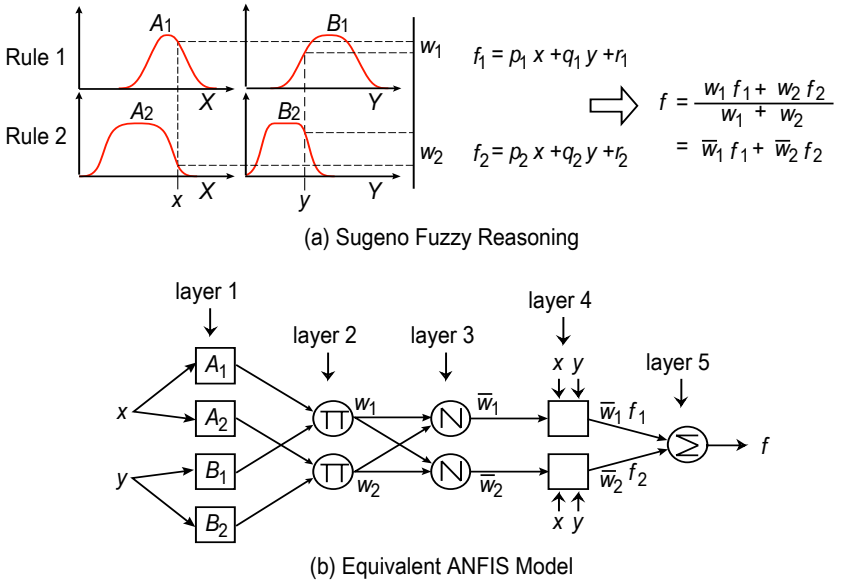


Fig. 8. Sugeno FIM with two inputs (X, Y) and one output (F).

In the context of prognostics and failure definition, Sugeno Fuzzy Inference Model (FIM), illustrated in Fig. 8(a), is particularly attractive for failure definition for three reasons:

1. It makes a provision for incorporating subjective knowledge of domain experts and experienced operators,
2. Model can be viewed as a feed-forward neural network (labeled Adaptive-Network based Fuzzy Inference Systems or ANFIS), and hence, can be adapted using empirical/historical data coupled with gradient search methods (Jang 1993), and
3. Computationally efficient for the absence of a de-fuzzification operator prevalent in other fuzzy inference models.

The illustrated two-input (X and Y) one-output (F) Sugeno FIM carries two membership functions for each of the two input variables, namely A_1, A_2 and B_1, B_2. The model is made of two rules. For example, *Rule-1* states that if X is A_1 and Y is B_1, then the output is given by $f_1 = p_1 x + q_1 y + r_1$. Here, A_1 and B_1 denote linguistic variables (such

as "thrust is low" or "vibration is high"). Even though the consequent of each rule constitutes a first-order model, the overall relationship is often highly nonlinear. The equivalent ANFIS model is shown in the illustration as well.

For the application considered here, typically, the number of input variables for the Sugeno FIM will be equal to the number of degradation signals under investigation, and there is one output variable predicting the reliability of the unit (i.e., $r(t)$). The number of membership functions and the number of rules needed to fully describe the failure definition will be dictated by the specific application and input from domain experts. In the absence of first-principles models, rules can be initially formulated with the help of domain experts and experienced operators. All the parameters of the Sugeno FIM can be adapted to best describe any historical dataset using the ANFIS framework. For more details regarding Sugeno fuzzy inference models or their ANFIS equivalents (Jang et al. 1997).

9.5 Online Reliability Monitoring with Neural Networks

In general, artificial neural networks are composed of many non-linear computational elements, called nodes, operating in parallel and arranged in patterns reminiscent of biological neural nets (Lippmann 1987). These processing elements are connected by weight values, responsible for modifying signals propagating along connections (also called synapses) and used for the training process. The number of nodes plus the connectivity define the topology/structure of the network, and is intimately connected with the learning algorithm used to train the network Haykin (1999). The higher the number of nodes per layer and/or the number of layers, the higher the ability of the network to extract higher-order statistics (Churchland and Sejnowski 1992) and approximate more complex relationships between inputs and outputs.

One of the most significant properties of a neural network is its ability to learn from its environment that normally involves an iterative process of adjustments applied to the synaptic weights. There is no unique learning algorithm for the design of neural networks and they differ from each other in the way in which the adjustment of synaptic weights takes place. Two popular learning algorithms are the error-correction learning algorithm (in essence a stochastic gradient-descent search technique) used normally for training FFNs such as FIR MLPs discussed in Section 5.2 and the competitive learning algorithm used for training networks such as SOMs discussed in Section 5.3. It is beyond the scope of this chapter to discuss the nature

of these learning algorithms in detail, and the reader is referred to Haykin (1999).

Two popular learning paradigms for neural networks involve *supervised* learning and *unsupervised* (self-organized) learning. FFNs, such as the FIR MLP, are trained in a supervised learning mode. In essence, there is a teacher present (metaphorically speaking) to guide the network toward making accurate predictions. Unsupervised learning is performed in a self-organized manner in that no external teacher or critic is required to instruct synaptic changes in the network, and is the case with SOMs. For a more thorough treatment of the general topic of neural networks, the reader is referred to Haykin (1999).

9.5.1 Motivation for Using FFNs for Degradation Signal Modeling

FFNs have proven to be very effective in function approximation and time series forecasting (Wan 1994; Sharda and Patil 1990; Tang et al. 1991; Harnik et al. 1989; Haykin 1999; Cheng and Titterington 1994; Balazinski et al. 2002). They are flexible models that are widely used to model high dimensional, nonlinear data (De Veaux et al. 1998). In fact, FFNs with nonlinear sigmoidal nodal functions are universal approximators (proved by Hornik et al. (1989), using the Stone-Weierstrass theorem), meaning that a network with finite number of hidden layers and finite number of nodes per hidden layer can approximate any continuous function (R^N, R^M) over a compact subset of R^N to arbitrary precision. However, since the FFN model parameters are generally not interpretable, they are not recommended for process understanding. However, if the emphasis is simply on accurate prediction, they tend to be extremely good and tend to outperform most traditional methods. This is not to say that traditional methods cannot be effectively used for modeling degradation signals. The method for online estimation of individual component reliability introduced in this chapter is compatible with traditional methods of modeling degradation signals as well. However, there are other motivations for using FFNs for degradation signal modeling. These include their nonparametric properties and superior ability to adapt to changes in surrounding environment (neural network trained to operate in a specific environment can be easily retrained to deal with minor changes in the environmental conditions).

9.5.2 Finite-Duration Impulse Response Multi-layer Perceptron Networks

A FIR MLP is an extension to the popular MLP network in which each scalar synaptic weight in an MLP network is replaced by a FIR synaptic filter. The additional memory allows the network *dynamic* properties necessary to make the network responsive to time-varying signals. A standard MLP network trained using an algorithm such as back-propagation is only capable of learning an input-output mapping that is static, and hence, is only capable of performing nonlinear prediction on a stationary time series (Haykin 1999). However, most degradation signals measured from physical systems, as they degrade with time, tend to be non-stationary. A typical FIR MLP network with an input layer, an output layer, and two hidden layers is shown in Fig. 9(a), whose synaptic FIR filter structure is defined by the signal-flow graph of Fig. 9(b). Here $\mathbf{x}^T = [x_1, x_2, ..., x_n]$ denotes the input vector while $\mathbf{y}^T = [y_1, y_2, ..., y_m]$ is the output vector. $\mathbf{v}^T = [v_1, v_2, ..., v_p]$ and $\mathbf{z}^T = [z_1, z_2, ..., z_q]$ are the outputs at the first and second hidden layers, respectively. $\{\mathbf{w}_{ij}^1\}_{pxn}$, $\{\mathbf{w}_{ki}^2\}_{qxp}$, $\{\mathbf{w}_{lk}^3\}_{mxq}$ are matrices of FIR weight vectors associated with the three layers. For example, $\mathbf{w}_{ij}^1 = [w_{ij}(0), w_{ij}(1), ..., w_{ij}(M)]$, where $w_{ij}(r)$ denotes the weight connected to the rth memory tap of the FIR filter modeling the synapse that connects the input neuron j to neuron i in the first hidden layer. As shown in Fig. 9(b), the index r ranges from 0 to M, where M is the total number of delay units (element z^{-1} represents unit time delay in Fig. 9(b) and s denotes a discrete-time variable) built into the design of the FIR filter. The vectors $\overline{\mathbf{v}} \in R^p$, $\overline{\mathbf{z}} \in R^q$, and $\overline{\mathbf{y}} \in R^m$ are as shown in Fig. 9(a) with $\gamma(\overline{v}_i) = v_i$, $\gamma(\overline{z}_k) = z_k$, and $\gamma(\overline{y}_l) = y_l$ where $\overline{v}_i$, $\overline{z}_k$, and $\overline{y}_l$ are elements of $\overline{\mathbf{v}}$, $\overline{\mathbf{z}}$, and $\overline{\mathbf{y}}$ respectively. Here γ is a sigmoidal nonlinear operator[1], $\overline{v}_i = \sum_{j=1}^{n} \overline{v}_{ij}$, $\overline{z}_k = \sum_{i=1}^{p} \overline{z}_{ki}$, and $\overline{y}_l = \sum_{k=1}^{q} \overline{y}_{lk}$.

During network training, the weights of the network are adjusted using an adaptive algorithm based on a given set of input-output pairs. An error-correction learning algorithm will be briefly discussed here, and readers can see Haykin (1999) for further details and information regarding other training algorithms. If the weights of the networks are considered as ele-

[1] The most popular non-linear nodal function for FIR MLP networks is the sigmoid [*unipolar* $\rightarrow f(x) = 1/(1 + e^{-x})$ and *bipolar* $\rightarrow f(x) = (1 - e^{-x})/(1 + e^{-x})$].

ments of a parameter vector $\boldsymbol{\theta}$, the error-correction learning process involves the determination of the vector $\boldsymbol{\theta}^*$ which optimizes a performance function J based on the output error.[2] In error-correction learning, the weights are adjusted along the negative gradient of the performance function as follows:

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} - \eta \frac{\partial J^{(s)}}{\partial \boldsymbol{\theta}^{(s)}} \tag{10}$$

where η is a positive constant that determines the rate of learning and the superscript refers to the iteration step. In the literature, a method for determining this gradient for FIR MLP networks is the temporal back-propagation learning, which is not repeated here due its complexity. For further information on the algorithm, see Haykin (1999) or Wan (1990).

9.5.3 Self-Organizing Maps

The principal goal of the SOM developed by Kohonen (1982) is to transform an incoming signal pattern of arbitrary dimension into a one- or two-dimensional discrete map, and to perform this transformation adaptively in a topological ordered fashion. The presentation of an input pattern causes a corresponding "localized group of neurons" in the output layer of the network to be active (Haykin 1999), introducing the concept of a neighborhood.

Let Φ denote a non-linear SOM transformation which maps the spatially *continuous input space X* onto a *spatially discrete output space* (made up of a set of N computation nodes of a lattice) A. Given an input vector $\mathbf{x}$, the SOM identifies a best-matching neuron $i(\mathbf{x})$ in the output space A, in accordance with the Map Φ. For information on the unsupervised competitive learning algorithm typically used for training SOMs (Haykin 1999). For a typical SOM, trained in such a fashion, the map Φ has the following properties (Haykin 1999):

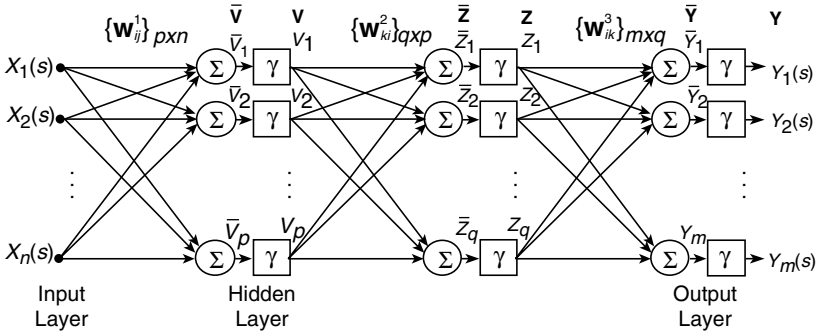
Property 1: Approximation of the Input Space–The SOM Φ, represented by the set of synaptic weight vectors $\{\mathbf{w}_j | \ j = 1, 2,..., N\}$, in the output space A, provides a good approximation of the input space X.

Property 2: Topological Ordering–The Map Φ computed by the SOM algorithm is topologically ordered in the sense that the spatial location of a

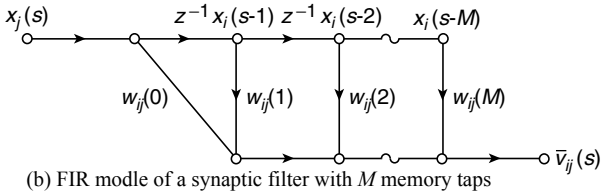
[2] A popular performance function in the literature is the sum of the squared values of the prediction error for all training patterns.

neuron in the lattice corresponds to a particular domain or feature of input patterns.

Property 3: Density Matching–The Map Φ reflects variations in the statistics of the input distribution.



(a) Architectural flow graph of a 3-layer FIR multi-layer perceptron network



(b) FIR modle of a synaptic filter with M memory taps

Fig. 9. A typical network and its structure

9.5.4 Modeling Dispersion Characteristics of Degradation Signals

The proposed approach for on-line performance reliability estimation of physical systems calls for modeling degradation signals as well as the dispersion characteristics of the signals around the degradation models. Globally generalizing neural networks such as FFNs do not easily lend themselves for modeling dispersion characteristics. In contrast, locally generalizing networks such as radial-basis function (RBF) networks and cerebellar model arithmetic computer (CMAC) networks have a naturally well-defined concept of local neighborhoods and lend themselves for modeling dispersion. Such networks have been extended in the literature to include dispersion attributes such as prediction limits (PLs). For example, the validity index (VI) network derived from an RBF network, fits functions (Park and Sandberg 1991) and calculates PLs/error bounds for its predictions (Leonard et al. 1992).

Since globally generalizing neural networks concentrate on global approximation and tend to produce highly compact and effective models, we need a method for estimating prediction limits for these networks. Most of the very few extensions to FFNs discussed in the literature that facilitate estimation of prediction intervals make the strong assumption that residual dispersion or variance in the output space is constant (for example, see De Veaux et al. 1998; Chryssolouris et al. 1996). Our experience with degradation signals reveals that this is not true. This section introduces an approach that integrates SOMs with FFNs to facilitate modeling of dispersion characteristics using FFNs without making such an assumption. In addition, the method allows the dispersion properties in the output space to be modeled using non-diagonal covariance matrices in those cases where there are multiple output variables. The intent is to utilize a SOM to introduce the concept of a "local neighborhood" even with globally approximating FFNs, critical for modeling dispersion characteristics.

Let M represent the total number of training patterns spanning the entire input space X. Let M_j the "membership" of neuron j in the discrete output space A represent the subset of training patterns from input space X that activate it. This is shown by:

$$i(\mathbf{x}) = j \text{ for all } \mathbf{x} \in M_j, \quad j = 1, 2, ..., N. \tag{11}$$

It is also true that the sum of the memberships of the neurons in the lattice output space must equal the total number of training patterns for the SOM, as shown by:

$$\sum_{j=1}^{N} M_j = M. \tag{12}$$

The three properties exhibited by SOMs (discussed earlier) provide the motivation to utilize the SOM to break the input space X into N distinct regions (denoted by X_j) that are mutually exclusive, and hence satisfy the following relationship:

$$\sum_{j=1}^{N} X_j = X. \tag{13}$$

All the signal patterns from any given distinct region X_j, when provided as input to the Map Φ, will activate the same output neuron j. This is shown by:

$$i(\mathbf{x}) = j \text{ for all } \mathbf{x} \in X_j, \ j = 1, 2, ..., N. \tag{14}$$

Thus, using SOMs, one can introduce the concept of a "local neighborhood," the resolution depending on the number of neurons (N) in the discrete output space.

From the above definition of local neighborhood, input signal patterns can be associated unambiguously with one of the distinct regions X_j. Assuming that a FFN is being used for function approximation or time series forecasting, an estimate of the covariance matrix for the FFN model residuals within the domain of region X_j is given by:

$$Cov_j = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1O} \\ S_{21} & S_{22} & \cdots & S_{2O} \\ \vdots & \vdots & \ddots & \vdots \\ S_{O1} & S_{O2} & \cdots & S_{OO} \end{bmatrix} \tag{15}$$

where:

$S_{pq} = \frac{1}{(M_j - 1)} \sum_{k=1}^{M_j} (E_{kp})(E_{kq})$, and denotes the covariance between output variables p and q,

E_{kp} denotes the FFN model residual for output variable p for pattern k,

O denotes the number of output variables predicted by the FFN.

Assuming that the residuals are independent and Gaussian distributed with a constant covariance matrix over the domain of any region but varying from domain to domain, one can even estimate the PLs. In fact, the 1-α quantile is given by the point $\mathbf{x}$ satisfying the following condition:

$$(\mathbf{x} - \boldsymbol{\mu}_j)^T Cov_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) \leq \chi_O^2(\alpha) \tag{16}$$

where:

$\chi_O^2(\alpha)$ denotes the (1-α) quantile of the Chi-Square distribution with O degrees of freedom.

$\boldsymbol{\mu}_j$ denotes the mean residual vector for domain X_j, and

Cov_j^{-1} is the inverse of the matrix Cov_j.

Investigation through simulation studies and statistical analysis have revealed that the residuals do tend to exhibit Gaussian distribution in different neighborhoods as long as the overall noise in the data is Gaussian. If the FFN has adequate representational capacity, the fit should not be significantly biased, and the mean residual vector can be a null vector. In a similar fashion, one could also determine the limits of the dispersion of the mean, i.e., the range of possible values for the mean predicted value, rather than the value for a single sample.

9.6 Drilling Process Case Study

A drilling operation was chosen as the physical test-bed for the reason that it is a commonly used machining process. El-Wardany et al. (1996) note that of all the cutting operations performed in the mechanical industries, drilling operations contribute approximately 40%. Broad steps involved and methodology used is briefly described in Fig. 10 as a guide to the case study followed-up with detailed description of the last two steps.

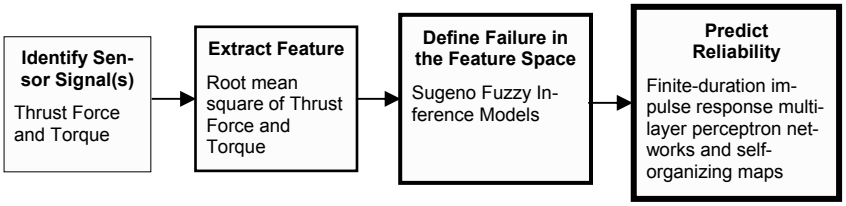


Fig. 10. Broad steps and methodology for online reliability prediction for drilling process

The development of reliable tool-wear sensors has been an active area of metal cutting research (Andrews and Tlusty 1983). Machining literature has shown that there is a strong correlation between thrust-force (and torque) acting on a drill-bit and the bit's future life expectancy (Kim and Kolarik 1992, Dimla 2000, Dimla and Lister 2000). Dimla and Lister (2000) applied multi-layer perceptron neural network for tool-state classification using online data on the cutting forces and vibration, and reported achieving approximately 90% accuracy in tool-state classification. Jantunen (2002) summarizes monitoring methods that have been studied for tool condition monitoring in drilling with thrust force and torque being most popular. Hence, thrust force and torque signals are appropriate degradation signals for estimating on-line drill-bit reliability.

As thousands of data points are recorded for thrust force and torque for each hole drilled, the data are condensed using root mean square value after systematically grouping the data points for each drilled hole. Sun et al (2006) discuss a systematic procedure for sampling the training data that helps to reduce the size of the training data without trading off the generalization performance. For defining failure in feature space, Sugeno fuzzy inference model as explained in Section 4 is used and demonstrated for the drilling process. For reliability prediction, a methodology using finite-duration impulse response multi-layer perceptron neural networks along with self-organizing maps as detailed in Section 5 is demonstrated.

9.6.1 Experimental Setup

A dynamometer was available in-house for measuring on-line the thrust-force and torque acting on the drill-bit. The experimental setup consists of a HAAS VF-1 CNC milling machine, a workstation with LabVIEW software for signal processing, a Kistler 9257B piezo-dynamometer for measuring thrust-force and torque, and a National Instruments PCI-MIO-16XE-10 card for data acquisition. The experimental setup is depicted in Fig. 11.

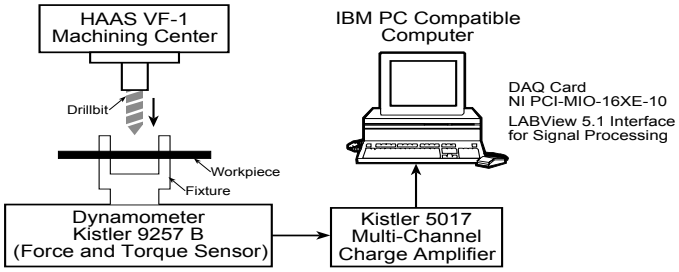


Fig. 11. Experimental setup for capturing thrust-force and torque degradation signals from a ¼" HSS drill-bit.

9.6.2. Actual Experimentation

A series of drilling tests were conducted using quarter-inch drill-bits on a HAAS VF-1 Machining Center. Stainless steel bars with quarter-inch thickness are used as specimens for the tests. The drill-bits were high-speed twist drill-bits with two flutes, and were operated under the following conditions without any coolant: feed-rate of 4.5 inches-per-minute (ipm) and spindle-speed of 800 revolutions-per-minute (rpm).

Twelve drill-bits were used in the experiment. Each drill-bit was used until it reached a state of physical failure, either due to macro chipping or gross plastic deformation of the tool tip due to excessive temperature. Collectively, the drill-bits demonstrated significant variation in life (varying between eight and twenty five successfully drilled holes) even though they came from the same manufacturer in the same box. This further validates the need to develop good on-line reliability estimation methods to help end users arrive at optimal tool or component replacement strategies.

The thrust-force and torque data were collected for each hole from the time instant the drill penetrated the work piece through the time instant the drill tip protruded out from the other side of the work piece. The data was initially collected at 250 Hz and later condensed using RMS techniques to

24 data points per hole, considered normally adequate for the task at hand. Throughout the rest of this paper, in all illustrations, one time unit is equivalent to the time it takes to drill $1/24^{th}$ of a hole. For illustrative purposes, data collected from drill-bit #8 is depicted in Fig. 12.

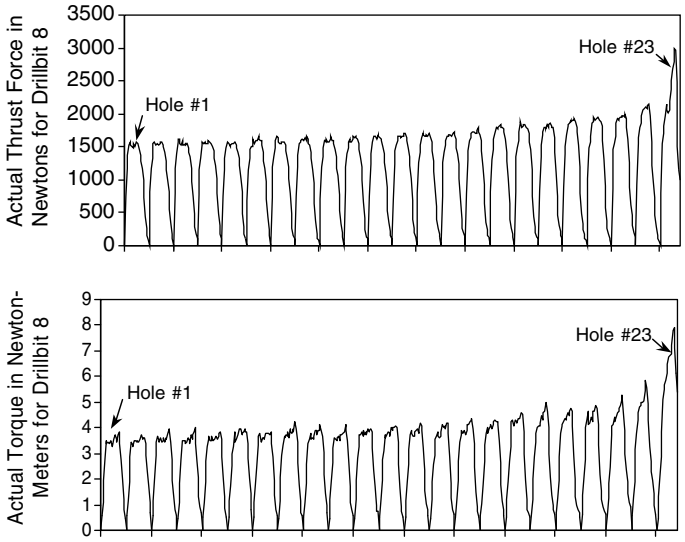


Fig. 12. Plots of thrust-force and torque signals collected from drill-bit #8.

9.6.3 Sugeno FIS for Failure Definition

Experimental data has revealed a lot of variation between drill-bits, in the amplitudes of thrust-force and torque observed during the final hole. This invalidates the concept of a deterministic critical limit for establishing failure definition in the thrust-force and torque signal space. While one could potentially introduce a probabilistic critical plane, here, we utilize fuzzy logic to introduce an FIS failure definition model in the degradation signal space.

It was decided initially to use two membership functions for representing the "low" and "high" linguistic levels for each of the degradation signals. Sigmoid membership functions were considered appropriate for three reasons:

1. They are open-ended on one side,
2. They are monotonous functions (always increase or decrease but not both), and

3. They are compatible with most ANFIS training algorithms.

Past experience suggested that thrust-force typically varies between 0 to 3000 Newtons for the drilling operation at hand. Similarly, it was common to see torque vary between 0 to 6 Newton-meters. Initially, the membership functions were set up to equally divide the ranges of the variables, as illustrated in Fig. 13(a) and 13(c) for thrust-force and torque, respectively. It was expected that ANFIS training would address any misrepresentations in these membership functions.

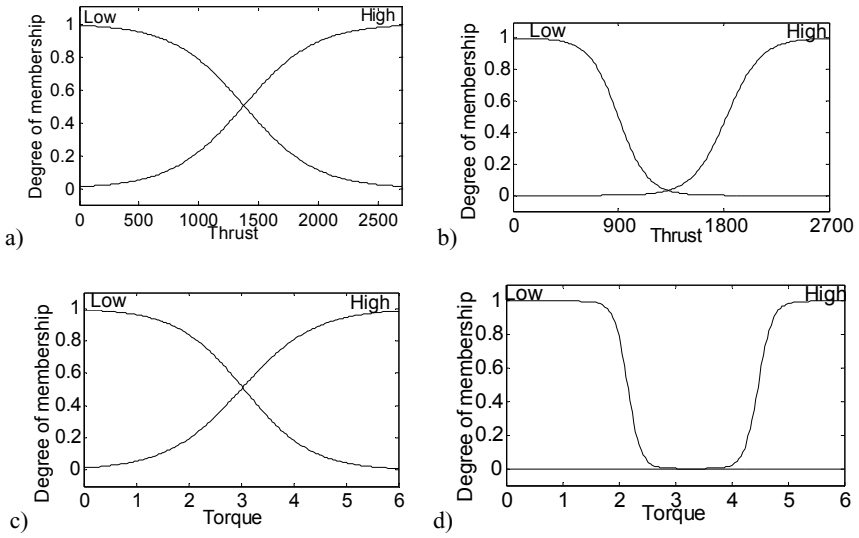


Fig. 13. Plots of membership functions before and after ANFIS training (a) Thrust-force MFs – before training. (b) Thrust-force MFs – after training. (c)Torque MFs – before training. (d) Torque MFs – after training.

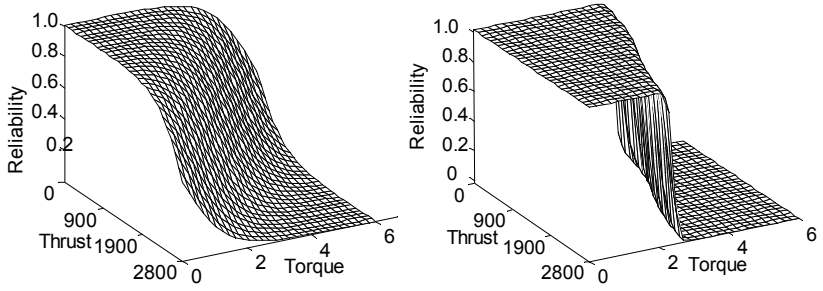
Two rules were initially formulated with the understanding that more rules can be added to address any serious violations by the FIS model. The rules are as follows:

IF thrust-force is *low* AND torque is *low*, THEN, drill-bit reliability = 1.0.

IF thrust-force is *high* AND torque is *high*, THEN, drill-bit reliability = 0.0.

Thus, the consequent of each rule constitutes a zero-order model. The resulting FIS model relationship is illustrated in Fig. 14(a). It is clear that the overall relationship is highly non-linear and certainly seems plausible. At this stage, it was decided to extract training data to further refine the FIS model using the ANFIS framework. The training, validation, and testing datasets used for developing the forecasting model were once again exploited to refine the FIS model. The reasoning behind the generation of training data is as follows. Given any drill-bit and the provided operating

conditions, it is totally reasonable to assume that the drill-bit will survive the very first hole. This implies that the FIS model should estimate the drill-bit reliability to be 1.0 when exposed to the sort of thrust-force and torque conditions witnessed during the machining of the first hole for all the eight training set drill-bits. Similarly, it is only reasonable to expect the FIS model to estimate the drill-bit reliability to be 0.0 when exposed to the sort of thrust -force and torque conditions witnessed during the machining of the last hole for all the eight training set drill-bits. Thus, in total, 16 data points were developed from the eight training set drill-bits. Validation and testing datasets were also developed similarly, using the corresponding drill-bit data. Note that while labeled data could be generated for representing extreme states of the drill-bits, it is not easily possible to develop any such data for intermediate states (i.e., states other than those representing either an extremely sharp/good or extremely dull/bad drill-bits).



(a) FIM Surface – before training. (b) FIM Surface – after training.

Fig. 14. Failure definition model surface.

Training the ANFIS formulation of our FIS model using these datasets resulted in the final relationship illustrated in Fig. 14(b). The corresponding changes to the membership functions by the ANFIS training algorithms are also illustrated in Fig. 13. Close observation of Fig. 13 and 14 reveals that the FIS model is predominantly utilizing the torque degradation signal in comparison with the thrust-force signal for estimating the online reliability of the drill-bit. This is partially attributed to the fact that torque exerted on a drill-bit is more sensitive to most of the failure modes that dominate drilling operations (i.e., it offers better signal-to-noise ratio in comparison with the thrust-force signal).

9.6.4 Online Reliability Estimation using Neural Networks

Structured experiments revealed that a FIR MLP network with the following configuration appeared to be good at maintaining generalization with respect to predicting the thrust force and torque levels into the future:

1. Input layer:
   - Number of neurons: 2
     (One for each of the two (thrust force & torque) degradation signals.)
   - Number of taps per synaptic filter: 15
2. Hidden layer
   - Number of neurons: 25
   - Number of taps per synaptic filter: 5
3. Output layer
   - Number of neurons: 6
     (Three neurons for each of the two degradation signals. First neuron is used for predicting one-step into the future, second neuron for predicting three-steps into the future, and the third neuron is used for predicting six-steps into the future. Networks with simultaneous multi-step predictions into the future outperformed networks with just one-step ahead prediction, in terms of generalization.)

Of the information collected from the 16 drill-bits, information from 12 randomly picked drill-bits was used for training purposes (labeled #1 to #12), and the information from the remaining 4 drill-bits was used for testing purposes (labeled #13 to #16). The network was designed to reduce the mean-square-error associated with testing patterns.

A SOM with a two-dimensional lattice of neurons (8×8) was used in dividing the 42 dimensional continuous input space into 64 distinct regions in an adaptive, topologically ordered fashion. The 42 dimensions are made up of (20+1) dimensional thrust force input vector and (20+1) dimensional torque input vector. The adaptive training scheme and parameter selection process discussed by Haykin (1999) was utilized in training the network. The covariance matrix, for each of the 64 distinct regions, for the FIR MLP model residuals in the output space, has been computed as per the procedure discussed in Section 5.4. Statistical analysis using Chi-Square tests and normal probability plots revealed that the residuals in distinct SOM neighborhoods tend to follow a Gaussian distribution. The reliability integral shown in equation (13) is calculated for these experiments using the Romberg method (Press et al. 1988).

The conditional performance reliability predictions for drill-bit #16 used for testing is shown in Fig. 15. All the conditional performance reliabilities are based on the assumption that the critical plane for the drill-bit with

respect to thrust is 700 pounds and critical plane for the drill-bit with respect to torque is 42.5 inch-pounds, levels set from data available from the drill-bit manufacturer and laboratory experiments. Here again, the conditional performance reliability is equivalent to mission reliability where the mission constitutes the probability of successfully drilling the hole for the next $\Delta T = T_f - T_c$, given that it has survived thus far (T_c).

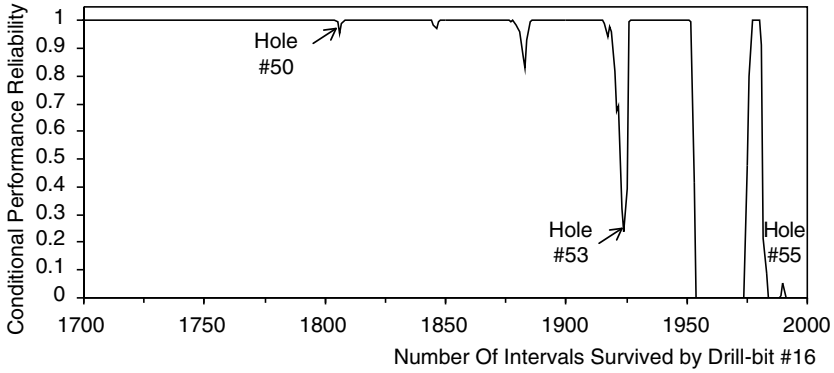


Fig. 15. Conditional performance reliability exhibited by drill-bit #16.

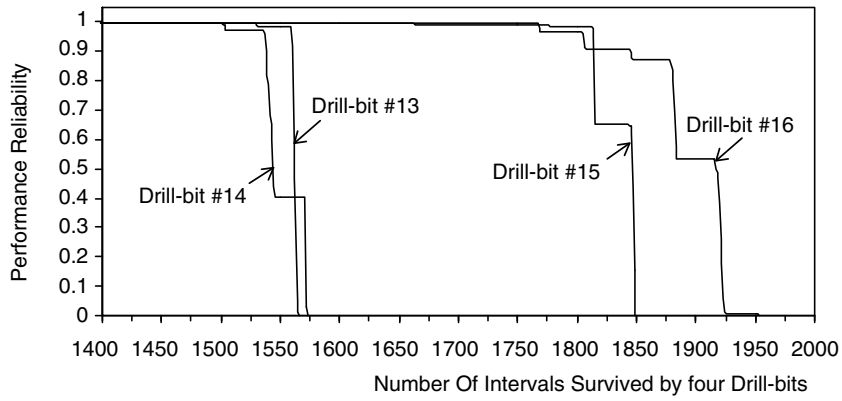


Fig. 16. Performance reliability exhibited by drill-bits #13, #14, #15, and #16.

Fig. 16 depicts the changes in the performance reliabilities (unconditional) for the 4 drill-bits used for testing. These unconditional performance reliabilities are calculated from their respective conditional reliabil-

ities developed above. For example, the performance reliability that a particular drill-bit would survive Z intervals (interval width or sampling period in time units = 1/sampling rate = 1/150 = 0.0067 seconds) is the product of the conditional performance reliabilities of surviving each of these Z intervals (basically the product of Z conditional performance reliabilities). Since the FIR MLP is of prediction order 21, these calculations are based on the assumption that each drill-bit will survive the first 21 intervals (i.e., 0.14 seconds) with certainty.[3]

9.7 Summary, Conclusions and Future Research

Traditional approaches to reliability analysis are based on life tests that record only time-to-failure. With very few exceptions, all such analyses are aimed at estimating a population characteristic or characteristics of a system, subsystem, or component. For some components, it is possible to obtain degradation measurements over time, and these measurements contain useful information regarding component reliability. Then, one can define component failure in terms of a specified level of degradation, and estimate the reliability of that "particular" component based on its unique degradation measures. This chapter demonstrates that fuzzy inference models can be used to introduce failure definition in the degradation signal space using expert opinion and/or empirical data. This is particularly valuable for carrying out prognostics activities in the absence of sound knowledge for the mechanics of degradation and/or lack of adequate failure data. The specific application considered is in-process monitoring of the condition of the drill-bit in a drilling process utilizing the torque and thrust signals. The drilling process case study demonstrates the feasibility of on-line reliability estimation for individual components using the neuro-fuzzy approach. Further, this chapter provides an approach that allows the determination of a component's reliability as it degrades with time by monitoring its degradation measures. The concepts have been implemented using finite-duration impulse response multi-layer perceptron neural networks for modeling degradation measures and self-organizing maps for modeling degradation variation. An approach to compute prediction limits for any feed-forward neural network, critical for on-line performance reliability monitoring of systems using neural networks, is also introduced by combining the net-

[3] The prediction order for a FIR MLP is equal to the sum of the memory taps for the input and hidden layers plus one for the current state. For this particular network, $p = 15 + 5 + 1 = 21$.

work with a self-organizing map. Experimental results reveal that neural networks are effective in modeling the degradation characteristics of the monitored drill-bits, and predicting conditional and unconditional performance reliabilities as they degrade with time or usage. In contrast to traditional approaches, this approach to on-line performance reliability monitoring opens new avenues for better understanding and monitoring systems that exhibit failures through degradation. Essentially, implementation of the proposed performance reliability monitoring approach reduces overall operations costs by facilitating optimal component replacement and maintenance strategies.

However, there are still several unanswered questions. For example, there is no evidence that all types of failure modes prevalent in critical equipment could be adequately captured by the proposed Sugeno FIS model. Secondly, the inability to easily generate labeled training data for the ANFIS model from intermediate states (i.e., when the unit is neither brand new nor completely worn out) might jeopardize the interpolation capability of the FIS model. This issue, however, may not be significant from a practical perspective, for in general, there isn't a lot of interest in the intermediate states, at least from the standpoint of CBM. Typically, there is no provision to estimate MRL using the proposed method for the suggested neural network forecasting models are not capable of making long-term forecasts. This is beginning to change with the introduction of the so-called structural learning neural networks (Zimmerman et al. 2002). Means to develop confidence intervals is of paramount importance as well, without which, there is no provision to gauge the accuracy of the overall prognostics procedure.

References

Akay M, Mello C (1998) Time-Frequency and Time-Scale (Wavelets) Analysis Methods: Design and Algorithms, *International Journal of Smart Engineering System Design*, Vol. 1, No. 2, pp. 77-94.

Aminian F, Aminian M (2001) Fault Diagnosis of Nonlinear Analog Circuits Using Neural Networks with Wavelet and Fourier Transforms as Preprocessors, Journal of Electronic Testing: Theory and Applications, Vol. 17, pp. 471-481.

Aminian M, Aminian F (2000) Neural-Network Based Analog-Circuit Fault Diagnosis Using Wavelet Transform as Preprocessor, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Vol. 47, pp. 151-156.

Andrews G, Tlusty J (1983) A Critical Review of Sensors for Unmanned Machining", *Annals of the CIRP*, vol 32, pp. 563-572.

Aussem A, Murtagh F (1997) Combining Neural Network Forecasts on Wavelet-Transformed Time Series, *Connection Science*, Vol. 9, pp. 113-121.

Balazinski M, Czogala E, Jemielnia K, Leski J (2002) Tool condition monitoring using artificial intelligence methods, *Engineering Applications of Artificial Intelligence*, 15, 73-80.

Bashir Z, El-Hawary ME (2000) Short Term Load Forecasting by Using Wavelet Neural Networks, *Canadian Conference on Electrical and Computer Engineering*, Vol. 1, pp. 163-166.

Begg C, Merdes T, Byington CS, Maynard KP (1999) Mechanical System Modeling for Failure Diagnostics and Prognosis, *Maintainability and Reliability Conference* (MARCON 99), Gatlinburg, Tennessee, May 10-12.

Byington CS, Garga AK (2001) Data Fusion for Developing Predictive Diagnostics for Electromechanical Systems," in *Handbook of Multisensor Data Fusion*, D.L. Hall and J. Llinas eds., CRC Press, FL: Boca Raton.

Churchland PS, Sejnowski TJ (1992) *The Computational Brain*, Massachusetts: MIT Press.

Cheng B, Titterington DM (1994) Neural Networks: A Review from a Statistical Perspective, *Statistical Sciences*, vol. 9, pp. 2-54.

Chinnam RB (1999) On-line Reliability Estimation of Individual Components Using Degradation Signals, *IEEE Transactions on Reliability*, Vol. 48, No. 4, pp. 403-412.

Chinnam RB, Baruah P (2004) A Neuro-Fuzzy Approach for Estimating Mean Residual Life in Condition-Based Maintenance Systems, Vol. 20, Nos. 1-3, pp. 166-179.

Chinnam RB, Kolarik WJ, Manne VC (1996) Performance Reliability of Tools in Metal Cutting Using the Validity Index Neural Network, *International Journal of Modeling and Simulation*, Vol. 16, No. 4, pp. 210-217.

Chinnam RB, Mohan P (2002) Online Reliability Estimation of Physical Systems Using Neural Networks and Wavelets, Smart Engineering System Design, 4, 253-264.

Chryssolouris G, Lee M, Ramsey A (1996) Confidence Interval Prediction for Neural Network Models, *IEEE Trans. on Neural Networks*, vol 7, pp. 229-232.

Cristea P, et al. (2000) Time Series Prediction with Wavelet Neural Networks, Proceedings of the 5th Seminar on Neural Network Applications in Electrical Engineering. NEUREL 2000, Piscataway, NJ, USA & Belgrade, Yugoslavia.

Daubechies I (1990) The Wavelet Transform, Time-Frequency Localization and Signal Analysis, IEEE Transactions on Information Theory, Vol. 36, pp. 961-1005.

DeVeaux RD, et al. (1998) Prediction Intervals for Neural Networks via Nonlinear Regression, *Technometrics*, vol. 40, pp. 273-282.

Devore RA, Lucier BJ (1992) Wavelets, Acta Numerica, A. Iserles, ed., Cambridge University Press, Vol. 1, pp. 1-56.

Dimla ED (2000) Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods, *International Journal of Machine Tools & Manufacture*, 40, pp. 1073-1098.

Dimla DE, Lister PM (2000) Online metal cutting tool condition monitoring. I: tool-state classification using multi-layer perceptron neural networks, *International Journal of Machine Tools and Manufacture Design, Research and Applications,* Vol. 40, 739-768.

Dimla DE, Lister PM (2000) Online metal cutting tool condition monitoring. II: tool-state classification using multi-layer perceptron neural networks, *International Journal of Machine Tools and Manufacture Design, Research and Applications*, Vol. 40, 769-781.

Dimla DE, Lister PM, Leighton NJ (1997) Neural network solutions to the tool condition monitoring problem in metal cutting - A critical review of methods, *International Journal of Machine Tools & Manufacture*, 37(9), 1219-1241.

Dzenis YA, Qian J (2001) Analysis of microdamage evolution histories in composites, *Int. J. Solids Struct*., vol. 38, pp.1831-1854.

Elsayed EA (1996) *Reliability Engineering*, Massachusetts: Addison Wesley.

Elverson B (1997) Machinery fault diagnosis and prognosis, MS Thesis, The Pennsylvania State University.

El-Wardany TI, Gao D, Elbestawi MA (1996) Tool condition monitoring in drilling using vibration signature analysis, *International Journal of Machine Tools Manufacturing*, 36, 687-711.

Fan C, et al. (2001) Detection of Machine Tool Contouring Errors Using Wavelet Transforms and Neural Networks, *Journal of Manufacturing Systems*, Vol. 20, pp. 98-112.

Fang H, et al. (2000) Adaptive Control Using Wavelet Neural Networks, Hsi-An Chiao Tung Ta Hsueh/Journal of Xi'an Jiaotong University, Vol. 34, pp. 75-79.

Flores-Pulido L, et al. (2001) Classification of Segmented Images Combining Neural Networks and Wavelet Matching, Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 4305, pp. 90-96.

Gebraeel N, Lawley M, Liu R, Parmeshwaran V (2004) Residual life predictions from vibration-based degradation signals: a neural network approach, *Industrial Electronics, IEEE Transactions on*, 51(3), 694-700.

Gu D, Hu H (2000) Wavelet Neural Network Based Predictive Control for Mobile Robots, Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Piscataway, NJ, USA.

Haykin S (1999) Neural Networks-A Comprehensive Foundation, 2nd Edition, Prentice Hall.

Hong GS, et al. (1996) Using Neural Network for Tool Condition Monitoring Based on Wavelet Decomposition, *International Journal of Machine Tools & Manufacture*, Vol. 36, pp. 551-566.

Hornik K, Stinchcombe M, White H (1989) Multi-Layer Feed Forward Networks are Universal Approximators, *Neural Networks*, Vol. 2, pp. 359-366.

Huang NE, et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. of R. Soc. Lond.*, Vol. 454, pp. 903-995.

Jang JR, Sun C, Mizutani E (1997) Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall, NJ: Upper Saddle River.

Jang JR (1993) ANFIS: Adaptive-Network based Fuzzy Inference Systems, *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 23, pp. 665-685.

Jantunen E (2002) A summary of methods applied to tool condition monitoring in drilling, *International Journal of Machine Tools & Manufacture*, 42, pp. 997-1010.

Jun Z, Fei-Hu Q (1998) A Wavelet Transformation-Based Multichannel Neural Network Method for Texture Segmentation, *Journal of Infrared and Millimeter Waves*, Vol. 17, pp. 54-60.

Kacprzynski GJ, Roemer MJ (2000) Health Management Strategies for 21st Century Condition-Based Maintenance Systems, 13th International Congress on COMADEM, Houston, TX, December 3-8.

Kapur KC, Lamberson LR (1977) *Reliability in Engineering Design*, New York: John Wiley.

Karam M, Zohdy MA (1997) Augmentation of Optimal Control with Recurrent Neural Network and Wavelet Signals, Proceedings of the American Control Conference, Vol. 3, pp. 1551-1555.

Karunaratne PV, Jouny II (1997) Neural Network Face Recognition Using Wavelets, Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 3077, pp. 202-213.

Katic D, Vukobratovic M (1997) Wavelet Neural Network Approach for Control of Non-Contact and Contact Robotic Tasks, Proceedings of 12th IEEE International Symposium on Intelligent Control, New York, NY, USA.

Ke Y, et al. (1998) The Application of Wavelet Neural Network in Spectral Data Compression and Classification, Journal of Infrared and Millimeter Waves, Vol. 17, pp. 215-220.

Kim SY, Kolarik WJ (1992) Real-time Conditional Reliability Prediction from On-line Tool Performance Data, *International Journal of Production Research*, Vol. 30, pp. 1831-1844.

Kobayashi K, et al. (1994) A Wavelet Neural Network with Network Optimizing Function, Transactions of the Institute of Electronics, Information and Communication Engineers D-II, Vol. J77D-II, pp. 2121-2129.

Kohonen T (1982) Self-Organized Formation of Topologically Correct Maps, *Biological Cybernatics*, vol 43, 1982, pp. 59-69.

Kuo RJ (2000) Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network, *Engineering Applications of Artificial Intelligence*, 13, 249-261.

Lawless JF (1982) Statistical Models and Methods for Lifetime Data, New York: John Wiley.

Lebold M, et al. (2000) Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics, *Proceedings of the 54th Meeting of Society for Machinery Failure Prevention Technology*, Virginia Beach, VA, May 1-4, pp. 623-634.

Leonard JA, Kramer MA, Ungar LH (1992) Using Radial Basis Functions to Approximate a Function and Its Error Bounds, *IEEE Trans. on Neural Networks*, vol 3, pp. 624-627.

Lewis EE (1987) *Introduction to Reliability Engineering*, New York: John Wiley.

Lippmann RP (1987) An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, vol 4, pp. 4-22.

Lu JC, Meeker WQ (1993) Using Degradation Measures to Estimate a Time-to-Failure Distribution, *Technometrics*, Vol. 35, pp. 161-172.

Lu S, Lu H, Kolarik WJ (2001) Multivariate Performance Reliability Prediction in Real-time, *Reliability Engineering and System Safety*, Vol. 72, No. 1, pp. 39-45.

Lu JC, Pantula SG (1989) A Repeated-Measurements Model for Over-Stressed Degradation Data, Technical Report, North Carolina State University, Dept. of Statistics.

Lu CJ, Park J, Yang Q (1997) Statistical Inference of a Time-to-Failure Distribution Derived From Linear Degradation Data, *Technometrics*, vol 39, pp. 391-400.

Mathur A, et al. (2001) Reasoning and Modeling Systems in Diagnosis and Prognosis, *Proceedings of the SPIE AeroSense Conference*, Orlando, FL, April 16-20.

Matsumoto K, et al. (2001) Machine Fault Diagnosis Using a Neural Network Based on Autocorrelation Coefficients of Wavelet Transformed Signals, Transactions of the Institute of Electrical Engineers of Japan, Part C, Vol. 121-C, pp. 167-176.

McAulay AD, Li J (1992) Wavelet Data Compression for Neural Network Preprocessing, Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 1699, pp. 356-365.

Meeker WQ, Escobar LA, Lu JC (1998) Accelerated Degradation Test: Modeling and Analysis, *Technometrics*, Vol. 40, pp. 89-99.

Meeker WQ, Escobar LA (1998) Statistical Methods for Reliability Data. John Wiley, New York.

Nelson W (1981) Analysis of Performance Degradation Data from Accelerated Life Tests, *IEEE Transactions on Reliability,* Vol. 30, pp.149-155.

Nelson W (1982) Applied Life Data Analysis, New York: John Wiley.

Nelson W (1990) Accelerated Testing-Statistical Models, Test Plans, and Data Analysis, John Wiley & Sons.

Nur TI, et al. (1995) Detection of Tool Failure in End Milling with Wavelet Transformations and Neural Networks (Wt-Nn), International Journal of Machine Tools & Manufacture, Vol. 35, pp. 1137-1147.

Park J, Sandberg IW (1991) Universal Approximation Using Radial-Basis-Function Networks, *Neural Computation*, vol 3, pp. 246-257.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical Recipes in C: The Art of Scientific Computing.

Pittner S, et al. (1998) Wavelet Networks for Sensor Signal Classification in Flank Wear Assessment, *Journal of Intelligent Manufacturing*, Vol. 9, pp. 315-322.

Qian S, Chen D (1996) Joint Time-Frequency Analysis, Prentice Hall.

Qing W, et al. (1996) The Application of Wavelet Neural Networks in Optimizing Electromagnetic Devices, Proceedings of the CSEE, Vol. 16, pp. 83-86.

Quan Y, Zhou M, Luo Z (1998) On-line robust identification of tool-wear via multi-sensor neural-network fusion, Engineering Applications of Artificial Intelligence, 11(6), 717-722.

Rai BK, Chinnam RB, Singh N, (2004) Tool-condition monitoring from degradation signals using MTS analysis, Conference proceedings of 21st Robust Engineering Symposium at Novi (Michigan), 343-351.

Schumacher P, Zhang J (1994) Texture Classification Using Neural Networks and Discrete Wavelet Transform, Proceedings of 1st International Conference on Image Processing, Los Alamitos, CA, USA.

Sharda R, Patil RB (1990) Neural Networks as Forecasting Experts: An Empirical Test, *Proceedings of IJCNN Meeting*, pp. 491-494.

Shashidhara HL, et al. (2000) Function Learning Using Wavelet Neural Networks, Proceedings of IEEE International Conference on Industrial Technology 2000, Mumbai, India.

Sun J, Hong GS, Wong YS, Rahman M, Wang ZG (2006) Effective training data selection in tool condition monitoring system, *International Journal of Machine Tools & Manufacture*, 46, pp. 218-224.

Sung DA, Kim CG, Hong CS (2002) Monitoring of impact damages in composite laminates using wavelet transform, *Composites: Part B*, vol. 33, pp.35-43.

Szu HH, et al. (1992) Neural Network Adaptive Wavelets for Signal Representation and Classification, Optical Engineering, Vol. 31, pp. 1907-1916.

Taguchi G, Jugulum R (2002) The Mahalanobis-Taguchi Strategy – A pattern technology system. John Wiley & Sons, New York.

Tan Y, et al. (2000) Dynamic Wavelet Neural Network for Nonlinear Dynamic System Identification, IEEE Conference on Control Applications - Proceedings, Vol. 1, pp. 214-219.

Tang Z, Almeida CD, Fishwick PA (1991) Time Series Forecasting using Neural Networks vs. Box-Jenkins Methodology, *Simulation*, vol 57, pp. 303-310.

Tang LC, Chang DS (1995) Reliability prediction using nondestructive accelerated-degradation data: case study on power supplies. *IEEE Transactions on Reliability*; **44**:562-566.

Ting W, Sugai Y (2000) A Wavelet Neural Network for the Approximation of Nonlinear Multivariable Functions, Transactions of the Institute of Electrical Engineers of Japan, Part C, Vol. 120-C, pp. 185-193, 2000.

Tomsky J (1982) Regression Models for Detecting Reliability Degradation, *Proceedings of the Annual Reliability and Maintainability Conference*, pp. 238-244.

Tseng ST, Hamada MS, Chiao CH (1995) Using Degradation Data to Improve Fluorescent Lamp Reliability, *Journal of Quality Technology*, Vol. 27, No. 4, pp. 363-369.

Vachtsevanos G, Wang P, Khiripet N (1999) Prognostication: Algorithms and Performance Assessment Methodologies, *ATP Fall National Meeting Condition-Based Maintenance Workshop*, San Jose, California, November 15-17.

Wan EA (1990) Temporal Backpropagation for FIR Neural Networks, *Proc. of the IEEE Int'l Joint Conf. on Neural Networks*, vol 1, San Diego, CA, pp. 575-580.

Wan EA (1994) Time Series Prediction by Using a Connectionist Network with Internal Delay Lines, *Time Series Prediction: Forecasting the Future and Understanding the Past* (A.S. Weigend, N.A. Gershenfeld, *Ed*), pp. 195-217.

Wu S-J, Shao J (1999) Reliability analysis using the least-squares method in nonlinear mixed-effect degradation models, *Statistica Sinica*; **9**:855-877.

Wu S-J, Tsai T-R (2000) Estimation of time-to-failure distribution from a degradation model using fuzzy clustering, *Quality and Reliability Engineering International*; **16**:261-267.

Xiaoli L, et al. (1997) On-Line Tool Condition Monitoring System with Wavelet Fuzzy Neural Network, *Journal of Intelligent Manufacturing*, Vol. 8, pp. 271-276.

Xu J, Ho DWC (1999) Adaptive Wavelet Networks for Nonlinear System Identification, Proceedings of the 1999 American Control Conference, Piscataway, NJ, USA.

Yang MY, et al. (1997) Automatic ECG Classification Using Wavelet and Neural Networks, Chinese Journal of Medical and Biological Engineering, Vol. 17, pp. 265-275.

Yen GG, Lin K (2000) Wavelet Packet Feature Extraction for Vibration Monitoring, *Industrial Electronics, IEEE Transactions on*, 47(3), 650-667.

Ying J, Licheng J (1995) Wavelet Neural Networks for Functional Optimization Problems, Proceedings of International Conference on Neural Information Processing - ICONIP `95, Beijing, China.

Yong F, Chow TWS (1997) Neural Network Adaptive Wavelets for Function Approximation, Proceedings of European Symposium on Artificial Neural Networks. ESANN '97, Brussels, Belgium.

Yu IK, et al. (2000) Industrial Load Forecasting Using Kohonen Neural Network and the Wavelet Transform, Proceedings of 2000 Universities Power Engineering Conference, Belfast, Ireland.

Zadeh LA (1965) Fuzzy sets, *Information and Control*, Vol. 8, pp. 338-353.

Zhang J, et al. (1995) Wavelet Neural Networks for Function Learning, IEEE Transactions on Signal Processing, Vol. 43, pp. 1485-1497.

Zhang Q, Benveniste A (1991) Approximation by Nonlinear Wavelet Networks, ICASSP 91: 1991 International Conference on Acoustics, Speech and Signal Processing (Cat. No.91CH2977-7), New York, NY, USA.

Zhou Q, Hong GS, Rahman M (1995) A New Tool Life Criterion For Tool Condition Monitoring Using a Neural Network, *Engineering Applications of Artificial Intelligence*, 8, 578-588.

Zimmerman HG, Neuneir R, Gothmann R (2002) Modeling of Dynamic Systems by Error Correction Neural Networks, in Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics, A. S. Soofi, L. Cao eds., Kluwer Publishing, MA: Boston, March 2002.