

# Electricity consumption forecasting in office buildings: an artificial intelligence approach

Aria Jozi<sup>1</sup>, Tiago Pinto<sup>1</sup>, *Member IEEE*, Goreti Marreiros<sup>1</sup>, Zita Vale<sup>2</sup>, *Senior Member, IEEE*

<sup>1</sup>GECAD research group, Polytechnic of Porto (ISEP/IPP)

Porto, Portugal

{arjoz, tmcfp, mgt}@isep.ipp.pt

<sup>2</sup>Polytechnic of Porto (ISEP/IPP)

zav@isep.ipp.pt

**Abstract**—The rising needs for increased energy efficiency and better use of renewable energy sources bring out the necessity for improved energy management and forecasting models. Electricity consumption, in particular, is subject to large variations due to the effect of multiple variables, such as the temperature, luminosity or humidity, and of course, consumers' habits. Current forecasting models are not able to deal adequately with the influence and correlation between the multiple involved variables. Hence, novel, adaptive forecasting models are needed. This paper presents a novel approach based on multiple artificial intelligence-based forecasting algorithms. The considered algorithms are artificial neural networks, support vector machines hybrid fuzzy inference systems, Wang and Mendel's fuzzy rule learning method and a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology. These algorithms are used to forecast the electricity consumption of a real office building, using multiple input variables and consumption disaggregation.

**Keywords:** artificial intelligence, electricity consumption, forecasting, office building

## I. INTRODUCTION

The increase of the technology and energy based systems in nowadays lifestyles lead to a larger daily energy consumption which results as the unpredictability of the energy demand [1]. On the other hand, the limitation of renewable energy sources and energy storage makes it important and necessary for the energy operators to have a trustable profile of energy consumption in the future [2]. A more reliable energy consumption perspective helps the energy operators to have better control on the energy demand and the power distribution networks that terminate as a better control on the market price and unnecessary consumptions [3]. This way, The European Union (EU) has assumed a pioneer and leading role in this field and has committed to reaching ambitious targets [4] and a huge investment has already been made, e.g. in smart meters [5].

Many studies have been published during the past years to improve the prediction models in power energy systems. In [6] has been presented a study that uses forecasting methods to identify correlations between electricity consumption behavior and distributed photovoltaic (PV) output. Or in [7] where the authors present a hybrid model that incorporates multiple linear regression with artificial neural networks for a daily energy consumption forecasting. Further relevant related works using different approaches can be founded in [8]–[10]. Current

forecasting models are, however, not able to deal adequately with the influence and correlation between the multiple involved variables, such as luminosity and temperature, and therefore, novel, adaptive forecasting models are needed.

This paper presents a day-ahead energy consumption forecasting approach for office buildings. The forecasting model contains five forecasting algorithms, namely Artificial Neural Network (ANN), Support Vector Machines (SVM) and three Fuzzy Rule Based Systems (FRBS), which are: hybrid fuzzy inference systems (HyFIS), Wang and Mendel's Fuzzy Rule Learning Method (WM) and a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL). Based on this approach, the total energy consumption of the building is disaggregated considering the different types of existent energy consumption in the building, namely HVAC, lights and electrical sockets. The forecasting methods are trained separately by different data sets to predict these three consumption values. Also, the influence of the different environmental variables on the consumption of every consumption type is included in the learning process.

## II. MATERIAL AND METHODS

This work proposes a forecasting model for office buildings that includes five forecasting methods namely, Artificial Neural Network (ANN), Support Vector Machines (SVM) and three Fuzzy Rule Based Systems (FRBS), which are: hybrid fuzzy inference systems (HyFIS), Wang and Mendel's Fuzzy Rule Learning Method (WM) and a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL). These methods are implemented in the R programming language and are based on different data mining techniques. This improves the performance of this forecasting model and ensures that the obtained results are the most reliable results. The proposed model in this study uses the data from building N of the GECAD research center located in ISEP/IPP, Porto, Portugal. This building has a SQL database that includes the electrical information of the building as well as meteorological variables of the related location. The electrical data are collected and stored through SOICAM (SCADA Office Intelligent Context Awareness Management [11], a system that is used to manage and simulate the GECAD campus microgrid. The meteorological are received from the ISEP meteorological service [12].

### A. Artificial Neural Network

ANNs are one of the most used and known forecasting methods. This method is inspired by the human brain and their number of neurons with high interconnectivity. ANNs are several combined nodes or neurons, divided into different levels and interconnected by numeric weights. They resemble the human brain in two fundamental points: the knowledge being acquired from the surrounding environment, through a learning process; and the network's nodes being interconnected by weights (synaptic weights), used to store the knowledge. The ANN is implemented in R using the "neuralnet" package [13].

### B. Support Vector Machines

SVM's are a field of supervised machine learning methods and are one of the most known methods in the area of forecasting. The first running kernel of SVM was created in the sequence of [14] by Vapnik, implementing a generalization of the nonlinear algorithm Generalized Portrait and only for classification and linear problems. Vapnik developed the statistical learning theory further in 1979. Finally, the current form of the SVM approach was presented in 1992, with a paper at the COLT conference [15]. In this study, the "e1071" package [16] based on R language is used to implement SVM.

### C. Fuzzy Rule Based Systems

Fuzzy rule-based systems (FRBSs) are a set of different forecasting algorithms based on the fuzzy concept [17]. These methods are implemented based on various combinations between fuzzy rules and data mining approaches such as space partition, neural networks, clustering, genetic algorithms, and gradient descent methods. This study takes advantage of using the "frbs" package [18] based on the R language which contains 15 FRBSs methods. For this work, three methods from this package have been chosen to be used namely as Hybrid Neural Fuzzy Interface System (HyFIS), Wang and Mendel's Fuzzy Rule Learning Method (WM) and a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL).

HyFIS uses a combination of neuronal networks with fuzzy rules to implement a forecasting algorithm and is the most known method of the FRBS package which includes two learning phases [19]:

- The first phase concerns the structure learning, i.e., finding the rules by using the knowledge acquisition module;
- The second phase regards the parameter learning phase for tuning fuzzy membership functions to achieve the desired level of performance [20].

The advantage of using HyFIS is that the fuzzy rule base can be easily updated when there is newly available data. When there is a newly available pair data, a rule is created for this data, and this new rule updates the fuzzy rule base. A detailed explanation of this method's implementation can be found in [21].

The second chosen FRBS method for this work is WM proposed by Wang and Mendel in 1991 [22]. The process of fuzzy rule bases generation in this method includes four steps:

1. Divide the Input and Output Spaces into Fuzzy Regions.
2. Generate Fuzzy Rules from Given Data Pairs

3. Assign a degree to each rule
4. Create a Combined Fuzzy Rule Base

This method has been known because of having a simple structure with good performance. In [23] has been presented with a complete explanation of the steps of this method as well as the details of the implementation process.

GFS.FR.MOGUL is a forecasting method that implements a genetic algorithm determining the structure of the fuzzy IF-THEN rules and the membership function parameters. Two general types of fuzzy IF-THEN rules are considered:

- Descriptive rules.
- Approximate/free semantic approaches.

In the first type, the linguistic labels illustrate a real-world semantic, and the linguistic labels are uniformly defined for all rules. In contrast, in the approximate approach, there is any associated linguistic label. The work presented in [24] includes a complete explanation about the structure of this method.

## III. RESULTS AND DISCUSSION

This study proposes a day-ahead energy consumption forecasting approach for the energy consumption of an office building, which the energy consumption of the building N of GECAD facilities located in Porto, Portugal has been chosen for this purpose. The proposed model includes five forecasting algorithms namely as ANN, SVM, HyFIS, WM, and GFS.FR.MOGUL in order to obtain the best possible results.

The proposed forecasting approach contains two forecasting strategies that are based on using different variables and different data structures to train the forecasting methods. Both strategies divide the total energy consumption of the building into three values of consumption which correspond as the three existed types of consumers in the building such as HVAC, Lights, and Sockets. The Energy consumption of each one of these consumers will be forecasted separately and the methods will be trained ones for every type of consumers. For the case of Lights only the consumption from 11:00 to 19:00 is considered because out of this interval the lights consumption is zero. These strategies first selected 14 days before the target day that have the same day type (official day, weekend) as the target day. The test data set which is the main input of the methods includes the data from these selected 14 days. And as the train data, the same day of the week as the target day from the past 10 weeks are considered as the targets. This way, for all these 10 days the same data set as the test data set is created and the methods train based on this data set. A different data set is created for every hour of the target day and every consumer type. This means to predict the energy consumption of the next 24 hours; 72 different data sets are created, and the methods will be trained 72 times by these data sets.

In the first strategy, the input data consider only the historical energy consumption of the building. For every consumer, the train and the test data sets include the value of energy consumption of the intended consumer. The forecasting methods predict three consumption values for every hour and the sum of these values presents the predicted total consumption. Figure 1 presents the total consumption values by the five proposed forecasting methods as well as the real energy consumption of 17/5/2018.

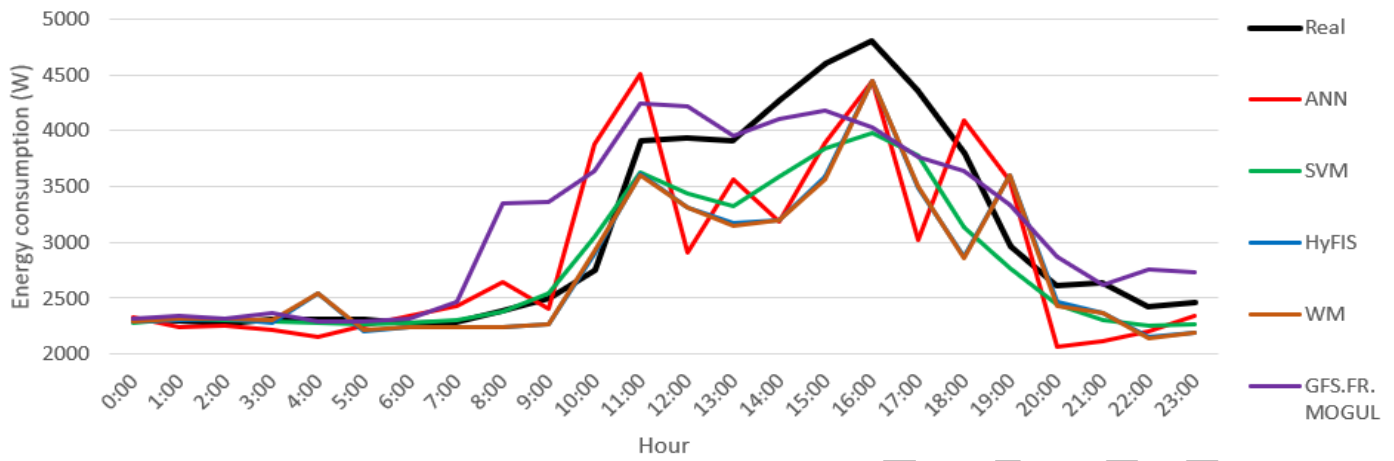


Figure 1 – Forecasted total consumption values for 17/5/2018 based on the first forecasting strategy

As can be seen in figure 1 the methods present different performances but all the forecasted values at the peak hours of the consumption are lower than the real consumption of the building. Between these five methods SVM by average MAPE error of 7.35% presents the most reliable results, followed by GFS.FR.MOGUL by 9.87%. HyFIS and WM have a similar result by average MAPE errors of 10.13% and 10.36% and finally, the ANN by the error of 12.27% presents the highest average error while the first proposed forecasting strategy is used.

The Second forecasting strategy proposes the usage of a second variable in the process of the training. This second variable is chosen based on the type of consumer. In this process in the case of the HVAC, the input data set includes a combination of the energy consumption and environmental temperature of the related place which in this case is the environment temperature of the Porto. It is obvious that the usage on the HVAC system has a direct influence by the temperature of the place and the methods can improve their result by recognizing this relation. To forecast the energy consumption of the Lights the value of the solar radiation of the

related place is considered the energy consumption as well in the training data set. As same as the influence of the environmental temperature on the HVAC's consumption, the solar radiation can affect the consumption of the Lights and using this combination in order to estimate this consumption can improve the results. The process of the forecasting the consumption of Sockets in this strategy is the same as the first strategy and only the energy consumption of the sockets during the selected days is used to train the methods. Figure 2 presents the structure of this forecasting strategy.

In order to compare two proposed forecasting strategies, the energy consumption of the same date as the presents results for the first strategy has been chosen to evaluate the efficiency of the second forecasting strategy. Figure 2 presents the predicted total consumption values by the forecasting methods based on the second strategy as well as the real values for 17/5/2018.

As it is visible in figure 3, while the second strategy is used the method are able to predict the more reliable results during the peak hours of the consumption.

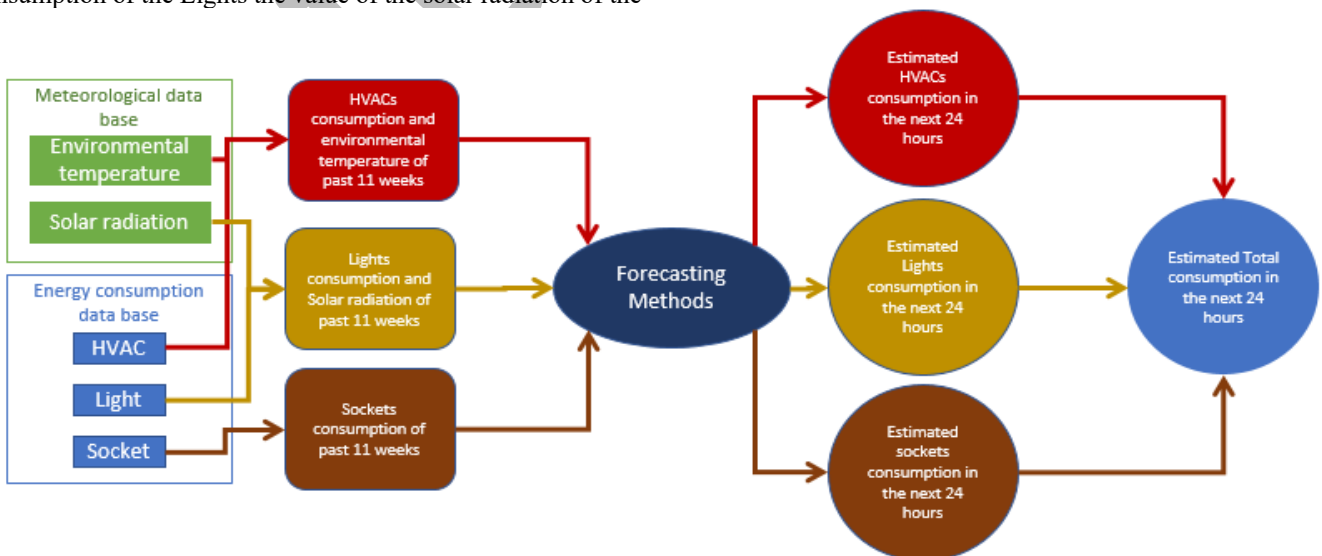


Figure 2 - The structure of the second proposed forecasting strategy

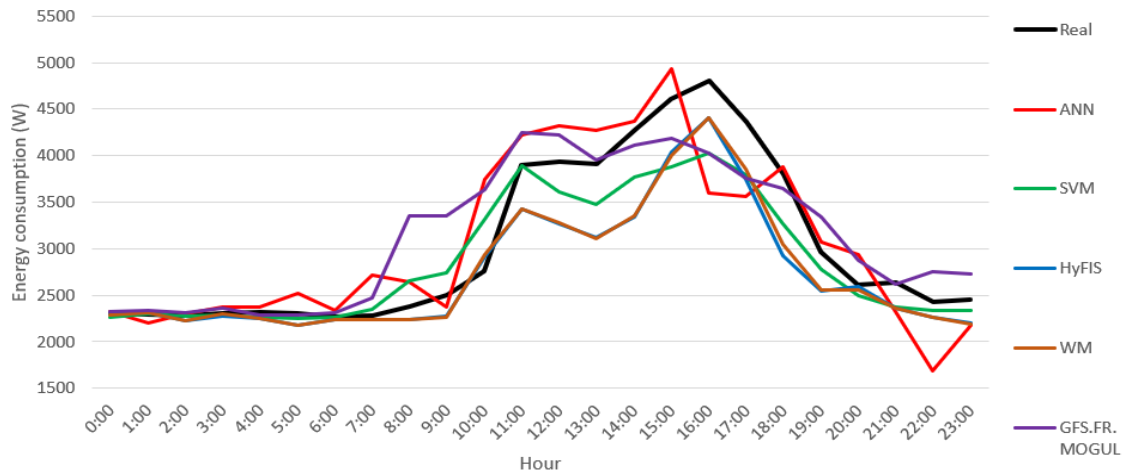


Figure 3 - Forecasted total consumption values for 17/5/2018 based on the second forecasting strategy

Table 1 – Average MAPE errors of the forecasting methods

		ANN	SVM	HyFIS	WM	GFS.FR.MOGUL
Strategy 1	HVAC	35.15%	20.71%	27.44%	27.40%	26.17%
	Lights	32.24%	15.66%	31.06%	31.55%	31.11%
	Sockets	9.98%	5.52%	7.02%	7.10%	4.87%
	<b>Total</b>	<b>12.27%</b>	<b>7.35%</b>	<b>10.13%</b>	<b>10.26%</b>	<b>9.87%</b>
Strategy 2	HVAC	27.39%	19.60%	23.33%	23.19%	26.17%
	Lights	47.23%	15.67%	35.10%	30.08%	31.11%
	Sockets	9.98%	5.52%	7.02%	7.10%	4.87%
	<b>Total</b>	<b>10.23%</b>	<b>7.19%</b>	<b>8.71%</b>	<b>8.58%</b>	<b>9.87%</b>

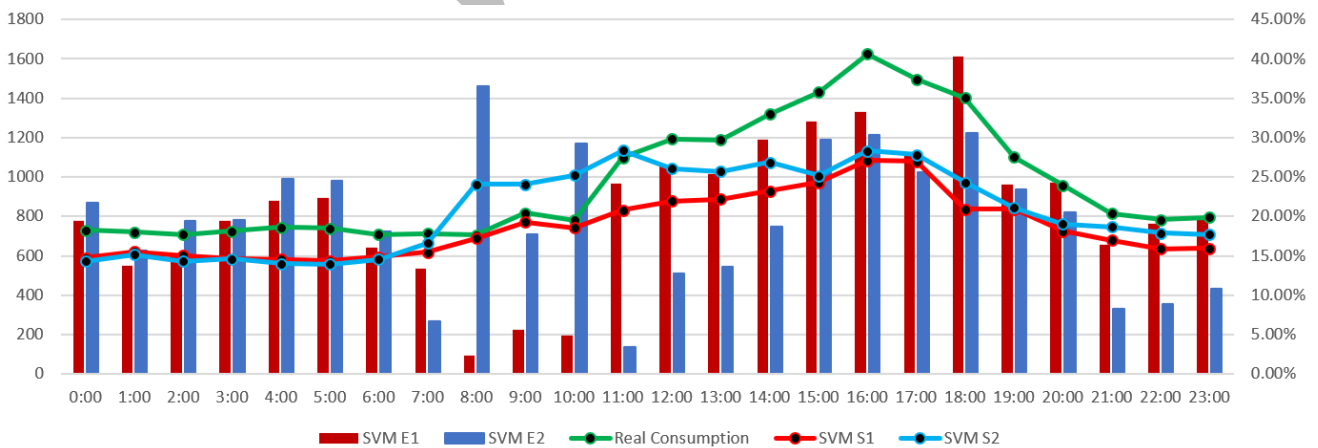


Figure 4- Forecasted values and MAPE errors of energy consumption of the HVAC system by SVM for 17/5/2018

\*S: strategy, E: MAPE error

Based on the showed results the SVM by the average MAPE error of 7.19% is the most trustable methods. WM by 8.58%, HyFIS by 8.71% and ANN by 10.23% are also presenting the more trustable results than their own results when the first strategy is used. GFS.FR.MOGUL by an average error of 9.87% presents the same results as the first strategy. Table 1 presents the average MAPE errors of the forecasting methods based on the first and second strategy.

The comparison of the presented average errors for every type of consumer separately proves that the SVM also in the case of the consumption of each type of consumer also has the best performance. Figure 4 presents the forecasted values for the HVAC system by the SVM for the 17/5/2018 as well as the MAPE error of each value. Figure 4 shows that in some specific hours the SVM based on the first strategy presents a lower error but, in most hours of the day, especially during the peak hours of consumption, the second strategy has the better results.

In the case of the consumption of the Lights, SVM is also the most trustable method. As has been explained, for the consumption of the lights only the consumption from 11:00 to 19:00 is considered and SVM by the average MAPE error of 15.66% for strategy 1 and 15.67% for strategy 2 has the most reliable results during these hours. Based on the average MAPE error using the solar radiation values in the train data of the method does not make a positive influence on the results, but the analyses of the standard deviation of the obtained error disaffirm this theory. The standard deviation of the errors when the first strategy is used is 10.3%. while this value when the second strategy is used is 9.06%. Which indicate that majority of the forecasted values based on the second strategy present a lower error than the forecasted values based on the first strategy. Figure 5 presents the predicted values by SVM for the consumption of the lights as well as the MAPE error for every value.

For the consumption the electrical sockets the situation is different. As there is no available environmental variable that has an influence on this consumption, the only used variable in the process of both strategies is the historical energy consumption. However, the consumption of the sockets is the only case which SVM does not present the best forecasting results. The average MAPE errors of SVM for consumption of

the sockets is 5.52% and this error for GFS.FR.MOGUL is 4.87%. as can be seen in the table1 using a second variable in the process of forecasting the consumption of HVAC and lights has no influence on the results of the GFS.FR.MOGUL. But in the other hand, in the case of the sockets, as the value of this consumption is more stable during the 24 hours of the day and there is no other variable to be used in the method training process, the GFS.FR.MOGUL is the most trustable method in this case. Figure 6 presents the forecasted values by GFS.FR.MOGUL and SVM for sockets consumption as well as the MAPE errors of these values.

As the presented results and figures show, for a day-ahead energy consumption forecasting, dividing the value of the consumption in to consumption values of different types of consumer and using different variables related to the intended consumer type to train the forecasting method is able to improve the final results and describe a more trustable energy consumption profile for the next 24 hours. Based on the obtained results between the five proposed forecasting methods, SVM calculates the most closed estimated values to the real value by the average MAPE error of 7.19% for total consumption.

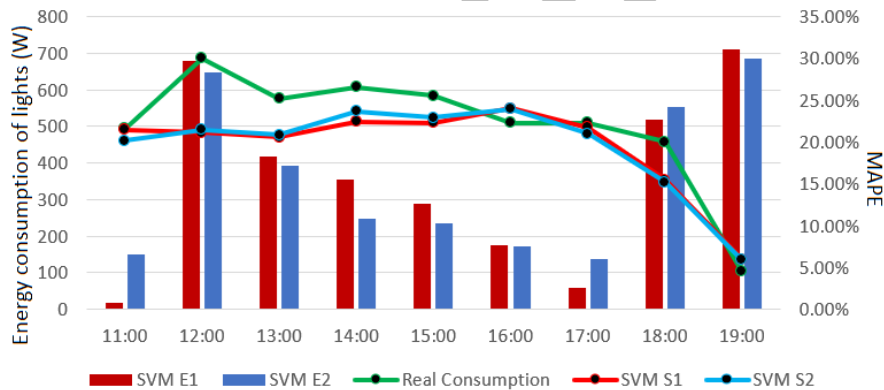


Figure 5 - Forecasted values and MAPE errors of energy consumption of the HVAC system by SVM for 17/5/2018

\*S: strategy, E: MAPE error

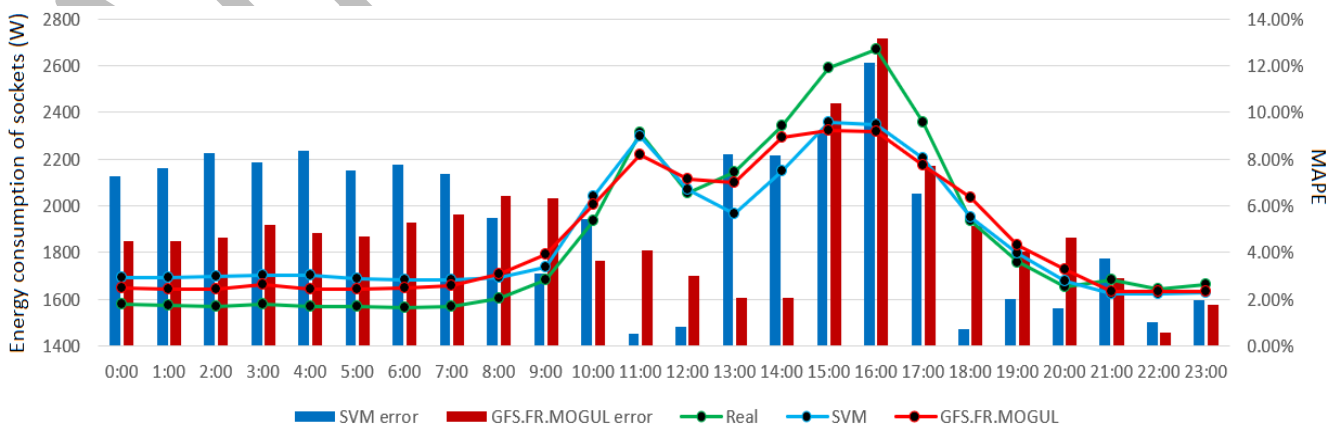


Figure 6 - Forecasted values and MAPE error of energy consumption of the HVAC system by SVM and GFS.FR.MOGUL for 17/5/2018

#### IV. CONCLUSIONS

The presented paper as the main target considers the improvement of day-ahead energy consumption forecasting for a building. This work proposes a forecasting model that includes five forecasting methods namely as Artificial Neural Network (ANN), Support Vector Machines (SVM) and three Fuzzy Rule Based Systems (FRBS), which are: hybrid fuzzy inference systems (HyFIS), Wang and Mendel's Fuzzy Rule Learning Method (WM) and a genetic fuzzy system for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL), implemented in R programming language. Also, two forecasting strategies have been presented in this study in order to forecast the energy consumption of the building N of GECAD facilities. These forecasting strategies divide the energy consumption into three consumption values which correspond as the consumption of HVAC, Lights, and Sockets. And train the methods separately for each one of these values to have a more exact predicted consumption. The first strategy uses only the historical data of the energy consumption and the second strategy uses a second variable as well as the energy consumption in the training process of the methods. This second variable is chosen based on their influence on the consumption of the related consumer type. Which for the HVAC system is the environmental temperature and for the lights is the solar radiation of the related place.

The presented results in this paper prove that while the methods are trained separately to forecast the value for the different type of consumers instead of forecasting directly the total consumption, the obtained result is more trustable. Also, the result confirms that usage of a second related variable in the training process helps the methods to estimate a more reliable perspective in the energy consumption during the next 24 hours.

Between the five presented forecasting method, four of them were able to recognize the relation between the used second variable and used this relation to calculate more exact results. These four methods are ANN, SVM, HyFIS, and WM. The results of GFS.FR.MOGUL were not improved when the second variable has been used but also no negative influence was recorded. Between these methods, the SVM by the average MAPE error of 7.19% based on the second strategy presents the most reliable results. Also in the case of the consumption of each type of consumer separately the SVM by the average error of 19.6% for HVAC systems and 15.67% for lights based on the second strategy has the lowest errors between the methods. For the case of the consumption of the sockets the best performed results belongs to GFS.FR.FS by the average error of 4.87% followed by SVM by 5.52%.

As future work, energy consumption prediction for longer time intervals are considered as well as the implementation of the proposed model for outer types of buildings by different consumer types.

#### REFERENCES

- [1] J. Conejo, Antonio, Carrión Miguel, Morales, "Decision Making Under Uncertainty in Electricity Markets," 2010.
- [2] S. X. Yang and Y. Wang, "Applying support vector machine method to forecast electricity consumption," *2006 Int. Conf. Comput. Intell. Secur. ICCIAS 2006*, vol. 1, pp. 929–932, 2007.
- [3] M. S. Mohamed Othman, D. Johari, I. Musirin, T. K. Abdul Rahman, and N. F. Nik Ismail, "Artificial neural network-based forecast for electricity consumption in Malaysia," *PECon2010 - 2010 IEEE Int. Conf. Power Energy*, pp. 24–28, 2010.
- [4] European Commission, "A 2030 framework for climate and energy policies," pp. 1–16, 2013.
- [5] European Commission, "Benchmarking smart metering deployment in the EU-27 with a focus on electricity," *Europe*, no. 2013, pp. 1–8, 2014.
- [6] Y. Zhang, R. Yang, K. Zhang, H. Jiang, and J. J. Zhang, "Consumption Behavior Analytics-Aided Energy Forecasting and Dispatch," *IEEE Intell. Syst.*, vol. 32, no. 4, pp. 59–63, 2017.
- [7] J. Teixeira, S. Macedo, S. Gonçalves, A. Soares, M. Inoue, and P. Cañete, "Hybrid model approach for forecasting electricity demand," vol. 2017, no. June, pp. 2316–2319, 2017.
- [8] D. Van Der Meer, et al, "Energy Management System with PV Power Forecast to Optimally Charge EVs at the Workplace," *IEEE Trans. Ind. Informatics*, vol. 14, no. 1, pp. 311–320, 2018.
- [9] O. Valgaev, F. Kupzog, and H. Schmeck, "Building power demand forecasting using K-nearest neighbours model – practical application in Smart City Demo Aspern project," *CIREC - Open Access Proc. J.*, vol. 2017, no. 1, pp. 1601–1604, 2017.
- [10] P. Li, B. Zhang, Y. Weng, and R. Rajagopal, "A Sparse Linear Model and Significance Test for Individual Consumption Prediction," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4489–4500, 2017.
- [11] L. Gomes, P. Faria, H. Morais, Z. Vale, and C. Ramos, "Distributed, agent-based intelligent system for demand response program simulation in smart grids," *IEEE Intell. Syst.*, vol. 29, no. 1, pp. 56–65, 2014.
- [12] "ISEP meteo website: <http://meteo.isep.ipp.pt>."
- [13] S. M. Stefan Fritsch, Frauke Guenther, Marc Suling, "Package 'neuralnet.'"
- [14] A. Vapnik, V.; Lerner, "Pattern recognition using generalized portrait method," *Autom. Remote Control*, vol. 24, pp. 774–780, 1963.
- [15] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proc. fifth Annu. Work. Comput. Learn. theory - COLT '92*, pp. 144–152, 1992.
- [16] D. Meyer, A. W. Evgenia Dimitriadou, Kurt Hornik, F. Leisch, C.-C. Chang, and C.-C. Lin, "Package 'e1071.'"
- [17] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
- [18] L. S. Riza, C. Bergmeir, F. Herrera, and J. M. Benítez, "{frbs}: Fuzzy Rule-Based Systems for Classification and Regression in {R}," *R Packag. version 3.1-0*, 2015.
- [19] J. Kim and N. Kasabov, "HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems," *Neural Networks*, vol. 12, no. 9, pp. 1301–1319, 1999.
- [20] W. Pedrycz, P. Ekel, and R. Parreiras, "Notions and Concepts of Fuzzy Sets: An Introduction," *Fuzzy Multicriteria Decis.*, pp. 21–62, 2010.
- [21] A. Jozi, T. Pinto, I. Praca, F. Silva, B. Teixeira, and Z. Vale, "Energy consumption forecasting based on Hybrid Neural Fuzzy Inference System," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–5.
- [22] X. Zhang, J. Wang, K. Zhang, "Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm," *Electr. Power Syst. Res.*, v. 146, 270–285, 2017.
- [23] A. Jozi, et al, "Wang and Mendel's Fuzzy Rule Learning Method for Energy Consumption Forecasting considering the Influence of Environmental Temperature," in *Giis 2016*, 2016.
- [24] A. Jozi, T. Pinto, I. Praca, and Z. Vale, "Day ahead electricity consumption forecasting with MOGUL learning model," in *IEEE World Congress on Computational Intelligence*, 2018.