One Day Ahead Peak Electricity Load Prediction

Jelena Milojković and Vančo Litovski
Faculty of Electronic Engineering
University of Niš, Niš, Serbia
jelena@venus.elfak.ni.ac.rs, vanco@elfak.ni.ac.rs

Abstract—**One step ahead prediction method for peak daily electricity loads based on artificial neural networks (ANN) is presented. Two architectures of ANN were implemented to produce predictions that were used to generate the final value as an average. Examples will be given confirming both the feasibility of the method and the need for further elaboration of the procedure.**

Keywords-Peak electricity load; prediction; artificial neural networks.

I. Introduction

The necessity of load forecasting is nowadays broadly recognized. Precise load forecasting helps the electric utility to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly. In addition, corrective actions may be prepared, such as to avoid load shedding, planning power purchases and bringing peaking units on line. Especially, as noted in [1] accurate short-term forecasts are needed by both generators and consumers of electricity particularly during periods of abnormal peak load demand.

The electricity forecasting period may span from several tenths of minutes to several years so very short term (at tenths of minutes level), short term (hourly), daily, weekly, monthly and yearly load forecasts may be encountered. The proceedings presented here are based on our previous results related to short term prediction [2,3,4]. Here we implement similar methods to generate the forecast of a daily peak value for a given load. Data were extracted from the 1999 UNITE competition [5].

Our method is based on several hypotheses. First we claim that main influence to the future value may have the most recent observables. These contain most recent information on trend, season and weather. Second, we believe that if quality forecast is to be obtained, one that may be used for action, one is not supposed to try many time steps in advance. One or two time steps are the best one can afford. That is why the time series is presented here as deterministic and one-step-ahead prediction is planned. To help the prediction, however, in an appropriate way, we introduce past values e.g. loads for the same day but in previous weeks. That is in accordance with existing experience claiming that every day in the week has its own general consumption profile [6].

In many load forecasting procedures weather data are used as basic input together with the load time series. We here have a specific opinion about the use of weather data. First, as can be seen from experiments [7] it is not easy to establish a significant correlation between the weather parameters and the peak load value. Second, for the prediction instant no weather data are available. These are to be generated by pre-diction with equal uncertainty as the main prediction. Finally, the known load values already contain information on the weather if any correlation exists.

As an example of the uncertainty of long term weather prediction based on abundant amount of data let us consider the day of December 19, 2011 which is celebrated as the St. Nicolas by the Orthodox Church exercising the Julian calendar. It is the most celebrated Serbian “Slava” and while “half of the people celebrate the other half is visiting the families that celebrate”. There is always snow at St. Nicolas. There are even proverbs related to the snow at St. Nicolas. On the last St. Nicolas day, however, there was no snow and the temperature was above zero all day. Nobody could predict that state one month earlier at November 19, 2011, while everyone could do that at December 18, 2011.

The problem of daily peak load forecasting was considered many times in the literature [8,9,10]. As can be seen statistical methods are used.

One of the approaches to load forecast is implementation of artificial neural networks (ANN) [11,12]. The main advantage of the method is related to the property of the ANN to be an universal approximator meaning the main problem of regression: the choice of the approximating function, is solved in advance. A common feature, however, of the existing application is that they ask for a relatively long time series to become effective. Typically it should be not shorter than 50 data points [11].

Following these considerations new forecasting architectures were developed [2,3,4]. Namely, prediction is an activity that is always related to uncertainty. One is supposed to have at least two solutions for them to support each other. The structures developed were named Time Controlled Recurrent (TCR) and Feed Forward Accommodated for Prediction (FFAP). Both were implemented successfully for prediction in modern developments in micro electronics [2] as well as in other areas including hourly [3,4] and yearly [8] load prediction.

Here we present extensions of the TCR and FFAP ANNs that allow for implementation in daily peak load prediction. These are named extended TCR (ETCR) and extended FFAP (EFFAP).

The structure of the paper is as follows. After general definitions and statement of the problem we will give a short

This research was partially funded by The Ministry of Education and Science of Republic of Serbia under contract No. TR32004.

description of the two solutions. After presenting the experimental results a short discussion of the results and consideration related to future work will be given.

II. Problem Formulation

A time series is a number of observations that are taken consecutively in time. A time series that can be predicted precisely is called deterministic, while a time series that has future elements which can be partly determined using previous values, while the exact values cannot be predicted, is said to be stochastic. We are here addressing only deterministic type of time series.

Consider a scalar time series denoted by y_i, $i=1,2, \ldots m$. It represents a set of observables of an unknown function $\hat{y} = \hat{f}(t)$, taken at equidistant time instants separated by the interval Δt i.e. $t_{i+1}= t_i+\Delta t$. One step ahead forecasting means to find such a function f that will perform the mapping

$$y_{m+1} = f(t_{m+1}) = \hat{y}_{m+1} + \varepsilon, \qquad (1)$$

where $\hat{y}_{m+1}$ is the desired response, with an acceptable error ε.

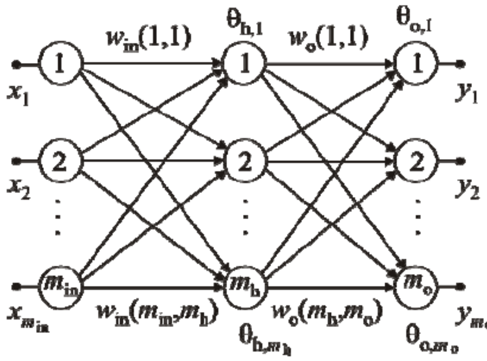

Figure 1. Fully connected feed-forward artificial neural network with one hidden layer and multiple outputs.

In the next, we will first briefly introduce the feed-forward neural networks that will be used as a basic structure for prediction throughout this paper.

The network is depicted in Fig. 1. It has only one hidden layer, which has been proven sufficient for this kind of problem [14]. Indices: "in", "h", and "o", in this figure, stand for input, hidden, and output, respectively. For the set of weights, $w(k, l)$, connecting the input and the hidden layer we have: $k=1,2,\ldots, m_{in}$, $l=1,2,\ldots, m_h$, while for the set connecting the hidden and output layer we have: $k=1,2, \ldots m_h$, $l=1,2,\ldots, m_o$. The thresholds are here denoted as θ_x,ω_r, $r=1,2, \ldots, m_h$ or m_o, with x standing for "h" or "o", depending on the layer. The neurons in the input layer are simply distributing the signals, while those in the hidden layer are activated by a sigmoidal (logistic) function. Finally, the neurons in the output layer are activated by a linear function. The learning algorithm used for training is a version of the steepest-descent minimization algorithm [15]. The number of hidden neurons, m_h, is of main concern. To get it we applied a procedure that is based on proceedings given in [16].

In prediction of time series, in our case, a set of observables (samples) is extracted (one peak value per day) from the UNITE 1997 file. According to (1) we are predicting one quantity at a time. To make the forecasting problem numerically feasible we performed a small transformation in the response. Namely the samples are reduced in the following way

$$y=y^* - M \qquad (3)$$

where y^* stands for the measured value of the target function, M is a constant (here M=600 kW).

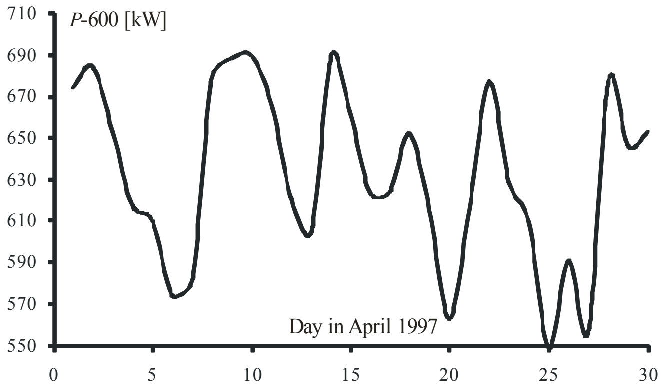

Figure 2. Load peak values for April/May 1997.

If the architecture depicted in Fig. 1 was to be implemented (with one input and one output terminal) the following series would be learned: $(t_i, f(t_i))$, $i=1,2,\ldots$.

The observables are illustrated by Fig. 2. It represents the daily peak values in the period from April 07, to May 06, 1997.

III. The ETCR Solution

Starting with the basic structure of Fig. 1, in [2,3,4] possible solutions were investigated and two new architectures were suggested to be the most convenient for the solution of the forecasting problem based on short prediction base period.

The first one, named *extended time controlled recurrent* (ETCR) was inspired by the time delayed recurrent ANN. It is a recurrent architecture with the time as input variable so controlling the predicted value. Our intention was to benefit from both: the generalization property of the ANNs and the success of the recurrent architecture. Its structure is depicted in Fig. 3. In this figure i stands for the sample counter and in fact represents the time variable i.e. the day. t_i stands for the daily peak value time while y_i is the daily peak value. Here in fact, the network is learning two sets of variables. The first is the output value representing the daily peak power consumption for the next day is controlled by the present time (variable i) and by its own previous instances. The second is the daily peak value time which is controlled by the same data:

$$y_i = f(i, y_{i-1}, y_{i-2}, y_{i-3}, y_{i-4}, y_{i-7}, y_{i-14}),\quad i=8,9,... \qquad (2)$$

In these first proceedings we chose four recent samples and one one-week old to control the output. That choice was confirmed by the results obtained so no new attempts were made to complicate the training set of the ANN.

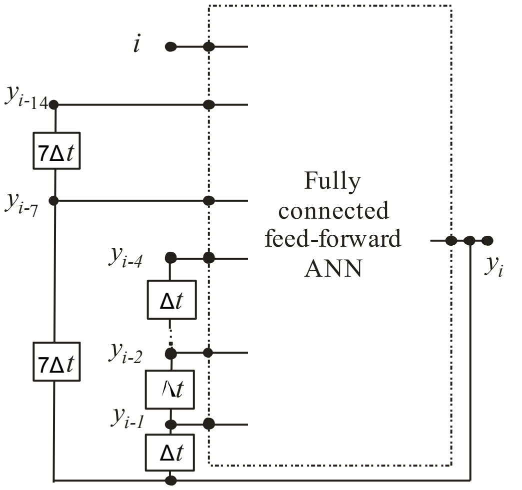

Figure 3. ETCR. Extended time controlled recurrent ANN.

TABLE I. TRAINING DATA FOR THE ETCR ANN AT APRIL 30, 1997.

i	y^{i-1}	y^{i-2}	y^{i-3}	y^{i-4}	y^{i-7}	y^{i-14}	y^{i}
8	84	74	82	13	103	167	53
9	53	84	74	82	97	140	17
10	17	53	84	74	51	137	10
11	10	17	53	84	13	76	-26
12	-26	10	17	53	82	135	-19
13	-19	-26	10	17	74	21	79
14	79	-19	-26	10	84	94	88
15	88	79	-19	-26	53	103	89
16	89	88	79	-19	17	97	66
17	66	89	88	79	10	51	20
18	20	66	89	88	-26	13	05
19	05	20	66	89	-19	82	89
20	89	05	20	66	79	74	63
21	63	89	05	20	88	84	24
22	24	63	89	05	89	53	24
23	24	24	63	89	66	17	52
24	52	24	24	63	20	10	15
25	15	52	24	24	05	-26	-37
26	-37	15	52	24	89	-19	15
27	15	-37	15	52	63	79	77
28	77	15	-37	15	24	88	29
29	29	77	15	-37	24	89	

According to this definition when preparing the training data for the ETCR ANN, sets of vectors were created by extracting data from the original similarly to the time series reconstruction technique that stems from the embedding theorem developed in [17,18]. The ith input training vector would be:

$$\mathbf{x}_i=\{i, y_{i-1}, y_{i-2}, y_{i-3}, y_{i-4}, y_{i-7}, y_{i-14}\},$$

while the corresponding training output vector would be

$$\mathbf{z}_i=\{y_i\}.$$

In this proceedings $i\in\{8,28\}$. Namely, 21 training lessons were used. The training data are given in Table I.

The task was to predict the peak value at April 30, 1997, which was given in the literature [5] to be 609 kW. The resulting ANN had 7 input, two output, and 5 hidden neurons. After proper excitation the prediction was $y_{29}=\{625.3241\}$, as can be seen from Table II.

Summarizing the example, in order to get a picture about the research that should be done, we will state here the list of parameters that are to be set for the prediction procedure to become stable, repeatable and reliable. There are two domains, being interrelated, that are to be parameterized: the domain of data and the domain of the ANN. In the data domain we are to define the number of samples presented to the ANN as input in every training lesson, q. Then, the complete set of input data is to be defined. Its length will be denoted by p. The number of lessons is p-q since, while creating the next lesson, we shift the data window by one to the future. As we can see from Table I, in the example above q=4 and p=25 (The set starts at i=4 and ends at i=28). Further, since we want to use the value from the previous week we need to have more samples in the past so that $s+q$=7. In cases where more than one week backward is defined as necessary data, we will extend the input set by $7\cdot(r-1)+s$ data, r being the number of weeks backward. All together, from the data point of view we are to choose three parameters: r, p and q. The value of q partly defines the number of ANNs inputs related to one of the variables to be predicted as can be seen from Fig. 3 for q=4. Similarly r defines the number of additional inputs for a week old data. In this proceedings we use r=2. Having all that in mind we may conclude that the number of input and output terminals of the ANN is fixed by the data structure. The remaining free parameter is the number of hidden neurons, m_h. That makes four parameters necessary to be estimated in order for the method to be completely defined. Here, however, we will show only preliminary results confirming the feasibility of the method and encouraging for a serious effort to be done for the estimation of the four parameters appropriate for the proper implementation.

IV. THE EFFAP SOLUTION

The second structure was named *extended feed forward accommodated for prediction* (FFAP) and depicted in Fig. 4. We use the same notation as in Fig. 3. Our idea was here to force the neural network to learn the same mapping several times simultaneously but shifted in time. In that way, we suppose, the previous responses of the function will have larger influence on the $f(t)$ mapping. Note that y_{i+1} is learned meaning a set of data shifted by one in time was used in this case.

In that way for the approximation function we may write the following

TABLE II. PREDICTION RESULTS WITH TWO WEEKS BACKWARD DATA (APRIL 30,1997).

No.	Expected value	ETCR	%	EFFAP	%	Averaged prediction	%	Hidden neurons	
								ETCR	EFFAP
1	609	625,3241	2.68	653.2675	7.27	639.2958	4.975	5	5
2	549	655.4755	19.39	628.6489	14.51	642.0622	16.95	5	5
3	591	678.138	14.74	586.2964	-0.80	632.2172	6.97	6	6
4	557	569.3818	2.22	659.0633	18.32	614.2226	10.27	4	4
5	677	634.3647	-6.30	538.9573	-20.39	586.661	-13.34	5	5
6	646	540.3626	-16.35	577.984	-10.53	559.1733	-13.44	5	5
7	653	583.5895	-10.63	650.1992	-0.43	616.89	-5.53	5	5

$$\{y_{i+1}, y_i, y_{i-1}, y_{i-2}, y_{i-3},\} = \mathbf{f}(i, y_{i-6}, y_{i-13}),$$

i=7,8, ... (3)

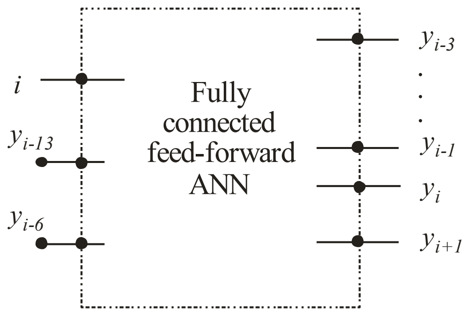

Figure 4. The Extended feed forward accommodated for prediction ANN (EFFAP) according to (3).

The new network is approximating the future (unknown) value y_{i+1}, based on the actual time i, the actual peak value coordinates (y_i), three past peak value coordinates (y_{i-k}, k=1,2,3), and the past peak value coordinates for the same day of two previous weeks (y_{i-6},y_{i-13}).

The resulting ANN for Appril 30, 1997, had 3 input, 5 outputs, and 5 hidden neurons. After proper excitation the prediction was y_{29}= {653.2675}. As can be seen from Table II, now the prediction is slightly worse in comparison with the ETCR solution for the same day.

To get a complete picture about the capabilities of the method, the same procedure, for both ETCR and EFFAP, was applied for the nex six days. The results obtained are depicted in Table II. Lookin for the relative error depicted in columns 4 and 6 of that table we may conclude that none of them may be pronounced better. In both cases the maximum error in prediction is about 20%.

After this, one more aspect of the procedure may be discussed here. Namely, since in reality we have no knowledge of the quality of the prediction we need some kind of stoppage criterion in the estimation of the parameters mentioned. In that sense we insist the two predictions i.e. EFFAP and ETCR, to support each other (i.e. to be of similar value) and the predicted values not to abandon a foreseen interval established by examining the complete set of input data. In this example, looking into Table I, we may adopt the upper limit of the prediction to be 735•1.1=808.5, while the lower limit to be 563•0.9=506.7. A 10% margin was added to the maximum value in the data of Table I for the top limit, while 10% margin was subtracted for the lower limit.

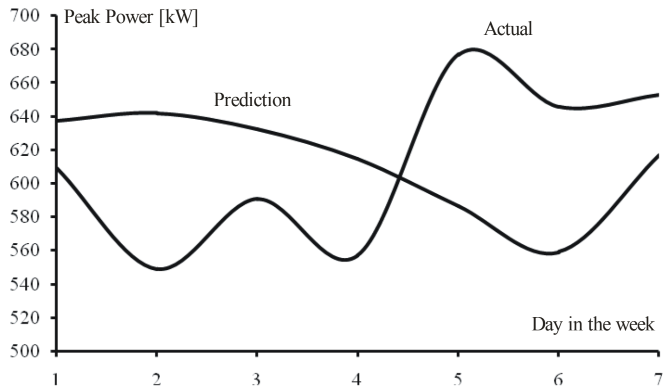

Figure 5. Seven consecutive predictions compared with the actual values of peak daily electricity consumption (April 30 to May 06, 1997).

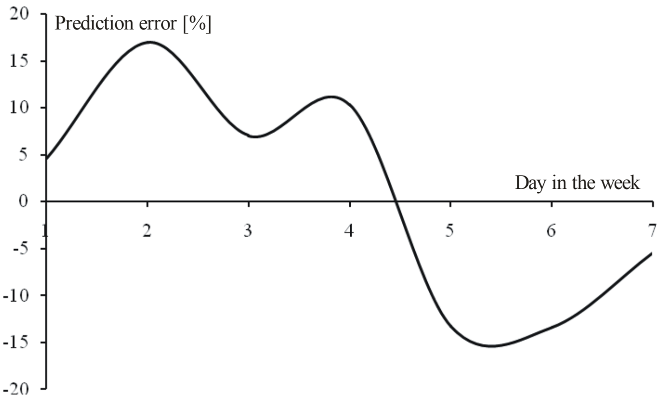

Figure 6. Relative discrepances between actual and predicted peak daily electricity consumption.

V. Summary and Conclusion

Preliminary results on implementation of the ETCR and EFFAP ANNs for prediction of peak daily electricity loads were reported above.

As expected the prediction obtained by application of the two ANN structures differ. These are both necessary, however, in order to mutually support since, in prediction, no other reference is available. The main criterion for acceptance is the mutual similarity of the results produced by different methods. Since, however, none of them may be considered better in advance one is to profit of both by using the average. The averaged prediction values and the corresponding errors are also given in Table II. Note the worst case prediction is now about 17%.

These results are visualized in Fig. 5, where the actual and the average prediction value are shown for seven consecutive prediction produced fully independently from each other. Fig. 6, depicts the relative discrepancies between the curves of Fig. 5.

Based on the preliminary results reported above we find the method proposed feasible for implementation in short term prediction of daily peak loads with no use of environmental data. The number of previous days and weeks used for prediction will be considered in more detail in the future. Multistep ahead prediction will be investigated, too.

References

[1] Taylor, J.W., "An evaluation of methods for very short-term load forecasting using minute-by-minute British data", *International Journal of Forecasting (Energy Forecasting)*, Vol. 24, No. 4, 2008, pp.645–658.

[2] Milojković, J.B., Litovski, V.B., "Short term forecasting in Electronics", *International Journal of Electronics*, ISSN: 0020-7217, Vol. 98, No. 2, 2011, pp. 161-172.

[3] Milojković, J, and Litovski, V.B., „Dynamic Short-Term Forecasting of Electricity Load Using Feed-Forward ANNs", *Int. J. of Engineering Intelligent Systems for Electrical Engineering and Communication*, ISSN 1472-8915, Vol. 17, No. 1, March 2009, pp. 38-48.

[4] Milojković, J., and Litovski, V., „Dynamic One Step Ahead Prediction of Electricity Loads at Suburban Level", *Proc. of the First IEEE Int. Workshop on Smart grid Modeling and Simulation – at IEEE SmartGridComm 2011*, SGMS2011, Brussels, October 2011, Proc. on disc, paper no. 25.

[5] World wide competition within the eunite network. (2001). [Online] Available: http://neuron.tuke.sk/competition.

[6] Murto P., '*Neural Network Models for Short-Term Load Forecasting*", MS Thesis, Helsinki University of Technology, 1998.

[7] Hyndman, R.J., and Fan, S., "*Forecasting long-term peak half-hourly electricity demand for South Australia*", Report for Electricity Supply Industry Planning Council (South Australia, Monash University, Australia, 2009.)

[8] Elattar, E.E., Goulermas, J., and Wu, Q. H., "Electric Load Forecasting Based on Locally Weighted Support Vector Regression", *IEEE Transactions On Systems, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 40, No. 4, July 2010, pp. 438-447.

[9] As'ad, M., "Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand", *Applied Statistics education and Research Collaboration (ASEARC) – Conf. Papers,* Uni. of Wollongong, Australia, 2012, Paper 11.

[10] Sigauke, C., and Chikobvu, D., "Prediction of daily peak electricity demand in South Africa using volatility forecasting models", *Energy Economics*, Elsevier, Vol. 33, No. 5, Sept. 2011, pp. 882-888.

[11] Zhang, B.G., "Forecasting with artificial neural networks: The state of the art", *Int. J. of Forecasting*, Vol. 14, No. 1, March 1998, pp. 35-62.

[12] Hippert, H. S., Pedreira, C. E., and Souza, R. C., "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, pp. 44–55, February 2001.

[13] Milojković, J., Litovski, V., Nieto-Taladriz, O., and Bojanić, S., "Forecasting Based on Short Time Series Using ANNs and Grey Theory – Some Basic Comparisons", *In J. Cabestany, I. Rojas, and G. Joya (Eds.): Part I, LNCS 6691*, pp. 183–190, 2011, © Springer-Verlag, Berlin, Heidelberg.

[14] Masters, T., "*Practical Neural Network Recipes in* C++", Academic Press, San Diego, 1993.

[15] Zografski, Z., "A novel machine learning algorithm and its use in modeling and simulation of dynamical systems", *Proc. of 5th Annual European Computer Con.*, COMPEURO '91, Hamburg, Germany, 1991, pp. 860-864.

[16] Baum, E.B., and Haussler, D., "What size net gives valid generalization", *Neural Computing*, 1989, Vol. 1, pp. 151-160.

[17] Takens, F., "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics* (Springer Berlin), Vol. 898, 1981, pp. 366–381.

[18] Sauer, T., Yorke, J. A., and Casdagli, M., "Embedology," *Journal of Statistical Physics*, Vol. 65, May 1991, pp. 579–616.