Day ahead electricity consumption forecasting with MOGUL learning model

Aria Jozi[1], Tiago Pinto[2], Isabel Praça[1], Zita Vale[1], João Soares[1]
[1]GECAD research group, Polytechnic of Porto (ISEP/IPP) Porto, Portugal {arjoz, icp, zav, joaps}@isep.ipp.pt [2]BISITE research group, University of Salamanca
Salamanca, Spain tpinto@usal.es

***Abstract*— Due to amount of today's electricity consumption, one of the most important tasks of the energy operators is to be able to predict the consumption and be ready to control the energy generation based on the estimated consumption for the future. In this way, having a trustable forecast of the electricity consumption is essential to control the consumption and maintain the balance in energy distribution networks. This study presents a day ahead forecasting approach based on a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL). The proposed approach is used to forecast the electricity consumption of an office building in the following 24 hours. The goal of this work is to present a more reliable profile of the electricity consumption comparing to previous works. Therefore, this paper also includes the comparison of the results of day ahead forecasting using GFS.FR.MOGUL method against other fuzzy rule based methods, as well as a set of Artificial Neural Network (ANN) approaches. This comparison shows that using the GFS.FR.MOGUL forecasting method for day-ahead electricity consumption forecasting is able to estimate a more trustable value than the other approaches.**

Keywords: day-ahead forecasting; electricity consumption; MOGUL learning methodology; office building

I. INTRODUCTION

One of the most important goals of smart grid operators is to have a better control on the power distribution network. Having a trustable estimated profile of the electricity consumption is one of the important tasks to enable achieving this target [1]. A trustable profile of the electricity consumption can also help the management of reserved electricity for emergency use. this way the generation cost for the electricity can be minimized and electricity tariff can be kept under controlled values [2].

Many studies have been published proposing different shorttime forecasting approaches to forecast the electricity consumption. For example, [3] presents a study that uses an artificial neural network (ANN) to predict the electricity consumption and results proves that the structure of ANN has a large influence on the quality of a short-time forecasting result. In [4] an ANN is also used for a day-ahead electricity consumption forecasting; however, in this work the forecasting method predicts the consumption value of every type of consumption in the building and then the sum of this values presents the final estimated consumption. In this work, the best time interval of data and best available variables are used in the case of each types of consumers. In [5] a method named as Hybrid Neural Fuzzy Inference System (HyFIS) is used in order to reach an hour-ahead forecasted profile of the electricity consumption and the results of this work shows that the HyFIS can forecast a more trustable value than the other methods, namely ANN based methods.

The objective of this study is to predict a more reliable dayahead profile of the electricity consumption of an office building. A genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL) has been implemented to be used as the forecasting method of this process. The electricity consumption of the building N of the Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD) is chosen to be forecasted in this work. The total electricity consumption of this building contains the consumption of three different types of consumption, namely air conditioning systems, lights and electrical sockets. For every target hour, the forecasting method will predict a consumption value for these three types of consumption and the sum of these three values will represent the final consumption values of the intended hour. The GFS.FR.MOGUL is used to predict these values. Moreover, the same process is applied to the HyFIS and Wang and Mendel's Fuzzy Rule Learning Method (WM), so that results can be compared. Additionally, the results of this work are also compared to the results from previous studies.

After this introductory section, section II describes the implementation of the GFS.FR.MOGUL forecasting method and includes the description of the used data base in this work. The experimental tests and the results are presented in section

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 641794 (project DREAM-GO) and grant agreement No 703689 (project ADAPT); and from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013

III. Finally, section IV presents the most relevant conclusions and contributions of this work.

II. MATERIAL AND METHODS

A genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL) is implemented in this work to forecast the electricity consumption of an office building for the next 24 hours. The electricity consumption form building N of the GECAD research center located in ISEP/IPP, Porto, Portugal has been chosen to be used in this work. These data are collected and stored through SOICAM (SCADA Office Intelligent Context Awareness Management) [6], a system that is used to manage and simulate the GECAD campus microgrid. This study uses the R programming language to implement the GFS.FR.MOGUL method. The implementation details and results of this work are discussed and compared in the following sections.

A. GFS.FR.MOGUL methodology

The GFS.FR.MOGUL is a forecasting method which implements a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology to specify the structure of the fuzzy IF-THEN rules and the membership function parameters. In this process two types if fuzzy rules are considered namely the descriptive and approximate/free semantic approaches. In the descriptive approaches, the linguistic labels represent a realworld semantic, and are uniformly defined for all rules. But in the other hand, the approximative approach does not have any associated linguistic label. This way every rule has the own membership function values. This scenario is proved by modeling a fuzzy IF-THEN rule on a chromosome which consists of the parameter values of the membership function. A population consists many such generated chromosomes, based on the iterative rule learning approach (IRL). IRL means that the chromosomes will be generated one by one, considering the fitness value and covering factor, until there are sufficient chromosomes in the population. the genetic algorithm will be started after obtaining the population [7].

In order to have a good behavior from a FRBS, many important statistical attributes need to be verify [8]. In the Generated Fuzzy Rule Bases (GFRB) obtained from MOGUL, the satisfaction of two of these statistical attributes will be considered. These two attributes are:

- completeness
- consistency

As an inductive approach to building GFRBSs is considered, both properties will be based on the existence of a training data set, Ep, composed of p numerical input-output problem variable pairs. These examples will present the following structure (1):

$$e^{"} = (ex_{\&}^{"}, ex_{(}^{"}, \dots, ex_{*}^{"}, ey^{"}), \quad l = 1, \dots, p \qquad (1)$$

It is clear that an FRBS most be always able to infer a proper output for all possible system input [9]. This property is called τ-completeness in the field of inductive learning and it may be mathematically formulated using a real value t by means of the following expressions (2), (3), (4) [10]:

$$C_R(e_l) = \bigcup_{i=1\dots t} R_i(e_l) \geq \tau, \quad l = 1, \dots, p \qquad (2)$$

$$R_i(e_l) = *\,(A_i(ex^l), B_i(ey^l)) \qquad (3)$$

$$A_i(ex^l) = *\,(A_{i1}(ex_1^l), \dots, A_{5*}(ex_*^{"})) \qquad (4)$$

where * is a t-norm, and $R_i(e^{"})$ is the compatibility degree between the rule Ri and the example e_l.

For the case of the Consistency of a Fuzzy Rule Base, a generic set of IF-THEN rules is consistent if it does not contain contradictions. It is necessary to relax the consistency property to consider the fuzzy rule bases. This is done by means of the positive and negative examples concepts. Equations (5) and (6) are the examples of the positive and negative set for the rule R_i.

- Positive: (5)

$$E_{B}(R_5) = \{e_l \in \frac{p}{} \geq 0_{2GE(HI)} \quad K$$

- Negative:

$$E^{N}(R_i) = \{e_l \in \frac{}{R_i(e_l)} = 0 \;\; and \;\; A_5(ex^{"}) > 0 \quad LS \qquad (6)$$

And by giving the value and equations (7) and (8):

$$n_{2}{}^{B}{}_{G} = |E^{B}(R_5)| \qquad (7)$$

$$n_{2}{}^{L}{}_{G} = |E^{L}(R_5)| \qquad (8)$$

We get that:

$$R_5 \text{ is } k-consistent \text{ when } n^{-}_{R_i} \leq k . n^{+}_{R} \qquad (9)$$

So, the way to incorporate the satisfaction of this property in the designed GFRBSs is to encourage the generation of kconsistent rules. Those rules not verifying this property will be penalized so as not to allow them to be in the FRB finally generated.

The GFS.FR.MOGUL in this work is implemented in R programming language and this implementation takes the advantage of using the FRBS package. The implemented GFS.FR.MOGUL method requires the specification of 6 variables. Table 1 presents these 6 variables as well as the values used in this work for each variable.

B. Data base

This work uses the electricity consumption of the Building N of GECAD facilities located in Porto, Portugal. These data are available in the SQL server of GECAD. This server includes many data bases with different information from the GECAD facilities. The data base of the building N is used in this work.

The electricity consumption of the building N is stored by five energy meters. Which energy meter stores the electricity consumption of a specific part of the building by 10 second time interval. Also, at any time the consumption is recorded from 3 different types of consumptions, namely the lights, the electrical sockets and the air conditioning system.

In order to collect the data from the data base, a java based application has been developed, which is connected to the SQL server. This application calculates the hourly consumption of each type of consumers and creates a .csv file that can be used as the input of the forecasting method.

Table 1 – GFS.FR.MOGUL input variables

Variable	Description	Used values
Persen_cross	the probability of crossover	0.6
Persen_mutant	the probability of mutation	0.3
Max.iter	The maximum number of iterations	10
Max.gen	The maximum number of generations of the genetic algorithm	10
Max.tune	The maximum number of tuning iterations	10
Epsilon	the boundary of covering factor	0.4

III. RESULTS AND DISCUSSION

This study aims at reaching a more reliable prediction value of the electricity consumption of an office building in the next 24 hours, comparing to the results of previous works. In this way, the electricity consumption of 24 hours of an official day will be forecasted and the GFS.FR.MOGUL has been chosen to forecast these values. Also, to compare the performance of this method to the other fuzzy rule based methods, the electricity consumption of these hours will be forecasted by 2 other forecasting methods. Namely as Hybrid Neural Fuzzy Inference System (HyFIS) and Wang and Mendel's Fuzzy Rule Learning Method (WM). The HyFIS is a combination of fuzzy interface systems and Artificial Neural Networks [5] and the WM is a fuzzy interface system based on Wang and Mendel's model [11].

The total electricity consumption of the building N consists of the consumption of three types of consumption, namely lights, electrical sockets and air conditioning system. The consumption of these types is recorded by five energy meters that each one stores the consumption of a different part of the building. Therefore, to forecast the total consumption of the building the methods will predict a value of consumption for every type of consumers and the sum of these three values presents the final predicted consumption value. Figure 1 demonstrates the processes of data collection and data usage in this implementation.

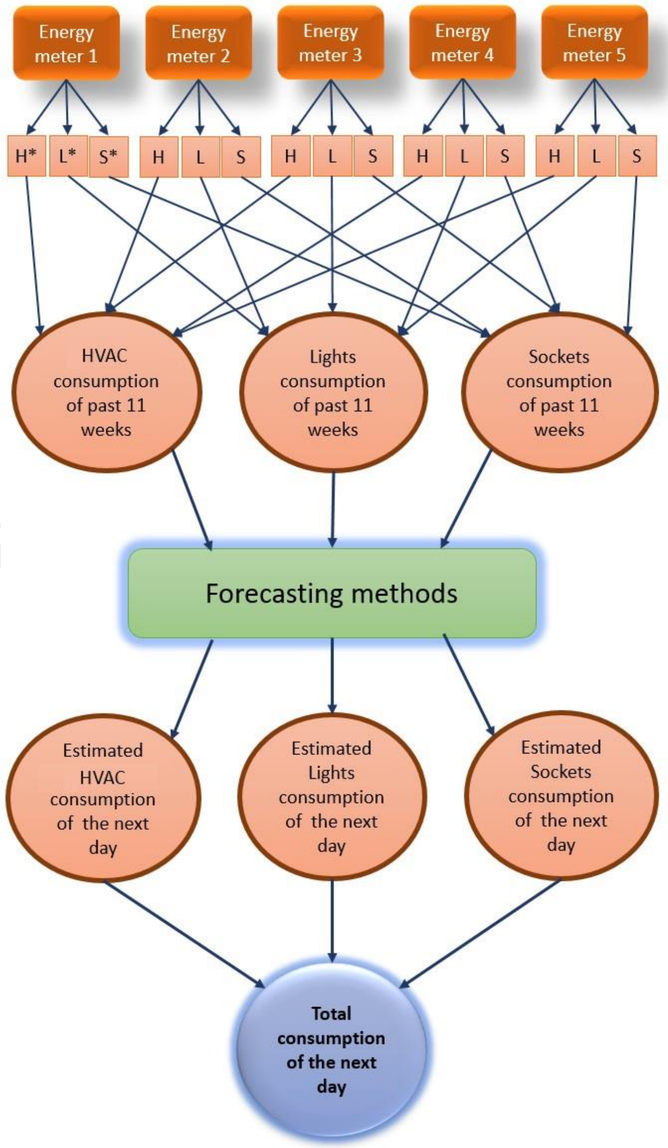

Figure 1 - Data collection and data usage in this implementation *H: HVAC, L: Lights, S: Sockets

The forecasting methods receive as input a .csv file, which is created by a java based application that is connected to the GECAD's data base server. The java application creates three .csv file to predict the consumption of every hour. Each one of these 3 files include the needed data to forecast the consumption of one type of consumers in the target hour. This way, to forecast the total electricity consumption of building for a target hour, the method is trained three time by different data to predict the consumption of three types of consumers.

Every .csv file includes 3 data tables: Train-input, Trainoutput and Test-input. The Train-input and Train-output tables are meant to be used to train the forecasting methods. These tables have the consumption data of the Intended consumers from past 3 months and by these data the method will be trained 10 times to be prepared to make the final decision. The Testinput table is the main input of the method. This table includes the electricity consumption value of the intended consumers of the same hours as the target hour from the past 14 days. Table 2 presents a brief example of the Test-input table in case of forecasting the consumption of electrical sockets in 13h of 16/11/2016.

Table 2 - Brief example of the Test-input table to forecasting the consumption of electrical sockets in 13h of 16/11/2016

Date→ / Hour↓	2/11/2016	3/11/2016	…	14/11/2016	15/11/2016
13:00	1860W	2299W	…	1466W	1605W

In order to test this strategy, the electricity consumption of 24hour of 16/11/2016 has been chosen to be forecasted. Table 3

presents the forecasted values for the electricity consumption during these hours by GFS.FR.MOGUL forecasting method.

Table 3 – Forecasted electricity consumption values by GFS.FR.MOGUL

Hour	GFS.FR.MOGUL			
	HVAC (W)	Lights (W)	Sockets (W)	Total (W)
0:00	587	0	1097	1683
1:00	531	0	1254	1785
2:00	520	0	1208	1728
3:00	531	0	1221	1752
4:00	514	0	1058	1572
5:00	499	0	1270	1769
6:00	502	0	1053	1555
7:00	509	0	1298	1807
8:00	471	0	1128	1598
9:00	533	323	1505	2360
10:00	530	539	1437	2505
11:00	563	660	1611	2834
12:00	893	684	1607	3184
13:00	1104	667	1461	3232
14:00	1174	701	1681	3556
15:00	1208	833	1839	3880
16:00	1262	804	1657	3722
17:00	1252	786	1547	3585
18:00	926	438	1338	2702
19:00	664	299	1178	2140
20:00	531	93	1230	1854
21:00	529	121	1373	2022
22:00	523	229	1245	1997
23:00	517	5	1292	1814

As Table 3 shows, during hours 0:00 to 8:00, the electricity consumption of the lights is predicted as zero, because during these hours usually the consumption of the lights are zero or close to zero.

All of the target values have been forecasted also by HyFIS and WM forecasting methods with the same training approach as the forecasted values by the GFS.FR.MOGUL. Figure 2 presents a comparison between the forecasted values of total electricity consumption by the forecasting methods as well as the real consumption value in every hour.

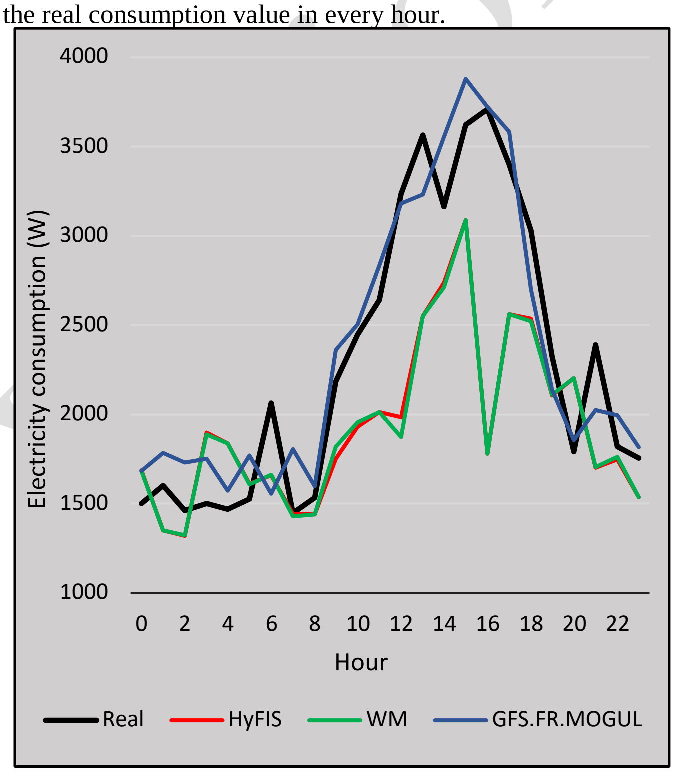

Figure 2 - comparison between forecasted and real electricity consumption values

As Figure 2 shows, during the night (from 0:00 to 6:00) the consumption has a stable value between 1450W and 1600W. As it is visible, the variation of the forecasted values by GFS.FR.MOGUL is much closer to the variation of the real values than the variation of the other methods and during the hours in which the consumption is higher; thus the GFS.FR.MOGUL provides a more reliable value. The Mean absolute percentage error (MAPE) has been chosen to validate the accuracy of each forecasted value. Table 4 presents the MAPE error of every forecasted value for each hour, as well as the average error for each forecasting method.

Although the HyFIS and the WM are forecasting a more exact value in some hours, for example at 7:00, the average error of these three methods proves that the GFS.FR.MOGUL has the most acceptable performance to predict the electricity consumption of the next 24 hours of an office building.

Table 4- MAPE error of every forecasted value

Hour	HyFIS	WM	Mogul
0:00	12.43%	12.03%	12.10%
1:00	15.73%	15.73%	11.39%
2:00	9.73%	9.50%	18.26%
3:00	26.75%	25.93%	16.91%
4:00	25.16%	25.16%	7.17%
5:00	5.71%	5.48%	15.99%
6:00	19.65%	19.59%	24.68%
7:00	0.26%	1.18%	24.85%
8:00	6.18%	6.18%	4.19%
9:00	19.89%	16.72%	7.96%
10:00	21.05%	20.03%	2.38%
11:00	23.81%	23.81%	7.35%
12:00	38.64%	42.11%	1.59%
13:00	28.47%	28.47%	9.39%
14:00	13.46%	14.28%	12.40%
15:00	14.69%	14.69%	7.14%
16:00	52.04%	52.04%	0.32%
17:00	24.64%	24.64%	5.47%
18:00	16.33%	16.82%	10.88%
19:00	9.33%	9.22%	7.91%
20:00	23.12%	23.12%	3.52%
21:00	28.71%	28.61%	15.31%
22:00	3.98%	3.21%	9.69%
23:00	12.48%	12.47%	3.29%
Average	**18.84%**	**18.79%**	**10.01%**

In [7] has been presented a study that uses Artificial Neural Network (ANN) to forecast the electricity consumption of the building N in the next 24 hours. This work proposes 5 different approaches to execute this forecast. These 5 approaches use different time intervals to prepare the input data and also use different environmental data to train the ANN. Table 5 shows the details of these forecasting approaches. Some of these approaches forecast the total consumption of the building by forecasting the consumption of every type of consumers, and then the sum of these values represents the predicted value for total consumption. The approaches considering the separation of the three types of consumption are referred in Table 5 as using strategy 2. On the other hand, the other approaches forecast the total consumption of the building without forecasting the consumption of every type of consumers and the method only predicts one value which is the predicted value for total consumption. These approaches have 1 as their strategy.

The ANN2-2 in case of forecasting the consumption of each type of consumers, uses the best time interval of the training for intended consumers and also takes the advantage of using the environmental data to train the method. In order to compare the performance of these forecasting approaches, Figure 3 presents the average MAPE error of forecasted values of next 24 hours of an official day by every forecasting approaches.

Table 5 - Details of the forecasting approaches [7]

Method	Time interval	Strategy	Training variables
ANN1-1	6 months	1	Electricity Consumption
ANN1-2	3 months	1	Electricity Consumption
ANN1-3	1 month	1	Electricity Consumption
ANN2-1	1 month	2	Electricity Consumption
ANN2-2	1 month for Air conditioning system and Sockets 6 months for lights	2	Electricity Consumption + Environmental data
GFS.FR. MOGUL	3 months	2	Electricity Consumption
HyFIS	3 months	2	Electricity Consumption
WM	3 months	2	Electricity Consumption

As the Figure 3 shows, the GFS.FR.MOGUL has the most trustable result between these approaches, with an average error of 10.1%. The GFS.FR.MOGUL not only has the best result but it only uses the electricity consumption data to train the method while ANN2-2 is using the consumption data and the environmental data and still has a higher error. This way the GFS.FR.MOGUL needs less data to make the forecasting process. Using the environmental data as well might help this method to have even more reliable result.

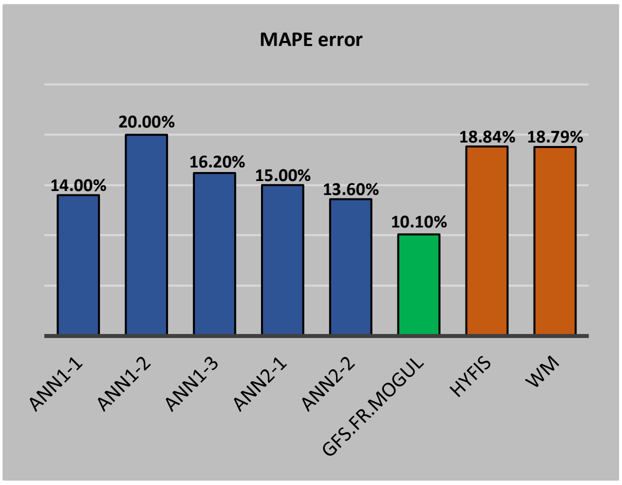

Figure 3 - Average MAPE error of the forecasting approaches

IV. CONCLUSIONS

The objective of this work is to predict a more reliable estimated profile for the electricity consumption of an office building during an official day comparing to other previous solutions. For this purpose, a Genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL) has been implemented to be used as the forecasting method of this work. In this process, the hourly electricity consumption of different types of the consumers of the building has been forecasted and the sum of these forecasted values presents the total estimated electricity consumption. The comparison of results of this forecasting method against the results of other fuzzy rule based methods and also to some proposed approaches based on ANN shows that the GFS.FR.MOGUL is able to forecast a more trustable value for electricity consumption than the other methods. The GFS.FR.MOGUL is providing a more reliable value by using even less historical data and less input variables.

The future works will include analyzing the influence of using other variables to train the GFS.FR.MOGUL, such as environmental data and using this method to predict the electricity consumption of longer time intervals.

REFERENCES

[1] S. x. Yang and Y. Wang, "Applying Support Vector Machine Method to Forecast Electricity Consumption," Computational Intelligence and Security, 2006 International Conference on, Guangzhou, 2006, pp. 929932.

[2] M. S. M. Othman, D. Johari, I. Musirin, T. K. A. Rahman and N. F. N. Ismail, "Artificial neural network-based forecast for electricity consumption in Malaysia," Power and Energy (PECon), 2010 IEEE International Conference on, Kuala Lumpur, 2010, pp. 24-28.

[3] D. Baczynski and M. Parol, "Influence of artificial neural network structure on quality of short-term electric energy consumption forecast," in IEE Proceedings - Generation, Transmission and Distribution, vol. 151, no. 2, pp. 241-245, 2 March 2004.

[4] E. Vinagre, L. Gomes and Z. Vale, "Electrical Energy Consumption Forecast Using External Facility Data," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, 2015, pp. 659-664.

[5] A. Jozi, T. Pinto, I. Praça, F. Silva, B. Teixeira and Z. Vale, "Energy consumption forecasting based on Hybrid Neural Fuzzy Inference System," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-5.

[6] Luís Gomes, Pedro Faria, Hugo Morais, Zita Vale, Carlos Ramos, “Distributed, Agent-Based Intelligent System for Demand Response Program Simulation in Smart Grids”, IEEE Intelligent Systems, vol. 29, pp. 56-65, January/February 2014

[7] Lala Septem Riza, Christoph Bergmeir, Francisco Herrera, and Jose Manuel Benitez, Fuzzy Rule-Based Systems for Classification and Regression Tasks, 2015

[8] Driankov, D.; Hellendoorn, H.; Reinfrank, M. An Introduction to Fuzzy Control; Springer-Verlag: Berlin, 1993.

[9] O. Cordon, M.J. del Jesus, F. Herrera, and M. Lozano, "MOGUL: A methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach", International Journal of Intelligent Systems, vol. 14, pp. 1123 - 1153 (1999).

[10] Herrera, F.; Lozano, M.; Verdegay, J. L. Fuzzy Sets and Systems 1998, 100, 143-158.

[11] L. X. Wang, J. M. Mendel, “Generating fuzzy rules by learning from examples”, IEEE Transactions on Systems, Man and Cybernetics, vol.22, no.6, p.1414-1427, 1992.