Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Control Engineering Practice 12 (2004) 1151–1165
CONTROL ENGINEERING PRACTICE
www.elsevier.com/locate/conengprac

Fault diagnosis for a turbine engine[☆]

Yixin Diao[a,*], Kevin M. Passino[b]

[a] IBM Research, IBM T.J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, USA
[b] Department of Electrical Engineering, Ohio State University, 2015 Neil Avenue, Columbus, OH 43210, USA

Received 16 February 2000; accepted 28 November 2003

Abstract

Fault detection and diagnosis for jet engines is complicated by the presence of engine-to-engine manufacturing differences and engine deterioration during normal operation, the complexity of an accurate engine model, and our inability to directly measure certain engine variables. Here, we work with a sophisticated component level model (CLM) simulation of a turbine engine (the General Electric XTE46) that can simulate the effects of manufacturing and deterioration differences, in addition to a variety of faults. To develop a fault diagnosis system we begin by using the CLM to generate data that is used by a Levenberg–Marquardt method to train a Takagi–Sugeno fuzzy system to represent the engine. The resulting nonlinear model provides a reasonably accurate representation of manufacturing differences, engine deterioration, and fault effects. We use multiple copies of this nonlinear model, each representing a different fault, to generate error residuals by comparing them to the engine output. In fact, we manage the composition of the set of models with a "supervisor" that ensures that the appropriate models are on-line, and processes the error residuals to detect and identify faults. Robustness and fault sensitivity of the proposed approach are studied in the paper and the component model level simulation of the XTE46 engine is used to illustrate the effectiveness of the fault diagnosis scheme.

Keywords: Fault diagnosis; Fuzzy systems; Multiple models; Robustness analysis; Jet engines

1. Introduction

The last two decades saw continuous improvement in control techniques resulting from the spectacular progresses in control theory and computer technology. Meanwhile, stimulated by the growing demand for improving the reliability and performance of systems, many fault diagnosis methods, in particular, model-based analytical redundancy methods, have been developed, which have the capability of detecting the occurrence of faults and determining the types (or locations) of the faults. Survey papers by Gertler (1988), Frank (1990) and Patton (1997) present excellent overviews of advances in model-based fault diagnosis methods. Due to the complexity and importance of jet engines, the study of fault diagnosis for jet engines has become a very active research topic for both theoretical and practical reasons. In Merrill (1985), Merrill, DeLaat, and Bruton (1988), and Merrill, DeLaat, and Abdelwahab (1991), the authors studied sensor failure detection for jet engines using a Kalman filter with a generalized likelihood ratio testing-based scheme. In Duyar and Merrill (1992) and Duyar, Eldem, Merrill, and Guo (1994) the authors derived linearized models of jet engine systems via the α-canonical form parameterization identification method and applied a parameter estimation approach in fault detection and isolation (FDI) for the space shuttle main engine. In Patton and Chen (1992, 1997) and Patton, Chen, and Zhang (1997), the authors studied fault detection of jet engine sensor systems using an eigenstructure assignment technique to design observer-based residual generators, and they also studied its robustness. The model based FDI techniques have also been applied to many other complex applications. In Menke and Maybeck (1995), the authors

[☆] This work was supported by the NASA Glenn Research Center under grant NAG3-2084. The authors would like to thank Dr. S. Adibhatla at General Electric Aircraft Engines for providing the CLM for the engine, technical advice on how to use it, and feedback on the development of the modeling and control strategies. (Also, this work was done while the first author was at the Dept. of Electrical Engineering, The ohio state University, 2015 Niel Ave., Columbus, OH 43210, USA)

*Corresponding author. Tel.: +1-914-784-7226; fax: +1-914-784-6040.

E-mail addresses: diao@us.ibm.com (Y. Diao), passino@ee.eng.ohio-state.edu (K. M. Passino).

0967-0661/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2003.11.012

detected sensor and actuator faults of flight control systems using multiple model adaptive estimation. The multiple model fault diagnosis approach were also studied in Yen and Ho (2001); Fu, Shen, Zhang, and Liu (2001); Patton, Toribio, and Simani (2001); Boskovic, Li, and Mehra (2001); and Tu, Pattipati, Deb, and Malepati (2003). In Gertler, Costin, Fang, Hira, Kowalczuk, and Luo (1993), the authors used orthogonal parity equations for automotive engines. In Tylee (1983), the authors detected instrument failures in nuclear power plant instrumentation using Kalman filters, and in Wang, Kropholler, and Daley (1993), the authors applied robust observer-based methods to a distillation column. In the application of FDI, process disturbances, sensor noise, and model inaccuracy are inevitable. Since residuals are not only affected by faults but also by these factors, robustness is a very important problem in model-based FDI (Chen & Patton, 1999). To reduce the sensitivity of the residuals to modeling errors, many methods have been developed by decoupling the effects of faults from those of modeling error, which are based on unknown input observers (Seliger & Frank, 1991; Chen & Zhang, 1991; Hou & Muller, 1994), eigenstructure assignment (Patton & Chen, 1991, 1997a), or orthogonal parity relations (Gertler & Luo, 1989; Gertler & Singer, 1990). Another class of robust FDI method is achieved by making the threshold adaptive to the input (Emani-Naeini, Athter, & Rock, 1988; Frank & Ding, 1994, 1997).

Recently, the research on FDI has focused more on nonlinear system fault diagnosis. Some nonlinear techniques such as nonlinear observer methods (Frank, 1994; Garcia & Frank, 1997) and nonlinear parity relations (Krishnaswami, Luh, & Rizzoni, 1995) have been applied. Moreover, artificial intelligence has received increasing attention and been applied in the field of nonlinear FDI (e.g., see the discussion in Antsaklis & Passino, 1993). The artificial intelligent methods such as fuzzy systems, neural networks, and expert systems have the potential to "learn" the plant model from input–output data or "learn" fault knowledge from past experience, and they can be used as function approximators to construct the analytical model for residual generation, or as supervisory schemes to make the fault analysis decisions. The nonlinear modeling ability of neural networks has been utilized for nonlinear FDI problems (Watanabe, Matsuura, Abe, Kubota, & Himmelblau, 1989; Naidu, Zafiriou, & McAvoy, 1990; Sorsa, Koivo, & Koivisto, 1991; Sorsa & Koivo, 1993; Maki & Loparo, 1997; Kavuri & Venkatasubramanian, 1994). In addition, the learning ability has also been studied and successfully used in nonlinear robust FDI (Polycarpou & Vemuri, 1995; Polycarpou & Helmicki, 1995; Vemuri & Polycarpou, 1997a, b; Vemuri, Polycarpou, & Diakourtis, 1998; Demetriou & Polycarpou, 1998), where robustness, fault sensitivity and stability conditions of the learning scheme are also studied. Meanwhile, expert systems, fuzzy logic and genetic algorithms have been used in model based FDI (Lopez, Benkhedda, & Patton, 1997; Fussel, Balle, & Isermann, 1997; Passino & Antsaklis, 1988; Laukonen & Passino, 1995; Laukonen, Passino, Krishnaswami, Luh, & Rizzoni, 1995; Gremling & Passino, 1997).

This paper presents a multiple-model-based fault diagnosis method that utilizes fuzzy modeling and an expert supervisory scheme. Note that most existing multi-model fault diagnosis methods are based on linear models, while nonlinear multiple models are rarely used. This paper presents a FDI method using multiple nonlinear models and the robustness and fault sensitivity of the proposed method are also studied. The analytical system models in the form of Takagi–Sugeno fuzzy systems are developed using nonlinear system identification techniques (Section 2). In Section 3 a fault diagnosis system is described with a bank of multiple models and a supervisory scheme determines the proper model bank to generate residuals and analyzes the residuals to detect and identify faults. Robustness and fault sensitivity of the proposed approach are also studied in this section. In Section 4, the proposed method is applied to a turbine engine (the General Electric XTE46) and its effectiveness is demonstrated via component level model (LH) simulation examples.

2. Model development using Takagi–Sugeno fuzzy systems

2.1. The XTE46 turbine engine

The General Electric XTE46 engine, as shown in Fig. 1, is a simplified, unclassified version of the original IHPTET engine (Adibhatla & Lewis, 1997). To develop a model-based fault diagnosis scheme an accurate representation of the engine dynamics is desired. This model can be used to generate residuals by comparing its output to the engine output. However, modeling a turbine engine is undoubtedly a very difficult problem in the fact that the jet engine system has an iterative structure, which means that the model cannot be written down in a differential–algebraic equation form. Fortunately, a thermodynamic simulation package, the component level engine cycle model (CLM) of the XTE46 engine, was provided by General Electric Aircraft Engines (GEAE). This is a sophisticated highly nonlinear dynamic model where each engine component is simulated. The CLM executes one pass within the digital control's sampling time and thermodynamic states are assumed to be in equilibrium after each pass through the simulation. The CLM is a low-frequency transient turbofan engine simulation and volume

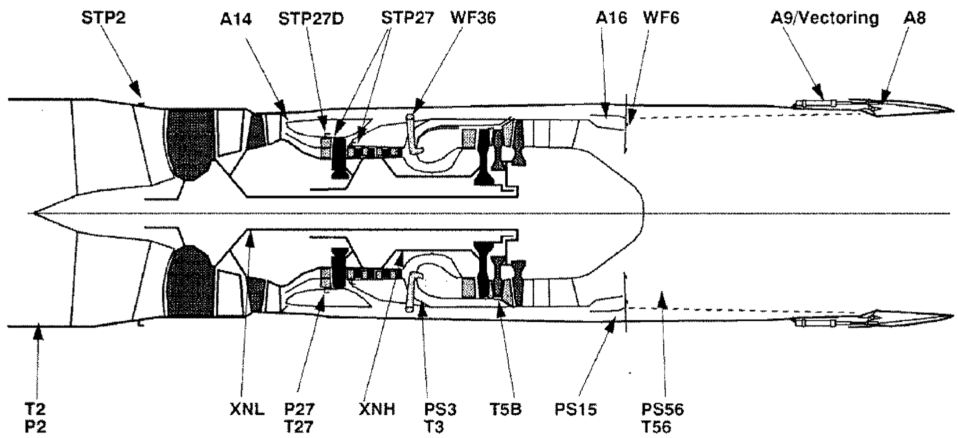

Fig. 1. Schematic of the XTE46 turbine engine.

dynamics and airflow storage affects, which are high-frequency phenomenon, are not included. The operating condition of the engine is defined by the altitude (ALT), the Mach number (XM), the difference of temperature (DTAMB), and the throttle setting represented by power code (PC). The health of the engine is described by 10 quality parameters which include the flows and efficiencies of the fan, the compressor and turbines. The model has three state variables, including the fan rotor speed (XNL), the core rotor speed (XNH), and the temperature at combustor inlet (TMPC). There are six actuators, but the major control variables are the combustor fuel flow (WF36), the exhaust nozzle area (A8), and the variable area bypass injector area (A16). (Refer to Table 1 for list of acronyms.)

In the course of developing fault diagnosis schemes, the use of analytical redundancy implies that a mathematical model of the system is used to describe the inherent relationship (or redundancy) contained among the system inputs and outputs, which can be used to generate the residuals for fault diagnosis. The resulting approaches are usually referred to as "analytical redundancy-based" fault diagnosis or "model-based" methods. From the point of view of theoretical studies in fault diagnosis, however, the CLM is too complicated. Although the CLM does provide a driver to trim the model to specified operating conditions and generate linearized models, studies show that the accuracy of the linear models is not adequate for our purposes (where we consider faults with significant nonlinear effects). Here, we developed a nonlinear model with Takagi–Sugeno fuzzy systems using a system identification methodology that utilized nonlinear transient data generated by the CLM. Note that other nonlinear modeling methods such as neural networks can also be used for modeling the engine dynamics. In this paper we choose to use the Takagi–Sugeno fuzzy systems because of its clear model interpretation, constructed from nonlinearly interpolated linear models. Besides modeling the engine behaviors at each operating node, we also use Takagi–Sugeno fuzzy systems to construct the global engine model by interpolating among local models.

Table 1
Table of acronyms

CLM	Component level model
XNL	Fan rotor speed
XNH	Core rotor speed
TMPC	Temperature at combustor inlet
ALT	Altitude
XM	Mach number
PC	Power code
WF36	Combustor fuel flow
A8	Exhaust nozzle area
A16	Variable area bypass injector area
ZSW2	Fan flow
SEDM2	Fan efficiency
ZSW7D	Compressor tip flow
SEDM7D	Compressor tip efficiency
ZSW27	Compressor hub flow
SEDM27	Compressor hub efficiency
ZSW41	High pressure turbine flow
ZSE41	High pressure turbine efficiency
ZSW49	Low pressure turbine flow
ZSE49	Low pressure turbine efficiency

2.2. The Takagi–Sugeno fuzzy system

Developing mathematical models for nonlinear systems can be quite challenging. Takagi–Sugeno fuzzy systems are capable of serving as the analytical model for nonlinear systems due to the universal approximation property, that is, any desired approximation accuracy can be achieved by increasing the size of the approximation structure and properly defining the parameters of the approximator (Passino & Yurkovich, 1998). A Takagi–Sugeno fuzzy system can be defined by

$$y = F_{ts}(x, \theta) = \frac{\sum_{i=1}^{R} g_i(x)\mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)}, \tag{1}$$

$$g_i(x) = a_{i,0} + a_{i,1}x_1 + \cdots + a_{i,n}x_n, \quad (2)$$

$$\mu_i(x) = \prod_{j=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right), \quad (3)$$

where y is the output of the fuzzy system, $x = [x_1, x_2, \ldots, x_n]^\top$ holds the n inputs, and $i = 1, 2, \ldots, R$ represent R different rules. The shapes of the membership functions are chosen to be Gaussian, and center-average defuzzification and product are used for the premise and implication in the structure of the fuzzy system. The $g_i(x)$, $i = 1, 2, \ldots, R$ are called consequent functions of the fuzzy system, where the $a_{i,j}$ are constants. The premise membership functions $\mu_i(x)$ are assumed to be well defined so that $\sum_{i=1}^{R} \mu_i(x) \neq 0$ for all x. The parameters that enter in a nonlinear fashion are c_j^i and σ_j^i, which are the centers and relative widths of the membership functions, respectively, for the jth inputs and the ith rules. Actually, the Takagi–Sugeno fuzzy system, where the consequent parts are chosen to be affine functions, is a special case of "functional fuzzy systems" (Passino & Yurkovich, 1998). If other functions such as polynomial or neural networks are used as consequent functions, different kinds of functional fuzzy systems will be generated.

The tunable parameter vector θ in (1) can be composed of both premise membership function parameters (c_j^i and σ_j^i) and consequent function parameters ($a_{i,j}$). This is referred to as *nonlinear in the parameter* case. A *nonlinear in the parameter* Takagi–Sugeno fuzzy system can be tuned by a variety of gradient methods such as the steepest descent method and Levenberg–Marquardt method. Alternatively, the parameter vector θ can be composed of only the consequent function parameters so that y is a linear function of θ. To tune the fuzzy approximator for this *linear in the parameter* case, a linear least-squares method will normally be suitable.

2.3. Fuzzy modeling for the XTE46 engine

The CLM for the XTE46 aircraft engine is a complicated multiple-input multiple-output (MIMO) nonlinear system (involving schedules, look-up tables, and partial differential equations). We assume that its fundamental dynamic characteristics, however, can be represented by a single-input single-output (SISO) system in the form

$$\dot{x} = f(x, u, c, p) + \eta(x, u, c, p), \quad (4)$$

$$y = h(x, u, c, p), \quad (5)$$

where $x = [\text{XNL}, \text{XNH}, \text{TMPC}]^\top$ is the state vector, $u = \text{WF36}$ is the input variable, $y = \text{XN2}$ is the output of the engine, $c = [\text{ALT}, \text{XM}, \text{DTAMB}, \text{PC}]^\top$ represents the operating condition of the engine, $p =$[ZSW2, SEDM2, ZSW7D, SEDM7D, ZSW27, SEDM27, ZSW41, ZSE41, ZSW49, ZSE49]$^\top$ represents the quality parameter vector, $f(\cdot)$ denotes the unknown function representing the nonlinear characteristics of the engine, $h(\cdot) = \text{XNL}$ because the output variable XN2 is the measurement of the state variable XNL, and $\eta(\cdot)$ represents model uncertainty caused by describing the MIMO engine by a SISO system.

The analytical model of the engine is developed in two steps. Fuzzy identification is applied first to generate (a grid of) "node" models specified by operating conditions and quality parameters. Afterwards, the "global" model can be constructed by fuzzy interpolation on these node models. (We use this two-step method rather than identifying the model directly from the data collected from the whole space of operating conditions and quality parameters in that it is practically impossible to train an approximator using such a large amount of data, due to the limitations of computing resources.) Given a specific operating condition (c_i) and fixed values of quality parameters (p_i), the node model for the XTE46 engine can be obtained using nonlinear identification technique as

$$\text{X}\widehat{\text{NL}}\text{dot}(k) = F_{ts}^1(\widehat{\text{XNL}}(k), \widehat{\text{XNH}}(k), \text{WF36}(k), \theta^1(c_i, p_i)), \quad (6)$$

$$\text{X}\widehat{\text{NH}}\text{dot}(k) = F_{ts}^2(\widehat{\text{XNL}}(k), \widehat{\text{XNH}}(k), \text{WF36}(k), \theta^2(c_i, p_i)), \quad (7)$$

$$\widehat{\text{XNL}}(k+1) = \widehat{\text{XNL}}(k) + T_s \text{X}\widehat{\text{NL}}\text{dot}(k), \quad (8)$$

$$\widehat{\text{XNH}}(k+1) = \widehat{\text{XNH}}(k) + T_s \text{X}\widehat{\text{NH}}\text{dot}(k), \quad (9)$$

$$\widehat{\text{XN2}}(k+1) = \widehat{\text{XNL}}(k+1), \quad (10)$$

where two multiple-input single-output (MISO) Takagi–Sugeno fuzzy systems, F_{ts}^1 and F_{ts}^2, are specified with corresponding parameters θ^1 and θ^2, respectively. The variables $\widehat{\text{XNL}}(k)$ and $\widehat{\text{XNH}}(k)$ denote the estimated values of XNL(k) and XNH(k) ($\widehat{\text{XNL}}(0) = \text{XNL}(0)$ and $\widehat{\text{XNH}}(0) = \text{XNH}(0)$ to let the fuzzy model have the same initial values as the engine), and $\widehat{\text{XN2}}(k)$ is the estimated value of XN2(k). $\text{X}\widehat{\text{NL}}\text{dot}(k)$ and $\text{X}\widehat{\text{NH}}\text{dot}(k)$ are the outputs of the fuzzy systems. T_s is the sample time, which is 0.02 s. The fuzzy systems are trained using engine data generated by the transient driver of the CLM simulator of the XTE46 engine. One thousand engine input–output data pairs are collected which reflect the transient performance of the engine for 20 s (sampled every 0.02 s) at a specific "node" (of operating conditions and quality parameters). For the kth, experimental data pair, the input data are the state variables XNL(k) and XNH(k), and the input variable WF36(k). The output data are XNLdot(k) and XNHdot(k), which denote the derivatives of XNL(k) and XNH(k), respectively.

The structure of fuzzy approximators is selected by having the premise input be $\widehat{\mathrm{XNL}}(k)$ only because most nonlinearity can be captured by $\widehat{\mathrm{XNL}}(k)$ (and this selection will result in an affine model, which can be beneficial to fault tolerant control using stable adaptive fuzzy/neural approach). The Takagi–Sugeno fuzzy system can be written in the form of

$$F_{ts}(\widehat{\mathrm{XNL}}(k), \widehat{\mathrm{XNH}}(k), \mathrm{WF36}(k), \theta) = \frac{\sum_{j=1}^{R} (a_{j,0} + a_{j,1}\widehat{\mathrm{XNL}}(k) + a_{j,2}\widehat{\mathrm{XNH}}(k) + a_{j,3}\mathrm{WF36}(k))\mu_j(\widehat{\mathrm{XNL}}(k))}{\sum_{j=1}^{R} \mu_j(\widehat{\mathrm{XNL}}(k))}.$$

Using trial and error, fuzzy models with $R = 3$ rules each were selected. The models are tuned by the Levenberg–Marquardt method. Note that since the engine models are identified off-line, the stability and accuracy of the models can be validated prior to their applications to fault diagnosis.

Notice that we did not use TMPC as one of the model inputs because it is not measurable and the importance of TMPC to the model accuracy is insignificant, which is verified by an input selection method referred to as "regressor analysis" (where the regression models are constructed and the regression coefficients are analyzed to determine the importance of each input), so that it is not necessary to estimate TMPC either. Also note that the fuzzy models are running in an "open-loop manner", i.e., the outputs of the models will be feeded back into the models as the inputs, and only the basic behavior of the engine can be obtained due to the drifting effects caused by the accumulation of approximation errors. We use this approach because sometimes this analytical model is desired to run as the "truth engine" in the simulation study, where the CLM and thus the engine states XNL and XNH are not available. Certainly, when we utilize the model in the fault diagnosis of the CLM, we can use the (measurable) real engine states to be the inputs of fuzzy systems and expect an improvement in model accuracy.

By nonlinearly interpolating between a grid of node models obtained above from nonlinear system identification, the global model can then be constructed which is in the form

$$\widehat{\mathrm{XNLdot}}(k) = \frac{\sum_{i=1}^{N} F_{ts}^{1}(\widehat{\mathrm{XNL}}(k), \widehat{\mathrm{XNH}}(k), \mathrm{WF36}(k), \theta^1(c_i, p_i))\mu_i(z)}{\sum_{i=1}^{N} \mu_i(z)}, \quad (11)$$

$$\widehat{\mathrm{XNHdot}}(k) = \frac{\sum_{i=1}^{N} F_{ts}^{2}(\widehat{\mathrm{XNL}}(k), \widehat{\mathrm{XNH}}(k), \mathrm{WF36}(k), \theta^2(c_i, p_i))\mu_i(z)}{\sum_{i=1}^{N} \mu_i(z)}, \quad (12)$$

$$\mu_i(z) = \prod_{j=1}^{14} \exp\left(-\frac{1}{2}\left(\frac{z_j - c_j^i}{\sigma_j^i}\right)^2\right), \quad (13)$$

where $i = 1, 2, \ldots, N$ represent N different models and $z = [c^\top, p^\top]^\top$ is the premise input vector including 14 variables.

We choose to focus on fault diagnosis system development for the "climb" region (which is defined as $\mathrm{ALT} \in [12\,500, 17\,500]$, $\mathrm{XM} \in [0.6, 0.8]$, $\mathrm{DTAMB} \in [-35, 35]$, and $\mathrm{PC} \in [45, 50]$). We partition each operating condition variable into three regions to define our grid. In this way, we have $3^4 = 81$ models to describe the nonlinearity presented in the climb region. The values of quality parameters are composed of four parts: the nominal value, the initial engine variation due to manufacturing differences, the quality parameter adjustment resulted from engine deterioration, and the quality parameter change due to the faults. Note that the effects of engine deterioration and faults are larger than the initial engine variation, and we will try to capture the characteristics of these two factors and leave the effect of initial engine variation to be model uncertainty. By assuming that the engine deterioration affects 10 quality parameters in the same way (so that we may have three grids to represent no deterioration, half deterioration, and full deterioration, respectively) and considering four sizes of faults (no fault, small, medium, or large fault), we have $3 \times 4 \times 4 = 48$ models to describe the nonlinearity presented in the quality parameters (for simplicity, we only consider two possible faults, fan fault and compressor hub fault, here.) In total, we have $81 \times 48 = 3888$ (nonlinear) node models to describe the nonlinearity in the climb region. We need this level of complexity to obtain a reasonably accurate "design model" for the development of our fault diagnosis scheme.

The general form of the model can be described as

$$\dot{x} = f(x, u, c, p) + \eta(x, u, c, p), \quad (14)$$

where

$$f(x, c, p) = \frac{\sum_{i=1}^{N} \underline{f}(x, c_i, p_i)\mu_i(c, p)}{\sum_{i=1}^{N} \mu_i(c, p)}, \quad (15)$$

$$\underline{f}(x, c_i, p_i) = \frac{\sum_{j=1}^{R} [a_{j,0}(c_i, p_i) + a_{j,1}(c_i, p_i)x_1 + a_{j,2}(c_i, p_i)x_2 + a_{j,3}(c_i, p_i)u]\underline{\mu}_j(x_1)}{\sum_{j=1}^{R} \underline{\mu}_j(x_1)}, \quad (16)$$

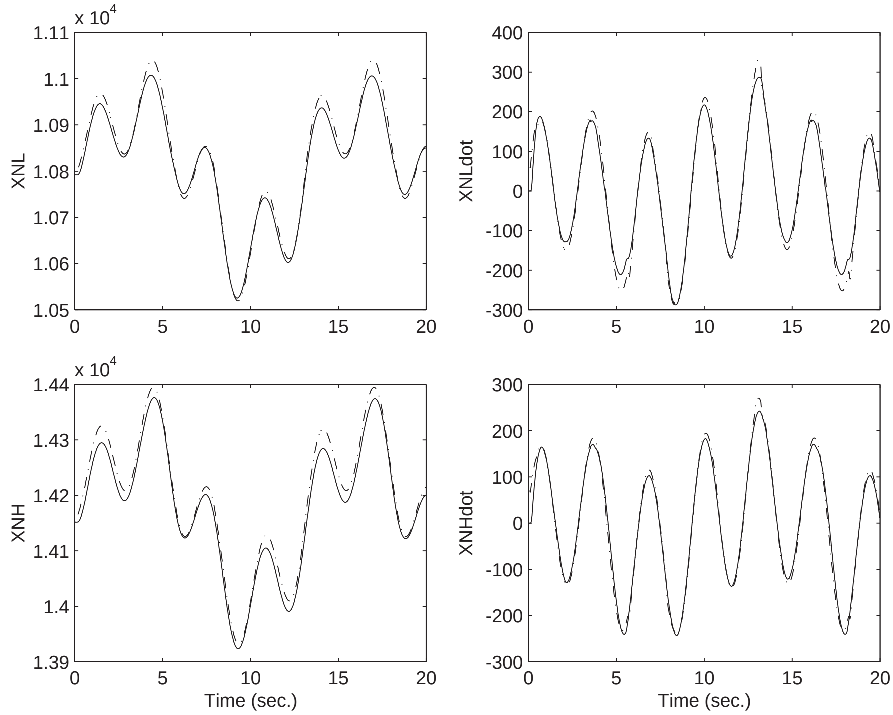

Fig. 2. Illustrative example of the model performance.

where $u = \mathrm{WF36}$ and $x = [x_1, x_2]^\top = [\widehat{\mathrm{XNL}}, \widehat{\mathrm{XNH}}]^\top$. The value of c is the known operating condition vector, p is unknown quality parameter vector, and c_i, p_i are to specify nodes where we establish "node models". In addition, f and $\underline{f}$ represent 2×1 function vectors. The model parameters $a_{j,0}, a_{j,1}, a_{j,2}, a_{j,3}$ are 2×1 vectors of linear parameters of Takagi–Sugeno fuzzy systems, and $\underline{\mu}_j(x_1)$ are membership functions describing the nonlinearity with respect to x_1. They are identified using the Levenberg–Marquardt method, separately for each node. The function $\mu_i(c,p)$ are membership functions of fuzzy interpolation between different operating conditions and quality parameters, whose parameters are also identified using the Levenberg–Marquardt method.

The function η represents model uncertainty caused by using finite size approximators to approximate the engine dynamics. Note that we represent the relationship between the model and model uncertainty to be additive but the actual relationship may not be so. This is because no matter what kind of systems we consider and what kind of model is obtained, the model uncertainty (η) can always be represented as the difference between the system dynamics (F) and the model dynamics (f), i.e. $\eta = F - f$.

The resulting nonlinear model provides a reasonably accurate representation of engine dynamics (GE Aircraft Engines verified this for us). Here we give an example at a point different from the nodes where we generated the model. The engine operating conditions are $\mathrm{ALT} = 16\,000$, $\mathrm{XM} = 0.75$, $\mathrm{DTAMB} = 0$, and $\mathrm{PC} = 46$. The quality parameters are defined by considering some initial engine variation, nearly half engine deterioration, and a fan fault a little bit larger than medium size. Fig. 2 compares system responses between the nonlinear model and the CLM and indicates their similarity (where the solid lines represent the system response of the CLM and the dashed lines represent that of the analytical model). We also conducted many other such simulations to verify the quality of our design model; however, in the interest of brevity we do not include those plots here. Compared to modeling engine dynamics using local linear models (either through a CLM driver to trim the model to specified operating conditions and generate linearized models, or by using fuzzy models with $R = 1$), using local nonlinear models generally provides more accurate modeling results, especially when the input variable WF36 has a large variation magnitude and the state variables XNL and XNH are far away from the operating point.

3. Fault diagnosis by interfacing multiple models with a supervisory scheme

3.1. Residual generation using multiple models

The engine simulation can represent many situations. For example, it can simulate what happens for several types of faults, possibly with manufacturing differences and engine deterioration. Usually, engine faults result in changes of the input–output characteristics of the system so that the faults may be detected by evaluating the residuals generated by comparing the model output and the engine output. Some faults, however, may affect all such residuals, so that if we generate residuals with only a single model, we can only detect but not isolate such faults. In such situations, a "multiple model" method may be preferable. In this approach we utilize different models to identify different fault situations, and isolate faults by comparing the residuals corresponding different models. The structure of the fault diagnosis scheme using multiple models and a supervisory expert system is shown in Fig. 3.

Assume that there are N possible fault situations and we have $N+1$ models $\{M_i\}_{i=0}^{N}$, where the nominal model M_0 corresponds to the situation where there is no fault, and the fault models M_i, $i=1,2,\dots,N$ represent the ith fault situations. These models can be obtained by using nonlinear system identification with Takagi–Sugeno fuzzy systems (as explained in the previous section). Note that the faults affect the engine dynamics through engine quality parameters, and the engine quality parameters are also affected by initial engine variation (due to manufacturing differences) and engine deterioration, so that the quality parameter vector p in $f(x,u,c,p)$ of (14) is actually unknown (and unmeasurable as well). However, we can represent the engine dynamics of different fault situations as $f(x,u,c,p_i^0)$ with p_i^0 indicating the expected quality parameters with respect to the corresponding fault situation, and consider the effects of initial engine variation and engine deterioration as model uncertainty. By running a bank of models on-line (which are selected by the expert supervisory scheme as we will explain later), the residuals r_i are generated by $r_i = y - \hat{y}_i$, where y is the engine output and $\hat{y}_i$ is the output of the model M_i which corresponds to the ith fault situation. (More discussion on the forms of the system and models will be given later in the section of robustness analysis.)

3.2. Expert supervisory scheme

The determination of the on-line model bank composition and the evaluation of residual signals are conducted using an expert supervisory scheme. The advantage of using an expert system is that the heuristic knowledge of faults and our experience in handling faults can be easily incorporated into the expert system in the form of rules, and the operation of the expert system is transparent so that we can investigate the residual evaluation process of the expert system. Basically, the rule-based expert supervisory system of our fault diagnosis scheme performs the following functions:

1. *Determine the on-line models to generate the residuals*: Our experience of working with the engine provide us some a priori knowledge of the faults, for instance, the possible types of the faults and the possible combinations of these faults. According to this, $N+1$ models have been established that correspond to different no-fault or fault situations. Besides, we know there exist certain relationships among these situations. For instance, suppose fault i is happening at present, then the next possible fault situation can be the existence of fault i together with fault j (a fault that could occur in the presence of fault i), but not the situations where there exist three faults under the assumption that no two faults can occur simultaneously, or no fault exists if we know that the fault cannot be self-recovered. Therefore, at each time instant only a subset of $N+1$ situations

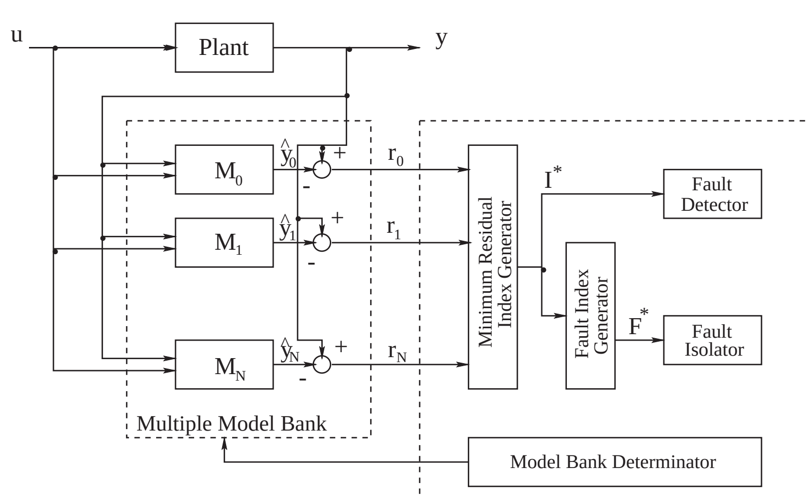

Fig. 3. Structure of the fault diagnosis scheme.

can happen and thus only a subset of models are required to be on-line to generate the residuals, which can be summarized as the rule: *IF (fault situation i is currently in existence) THEN (a subset of models (S_i) will be used as the on-line model bank to generate residuals afterwards)*. In particular, the rules can include the following:

- IF (there is no fault in existence) THEN (the model bank consists of no fault model and all possible single-fault models).
- IF (there is a single fault in existence) THEN (all single-fault models except the one that has been isolated will be removed from the model bank, and the multiple-fault models including the isolated fault and another possible fault will be added).
- IF (there are two faults in existence) THEN (the on-line model bank includes the model corresponding to the situation with these two faults, and the models corresponding to the situations with three faults, i.e., these two faults and another possible fault).

The advantage of using the fault knowledge to select the on-line model bank is to reduce the computational complexity caused by using this multiple model fault diagnosis approach.

2. *Fault detection*: Once the residuals are generated, the fault decision logic is used to diagnose the faults. A sequence of "minimum residual indices" is first generated by

$$I^*(k) = \arg \min_{i \in S_i} (r_i(k)), \tag{17}$$

where $I^*(k)$ denotes the index of the model whose residual is the minimum, which corresponds to the most appropriate model to represent the engine at time instant k. Note that the change of fault situations results in the change of the input–output characteristics of the system. Therefore, the change of index, i.e., the change of most suitable model, may be used to indicate the change of fault situations, i.e., the occurrence of new faults.

3. *Fault isolation*: The change of index can serve as a fault detector, but not a fault isolator. Actually, the new index may not indicate that the new fault is the one corresponding to this index. This is because during the transient phase the residuals corresponding to the on-line models may change drastically and some of them may happen to be very small for a short time (and become large afterwards since they are not the one corresponding to the new fault situation). In order to isolate the fault situations correctly, some logic rules are used to guarantee that a fault will be isolated only if its corresponding index is the minimum index at least for T_0 seconds (to indicate the suitability of the model). The time delay term T_0 will be used to ensure robustness of fault diagnosis (which we will discuss later), and its value can be obtained by trial and error to balance the robustness and sensitivity of fault isolation. The results of isolation is represented by a fault index

$$F^*(k) = \arg \min_{i \in S_i, k-T_0 \leqslant j \leqslant k} (r_i(j)), \tag{18}$$

where $F^*(k)$ denotes the index of the model whose residual is minimum for a period of T_0. Actually, the fault diagnosis strategy of using a time delay term may affect fault sensitivity to some extent but it is often useful in isolating faults accurately (which is more important), and it generally does not affect the sensitivity of fault detection. As a result, we will have a fault diagnosis scheme that can detect a fault relatively fast, but isolate the fault type a little more slowly but accurately, which is quite reasonable.

3.3. Robustness analysis

The robustness of fault diagnosis refers to the ability to prevent false alarms in the presence of modeling uncertainties, i.e., if the system is in the ith fault situation, the fault diagnosis system should indicate the ith fault situation rather than the jth fault situation where $j \neq i$. The robustness of the proposed fault diagnosis system is achieved by using the time delay term T_0 in the fault isolation scheme.

We study the nonlinear dynamic system in the ith fault situation that can be described by

$$\dot{x} = f(x, u, c, p_i^0) + \eta_i(x, u, c, p_i), \tag{19}$$

$$y = Cx, \tag{20}$$

where $C = [c_1, c_2, \ldots, c_n]$, $c_l > 0$, $l = 1, 2, \ldots, n$ and $f = [f_1, f_2, \ldots, f_n]^\top$ denotes the known model dynamics, which are characterized by the developed Takagi–Sugeno model with "nominal" quality parameter vector p_i^0 (corresponding to the ith fault situation but without considering initial engine variation and engine deterioration). The functions $\eta_i = [\eta_{i,1}, \eta_{i,2}, \ldots, \eta_{i,n}]^\top$ represent modeling uncertainties in the system (e.g., caused by ignoring initial engine variation and engine deterioration, and by representing the system dynamics with finite size approximators).

The model corresponding to the jth fault situation can be represented by

$$\dot{\hat{x}}_j = f(\hat{x}_j, u, c, p_j^0), \tag{21}$$

$$\hat{y}_j = C\hat{x}_j, \tag{22}$$

where $\hat{x}_j$ is the estimated state vector of the jth model and $\hat{y}_j$ is the estimated output.

To study the robustness of the fault diagnosis scheme for the above nonlinear system, we have the following assumptions.

(A1) The modeling errors are additive, unstructured, and bounded, and the accumulation effect of modeling uncertainties can be described by

$$\left|\int_0^t \eta_{i,l}(x,u,c,p_i)\,\mathrm{d}\tau\right| \leqslant \eta_{i,l}^0, \tag{23}$$

where $l = 1,2,\ldots,n$, and $\eta_i^0 = [\eta_{i,1}^0, \eta_{i,2}^0, \ldots, \eta_{i,n}^0]^\top$ are constant bounds on modeling uncertainties.

(A2) The ith model can represent the system dynamics in the ith fault situation accurately enough compared to the modeling uncertainties, i.e.,

$$\left|\int_0^t (f_l(x,u,c,p_i^0) - f_l(\hat{x}_i,u,c,p_i^0))\,\mathrm{d}\tau\right| \leqslant \bar{\alpha}\eta_{i,l}^0. \tag{24}$$

(A3) The difference between the jth and ith model ($j \neq i$) is large enough compared to the modeling uncertainties, i.e.,

$$\max_{t-T_0 \leqslant t' \leqslant t}\left|\int_0^{t'} (f_l(x,u,c,p_i^0) - f_l(\hat{x}_j,u,c,p_j^0))\,\mathrm{d}\tau\right| \geqslant \underline{\alpha}\eta_{i,l}^0. \tag{25}$$

The constants $\bar{\alpha}$ and $\underline{\alpha}$ are parameters that satisfy the inequality $\underline{\alpha} \geqslant \bar{\alpha} + 2$ (which will be used later for robustness analysis). We use "max" to represent that $|\int_0^{t'} (f_l(x,u,c,p_i^0) - f_l(\hat{x}_j,u,c,p_j^0))\,\mathrm{d}\tau|$ may only be smaller than $\underline{\alpha}\eta_{i,l}^0$ for no longer than T_0 seconds, which could happen when the estimated output of a "wrong" model approaches the engine output for a short period of time and then leaves again. Note that the effects of the faults on the quality parameters are usually larger than those of initial engine variation and engine deterioration, and the inaccuracy that arises from the accumulation of drifting errors is usually smaller than that from modeling uncertainties; hence, the above assumptions can be satisfied in real applications.

With the above assumptions the robustness problem of the fault diagnosis system is that in the ith fault situation, the residual of the jth model ($r_j = y - \hat{y}_j$, $j \neq i$) cannot be smaller than the residual of the ith model ($r_i = y - \hat{y}_i$) for a period of T_0 time, i.e.,

$$\max_{t-T_0 \leqslant t' \leqslant t} |r_i(t')| \leqslant \max_{t-T_0 \leqslant t' \leqslant t} |r_j(t')|, \tag{26}$$

so that the fault diagnosis system will still indicate "i" as the current fault situation. The following analysis ensures robustness of the proposed fault diagnosis scheme. Related analysis, but for a different problem, can be found in Vemuri and Polycarpou (1997a).

For the ith model, defining the state estimation error to be $e_i = x - \hat{x}_i$, the error dynamics are described by

$$\dot{e}_i = f(x,u,c,p_i^0) - f(\hat{x}_i,u,c,p_i^0) + \eta_i(x,u,c,p_i), \tag{27}$$

$$r_i = Ce_i. \tag{28}$$

By processing the differential equation we obtain

$$\begin{aligned}|r_i(t)| &= \left|C\int_0^t (f(x,u,c,p_i^0) - f(\hat{x}_i,u,c,p_i^0) + \eta_i(x,u,c,p_i))\,\mathrm{d}\tau\right| \\ &\leqslant C\left|\int_0^t (f(x,u,c,p_i^0) - f(\hat{x}_i,u,c,p_i^0))\,\mathrm{d}\tau\right| + C\left|\int_0^t \eta_i(x,u,c,p_i)\,\mathrm{d}\tau\right| \\ &\leqslant C\bar{\alpha}\eta_i^0 + C\eta_i^0 \\ &= (\bar{\alpha}+1)C\eta_i^0\end{aligned}$$

and thus

$$\max_{t-T_0 \leqslant t' \leqslant t} |r_i(t')| \leqslant (\bar{\alpha}+1)C\eta_i^0. \tag{29}$$

The error dynamics of the jth model ($j \neq i$) can be described by

$$\dot{e}_j = f(x,u,c,p_i^0) - f(\hat{x}_j,u,c,p_j^0) + \eta_i(x,u,c,p_i), \tag{30}$$

$$r_j = Ce_j. \tag{31}$$

From the differential equation we obtain

$$\begin{aligned}\max_{t-T_0 \leqslant t' \leqslant t} |r_j(t')| &= \max_{t-T_0 \leqslant t' \leqslant t}\left|C\int_0^{t'} (f(x,u,c,p_i^0) - f(\hat{x}_j,u,c,p_j^0) + \eta_i(x,u,c,p_i))\,\mathrm{d}\tau\right| \\ &\geqslant C\max_{t-T_0 \leqslant t' \leqslant t}\left|\int_0^{t'} (f(x,u,c,p_i^0) - f(\hat{x}_j,u,c,p_j^0))\,\mathrm{d}\tau\right| - C\max_{t-T_0 \leqslant t' \leqslant t}\left|\int_0^{t'} \eta_i(x,u,c,p_i)\,\mathrm{d}\tau\right| \\ &\geqslant \underline{\alpha}C\eta_i^0 - C\eta_i^0 \\ &= (\underline{\alpha}-1)C\eta_i^0.\end{aligned}$$

Therefore, $\max_{t-T_0 \leqslant t' \leqslant t} |r_i(t')| \leqslant \max_{t-T_0 \leqslant t' \leqslant t} |r_j(t')|$ when $\underline{\alpha} \geqslant \bar{\alpha} + 2$, which means the robustness of the fault diagnosis system is guaranteed and no false alarms will be generated.

3.4. Fault sensitivity analysis

Fault sensitivity of fault diagnosis refers to the ability to correctly determine the existence of a fault and then isolate its type. In particular, if a fault happens at time T_f and thus the system changes from the ith fault situation to the kth fault situation ($k \neq i$) at that time, the fault diagnosis system can isolate the kth situation after an *isolation time* T_i if for the kth model and

$t \geqslant T_f + T_i - T_0$

$$\left| \int_0^t (f_l(x,u,c,p_k^0) - f_l(\hat{x}_k,u,c,p_k^0))\,\mathrm{d}\tau \right| \leqslant \bar{\beta}\eta_{k,l}^0 \tag{32}$$

and for jth model ($j \neq k$ but j may be equal to i) and $T_f + T_i - T_0 \leqslant t \leqslant T_f + T_i$

$$\left| \int_0^t (f_l(x,u,c,p_k^0) - f_l(\hat{x}_j,u,c,p_j^0))\,\mathrm{d}\tau \right| \geqslant \underline{\beta}\eta_{k,l}^0, \tag{33}$$

where $\bar{\beta}$ and $\underline{\beta}$ are parameters that satisfy the inequality $\underline{\beta} \geqslant \bar{\beta} + 2$ (which will be used later for fault sensitivity analysis). Under above conditions, using the triangle inequality the residual corresponding to the kth model has the following relation (when $t \geqslant T_f + T_i - T_0$):

$$\begin{aligned}
|r_k(t)| &= \left| C \int_0^t (f(x,u,c,p_k^0) - f(\hat{x}_k,u,c,p_k^0) \right. \\
&\qquad \left. + \eta_k(x,u,c,p_k))\,\mathrm{d}\tau \right| \\
&\leqslant C \left| \int_0^t (f(x,u,c,p_k^0) - f(\hat{x}_k,u,c,p_k^0))\,\mathrm{d}\tau \right| \\
&\qquad + C \left| \int_0^t \eta_k(x,u,c,p_k)\,\mathrm{d}\tau \right| \\
&\leqslant C\bar{\beta}\eta_k^0 + C\eta_k^0 \\
&= (\bar{\beta}+1)C\eta_k^0
\end{aligned}$$

and for $T_f + T_i - T_0 \leqslant t \leqslant T_f + T_i$ we have

$$\begin{aligned}
|r_j(t)| &= \left| C \int_0^t (f(x,u,c,p_k^0) - f(\hat{x}_j,u,c,p_j^0) \right. \\
&\qquad \left. + \eta_k(x,u,c,p_k))\,\mathrm{d}\tau \right| \\
&\geqslant C \left| \int_0^t (f(x,u,c,p_k^0) - f(\hat{x}_j,u,c,p_j^0))\,\mathrm{d}\tau \right| \\
&\qquad - C \left| \int_0^t \eta_k(x,u,c,p_k)\,\mathrm{d}\tau \right| \\
&\geqslant \underline{\beta}C\eta_k^0 - C\eta_k^0 \\
&= (\underline{\beta}-1)C\eta_k^0.
\end{aligned}$$

Therefore, in the time interval $[T_f + T_i - T_0,\ T_f + T_i]$, we have $|r_k(t)| \leqslant |r_j(t)|$ since $\underline{\beta} \geqslant \bar{\beta} + 2$, which implies that the fault isolation scheme can indicate the kth fault situation at the time $T_f + T_i$. (Related analysis, but for a different problem, is given in Vemuri and Polycarpou (1997a).) Note that the above analysis on robustness and fault sensitivity are obtained under worst-case conditions, which are, in other words, sufficient but not necessary conditions to guarantee the robustness and fault sensitivity of the fault diagnosis system, so that the analysis are quite conservative.

4. Component level model simulation

To study the effectiveness of the proposed fault diagnosis method we pick an engine in the climb region. Specifically, the operating condition variables are chosen to be ALT = 15 000, XM = 0.7, DTAMB = 0, and PC = 46. For the quality parameters of the engine, we set the initial engine variation to be p_{iev} = [0.1%, 0.1%, 0.2%, 0.1%, −0.1%, 0, −0.3%, 0.3%, −0.1%, 0.1%], and the engine deterioration index to be 0.1. For simplicity, we only consider two fault types: fan fault and compressor hub fault, and each fault may have three different levels: small, medium, and large. Using Takagi–Sugeno fuzzy systems we generated 16 models including one nominal model (no fault), six single fault models (three models corresponding to small, medium, and large fan fault, respectively, and three models of small, medium, and large compressor hub fault), and nine double fault models (small fan fault with small compressor hub fault, small fan fault with medium compressor hub fault, etc.). The fault diagnosis results for four different fault situations are given as below. We also conducted many other such simulations to verify the effectiveness of our fault diagnosis scheme; however, in the interest of brevity we do not include them here.

4.1. No fault

First, we study the case where there is no fault. The expert supervisory scheme determines the initial bank of on-line multiple models to consist of six single-fault models and one no-fault model. The residuals will be generated by comparing the outputs of these seven models and the engine output. Afterwards, the residuals are analyzed by the expert supervisory scheme to diagnose whether there are any faults in the engine. The output of the engine (XN2) and the estimated outputs (and residuals) of multiple models are shown in Fig. 4, where the wide solid lines represent the engine output, the thin solid lines correspond to the nominal model (model 0, or index 0), the thin dotted lines represent the small fan fault model (model 1, or index 1), the thin dash–dotted lines represent the medium fan fault model (model 2, or index 2), the thin dashed lines represent the large fan fault model (model 3, or index 3), the wide dotted lines represent the small compressor hub fault model (model 4, or index 4), the wide dash–dotted lines represent the medium compressor hub fault model (model 5, or index 5), and the wide dashed lines represent the large compressor hub fault model (model 6, or index 6). Note that there is some oscillation in the beginning of the simulation, which can be seen in the plot of "minimum residual index". However, the oscillation has been attenuated by the isolation scheme, where the time delay term is designed to be $T_0 = 1.2$ s by trial and error in order to let the fault diagnosis scheme

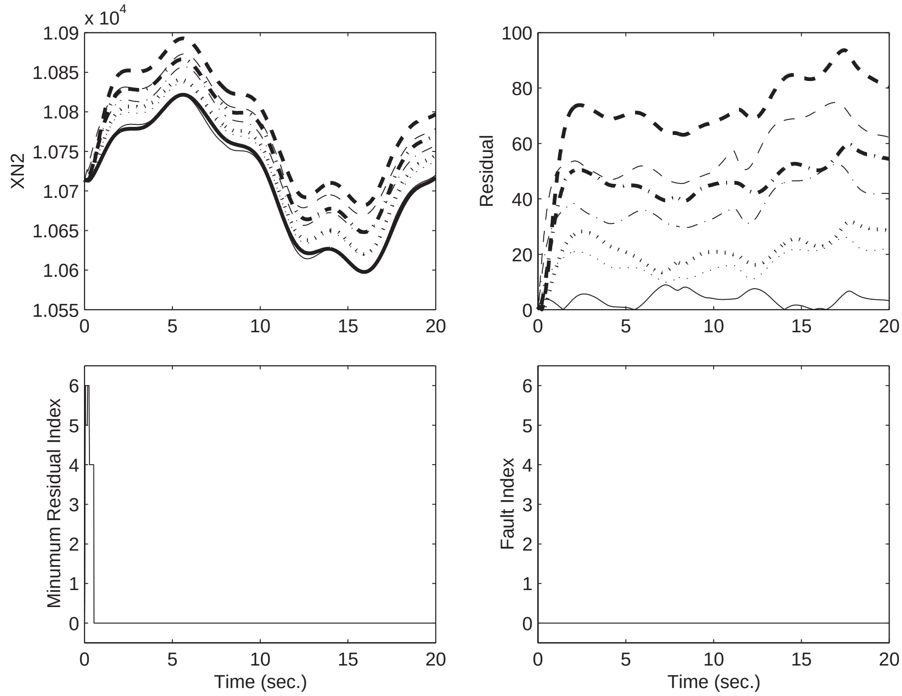

Fig. 4. No-fault situation.

be robust to the modeling uncertainties and sensitive to faults. Thus, the "fault index", as seen in Fig. 4, is always zero, which indicates that the nominal (no-fault) model always serves as the most suitable model and the engine is running in the no-fault situation.

From this case (no-fault, $i = 0$) we can also study how the assumptions for robustness analysis (A1, A2, and A3) are hold. From the "Residual" plot in Fig. 4, we notice that $|r_0(t)| \leqslant 9$, i.e., we should be able to find $\bar{\alpha}$ and η_0^0 such that $(\bar{\alpha}+1)\eta_0^0 \geqslant 9$. For the residuals generated from other models ($j = 1, 2, \ldots, 6$) we have $\max_{t-T_0 \leqslant t' \leqslant t} |r_j(t')| \geqslant 10$, which implies that $(\underline{\alpha}-1)$ $\eta_0^0 \leqslant 10$. We also know that the residual between the nominal (no-fault) model and the nominal engine (not shown in this figure) is smaller than 2.8, which means we should have $\bar{\alpha}\eta_0^0 \geqslant 2.8$. Therefore, if we pick $\eta_0^0 = 6$, $\bar{\alpha} = 0.5$, and $\underline{\alpha} = 2.6$, all the above inequalities will hold and $\underline{\alpha} \geqslant \bar{\alpha} + 2$ also holds. Studies of the validity of assumptions for other cases and for the assumptions of fault sensitivity analysis (32) and(33) are similar to the above analysis.

4.2. Small fan fault

Next, we study the case where only the small fan fault occurs in the "take-off" region and there is no compressor hub fault in the "climb" region. The initial bank of on-line multiple models is chosen to be composed of six single-fault models and one no-fault model. As shown in Fig. 5, the fault is isolated at 2.44 s. Note that the residuals corresponding to the small fan fault model and the small compressor hub fault model are interlaced, which imply that these two small faults have some similar signatures. However, since the small fan fault is isolated at 2.44 s, the small compressor hub fault model will not be used after that to generate the residual so that it will not cause any confusion. (Here, we provide them just for the purpose of illustration.) Also note that the multiple-fault models are not used after the small fan fault was isolated. This is because we want to keep this example simple and thus assume that only single fault can occur here. Studies of the multiple-fault cases will be given below.

4.3. Multiple faults: case 1

The effectiveness of the fault diagnosis scheme for multiple-fault situations is illustrated in the example, where there is no fault for the first 5 s, a small compressor hub fault occurs and lasts for 10 s, and at 15 s the size of the compressor hub fault changes from small to large. Here, six single-fault models and one no-fault model are used as the initial bank of on-line models. As shown in the "minimum residual index" plot in Fig. 6, a fault (index 1, small fan fault) is detected at 5.3 s. However, this is just the transient phenomenon that occurred after the change of the fault situation. After it lasts for 0.86 s, the minimum residual index changes from 1 to 4, which indicates a small compressor hub fault. This index is the minimum for more than 1.2 s. At 7.38 s the small compressor hub fault is claimed to be isolated by the fault isolation scheme (using the

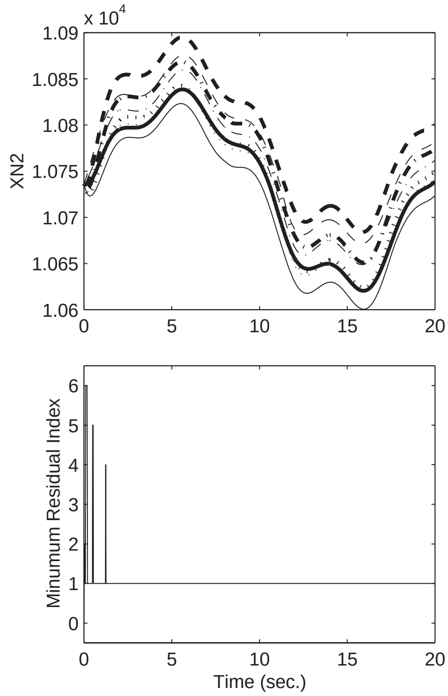

Fig. 5. Small fan fault situation.

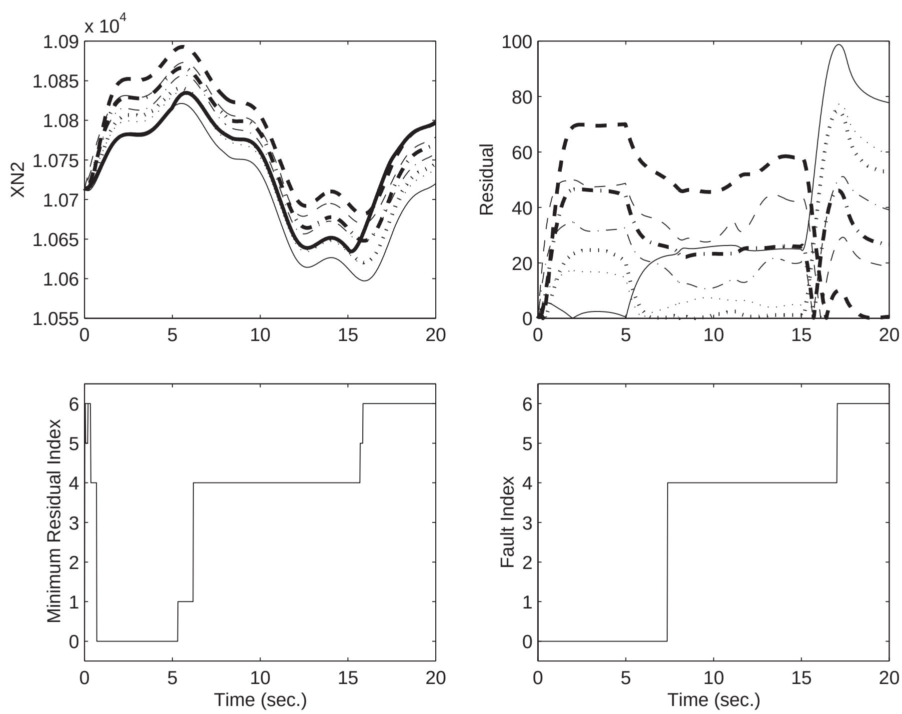

Fig. 6. Small compressor hub fault followed by large compressor hub fault.

time delay term T_0), as shown in the ‘‘fault index’’ plot. Afterwards, only three single-fault models (small, medium, and large compressor hub fault models) will be used as on-line models to generate the residuals and at 15.7 s a fault is declared to be detected, as shown in the ‘‘minimum residual index’’ plot and at 17.04 s the large compressor fault is isolated, as shown in the ‘‘fault index’’ plot. Clearly, a fault can be detected shortly after it occurs, and be isolated correctly a bit later to help maintain robustness.

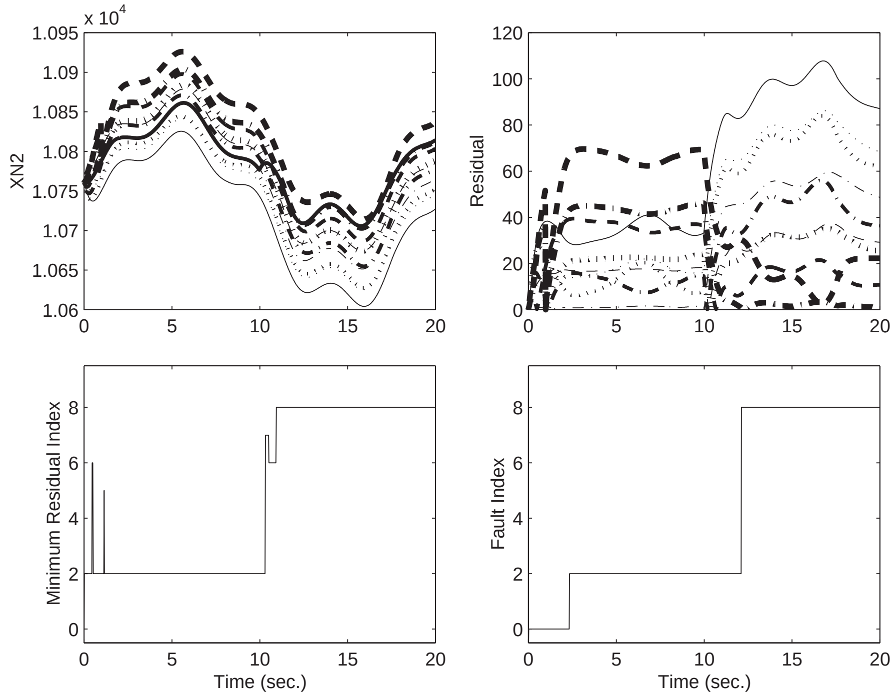

Fig. 7. Medium fan fault followed by medium compressor hub fault.

4.4. Multiple faults: case 2

Another multiple-fault case we have studied is shown in Fig. 7, where the medium fan fault exists in the beginning and the medium compressor hub fault occurs at 10 s. The medium fan fault is isolated by the fault diagnosis system at 2.34 s. Note that there are significant oscillations for the first 2 s. After that, a bank of four models (which correspond to medium fan fault only, medium fan fault with small compressor hub fault, medium fan fault with medium compressor hub fault, and medium fan fault with large compressor hub fault) are used to generate residuals. A fault is then detected at 10.3 s and the medium compressor hub fault is isolated at 12.12 s.

4.5. Robustness and sensitivity issues

To obtain a sensitive fault detection scheme, any changes in the minimum residual index (i.e., the change of the most suitable model) will be considered as indicating the occurrence of new faults, and the minimum residual index is used to detect the fault promptly. In our simulation examples, the detection delay time is less than 1 s. Furthermore, considering the presence of modeling uncertainties, a time delay term T_0 is included to achieve the robustness of the fault isolation scheme. In general, there is a trade-off between robustness and sensitivity of fault isolation. Improving the robustness to modeling uncertainties may cause the fault isolation scheme to be insensitive. Here, T_0 is chosen to be 1.2 s via a priori knowledge on modeling errors, and the isolation delay in our simulation is less than 2.5 s.

5. Conclusion

A fault diagnosis method that interfaces fuzzy models with an expert supervisory scheme has been presented. This method can recognize the current fault situation of the plant by comparing a bank of nonlinear analytical models formed by Takagi–Sugeno fuzzy systems. Under the supervision of an expert system, the proper bank of models are applied to generate residuals which are analyzed on-line to indicate the occurrence of faults and identify them. Both robustness and fault sensitivity are analyzed and the XTE46 engine simulation example is used to show the reliability of the method.

The effectiveness of the proposed fault diagnosis method has been demonstrated via the CLM simulation of the XTE46 engine. Unlike the typical engine models that are used in some of the literature, this XTE46 simulator has been developed by GEAE to be very complicated and accurate (through thermodynamic simulation for each engine component) so that the simulation conducted on this simulator is very close to that on the real engine for actual flights.

Nonlinear modeling is important to model-based fault diagnosis for nonlinear systems. Having a hierarchical modeling structure and interpolating among nonlinear local models helps to model a complex system in a manageable way. A similar hierarchical structure can also be used to organize the multiple models for fault diagnosis so as to scale up for the cases where a large number of faults are involved.

Acknowledgements

We would like to thank S. Adibhatla at General Electric Aircraft Engines for his assistance and advice on many aspects of this project. Moreover, we would like to thank T.H. Guo at NASA Glenn Research Center for his support.

References

Adibhatla, S., & Lewis, T. (1997). Model-based intelligent digital engine control. In *AIAA-97-3192, 33rd joint propulsion conference*.

Antsaklis, P. J., & Passino, K. M. (Eds.), (1993). *An introduction to intelligent and autonomous control*. Norwell, MA: Kluwer Academic Publishers.

Boskovic, J., Li, S.-M., & Mehra, R. (2001). On-line failure detection and identification (fdi) and adaptive reconfigurable control (arc) in aerospace applications. In *Proceedings of the American control conference*, Arlington, VA.

Chen, J., & Patton, R. J. (Eds.), (1999). *Robust model-based fault diagnosis for dynamic systems*. Norwell, Massachusetts: Kluwer Academic Publishers.

Chen, J., & Zhang, H. (1991). Robust detection of faulty actuators via unknown input observers. *International Journal of Systems Science*, *22*(10), 1829–1839.

Demetriou, M., & Polycarpou, M. (1998). Incipient fault diagnosis of dynamical systems using on-line approximators. *IEEE Transactions on Automatic Control*, *43*(11), 1612–1617.

Duyar, A., Eldem, V., Merrill, W. C., & Guo, T. (1994). Fault detection and diagnosis in propulsion systems—a fault parameter estimation approach. *Journal of Guidance, Control and Dynamics*, *17*(1), 104–108.

Duyar, A., & Merrill, W. C. (1992). Fault diagnosis for the space shuttle main engine. *Journal of Guidance, Control and Dynamics*, *15*(2), 384–389.

Emani-Naeini, A., Athter, M., & Rock, S. (1988). Effect of model uncertainty on failure detection: The threshold selector. *IEEE Transactions on Automatic Control*, *33*(12), 1106–1115.

Frank, P. (1994). Online fault detection in uncertain nonlinear systems using diagnostic observers—a survey. *International Journal of Systems Science*, *25*(12), 2129–2154.

Frank, P., & Ding, X. (1994). Frequency domain approach to optimally robust residual generation and evaluation for model based fault diagnosis. *Automatica*, *30*(4), 789–804.

Frank, P., & Ding, X. (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control*, *7*(6), 403–424.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automatica*, *26*, 459–474.

Fu, Q., Shen, Y., Zhang, J. Q., & Liu, S. (2001). An approach to fault diagnosis based on a hierarchical information fusion scheme [and turbine application]. In *Proceedings of the 18th IEEE instrumentation and measurement technology conference*, Budapest, Hungary (pp. 875–878).

Fussel, D., Balle, P., & Isermann, R. (1997). Closed loop fault diagnosis based on a nonlinear process model and automatic fuzzy rule generation. In *IFAC fault detection, supervision and safety for technical processes*, Kingston Upon Hull, UK (pp. 349–354).

Garcia, E., & Frank, P. (1997). Deterministic nonlinear observer-based approaches to fault diagnosis: A survey. *Control Engineering Practice*, *5*(5), 663–670.

Gertler, J. (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*, *8*(6), 3–11.

Gertler, J., Costin, M., Fang, X., Hira, R., Kowalczuk, Z., & Luo, Q. (1993). Model-based on-board fault detection and diagnosis for automotive engines. *Control Engineering Practice*, *1*(1), 3–17.

Gertler, J., & Luo, Q. (1989). Robust isolable models for failure diagnosis. *American Institute of Chemical Engineers Journal*, *35*(11), 1856–1868.

Gertler, J., & Singer, D. (1990). A new structural framework for parity equation based failure detection and isolation. *Automatica*, *26*(2), 381–388.

Gremling, J., & Passino, K. M. (1997). Genetic adaptive failure estimation. In *Proceedings of the American control conference*, Albuquerque, NM (pp. 908–912), June.

Hou, M., & Muller, P. (1994). Fault detection and isolation observers. *International Journal of Control*, *60*(5), 827–846.

Kavuri, S., & Venkatasubramanian, V. (1994). Neural network decomposition strategies for large scale fault diagnosis. *International Journal of Control*, *59*(3), 767–792.

Krishnaswami, V., Luh, G., & Rizzoni, G. (1995). Nonlinear parity equation based residual generation for diagnosis of automotive engine faults. *Control Engineering Practice*, *3*(10), 1385–1392.

Laukonen, E. G., & Passino, K. M. (1995). Training fuzzy systems to perform estimation and identification. *Engineering Applications of Artificial Intelligence*, *8*(5), 499–514.

Laukonen, E. G., Passino, K. M., Krishnaswami, V., Luh, G.-C., & Rizzoni, G. (1995). Fault detection and isolation for an experimental internal combustion engine via fuzzy identification. *IEEE Transactions on Control Systems Technology*, *3*, 347–355 September.

Lopez, C. J., Benkhedda, H., & Patton, R. J. (1997). Fuzzy observers for FDI: Application to bilinear systems. In *IFAC fault detection, supervision and safety for technical processes*, Kingston Upon Hull, UK (pp. 1147–1151).

Maki, Y., & Loparo, K. (1997). A neural network approach to fault detection and diagnosis in industrial processes. *IEEE Transactions on Control Systems Technology*, *5*(6), 529–541.

Menke, T., & Maybeck, P. (1995). Sensor/actuator failure-detection in the vista F-16 by multiple model adaptive estimation. *IEEE Transactions on Aerospace an Electronic Systems*, *31*(4), 1218–1229.

Merrill, W. C. (1985). Sensor failure detection for jet engines using analytical redundancy. *Journal of Guidance*, *Control and Dynamics*, *8*(6), 673–682.

Merrill, W. C., DeLaat, J., & Bruton, W. (1988). Advanced detection, isolation, and accommodation of sensor failures—a real-time evaluation. *Journal of Guidance*, *Control and Dynamics*, *11*(6), 517–526.

Merrill, W. C., DeLaat, J. C., & Abdelwahab, M. (1991). Turbofan engine demonstration of sensor failure-detection. *Journal of Guidance, Control and Dynamics*, *14*(2), 337–349.

Naidu, S., Zafiriou, E., & McAvoy, T. (1990). Use of neural networks for failure detection in a control system. *IEEE Control Systems Magazine*, *10*, 49–55.

Passino, K. M., & Antsaklis, P. J. (1988). Fault detection and identification in an intelligent restructurable controller. *Journal of Intelligent and Robotic Systems*, *1*, 145–161.

Passino, K. M., & Yurkovich, S. (Eds.), (1998). *Fuzzy control*. Menlo Park, CA: Addison-Wesley Longman.

Patton, R., & Chen, J. (1991). Robust fault detection using eigenstructure assignment: A tutorial consideration and some new results. In *Proceeding of the 30th IEEE conference on decision and control*, Brighton, UK (pp. 2242–2247).

Patton, R., & Chen, J. (1992). Robust fault detection of jet engine sensor systems using eigenstructure assignment. *Journal of Guidance, Control and Dynamics*, *15*(6), 1491–1497.

Patton, R., & Chen, J. (1997). Modelling methods for improving robustness in fault diagnosis of jet engine system. In *Proceeding of the 31st IEEE conference on control and decision*, Tucson, Arizona (pp. 2330–2335).

Patton, R., Chen, J., & Zhang, H. Y. (1997). Observer-based fault detection and isolation: Robustness and applications. *Control Engineering Practice*, *5*(5), 671–682.

Patton, R., Toribio, C. L., & Simani, S. (2001). Robust fault diagnosis in a chemical process using multiple-model approach. In *Proceedings of the 40th IEEE conference on decision and control*, Orlando, FL, (pp. 149–154).

Patton, R. J. (1997). Fault-tolerant control: The 1997 situation. In *IFAC fault detection, supervision and safety for technical processes*, Kingston Upon Hull, UK (pp. 1029–1051).

Patton, R. J., & Chen, J. (1997a). Observer-based fault detection and isolation: Robustness and applications. *Control Engineering Practice*, *5*(5), 671–682.

Polycarpou, M., & Helmicki, A. (1995). Automated fault detection and accommodation—a learning systems approach. *IEEE Transactions on Systems, Man, and Cybernetics*, *25*(11), 1447–1458.

Polycarpou, M., & Vemuri, A. (1995). Learning methodology for failure detection and accommodation. *IEEE Control Systems Magazine*, *15*(3), 16–24.

Seliger, R., & Frank, P. M. (1991). Fault diagnosis by disturbance decoupled nonlinear observers. In *Proceedings of the 30th IEEE conference on decision and control*, Brighton, UK (pp. 2248–2253).

Sorsa, T., & Koivo, H. (1993). Application of artificial neural networks in process fault diagnosis. *Automatica*, *29*(4), 843–849.

Sorsa, T., Koivo, H., & Koivisto, H. (1991). Neural networks in process fault diagnosis. *IEEE Transactions on Systems, Man and Cybernetics*, *21*(4), 815–825.

Tu, F., Pattipati, K., Deb, S., & Malepati, V. (2003). Computationally efficient algorithms for multiple fault diagnosis in large graph-based systems. *IEEE Transactions on Systems, Man and Cybernetics*, *33*(1), 73–85.

Tylee, J. (1983). On-line failure detection in nuclear power plant instrumentation. *IEEE Transactions on Automatic Control*, *28*(3), 406–415.

Vemuri, A., & Polycarpou, M. (1997a). Robust nonlinear fault diagnosis in input–output systems. *International Journal of Control*, *68*(2), 343–360.

Vemuri, A., & Polycarpou, M. (1997b). Neural network based robust fault diagnosis in robotic systems. *IEEE Transactions on Neural Networks*, *8*(6), 1410–1420.

Vemuri, A., Polycarpou, M., & Diakourtis, S. (1998). Neural network based fault detection and accommodation in robotic manipulators. *IEEE Transactions on Robotics and Automation*, *14*(2), 342–348.

Wang, H., Kropholler, H., & Daley, S. (1993). Robust observer based FDI and its application to the monitoring of a distillation column. *Transactions of the Institute of Measurement and Control*, *15*(5), 221–227.

Watanabe, K., Matsuura, I., Abe, M., Kubota, M., & Himmelblau, D. (1989). Incipient fault diagnosis of chemical processes via artificial neural networks. *American Institute of Chemical Engineers Journal*, *35*(11), 1803–1812.

Yen, G., & Ho, L.-W. (2001). On-line multiple-model based fault diagnosis and accommodation. In *Proceedings of the IEEE international symposium on intelligent control*, Mexico City, Mexico (pp. 73–78).