



Monitoring System for an Experimental Facility Using GMDH Methodology

Iraci Martinez Pereira, Davi Almeida Moraes

IPEN Instituto de Pesquisas Energéticas e Nucleares

iraci.martinez.pereira@gmail.com

ABSTRACT

This work presents a Monitoring System based on the GMDH - Group Method of Data Handling methodology applied in an Experimental Test Facility. GMDH is a combinatorial multi-layer algorithm in which a network of layers and nodes is generated using a number of inputs from the data stream being evaluated. The GMDH method is based on an underlying assumption that the data can be modeled by using an approximation of the Volterra Series or Kolmogorov-Gabor polynomial. The Fault Test Experimental Facility was designed inspired in a PWR nuclear power plant and is composed by elements that correspond to the pressure vessel, steam generator, pumps of the primary and secondary reactor loops. The nuclear reactor core is represented by an electrical heater with different values of power. The experimental plant is fully instrumented with sensors and actuators. The Fault Test Experimental Facility can be operated to generate normal and faulty data. These failures can be added initially with small magnitude, and their magnitude being increasing gradually in a controlled way. The database will interface with the plant supervisory system SCADA (Supervisory Control and Data Acquisition) that provides the data through standard interface.

Keywords: GMDH, monitoring, experimental bench.

1. INTRODUCTION

Group Method of Data Handling was applied in a great variety of areas for data mining and knowledge discovery, forecasting and systems modeling, optimization and pattern recognition. Inductive GMDH algorithms give possibility to find automatically interrelations in data, to select optimal structure of model or network and to increase the accuracy of existing algorithms [1].

This original self-organizing approach is substantially different from deductive methods used commonly for modeling. It has inductive nature - it finds the best solution by sorting-out of possible variants. By sorting of different solutions GMDH algorithms aims to minimize the influence of the author on the results of modeling. Computer itself finds the structure of the model and the laws which act in the system.

In mathematical statistics it is need to have a priori information about the structure of the mathematical model. In neural networks the user estimates this structure by choosing the number of layers and the number and transfer functions of nodes of a neural network. This requires not only knowledge about the theory of neural networks, but also knowledge of the object nature and time. Besides this the knowledge from systems theory about the systems modeled is not applicable without transformation in neural network world. But the rules of translation are unknown. These problems can be overcome by GMDH that can pick out knowledge about object directly from data sampling. The Group Method of Data Handling is the inductive sorting-out method, which has advantages in the cases of rather complex objects, having no definite theory, particularly for the objects with fuzzy characteristics. [1][2]

The purpose of this work is to develop a GMDH - Group Method of Data Handling model applied to an Experimental Test Facility specifically developed to generate normal and faulty data.

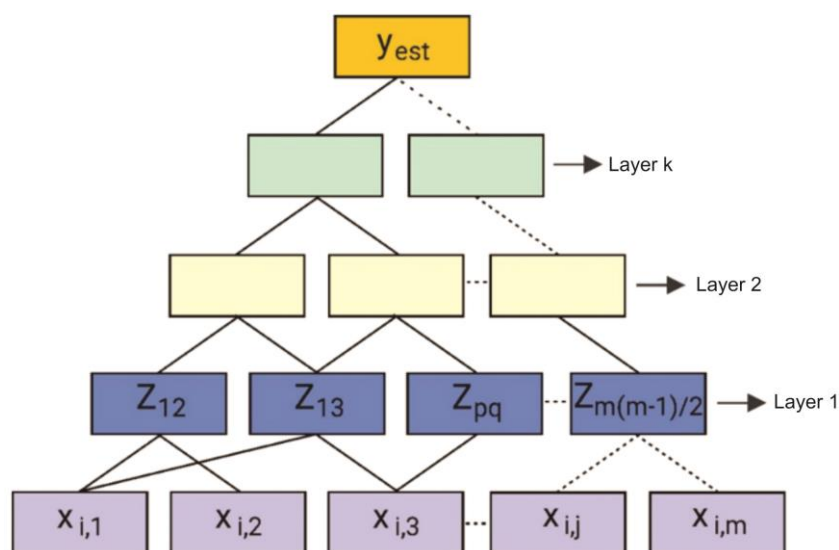
This work is part of a study of the sensitivity and robustness of Fault Detection methods using Artificial Intelligence methodology.

2. GROUP METHOD OF DATA HANDLING - GMDH

The Group Method of Data Handling – GMDH method is composed by an algorithm proposed by Ivakhnenko [3]. The methodology can be considered as a self-organizing algorithm of inductive propagation applied at the solution of many complex practical problems. Moreover, it is possible to get a mathematical model of the process from observation of data samples, which will be used in identification and pattern recognition or even though to describe the process itself.

The network constructed using the GMDH algorithm is an adaptive, supervised learning model. The architecture of a polynomial network is formed during the training process. The node activation function is based on elementary polynomials of arbitrary order. This kind of networks is shown in Figure 1.

Figure 1: Self-organizing GMDH structure with m inputs and k layers



This method solves the multidimensional problem of model improvement by the choice procedure and selection of models chosen from a set of candidate models in accordance with a supplied criterion. The majority GMDH algorithms use reference polynomial functions. A generic connection

between inputs and outputs can be expressed by the series functions of Volterra which is the discrete analogous of the polynomial of Kolmogorov-Gabor [4], equation (1):

$$y = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \dots \quad (1)$$

Where:

{x1, x2, x3 ...}: inputs

{a, b, c...}: polynomials coefficients

y: the node output

The components of input matrix can be changeable independent, functional forms or terms of finite differences, moreover, can be used other nonlinear reference functions. The methods still allow, simultaneously finding the model structure and the output system dependence as a function of the most important inputs system values.

The following procedure is used for a given set of n observations of the m independent variables {x1, x2, ..., xm} and their associated matrix of dependent values {y1, y2, ... yn} [3] [4].

- Subdivide the data into two subsets: one for training and other for testing;
- Compute the regression polynomial using the equation (2), for each pair of input variables xi and xj and the associated output y of the training set which best fits the dependent observations y in the training set. From the observations, m(m-1)/2 regression polynomials will be computed from the observations;

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_i x_j \quad (2)$$

- Evaluate the polynomial for all n observations for each regression. Store these n new observations into a new matrix Z. The other columns of Z are computed in a similar manner. The Z matrix can be interpreted as new improved variables that have better predictability than those of the original generation x1, x2, ..., xm;

- Screening out the last effective variables. The algorithm computes the root mean-square value (regularity criterion – r_j) over the test data set for each column of Z matrix. The regularity criterion is given by the equation (3);

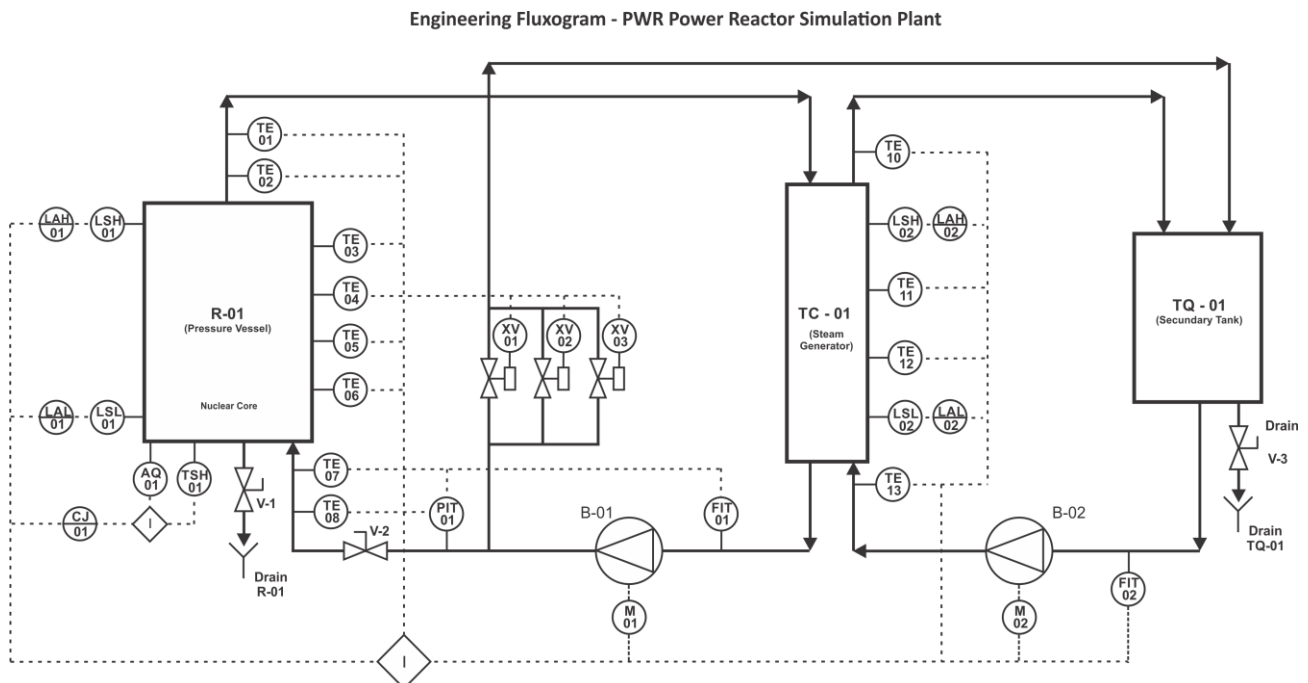
$$r_j^2 = \frac{\sum_{i=1}^{nt} (y_i - z_{ij})^2}{\sum_{i=1}^{nt} y_i^2} \quad (3)$$

- Order the columns of Z according to increasing r_j , and then pick those columns of Z satisfying $r_j < R$ (R is some prescribed value chosen by the user) to replace the original columns of X;
- The above process is repeated and new generations are obtained until the method starts overfitting the data set. One can plot the smallest of the r_j 's computed in each generation and compare it with the smallest r_j 's of the most recent generation start to have an increasing trend.

3. PLANT FOR DATA GENERATION AND ANALYSIS

Using an experimental process plant specially developed for this work, it was possible to carry out the tests of control systems in different situations of normal and faulty operation. [5]

Figure 2 shows the Engineering Flowchart of the experimental process plant. The plant is composed of elements inspired on the behavior of the pressure vessel, steam generator, primary and secondary circuit pumps of a Nuclear Power Plant. The experimental plant was built with actuator technologies, digital networks or connectivity using a Fieldbus system [6], signal processing devices (Programmable Logic Controllers), electric and electronics assembly of all system, as well as the mechanical structure in carbon steel, steel and aluminum.

Figure 2: Engineering flowchart – Experimental plant

The Supervisory System of the plant is a SCADA (Supervisory Control and Data Acquisition) and it makes the data available through a standard interface. SCADA system is a software that interfaces with industrial processes and / or machines of the most varied sectors. [7].

The supervisory system developed has the following functions:

- Recording and temporal analysis of the process variables of pressure, level, temperature and flowrate;
- Analysis and storage of more than 40 electrical variables;
- Analysis of useful life and predictive maintenance of sensors and actuators;

The interface developed for the Supervisory is shown in Figure 3, and a list of the variables monitored is described in Table 1.

Figure 2: Experimental Test Facility

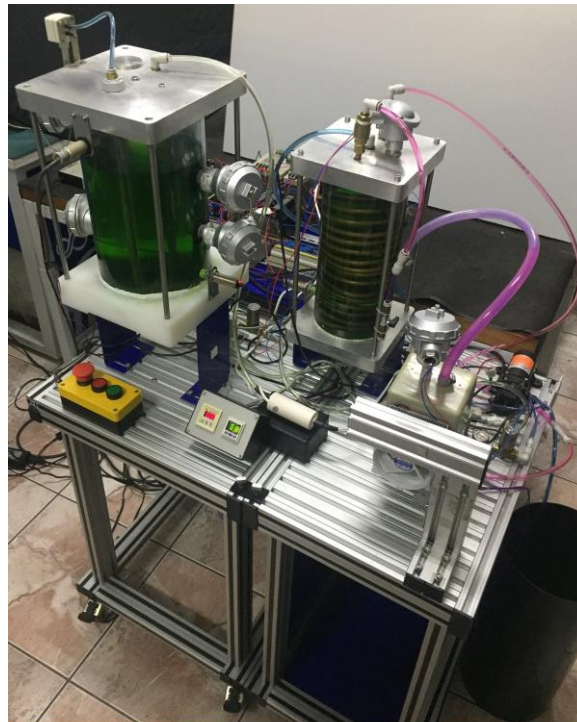


Figure 3: SCADA Interface

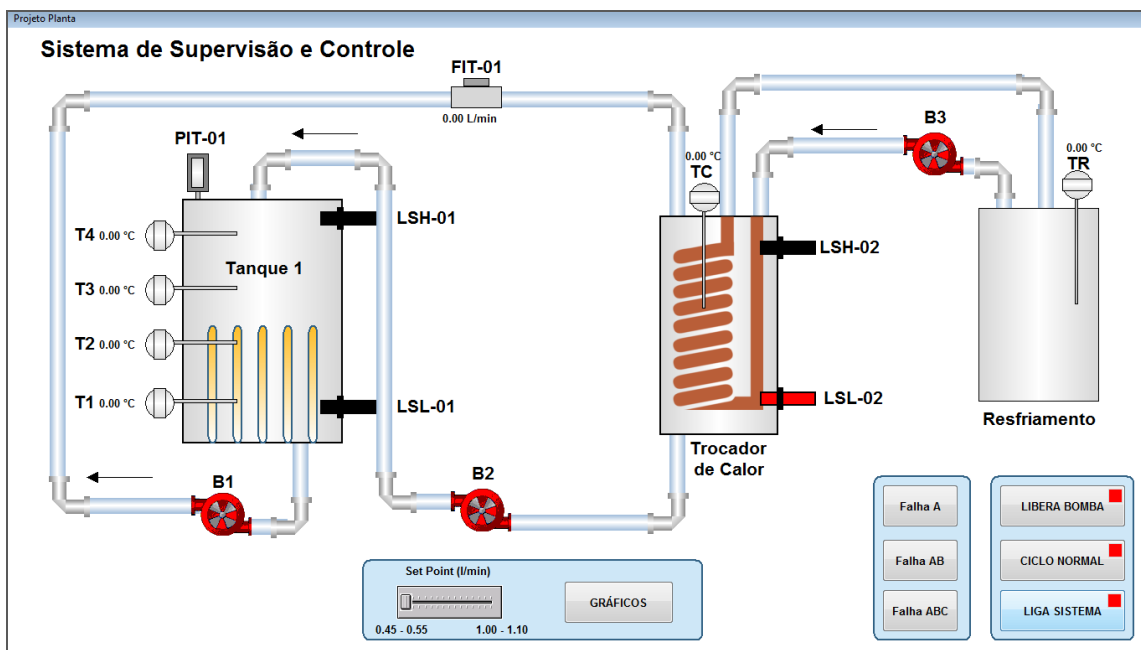


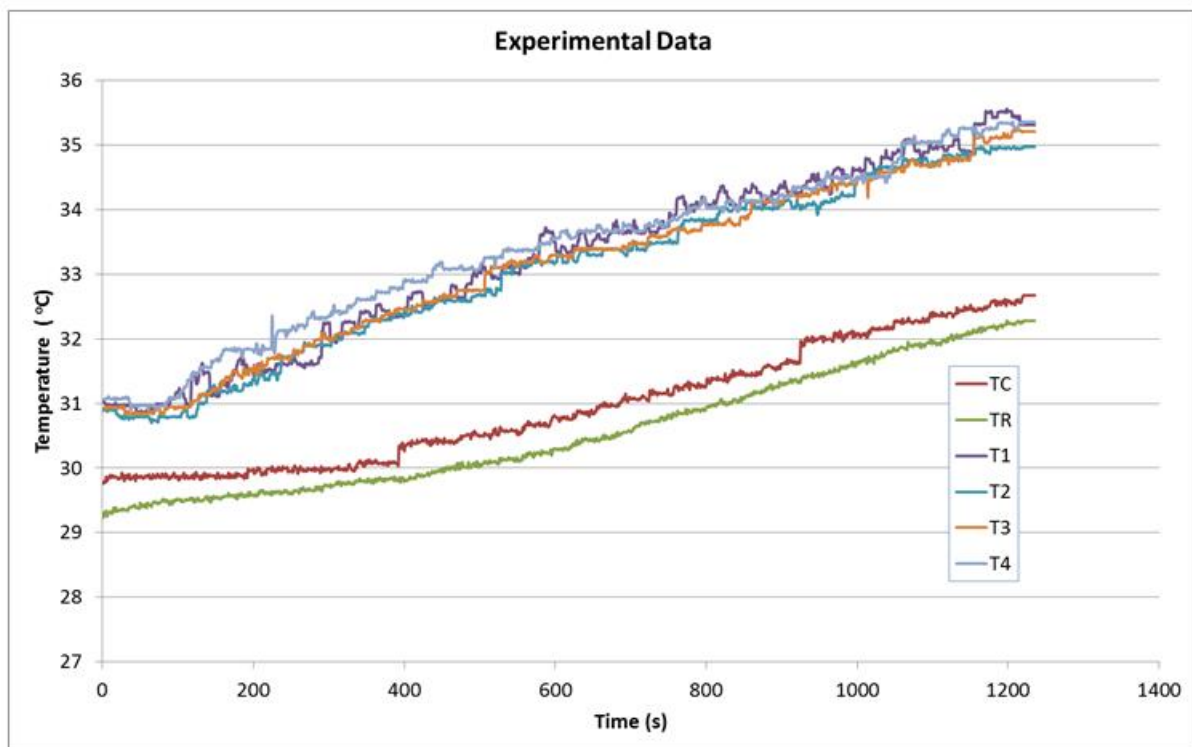
Table 1: Experimental Test Facility Variables

Item	Variable
T1	Temperature Sensor 01
T2	Temperature Sensor 02
T3	Temperature Sensor 03
T4	Temperature Sensor 04
TR	Temperature Cooler
TC	Temperature Heat Exchanger
FIT-01	Flowmeter
PIT-01	Pressure Sensor
LSH-01	High Level Sensor Tank 1
LSH-02	High Level Sensor Heat Exchanger
LSL-01	Low Level Sensor Tank 1
LSL-02	Low Level Sensor Heat Exchanger
B1	Pump 01 – Primary Loop
B2	Pump 02 – Primary Loop
B3	Pump 03 – Secondary Loop

4. METHOD AND RESULTS

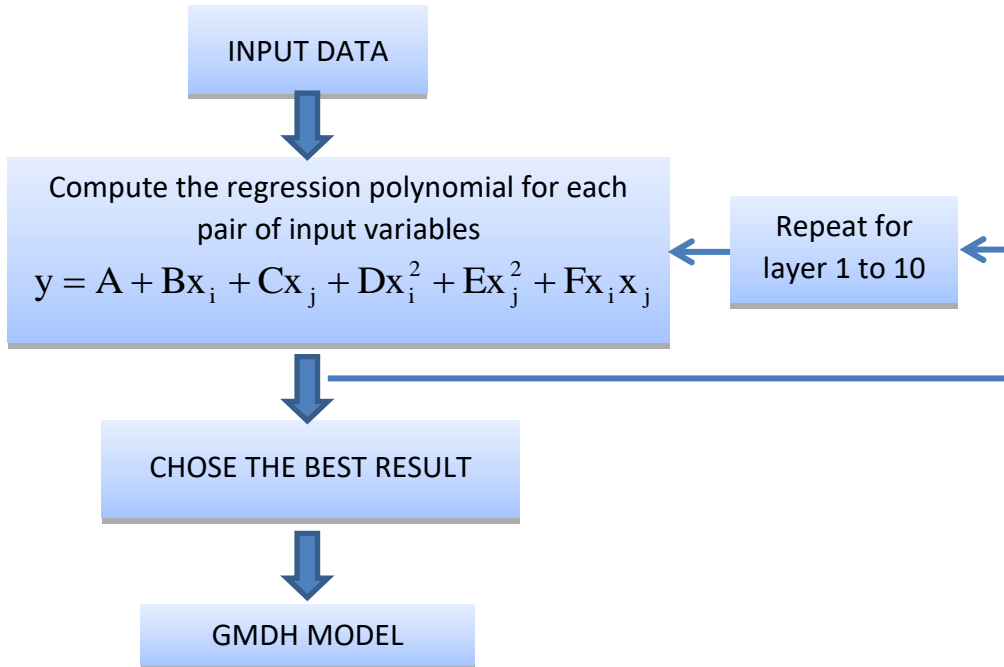
The GMDH Monitoring System was built using data generated by the Experimental Test Facility operating at normal condition.

Figure 4: *Experimental Data*



The GMDH methodology was developed for monitoring the temperature T1. According to GMDH theory, the input variables are combined two by two and the coefficients of equation 2 are calculated. The procedure is repeated for 10 layers and the layer correspondent to the lowest error is chosen. Figure 5 shows the GMDH model development for the Experimental Test Facility variables.

Figure 5: Determination of the GMDH Model



In order to determine the GMDH Model performance, the results of the GMDH model calculated for the each variable are compared with the actual value. The results are show at Figure 7 where we have also the percent mean error.

Figure 6: Comparison between actual value and GMDH

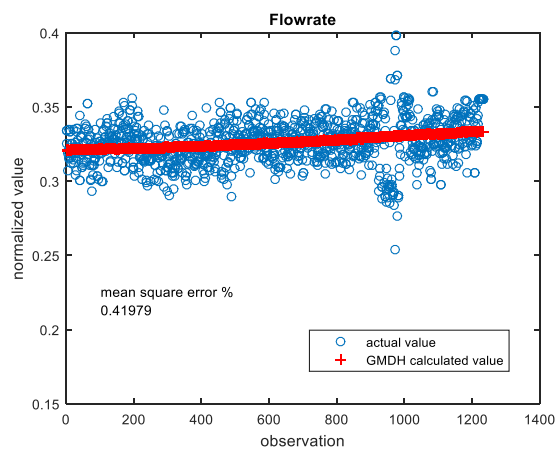
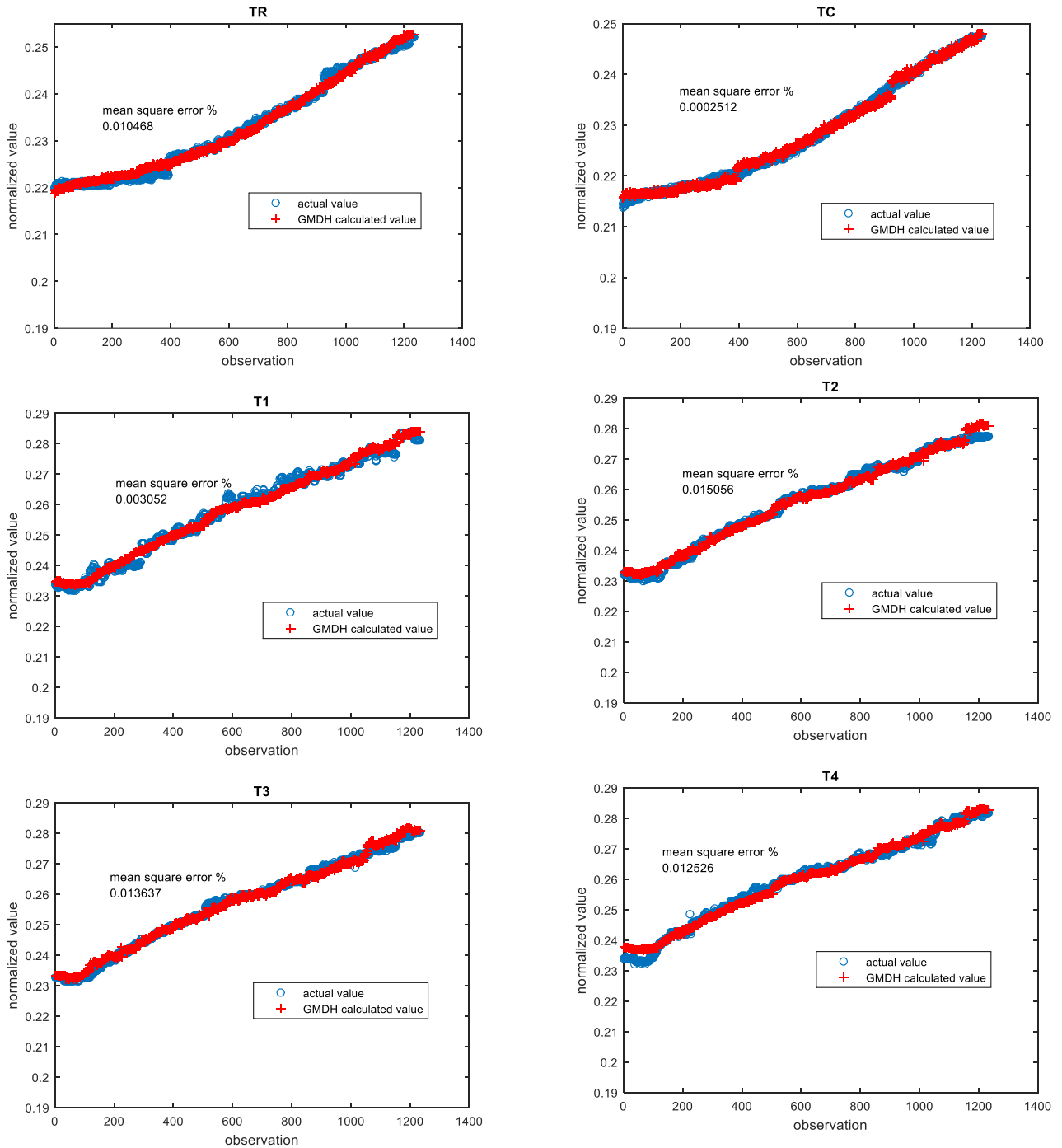


Figure 7: Comparison between actual value and GMDH



5. CONCLUSION AND FURTHER WORK

The GMDH model developed was applied to monitor seven Experimental Facility variables. For Flowrate monitoring, the Mean Square Error (MSE) was 0.42% and for the six Temperature variables, the mean MSE value was less than 0.01%. As further work, GMDH model will be improved to include more variables of the Experimental Facility, and also different operational conditions.

REFERENCES

1. FARLOW, S. J. **Self-organizing methods in modeling: GMDH-type algorithms**, New York: M. Dekker, 1984.
2. Onwubolu G. et al. **GMDH-Methodology and implementation in Matlab**. London, Imperial College Press, 2016.
3. IVAKHNENKO, A. G. **Self-teaching Systems of Recognition and Automatic Control**, Moscou: Tekhnika, 1969.
4. IVAKHNENKO, A. G.; YARACHKOVSKIY, Yu. P. **Self-organization at a System of Complex Models. Radio and Signal**, Moscou, 1981.
5. PEREIRA, I. M., MORAES, D. A. Development of a fault test experimental facility model using Matlab. In: **INTERNATIONAL NUCLEAR ATLANTIC CONFERENCE**, São Paulo, 2015.
6. FRENCH ASSOCIATION FOR STANDARDIZATION (FIP), **Bus for Exchange of Information Between Transmitter, Actuators and Programmable Controllers**, NF C46 601-607, Mar. 1990.
7. **O que é um sistema SCADA (What is a SCADA System)**. Available at: <<http://www.scadabr.org.br/?q=node/1>>. Last accessed: March 2014.