# Comparative Analysis of Statistical Inference Methods for Fault Detection and Diagnosis Using Nonlinear Models Applied to the Hydraulical Benchmark System

Patricio Cofre[1] and Aldo Cipriano[1]

[1] College of Engineering, Pontificia Universidad Católica de Chile
PO Box 306, Santiago 22, Chile
Phone: 56-2-6864281, fax 56-2-5522563
aciprian@ing.puc.cl

**Abstract.** This work compares three nonlinear models for a three-tank hydraulic system. From each model, nonlinear predictors are designed for the generation of residuals representing the fault conditions. The residuals feed three different statistical inference modules to be compared in search for the best solution for fault detection and diagnosis.

## 1 Introduction

The automatization of industrial processes and the growing complexity of modern control systems have made necessary the development of functions that detect the presence of a fault in an opportune and reliable manner. In this way, a diagnosis is made that allows a quick restitution of the normal process operation.

Figure 1 shows a classic scheme for fault detection and diagnosis. The input variables are fed to the plant and to the models trained under different operating conditions. In this case the conditions correspond to a normal operation, fault operation 1, fault operation 2 and unknown fault. The scenario that best represents the operating condition is the one that gives the output that resembles the most to the real one. In order to decide which is the most representative, an inference module is used that from the residues ($r_0$, $r_1$, $r_2$), determined by the difference between the real output and the models output, generates signals ($f_0$, $f_1$, $f_2$, $f_3$) indicating in which operating scenario is the plant. In this work we consider statistical decision tools.

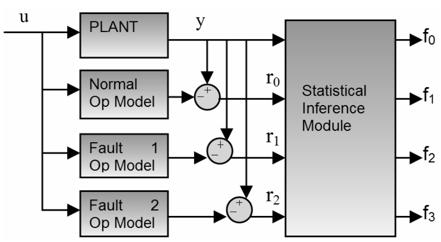


**Fig. 1.** Classical layout of a fault detection and diagnosis system.

On the other hand, the most industrial processes present highly nonlinear behaviors which need to be considered at the moment of identifying models for the generation of residuals. Among the different nonlinear models that can be used for such goals are polynomial models, neural models, and fuzzy models. In this study we consider the following:

- The polynomial Hammerstein model
- The Neuronal model
- The ANFIS Dynamical Fuzzy model

The fault detection and diagnosis based in these residuals face the presence of random disturbances that can cause false alarms. To avoid such effect, it is necessary to pay special attention to the design of the inference modules. In this work, we study and compare three statistical methods used for such goals:

- The filtered derivative method [2]
- The exponential media mobile method [2]
- The Scheffe's test [10]

The paper starts with a review of the work done in nonlinear modeling and statistical inference, followed by the formulation of the modeling and inference methods that we seek to compare. Next, an hydraulic benchmark system simulator is described which is used for the comparison. Finally the results obtained are analyzed and the main conclusions are drawn.

## 2 Nonlinear Modeling for Residual Generation

For residual generation, diverse nonlinear models can be used, the most basic being the Volterra, Hammerstein, and Wiener [4][5] which include powers and products of the input and output data as regressors. The NARX and NARMAX models [3] can also be employed because they allow any input/output nonlinear function as regressors. Other models such us Takagi-Sugeno are based in fuzzy logic and assign a linear submodel to each operating zone of the plant; such zones are determined using the Fuzzy Clustering technique [7]. Another option is to use a feed-forward type neural network in which the regressors include previous inputs and outputs that allow improving the fit. Also, a mixed neural-fuzzy technique called ANFIS [6] has been used that gets a better adjustment thanks to its architecture and hybrid training.

Next, we present the three models that are compared: the Hammerstein polynomial model, the neural model, and the ANFIS (Adaptive Network Based Fuzzy Inference System) model.

### 2.1 The Hammerstein Model

The Hammerstein model is characterized for separating the model in a static nonlinear component represented by *f*, applied to the manipulated variable, and a transference function *G(z),* see Figure 2:

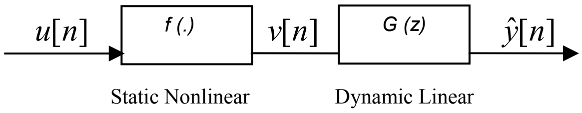


**Fig. 2.** The Hammerstein Model.

The nonlinear function is of the polynomial type so the estimation of the output ( $\hat{y}[n]$ ) is represented by:

$$\hat{y}[n] = \sum_{i=1}^{na} a_i y[n-i] + \sum_{k=0}^{M} \sum_{j=1}^{nb} b_k[j] \cdot u^k[n-j-d] \qquad \textbf{(1)}$$

Where *M* is the order of the polynomial, *na* and *nb* are the input and output regressors window lengths, and *d* is the delay of the transference function. The first sum is a moving average of the measured variable *y*[*n*] in order to improve the fit. The coefficients *a* and *b* are determined by numerical minimization and the number of regressors is selected using the Akaike information criteria that indicates until what point it is significant to add a regressor in order to make a better adjustment [8].

### 2.2 The Neural Model

The neural model allows the existence of different activation functions, such as sigmoids or hyperbolic tangents. For this study we consider a network of two layers, one nonlinear of 10 neurons with sigmoidal activation function and a linear output layer as shown in Figure 3.

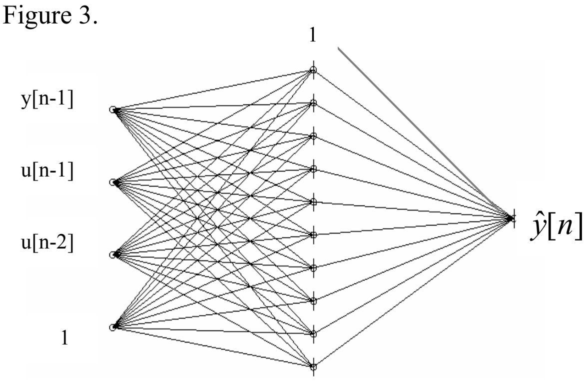


**Fig. 3.** A two layers neural network

The two-layer dynamical estimation of the output $\hat{y}[n]$ for a net is shown in (2), and it is expressed in terms of the interconnection weights that for this comparative study are estimated using the Levenberg-Maruardt algorithm [9]:

$$\hat{y}[n] = \sum_i W_i \cdot f(\sum_j w_{ij} \cdot \phi_j) \qquad \textbf{(2)}$$

Where:

$\phi$ is the vector of regressors: $\phi$ =[ $y$[$n$-$1$] $u$[$n$-$1$] $u$[$n$-$2$] 1]

$w_{ij}$ are the first layer weights

$W_i$ are the second layer weights

$f$ is a nonlinear function, in this case a sigmoid.

### 2.3 The ANFIS Model

The ANFIS (Adaptative Network Based Fuzzy Inference System) structure is shown in Figure 4.

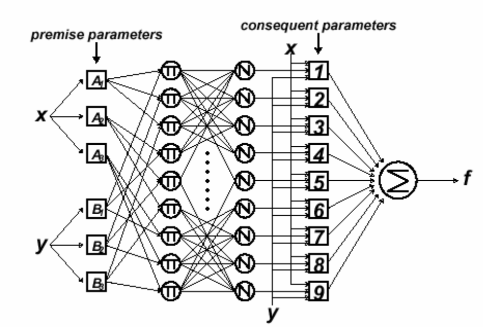


**Fig. 4.** The ANFIS structure

The first layer corresponds to a fuzzy ruleset that directs the signal according its membership to various operation zones; the second layer multiplies the membership degree to each scenario giving the strength with which the fuzzy rule is fulfilled; the third layer normalizes such values, and the fourth assigns a linear model to each operating zone. The last layer is a weighted average of each linear model. The parameters of this model are estimated recursively by a hybrid algorithm which combines the gradient method and least squares methods [6].

The dynamical estimation of the output $\hat{y}[n]$ is given by:

$$\hat{y}[n] = \sum_{i=1}^{na} \sum_{j=1}^{nb} w_{ij} \cdot f_{ij} \qquad \textbf{(3)}$$

$$w_{ij} = \frac{\mu_{Ai}(y[n-1]) \cdot \mu_{Bj}(u[n-1])}{\sum_{k=1}^{na}\sum_{l=1}^{nb} \mu_{Ak}(y[n-1]) \cdot \mu_{Bl}(u[n-1])} \tag{4}$$

$$f_{ij} = p_{ij} \cdot y[n-1] + q_{ij} \cdot u[n-1] + r_{ij} \tag{5}$$

## 3 Statistical Inference Methods for Residual Evaluation

One of the first works that used statistical inference for detection and diagnosis where based on control charts as the Shewhart, mobile media techniques (GMA, EWMA, FMA), and the filtered derivative method [2]. The other works include the CUMSUM and GLR variants and algorithms derived from the Bayes law [2].

These methods are used to decide, from the residuals obtained (7), whether the model represents the operating condition. For this purpose, the model that gives a smaller residual is searched but this decision is not simple because the presence of noise that can cause false alarms. In this work we analyze three known statistical models.

### 3.1 The Filtered Derivative Algorithm

The filtered derivative algorithm consists in filtering the residual and then applying a gradient operator that detects abrupt changes or a threshold in case of progressive changes. The filter can be of any type: exponential with forgetting factor, triangular, or simply of mobile media. It is also possible to use analog filters of the Chevyschev or Butterworth types. The disadvantage of this method is the great number of samples required in order to filter that produces a fault detection delay.

The expression for the filtered residual is given by:

$$g[n] = \sum_{i=0}^{N-1} \gamma[i] \cdot r[n-i] \tag{6}$$

$$\text{With } r[n] = y[n] - \hat{y}[n] \tag{7}$$

The filtered residual correspond to the temporal convolution between $\gamma[n]$ and the residual $r[n]$. The fault is detected when $g$ crosses for the first time a threshold that for this study was fixed in the +/- 3 standard deviations criteria of which is equivalent to a confidence level of 99.73%. The residual $r[n]$ is determined by the difference between the real system output $y[n]$ and the estimation of the output $\hat{y}[n]$.

### 3.2 The EWMA Method

The EWMA (Exponentially Weighted Moving Average) method uses a logarithmic likelihood ratio as a fault detector:

$$g[n] = \sum_{i=0}^{\infty} \gamma[i] \ln\left(\frac{p\theta_1(r[n-i])}{p\theta_0(r[n-i])}\right) \tag{8}$$

Where $\gamma[i]$ is the weight vector of the likelihood function which in this case is of the exponential type, that is

$$\gamma[i] = \alpha(1-\alpha)^i \tag{9}$$

In this form $g[n]$ can be written as a recursion

$$g[n] = (1-\alpha)g[n-1] + \alpha \ln\left(\frac{p\theta_1(r[n])}{p\theta_0(r[n])}\right) \tag{10}$$

Where:

$\alpha$ is the innovation weight
$p\theta_0$ is the probability of getting into a normal operation
$p\theta_1$ is the probability of getting into a fault operation.
$r$ is the residual

We note in (10) and in Figure 5 that if the membership is mainly in the fault regions, the feedback is positive because the quotient in the logarithm is greater than 1 giving a high statistical growth in g; otherwise $g[n]$ will remain around zero. Thus a threshold can certainly be applied because all faults are characterized by a high growth of the indicator, increasing significantly the signal to noise ratio of the statistical with respect to the non-processed residual. In this work the probability density functions (or membership functions seen from a fuzzy logic point of view) were chosen to be gaussian and $\alpha$ is chosen to be 0.05 in order to assure the smoothness of the statistic which constitutes a compromise between detection speed and the possibility of a false alarm when weighting more the innovation.

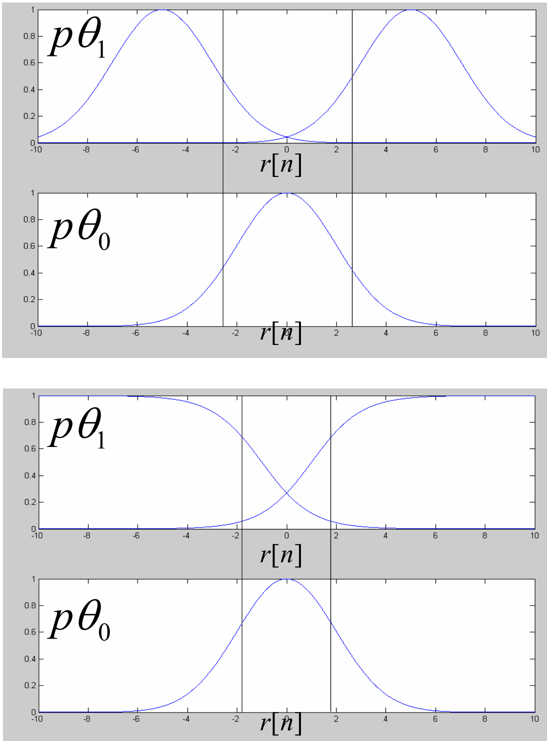


**Fig. 5.** Probability density functions (membership functions) for the EWMA.

### 3.3 The Sheffe´s Test

The Scheffe's test is another option to perform the inference because it can be applied online and provides also a high signal to noise ratio with respect to the residual. The test verifies whether the hypothesis that states the mean of the residual is zero, or, in other words, whether the model represents the real operating condition. In order to apply the Scheffe's test, a statistical with distribution F is defined according (11) and (12):

$$-k_i \le \frac{\bar{r}_i(k)}{\sigma_i(k)/\sqrt{N}} \le k_i, \; i = 1,\ldots, m \tag{11}$$

$$\pm k_i = \pm\sqrt{\frac{(N-1)m}{N-m} F^{\alpha}_{m,N-m}} \tag{12}$$

Where:

| | |
|---|---|
| 1- $\alpha$ | is the confidence level, |
| m | m is the number of variables to be tested, |
| N | is the data window dimension, |
| $\bar{r}_i$ | is the mean of the i-th residual. |
| $\sigma_i$ | is the standard deviation of the i-th residual. |
| $F^{\alpha}_{m,N-m}$ | Is the Fisher distribution, with confidence level. 1- $\alpha$ and m, N-m degrees of freedom. |

## 4 Diagnostic Mechanism

By processing the residuals, each inference method generates a binary number $R_i$ {i=0,1,2} that indicates the degree of representation ($R_i$=0) or not representation ($R_i$=1) of this model for the real operating condition. For the studied case, it is possible to obtain table 1 that indicates the plant operating scenario.

**Table 1.** Truth table for the diagnosis

| $R_0$ | $R_1$ | $R_2$ | Operating Scenario |
|---|---|---|---|
| 0 | 0 | 0 | Unknown Fault |
| 0 | 0 | 1 | Unknown Fault |
| 0 | 1 | 0 | Unknown Fault |
| 0 | 1 | 1 | Normal Operation |
| 1 | 0 | 0 | Unknown Fault |
| 1 | 0 | 1 | Type 1 Fault |
| 1 | 1 | 0 | Type 2 Fault |
| 1 | 1 | 1 | Unknown Fault |

In order to avoid false alarms an extra confirmation is performed which consists in demanding that the operating condition repeat itself in two consecutive samples for it to be considered, otherwise, we assume the change is a product of a disturbance and not a fault.

## 5 Simulator Description

For the comparison of the models and the inference methods, a simulator of the MISO TTS (Three Tank System) hydraulic benchmark system will be used (See figure 6). In this study the level $h_1$ is considered the output of the system.

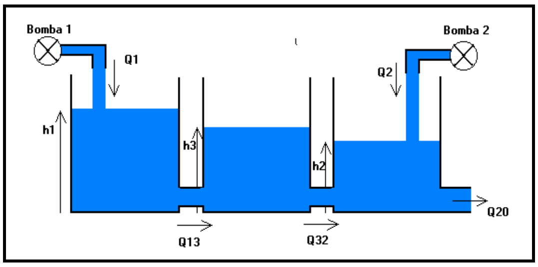


**Fig. 6.** Hydraulic System Benchmark

The system in not linear because it responds to the Bernoulli fluid law that states that the flux between tanks is expressed by:

$$S \cdot sign\,(h_i - h_j) \cdot \sqrt{2g\left|h_i - h_j\right|} \tag{13}$$

Where S is the transversal section of the interconnection tubing. The simulator allows the introduction of various faults from which the following will be considered in this study:

F1: Corresponds to a blocking of the 40% of the connection tubing between tanks 1 and 3.
F2: Corresponds to a blocking of 10% of the outbound conducts.

## 6 Comparison of the Models

In order to compare the different models the RMSE index is adopted as a performance index; the index is expressed by:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(y[n]-\hat{y}[n])^2}{N}} \tag{14}$$

This index is evaluated for all data that allows training the model in normal operating condition and for data that allows validating the obtained model. Table 2 presents the errors obtained in such case.

**Table 2.** Comparison of models using RMSE index

| MODEL | RMSE TRAINING | RMSE VALIDATION |
|---|---|---|
| Hammerstein | $3.624\ 10^{-4}$[m] | $3.366\ 10^{-4}$[m] |
| Neural | $3.317\ 10^{-4}$[m] | $3.469\ 10^{-4}$[m] |
| ANFIS | $3.164\ 10^{-4}$[m] | $3.177\ 10^{-4}$[m] |

# 7 Comparison of the Inference Methods

This section presents the results of the comparison of the three inference methods. Each method was applied to the residuals obtained with all three models and at the same time all these combinations were tested for the three possible scenarios. Table 3 presents the delays in diagnosis obtained by applying the different methods for each model in the three operating conditions.

## 7.1 Inference using Filtered Derivative Algorithm

Figures 7, 8, and 9 present the results obtained by applying the filtered derivative method to the residuals obtained with the Hammerstein, Neural and ANFIS models respectively; this method was performed for the activation of a process fault of type 1 in 1000 seconds.

In all cases it is observed that in the initial seconds, an unknown fault is detected before the normal operation begins; this corresponds to a transient that occurs before the statistics of each inference method reaches a permanent regimen indicating the normal operation.

From the moment the fault occurs until its detection, there exists a false alarm period of unknown or type 2 faults of small duration. In all cases the presence of fault 1 is detected appropriately. It stresses the cleanliness of the results obtained with the Hammerstein model that is also in evidence in table 3.

## 7.2 Inference using EWMA

Figure 10, 11, and 12 show the results obtained for a type 2 fault. As mentioned previously, the inference with the EWMA method requires to adjust the parameter $\alpha$ , which in this work it is fixed at 0.05 based on trial and error with the knowledge that if it is small, we are privileging the certainty over the detection speed.

## 7.3 Inference using Scheffe´s Test

In order to study the application of the Scheffe's test, the system was simulated for normal conditions during all the simulation. Figures 13, 14, and 15 show the results obtained in the simulation. It can be seen that ANFIS performs the fastest detection.

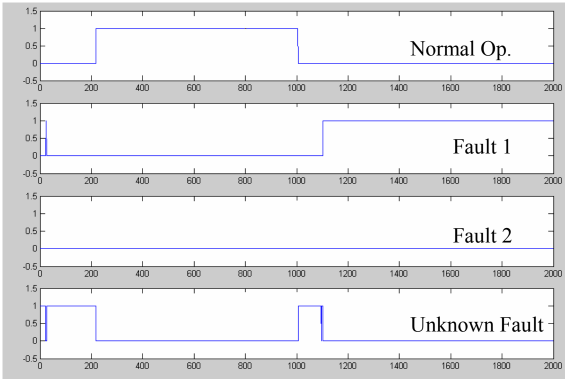


**Fig. 7.** Application of the filtered derivative algorithm using the Hammerstein model for fault 1

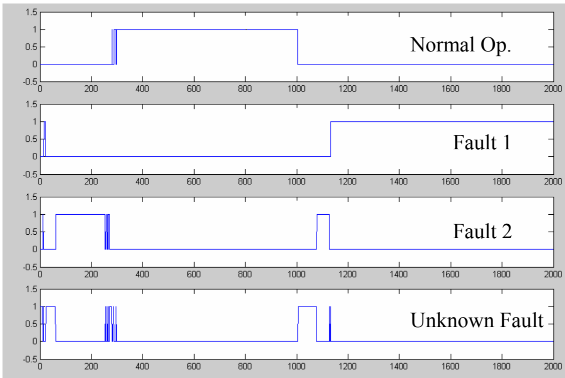


**Fig. 8.** Application of the filtered derivative algorithm using the Neural model for fault 1

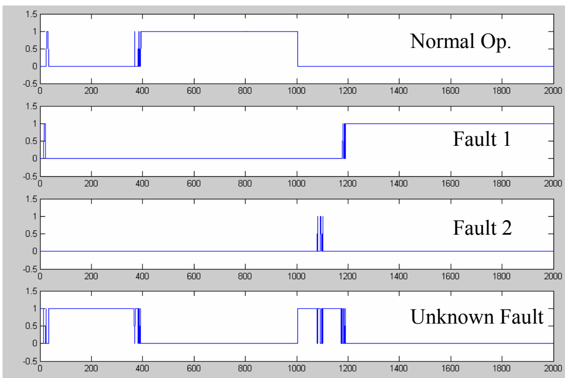


**Fig. 9.** Application of the filtered derivative algorithm using the ANFIS model for fault 1

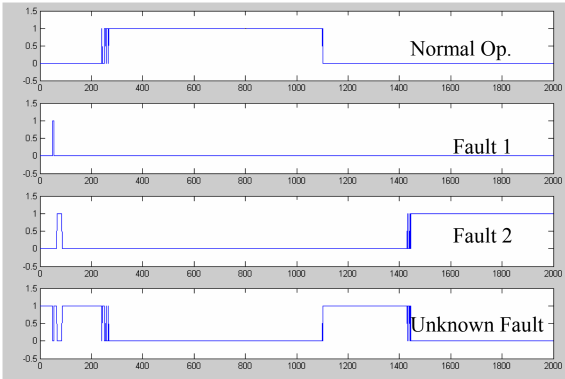


**Fig. 10.** Application of EWMA using Hammerstein model for fault type 2

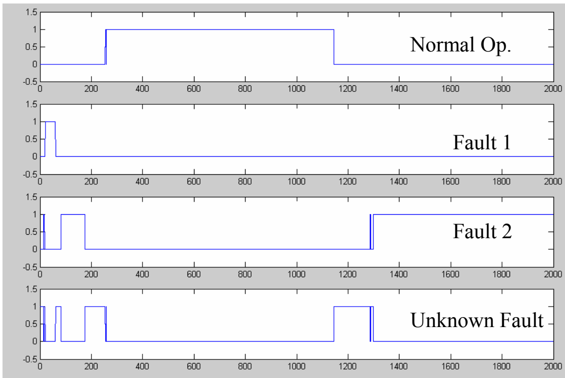


**Fig. 11.** Application of EWMA using Neural model for fault type 2

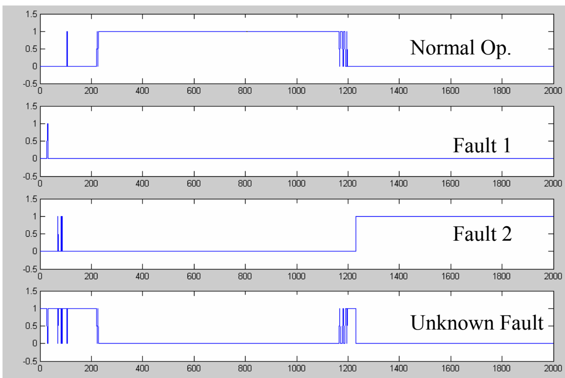


**Fig. 12.** Application of EWMA using ANFIS model for fault type 2

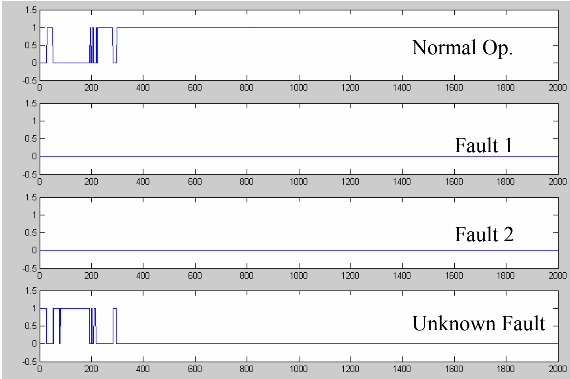


**Fig. 13.** Application of Scheffe´s using Hammerstein model for normal operation.

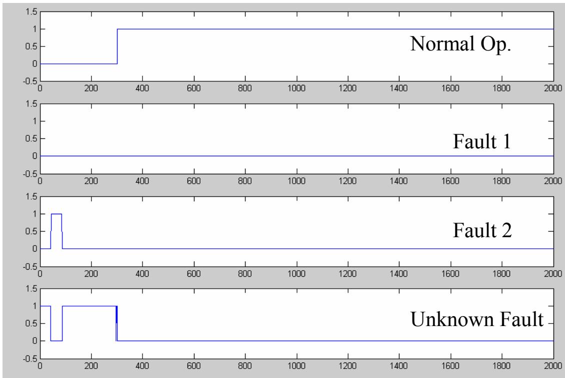


**Fig. 14.** Application of Scheffe´s using Neural model for normal operation.

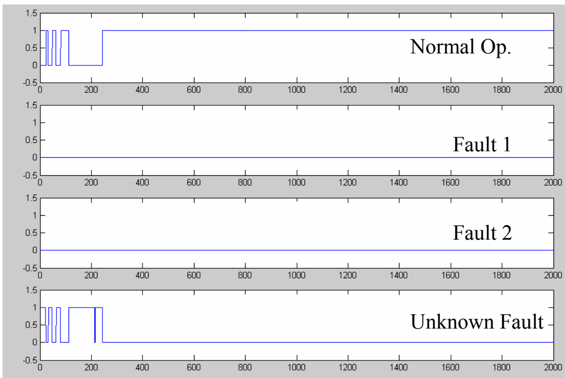


**Fig. 15.** Application of Scheffe´s using ANFIS model for normal operation.

Table 3 presents the detection times obtained for each inference method with the different models and the three operation modes.

**Table 3.** Delay times for each type of operating condition, model and inference method, in seconds

| | | | |
|---|---|---|---|
| Fault 1 | F y D | EWMA | Scheffe |
| HAMMER. | 102.3 | 388.6 | 109.5 |
| NEURONAL | 132.4 | 152.3 | 137.2 |
| ANFIS | 190.2 | 205.4 | 166.5 |
| | | | |
| Fault 2 | F y D | EWMA | Scheffe |
| HAMMER. | 329.2 | 444.5 | 215.1 |
| NEURONAL | 456.2 | 298.1 | 391.3 |
| ANFIS | 197.4 | 230.4 | 245.6 |
| | | | |
| Normal Op | F y D | EWMA | Scheffe |
| HAMMER. | 218.4 | 267.1 | 297.1 |
| NEURONAL | 298.2 | 258.2 | 301.3 |
| ANFIS | 392.2 | 226.1 | 243.3 |

Defining the addition of the detection times for the operating condition as an indicator, table 4 is obtained showing that the Hammerstein model and the Scheffe's test provide the best results.

**Table 4.** Sum of the detection delays for each model and inference method.

| | F y D | EWMA | Scheffe | **TOTAL** |
|---|---|---|---|---|
| HAMMER. | 649.9 | 1100.2 | 621,7 | **2371.8** |
| NEURONAL | 886.8 | 708.6 | 829,8 | **2425.2** |
| ANFIS | 779.8 | 661.9 | 655,4 | **2097.1** |
| **TOTAL** | **2316.5** | **2116.8** | **2106.9** | |

## 8 Conclusions

This paper presents a comparison of models and statistical inference methods for the fault detection and diagnosis using a simulator of three tanks hydraulic system.

Three nonlinear models were evaluated for the generation of residuals. All show similar performance being ANFIS the one that minimizes the training as well as the validation errors for a normal operation simulation.

Also, three inference methods were evaluated which processed the residuals obtained with the models. Scheffe's test was shown to have a favorable detection speed. This characteristic is due to the fact that a small data number is required in order to give a high signal to noise ratio in contrast to the other two methods of medians that use all the history of data in order to give a similar signal to noise ratio in the processed residuals. In all studied cases the statistical detection and the average detection time have been satisfactory and the average detection time has been maintained in a similar range at half the dominant time constant of the system.

The false alarms are due to temporal transits of the inference statistics within the confidence zones of the other faults when the fault has recently occurred; this is the result of the uncertain output that gives a model identified for a given operation condition when the system is in presence of another.

## Acknowledgments

The authors would like to acknowledge to FONDECYT for the financing given to project 1020741, Detection and Fault Diagnostics in Nonlinear Time-varying Systems, and to Dr. Martin Murillo for his help with the translation of this paper.

## References

1. Goycoolea E., Araya J.F., Cipriano A. "Fault detection and diagnosis in the hydraulic benchmark system using nonlinear observer and fuzzy inference". *Proc. of the XV Congress of the Chilean Automatic Control Society*. Santiago 2002, pp 57-63
2. Basseville M., Nikiforov I. "Detection of abrupt changes Theory and application". Prentice Hall.
3. Norgaard M. "Neural network based system identification toolbox", *Technical Report Department of Automation*, Technical University of Denmark 2000.
4. Westwick D., Kearney R., "Identification of a Hammerstein model of the stretch reflex EMG using separable last squares", *Proc. IEEE EMBS 22*, Chicago 2000.
5. Paoli G and Koppl H. "Non linear identification and modeling of large scale analog integrated circuitries for DMT based applications*", Proc. Electronic Circuits and Systems Conferenc*e, Bratislava, 2001.
6. Jang R., "ANFIS: Adaptive-Network-based Fuzzy Inference Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp. 665-685.
7. Babuska R, "Fuzzy Modeling and Identification Toolbox User's Guide" 1998
8. Ljung L. "System Identification Toolbox User's Guide", The *Mathworks*, Inc., 2002
9. Howard D, Mark B, "Neural Network Toolbox User's Guide" The Mathworks, Inc.
10. Gomez E., "Model based fault detection and diagnosis in MIMO systems using statistical methods and Fuzzy Logic", *Ph.D Thesis*, Ruhr Universitat Bochum, 1995.