41

Nondeterministic Hybrid Architectures for Vehicle Health Management

Joshua Altmann
Vipac Engineers & Scientists Ltd.

Dan M. Ghiocel
Ghiocel Predictive Technologies Inc.

41.1 Introduction

This chapter presents two nondeterministic applications where hybrid architectures have been employed to provide enhanced health-management reasoning for the detection, diagnosis, and prognosis of faults (that are typically defined by some loss of system functionality).

The first application illustrates a prognostic health management (PHM) system capable of predicting faults of air or ground vehicle engines under highly transient in-operation conditions. The system's predictions also include the associated confidence or risk levels. To adequately address the complex problem of probabilistic in-operation diagnostics and prognostics, a hybrid stochastic-neuro-fuzzy inference system was developed that is a combination of stochastic parametric and nonparametric modeling techniques. This hybrid nondeterministic inference system, named StoFIS, is an integration of multivariate stochastic space-time process models with adaptive network-based fuzzy inference system models using clustering techniques [1]. StoFIS provides a hierarchical data-fusion modeling to maximize the extracted information used for diagnostic–prognostic reasoning. StoFIS is used to quantify the fault risks of an engine system at any given time and project their risk evolution in the future for risk-based prognostics.

0-8493-1180-X/04/$0.00+$1.50

The second nondeterministic application illustrates the application of discrete wavelet analysis in conjunction with an adaptive network-based fuzzy inference system to provide automated fault detection and diagnosis of rolling-element bearings. The proposed method involves the automatic extraction of wavelet packets containing bearing-fault-related features from the discrete wavelet packet analysis representation of machine vibrations. The resultant signal extracted by this technique is essentially an optimal multiple-band-pass filter of the high-frequency bearing impact transients. The discrete wavelet packet analysis multiple-band-pass filtering of the signal results in improved signal-to-noise ratios, with an exceptional capacity to exclude contaminating sources of vibration.

41.2 Application 1: Vehicle Engine Prognostics Health Management (VEPHM)

Many VEPHM technologies over the last decade have focused on the ability to classify engine performance faults as predicted by either gas-path-analysis models or maintenance personnel experience. Different technologies implemented in the past have included various advanced techniques, such as neural-network architectures, expert systems, fuzzy inference systems, empirical-based lifing algorithms, and more recently, probabilistic or stochastic modeling techniques. Each of the implementations of these VEPHM technologies brings benefits in terms of their capabilities for detection, diagnostics, and prognostics of engine faults. None of these technologies provides a complete solution to the challenge of developing a VEPHM system that is robust, reliable, and yet sensitive. The focus of more recent work in the field of VEPHM has shifted to utilizing a combination of the aforementioned technologies, and providing diagnostic reasoning based on data fusion. As greater knowledge of system behavior becomes available with the increased data collection and dissemination, improved system integration will become feasible, including the data and feature fusion of nondeterministic performance-based and vibration-based diagnostic reasoning.

At this time, there is a significant need to move to in-operation-capable VEPHM systems that can be used on a regular ongoing basis, thus providing a proactive approach rather than a reactive approach to vehicle engine maintenance. In-operation engine diagnostics and prognostics offer critical information on the engine state and functionality that is of key importance for quick, cost-effective decisions of the vehicle pilot. Also, in-operation diagnostics offer extremely useful information to maintenance engineers for preventive actions and, in some potentially catastrophic situations, to vehicle pilots for avoiding the accidental loss of the vehicle.

The application presented in this section addresses the air VEPHM problem. The previous generation of air VEPHM systems suffered from several shortcomings, which limited their use to ground test environments. In some cases, air VEPHM systems assumed that the relationship between engine parameters and rotor operating speed (or corrected speed) could be represented by one-dimensional high-order polynomial functions [2]. This polynomial fitting is suited for simple, well-controlled engine tests with constant or slowly varying operational conditions, but not for a realistic engine environment. In a realistic engine environment characterized by large and rapid variations in speed and engine inlet conditions, the polynomial fitting provides a poor approximation for the in-operation engine problem. Another significant shortcoming of the previous generation of air VEPHM systems that needs to be addressed is the limited prognostic capabilities available. This was partially due to the diagnostic reasoning of these systems concentrating on the type of fault detected, with limited assessment of the fault severity. Prognostic capabilities are thus usually limited to trending of nondimensional trend parameters taken under specific operating conditions.

These issues are addressed by the development of stochastic models that treat the progression of engine performance faults as movement into and along fault basins of attraction. The fault basins are based on multivariate stochastic modeling of performance parameter deviations from normal operating condition for given engine faults or combinations of faults.

41.2.1 Stochastic VEPHM Tools Development

The development of a probabilistic framework for engine diagnostics and prognostics, based on parameter deviations from transient gas-path-analysis (GPA) engine models, provides the basis for in-operation risk-based assessment of engine condition at any given time. By incorporating transient engine models, appropriate feature extraction, data filtering, and probabilistic reasoning, an assessment of engine condition in terms of risk can be ascertained given prior risk-association data (risk associated with given fault type and severity).

To enable robust and sensitive system performance, regardless of whether the system is operating in a ground test environment or a highly transient flight profile, nondeterministic GPA models were developed using adaptive network-based fuzzy inference system (ANFIS) models. These ANFIS GPA models were incorporated into the probabilistic reasoning model. This was in recognition of the fact that functional variations are significantly greater than the random variations in performance parameters from their normal operating conditions. The input parameters selected were based on a typical analytical GPA model, with additional parameters (power lever angle) used as substitutes for parameters that are not currently measured (fuel flow rate). The remaining error patterns described by uncertain parameter fluctuations were substantially a result of random quantities, although some degree of functional variability remains due to input factors not modeled and a limited size of the statistical database. A robust probabilistic diagnostic-prognostic VEPHM system has to consider both the functional and random aspects of the fault patterns. Three different classes of engine performance models were developed, namely the overall quasi-stationary model (OQS), the overall transient model (OT), and the partial transient model (PT).

Three levels of data fusion are embedded into StoFIS. The data- and feature-level fusions are nondeterministic data-fusion processes, whereas the decision-level fusion method is deterministic. This procedure enables in-operation variability and data uncertainty to be taken into account, and provides a robust diagnostic-prognostic output of engine health. The three major components that form the basis for the StoFIS development are outlined below.

41.2.1.1 Feature Vector Extraction

The first component of StoFIS is a data-level fusion procedure, where data from multiple sensors are fused into ANFIS engine models prior to feature extraction. This fusion process utilizes data from a sparse sensor array to track performance parameter deviations from normal operating conditions under typical real-life environments. Basically, in StoFIS, engine behavior is nondeterministically modeled using an ANFIS model, with the system design derived from a physics-based quasi-stationary GPA model. It involves a two-step computational process; the initial membership functions (likelihood functions) and fuzzy-logic rules are formed through subclustering, followed by a fine-tuning of the system with a loosely coupled procedure using least squares and a backward-propagation neural network. The procedure enables system adaptation to include *engine–engine variability*, as well as the provision for *quasi-stationary,* highly *transient,* and *partial-engine transient* implementations of the models. Synergies derived from the multiple modeling procedures enable more-sensitive fault detection under adverse conditions, increased confidence in diagnostics, and improved multiple-fault discrimination.

41.2.1.2 Multivariate Reliability Analysis

The second component of StoFIS involves the reliability analysis using multivariate stochastic feature vectors extracted from the ANFIS engine models. Multisensor stochastic feature vectors are used to characterize the system and arrive at a diagnostic-prognostic output for each fault type. This is the most critical step in the fusion process, with diagnostic and prognostic reasoning based on stochastic parameter deviations due to off-specification conditions. This stochastic modeling has been termed as the stochastic fault basin, as it models the stochastic deviations in multidimensional parameter space, with the development of faults treated as progression into and along a fault basin (not a specific center point for a given severity level) of attraction. Prognostics are based on the location with respect to and rate of progression toward a fault basin of attraction. The outputs from the stochastic feature-level fusion are a reliability index (RI), used for

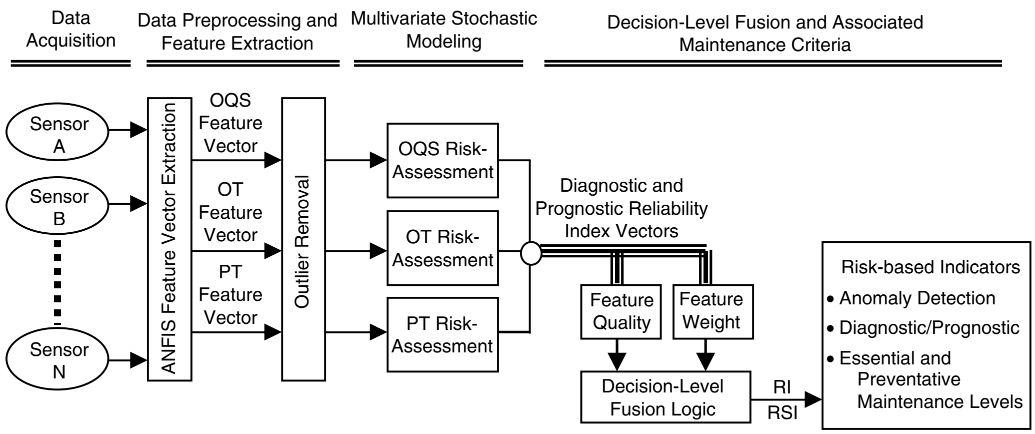

FIGURE 41.1 Illustration of StoFIS architecture.

diagnostics, and a reliability sensitivity index (RSI), used for prognostics. These two reliability indices can be associated with the essential maintenance and preventive maintenance events, respectively. Separate outputs are available from the three basic nondeterministic ANFIS GPA engine models.

41.2.1.3 Decision-Level Fusion

The third component of StoFIS is the decision-level fusion, where diagnostic outputs from separate engine models, OQS, OT, and PT, are fused after diagnostic reasoning to enhance the confidence of the final output. The purpose of this procedure is to take advantage of the synergies provided by using the multiple engine models. The method employed is a heuristic fusion procedure based on a scoring method that incorporates knowledge-based rules. The reliability indices computed using the three separate engine models are weighted based on the fault-detection index and the fault-discrimination sensitivity index of the engine models for given fault conditions. The combined effects of the data-level and decision-level fusion technologies developed during this project have indicated improvements in the fault-diagnostic resolution for in-operation conditions of between 100 and 1000%, depending on the location and type of fault in the engine compared with corrected rotor-speed-based VEPHM systems. Figure 41.1 shows a schematic illustration of the StoFIS architecture.

41.2.2 Nondeterministic Adaptive Network-Based Fuzzy Inference Models

The adaptive network-based fuzzy inference system, which is utilized in both of the studies illustrated in this chapter, is a transformational model of integration where the final fuzzy inference system is optimized via artificial neural-network training. ANFIS has the ability to either incorporate expert knowledge or use subtractive clustering to form its initial rule base. In both cases, ANFIS maintains system transparency while allowing tuning of the fuzzy inference system via neural training to ensure satisfactory performance. The validity of the expert knowledge and the suitability of the input data chosen can then be verified by examining the structure and the performance of the final fuzzy inference system. This section describes the design and operation of an ANFIS [3].

The initial membership functions and rules for the fuzzy inference system can be designed by employing human expertise about the target system to be modeled. ANFIS can then refine the fuzzy if–then rules and membership functions to describe the input/output behavior of a complex system.

If human expertise is not available, it is possible to intuitively set up reasonable membership functions and then employ the neural training process to generate a set of fuzzy if–then rules that approximate a desired data set. Sugeno-type fuzzy inference systems have been used in most adaptive techniques for constructing fuzzy models because they are more compact and provide a more computationally efficient

1180_C41.fm Page 5 Thursday, October 21, 2004 7:56 PM

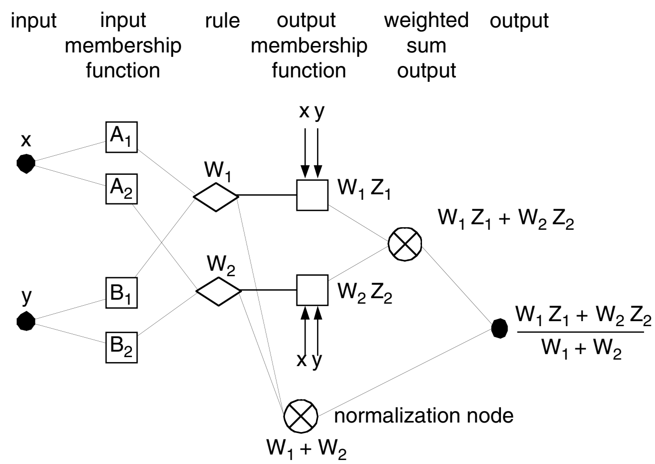

FIGURE 41.2 ANFIS Sugeno fuzzy model.

representation of data than the Mamdani or Tsukamato fuzzy systems. A typical fuzzy rule in a zero-order Sugeno fuzzy system has the form:

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B\text{, then } z = c \tag{41.1}$$

where A and B are fuzzy sets in the antecedent, and z is a crisply defined function in the consequent. It is frequently the case that the singleton spike of the crisply defined consequent is completely sufficient to cater to a given problem's needs. If required, the more general first-order Sugeno can be employed by setting the consequent to $z = px + qy + c$. Higher-order Sugeno systems add an unwarranted level of complexity, with minimal remuneration. A zero-order Sugeno fuzzy inference system is used in this investigation. The equivalent ANFIS architecture for a Sugeno fuzzy inference system is illustrated in Figure 41.2.

The nodes in the *input membership function layer* are adaptive. Any appropriate membership functions can be used to describe the input parameters. The outputs of this layer, $\mu_{A_i}(x)$ and $\mu_{B_i}(x)$, are the membership values of the premise where x and y are the node inputs, and A_i and B_i are the associated fuzzy sets. The output of the fixed nodes in the *rule layer* represents the fuzzy strengths of each rule. Either the product or minimum rules can be used to calculate the weighting function for the fuzzy operator "AND" of a Sugeno fuzzy inference system.

$$\text{Product: } W_i = \mu_{A_i}(x) \times \mu_{B_i}(x) \tag{41.2}$$

$$\text{Minimum: } W_i = \min\{\mu_{A_i}(x), \mu_{B_i}(x)\} \tag{41.3}$$

The adaptive nodes in the *output membership function layer* calculate the weighted output of the consequent parameters, as given by

$$W_i Z_i = W_i C_i \tag{41.4}$$

The *weighted-sum output layer* consists of a single fixed node. The weighted-sum output is the summation of the weighted output of the consequent parameters,

$$\sum_i W_i Z_i \tag{41.5}$$

The final *output layer* is the normalized weighted output given by

$$\frac{\sum_i W_i Z_i}{\sum_i W_i} \tag{41.6}$$

The *normalization node* connects the rule layer to the output layer in order to normalize the final output. The normalization factor is calculated as the sum of all weight functions.

$$\sum_i W_i \tag{41.7}$$

Although any feed-forward network can be used in an ANFIS, Jang and Sun [4] implemented a hybrid learning algorithm that converges much faster than training that relies solely on a gradient-descent method. During the forward pass, the node outputs advance until the output membership function layer, where the consequent parameters are identified by the least-squares method. The backward pass uses a back-propagation gradient-descent method to upgrade the premise parameters, based on the error signals that propagate backward. Under the condition that the premise parameters are fixed, the consequent parameters determined are optimal. This reduces the dimension of the search space for the gradient-descent algorithm, thus ensuring faster convergence. This hybrid learning system is used in the training of the fuzzy inference systems used for both applications presented. A more-detailed explanation of ANFIS can be found in the literature [3, 4].

41.2.3 Feature Vector of Engine Behavior

Figure 41.3 shows a sketch of the investigated turbofan jet engine including typically installed sensors. Figure 41.4 and Figure 41.5 show pressure variations as a function of the high-pressure shaft speed for testing data and in-operation conditions, respectively. It is obvious from these figures that the pressure closely follows a nonlinear relationship with shaft speed for slowly varying testing conditions. However, for in-operation conditions, the pressure deviates from this nonlinear path due to highly transient conditions and significant changes in the inlet conditions, namely inlet pressure, temperature, and mass flow. This means that using deviations from a fitted polynomial regression line for diagnostics, as is commonly used in engine health-monitoring applications based on test data, is not suited to in-operation conditions. In fact, a large stochastic

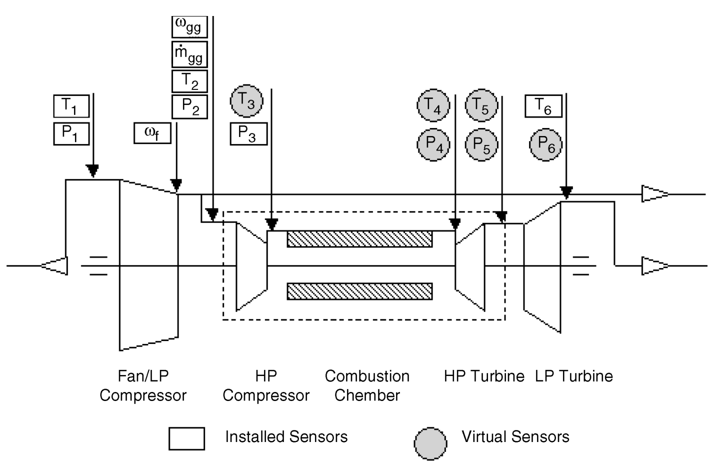

FIGURE 41.3 Schematic of a typical sparse sensor array. (T-temperature, P-pressure, ω_f-fan rotor speed, ω_{gg}-gas generator rotor speed, $\dot{m}_{gg}$-gas generator mass flow rate.)

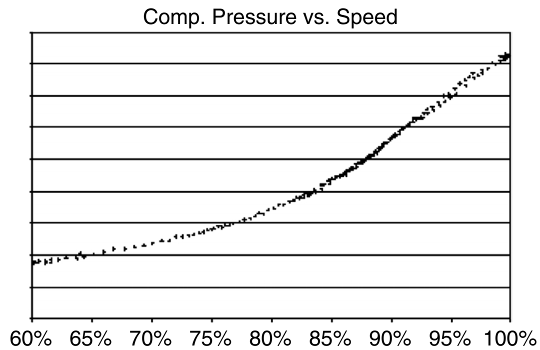

FIGURE 41.4 P_3 vs. ω_{gg} for ground tests.

variability projected on the pressure–speed plane in Figure 41.5 is apparent. This large variability is mostly due to the transient variations induced by the vehicle's pilot maneuvers. A key aspect for getting realistic predictions for in-operation conditions is to separate the *true statistical variabilities* (random part) from the *functional variabilities* introduced by engine transient behavior. For real, in-operation transient conditions, the functional dependence between engine performance parameters becomes complex and highly nonlinear. If these transient complex functional dependencies between multiple parameters are ignored, then the statistical variability is overestimated and the computed fault risks are unreliable.

Physics-based, analytical GPA models only cater to quasi-stationary engine operation, which provides limited ability to track engine parameters during the transitory operating conditions encountered during typical operation of a vehicle. An alternative scheme capable of including the highly transient in-flight conditions has been developed based on deviations from physics-based empirical GPA models of the performance parameters. This leads to the formation of a stochastic engine GPA model, developed by building and calibrating a generic analytical GPA model for a given engine and then tuning its stochastic input–output using the nondeterministic ANFIS GPA model based on a statistical training data set from typical operational data.

The ANFIS model was developed based on subtractive clustering using the modified mountain technique [5]. Subtractive clustering is a fast one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data and provides a powerful tool to form the initial fuzzy inference system (FIS). Then, using a neural network (NN), the FIS was fine-tuned using least-squares training on the forward-pass and back-propagation on the backward pass, resulting in a final FIS in the ANFIS model that includes the NN-based correction. The final ANFIS GPA engine model provides an engine-specific

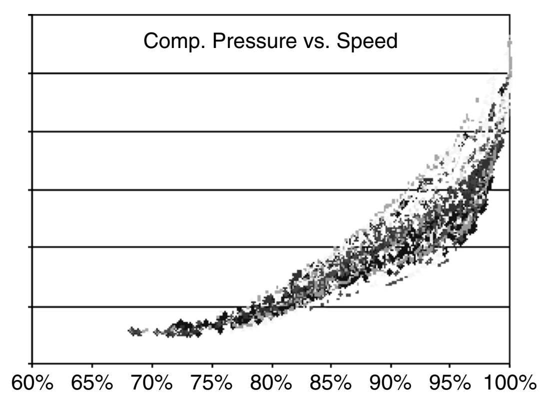

FIGURE 41.5 P_3 vs. ω_{gg} for in-operation (flight) conditions.

physics-based empirical GPA model that is capable of accurately predicting normal-condition pressures and temperatures at each section of the engine under transient operating conditions. The ANFIS GPA models were developed and tested with nine different operating-condition spectra. Three operating spectra were used for the NN training; three were used to check the fuzzy inference system for overtraining; and three were used to test the final system. The developed ANFIS models performed robustly for all of the data sets considered, with only minor flight-profile dependencies. The ANFIS models provide two functions: (1) an indicator of engine or sensor malfunction based on statistical deviations from measured performance parameters; and (2) virtual sensors in the event that physical sensors are not mounted at certain compartments in production models of an engine. Virtual sensors are normally used to provide lifing estimates for the critical engine components.

Updating of ANFIS to cater to variations due to engine-to-engine variability for engines within a given engine class can be accommodated by fine-tuning based on tests prior to in-operation implementation. Alternatively, calibration can be performed based on in-operation data collected from a specific engine. In this case, the vehicle's pilot-to-pilot variability in addition to operation-to-operation variability can be included. The deviations from the original ANFIS GPA models can be used to rate an engine's performance in relation to the standard engine characteristics for a given engine class. This provides an important quality-control function for the engine-manufacturing process. The deviation from the calibrated ANFIS GPA models can be used to detect, diagnose, and provide prognosis of deterioration in the performance of the specific engine.

41.2.4 Sparse Sensor Array Studies

Three key issues need to be resolved in the modeling of the engine performance: (1) maximization of the useful information expressed in terms of system outputs available for diagnostic–prognostic reasoning; (2) incorporation of multiple engine faults for detection; and (3) isolation of their respective contributions for diagnostic–prognostic purposes. Due to production engines currently having a very limited number of onboard sensors, it was of interest to develop a sparse-sensor-array implementation that could satisfy these three key issues. For example, we assume that a typical production engine can be confined to an output parameter space of P_2, T_2, P_3, T_6, ω_f, and $\dot{m}_{gg}$. In order to maximize the available information, the three nondeterministic ANFIS GPA models, OQS, OT, and PT, were used.

The two overall engine models (OT and OQS) have a qualitatively different behavior. The OT model eliminates feedback of downstream faults on the predicted performance parameters, while the OQS model includes this feedback effect, thus providing additional parameters and more-complex interactions for diagnostic reasoning. Detection of multiple cascaded engine faults is not easily achieved when limited to overall engine models, especially when a downstream engine fault is shadowed by an upstream engine fault. Thus a partial-engine model (PT) was also introduced. The functional basis for the three GPA models is described in the following three subsections.

41.2.4.1 Overall Quasi-Stationary (OQS) Engine Model

The OQS engine model is similar to measuring performance parameters as a function of corrected high-pressure rotor speed. It has the advantage of additional outputs compared with the transient models above, in the form of mass-flow-rate and low-pressure shaft speed, however it does not perform as well under highly transient operating conditions. The following relationship was assumed in the OQS model:

$$P_n, T_n, \dot{m}_{gg}, \omega_f = f(P_1, T_1, \omega_{gg}, \text{PLA}) \tag{41.8}$$

Parameters P_n and T_n represent the compartment static pressure and temperature parameters indicated in Figure 41.3. Significant functional errors were present due to the inability of the quasi-stationary model to cater to the highly transient conditions encountered in operation. To reduce the function errors in the quasi-stationary model, the available parameters were assessed for their ability to decrease the residual errors. The power lever angle (PLA) was the only parameter that significantly reduced the model errors.

41.2.4.2 Overall Transient (OT) Engine Model

The OT engine model takes transients into account by including the low-pressure shaft speed and flow rates. The air mass flow is denoted by $\dot{m}_{gg}$, and ω_f and ω_{gg} denote the fan and gas-generator rotor speeds, respectively. The following relationship was assumed in the OT model:

$$P_n, T_n = f(P_1, T_1, \dot{m}_{gg}, \omega_f, \omega_{gg}) \tag{41.9}$$

Significant functional errors were present in the postcombustion chamber the OT GPA model-predicted temperatures. This error is due to the lack of a fuel flow-rate measurement, which currently cannot be directly measured. Unlike the quasi-stationary model, PLA only provided a marginal reduction in the residual errors, with the errors in some parameters increasing.

41.2.4.3 Partial-Engine Transient (PT) Model

The PT model uses the previous temperature and pressure available in the sensor array as inputs to the ANFIS GPA model. This isolates contributions to deviations in the measured parameters from model predictions due to faults at earlier stages of the engine. The benefits of the PT model are highly dependent on the number of sensors installed on the engine. As the number of sensors is increased, the capability of the PT model to distinguish between faults and to identify multiple faults in the engine is augmented significantly. The PT models can be written in their functional form as:

$$P_n, T_n = f(P_{n-p}, T_{n-q}, \dot{m}_{gg}, \omega_f, \omega_{gg}) \tag{41.10}$$

where *n-p* and *n-q* represent the location of the previous sensor in the array. Although the PT models do not provide a true compartmentalization of the turbofan engine, they still bring important complementary information that adds to the overall engine models and can be used to enhance engine diagnostics.

Figure 41.6 and Figure 41.7 illustrate the mean deviations in the measured parameters for each of the models for two different fault conditions of a turbofan engine. Two important benefits of the multiple-model approach are illustrated in Figure 41.6. First, although the OQS model indicates a probable fault, it would not provide as accurate an indicator of severity as the OT model due to the high scatter in

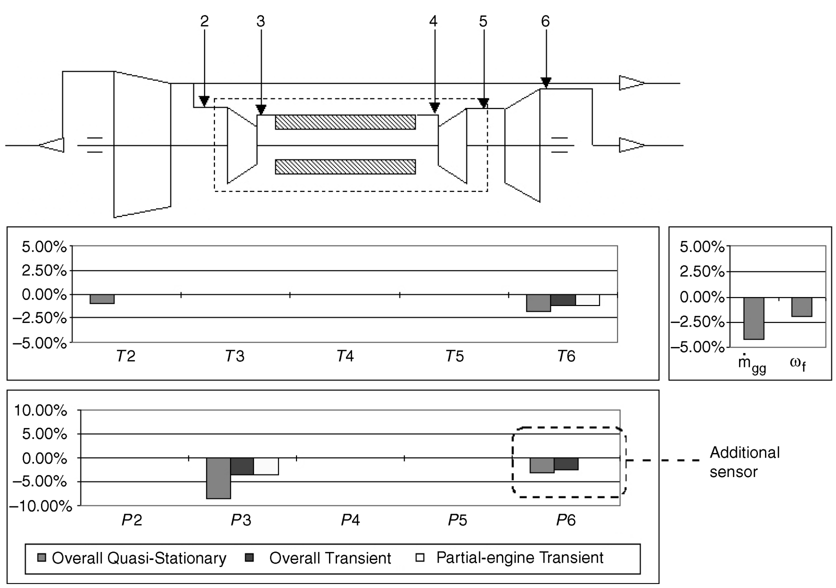

FIGURE 41.6 Sparse sensor array mean parameter deviations for 2% drop in HPC capacity.

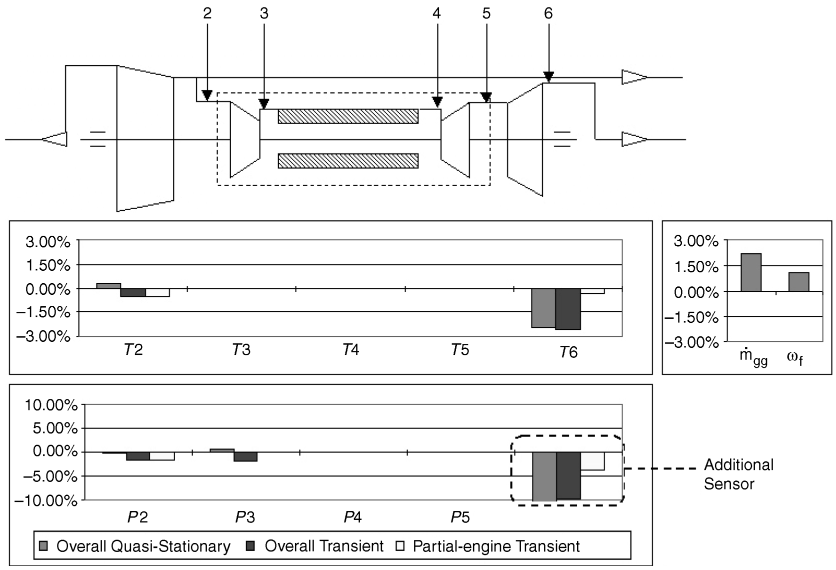

FIGURE 41.7 Sparse sensor array mean parameter deviations due to a combined fault (2% drop in fan capacity and a 1% drop in HPT efficiency).

parameter deviations. In addition, the OQS model would not sufficiently discriminate between the following faults: High Pressure Compressor (HPC) capacity drop, Low Pressure Compressor (LPC) capacity/efficiency drop, and High Pressure Turbine (HPT)/Low Pressure Turbine (LPT) capacity increase. This is due to the relatively small, normalized sensitivity indices of the OQS model for these faults. In contrast, the OT and PT models clearly isolate the fault as being located in the HPC compartment, although they are not as sensitive in distinguishing between an efficiency- and capacity-related fault. A logical combination of these results allows the fault to be isolated as a HPC capacity drop.

The benefits of the multiple-model approach at detecting and discriminating multiple faults are illustrated in Figure 41.7. In this example, the OT and PT models strongly support the supposition that there is a drop in fan capacity, while the OQS model is less definitive due to the feedback from the HPT efficiency drop confusing the issue. In this situation, the characteristic deviations for the OQS model would lie somewhere between the drop in fan-capacity and HPT-efficiency basins of attraction in multidimensional parameter space. This could result in a reduction in confidence of the diagnostics based on the OQS model due to potential confusion with other faults or combinations of faults.

The benefit of this is that the OQS model indicates that there is a high probability of faults existing in more than one compartment of the engine. As the OT model is relatively insensitive to the HPT fault, the OT model provides a more robust indicator of the drop in fan capacity, but it is unable to detect the existence of a second fault. The PT model, on the other hand, is able to isolate the drop in fan capacity and indicates either a drop in LPT/HPT efficiency or an increase in LPT/HPT capacity. Linking this knowledge with the deviations in the OT and OQS model, it is possible to eliminate LP capacity increase as a potential fault. Thus diagnostic reasoning after preliminary fusion of the three models could isolate the primary fault as a drop in fan capacity, with a secondary fault located in the HPT compartment (either efficiency drop or capacity increase).

41.2.5 Comparison of ANFIS Model Fault-Detection Sensitivity

Figure 41.8 illustrates the normalized sensitivity index of the three ANFIS models for fault detection under typical transient in-operation conditions, based on the sparse sensor array illustrated in Figure 41.3. The normalized sensitivity index is an indicator of each model's sensitivity in detecting the presence of an engine fault in a given compartment, relative to corrected rotor-speed-based measurements. The fault-sensitivity

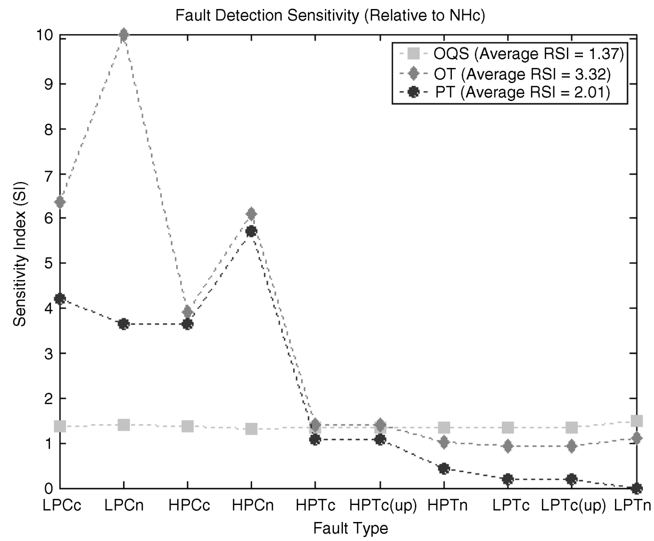

FIGURE 41.8 Relative model sensitivity: OQS, OT, and PT.

index is illustrated for changes in capacity and efficiency of each engine compartment, with the average model sensitivity (no weighting given for the importance of fault location) displayed in the legend.

The sensitivity index (*SI*) is based on the Euclidean distance between the mean deviation vectors in transformed space. The definition of SI for a given fault (*f*) condition is as follows:

$$SI(f) = \sqrt{\sum_{i=1}^{n} (\mu(f)_i^* - \mu(nc)_i^*)^2} \tag{41.11}$$

where *i* represents the measurable parameters for fault detection (i.e., pressures, temperatures, rotor speed, and flow rate), $\mu(f)^*$ is the mean parameter deviation in transformed space for a given fault condition, and $\mu(nc)^*$ is the mean parameter deviation in transformed space for an engine in normal condition (≈0).

The transformed space is based on the relative standard deviation of the normal-condition parameter errors for the three ANFIS models. The transformation enables the mean parameter variations to be considered with an equal weighting of importance, with the models using a common reference σ_{ref} to allow the performance of the models to be compared directly. The definitions of μ_i^* and σ_{ref} are given as:

$$\mu_i^* = \frac{\sigma_{\text{ref}}}{\sigma(nc)_i} \tag{41.12}$$

$$\sigma_{\text{ref}} = \min\left\{\sigma(nc)_{i,\text{Model}}/; i \in 1{:}n, \text{Model} \in [\text{OQS}, \text{OT}]\right\} \tag{41.13}$$

The fault-detection capabilities of the models are dependent on both the type of fault and the number/location of the onboard sensors. The OT and PT models are more sensitive that the OQS model for detecting compressor faults, whereas the OQS model is the most sensitive for turbine faults. This indicates the critical need for an in-operation VEPHM system to use intelligent fusion of the three models to take advantage of the provided synergies.

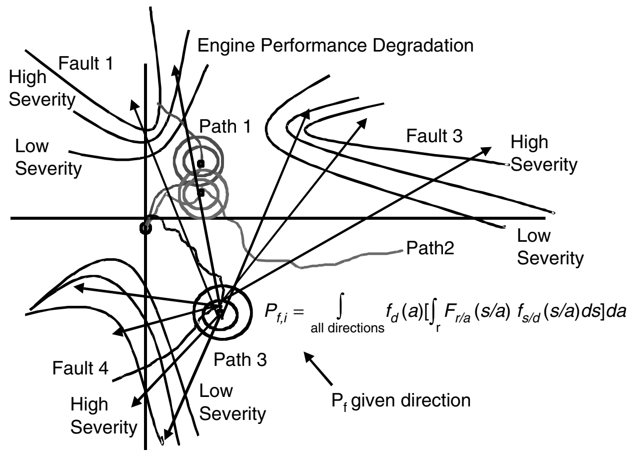

FIGURE 41.9 Engine performance degradation.

41.2.6 Multivariate Reliability Analysis Using Stochastic Fault-Basin Concept

As discussed in the previous section, engine parameters measured on-line include pressures, temperatures, and fuel flows in different compartments of an engine. The proposed stochastic fault-diagnostic–prognostic procedure is illustrated in Figure 41.9 using a two-dimensional stochastic parameter space representation [1, 2]. As shown in Figure 41.9 for the usage path 3, the engine condition at a given time can be diagnosed by evaluating all the risks of potential engine faults. Figure 41.10 shows the engine performance degradation from usage point P1 to usage point P2. This degradation is shown in the original parameter space, X-space, and in transformed standard Gaussian space, U-space that is used typically for reliability estimate calculations. Herein, the engine reliability is measured by the fault-reliability index, which is similar in concept to the traditional reliability index used in structural reliability theory computed in a transformed standard Gaussian parameter space.

To compute the fault-reliability index, the performance safety margin or the performance function first needs to be defined. The performance safety margin in engine-parameter space was simply defined by the stochastic distance between the measurement-variability ellipsoid (cluster) and the fault-variability ellipsoid (cluster), as shown in Figure 41.11. Figure 41.11 shows that this distance can be defined in two ways: (1) safety margin of Type A, a linear distance between the two multidimensional ellipsoids, and (2) safety margin of Type B, an arc length defined by the curvilinear usage trajectory. The curvilinear

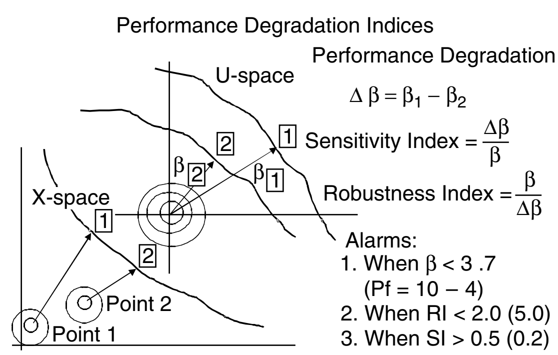

FIGURE 41.10 Standard space reliability model.

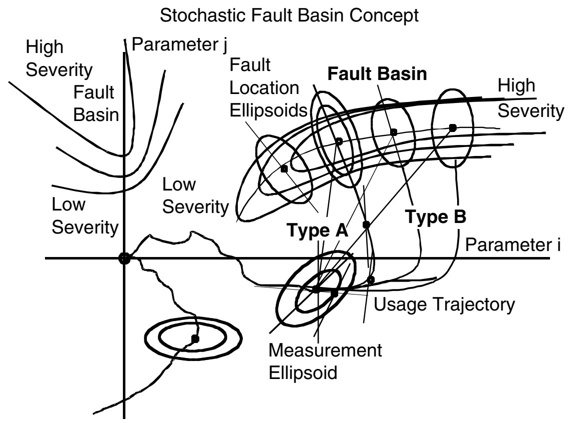

FIGURE 41.11 Illustration of stochastic fault-basin concept.

safety margin B, in comparison with the approach using linear safety-margin A, ensures a more complete compatibility between the measurement and the fault-complex patterns.

There are two important aspects related to stochastic fault diagnostics:

1. *Single fault pattern*: The single-fault-pattern modeling needs a continuous representation in the parameter space. This stochastic representation is called herein the stochastic fault-basin (SFB) concept. The fault-point location model is a truncated representation that can produce erroneous diagnostics and prognostics. More appropriately, faults should be represented in the parameter space as basins of attraction rather than point locations, as illustrated in Figure 41.11. Thus, the reliability index has to be computed with respect to a continuous fault path from low-severity to high-severity levels, and not with respect to a particular fault location.
2. *Multiple-fault pattern*: If multiple faults with different severities are simultaneously present, then it is necessary to decompose the multivariate statistical measurement in the fault patterns before defining the fault safety margins. The stochastic representation for multiple faults is called herein the stochastic multiple-fault map (SMFM) concept. Figure 41.12 illustrates that directly using the statistical measurement reference point M for reliability computations may hide the existence of simultaneous faults. The safety margins computed for location M are quite large, and there is no imminent fault detected. If the measurements in the two fault patterns are decomposed, the two reference measurement points M1 and M2 are determined, with one reference point for each fault

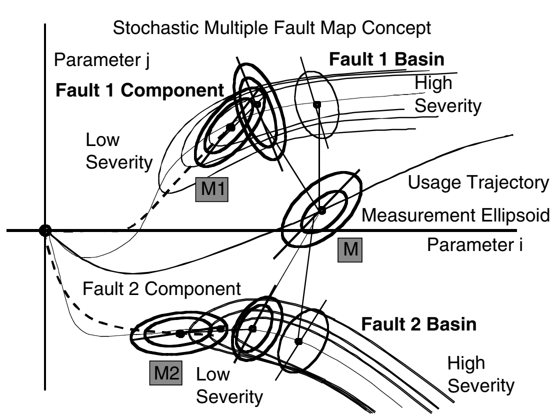

FIGURE 41.12 Illustration of multiple-fault-basin map concept.

(measurement fault component). If the reliability is computed using the two reference points M1 and M2 instead of the single measurement point M, the results are very different. In this last situation, the simultaneous faults 1 and 2 are detected with different severities.

To diagnose the engine fault and provide prognostics for in-operation conditions, reliability indices were computed for any point on the predicted trajectory within the fault basin of attraction. The reliability index computed for current measurement location is used for fault diagnostics. Reliability indices with computed locations on the future projected usage trajectory, from a predicted location to the fault location, are used for prognostics. The associated fault diagnostic and prognostic probabilities, P_f, in the multidimensional parameter space are approximated based on the computed reliability index, β

$$P_f \approx \Phi(-\beta) \tag{41.14}$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function.

To determine the usage rate in probabilistic terms (measurement location speed on the trajectory), the reliability index gradients are required. Specifically, two reliability-index sensitivity measures are introduced: a cumulative sensitivity index and an evolutionary sensitivity index. The cumulative reliability sensitivity index (CRSI) is defined by the "global" nondimensional variation of the reliability index, β (the relation between "failure probability," here read as fault-diagnostic probability, and reliability index is discussed on the next page) from initial state, at 0, to the final state, at t (over the interval $[0, t]$):

$$C_{0,t} = -\frac{\beta_t - \beta_0}{\beta_0} = -\frac{\Delta\beta_{ot}}{\beta_0} \tag{41.15}$$

The evolutionary reliability sensitivity index (ERSI) is defined by the "local" nondimensional variation of the reliability index from an intermediary state, at time ti, to another intermediary state, at time $ti+1$ (over the interval $[ti, ti+1]$)

$$E_{ti,ti+1} = -\frac{\beta_{ti+1} - \beta_{ti}}{\beta_{ti}} = -\frac{\Delta\beta_{ti,ti+1}}{\beta_{ti}} \tag{41.16}$$

These two reliability-sensitivity indices indicate (in percentage) the changes in engine reliability. A zero value indicates no safety (performance) degradation, while a positive value indicates a safety (performance) degradation, and a negative value indicates a safety improvement. Robustness indices (RI) can be defined as the inverse of sensitivity indices (SI). After model calibration, "red" alarms can be set to a lower bound of reliability index of 3.70 (equivalent to fault probability of 0.0001). A CRSI of 0.5 or, equivalently, a CRRI of 2.0, and an ERSI of 0.2 or, equivalently, an ERRI of 5.0, can be set as "yellow" alarms.

Figure 41.13 through Figure 41.16 show the computed reliability index and cumulative sensitivity reliability index (CRSI) for a fan fault that produces a 3% efficiency drop. The reliability computations are done using the dimensional variations of engine parameters. It should be noticed from these figures that the reliability index value is zero and the CRSI value is unity for this fault type. The computed values indicate that the fan fault defined by 3% efficiency has certainly occurred if the measurement and fault location for a given overlap (probability of occurrence is 50%).

It is interesting to note — by comparing Figure 41.13 and Figure 41.14 with Figure 41.15 and Figure 41.16 — the different qualitative behavior of the OQS and the OT models. The OQS GPA model incorporates a significant "feedback" effect that is visible in the reliability estimate profile computed along the engine profile (Fault 7 is at inlet, Fault 1 is at outlet). This is due to the fact that transients are not captured well by this model, so that they add larger variability in the statistical deviations of engine parameters.

In contrast, the OT GPA model shows as a "forward" model for which the reliability estimate profile indicates a monotonic growth that shows a gradual increase of statistical deviations. Higher reliability indices indicate that the statistical deviations are much larger than those that correspond to fault-severity levels between 1 and 3% in the compartments other than the fan compartment. This can be a critical

1180_C41.fm Page 15 Thursday, October 21, 2004 7:56 PM

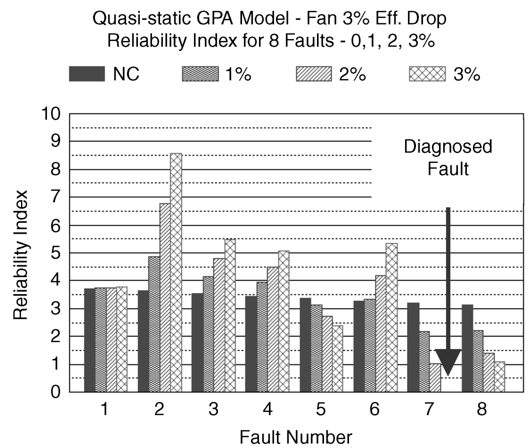

FIGURE 41.13 Computed reliability index using the OQS GPA model.

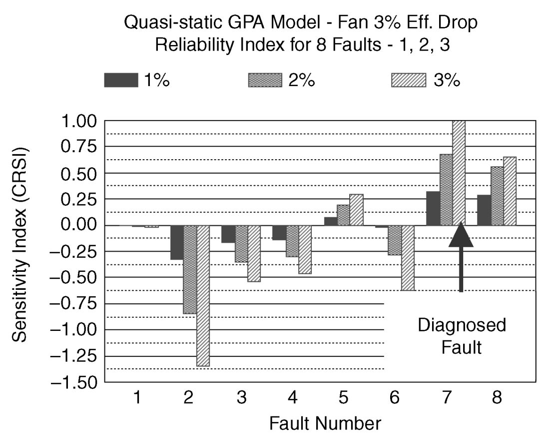

FIGURE 41.14 Reliability sensitivity index using the OQS GPA model.

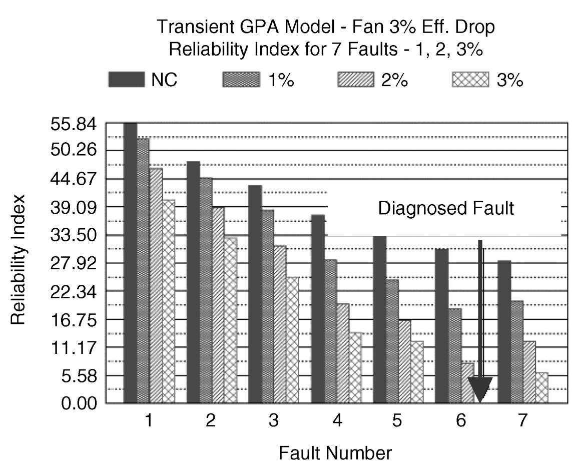

FIGURE 41.15 Computed reliability index using the OT model.

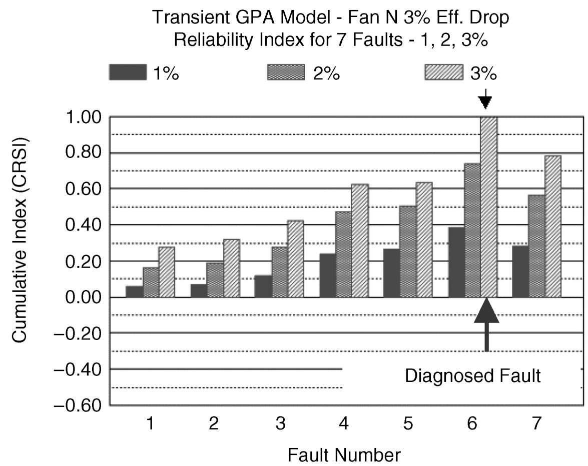

FIGURE 41.16 Reliability sensitivity index using the OT model.

issue when an incipient fan fault exists, since this can project unreal, "phantom" faults in the other compartments. This aspect can be fixed by using the OT model in conjunction with the PT model.

One important aspect that can reflect on the accuracy of a fault diagnostic–prognostic procedure is the fact that statistics of different flight profiles are quite different, and they are also different than the statistics of the overall ensemble of all flight-profile types (Table 41.1). In statistical terms, this indicates that the ensemble of flight profiles (including all profile types) is not ergodic. Thus, the fault diagnostics–prognostics have to be approached by operational spectrum-type statistical modeling.

41.2.7 Decision-Level Fusion

The decision-level fusion presented here is based on diagnostic–prognostic reasoning from the individual models. The benefit of this approach is that it maintains separation of the models, thus providing a plug-and-play scenario, where one, two, or all three modeling approaches could be used to assess the engine's health, depending on the engine's requirements. The decision-level fusion uses the diagnostic outputs from the separate engine models, (OQS, OT, PT) and combines them using a heuristic scoring approach to enhance the confidence of the final output. The purpose of the fusion procedure is to take advantage of the synergies provided by using the multiple engine models. For example, the PT model is good for detecting multiple faults, although it is limited in its ability to discriminate between certain turbine faults; the OT model is good for detecting and discriminating between compressor faults; and the OQS model is good for detecting turbine faults.

The reliability-based indices from three separate engine models are weighted, based on the fault-detection and fault-discrimination sensitivity indices of the engine models for given fault conditions, to provide a joint reliability index (RI). The scoring method used employs the sensitivity and normalized sensitivity indices for the three models. This applies weighting to the model outputs based on their fault-detection and fault-discrimination capabilities. The scoring model implemented computes the sum

$$S(RI_i) = \frac{1}{C_i} \sum_{j=1}^{3} RI_{ij}\left(\frac{3}{4} SI_{ij} + \frac{1}{4}\mathrm{NSI}_{ij}\right) \tag{41.17}$$

where C_i is the normalization constant.

An example based on the single-fault implementation of the fault-basin approach is presented in Figure 41.17 through Figure 41.20 to illustrate the benefits of decision-level fusion in the diagnostic process. In this case, the OQS model performs relatively consistently across the faults considered. However, the reliability indices are relatively low with the exception of HPT efficiency and LPT capacity faults.

TABLE 41.1 Mean and Standard Deviation of Aircraft Engine Parameters

Table of Mean Values

Profile Type	P2	P3	P4	P5	P6	T2	T3	T4	T5	T6
All Types	0.000	0.000	−0.032	0.024	0.024	−0.048	0.720	−0.162	0.000	−0.141
Type A	−0.003	0.046	−0.009	0.012	0.012	−0.038	0.299	−1.217	−0.955	−0.672
Type B	0.002	−0.017	−0.015	0.010	0.010	−0.054	0.604	−1.873	−1.494	−1.484
Type C	−0.001	−0.054	−0.050	0.015	0.015	−0.072	1.100	−0.283	−0.126	−0.159
Type D	0.005	0.091	0.008	0.046	0.046	0.035	0.660	−0.131	0.053	0.700
Type E	0.005	−0.522	−0.057	0.074	0.074	−0.082	0.607	0.783	0.940	0.792
Type F	0.005	0.095	0.044	0.042	0.042	−0.043	0.975	−0.189	0.025	−0.605
Type G	0.005	−0.181	−0.227	0.007	0.007	−0.145	0.847	2.561	2.271	1.218
Type H	0.000	0.019	−0.016	0.031	0.031	−0.026	0.795	0.189	0.299	−0.113
Type 1	−0.009	−0.024	−0.017	0.005	0.005	−0.049	0.566	−0.369	−0.153	−0.299
Maximum	0.005	0.095	0.044	0.074	0.074	0.035	1.100	2.561	2.271	1.218
Minimum	−0.009	−0.522	−0.227	0.005	0.005	−0.145	0.299	−1.873	−1.494	−1.484
Mean	0.001	−0.061	−0.038	0.027	0.027	−0.053	0.717	−0.059	0.096	−0.069
STD	0.005	0.192	0.077	0.023	0.023	0.048	0.240	1.247	1.074	0.844

Table of Standard Deviations

Profile Type	P2	P3	P4	P5	P6	T2	T3	T4	T5	T6
All Types	0.054	0.406	0.369	0.129	0.129	0.468	1.481	6.367	5.290	4.854
Type A	0.060	0.237	0.259	0.099	0.099	0.473	0.931	7.198	5.874	4.885
Type B	0.041	0.260	0.242	0.083	0.083	0.354	1.114	6.181	5.032	4.995
Type C	0.047	0.312	0.289	0.104	0.104	0.425	1.640	6.072	5.033	4.474
Type D	0.066	0.383	0.336	0.132	0.132	0.507	1.531	5.879	4.768	4.837
Type E	0.047	0.298	0.326	0.090	0.090	0.776	1.659	7.775	6.747	5.251
Type F	0.046	0.329	0.292	0.103	0.103	0.353	1.591	5.075	4.501	4.362
Type G	0.063	0.703	0.627	0.194	0.194	0.624	1.642	6.855	5.649	5.515
Type H	0.055	0.408	0.395	0.140	0.140	0.418	1.525	5.798	4.864	4.411
Type I	0.054	0.450	0.388	0.155	0.155	0.379	1.496	6.307	5.232	4.869
Maximum	0.066	0.703	0.627	0.194	0.194	0.776	1.659	7.775	6.747	5.515
Minimum	0.041	0.237	0.242	0.083	0.083	0.353	0.931	5.075	4.501	4.362
Mean	0.053	0.376	0.351	0.122	0.122	0.479	1.459	6.349	5.300	4.844
STD	0.009	0.141	0.116	0.036	0.036	0.141	0.258	0.811	0.690	0.387

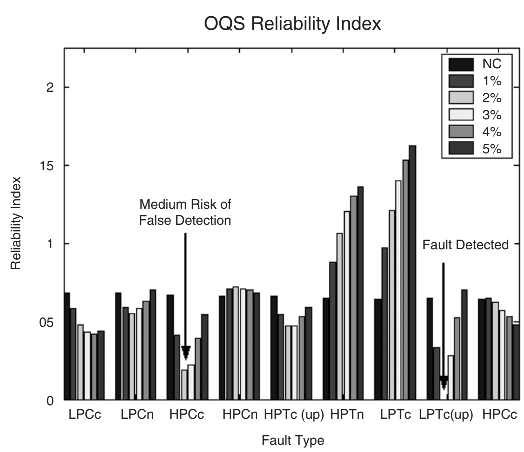

FIGURE 41.17 OQS reliability indices LPT 2% capacity increase.

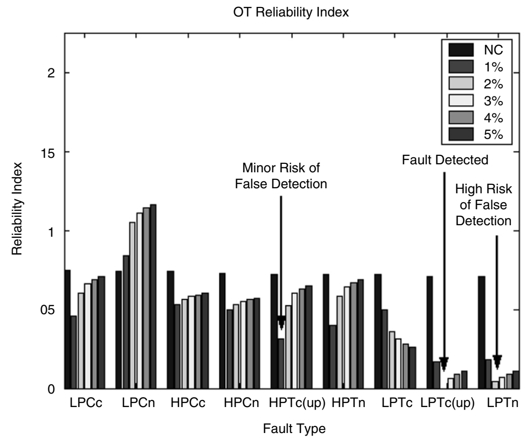

FIGURE 41.18 OT reliability indices LPT 2% capacity increase.

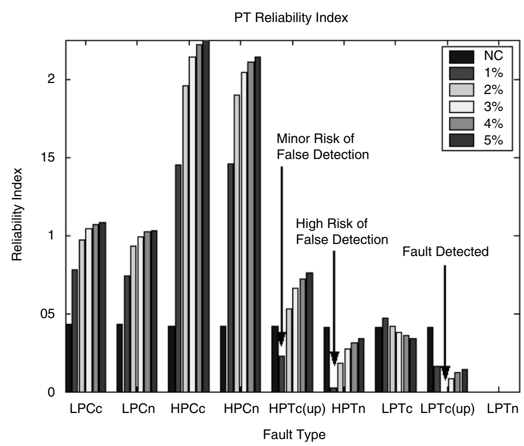

FIGURE 41.19 PT reliability indices LPT 2% capacity increase.

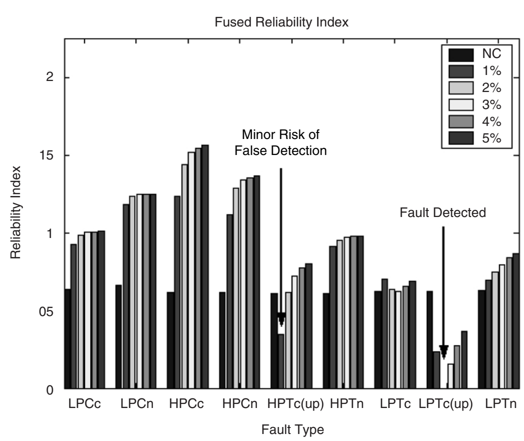

FIGURE 41.20 Fused reliability indices LPT 2% capacity increase.

TABLE 41.2 Benefits of Feature Vector Extraction (ANFIS)

Category of Benefit	General Benefit	Operational Advantages
Feature extraction	Converts data into useable features that describe the state of the engine	Provides deviations from normal operating conditions
Robust operational performance	In the event of a reduced sensor array (one or more nonoperational sensors), the engine models still extract features. Loss of output sensor results in reduced dimensionality of feature space. Loss of input sensor results in a reduction of model performance	Allows continued operation despite reduced sensor array (input or output)
Engine–engine variability	Models can be readily tuned to fit a specific engine, based on initial generic ANFIS models for a given engine class	More accurate model of a specific engine enables assessment of engine performance relative to fleet
In-flight capabilities	Transient engine models extend the capabilities of air VEHM systems to be a true flight-capable system. Quasi-stationary models are improved about 40% through the inclusion of PLA	More up-to-date information regarding engine condition means less reliance on ground-based testing, leading to a reduction in maintenance costs
Improved detection	Increased sensitivity index of engine models results in earlier and more confident fault detection	Earlier detection of fault onset
Improved diagnosis	Increased fault discrimination of the engine models results in improved diagnostic output	Accuracy of fault diagnostics enables maintenance and parts to be planned prior to servicing

The OT model provides additional evidence against a Fan/LPC fault; however, it is unable to clearly isolate the severity of the fault, and there is potential for false detection of an HPT capacity or an LPT efficiency fault. The PT model clearly eliminates the fan/compressor sections as a source of the fault. However, it has difficulties discriminating between the source of the turbine fault and its severity. In this case, the fused reliability indices provide clear advantages, with the source of the fault identified as resulting from a 2% increase in the LPT capacity. There is still a small probability of false detection of a 1% HPT capacity increase. Nevertheless, the fused reliability indices perform significantly better than the individual models, both at fault diagnosis and determination of the fault severity.

Embedded into StoFIS are three component technologies. The first is the feature vector extraction (data-level fusion), where data from multiple sensors are fused into the nondeterministic ANFIS GPA engine models prior to feature extraction. The second is multivariate stochastic analysis of the extracted feature vectors (feature-level fusion), where parameter deviations from the ANFIS models are used to characterize the system and arrive at a diagnostic–prognostic output. The third is decision-level fusion, where diagnostic outputs from separate nondeterministic ANFIS GPA engine models are fused after diagnostic reasoning to enhance the confidence of the final output. The benefits of the three component technologies of StoFIS are described in Table 41.2 through Table 41.4.

41.3 Application 2: DWPA Multiple-Band-Pass Demodulation and Automated Diagnostics

Rolling-element bearings are the most common cause of small machinery failure, and overall vibration-level changes are virtually undetectable in the early stages of deterioration. However, due to the characteristics of rolling-element bearing faults, vibration analysis techniques have proven to be an effective tool for the detection and diagnosis of incipient faults. The demodulated spectrum is the most common technique used for the detection of localized bearing faults. However, for low-speed rolling-element bearings, demodulation can be unreliable for the detection and diagnosis of faults [6]. Difficulties include spectral smearing due to speed fluctuations and skidding of rolling elements, poor performance under

TABLE 41.3 Benefits of Multivariate Stochastic Analysis (Fault Basin)

Category of Benefit	General Benefit	Operational Advantages
Robust operational performance	Multidimensional mapping of parameter deviations is robust in the event of the loss of one or more parameters	Allows continued operation despite reduced sensor array (input or output)
More-refined feature mapping	Extraction of stochastic features provides more-refined diagnostics and prognostics (e.g., two faults may have similar mean deviations, but their ellipsoid clusters may have significantly different correlation structure and thus orientation in space)	Earlier detection of anomalies and improved fault discrimination
Incorporates data uncertainty	Transformation into standard Gaussian space enables engine reliability to be assessed in terms of risk	Enables calculation of the risks of potential engine faults
Fault basins of attraction	The reliability index is computed with respect to a continuous fault path from low-severity to high-severity levels and not with respect to a particular fault location	Improved robustness and more accurate indicator of fault severity; more-refined and accurate diagnostics/prognostics
Reliability-based reasoning	Diagnostic–prognostic output is provided in terms of reliability-based indices	Operational risks can be assessed, thus assisting in operational status and maintenance decisions for aircraft
Multiple fault detection	Scanning of multiple fault deviations in multidimensional parameter space provides the ability to assess the risk of multiple faults	Improved robustness for multiple faults
Prognostic output	Projected usage trajectories in multidimensional space are used to provide reliability-based prognostic output of fault degradation	Provides knowledge of the speed of progression for a given fault and assists with maintenance planning

high levels of noise, and difficulties in identifying and extracting the regions of bearing resonance. This section presents a nondeterministic method to surmount these problems by combining several techniques, including time-frequency decomposition, autoregressive (AR) stochastic process spectral analysis, and nondeterministic ANFIS.

TABLE 41.4 Benefits of Decision-Level Fusion

Category of Benefit	General Benefit	Operational Advantages
Robust operational performance	Model redundancy allows for the situation where one model is unable to adequately detect and diagnose a fault In the event of a reduced sensor array (one or more nonoperational sensors), the most robust model can provide information while others are not performing adequately. Adaptive weighting can be used to reflect the relative performance of each of the models under adverse conditions or reduced sensor arrays	Facilitates the best possible diagnostic accuracy for each fault class Enables optimum operation for a given array of sensors
Increased fault-detection sensitivity	Sensitivity of fault detection is limited by the most sensitive model for a given fault condition. Depending on the fault location and type, this may be the OQS, OT, or PT model	Earlier detection of performance degradation
Improved system reliability	Two or three models can confirm the same engine fault condition	Reduced probability of false detection
Reduced ambiguity	Reduced probability of uncertainty in fault diagnosis, as each of the models provides different levels of fault discrimination from other fault sources	Increased plausibility in the fault type diagnosed and corresponding severity level

The extraction of attenuated resonant vibrations due to impacts from localized faults in rolling-element bearings is normally achieved by high- or band-pass filtering of the vibration signal. The main problem with this approach is the difficulty in choosing an appropriate filter range of interest. This section presents an alternative to traditional approaches, which enables the automation of the selection and the inclusion of multiple frequency bands of interest.

The method for the extraction of high-frequency transients due to bearing impact resonance is achieved at an optimal time-frequency resolution via best-basis discrete wavelet packet analysis (DWPA) representation, using the Daubechies-20 wavelet [7]. Selection of the frequency band or bands of interest is achieved by analyzing the characteristics of each of the wavelet packets. The selection process is automated through the use of an ANFIS model, thus removing the need for the analyst to manually identify the bands of interest.

The best-basis DWPA provides an optimal time-frequency decomposition of the signal and facilitates the extraction and reconstruction of wavelet packets containing bearing fault-related information. For a signal component composed of wide-band transients, high time resolution and low frequency resolution would be required. On the other hand, a slowly varying narrow-band component of a signal requires better frequency resolution, with its time resolution being less important. By obtaining the best-basis DWPA of the signal prior to extraction of the wavelet packets of interest, two important objectives are accomplished. The most important of these is the improved time resolution of the bearing transients while maintaining isolation from other signal components. This improves the ability to resolve low-amplitude transient features over the noise floor level. Second, subsequent processing required to extract the relevant wavelet packets is substantially reduced. A detailed explanation of optimal time-frequency decompositions as well as the choice mother wavelet can be found in the literature [7–9].

DWPA multiple-band-pass filtering surmounts the problem of extracting regions of bearing resonance that are intertwined with continuous signals. Figure 41.21 illustrates how this method facilitates the extraction of bearing-fault-related components from a signal while rejecting the unwanted harmonics. The wavelet packets identified by the ANFIS model as containing bearing-fault-related features are indicated. To visualize the rejection of wavelet packets containing unwanted continuous signal components, the power spectral density is plotted along the vertical axis of the DWPA representation. Wavelet packets that contained the harmonic peaks present in the power spectral density plot were rejected by the adaptive network-based fuzzy inference system as containing excessive levels of signal contamination.

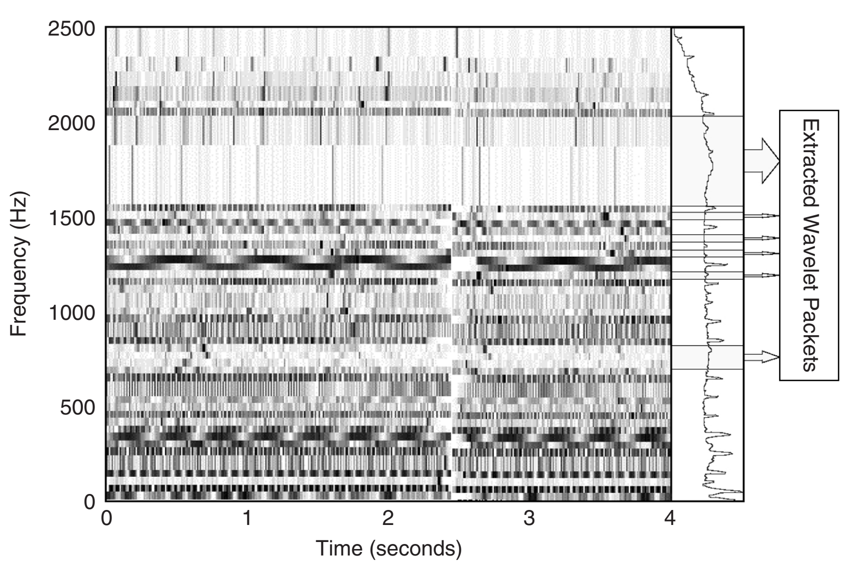

FIGURE 41.21 Selection and extraction of wavelet packets containing bearing-failure-related features. The extracted wavelet packets are (3, 7), (4, 10), (6, 17), (6, 26), (6, 27), (6, 30), (6, 49), (6, 50), and (6, 53).

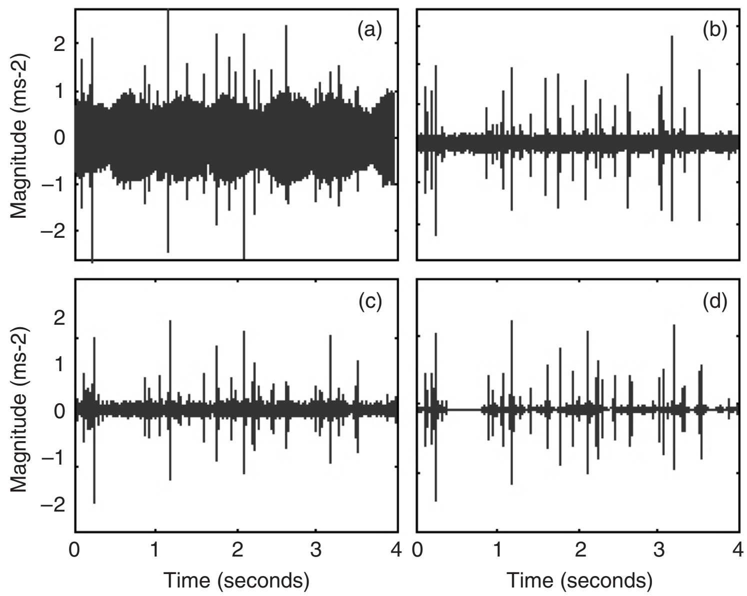

FIGURE 41.22 (a) High-pass filtered signal (order 40, [F > 500 Hz]); (b) FIR band-pass filter (order 40, [1500 Hz < $F_{\text{band-pass}}$ < 2500 Hz]); (c) reconstruction of extracted wavelet packets; (d) hard-threshold denoised reconstruction of extracted wavelet packets.

This clearly demonstrates the ability of DWPA multiple-band-pass filtering to extract only the wavelet packets composed predominantly of bearing-fault-related vibrations.

The ability of the multiple-band-pass technique to select more than a single band of interest enables sections of the signal that contain predominantly noise or contaminating sources of vibration to be excluded. This results in the extraction of a cleaner bearing-fault signal. Figure 41.22 shows a comparison between high-pass, manually optimized band-pass filters, and DWPA multiple-band-pass filtering (with and without noise reduction). Visual examination of the filtered signals indicates a significant decrease in the contaminating effects of noise and other sources of vibration when the DWPA multiple-band-pass filtering is applied. The DWPA reconstructed signal has a marginally lower level of sinusoidal contamination than the best possible band-pass filter, and the bearing-fault-related transients are also stronger. Hard threshold denoising almost eliminated the remaining polluting sources of vibration. This further enhances the ability of DWPA multiple-band-pass enveloped spectra to accurately diagnose the location and magnitude of bearing defects.

41.3.1 Implementation of ANFIS for Automatic Feature Extraction

In order to design and develop a robust and reliable nondeterministic identification of wavelet packets of interest, an ANFIS was used to provide a statistical best-estimate based on the input parameters of the model. The parameters chosen must enable the neuro-fuzzy network to make intelligent decisions regarding the extraction of wavelet packets containing bearing-fault-related information. The input parameters that were chosen for this process were kurtosis and the spectrum peak ratio (SPR).

Kurtosis is an effective measure of the spikiness of a signal. A high kurtosis level indicates that the wavelet packet is impulsive in nature, as would be expected from a wavelet packet that contains bearing-fault-related features. Kurtosis is defined as:

$$\text{Kurtosis} = \frac{1}{NS_y^4}\sum_{i=1}^{N}(y(i)-\bar{y})^4 \tag{41.18}$$

S_y is the standard deviation, and $\bar{y}$ is the mean of data sample y. Kurtosis was chosen over other measures of spikiness (crest factor, impulse factor, and shape factor) due to its statistically robust nature.

The spectrum peak ratio is defined as the sum of the peak values of the defect frequency and its harmonics, divided by the average of the spectrum [10]. Shiroishi [10] used the spectrum peak ratio as a trending parameter to indicate the presence of localized bearing defects, which was found to be more robust than considering just the defect frequency.

$$\mathrm{SPR} = \frac{N \times \sum_{h=1}^{n} P_h}{\sum_{k=1}^{N} A_i} \tag{41.19}$$

P_h is the amplitude of the peak located at the defect frequency harmonic; A_i is the amplitude at any frequency; and N is the number of points in the spectrum. In order to differentiate between wavelet packets belonging to different classes of bearing faults, three autoregressive-based peak ratios are employed: spectrum peak ratio inner (SPRI), spectrum peak ratio outer (SPRO), and spectrum peak ratio rolling-element (SPRR). Calculation of the spectrum peak ratios was based on Yule-Walker autoregressive spectral estimates of the reconstructed wavelet packets using a model order of 125, equivalent to one shaft revolution. Autoregressive spectral analysis was used in preference to the FFT (fast Fourier transform), as this method has been shown to reduce the effect of spectral smearing and skidding for low-speed rolling-element bearings [11].

Seeded faults in a low-speed test rig and mathematical models of bearings containing localized faults [12] were used to construct a database of 2810 wavelet packets. These wavelet packets were individually assessed as to whether they contained bearing-fault-related features by visual examination of their time series and the envelope AR spectrum. They were then categorized for each fault class as containing fault-related features (1), probably containing fault-related features (0.66), probably not containing fault-related features (0.33), or not containing fault-related features (0). The wavelet packet data set included 444 containing inner-race fault-defect information, 221 containing rolling-element fault information, and 162 containing outer-race fault information. The wavelet packets were split into three data sets: a training data set of 1000 wavelet packets, a checking data set of 1000 wavelet packets, and a testing data set of 810 wavelet packets.

Given the training and checking input/output data sets, the membership function parameters were adjusted using a back-propagation algorithm in combination with a least-squares method. The checking data were used to cross-validate and test the generalization capability of the fuzzy inference system. This was achieved by testing how well the checking data fits the fuzzy inference system at each epoch of training, and the final membership functions were associated with the training epoch that has a minimum checking error. This was an important task, as it ensured that the tendency for the fuzzy inference system to overfit the training data, especially for a large number of epochs, was avoided.

Two different membership function structures were compared in this study. The first consisted of the kurtosis and spectrum peak ratio input amplitudes being transformed by two membership functions (small and large), and the second was split into three membership functions (small, medium, and large). The parameters for the initial input membership functions were determined by expert knowledge of the system being modeled. Figure 41.23 illustrates a three-membership function structure.

41.3.2 Performance of ANFIS Feature Extraction

A substantial reduction of the sum of squared errors for the testing data was apparent after training of the neuro-fuzzy systems. This reduction is quantified in Table 41.5, with an average decrease in the sum of squared errors of 34.9% for the neuro-fuzzy systems. The errors in the testing data are defined as the sum of the squared differences between the output of the fuzzy inference system and the fault-categorization

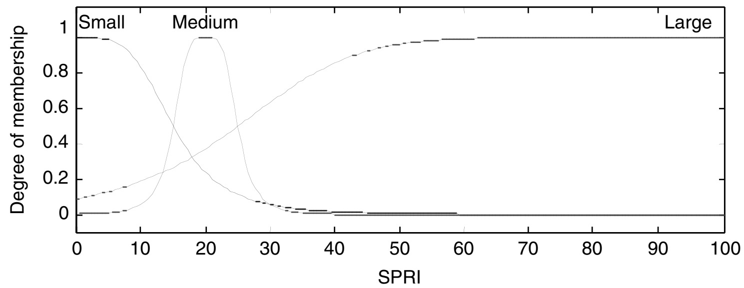

FIGURE 41.23 SPRI membership function (small, medium, and large).

scheme based on visual inspection. In order to classify the individual wavelet packets as either containing or not containing bearing-fault-related features, it is necessary to define a minimum crisp output value that would indicate the presence of high-frequency bearing transients. The minimum crisp output value was set as 0.5 for the correct classification rates presented in Table 41.6. The 3/3 neuro-fuzzy system had a marginally better correct classification rate than the 2/2 neuro-fuzzy system.

41.3.3 Automated Fault Classification

A multiple-band fusion technique was implemented to provide an automated diagnostic tool based on the DWPA feature-extraction process. This is similar in concept to the multisensor fusion used by Loskiewicz and Uhrig [13], where multiple sensor data were fused in order to increase the confidence factor for the final diagnosis compared with single-sensor diagnosis. The difference in this case is that the data-fusion concerns multiple bands of interest for a single sensor, with the intention of enhancing the confidence of correct diagnosis based on single-sensor vibration analysis.

The wavelet packet selection process using the nondeterministic ANFIS identified the existence and fault class of bearing transients for individual wavelet packets. The output for each wavelet packet can be viewed as the confidence factor relating to the existence of a particular bearing-fault class for the wavelet packet being examined. Through fusion of the confidence factors for each wavelet packet that indicated the presence of bearing-fault-related transients, a more confident assessment of the fault classification can be ascertained, with the signals requiring further manual analysis for definitive classification easily identified. Manual classification can be achieved by visual examination of the demodulated spectrum of the multiple-band-pass-filtered signal [14]. The use of multiple-band fusion of confidence factors for fault classification reduces both the probabilities of missing or of incorrectly classifying a stochastic faulty bearing.

Two parameters were used in the proposed fault-classification scheme: the average confidence level of fault classification and the number of wavelet packets that contained bearing-fault-related features. The average confidence level (ACL) was based on the average output from the nondeterministic ANFIS models

TABLE 41.5 Testing Sum of the Squared Errors for the Initial and Final Fuzzy Systems

	2/2 Fuzzy		3/3 Fuzzy	
	Initial	Trained	Initial	Trained
IRF	0.1239	0.0806	0.1383	0.0829
REF	0.1380	0.0926	0.1639	0.0958
ORF	0.0501	0.0390	0.0591	0.0380

IRF-inner race fault; REF-rolling-element fault; ORF-outer race fault.

TABLE 41.6 Correct Classification Rate (%) for the Neuro-Fuzzy Systems

	2/2 Fuzzy	3/3 Fuzzy
Fault	97.98%	98.85%
Probably fault	83.53%	82.18%
Probably no fault	90.45%	87.52%
No fault	99.96%	99.96%

for the wavelet packets that were identified as containing bearing-fault-related transients, and it was calculated for each fault class (inner-race fault, rolling-element fault, and outer-race fault).

$$\mathrm{ACL} = \frac{\sum_{i=1}^{n} \text{ANFIS Output}}{n} \tag{41.20}$$

where n is the total number of wavelet packets that contain bearing-fault-related features. Table 41.7 shows typical outputs obtained from this process.

Box plots of the outputs from the fusion of the multiple-band-pass fuzzy confidence factors are presented in Figure 41.24 for each of the fault classes. Box plots give you an idea of the distribution of data, especially in terms of symmetry and scale. The limits of the box correspond to the first and the third quartiles (Q1 and Q3), and the fences, respectively, to Q1 – 1.5(Q3-Q1) and Q3 + 1.5(Q3-Q1). The minimum and maximum data points are printed with a black circle. The distinction between each of the fault classes is evident. In the case of the combined inner-race/rolling-element fault class, the average confidence level was lower for both fault classes than for an individual fault. There were two factors at play that result in the reduced confidence levels. First, there was the interference due to the presence of bearing transients from both fault classes in the extracted wavelet packets. This reduced the spectral peak ratio for each fault class, and may have in turn reduced the fuzzy confidence factor indicating the presence of a fault for the individual wavelet packets. The second factor was a consequence of variation in the frequency and relative amplitude of the bearing impact transients, which was found to have a degree of dependence on the impact location. Due to this, certain bands contained bearing transients of significantly reduced amplitude from one of the fault classes, and in some cases contained transients from only one fault class. Although this resulted in a reduction of the average confidence level after fusion of the outputs from the ANFIS model, the existence of multiple faults was clearly indicated in each case.

Fault classification through fusion of multiple-band-pass fuzzy confidence factors resulted in a successful classification rate of 100% for the signals examined, with 95% requiring no additional analysis for verification of the fault classification. The vibration signals where the average confidence level was considered inadequate for a positive diagnosis to be made where limited to the condition of combined inner-race/rolling-element faults.

TABLE 41.7 Typical Outputs Obtained from Fusion of Multiple-Band-Pass Fuzzy Confidence Factors, and the Number of Wavelet Packets on Which the Calculations Were Based

Fault Type	Number of Wavelet Packets	ACL (IRF)	ACL (REF)	ACL (ORF)
IRF	6	98%	14%	0%
ORF	7	0%	0%	99%
REF	7	0%	93%	0%
IRF/REF	8	65%	85%	0%

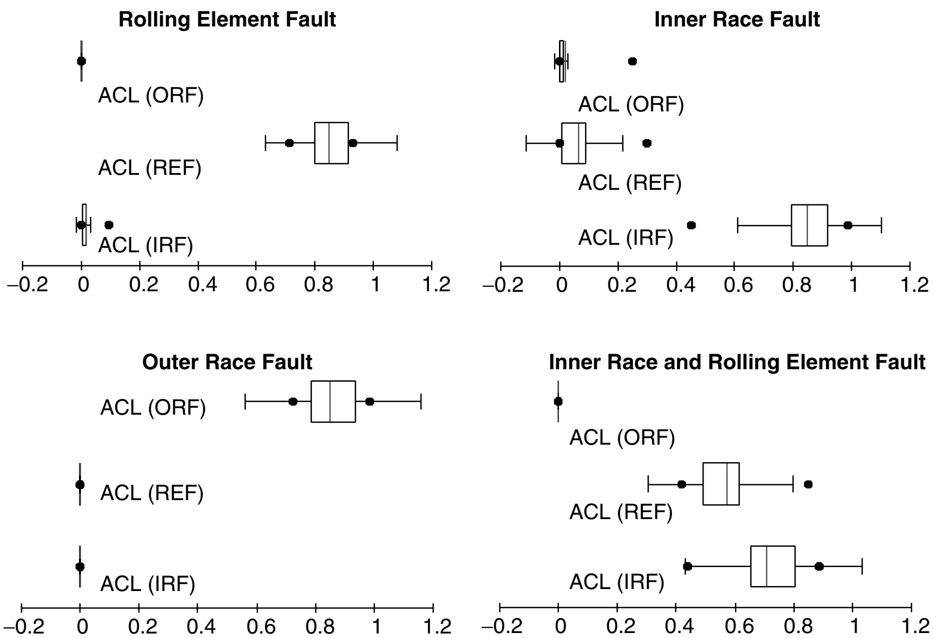

FIGURE 41.24 Box plots of the outputs obtained from the fusion of multiple-band-pass fuzzy confidence factors for each fault class considered.

41.3.4 Multiple-Band-Pass Fault-Severity Index

Obtaining an accurate determination and trending of fault severity is an integral part of a successful predictive maintenance program. Trending parameters can be used to indicate the general health of a machine, or that of a specific machine element. In order for trending parameters to be effectual, they must be sensitive enough to pick up changes in the condition of machine elements, yet not so sensitive that small variations in operating conditions trigger an alarm. A bearing-specific trending parameter, the wavelet peak index (WPI), based on the multiple-band-pass DWPA feature-extraction technique, is presented for this purpose.

The wavelet peak index is defined as the peak level of the combined reconstructed wavelet packets that contain bearing-fault-related features. The time-domain-based peak level was favored over trending of spectral peaks, as this allowed the modulating effects of load to be ignored, thus allowing direct comparison between the severity of different types of faults (inner-race, outer-race, and rolling-element faults). To test the performance of this trending parameter, a series of tests involving the introduction of an artificial crack on the inner race of a low-speed cylindrical rolling-element bearing were performed, simulating wear-out by deepening and widening the groove in several stages. The tests were conducted at 20, 60, and 120 rpm. The proposed trending parameter was then compared with a number of commonly used trending parameters for bearing-health determination, peak level, RMS (root mean square), crest factor, and kurtosis.

Figure 41.25 displays the results of the trending parameters tested for each of the series of low-speed artificial wear-out tests. For each series of tests, the wavelet peak index provided a clear and sensitive trend of the deteriorating condition of the bearing as the fault width was increased. This was true at all operating speeds, even for the smallest of the fault widths considered (0.38 mm). At operating speeds of 20 and 60 rpm, no discernible increase in either peak level or crest factor was noted for fault widths below 0.67 mm, and RMS provided no indication of the deteriorating bearing condition for any of the fault widths considered. The wavelet peak index is indisputably more sensitive to bearing transients of low amplitudes than either the peak level or crest factor, with the wavelet peak index more closely coupled

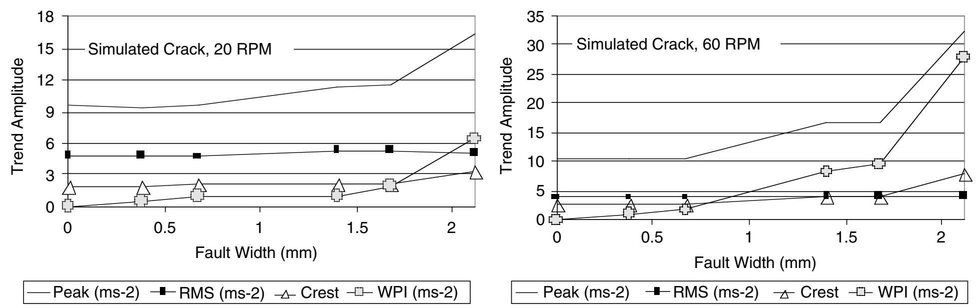

FIGURE 41.25 Trend parameters vs. crack size.

with the width of the bearing fault. Unlike the peak level, which is an overall indicator including vibration components unrelated to the bearing fault, the index WPI is specific to bearing-related faults, as it is the peak of the multiple-band-pass-filtered signal.

41.4 Concluding Remarks

Two nondeterministic hybrid architectures have been demonstrated to provide enhanced vehicle health-management reasoning for the detection, diagnosis, and prognosis of faults. The nondeterministic hybrid approaches have enabled more robust and comprehensive solutions to the challenges of developing vehicle health-management systems capable of predicting faults with associated confidence or risk levels. The utilization of advanced signal processing and multiple levels of data fusion to maximize the extracted information used for diagnostic-prognostic reasoning have also been illustrated.

References

1. Ghiocel, D.M. and Altmann, J., A Hybrid Stochastic-Neuro-Fuzzy Model-Based System for In-Flight Gas Turbine Engine Diagnostics, presented at 55th Machine Failure Prevention Technology Meeting, Society for Machinery Prevention Technology, Virginia Beach, VA, 2001.
2. Ghiocel, D.M., A new perspective on health management using stochastic fault diagnostic and prognostic models, *Int. J. Adv. Manuf. Syst.,* 4 (1), PAGE, 2001.
3. Jang, J.-S.R., ANFIS: adaptive network-based fuzzy inference systems, *IEEE Trans. Syst., Man, Cybernetics,* 23 (3), 665–685, 1993.
4. Jang J.-S.R. and Sun C.-T., Neuro-fuzzy modelling and control, *Proc. IEEE*, 83 (3), 378–406, 1995.
5. Ghiocel, D.M. and Altmann, J. Conceptual and Tool Advances in Machinery Preventive Diagnostics, presented at 56th Machine Failure Prevention Technology Meeting, Society for Machinery Failure Prevention Technology, Virginia Beach, VA, 2002.
6. Mathew, J. et al., Incipient damage detection in low speed bearings using the demodulated resonance technique, in *Proceedings of International Tribology Conference*, Monash University, Melbourne, Australia, 1987, pp. 366–369.
7. Daubechies, I., The wavelet transform, time-frequency localisation and signal analysis, *IEEE Trans. Inf. Theory*, 36 (5), 961–1005, 1990.
8. Misiti, M. et al., *Wavelet Toolbox MATLAB*, The MathWorks, Natick, MA, 1996.
9. Altmann, J. and Mathew J., Optimal configuration of the time-frequency representation of vibration signals, *Machine Condition Monitoring Res. Bull.*, 8, 13–24, 1996.
10. Shiroishi, J. et al., Bearing condition monitoring via vibration and acoustic emission measurements, *Mech. Syst. Signal Process.,* 11 (5), 693–705, 1997.
11. Mechefske, C., Fault detection and diagnosis in low-speed rolling-element bearings; 1: the use of parametric spectra, *Mech. Syst. Signal Process.*, 6, 297–307, 1992.

12. Altmann J. and Mathew J., Analytical modelling of vibrations due to localised defects in rolling-element bearings, in *Proceedings of COMADEM '98*, Monash University, Launceston, Australia, 1998, pp. 31–40.
13. Loskiewicz, A. and Uhrig, E., Decision fusion by fuzzy-set operation, in *Proceedings of IEEE World Congress on Computational Intelligence, Fuzzy Systems Conference*, IEEE, Orlando, FL, 1994, pp. 1412–1417.
14. Altmann J. and Mathew J., Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis, *Mech. Syst. Signal Process.*, 15 (5), 963–977, 2001.