Intelligent Energy Forecasting based on the Correlation between Solar Radiation and Consumption patterns

Eugénia Vinagre[1], Juan F. De Paz[2], Tiago Pinto[1], Zita Vale[1], Juan M. Corchado[2], Oscar Garcia[3]

[1]GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto, Portugal

[2]BISITE Research Centre, University of Salamanca, Salamanca, Spain

[3]Nebusens, S.L. R&D Department, Salamanca, Spain

empvm@isep.ipp.pt, fcofds@usal.es, tmcfp@isep.ipp.pt, zav@isep.ipp.pt, corchado@usal.es; oscar.garcia @nebusens.com

***Abstract*— The increasing penetration of renewable generation brings a significant escalation of intermittency to the power and energy system. This variability requires a new degree of flexibility from the whole system. The active participation of small and medium players becomes essential in this context. This is only possible by using adequate forecasting techniques applied both to the consumption and to generation. However, the large number of incontrollable factors, such as the presence of consumers in the building, the luminosity, or external temperature, makes the forecasting of energy consumption an arduous task. This paper addresses the electrical energy consumption forecasting problem, by studying the correlation between the solar radiation and the electrical consumption of lights. This study is performed by means of three forecasting methods, namely a multi-layer perceptron artificial neural network, a support vector regression method, and a linear regression method. The performed studies are analyzed using data gathered from a real installation – campus of the Polytechnic of Porto, in real time.**

***Index Terms*— Artificial Neural Network, Electricity Consumption, Solar Radiation, Support Vector Regression**

I. INTRODUCTION

The power and energy system operation is going through profound changes [1]. These are mainly due to the increasing penetration of renewable energy sources, as incentivized e.g. by the European Union (EU), through the H2020 Energy package [2]. Another significant change is related to the liberalization and restructuring of electricity markets [3], and consequent emergence of a diversity of new players, such as different types of aggregators [4].

The balance between consumption and generation is a critical issue when dealing with massive Distributed Generation (DG) [5]. This balance should take place in offline mode and also in an online mode. The online mode is when the balance and decisions take place in real-time. The offline mode takes the decisions before the event. The use of offline techniques requires adequate resource forecasting models (i.e. consumption and generation). The quality of offline methods is thereby closely linked to the quality of forecasting methods.

Energy consumption is dependent on several factors (e.g. weather conditions, annual seasons; consumer behavior, etc.), which makes its prediction a complex and difficult task [6-9]. Despite its complexity, the electrical consumption forecasting process should be evaluated in detail and not as a whole to ensure its full integration into demand response programs.

Demand response programs [10, 11], enable the active participation of small and medium players. One of the main goals is to enable this participation autonomously without harming the comfort of consumers and the efficient operation of energy services. This autonomy can be quickly achieved if the consumption forecast is analyzed in detail.

This paper focuses on studying the topic of energy consumption forecasting, by analyzing the correlation between consumption and external variable factors. The forecasting of resources is applied based on data gathered from an office building that uses a SCADA system. This SCADA system, the SCADA Office Intelligent Context Awareness Management (SOICAM), manages the consumption of the building and can also control HVAC systems [12]. SOICAM also stores the consumption data in a 10 seconds rate. The consumption data refers to different devices in separate, and, when not possible, to groups of loads (e.g. lights, HVAC, sockets), as discussed n section II. Some SCADA systems are limited regarding the collected data. They can store and acquire a huge amount of data. However, in general, this data refers to the system itself and does not include external variables to the system. In a forecasting point of view, the context of the data is important in order to understand the system. For instance, in a residential house, the data regarding the HVAC consumption is usually related to the outside temperature.

The main contribution of this paper is the development and implementation of a forecasting methodology that merges the data from the SCADA system with external and open data,

The present work has been developed under the EUREKA - ITEA2 Project M2MGrids (ITEA-13011), Project SIMOCE (ANI|P2020 17690), and has received funding from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013.

such as weather variables. In particular, the consumption from lights is associated to the solar radiation values. These different types of data are used by three distinct forecasting methods, namely, an Artificial Neural Network (ANN) [13], a Support Vector Regression (SVR) [14] and a Linear Regression Model (RM). Results achieved from this study are compared to those obtained in previous studies.

After this introductory section, section II presents the data infrastructure that supports the presented methods. Section III presents the used forecasting method, and section IV shows some achieved results. Finally, section V presents the most relevant conclusions of this work.

II. DATA INFRAESTRUCTURE

The data used in this study is measured from the campus buildings of the Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD) research center located in ISEP/IPP, Porto, Portugal. Environmental data from the same site is also used, in specific the environmental temperature [15].

The acquisition of data is performed through the SCADA (Supervisory Control and Data Acquisition) Office Intelligent Context Awareness Management (SOICAM). SOICAM is implemented in GECAD and the physical installations consist of four main spaces. Three of these are campus buildings where GECAD operates. These buildings include several offices, classrooms, kitchens, and bathrooms. The fourth place is a laboratorial controlled house. SCADA-House [16] is located in a GECAD laboratory, and contains a large set of different loads, normally used in a common house. These loads are connected to a SCADA management system, which is controlled by a software agent. Some resources are not available in our lab, making their physical integration in the system impossible. In order to overcome this limitation, OPAL-RT [17] is used to simulate resources that are not physically available. The integration between OPAL-RT and the remainder of the system is done through the Java API of OPAL-RT. Among many other resources, the OPAL-RT platform simulates wind generators making it possible to obtain outputs according to their electrical models, which can also be validated by using the platform capabilities. Additionally, OPAL-RT is also able to perform real time simulations of the components, loads and facilities that cannot be used or simulated in conventional systems. The integration of real loads in OPAL-RT is possible through the connection to software agents that represent different players in the electricity market (e.g. large consumers, large producers, and aggregators) and players connected to the distribution network (such as facilities and microgrids). This merge allows using different methods for management and control of the distribution network while the real time simulator analyses the impact of methods in the energy flows and transmission lines.

The GECAD buildings where SOICAM is implemented cover more than 30 researchers. SOICAM was implemented in June, 2014, therefore the starting of the historic of data is referent to this date. The system monitors all the consumption and generation of GECAD. The generation data (namely solar and wind based) is stored individually every 10 seconds. The consumption data is divided by three main types (Fig. 1): Illumination; Heating, Ventilation and Air Conditioning (HVAC); and electrical sockets. The consumption data is also stored every 10 seconds. All data is stored in a SQL Server database, allowing the study of consumption and generation in GECAD, as well as its management by software agents.

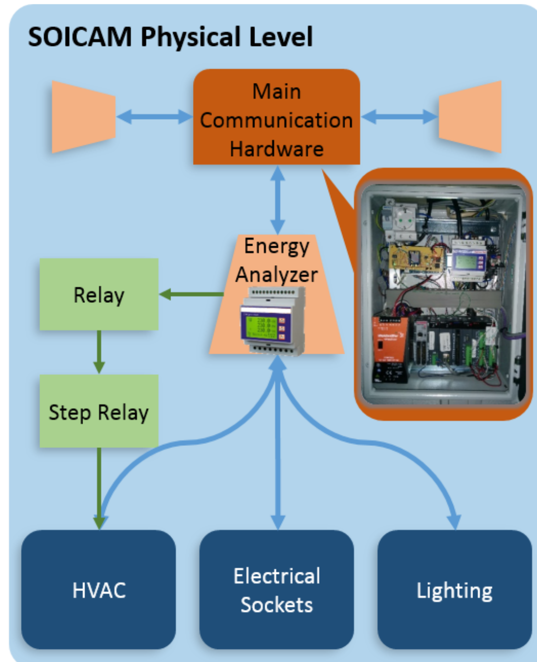

Fig. 1. SOICAM Physical Level architecture [18]

The Physical Level, in Figure 1, is composed by multiple electrical switchboards with three-phase energy analyzers. Each phase is used to measure one of the three load groups: Lighting Group; HVAC Group; and Electrical Sockets Group.

SOICAM is also able to control HVAC systems. This functionality is only available for one building, affecting 19 researchers. The possible control is only on/off for now. New hardware is being developed and implemented to allow individual load management and control. Using refined control over the load and not only on/off control, the SOICAM performance and utility is increased. The HVAC control is done using a digital auxiliary port, built in the energy analyzer, and a relay (24V/DC to 230V/AC) and a step relay in order to minimize the impact and deterioration of the relay. Each building contains a main communication hardware that can be a Programmable Logic Controller (PLC) or a custom built hardware. The main communication hardware communicates with the energy analyzers, located in several electrical switchboards, through RS-485 communication using the Modbus/RTU protocol. The time step of the measures is different for each building depending on the communication hardware (from 10 seconds to 40 seconds).

SOICAM also includes fully simulated players, which interact with software agents that control real hardware. This enables the development of a complex system capable of performing simulations with an agent society that contains both real infrastructures and simulated players, providing the means to test alternative approaches in a realistic simulation environment [12].

III. Forecasting method

The consumption forecast is carried out by means of a time series analysis. In this time series of historic data are also considered other set of parameters used to estimate consumption. Namely, the solar radiation, day of week, hour and the consumption in the previous 4 hours are the parameters that are used to predict the consumption. The selection of the parameters was carried out by means of the graphics representation, statistical test and analyzing the prediction capabilities of different methods.

The considered Artificial Neural Network (ANN) [19, 20] is characterized as a Multi-Layer Perceptron (MLP) feedforward neural network, receiving as inputs the hour, solar radiation and four input corresponding to the consumption of 4 previous hours. The output is the consumption forecast for one hour (in 1, 2, 3 or 4 hours in advance). The basis for this MLP topology has been presented in [21]. The training scheme of the MLP follows the structure presented in Figure 2.

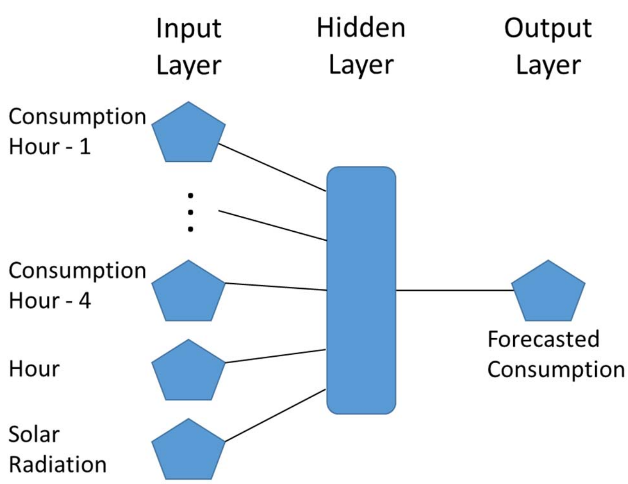

Fig. 2. Artificial Neural Network topology

Backpropagation using the gradient descent method [22] has been used as training algorithm for the ANN. This requires calculating the derivative of the squared error function with respect to the weights of the network. The squared error function E for the single output neuron is defined as in (1).

$$E = \frac{1}{2}(t - y)^2 \tag{1}$$

where t is the target output for a training sample, and y is the actual output of the output neuron.

For each neuron j, its output o_j is defined by feedforward calculation, as in (2).

$$o_j = f\left(\sum_{k=1}^{n} w_{kj} x_k\right) \tag{2}$$

where n is the number of input units to neuron j, and w_{kj} is the weight between neurons k and j. Hence, the input for the activation function f of a neuron is the weighted sum of outputs o_k of the previous neurons. The used activation function f is the logistic function, a log-sigmoid function, which can be defined as in (3) [23]. The activation function, also called the transfer function, determines the relationship between inputs and outputs of a node in a network. In general, the activation function introduces a degree of nonlinearity that is essential for most ANN applications.

$$f(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

The backpropagation algorithm is used as the training method of the designed artificial neural network. The backpropagation algorithm includes the following steps [24]:

1. Initialize weights as small random numbers;
2. Introduce training data to the ANN and calculate the output by propagating the input forward through the network using (2);
3. Calculate the error using (1)
4. Propagate the sensitivities backward through the network by simply taking the derivative of the activation function (3) with respect to the network parameters;
5. Calculate w_{kj} updates;
6. Update the values of w_{kj};

Repeat steps 2 to 6 until all examples are classified correctly.

The training of the MLP ANN also incorporates the learning rate, momentum and bias.

IV. Experimental findings

This section presents a cases study with the objective of experimenting and validating the proposed approach to forecast the electricity consumption. Besides the proposed method, forecasts are also performed using a Support Vector Regression (SVR) method, which has been presented in [25], and using a Linear Regression Method (LM). The inputs and outputs are defined in the same manner as in the case of MLP.

As a first step before making the prediction of the values, we proceeded to analyze data visually to determine the relationship between the parameters. By analyzing the variation of the lightning consumption throughout the hours, it can be seen that this relationship is not strong enough, because as can be seen in Figure 3, the consumption of lights is highly variable and there is no direct relationship between hours and consumption. For this reason, it has been determined that it would be necessary to include more parameters in the estimation. In Figure 3, each graph corresponds to a hour of day, the xx axis corresponds to the time, and the yy axis to the consumption.

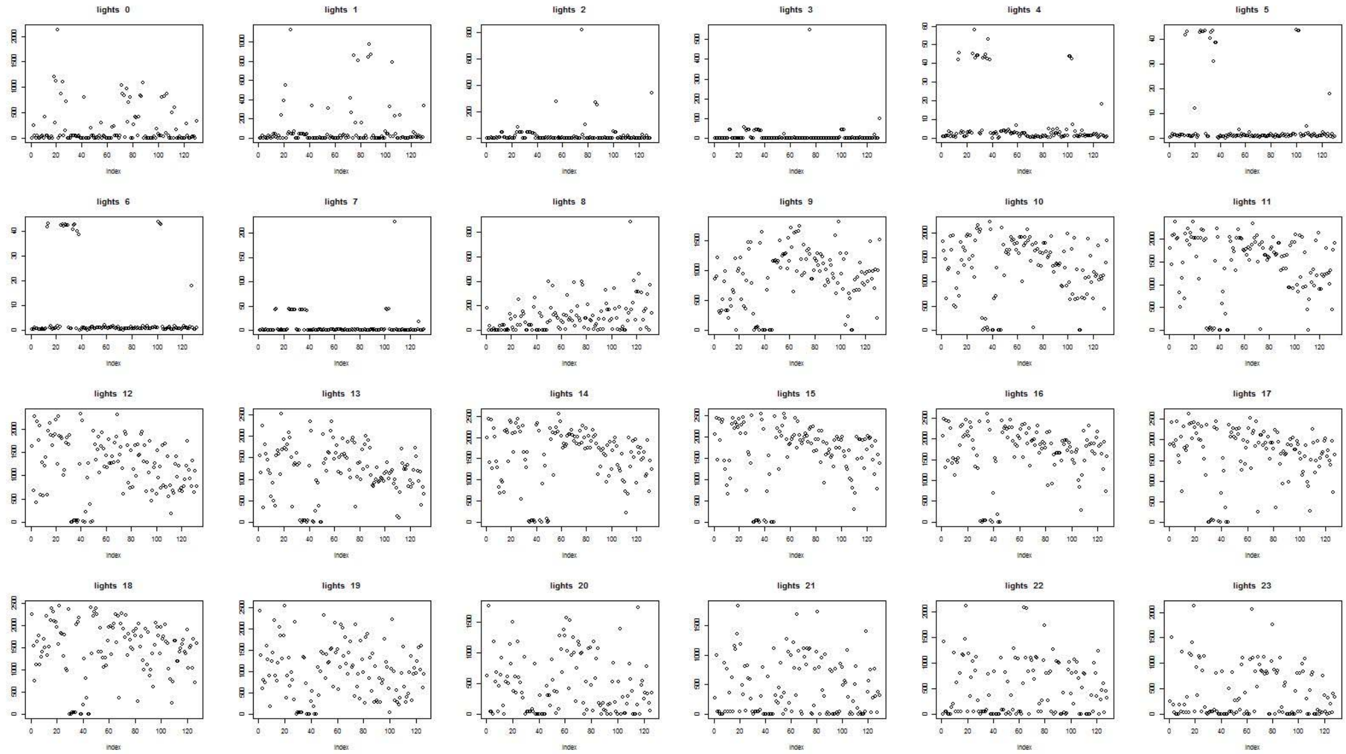

Fig. 3. Scatterplots of onsumption per hour

Additionally, we proceeded to represent the data by a diagram of boxes, in order to observe the data scattering more directly. In the boxplot of Figure 4, the *xx* axis represents the different hours and the *yy* axis represents the consumption for those hours. One can see that the consumption of lighting changes during the day. From hours 0 to 7 the consumption is close to zero. The problem is in estimating consumption for hours 9 to 23, since, as can be seen, the consumption is very variable.

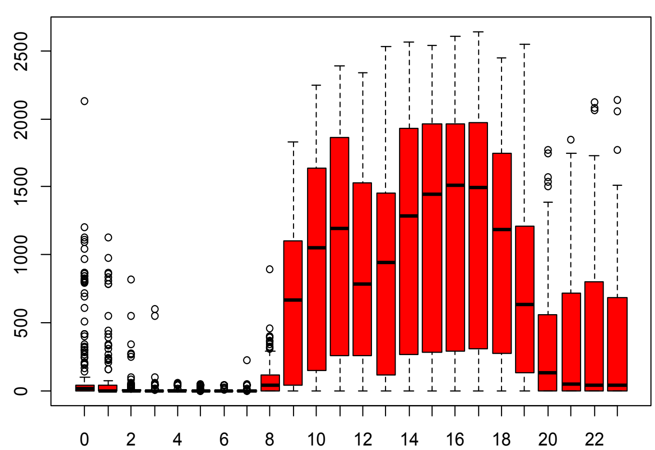

Fig. 4. Boxplots of consumption per hour

Another parameter that has been considered relevant to estimate consumption of lighting was the sunlight. In order to determine the relationship between solar radiation and the consumption, this connection has been analyzed graphically, by hour, as can be seen from Figure 5.

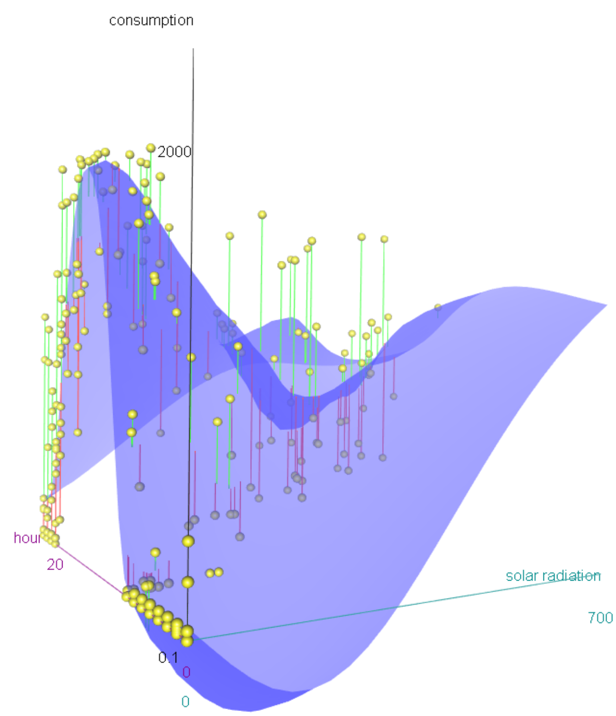

Fig. 5. Consumption according to the solar radiation and time

From Figure 5 it can be concluded that when the radiation is lower, then the consumption is higher in certain hours. For this reason, it is noticeable that this parameter may be important to collect relevant information to help the forecasting process. Thereby, it has been included in the estimation process.

After analyzing the parameters, the proposed ANN has been trained using the defined time series. As referred in section 3, the used input data are: the hour, the solar radiation and four inputs, corresponding to the consumption of the 4 previous hours. The output is the consumption forecast for one hour (1, 2, 3 or 4 hours ahead). To feed the training process, 150 patterns, corresponding to the 150 previous hours have been introduced in the neural network. Once the training process is concluded, we proceed to calculate the forecasts. In order to analyze the correct functionality of the ANN, we proceed to represent the real and predicted value as can be seen in Figure 6. In the same way, we proceed to make the training and the prediction using the SVR and the LM. The relationship between the real consumption values and the forecasted data using the three methods is shown in Figure 6.

Figure 7 shows the forecasting error, using the different methods, in different time horizon intervals.

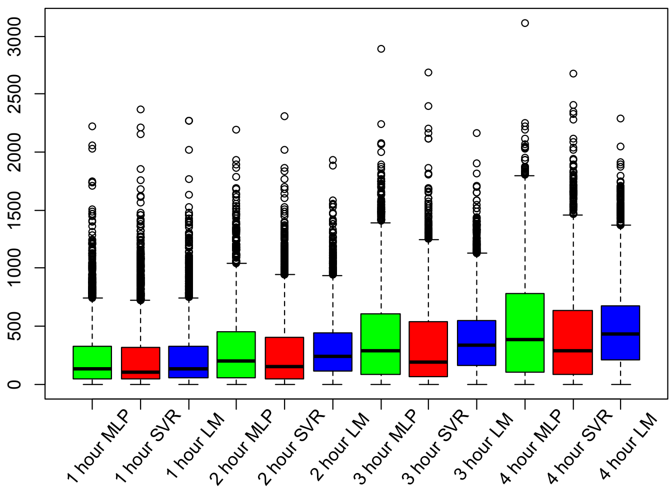

Figure 7. Boxplot with the forecast error of the consumption estimated by the three considered methods, for different time horizons

As can be seen from Figure 6, the best results are for 1 hour ahead forecast. From this interval onwards, the average error and the variance increase considerably. The values are detailed in Table 1, using the Mean Absolute Percentage Error (MAPE).

Table 1. MAPE average estimation error by the different methods for the different considered time horizons (%)

	Time-horizon			
Method	1 hour	2 hour	3 hour	4 hour
MLP	7.57	10.04	13.44	16.69
SVR	7.48	9.02	11.40	14.00
LM	7.87	10.33	13.31	16.20

Table 2 presents a Mann-Whitney test with H1: the error in the row *i* methods is lower than the error in the column *j*. With the statistic analysis presented in Table 2, we can determine that the method with better results is SVR, hence supporting the conclusions taken from Figure 7 and from Table 1.

In fact, all three considered methods are also able to achieve better results than other previously experimented forecasting approaches. The study presented in [18] has dealt with the electricity consumption forecast problem using a set of different ANN based approaches. The comparison of the MAPE forecasting error achieved by the methods considered in this paper and achieved from the methods proposed in [18], is presented in Table 3, for an hour-ahead time horizon. This table also presents a brief overview of the characteristics of these ANN, considering the different types of data that are used in the training process. Namely, Context data, which refers to: hour, day of the week and hour of work (high, medium and low permanency in the building); and External data, concerning: temperature, temperature felt, radiation, precipitation and humidity.

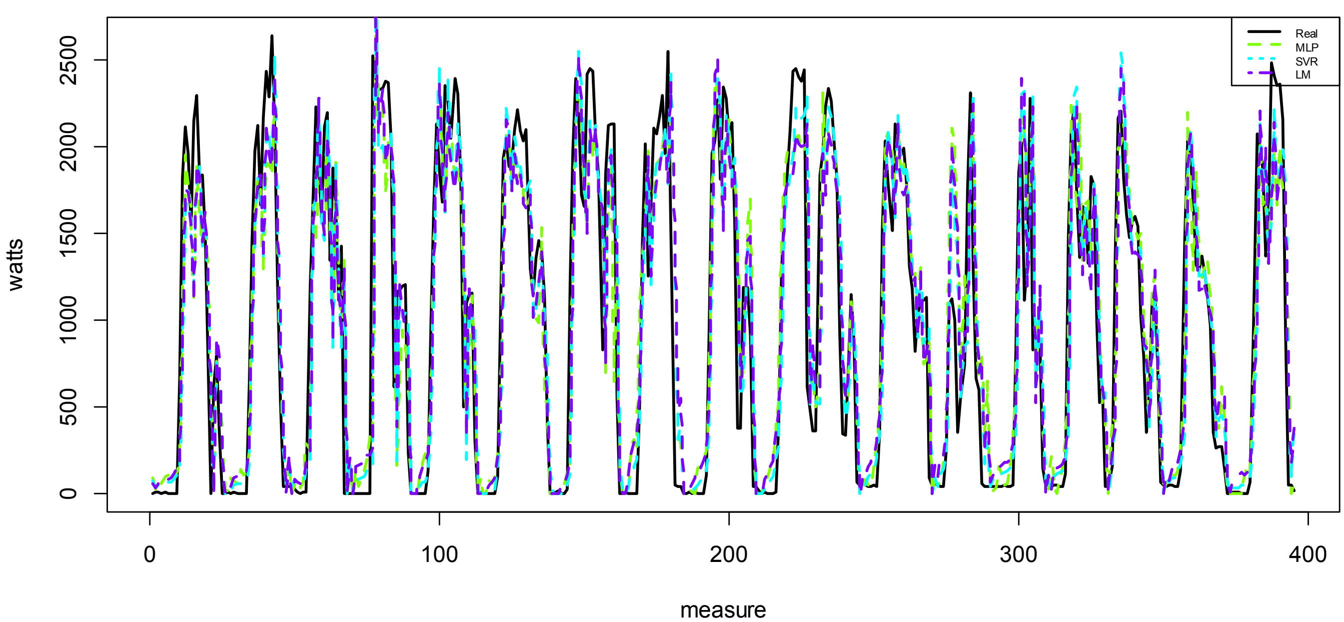

Fig. 6. Real consumption vs forecasted values, using the ANN, SVR and LM

Table 2. Mann-Whitney test results for the three considered forecasting methods

	Hour	1			2			3			4		
Hour	**Method**	**MLP**	**SVR**	**LM**	**MLP**	**SVR**	**LM**	**MLP**	**SVR**	**LM**	**MLP**	**SVR**	**LM**
1	**MLP**		0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SVR	0.02		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LM	0.99	1.00		0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	**MLP**	1.00	1.00	1.00		1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SVR	1.00	1.00	0.99	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00
	LM	1.00	1.00	1.00	1.00	1.00		0.00	1.00	0.00	0.00	0.00	0.00
3	**MLP**	1.00	1.00	1.00	1.00	1.00	1.00		1.00	0.00	0.00	0.07	0.00
	SVR	1.00	1.00	1.00	1.00	1.00	0.00	0.00		0.00	0.00	0.00	0.00
	LM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.00	1.00	0.00
4	**MLP**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00	0.00
	SVR	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00	0.00	0.00		0.00
	LM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 3 –MAPE forecasting errors of the methods (%)

Method	**MAPE(%)**	**Description**
ANN_1	16.9	Input: Contextual data Output (1): Total consumption
ANN_2	15.0	Input: Contextual data Outputs (3): Consumption of HVAC, lights, sockets
ANN_h	10.3	Input: Contextual data and external data Output: Consumption of HVAC
ANN_l	18.1	Input: Contextual data and external data Output: Consumption of lights
ANN_s	12.6	Input: Contextual data and external data Output: Consumption of sockets
MLP	**7.57**	**Input: Hour, solar radiation, consumption** **Output: Consumption of lights**
SVR	**7.48**	**Input: Hour, solar radiation, consumption** **Output: Consumption of lights**
LM	**7.87**	**Input: Hour, solar radiation, consumption** **Output: Consumption of lights**

As summarized by Table 3, all three considered forecasting methods are able to achieve better forecasting results than all the other strategies, for a hour-ahead consumption forecast. Thereby, the proposed methods prove to be an adequate solution for forecasting energy consumption in buildings.

V. CONCLUSIONS

The variability of energy consumption, especially due to a large number of incontrollable factors, such as the presence of consumers in the building, or external temperature, makes the forecasting of energy consumption an arduous task.

This paper addresses the electrical energy consumption forecasting problem, by using a data set gathered from a real installation, in real time. This physical site, located in the campus of the School of Engineering of the Polytechnic of Porto, provides a set of different types of data, which may be used for distinct analysis processes. In this paper, the correlation between the solar radiation and the electrical consumption of lights is studied. This study is performed by means of three forecasting methods, namely a multi-layer perceptron artificial neural network, a support vector regression method, and a linear regression method. These methods are fed with the historic consumption data and with the corresponding solar radiation data, in order to enable the methods to take extra information from the relationship between the data series, with the objective of improving the forecasting quality.

The achieved results are compared with those obtained by other methods, previously used to forecast the energy consumption using the same data. The comparison of these results indicated that the proposed methods are able to achieve a lower forecasting error. This allows concluding that the correlation between the consumption of lights and the solar radiation is fact, valid, and advantageous for the forecasting process.

References

[1]. Lund, H., “Renewable Energy Systems, Renewable Energy Systems – A Smart Energy Systems Approach to the Choice and Modeling of 100% Renewable Solutions”, Academic Press, 2nd Edition, May 2014

[2]. European Commission, “2030 framework for climate and energy policies”, 2014, Available: http://ec.europa.eu/clima/policies/2030/index_en.htm [accessed in March 2016

[3]. Pinto, T., Vale, Z., Rodrigues, F., Morais, H., Praça, I., "Strategic Bidding Methodology for Electricity Markets using Adaptive Learning", Modern Approaches in Applied Intelligence, vol. 6704, pp. 490-500, K. Mehrotra, et al (Eds.), Springer Berlin / Heidelberg, 2011

[4]. Pinto, T., Morais, H., Vale, Z., Praça, I., “Multi-Agent Negotiation for Coalition Formation and Management in Electricity Markets”, Negotiation and Argumentation in Multi-Agent Systems – Fundamentals, Theories, Systems and Applications, pp. 334- 360, F. Lopes, H. Coelho (Eds), Bentham e-books, 2014

[5]. Klessmann, C. et. al., “Status and perspectives of renewable energy policy and deployment in the European Union—What is needed to reach the 2020 targets?”, Energy Policy, Volume 39, Issue 12, pp.7637-7657, December 2011.

[6]. Rueda, J.L.; Guaman, W.H.; Cepeda, J.C.; Erlich, I; Vargas, A, "Hybrid Approach for Power System Operational Planning With Smart Grid and Small-Signal Stability Enhancement Considerations," Smart Grid, IEEE Transactions on, vol.4, no.1, pp.530,539, March 2013;

[7]. Nakawiro, W.; Erlich, I; Rueda, J.L., "A novel optimization algorithm for optimal reactive power dispatch: A comparative study," Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2011 4th International Conference on, vol., no., pp.1555,1561, 6-9 July 2011;

[8]. Venayagamoorthy, G.; Rohrig, K.; Erlich, I, "One Step Ahead: Short-Term Wind Power Forecasting and Intelligent Predictive Control Based on Data Analytics,"Power and Energy Magazine, IEEE , vol.10, no.5, pp.70,78, Sept.-Oct. 2012;

[9]. Fortmann, J.; Engelhardt, S.; Kretschmann, J.; Feltes, C.; Erlich, I, "New Generic Model of DFG-Based Wind Turbines for RMS-Type Simulation," Energy Conversion, IEEE Transactions on , vol.29, no.1, pp.110,118, March 2014;

[10]. Vale, Z.; Morais, H.; Ramos, S.; Soares, J.; Faria, P., "Using data mining techniques to support DR programs definition in smart grids," Power and Energy Society General Meeting, 2011 IEEE , pp.1,8, 24-29 July 2011;

[11]. Siano, P., “Demand response and smart grids—A survey,” Renewable and Sustainable Energy Reviews, vol. 30, pp. 461-478, Feb. 2014.

[12]. Gomes, L., Faria, P., Morais, H., Vale, Z., Ramos, C., “Distributed Agent-Based Intelligent System for Demand Response Program Simulation in Smart Grids,” IEEE Intelligent Systems, vol. 29,pp. 56-65, Jan/Feb 2014.

[13]. Pinto, T., Sousa, T.M., Vale, Z., "Dynamic Artificial Neural Network for Electricity Market Prices Forecast", IEEE 16th International Conference on Intelligent Engineering Systems (INES 2012), Costa de Caparica, Portugal, 13-15 June, 2012

[14]. Pinto, T., Sousa, T.M., Praça, I., Vale, Z., Morais, H., “Support Vector Machines for decision support in electricity markets' strategic bidding”, Neurocomputing 172: 438-445 (2016)

[15]. ISEP meteo website, available: http://meteo.isep.ipp.pt. Accessed on July 2016

[16]. Fernandes, F., et.al “Dynamic load management in a smart home to participate in demand response events”, Energy and Buildings, vol. 82, pp. 592-606, October 2014

[17]. OPAL-RT Technologies, “OP5600 off-the-shelf Hardware-in-the-Loop (HIL) simulator

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

[19]. Hao, Q., Srinivasan, D. and Khosravi, A., "Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals," IEEE Trans. on Neural Netw. and Learn. Syst., vol. 25, no. 2, pp. 303-315, Feb. 2014.

[20]. Alvarado-Pérez, J.C., Peluffo-Ordóñez, D.H., Therón, R., 2015. Bridging the gap between human knowledge and machine learning. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. 4, 54.

[21]. De Paz, J.F., et al., Mitigation of the Ground Reflection Effect in Real-Time Locating Systems Based on Wireless Sensor Networks by Using Artificial Neural Networks. Knowledge and Information Systems. 34(1), 193-217, 2013.

[22]. Rudd, K., Di Muro, G. and Ferrari, S., "A Constrained Backpropagation Approach for the Adaptive Solution of Partial Differential Equations," IEEE Trans. on Neural Networks and Learning Systems, vol. 25, no. 3, pp. 571,584, Mar. 2014.

[23]. Rumelhart, D. E; Hinton G. E; Williams, R. J., ”Learning representations of back- propagating error” London: Nature, v. 323, pp. 533-536, 1986

[24]. Pinto, T., et al “Adaptive Portfolio Optimization for Multiple Electricity Markets Participation” IEEE Transactions on Neural Networks and Learning Systems 27(8): 1720-1733 (2016)

[25]. De Paz, J. F., Bajo, J., González, A., Rodríguez, S. and Corchado, J. M., “Combining Case-based Reasoning Systems and Support Vector Regression to Evaluate the Atmosphere-Ocean Interaction”, Knowledge and Information Systems, 30, 155-177, 2012.