Analysis of Different Testing Parameters in Self-Organizing Maps for Short-Term Load Demand Forecasting in Spain

S. Valero *IEEE Member*(1), J. Aparicio(2), C. Senabre(1), M. Ortiz, *IEEE Student Member*(1), and J. Sancho
Department of Industrial Systems Engineering (1) and ORC(2)
Universidad Miguel Hernández de Elche.
Edif. Quorum-V, Elche, Spain.
svalero@umh.es

A. Gabaldon *IEEE Member.*
Department of Electrical Engineering.
Universidad Politécnica de Cartagena.
Cartagena, Spain.

Abstract— **Short-term forecasting is required by utility planners and electric system operators for tactical operational planning and day-to-day decision making. The forecasting is intended to obtain the system load demand over a period of hours or days, and it plays an important role in determining unit commitment, spinning reserve, economic power interchange, load management etc... Electrical load has a time-varying nature, and it is affected by various factors such as social, meteorological and financial conditions. Since the time horizon is relatively small in short-term load forecasting, the social and financial conditions have almost no influence on the forecasting process. In this research, the authors analyze the capability of a neural network, such as Self-Organizing Maps (SOM) for short-term load forecasting. The input data used concern the global load demand of Spain over several years and were obtained thanks to the Electrical Spanish System Operator: *Red Eléctrica Española* (REE). The study was focused on testing the first hours of the day to be forecasted in order to identify its common patterns with the historical database previously trained by the neural network. The input data has to be analysed beforehand to normalize them and filter anomalous days and holidays. Weekends were also excluded as their patterns are completely different to the rest of the week. After several simulations with different training parameters, three distinct tests were accomplished with the first 8, 10 and 12 hours of the day to be forecasted, and the errors between them were compared. Even in the case of 8 hours, the results show how the Self-Organizing Map is able to associate the evolution of the day with the most similar patterns in the database. The error in the case of the 10 hour test is lower and it reaches a minor value of around 1.6% when the test is carried out with 12 hours. Further tests will allow to select the best range of hours for creating the input data and will take into account the advantage of previously classifying the data depending on the season (summer, winter …).**

Keywords: Short-term Load Forecasting, Self-Organizing Map (SOM), Electricity Market, Load Estimation.

The work described in this paper is being supported by the Valencian Government, Department of Education under Research Project GV/2010/080. Directed by Dr. Sergio Valero Verdú

I. INTRODUCTION

Accurately forecasting load plays a very important role for electric utilities within a competitive environment created by the electric industry deregulation. An electric company is confronted with many financial and technical problems concerning the operation, planning, and control of an electric energy system as customers require high quality electric energy to be supplied in a secure and economic manner [1]. Load forecasting helps those running an electric utility to make important decisions on generating, interchanging, and purchasing electric power, load switching, and infrastructure development. Besides, load forecasting is crucial for energy suppliers, financial institutions, other organizations and institutions involved in electric energy generation, transmission, distribution, and markets [2]. Previous researches on electrical demand and inherent factors in different countries [3] and [4] allow to identify a group of basic variables explaining the evolution of the electrical demand. Load forecasting can be divided into three main categories: short-term forecasting, medium-term forecasting and long-term forecasting. In this work the authors focus on short-term forecasting. This category usually makes forecasts from one hour to several days. On the other hand, medium-term forecasting concerns the future electric load from a week to a month, and long-term forecasting often predicts the load of a one year or even longer. The short-term forecasting can be used for controlling and planning power generation, and also as input to study load-flow or carry out a contingency analysis. The load is a dynamic system [5] which is mainly affected by two factors: time of the day and weather conditions. The load's time dependence reflects the existence of a daily load pattern, which may vary for different weekdays and seasons. Among weather variables, temperature is usually the dominant weather factor influencing the load. For short-term load forecasting a variety of methods using statistical techniques or artificial intelligence algorithms, which include regression models, time series, neural networks, statistical learning algorithms, fuzzy logic, or expert systems have been developed. These methods have all succeeded in solving short-term load forecasting problems. The success of a forecasting technique not only depends on the approach but also on the quality of input data which could contain proper patterns representing the system dynamics. In general, the load presents two distinct patterns: weekday and weekend load patterns. Weekday patterns include Tuesday to Friday and weekend patterns include Sunday to Monday. In addition, holiday patterns are different from non-holiday patterns. However, in this work only weekday load patterns and a previous filter have been considered.

II. CASE OF STUDY

Historical data on real global load demand were used for the research. They were presented to the SOM as daily load value, in hours, from the workdays of the years 2002 and 2003, (see Fig. 1). The load curves were obtained from the Spanish Electrical System Operator [6] web page. For testing purposes, 2004 data were also collected. The main objective of the research is to use the capacity of SOM maps to classify historical data, followed by taking advantage of the memorization of this classification [7] to identify similarities between the trained map (years 2002 and 2003) and the first demand hours of a new day corresponding to year 2004. Certain pre-processing of the input data is needed in order to obtain good results as the demand evolves over the years. It is necessary to apply a small increase to the input data with the aim of expressing estimated demand growth in 2004. Vectors are also normalized using the maximum value of known demand.

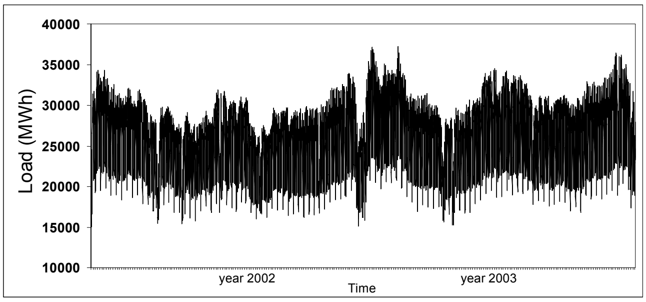

Figure 1. Global Load Demand (years 2002 & 2003).

III. SELF-ORGANIZING MAPS

A Self-Organizing Map (SOM) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional, discretized representation of an input space of training samples in a cell map. This methodology was introduced by Kohonen two decades ago [8].

The advantage of SOM is that the relationship between the original vectors is to some extent preserved in the output space, providing a visual format where a human operator can "easily" discover clusters, relations, and structures in the usually complex input space database. The map consists of a regular processing unit, neuron grid. A multidimensional observation model, possibly a vector consisting of features, is associated with each unit. Fig. 2 shows a simple outline of the SOM operation. The software used as a processing tool for the map creation was Matlab. Programming sentences have been modified to develop the neural model applied in this research.

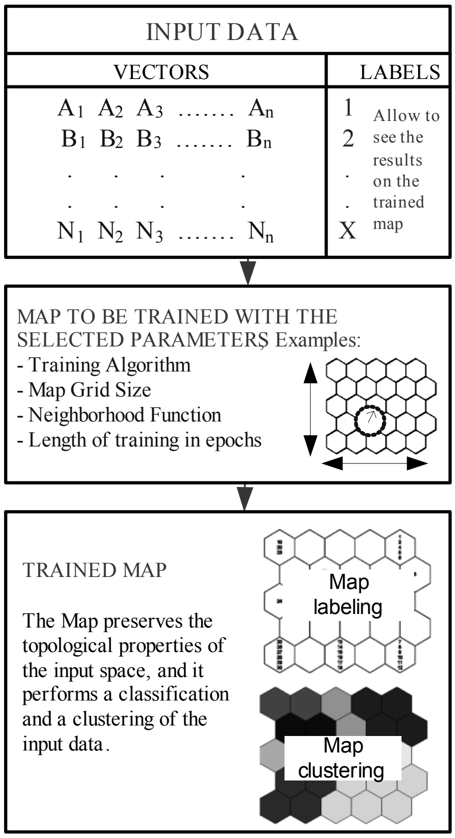

Figure 2. Steps for application of the SOM.

IV. METHODOLOGIES AND FIRST RESULTS

Once understood the operation of the SOM and their capacity, the authors show two different ways to face the process, although they have in fact carried out training on four different types of input data patterns. In order to express the accuracy of the tool, a measurement index is defined. This index is the Mean Absolute Percentage Error (MAPE), which measures the accuracy of fitted time series and forecasts [9].

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\left|L_{Ri} - L_{Pi}\right|}{L_{Ri}}\right]x100\% \qquad (1)$$

Where: $\boldsymbol{N}$ is the number of forecasted hours (in this case is 16 if we use the first 8 hours of the forecasting day), $\boldsymbol{L_{Ri}}$ is the real load value of the $\boldsymbol{i}$ hour and $\boldsymbol{L_{Pi}}$ is the forecasted load of the $\boldsymbol{i}$ hour.

A. *First Methodology*

The first method is based on the training of the network with daily load demand curves of the years 2002 and 2003. Afterwards, the maps are tested network bombing with some load curves of the year 2004, but simply using the first hours of the day to be forecasted. Different simulations were carried out using 8, 10 and 12 hours as input data. The objective is to associate the 2004 input data to the most similar days of the network, formed by 2002 and 2003 days, and then obtain the pattern for the most suitable day evolution. Input data are labelled as "mmddyy", i.e. month-day-year. After testing the network with the first hours of the day to be forecasted, the winner cell is chosen as the most similar curve shape by the software. This allows to estimate the evolution of the following hours. One of the pros of simulating three different size vectors (8, 10 and 12 hours) is to check if the map assigns the same or different winning cells. If the assignation remains always the same, the success is almost guaranteed. Fig. 3 identifies the trained map and the day labels assigned to each cell.

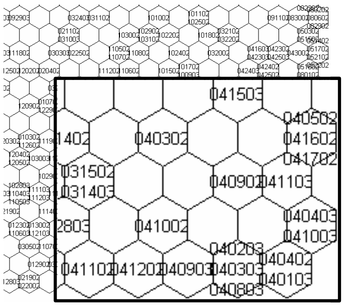

Figure 3. Trained SOM for the load curves of the years 2002 and 2003.

Fig. 4 shows the first 8 hours of the 16th of June 2004, which is the used testing day, and the winner cell containing the labels of the 3rd and 5th of June 2003.

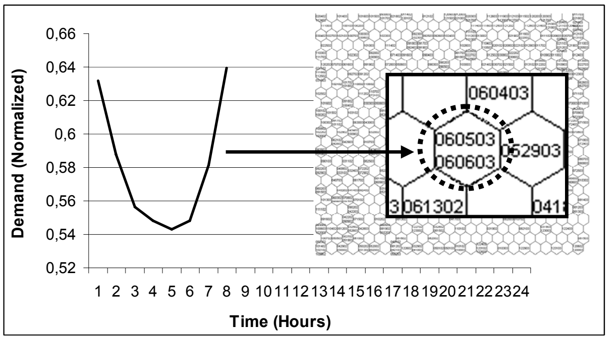

Figure 4. Example of a testing day, with the first 8 hours of the 06/16/2004 and the winner cells.

The authors have carried out successive trainings with different parameters, such as the initialization algorithm (random or linear), training algorithm (sequential or batch), map grid size (numbers of neurons), neighbor function (gaussian or bubble), length of rough training in epochs and length of fine-tuning in epochs, etc... The best results were achieved with the following parameters to train the map, (see Table I):

TABLE I. TRAINING PARAMETERS

Input data	Randinit
Initialization algorithm	Randinit
Training algorithm	Sequent
Network size	25 X 25
Neighbor function	Bubble
Iteration number	5000 and 3000

Fig. 5 shows the average demand load curve of both days and the real curve of the 16th of June. The winner cell was exactly the same for the three tests using 8, 10 and 12 hours as input.

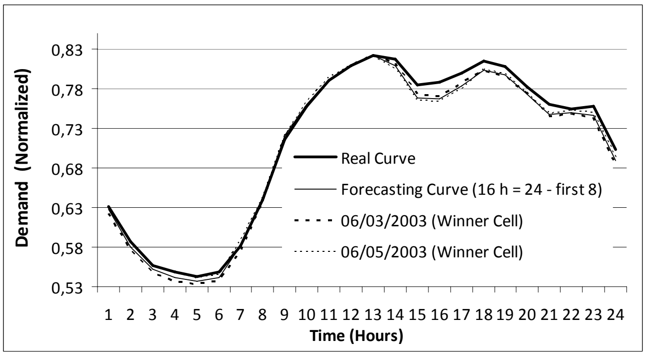

Figure 5. Real load demand (06/16/04) vs Estimated load curve (Average of the winner cell corresponding to labels 06/03/03 and 06/05/03).

Fig. 6 identifies the estimated curve for the 15th of July 2004 but with different solutions for the 8, 10 and 12 hours testing. MAPE index was better the more hours that were used as input for the test. The value obtained was 2.78 for 8 hours, 2.08 for 10 hours and 0.94 for 12 hours.

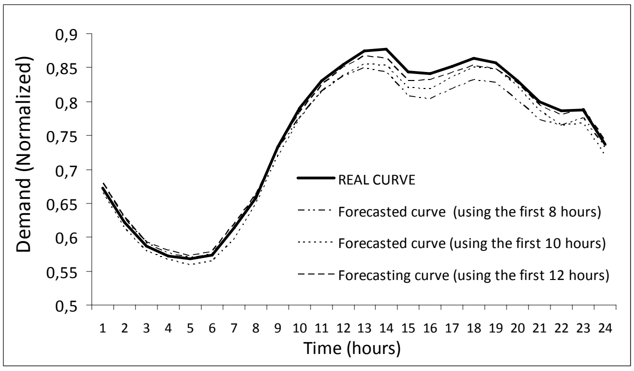

Figure 6. Real load demand curve vs Estimated load curve using 8, 10 and 12 hours as input data (07/15/04).

MAPE indexes can be observed in Table II. The monthly MAPE error averages 2.27 when the first 8 hours of a day are used. However, in this case, the error index slightly increases to 2.34 and 2.36 when 10 and 12 hours are used instead. Nevertheless, there are other months obtaining better results using 10 or 12 hours.

TABLE II. MAPE INDEX FOR SOME JUNE DAYS. FIRST METHODOLOGY.

Days	**MAPE**			**Daily Average** MAPE
	8 hours	***10 hours***	***12 hours***	
06/02/2004	2.44	3.88	4.26	3.52
06/03/2004	2.31	2.51	2.70	2.51
06/04/2004	0.91	1.02	1.14	1.02
06/07/2004	2.15	2.35	2.23	2.24
06/08/2004	4.09	2.44	2.61	3.05
06/09/2004	1.99	2.15	2.98	2.37
06/10/2004	0.67	0.58	0.56	0.60
06/11/2004	1.81	2.97	3.33	2.70
06/14/2004	1.25	4.39	4.98	3.54
06/15/2004	1.03	0.97	0.98	0.99
06/16/2004	1.17	1.27	1.45	1.30
06/17/2004	0.55	0.47	0.50	0.50
06/18/2004	2.84	6.44	2.92	4.06
06/21/2004	7.78	2.93	3.14	4.62
06/22/2004	1.96	2.20	2.40	2.19
06/23/2004	0.83	0.89	0.95	0.89
06/24/2004	1.20	1.30	1.29	1.26
06/25/2004	1.95	1.45	1.62	1.67
06/28/2004	4.74	3.83	4.07	4.21
06/29/2004	3.74	2.77	3.08	3.20
Average MAPE	2.27	2.34	2.36	2.32

Table III shows the MAPE indexes for July. In this case, the resulting index is 2.5, which is very similar to the one obtained for June.

TABLE III. MAPE INDEXES FOR SOME JULY DAYS. FIRST METHODOLOGY.

Days	**MAPE**			**Daily Average MAPE**
	8 hours	***10 hours***	***12 hors***	
07/01/2004	7.22	8.82	1.70	5.91
07/02/2004	3.33	3.12	2.10	2.85
07/05/2004	1.85	1.02	0.99	1.28
07/06/2004	3.67	4.17	4.65	4.16
07/07/2004	1.71	0.55	0.61	0.95
07/08/2004	4.39	4.93	2.14	3.82
07/09/2004	1.85	1.83	2.00	1.89
07/12/2004	1.86	2.04	2.13	2.01
07/13/2004	1.09	1.02	1.14	1.09
07/14/2004	1.18	1.09	1.03	1.10
07/15/2004	2.79	2.08	0.94	1.94
07/16/2004	4.97	1.98	1.94	2.96
07/19/2004	8.24	1.24	1.62	3.70
07/20/2004	1.55	1.62	1.13	1.43
07/21/2004	0.63	0.25	0.25	0.38
07/22/2004	1.19	1.22	1.37	1.26
07/23/2004	7.14	7.87	2.04	5.68
07/26/2004	2.70	1.89	2.11	2.24
07/27/2004	1.44	1.51	1.65	1.53
07/28/2004	6.80	1.48	1.72	3.34
07/29/2004	1.60	1.55	1.07	1.41
07/30/2004	1.16	1.05	0.91	1.04
Average MAPE	3.11	2.38	1.60	2.36

B. Second Methodology

The second method uses two consecutive days instead of simply one as input data for the training of the SOM, i.e. 48 data vectors (48 hours). Pairs of Mondays and Tuesdays of the years 2002 and 2003 are used for the research. The labels change now to a "wwyy" week and year format. The testing data are changed in the same way using the 24 hours of the previous day (Monday) and the 8, 10 or 12 hours of the day to be forecasted (Tuesday). Therefore, our objective is to estimate the rest of the hours of a Tuesday in 2004. Fig. 7 shows the trained map with the new labels. The authors have selected a size of map for this second methodology, which is smaller than the one used in the first. This is due to the fact that the input data are formed by two consecutive days of every week.

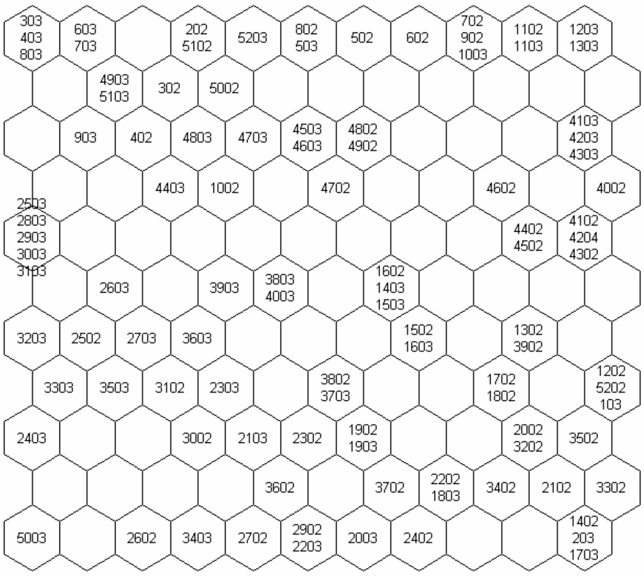

Figure 7. SOM trained with pairs of Mondays and Tuesdays in 2002/03.

In a similar way to the first methodology, when we test a vector (for example 24+8 h) on the trained map, neurons containing more than one week can be assigned. In this case, the average is obtained, and this value will be the forecasted curve. Fig. 8 shows the comparison between the real tested curve (32 h = 24 h of a Monday + first 8 h of a Tuesday) and the associated curves by the trained map. Fig. 9 shows forecasted curves with different testing input data patterns (24+8h, 24+10h, and 24+12h) vs. Real curve of the 7th week of 2004.

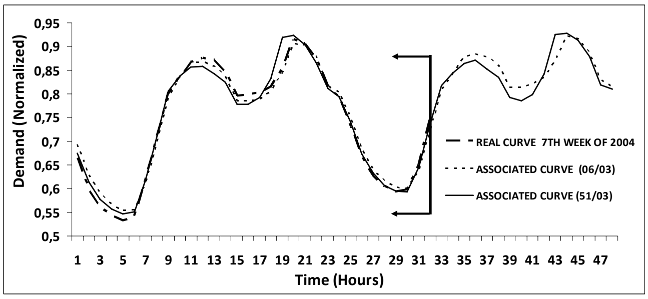

Figure 8. Associated curves by the trained map with testing 24 +8 hours.

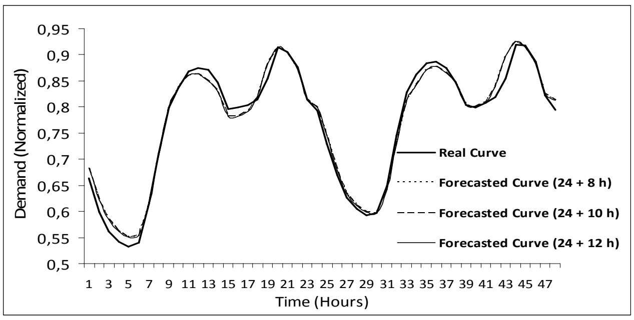

Figure 9. Forecasted curves with different testing input data patterns (24+8h, 24+10h, and 24+12 h) vs. Real curve of the7th week of 2004.

Table IV shows the error index obtained with the second methodology. The MAPE indexes correspond to weeks 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 of the year 2004. The average MAPE value for the weeks tested is 2.07 for the 24+8h, 2.11 for the 24+10h and 1.84 for the 24+12h testing.

TABLE IV. MAPE INDEX. SECOND METHODOLOGY.

Week	**MAPE**			**Daily average MAPE**
	8 hours	***10 hours***	***12 hours***	
3 / 2004	0.59	0.56	0.52	0.55
4 / 2004	1.09	0.92	0.87	0.96
5 /2004	1.26	1.32	1.22	1.27
6 / 2004	1.97	2.21	2.51	2.23
7 / 2004	1.21	1.10	1.08	1.13
8 / 2004	1.24	1.24	1.23	1.24
9 / 2004	4.07	4.46	1.66	3.40
10 / 2004	3.59	3.62	3.54	3.58
11 / 2004	2.70	2.69	2.65	2.68
12 / 2004	2.93	2.96	3.08	2.99
Average MAPE	2.07	2.11	1.84	2.00

V. CONCLUSIONS

The results obtained through this methodology show a low error index after the comparison of the estimated days with the real data of 2004 simply using historical data on weekdays of the years 2002 and 2003. Several parameter setups were tested which finally achieved an optimum combination that reduced the error to a 1.8-2% average, on some days. However, on other days of the year a higher error index was obtained. Nevertheless, the research objective from now onwards is to test new training configurations with gradually more complex input data, for instance adding weather conditions such as temperature, in order to improve the error indexes. Self-Organizing Maps (SOM) and the methodologies presented in this paper can be a useful tool for electrical companies or system operators to try to estimate short-term demand from one day to a few hours. In the future the authors will continue to use training maps with different input data patterns and compare the results.

ACKNOWLEDGMENT

This research was financed by the Research Project with reference GV/2010/080-Valencian Government, Department of Education, and directed by Dr. Sergio Valero Verdú.

REFERENCES

[1] Y.V. Makarov, V.I. Reshetov, A. Stroev, I. Voropai, "Blackout Prevention in the United States, Europe, and Russia: Proceedings of the IEEE", Volume (vol.) 93, issue11, pp. 1942-1955, Nov. 2005.

[2] J. H. Chow, "Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence", New York: Springer-Verlag, ch. 12, 2005.

[3] P. Murto, "Neural Network Models for short-term load forecasting", (Department of Engineering, Physics and Mathematics. Helsinki University of Technology, 1998).

[4] J. Bao, "Short-Term Load Forecasting bases on neural network and moving average", (Artificial Intelligence Lab, Dept. of Computer Science Iowa State University, 2000).

[5] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-Term Load Forecasting Using An Artificial Neural Network," IEEE Trans. on Power Systems, vol. 7, pp. 124-132, Feb. 1992.

[6] *REE, Red Eléctrica de España,* www.ree.es.

[7] Fan S., Chen L., "Short-term load forecasting based on an adaptive hybrid method", Power Systems, IEEE Transactions on vol. 21, issue 1, pp. 392-401, Feb. 2006.

[8] Kohonen, T., "Self-organisation and associative memory" (Springer-Verlag, 3rd edn., Berlin, 1989.

[9] Tafreshi S.M.M., Farhadi M., "Improved SOM based method for short-term load forecast of Iran power network", Power Engineering Conference, 2007. IPEC 2007. International, 3-6 Dec. 2007, pp. 1377-1384.