Comparison of Genetically Optimized Short-Term Price Forecasting Models for the Iberian Electricity Market

A. Yanguas-Peña, L.A. Fernández-Jiménez, P. Zorzano-Santamaría, E. Zorzano-Alba, E. Martínez-Prado

Department of Electrical Engineering
E.T.S. Ingeniería Industrial, University of La Rioja
Luis de Ulloa 4, 26004 Logroño (Spain)
phone:+34 941 299477, fax:+34 941 299478, e-mail: alberto.yanguas@alum.unirioja.es, luisalfredo.fernandez@unirioja.es, pedrojose.zorzano@unirioja.es, enrique.zorzano@unirioja.es

Abstract. The hourly price for the electric energy that will be settled in a day-ahead market constitutes very valuable information if it could be known in advance by the agents (producers, retailers and large consumers) operating in that market. This paper presents the comparison of the results obtained with a set of short-term electricity price forecasting models applied to the day-ahead hourly price forecasts in the Iberian Electricity Market. The studied set include artificial neural networks, adaptive neuro-fuzzy inference systems and support vector machines. The structure of the three kind of forecasting models were optimized by means of a genetic algorithm which also selected the input variables used by the forecasting model among a set of available input variables. The forecasting results obtained for an out-sample data test are quite similar for the best models of each kind, but with a slight better performance of the adaptive neuro-fuzzy system.

Key words

Price forecasting, electricity market, fuzzy logic, support vector machines, genetic algorithms.

1. Introduction

The deregulation process carried out in the electric industry and the introduction of competitive markets have changed the monopolistic and government-controlled power sectors. In most of the developed countries electricity is now traded under market rules although its singular characteristics make it different from other commodities markets [1].

The development of short-term electricity price forecasting (STEPF) models has been a very active research field in the last 15 years because the hourly price for the electric energy that will be settled in the pool constitutes very valuable information if it could be known in advance: any agent involved in the electricity market could use the forecasts to prepare his/her bids strategically in order to obtain the maximum profit. An accurate price forecast for an electricity market has a definitive impact on the bidding strategies and even on the price negotiation of bilateral contracts [2].

One of the biggest problem we have faced in order to design our STEPF models is the extreme price volatility and the remarkable price spikes in the Iberian electricity market, as it is showed in the Figure 1. The high proportion of outliers (unusual prices) is a consequence of a lesser degree of competition in the Iberian Electricity Market (compare with other markets), which in turn makes this market less predictable.

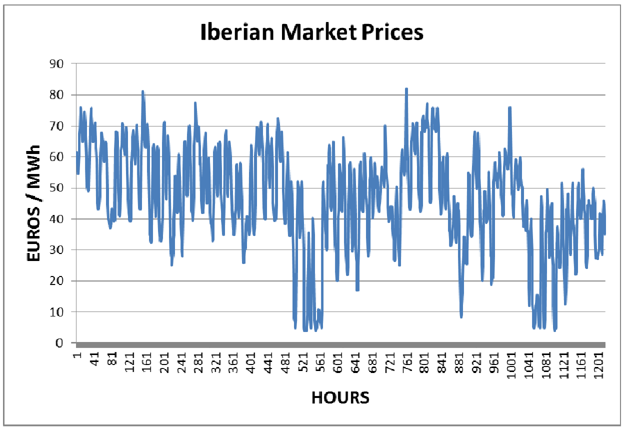

Fig. 1. Hourly prices for Janurary- February of 2015.

Tens of STEPF models have been described in the international literature [1]. The used techniques include traditional time series ones as autoregressive integrated moving average (ARIMA) [2], or artificial intelligence based ones as artificial neural networks (ANNs) [3, 4] and fuzzy inference systems (FIS) [5]. In general, most of the published papers are focused on the description of the forecasting technique. Only a few of published works include the analysis of the explanatory variables used to build the forecasting models [6].

This paper presents the comparison of the results obtained with a set of STEPF models applied to the day-ahead hourly price forecasts in the Iberian Electricity Market (MIBEL). The compared models include multilayer perceptron (MLP) neural networks, adaptive neuro-fuzzy systems (ANFIS), and support vector machines (SVM). The structure of the studied models is optimized by means of the application of a genetic algorithm which allows the selection of the input variable from the set of available ones (feature selection) and the value of the parameters that define the forecasting model.

2. Available input variables

The day-ahead hourly price forecasting can be influenced by different kinds of explanatory variables [6]:

1) Actual recorded hourly electricity prices.
2) Chronological variables: hour, week day, holiday, week number and month number.
3) Actual recorded hourly demands and hourly power generations aggregated by generation type.
4) Hourly weather forecasts, including wind speed, solar irradiance and temperature.
5) Power system hourly variable forecasts: power demand forecasts, wind power forecasts, solar power forecasts, hydropower forecasts, independent cogeneration forecasts, thermal power forecasts, etc.
6) Power market restriction variables: unavailable capacity for power generation, reserves of power generation and interconnection, volume of electric energy allocated in other electricity markets, and electricity futures market and bilateral contracts.

From the set of the above mentioned variables, in our work we have used the first three sets and the fifth one. The data corresponding to the actual recorded electricity prices were downloaded from the market's operator (OMIE) and the recorded demands and power generations were downloaded from the ENTSOE website (European Network of Transmission System Operators for Electricity) as well as the information regarding to the forecast power demands, and solar and wind energy production forecast for the two countries involved in the MIBEL, Spain and Portugal. We have not used in the work described in the paper weather forecasts such as temperature, radiation, wind speed, etc, since these forecasts has been already considered by the own transmission system operators (REE in Spain and REN in Portugal) to predict the demand and the wind power and solar power generations in their systems.

All the data used to build and compare the STEPF models correspond to the year 2015 and the first three months of 2016. Some of the records for this period weren't complete: only 9971 hourly records including values for all the input variables were available.

3. Studied short-term price forecasting models

Three families of artificial intelligence based models have been chosen in our comparative study: MLP neural networks, ANFIS and SVM for regression. The developed models, for each one of the families, were optimized: the structure and the input variables used are selected by means of a genetic algorithm. So, the number of neurons in the hidden layers of the MLPs, the number of membership functions and rules for the ANFIS, or the parameters defining the support vector regression model (cost, epsilon and gamma parameters) are elected according to the optimization process. Also the input variables used by the models are chosen from the set of available input variables.

A. Genetic algorithms

Genetic algorithms (GAs) can fit inside the class of stochastic search methods [7]. While most of these methods operate on a single solution, GAs operate on a population of solutions. The basic idea, inspired by biological evolutionary processes, is that the genetic content of a population potentially contains the solution, or a better solution to a given problem of adaptation. The application of GAs is mostly focused on large, complex and little understood search spaces. The basic idea of a GA is as follows: initially generate a set with some of the possible solutions or individuals, called the population; once the initial population is created, it will evolve according to the evaluation of each individual, obtaining new individuals (generations) associated with better solutions.

The information associated with an individual is composed of indivisible parts called chromosomes. Each chromosome has many genes, which correspond to separate parameters of the problem. To work with these genes in a computer program it is necessary to encode them in a numerical or alphanumeric string. During the evaluation the gene is decoded and becomes a series of parameters of the problem whose solution is intended to optimize. The solution is then achieved using these parameters and a score is calculated depending on how that solution is close to the best solution. This score is called fitness. As soon all the individuals from a population are evaluated, they reproduce using the genetic operators according their scores. The genetic operators allow the exchange of genetic material (genes) between individuals (crossover) or the spontaneous change of one or more genes in an individual (mutation). The result of the application of the genetic operators is the creation of a new population (new generation). The general performance of a GA is shown in Figure 2.

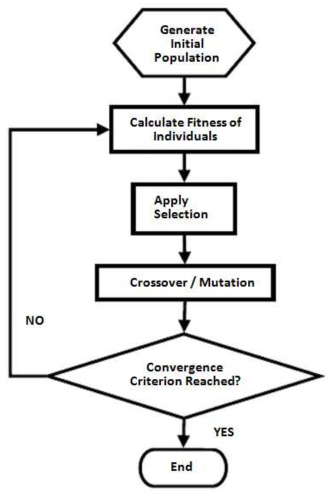

Fig. 2. Flowchart of a general Genetic Algorithm.

Two important parameters for the GA are:

- Population size: it must be sufficient to ensure the diversity of solutions and have to grow more or less with the number of genes in the chromosome.
- Termination condition: the most common are the achievement of a fitness value or the ending of a preset number of generations. The latter will be the termination condition in our study.

A GA based optimization process was carried out with all the studied models of this paper in order to obtain the best STEPF model corresponding to each family. The process selects the input variables used by the model and the parameters that defined each model.

B. Adaptive neuro-fuzzy systems

Fuzzy inference techniques provide a method for fuzzy modeling process. In addition, once a fuzzy inference system is built, it can be trained in a similar manner to artificial neural networks, constituting an adaptive neuro-fuzzy system (ANFIS) [8]. In an ANFIS, the training process adapts the parameters of membership functions and fuzzy rules, and the weights of each rule obtaining a better match between the output of the ANFIS and the desired output. The training process of an ANFIS is carried out using back propagation and least squares techniques, which give the system the ability to learn.

In order to reduce the number of rules of the inference system, we used subtractive clustering (SC) [9]. SC is a technique that estimates the centers of groups in a set of data. SC assumes that each point-data is a center of a potential group and it is assigned a potential based on the density of points (data) that surround it. The algorithm selects the point-data with the greatest potential as first group, and then deletes the potential of the points-data close to that first group center. Then, the algorithm selects the point-data with the largest potential remnants as the next center of group and returns to destroy the potentials of data in its area of influence. And so on.

The iterative cycle ends when the potential of all points falls below a threshold. The SC algorithm needs four parameters: the value of the radius that defines the influence (or neighborhood), the value of the radius that defines the area of reduction of potential when a center is chosen, a factor that represents the threshold for acceptance as a center (are accepted those possible centers with potential superior to the first center multiplied by this factor), and another factor that represents the threshold of rejection as a center.

C. Support Vector Machines

SVMs are a set of related methods for supervised learning, applicable to classification and regression problems. SVMs are able to learn with only a small number of free parameters, robust against outliers, and computational efficient [10]. In recent years SVMs have been shown as a powerful modelling tool used for many machine learning tasks and has been widely applied in different research fields. Although SVMs were initially proposed for classification problems, they were generalized for regression problems using the ε-intensive loss function [11]. The underlying idea in the SVM regression model is to find a function that presents at most ε deviation from the target values while being as flat as possible. Under that methodology, a SVM regression model can be defined by three parameters: ε, which define the radius of a tube around the regression function in which errors are ignored, C (called as cost) which represents a tradeoff between the prediction error and the tube's flatness, and γ is the width of the Gaussian function used as kernel. Parameters ε, C and γ account for a significant effect on the SVM regression model, so their values must be carefully chosen.

D. Multilayer perceptrons.

Multi-layer perceptrons (MLPs) are the most widely used network architecture in artificial neural networks [12]. A MLP consists of multiple layers of neurons (processing elements) with each neuron fully connected to all the neurons in the following layer. A MLP comprises at least three layers: an input layer, a hidden layer, and an output layer. A MLP network with a hidden layer is able to approximate any smooth nonlinear input-output mapping with an arbitrary degree of accuracy if a sufficient number of neurons is used in the hidden layer, and it is able to match the input-output mapping using two or more hidden layers [13].

In this work we have used two hidden layers MLPs trained with the back-propagation with momentum method [12]. The transfer function for the neurons of the hidden layer was the hyperbolic tangent function and the linear function for the neuron of the output layer. The number of neurons in the hidden layers, the learning factor and momentum for each layer and the selection of the input variables must be properly chosen in order to obtain the best forecasting model.

4. Available input variables

One of the most important tasks in developing a model, is the selection of relevant input variables. Unfortunately, there is no systematic method to follow. However, an acceptable practical solution is the iteration process of trial and error, where some new variables are added or other irrelevant are subtracted for a better model. In this context, the theory of linear regression can provide relevant information. So, automatic pruning methods are proposed that make it possible, starting from a model containing all possible input variables, those irrelevant are discarded by a sensitivity analysis.

A different approach is the one that has been followed in this work. The optimization process was controlled by a GA, which selected the input variables used for the model being optimized from all available input variables. The available input variables, related to the MIBEL electricity market, were the following:

E1.- Day.
E2.- Month.
E3.- Weekday (monday, tuesday…).
E4.- Hour.
E5.- Biomass energy generation (D-1).
E6.- Fossil Coal energy generation (D-1).
E7.- Fossil Gas energy generation (D -1).
E8.- Fossil oil energy generation (D -1).
E9.- Hydro energy generation (D-1).
E10.- Nuclear energy generation (D-1).
E11.- Other non-renewable energies (D-1).
E12.- Solar energy generation (D-1).
E13.- Wind energy generation (D-1).
E14.- Total energy generation (D-1).
E15.- Total energy generation (D-6).
E16.- Global energy generation forecast (D+1).
E17.- Scheduled Consumption forecast (D+1).
E18.- Renewable energy generation forecast (D+1).
E19.- Solar energy generation forecast (D+1).
E20.- Wind energy generation forecast (D+1).
E21.- Market price (€ / MWh) (D-1).
E22.- Market price (€ / MWh) (D-6).

All the STEPF models developed in this work presented as output the marginal price, in Euros/MWh, fixed by the market for the hour h of the day D+1. The forecast is supposedly carried out in the first hours of the day D and, since the marginal prices for all the hours of that day were fixed in the previous one (D-1), all those prices are known, as well as the prices fixed for the same day to the forecasted one in the previous week (day D-6). The inputs E5 to E11 corresponded to the energy generated at the hour h in the previous days. The input variables E16 to E20 corresponded to forecasted values for the hour h of the day D+1, but they are known in the day D.

5. Model performance evaluation

To develop the STEPF models described in this paper, we utilized recorded values for the MIBEL corresponding to the year 2015 and the first three months of the year 2016. The recorded data were divided into an in-sample data set used for training, and an out-sample data set used for testing and comparing the forecasting models. The out-sample data set was composed by complete weeks extracted along the two years of data in order to have a good representation of the different price behaviors along the year. The in-sample and out-sample data sets were defined as follows:

- In-sample data set: all the hours of the days in 2015 and three first months of 2016 except those included in the out-sample data set, totalizing 7909 cases (hours).
- Out-sample data set: all the hours of the weeks with numbers 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 in 2015, and weeks numbers 2, 7, and 12 in 2016, totalizing 2062 cases (hours).

Several measurements are used to examine the accuracy of the forecasting results. The most common measurement used to evaluate the performance is the mean absolute percentage error (MAPE). MAPE represents the average of the relation between the absolute prediction error and the real value, as it is expressed in (1)

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|Y_t - P_t|}{Y_t}100 \qquad (1)$$

where N represents the total of cases (hours), t an hour index, Y_t the real value for the hour t and P_t the forecasted one.

Another common measurement is the root of the mean square error (RMSE), expressed in (2)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y_t - P_t)^2} \qquad (2)$$

In general, for forecasting models of electric variables, the MAPE is a good performance indicator, however used in STEPF models can be disorienting: if the actual price value is small, as usually happen during night hours, the MAPE will increase considerably even if the difference between actual and forecasted values is small. In addition, if the forecasted value is small and actual value is large, the absolute percentage error will be near 100%. In order to avoid the adverse effect of very small prices, another error indicator called AMAPE was defined in the international literature (3). It is very similar to MAPE, but the divisor is the mean of the actual value on a daily basis (T=24), weekly basis (T=168), etc.

$$AMAPE = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{|Y_t - P_t|}{\left(\frac{1}{T}\sum_{t=1}^{T}Y_t\right)}\right)100 \qquad (3)$$

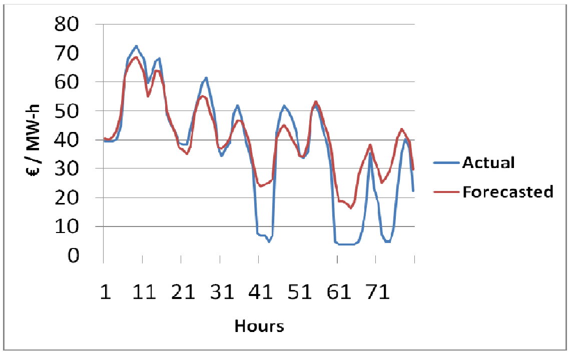

Fig. 3. Actual and forecasted available prices for four days in the out-sample data set.

As an example, in figure 3 are shown the actual and forecasted price available values for four days belonging to the out-sample data set. The STEPF model adequately follows the actual market behavior in the first 38 hours, but when an abrupt change in the actual price takes place (hours 38 to 44 in the figure), the forecasted values don’t follow the actual values correctly, shooting considerably the measured error. Hence, the importance of AMAPE indicator to evaluate the STEPF models [14].

6. Adopted solution

The MLP STEPF models were built using the software Neurosolutions [15]. This software includes the optimization process ruled by GA. The in-sample data set was randomly divided into two subsets, the first one with 75% of the data as the training data set, and the second subset with the remaining 25% as a cross-validation data set. The MLP to be optimized was defined with two hidden layers, with the selection of any of the available input variables, and with the selection of the learning factor and momentum factor for the two hidden layers and the output layer of the neural network. The number of epochs for each network was fixed in 2000, although the training process stopped when the error began to increase with the cross-validation data set. The fitness of the GA was inversely proportional to the mean square error with the data of the training data set. The final MLP obtained (best STEPF model with a MLP structure) had 35 neurons in both hidden layer and used 17 of the available input variables, as it is shown in Table I. Figure 4 shows the mean square error with the training data set obtained with the best MLP model in each generation: the optimal model was obtained approximately in the 17th generation, and although the evolution of the GA continued up to 50 generations, no better models were obtained.

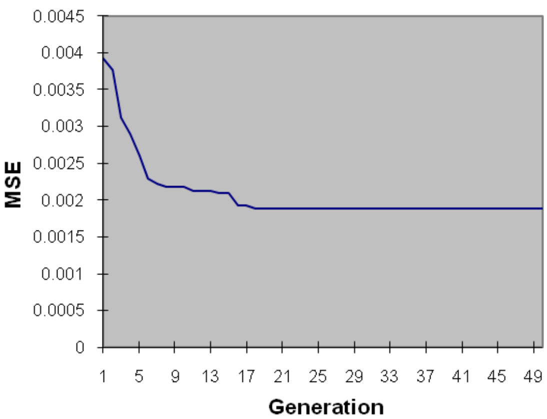

Fig. 4. Mean square error for the training data set for the best MLP model in each generation.

The ANFIS STEPF model was built using the software Matlab [16]. The optimization process was programmed and the ANFIS models were built and trained with functions of the fuzzy library. Initially all the available input data were normalized to values between 0 and 1 in order to apply the SC technique, which means that the range of values of all the data in each dimension (input in this case) was the same. The only parameter taken into account in the optimization process was the radius of influence defined for the SC, which could take values between 0 and the square root of the number of inputs used by the model. For this algorithm were taken constant threshold values of acceptance and rejection, with values of 0.15 and 0.5, respectively. The radius that defines the reduction potential was fixed in 25% greater than the radius of influence.

The structure used for the chromosome, which contains all the information necessary for the creation of the ANFIS model was divided into two parts: the first part is composed of 22 binary digits (with values 0 or 1) corresponding each one to the available input variable with the same index (the *n* digit represented to the input variable E*n*); the second part was a string with eight digits each one with values from 0 to 9. In the binary part, a digit with value 1 meant that the corresponding input variable was used by the model, otherwise a value 0 meant that the corresponding input variable wasn't used by the model. The last 8 digits of the chromosome (second part) represented the value of the radius of influence for the SC algorithm to be applied to the data corresponding to the selected (with the previous binary digits) input variables. The numerical value of this radio was obtained by multiplying the numeric value represented by those 8 digits by the square root of the number of the selected input variables and 1e-8. Thus, if the chromosome contains the numeric string "011000100000010011001122371901" it meant that the input variables selected for the ANFIS model were E2, E3, E7, E14, E17, E18, E21 and E22 and the value of the radius of influence for the subtractive algorithm was 0.63277292. The number of individuals and the number of generations was fixed in 50, and we applied elitism (the best individual was copied from the previous generation), obtained the 90% of the individuals of the new generation applying the crossover operator and the remaining to complete the population with mutation. The fitness function was inversely proportional to the RMSE obtained with the training data set. The characteristics of the best ANFIS model obtained after the optimization process are reflected in Table I and Table II.

Lastly, we used a very similar approach in the optimization of the SVM regression model. In this case we used the software R [17] with a library called "e1071". The structure of the chromosome was similar to the used for the ANFIS model, although in this case the second part of the chromosome had 24 digits, 8 for each one of the parameters to be optimized, ε, C and γ. In order to build the SVM model, the eight digits containing the information for the ε parameter were multiplied by 1e-8 (what lead to values in the range 0 to 1), multiplied by 1e-5 (values in the range 0 to 1000) for the parameter C, and multiplied by1e-7 (values in the range 0 to 10) for the parameter γ. The number of individuals and the number of generations was fixed also in 50, we applied elitism, obtained the 90% of the individuals of the new generation applying the crossover operator and the remaining to complete the population with mutation. The fitness function was inversely proportional to the RMSE obtained with the complete in-sample data set using 5-folds cross validation. The characteristics of the best SVM regression model obtained after the optimization process are reflected in Table I and Table II.

All the models had the same output variable, the forecast of the hourly electricity price settled by the MIBEL (Spanish price) corresponding to the hour h of the next day. Since one of the available input variables include information about the hour corresponding to the forecasting horizon, the same model can be used for all the hours of the day.

TABLE I.- Variables selected for the STEPF models after the optimization process.

Model	Selected input variables
MLP	E1, E2, E3, E4, E6, E8, E9, E10, E11, E14, E15, E16, E18, E19, E20, E21, E22 (17 inputs)
SVM	E1, E3, E4, E5, E6, E7, E8, E9, E11, E15, E16, E17, E18, E19, E20, E21, E22 (17 inputs)
ANFIS	E1, E3, E4, E6, E8, E10, E11, E13, E17, E18, E20, E21 (12 inputs)

TABLE II.- Parameters of the STEPF models obtained after the optimization process.

Model	Parameters
MLP	35 neurons in both hidden layers. Different learning factors and momentun factors for the two hidden layers and the output layer.
SVM	ε: 0.07413692 C: 880.34909 γ: 0.1094617
ANFIS	Radius of influence: 0.29854766 (what led to four membership functions and 4 rules)

Finally, the tree best STEPF models, one per each one of the studied families, were used to forecast the data corresponding to the out-sample data set. The forecasting results are presented in Table 3, where we can appreciate a better performance of the ANFIS based model, with a RMSE and AMAPE indicators with slightly lower values that those corresponding to the SVM regression model.

TABLE III. Forecasting results with the out-sample data set.

Model	RMSE (€/MWh)	MAPE	AMAPE
MLP	6.16	19.14%	10.26%
SVM	5.77	17.12%	9.59%
ANFIS	5.76	17.91%	9.49%

7. Conclusions

This paper presents the comparison of the results obtained with three models for electricity price forecast. The models were applied to the MIBEL. The models correspond to the best one build using three popular families: MLP neural networks, ANFIS, and SVMs for regression. The STEPF models were optimized by means of a genetic algorithm which allowed the selection of proper value of the parameters that define the model (number of neurons, number of membership functions, or parameters of the support vector regression model) and the selection of the input variables used by the model chosen from the set of available input variables. The available input variable included past days hourly electricity prices, chronological information, generation forecasts and past values of load demand and power generations in the region covered by the MIBEL (mainland of Spain and Portugal).

All the models achieved satisfactory forecasting results when applied to the real life data, with AMAPE error around 10%, allowing their use by any agent involved in the MIBEL. The STEPF model with the lowest AMAPE error for the out-sample data set was the ANFIS model, but with a very close value to the obtained by the best SVM regression model.

References

[1] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future”. Int. J. Forecast, Vol. 30, pp. 1030–1081, 2014.

[2] J. Contreras, R. Espínola, F.J. Nogales, and A.J. Conejo, “ARIMA models to predict next-day electricity prices”, IEEE Trans. Power Syst, Vol. 18, pp. 1014–1020, 2003.

[3] N.A. Shrivastava and B.K. Panigrahi, “A hybrid wavelet-ELM based short term price forecasting for electricity markets”, Int. J. Electr. Power Energy Syst, Vol. 55, pp. 41–50, 2014.

[4] C.H. Jin, G. Pok, Y. Lee, H.-W. Park, K.D. Kim, U. Yun and K.H. Ryu, “A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting”, Energy Convers. Manag, Vol. 90, pp. 84–92, 2015.

[5] G.J. Osório, J.C.O. Matias and J.P.S. Catalão, “Electricity prices forecasting by a hybrid evolutionary-adaptive methodology”, Energy Convers. Manag, Vol. 80, pp. 363–373, 2014.

[6] C. Monteiro, L.A. Fernandez-Jimenez and I.J. Ramirez-Rosado, “Explanatory information analysis for day-ahead Price forecasting in the Iberian electricity market”, Energies, Vol. 8, pp. 10464-10486, 2015.

[7] D.E. Goldberg, “Genetic algorithms in search, optimization and machine learning”, Adisson-Wesley, Reading, Massachussets, 1989.

[8] J.S.R. Jang, C.T. Sun and E. Mizutani, “Neuro-Fuzzy and Soft Computing”, Prentice Hall, Englewood Cliffs, NJ, 1997.

[9] S. Chiu, “Fuzzy model identification based on cluster estimation”, Journal of Intelligent and Fuzzy Systems, Vol. 2, pp. 267-278, 1994.

[10] I. Steinwart, A. Christmann, “Support Vector Machines”, Springer, New York, NY, 2008.

[11] V.N. Vapnik. “The nature of statistical learning theory”, Springer, New York, NY, 1995.

[12] S. Haykin. “Neural Networks. A Comprehensive Foundation”, Prentice Hall, New Jersey, 1999.

[13] H. White, “Artificial Neural Networks: Approximation and Learning Theory”, Blackwell, Cambridge, MA, 1992.

[14] L. Wu, M. Shahidehpour. “A hybrid model for day-ahead price forecasting”, IEEE T. on Power Systems, Vol. 25, pp 1519-1530, 2010.

[15] http://www.neurosolutions.com/

[16] http://es.mathworks.com/products/matlab/

[17] https://www.r-project.org/