Model Acceptability Measure for the Identification of Failures in Qualitative Fault Monitoring Systems

Antoni Escobet
Departament ESAII
Univ. Pol. de Catalunya
Edifici MN2 - Campus de Manresa
Av. Bases de Manresa 61-73
Manresa 08240, Spain
Phone: +34(93)877-7260
Fax: +34(93)877-7202
Toni@Bages.EUPM.UPC.ES

Angela Nebot
Departament LSI
Univ. Pol. de Catalunya
Mòdul C6 - Campus Nord
Jordi Girona Salgado, 1-3
Barcelona 08034, Spain
Phone: +34(93)401-5642
Fax: +34(93)401-7014
Angela@LSI.UPC.ES

François E. Cellier
Elect. & Comp. Engr. Dept.
University of Arizona
P.O.Box 210104
Tucson, AZ 85721-0104
U.S.A.
Phone: +1(520)621-6192
Fax: +1(520)621-8076
Cellier@ECE.Arizona.Edu

ABSTRACT

This paper deals with two of the main tasks of Fault Monitoring Systems (FMS): *fault detection* and *fault identification.* During fault detection, the FMS should recognize that the plant behavior is abnormal, and therefore, that the plant is not working properly. During fault identification, the FMS should conclude which type of failure has occurred. The first goal of this work is to consolidate a new fault detection technique, called *enveloping*, that was developed in the context of the Fuzzy Inductive Reasoning Fault Monitoring System (FIRFMS). The second and primary goal of this paper is to introduce the *model acceptability measure* as a tool to enhance and make more robust the fault identification process in the context of FIRFMS. The enveloping technique and the model acceptability measure are applied to an electric circuit model previously used for such purpose in the literature. It is shown that the new methods outperform the ones previously advocated in FIRFMS for that purpose[1].

Keywords: Fuzzy Systems, Inductive Reasoning, Fault Monitoring Systems.

INTRODUCTION

There has been a growing demand for fault monitoring systems (FMS) in recent years due to the increased complexity of modern engineering systems that are being governed by ever more complex and sophisticated control architectures. The increase in complexity entails an increase in the number of possible faults and the frequency of occurrence of these faults. In consequence, the demanded functionality of FMS have also grown over time. At present, a decent FMS needs to be able to at least detect, identify, and explain the different faults that may occur in the system through time.

There exists an intensive research activity in this area that includes quantitative as well as qualitative approaches. Quantitative approaches are primarily based on statistical techniques, first order logic, control theory, mathematical modeling, and computer simulation [Pau, 1981; Basseville and Nikiforov, 1993; Patton, 1989; Kumamaru *et al.*, 1984]. The main drawback of quantitative techniques is that they operate on a quantitative and precisely formulated plant model that is not always available. Also, human plant operators usually rely on heuristic knowledge that is easy to be captured by means of qualitative methodologies. There is a large amount of research done in the area of qualitative FMS, primarily making use of expert systems and neural networks [Boullart *et al.*, 1992; Crespo, 1993; Kandel, 1992; Miller *et al.*, 1990].

In this paper, the qualitative Fuzzy Inductive Reasoning (FIR) methodology has been chosen in order to introduce new approaches for the detection and identification of system faults. FIR was first used as a FMS methodology by [Albornoz, 1996]. Since then, the detection and identification phases have been enhanced, as will be shown in the present paper. In this section, a brief introduction of the FIR methodology follows. Also, a schematic explanation of the detection and identification processes as they had been introduced by Albornoz is provided.

[1] The research presented in this paper was supported by the TAP-96-0882 CICYT project

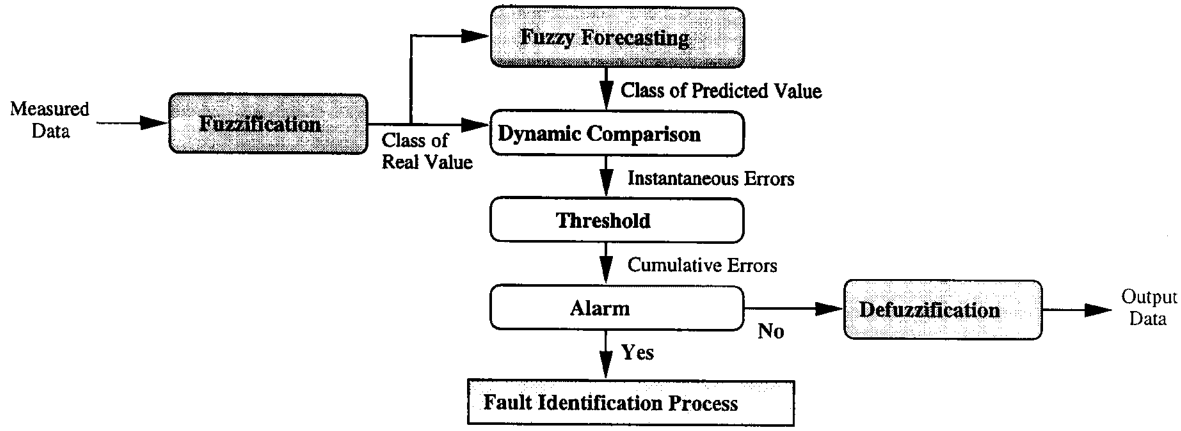

Figure 1: FIR Fault Detection Process

The new *enveloping* technique as well as the *acceptability measure*, that allow to reduce the time to failure detection and to increase the FMS robustness, are introduced in subsequent sections. Finally, an example of an electric circuit is presented to show the feasibility of the new approach.

Fuzzy Inductive Reasoning

The Fuzzy Inductive Reasoning methodology emerged from the General Systems Problem Solving (GSPS) methodology developed by Klir [Klir, 1985] and is composed of four main processes, namely: *fuzzification*, *qualitative model identification*, *fuzzy forecasting*, and *defuzzification*.

In the FIR fuzzification process, quantitative variables are fuzzified (discretized) into a fuzzy triple, consisting of a class, a membership, and a side value. The side function gives information about the position of the quantitative value with respect to the maximum of the membership function (left, middle, right) of the chosen class. It is important to note that the same information is contained in the qualitative triple as in the original quantitative value, hence no information is lost in the fuzzification process.

FIR is fed with data measured from the system under study, converted to fuzzy information by means of the previously described fuzzification function. The qualitative model identification process of the FIR methodology is responsible of finding spatial and temporal causal relations between variables and, therefore, of obtaining the best qualitative model that represents the system. A FIR model is composed by a so-called *mask* and the *behavior matrix*. The mask represents the structure of the model, whereas the behavior matrix is the associated "rule base."

The qualitative model identification process evaluates the possible masks and concludes which of them offers the highest quality from the point of view of an entropy reduction measure.

Once the best mask has been identified, it can be applied to the qualitative data obtained from the system resulting in a particular rule base (behavior matrix). Once the rule base and the mask are available, the system's prediction can take place using the FIR inference engine. This process is called fuzzy forecasting. The FIR inference engine is a specialization of the k-nearest neighbor rule, commonly used in the pattern recognition field. This technique computes a distance measure between the input pattern, for which the output prediction should be obtained, and all patterns stored in the behavior matrix that match (with regard to the class value) the input pattern. The predicted output is then computed as a weighted mean of the outputs associated with the k nearest neighbors, i.e., those neighbors that exhibit the smallest distance measure in the input space.

Defuzzification is the inverse process of fuzzification. It allows to convert the qualitative predicted output (a qualitative triple) into a quantitative variable that can then be used as input to an external quantitative model. For a deeper insight of the FIR methodology cf. [Cellier *et al.*, 1996].

Fault Detection

The fault detection process in the context of the FIR

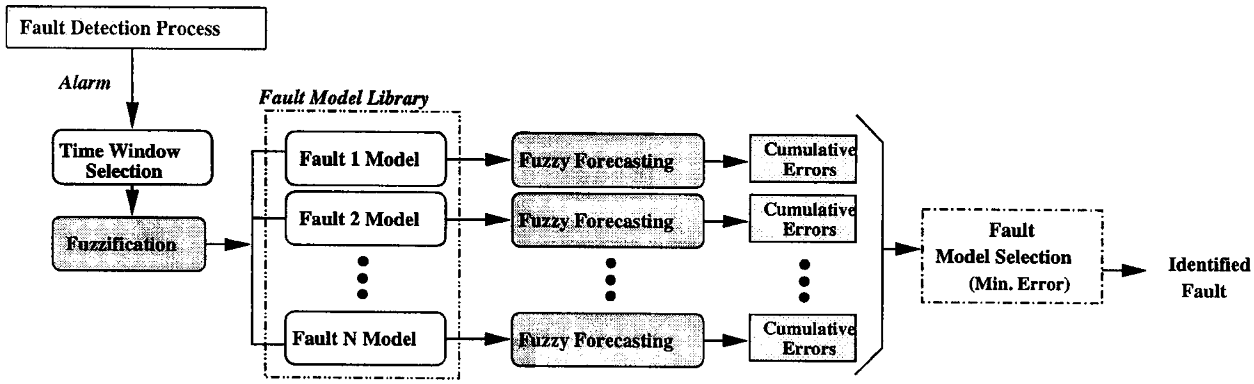

Figure 2: FIR Fault Identification Process

methodology is described in figure 1. In figure 1, the dark boxes represent FIR processes, whereas the white boxes constitute the fault detection procedure. The data measured from the system is converted to qualitative triples (class, membership, side) by means of the FIR fuzzification process. The fuzzy forecasting process predicts the next output value, also a qualitative triple, from the qualitative data using the model (mask) that represents the current behavior of the system. The detection procedure operates as follows:

- The class of the predicted output value is compared with the class of the real output value. The comparison is done by subtracting the class of the real value from that of the predicted value. The result of this subtraction is called *instantaneous error*.
- The instantaneous errors are stored in a matrix to which an error filter is applied. The error filter accumulates instantaneous errors over a movable time window with a predefined size. This window is shifted over the error matrix, and the instantaneous errors found within that time window are summed up to form the *accumulated error*.
- A threshold is specified by the modeler. When the accumulated error is greater than the threshold, an alarm is triggered, and it is then necessary to identify the fault that has occurred.

Fault Identification

Once a fault has been detected, it is necessary to identify it. This is accomplished in the FIR methodology by means of the fault identification process presented in figure 2. The fault identification procedure operates as follows:

- Once the alarm has been triggered because an abnormal behavior has been detected, a second time window is selected. The size of this time window defines the number of prediction values that will be used in order to identify the fault that has been produced. Therefore, the time window guides the prediction during the identification process. A narrow time window is desired because it implies a fast model identification.
- For each fault model stored in the fault model library, a prediction of future system behavior is made using the FIR fuzzy forecasting process for the duration indicated by the chosen time window.
- The prediction errors produced during each of the forecasting processes are accumulated. Therefore, each fault model stored in the library has associated a cumulative error. The model with the lowest cumulative error is selected as the one that best represents the new behavior of the system and, consequently, the detected fault has been identified.

The FMS approach described above was developed by [Albornoz, 96]. The *enveloping* technique and its associated *acceptability measure* offer an improvement of the fault detection and identification processes described above. In the next section, these two new concepts are introduced and explained in more detail.

ENVELOPING

The idea of using the enveloping concept in fault detection emerged from the fact that, in the previously used approach [Albornoz, 96], only a part of the available information, namely that contained in the class values, was being used, whereas the membership and

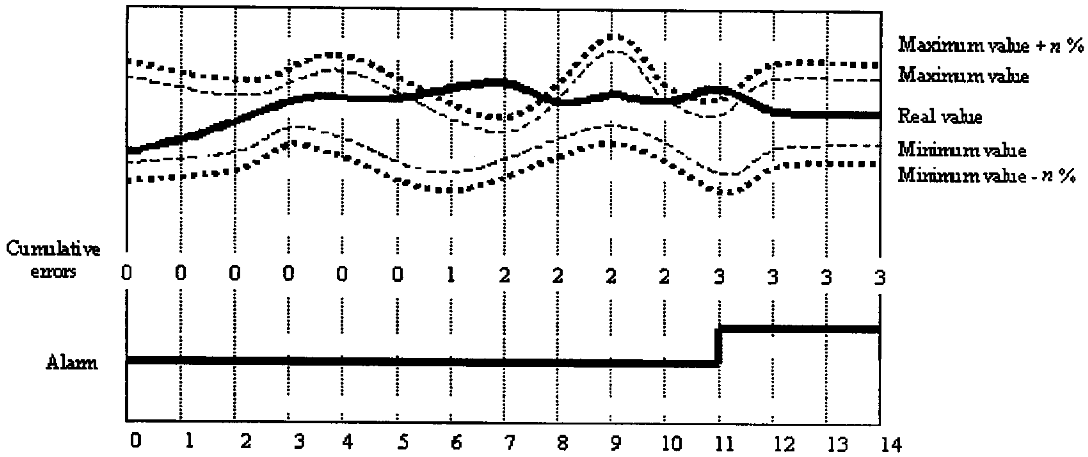

Figure 3: Example of FIR fault detection using the *enveloping* method (Time window=15)

side values were ignored. The decision that a fault has occurred was exclusively based on the class values. The enveloping concept is based on the *5*-nearest neighbors (quantitative values) that are computed inside the FIR inference machine by means of the *k*-nearest neighbor rule. As has been explained in the FIR section, a distance measure is computed between the input pattern, for which the output prediction should be obtained, and all patterns stored in the behavior matrix that match that input pattern. The five patterns that exhibit the shortest distance are selected as the *5*-nearest neighbors.

As shown in figure 3, the envelope is composed of an upper bound (maximum value) and a lower bound (minimum value) that delimit the area in which the real output signal is expected to lie. The envelope is computed by using the five nearest neighbors separately in predictions, i.e., without averaging. If the observed (real) value leaves the bounds specified by the envelope, an instantaneous error is recorded, meaning that the model used in the prediction no longer represents the system at that specific point in time. As in the fault detection procedure described in the previous section, the instantaneous errors recorded inside a predetermined time window are accumulated. When the cumulative error surpasses the specified threshold, an alarm is triggered, and it is then necessary to identify the fault that has occurred.

A narrow enveloping interval implies that the five neighbors are close to each other, meaning that the information available of the behavior of the system at that point in time is rich. In contrast, a wide enveloping interval means that there is not a lot of information available about the system at that point in time, and therefore, the nearest neighbors are far away from each other. The maximum and minimum values that determine the bounds of the envelope can be adjusted by means of an additional fudge factor, n, that represents the model prediction error. This factor is obtained by using the model in the prediction of known values that have been used in the model identification process. The fudge factor is added to the upper bound and subtracted from the lower bound of the envelope as shown in figure 3. The n value provides an idea of the goodness of the model relative to the goodness of the data available to identify it. The *goodness of the model* refers to the prediction error, whereas the *goodness of the data* refers to the richness of the data available for identification. The fudge factor adds flexibility to the enveloping concept by incorporating known information related to the quality of the data observed from the system. It is important to keep in mind that the FIR methodology is driven by the system's behavior rather than relying on structural knowledge, and therefore, the amount and richness of the data available from the system are crucial in order to assure the identification of an accurate and reliable model that represents it.

Figure 3 presents an example of FIR fault detection using the enveloping concept with a time window of 15 prediction points. The upper and lower dotted lines represent the upper and lower bounds of the envelope, respectively, whereas the continuous line is the real output signal. In the bottom part of the figure, the instantaneous errors are accumulated. As can be seen, the real value leaves the envelope for the first time at point number 6 where the observed value exceeds the upper bound of the envelope, causing an instantaneous error. The same occurs at points number 7 and 11. The threshold of accumulated errors specified in this example was 3, and therefore, the alarm is triggered in point number 11 when the third instantaneous error arrives. The next step is the identification of the fault that occurred and has been detected.

ACCEPTABILITY MEASURE

The identification process can have additional problems that have not been mentioned before, e.g., when the produced fault is not a foreseen fault and, therefore, is not available in the fault library, or when different faults have common traits that make it difficult to distinguish between them. The acceptability measure helps to deal with such problems. The acceptability measure is a metric that quantifies the relative suitability of the candidate models, i.e., the models contained in the fault library. This metric allows to determine when an observed fault is not available in the fault library, and it indicates to the user if multiple fault models can reasonably explain an observed fault.

It is possible to specify an acceptability measure of the i_{th} model by means of the following formula:

$$C_i = 1.0 - \frac{I_{a_i}}{I_{a_{max}}}$$

i.e., the acceptability measure of the i_{th} models, C_i, can be computed by use of the sum of instantaneous errors (the so-called *alarm indicators*) for that particular model, I_{a_i}, and the maximum number of alarm indicators possible (depends on the size of the time window), $I_{a_{max}}$. C_i is a *confidence measure*, i.e., a real-valued number in the range $[0, 1]$, where larger values indicate increased confidence.

Unfortunately, the proposed formula is not good enough, because it could happen that there are two almost perfect models, i.e., two separate models with C_i and C_j almost at 1.0. In the acceptability measure, it is important to take the dispersion among the C_i values into account. This can be accomplished by computing a relative confidence:

$$C_{rel_i} = \frac{C_i}{\sum_{k=1}^{5} C_k}$$

Now, if there is only one model with a high value of C_i, C_{rel_i} will still be very high, but if there are other models with high C_i values as well, C_{rel_i} will be much smaller.

Unfortunately, even this formula is not yet good enough. It could happen that there is no C_i with a large value, only one with a small value, while all others are zero. In this case, C_{rel_i} will be undeservedly high. Therefore, the following final formula is proposed as the acceptability measure that captures the quality of the model selection:

$$Q_i = C_i \cdot C_{rel_i}$$

The usefulness of the acceptability measure and the enveloping method is presented in the next section by means of their application to an electric circuit.

ELECTRIC CIRCUIT

The application that is presented in this section is an electric circuit that contains three binary switches that allow to define eight different structures depending on the switch positions. It can be considered that in a natural working regime all the switches are closed, and this structure is represented with the binary relation 000 ($SW_3 = 0$, $SW_2 = 0$ and $SW_1 = 0$). The electric circuit is described in figure 4.

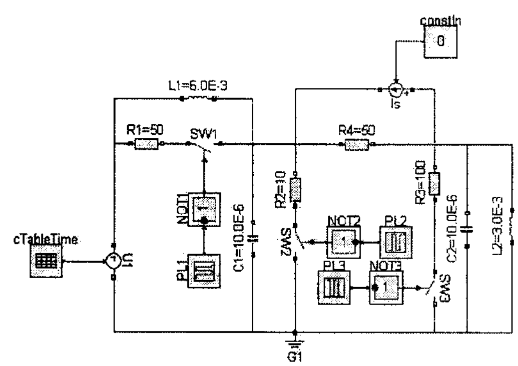

Figure 4: Electric circuit

The input signal of the circuit, U_1, produces a binary aleatory signal with values in the range $[0, 0.001]$ volts. The output of the circuit is the voltage measured across the resistance R_4. The quantitative model of the circuit has been built using Dymola [Elmqvist, 1995], and the continuous system simulation language ACSL is used for the simulation of the model.

The main goal of this application is to show the viability of the FIRFMS approach to detecting and identi-

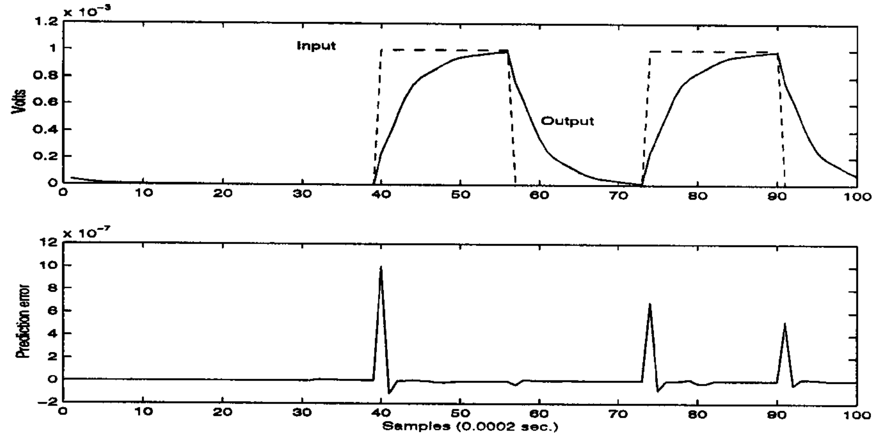

Figure 5: Identification of the model for circuit 000

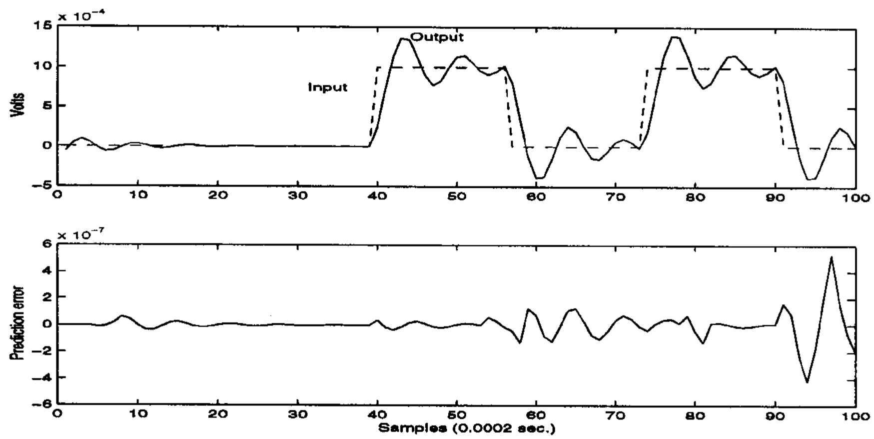

Figure 6: Identification of the model for circuit 011

fying structural changes in the circuit, i.e., the *qualitative circuit state* as specified by the status of the three switches. In order to achieve this goal, the first step is to obtain the FIR qualitative models that constitute the *fault library*. In this example, four different structures have been selected: 000, 001, 010 and 011. For each of them, the system is simulated over 0.4 seconds using ACSL with a sampling time of 0.0002 seconds. The first 1000 data points are used by FIR to identify the model and from those, the last 100 are used to verify the model and compute the fudge factor n, as outlined in the previous section. The mean square error in percentage is used to compute n, as described in equation 1.

$$MSE = \frac{E[(y(t) - \hat{y}(t))^2]}{y_{\text{var}}} \cdot 100\% \qquad (1)$$

where y_{var} denotes the variance of $y(t)$.

In order to illustrate this process, the prediction errors obtained for circuits 000 and 011 are shown in figures 5 and 6, respectively.

The upper plots of figures 5 and 6 show the input signal (dashed line), as well as the real (solid line) and predicted (dotted line) output signals. In both figures, the real and predicted output signals are indistinguishable, due to the high accuracy of the prediction. This is the reason why the prediction error (real minus predicted value) is shown in the bottom plots of the two figures. The MSE errors computed by means of equation 1 are $1.2 * 10^{-5}\%$, $1.1 * 10^{-5}\%$, $6.3 * 10^{-7}\%$, and $3.4 * 10^{-6}\%$ for circuits 000, 001, 010 and 011, respectively. These MSE errors constitute the n component of the enveloping procedure. The FIR models obtained for the four circuits were also validated with data not

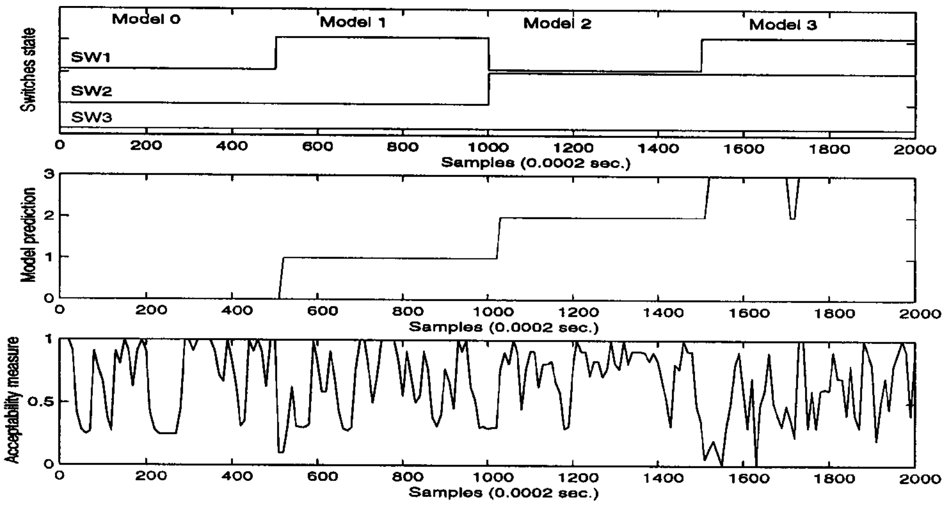

Figure 7: FIRFMS results when applied to the electric circuit using the enveloping technique together with the acceptability measure (TW=10 samples)

used in the identification process obtaining MSE errors almost as small as those previously shown. The FIR models (masks and behavior matrices) of the four circuits are stored in the fault model library and will be used during detection of structural changes occurring in the system, and during identification of the resulting new structures.

The experiment described in figure 7 is used to show the performance of the enveloping technique and model acceptability measure and to compare them with the detection and identification processes that were previously used [Albornoz, 96]. The top plot of this figure describes the structure of the system that corresponds to each one of the four models. Model 0 is equivalent to the circuit that has all three switches set to zero. Model 1 has SW3=0, SW2=0, and SW1=1. Model 2 is characterized by SW3=0, SW2=1, and SW1=0. Finally, model 3 corresponds to SW3=0, SW2=1, and SW1=1.

The top plot of figure 7 defines the experiment. Data observed while the system is operating in the four modes characterized by models 0, 1, 2, and 3 are sampled and stored. FIRFMS has to detect each structural change of the circuit and to identify the model that corresponds to the new structure as accurately as possible.

In the middle plot of figure 7, the FIRFMS results are presented. FIRFMS detects that the first structure corresponds to model 0. Shortly after sample 500, it detects a change of the structure of the system and identifies model 1 as the best model to represent the new structure. FIRFMS detects a new structural change just after sample 1000, identifying model 2 as the best model representing the system during the subsequent period. Finally, at sample 1500, FIRFMS detects the last change on the system's structure, and identifies model 3 as the one that best captures the behavior of the circuit in the final period of the simulation. As shown in the plot, FIRFMS made a mistake in the period associated to model 3. It detected a new structural change and identified model 2 as the new structure, when, in reality, no structural change in the circuit took place. This is due to the fact that during the last simulation segment, several of the models can be used to explain the behavior of the system.

The bottom plot of figure 7 shows the acceptability measure of the winning model. Each of the four models has its own acceptability measure, but only the largest of them (resulting in the selected model of the plot above) is shown. Whereas the acceptability measure varies a lot, on average, it assumes a value of somewhere around 0.6 during the first three simulation segments. During the final segment, the average acceptability measure is lower, somewhere around 0.4. This is because, during that segment, several of the models can reasonably explain the observed behavior. Yet, except for one mistake, FIRFMS picked consistently the correct model.

Notice that often the acceptability measure assumes a value of one. This means that only one model (the one identified by FIR) is able to predict correctly all the data points of the associated time window, whereas

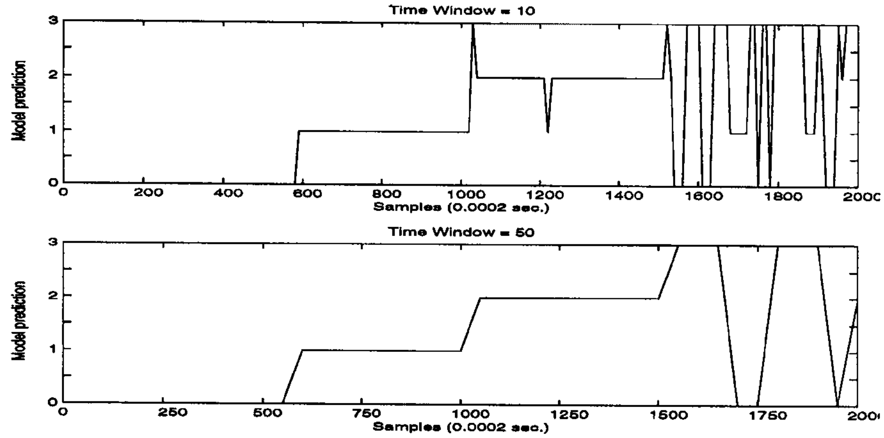

Figure 8: FIRFMS results when applied to the electric circuit using the Albornoz approach (TW=10 and 50 samples)

the other three models are not able to predict any of the data points. Usually, the acceptability measure assumes values between 0.25 and 1, corresponding to the cases where some models predict better than others the specific time window, although almost all of them are able to predict a portion of the data points that compose the associated time window. In these cases, FIR identifies the model with the smallest number of prediction errors. Values of the acceptability measure lower than 0.25 mean that, in general, all the models have problems to predict the time window. Therefore, it would be necessary to determine a threshold in the acceptability measure that indicates that the current behavior cannot reasonably be represented by any model contained in the model library. In our example, only some points corresponding to the last segment of the simulation exhibit this kind of low acceptability measure values, and this is due to the fact that the behavior of model 3 of the system is the one most difficult to identify, as will be shown later.

One final remark relating to the acceptability measure: whenever the real output signal is in steady-state, the acceptability measure assumes small values. This is due to the fact that all four circuits exhibit similar behavior in the vicinity of steady-state, because the gain has been normalized to one. The number of prediction errors obtained in this zone are similarly small for all models, and consequently, the acceptability measure is distributed among all of the models. This kind of situations could cause mistakes in the detection process if steady-state operation were allowed to continue over an extended period of time.

The top plot of figure 8 shows the results obtained when the detection and identification procedures of the Albornoz approach are used in the same experiment. As can be seen in the figure, the fault monitoring system is able to detect and identify correctly the first two models, though it needs more time than the new approach to detect the switch from model 0 to model 1. Yet, the technique is not capable to correctly identify models 2 and 3. During the third simulation period, the FMS makes two mistakes, whereas during the fourth and final segment (corresponding to model 3), the FMS detects a large number of structural changes in the circuit and identifies several models for different periods, making it impossible to know what is happening in the system. The time window chosen in this experiment was of 10 samples, and the threshold specified was of 2 cumulative errors.

The results can be somewhat improved by increasing the time window to 50 samples. The bottom plot of figure 8 shows these results. In this case, the first three models are identified reliably, but the FMS is still not able to identify correctly model 3. Moreover, due to the fact that the time window has been increased, the detection process slows down, and the FMS needs yet more time to recognize that a structural change has occurred in the system.

CONCLUSIONS

In this paper two new concepts, the *enveloping* technique and the *model acceptability measure*, are introduced in the context of the Fuzzy Inductive Reasoning Fault Monitoring System. These concepts allow to improve the fault detection and identification processes of FIRFMS significantly by reducing the time needed to detect a failure and by increasing the robustness of the FMS. The new FIRFMS is able to identify faults that have occurred in the system or its new structure in a more reliably fashion than previously proposed approaches. An example of an electric circuit is presented to show the feasibility of the new concepts and their performance in comparison with the previously used FIRFMS.

REFERENCES

Albornoz A., Inductive reasoning and reconstruction analysis, two complementary tools for qualitative fault monitoring of large-scale systems, Ph.D., Universitat Politècnica de Catalunya, 1996.

Basseville M. and I. Nikiforov, Detection of Abrupt Changes: Theory and Applications, Prentice Hall Information and Systems Science Series, USA, 1993.

Boullart L., A. Krijgsman and R.A. Vingerhoeds, Applications of Artificial Intelligence in Process Control, Pergamon Press, London, U.K., 1992.

Cellier F.E., A. Nebot, F. Mugica and A. De Albornoz, Combined qualitative/quantitative simulation models of continuous-time processes using fuzzy inductive reasoning techniques, Intl. J. General Systems 24 (1–2), pp. 95–116, 1996.

Crespo A., Real-Time Expert Systems, in Applications of Artificial Intelligence in Process Control, (L. Boullart, A. Krijgsman, and R.A. Vingerhoeds Eds.), Pergamon Press, UK, 1993.

Elmqvist, H., Dymola — Dynamic Modeling Language, User's Manual. Dynasim AB, Lund, Sweden, 1995.

Escobet A., A. Nebot and M. Vallverdú, A Qualitative Fault Monitoring System to Support Medical Decisions: An Application to the Central Nervous System, SCB'99: Intern. Conf. on Soft Computing in Biomedicine, Rochester, NY, 1999.

Kandel, Abraham, (Ed.), Fuzzy Expert Systems, CRC Press, USA. 1992.

Klir G.J., Architecture of Systems Problem Solving, Plenum Press, New York, 1985.

Kumamaru, K., S. Sagara and T. Söderström, Some Statistical Methods for Fault Diagnosis for Dynamical Systems, in (R. Patton, P. Frank, R. Clark eds.) Fault Diagnosis in Dynamic Systems, Theory and Applications, Prentice Hall, U.K., pp. 439-475, 1984.

Miller W.T., R.S. Sutton, P.J. Werbos (Eds), Neural Networks for Control, The MIT Press, Cambridge, MA, USA, 1990.

Patton, R.J. and S.M. Kangethe, Robust Fault Diagnosis Using Eigenstructure Assignment of Observers, in (R. Patton, P. Frank, R. Clark eds.) Fault Diagnosis in Dynamic Systems, Theory and Applications, Prentice Hall, U.K., pp. 99-154, 1989.

Pau, L.F., Failure Diagnosis and Performance Monitoring, Marcel Dekker, New York and Basel, 1981.