Short-Term Forecasting of Electricity Load Using Recurrent ANNs

Jelena B. Milojković and Vančo B. Litovski

Abstract—**One solution to the problem of one-step-ahead prediction of the electricity power consumption in sub-urban areas is proposed. It is based on implementation of artificial neural networks (ANN) that are properly structured for prediction. Architecture named Extended Time Controlled Recurrent (ETCR) ANN is introduced. A proper arrangement of training data is advised. After implementation, promising results were obtained giving, in some instances, predictions with an error less than 1%. In the worst situation observed, the discrepancy between the target and predicted values go as large as 14% what we consider as acceptable and in the same time as an inspiration for further research in improving the method. The method is intended to be integrated into a remote power reading and billing system giving an opportunity to the energy supplier to plan his business. The method proposed implements ANNs that are generally widely known but creates new architecture that is fully original. It is our opinion that it may be implemented with equal success to one-step-ahead prediction of broader class of time series exhibiting inherent quasi-periodical properties.**

Index Terms—**Electricity load, prediction, artificial neural networks.**

I. Introduction

IN an inspired paper, Prof. Mandel' [1] claims: "Prediction of short time series is a topical problem. Cases where the sample length N is too small for generating statistically reliable variants of prediction are encountered every so often. This form is characteristic of many applied problems of prediction in marketing, politology, investment planning, and other fields." Further he claims: "Statistical analysis suggests that in order to take carefully into account all components the prediction base period should contain several hundreds of units. For periods of several tens of units, satisfactory predictions can be constructed only for the time series representable as the sum of the trend, seasonal, and random components. What is more, these models must have a very limited number of parameters. Series made up by the sum of the trend and the random component sometimes may be predicted for even a smaller base period. Finally, for a prediction base period smaller than some calculated value $N_{\min}$, a more or less satisfactory prediction on the basis of observations is impossible at all, and additional data are required."

Among the fields not mentioned in [1], dealing with really small set of data or „prediction base period", we will discuss here hourly short-term prediction of electricity loads at suburban level or on the level of a low voltage transformer station. In fact, the amount of data available in this case, as depicted in Fig. 1, is large enough to apply any other forecasting method [2]–[5] but looking to the load diagram i.e. hourly load-value curves, we easily recognize that past values of the consumption are not very helpful when short-term prediction is considered. That stands even for data from the previous day and for data from the same day in the previous week. As an illustration of the claim in Fig. 2, we give three load diagrams representing one day consumption of the same load on a) Friday January 31, 2009, b) Thursday January 30, 2009, and c) Friday January 24, 2009. The numerical values are shown in Table I. The power is normalized by a factor of 200 being the turn ratio of the appropriate current transformer in the transformer station. One may notice the similarity of the general shape and the difference in main details confirming the paramount importance of the most recent data for prediction. Accordingly, we propose the problem of prediction of the load value in the next hour (one or two) to be performed as a deterministic prediction based on very short – one day – time series. To help the prediction, however, in an appropriate way, we

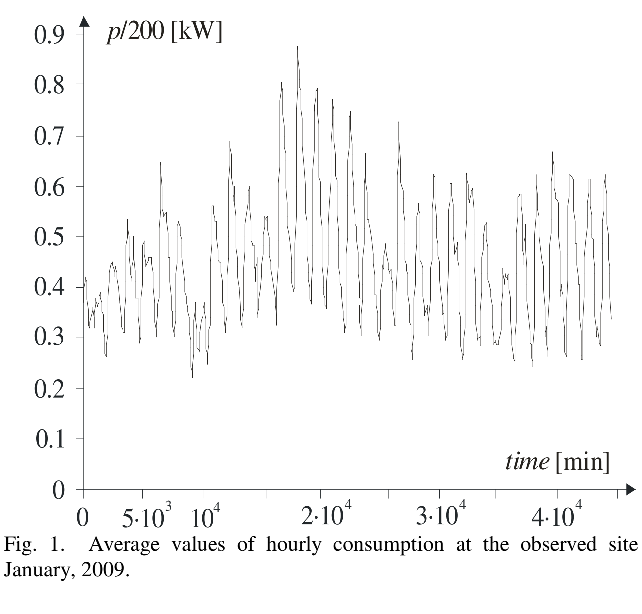

Fig. 1. Average values of hourly consumption at the observed site in January, 2009.

J. B. Milojković and V. B. Litovski are with the University of Niš, Faculty of Electronic Engineering, Niš, Serbia (e-mail: jelena@venus.elfak.ni.ac.rs, vanco@elfak.ni.ac.rs).

TABLE I
AVERAGE VALUES OF HOURLY CONSUMPTION (KW) ON THREE DAYS

No.	t (min)	p/200 31.01.09	p/200 30.01.09	p/200 24.01.09
1	0.06	0.40	0.42	0.31
2	146	0.32	0.28	0.32
3	284	0.34	0.27	0.30
4	422	0.30	0.44	0.44
5	561	0.50	0.56	0.50
6	700	0.58	0.63	0.54
7	837	0.64	0.63	0.44
8	977	0.58	0.52	0.42
9	1115	0.50	0.51	0.41
10	1255	0.44	0.44	0.38
11	1393	0.35	0.40	0.31

t=0 at midnight.

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 (Murto, 1998).

The idea is reminiscent to the substitution of the simple moving average (SMA) by the exponential moving average (EMA) method for prediction of trend [6]–[7]. The simple moving average is extremely popular among traders, but one argues that the usefulness of the SMA is limited because each point in the data series is equally weighted, regardless of its position in the sequence. It is common opinion that most recent data is more significant than the older and should have a greater influence on the final result.

Having all that in mind we undertook a project of developing an ANN based method that will be convenient for systematic implementation in stationary time series prediction with reduced set of data. Our first results were applied to prediction of environmental as well as technological data and published in [8]–[9]. Our main idea implemented was the following. If one wants to create neural network that may be used for forecasting one should enable this property during ANN's training by proper preparation of the training set. In

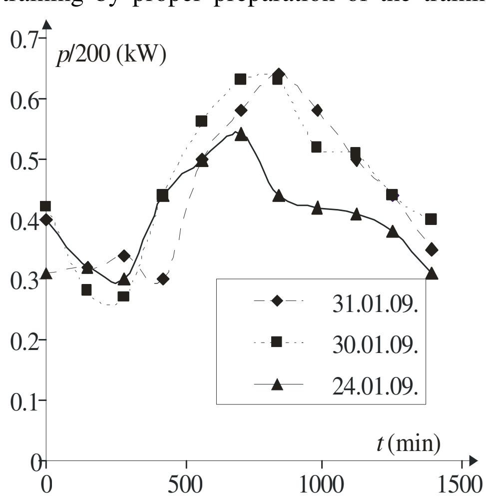

Fig. 2. Average values of hourly consumption on three days (Table I visualized).

addition, the ANN used has to have such an architecture to accommodate to the training process for prediction. Following these considerations new forecasting architectures were developed and implemented [8]–[9]. Here, however, we will upgrade one of them and, for the first time, describe its implementation in the field of short term electricity load forecasting.

The structure of the paper is as follows. After general definitions and statement of the problem we will give a short background related to ANNs application to forecasting. Then we will describe the solution for possible applications of ANNs aimed to the one-step-ahead forecasting task. Finally, short discussion of the results and consideration related to future work will be given.

II. PROBLEM FORMULATION AND SOLUTION

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, $i = 1, 2, \dots m$. 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. $t_{i+1} = t_i + \Delta t$. In the next, we will introduce $h = \Delta t$, for convenience. One-step-ahead forecasting means to find such a function f that will perform the mapping

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

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

The prediction of a time series is synonymous with modeling of the underlying physical or social process responsible for its generation. This is the reason of the difficulty of the task. There have been many attempts to find solution to the problem. Among the classical deterministic methods we may mention the k-nearest-neighbor [10], in which the data series is searched for situations similar to the current one each time a forecast needs to be made. This method asks for periodicity to be exploited that, as already discussed, here may be helpful but not decisively.

In the past decades ANNs have emerged as a technology with a great promise for identifying and modeling data patterns that are not easily discernible by traditional methods. Analysis as to why neural networks are implemented for prediction may be found in [9]. A comprehensive review of ANN use in forecasting may be found in [11]. Among the many successful implementations we may mention [12]. 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]. This is due to the fact that they all look for periodicity within the data. Very short time series were treated in [13]. Here additional "nonsample information" was added to the time series in order to get statistical estimation from deterministic data.

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. 3. 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,..., m_{in}, l=1,2,..., m_{h}, while for the set connecting the hidden and output layer we have: k=1,2, ...m_{h}, l=1,2,..., m_{o}. The thresholds are here denoted as $\theta_{x,r}$, r=1,2, …, 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 given (approximately every two hours or exactly twelve samples per day) meaning that only one input signal is available being the discretized time. According to (1) we are predicting one quantity at a time meaning one output is needed, too. The values of the output are numbers (average power for a period of approximately two hours). To make the forecasting problem numerically feasible we performed transformation in both the time variable and the response. The time was reduced by t_0 so that

$$t = t^* - t_0 \,. \tag{2}$$

Having in mind that t^* stands for the time (in minutes) during one day, this reduction gives the value of 0 to the time (t_0) related to the first sample. The samples are normalized in the following way

$$y = y^* / M \tag{3}$$

where y^* stands for the current value of the target function and M is a constant (for example M=200, being the turn ratio of the current transformer in the station).

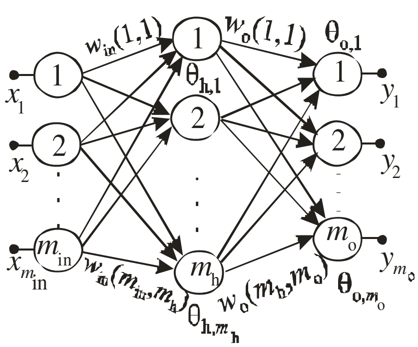

Fig. 3. Fully connected feed-forward neural network with one hidden layer and multiple outputs.

If the architecture depicted in Fig. 3 was to be implemented (with one input and one output terminal) the following series would be learned: (t_i, $f(t_i)$), i=1,..., m). It is our experience [8, 9], however, that such a simple architecture while excellent for interpolation can't extrapolate successfully. That is why we are going for such an architecture that includes more complex sample of input data. In fact, we will here exploit the advantages the k-nearest-neighbor idea and create such an architecture that combines most recent and data periodically dislocated backward in time.

Starting with the basic structure of Fig. 3, in [8] 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. Here, the implementation of one of these two architecture will be considered and accomodated to the situation when some additional data is available as mentioned in the introduction. Our intention was to benefit from both: the generalization property of the ANNs and the success of the recurrent architecture.

The first version of this architecture, named *time controlled reccurent* (TCR) was inspired by the time delayed recurrent ANN. It is a recurrent architecture with the time as the only input variable so controlling the predicted value. At the inputs where the output is feed back we introduce signals that are delayed by one sampling interval only.

We extend, now, this architecture so that we allow for the values of the power consumption at instants shifted for one step in future but of the same days in four previous weeks, to control the output. Hence, the term extended will be added. This structure will be referred from now on as the Extended Time Controlled Recurrent (ETCR) architecture. It is depicted in Fig. 4. for a set of chosen parameters. In the case depicted, three samples of previous instants of the given day plus four samples of previous weeks are used in order to create the prediction for a given instant. j stands for the number of samples per day while n is the order number of the week in the time series. In that way the values indexed with n are from the actual week, while the values indexed n-s, s=1,2,3,4, are from the previous weeks. i stands for the i-th sample in the day

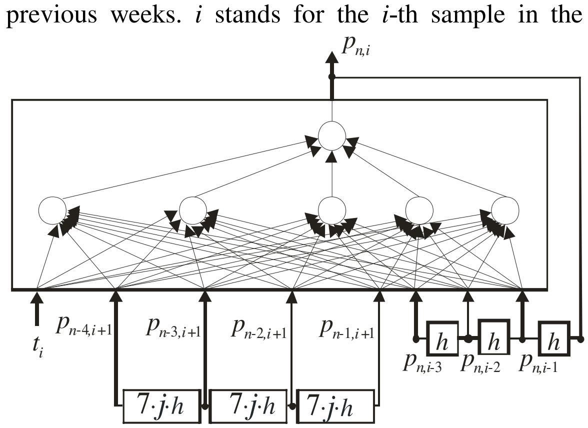

Fig. 4. ETCR (Extended Time Controlled Recurrent) ANN.

selected. Accordingly, for example, $p_{n\text{-}2,2}$ would mean: the second sample for the given day (say Friday) two weeks earlier of the nth week.

Here in fact, the network is learning a series in which the output value, representing the average power consumption for approximately two hours period in a given day of the week, is controlled by the present time and its own previous instances:

$$p_{n,i} = f(t_i, p_{n,i-1}, p_{n,i-2}, p_{n,i-3}, p_{n-1,i+1}, p_{n-2,i+1}, p_{n-3,i+1}, p_{n-4,i+1}), i = 4,...,m, \tag{4}$$

or

$$p_{n,i} = f(t_i, p(t_i - h), p(t_i - 2h), p(t_i - 3h), p(t_i - 7\cdot j\cdot h + h), p(t_i - 14\cdot j\cdot h + h), p(t_i - 21\cdot j\cdot h + h), p(t_i - 28\cdot j\cdot h + h)), i = 4,...,m. \tag{5}$$

The actual value $p_{n,i}$ is unknown and should be predicted.

In the next we will illustrate the implementation of the method in two phases. First we will select one time instant and show how we do prediction. Then we will implement the method to a complete day and show the effectiveness of the method.

III. Implementation Example

The implementation of the method is conceived so that it should create a prediction for the consumption in the next two hours. Just after the present measurement is finished new training data structure is to be created and training of the ANN lounched. After that the so obtained ANN is to be run with an input vector obtained from the last measurement and prediction is to be obtained. The predicted value is to be added to the consumption log for further processing.

The training data accomodated for training the ETCR network intended to be developed for forecasting the value of the consumption at 07.02 hours in the morning of Friday, January 31, 2009, is depicted in Table II. The location of that point at the consumption curve may be seen from Fig. 5 to be at the second minimim (t=422, note the different time window). The time in Table II is given in minutes so that the forcasting moment is 1523 minutes after the start of the time window. Here n=5 since it is the fifth Friday in the year. Note

TABLE I
Training Data Prepared for the ETCR Method

i	t_i	$p_{n-1,i+1}$	$p_{n-2,i+1}$	$p_{n-3,i+1}$	$p_{n-4,i+1}$	$p_{n,i-3}$	$p_{n,i-2}$	$p_{n,i-1}$	$p_{n,i}$
4	415	0.58	0.75	0.69	0.45	0.27	0.44	0.56	0.63
5	552	0.61	0.66	0.57	0.42	0.44	0.56	0.63	0.63
6	692	0.52	0.60	0.60	0.44	0.56	0.63	0.63	0.52
7	830	0.40	0.53	0.52	0.43	0.63	0.63	0.52	0.51
8	968	0.38	0.52	0.44	0.40	0.63	0.52	0.51	0.44
9	1107	0.31	0.38	0.40	0.37	0.52	0.51	0.44	0.40
10	1246	0.32	0.38	0.36	0.34	0.51	0.44	0.40	0.32
11	1383	0.30	0.32	0.32	0.31	0.44	0.40	0.32	0.34
12	1523	0.44	0.42	0.38	0.31	0.40	0.32	0.34	

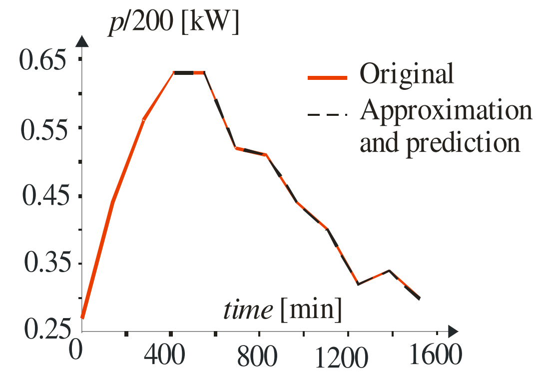

Fig. 5. The actual curve and the approximation (dashed black line) obtained by the ETCR network. The last segment of the dashed line finishes with the prediction.

that in the first row, enumerated i=4, the value $p_{n,i\text{-}3}$ is, in fact, the first value in the time series: $p_{n,1}$. The rightmost column separated by bold line is the target value for the given sample (lesson) used during training. The lowermost row, separated also by bold line, is related to the time instant where prediction should be peformed. The values given in that row will be used as excitation to the ANN obtained after training. The target value to be matched is **0.30**. One should fill the empty cell in Table II by the prediction.

By inspection of Table II despite the relatively large number of items in the table, we may conclude that the time series we are extrapolating is short in the sense that we use 11 samples in the main time frame only i.e. i=1,….,11.

After training the ETCR network with these data and, after training, exciting it as described above, the predicted value was **0.298699**. It is a miss of the target value by 0,43%. In addition, the ETCR ANN performs ideally in approximation of the load curve as can be seen from Fig. 10 where the input curve and the approximation overlap in the whole approximation interval $t\in\{0, 1383\}$. This result was obtained by an ANN with five neurons in the hidden layer as depicted in Fig. 4.

Having in mind the shape of the curve, the above prediction example may be stated as a very successful one. To check the behavior of the method on a larger set of examples we repeated the above process 11 times by moving the time window by one step to generate 11 consecutive predictions. The obtained results are presented in Fig 6. All predictions were obtained by ANNs with five neurons in the hidden layer.

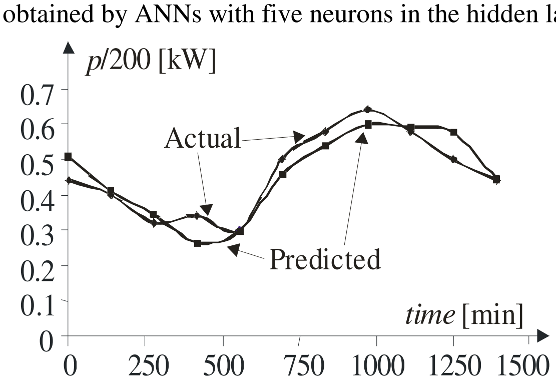

Fig. 6. Predicted and actual values (Thursday/Friday January 30/31, 2009).

As can be seen from fig. 6 not every prediction was as successful as the one related to the 07.02 hours in the morning. Nevertheless, most of the predictions are within several percent of the actual value and only two of them are significantly apart of the wanted one. The largest deviation is measured to be 14% only. One may notice from this diagram, however, that the error in prediction for a given instant does not influence the next prediction i.e. there is no accumulation of the error. The reason for that is the fact that every prediction step in our method represents a separate extrapolation task.

IV. Conclusion

The problem of short-term (hourly) forecasting of the electricity load at suburban level is considered. It is claimed first that despite "periodicity" of the phenomenon under consideration the data from previous days being from the given week or from the same day of the previous weeks are not convenient enough to be used directly. Then, architecture of ANN was proposed for the solution of the problem. Both, previous data for the given day and previous data from the same day of the previous weeks were used in the training set. The later is related to the habits of the consumers that generally influence the consumption.

Encouraging results were obtained. More detailed study will be performed in order to further improve the predictions. For example, the influence of the number of signals feed-back will be studied as well as the number of previous weeks.

It is our intention to integrate the prediction function into the remote-electricity-metering and billing system so enabling the electricity supplier to predict the load.

Finally, we want to stress the fact that we are here dealing wit extrapolation. That is to be opposed to generalization what the obvious property of the ANNs is. In our case the generalization is expressed by the excellent approximation of the input function. Namely, the ANN has the same response as is the input in between of the time interval given. That, however, is not forecasting. One should leave the input interval and predict the response value outside of the given time segment in order to achieve forecasting. That is what we do. This fact is stressed here since many of the published results are ambiguous in the sense that the term forecasting is used while interpolation is performed.

The method proposed implements ANNs that are generally widely known but creates new architecture that is fully original. It is our opinion that it may be implemented with equal success to one-step-ahead prediction of broader class of time series exhibiting inherent quasi-periodical properties.

References

[1] A. S. Mandel', "Method of analogs in prediction of short time series: an expert-statistical approach," Automation and Remote Control, Vol. 65, No. 4, 2004, pp. 634–641.

[2] G. Gross, F. D. Galiana, "Short-term load fore-casting," Proceedings of the IEEE, Vol. 75, No. 12, 1987, pp. 1558−1573.

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

[4] F. Cavallaro, "Electric load analysis using an artificial neural network," International Journal of Energy Research, Vol. 29, No. 5, 2005, pp. 377–392.

[5] A. Hainoun, "Construction of the hourly load curves and detecting the annual peak load of future Syrian electric power demand using bottom-up approach," Int. J. of Electrical Power and Energy Systems, Vol. 31, 2009, pp. 1–12.

[6] J. E. P. Box, G. Jenkins, "Time Series Analysis, Forecasting and Control," San Francisco, CA: Holden-Day, 1990.

[7] D. C. Montgomery, C. L. Jennings, M. Kulahci, "Introduction to Time Series Analysis and Forecasting", Hoboken, NJ: Wiley, 2008.

[8] J. Milojković, V. B. Litovski, "New methods of prediction implemented for sustainable development," in Proc. of the 51th Conf. ETRAN, Herceg Novi, Montenegro; June 2007, Paper no. EL1.8 (in Serbian).

[9] J. Milojković, V.B. Litovski, "Comparison of some ANN based forecasting methods implemented on short time series," in Proc. of the 9th IEEE Symposium NEUREL, Belgrade, 2008, pp. 175−178.

[10] E.A. Plummer, "Time series forecasting with feed-forward neural networks: guidelines and limitations", M.S. Thesis, University of Wyoming, Laramie, 2000.

[11] B.G. Zhang, "Forecasting with artificial neural networks: The state of the art," International Journal of Forecasting, Vol. 14, No. 1, 1998, pp. 35–62.

[12] J. Connor, R. D. Martin, "Recurrent neural networks and robust time series prediction," IEEE Trans. On Neural Networks, Vol. 5, No. 2, 1994, pp. 240–254.

[13] K. Brännäs, J. Hellström, "Forecasting based on Very Small Samples and Additional Non-Sample Information," Umeå Economic Studies 472, Umeå University, Sweden, 1998.

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

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

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