A. Bakirtzis, S. Kiartzis, V. Petridis and Ath. Kehagias.
"Short Term Load Forecasting using a Bayesian Combination Method"

This paper has appeared in the journal:
Int. Journal of Electrical Power and Energy Systems, Vol. 19, pp.171-177, 1997.

A BAYESIAN COMBINATION METHOD FOR SHORT TERM LOAD FORECASTING

A. BAKIRTZIS, S. KIARTZIS, V. PETRIDIS AND A. KEHAGIAS

Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki, Greece

Abstract: This paper presents the Bayesian Combined Predictor (BCP), a probabilistically motivated predictor for Short Term Load Forecasting (STLF) based on the combination of an artificial neural network (ANN) predictor and two linear regression (LR) predictors. The method is applied to STLF for the Greek Public Power Corporation dispatching center of the island of Crete, using 1994 data, and daily load profiles are obtained. Statistical analysis of prediction errors reveals that during given time periods the ANN predictor consistently forecasts better for certain hours of the day, while the LR predictors forecast better during for the rest. This relative prediction advantage may change over different time intervals. The combined prediction is a weighted sum of the ANN and LR predictions, where the weights are computed using an adaptive update of the Bayesian posterior probability of each predictor, based on their past predictive performance. The proposed method outperforms both ANN and LR predictions.

This paper appeared in Electrical Power and Energy Systems, Vol.19, pp.171-177, 1997

1. INTRODUCTION

The formulation of economic, reliable and secure operating strategies for a power system requires accurate short term load forecasting (STLF). The principal objective of STLF is to provide load predictions for the basic generation scheduling functions, the security assessment of a power system and for dispatcher’s information. A large number of computational techniques have been used for the solution of the STLF problem; these make use of statistical models, expert systems or artificial neural networks (ANN); in addition the hybrid method of fuzzy neural networks has appeared in the bibliography recently.

Statistical STLF models can be generically separated into regression models [1] and time series models [2]; both can be either static or dynamic. In static models, the load is considered to be a linear combination of time functions, while the coefficients of these functions are estimated through linear regression or exponential smoothing techniques [3]. In dynamic models weather data and random effects are also incorporated since autoregressive moving average (ARMA) models are frequently used. In this approach the load forecast value consists of a deterministic component that represents load curve periodicity and a random component that represents deviations from the periodic behavior due to weather abnormalities or random correlation effects. An overview over different statistical approaches to the STLF problem can be found in [4]. The most common (and arguably the most efficient) statistical predictors apply a linear regression on past load and temperature data to forecast future load. For such predictors, we will use the generic term Linear Regression (LR) predictors.

Expert systems have been successfully applied to STLF [5, 6]. This approach, however, presumes the existence of an expert capable of making accurate forecasts who will train the system.

The application of artificial neural networks to STLF yields encouraging results; a discussion can be found in [7]. The ANN approach does not require explicit adoption of a functional relationship between past load or weather variables and forecasted load. Instead, the functional relationship between system inputs and outputs is learned by the network through a training process. Once training has been completed, current data are input to the ANN, which outputs a forecast of tomorrow's hourly load. One of the first neural-network-based STLF models was a three-layer neural network used to forecast the next hour load [8]. A minimum-distance based identification of the appropriate historical patterns of load and temperature used for the training of the ANN has been proposed in [9], while both linear and non-linear terms were adopted by the ANN structure. Due to load curve periodicity, a non-fully connected ANN consisting of one main and three supporting neural networks has been used [10] to incorporate input variables like the day of the week, the hour of the day and temperature. Various methods were proposed to accelerate the ANN training [11], while the structure of the network has been proved to be system depended [12]. The most recent proposed ANN models for STLF tune the model performance efficiency based on the practical experience gained by the model implementation to Energy Management Systems (EMS), [13, 14, 15].

Hybrid neuro-fuzzy systems applications to STLF have appeared recently. Such methods synthesize fuzzy-expert systems and ANN techniques to yield impressive results, as reported in [16, 17].

Each of the methods discussed above has its own advantages and shortcomings. Our own experience is that no single predictor type is universally best. For example, an ANN predictor may give more accurate load forecasts during morning hours, while a LR predictor may be superior for evening hours. Hence, a method that combines various different types of predictors may outperform any single “pure” predictor of the types discussed above.

In this paper we present such a “combination” STLF method, the so called *Bayesian Combined Predictor (BCP),* which utilizes conditional probabilities and Bayes’ rule to combine ANN and LR predictors [18, 19, 23]. We proceed to describe the “pure” LR and ANN predictors and the BCP combination method. Then we present results and statistics of BCP forecasts for the Greek Public Power Corporation (PPC) dispatch center of the island of Crete during 1994.

2. STLF USING “PURE” PREDICTORS

The problem we are considering is the short term load forecasting for the power system of the island of Crete. In the summer of 1994 this system had a peak load of about 300 MW; power is supplied by PPC. Load and temperature historical data are available for the years 1989 to present. In this section we present three approaches to STLF which make use of so called “pure” predictors, namely two LR and one ANN predictor. We call these “pure” predictors, to distinguish them from the “combined” predictor which we will present in the next section.

2.1 "LONG PAST" LR PREDICTOR

This predictor performs a straightforward linear regression on two time series: daily loads (*for a given hour of the day*) and maximum daily temperature. There are $M+N$ inputs, where M is the number of past loads (*for the given hour of the day*) and N is the number of past temperatures used. Several values of M, between 21 and 56, have been employed. This means we use data from the last 21 to 56 days; hence the designation *"long past"*. (The best value turned out to be 35.) Output is tomorow’s load *for the given hour*. Hence, for a complete 24-hour load forecast, we need 24 separate predictors. The regression coefficients are determined by least squared error training; this is achieved using a standard matrix inversion routine, which takes between 1 and 2.5 secs (depending on the values of M and N) on a 66 Mhz 486 PC. The training phase is performed only once, offline. It should also be mentioned that the hourly load data were analysed and "irregular days", such as national and religious holidays, major strikes, election days, etc, were excluded from the training data set and replaced by equivalent regular days; of course this substitution was performed only for the training data. Training utilized load and temperature data for the years 1992 and 1993. Training error (computed as the ratio of forecast error divided by the actual load, averaged over all days and hours of the training set) was 2.30%. It must be mentioned that there was a “ceiling” effect as to the possible reduction of forecast error. While training error could be reduced below 2.30% by the introduction of more regression coefficients, this improvement was not reflected in the test error. This is the familiar “overfitting” effect and will be further discussed in Section 4.

2.2 "SHORT PAST" LINEAR PREDICTOR

This is very similar to the previous method. Again, it utilizes straightforward linear regression on the time series of loads; but now loads of all hours of the day are used as input., in addition to maximum *and* minimum daily temperature. There are ($24\times M+2\times N$) inputs, where M is the number of past loads (*for all hours of the day*) and N is the number of past temperatures used. Several values of M, between 1 and 8, have been employed. We have found that the best value of M is 4, which means data from four past days are used. For a given forecast day, we use the two immediately previous days and the same weekday of the previous two weeks.; hence this predictor uses a relatively *"short past"*, as compared to the one of Section 2.1. Output is tomorow's load *for every hour of the day*. The regression coefficients are determined by least squared error training; this is achieved using a standard matrix inversion routine, which takes between 0.1 and 1.5 secs (depending on the values of M and N) on a 66 Mhz 486 PC. The remarks of Section 2.1 on training and overfitting apply here as well. Training error (computed as the ratio of forecast error divided by the actual load, averaged over all days and hours of the training set) was 2.36%.

2.3 ANN PREDICTOR

A fully connected three layer feedforward ANN was used in this method. The ANN comprises of 57 input neurons, 24 hidden neurons and 24 output neurons representing next day's 24 hourly forecasted loads. The first 48 inputs represent past hourly load data for today and yesterday. Inputs 49-50 are maximum and minimum daily temperatures for today. The last seven inputs, 51-57, represent the day of the week, bit encoded. Other input variables were also tested but they did not improve the performance of our model. The ANN was trained by being presented with a set of input-desired output patterns until the average error between the desired and the actual outputs of the ANN over all training patterns is less than a predefined threshold. The minimization of the output error is achieved through a gradient algorithm. The well known back propagation algorithm [21] was used for the ANN training. The hourly load data were carefully analysed and all "irregular days", such as national and religious holidays, major strikes, election days, etc, were excluded from the training data set. Special logic for the treatment of missing data has also been incorporated in the data analysis software. The training data set consists of 90+4× 30=210 input/output patterns created from the current year and the four past years historical data as follows: 90 patterns are created for the 90 days of the current year prior to the forecast day. For every one of the 4 previous years, another 30 patterns are created around the dates of the previous years that correspond to the current year forecast day. Initial offline training takes between 4 and 9 secs on a 66 MHz 486 PC. The ANN parameters are then updated online, on a daily basis through the following procedure. A new round of ANN training is performed on the most recent input/output patterns; the ANN parameters are initialized to those of the previous day. Since the training data sets of two consecutive days differ by only a few patterns, daily model parameter updating is very efficient. Online training takes between 1 and 3 secs *per day*. The network is trained until the average error becomes less than 2.5%. It was observed that further training of the network (to an error 1.5% for example) did not improve the accuracy of the forecasts. Training of the ANN to a very small error may results in data overfitting.

3. THE BAYESIAN COMBINED PREDICTOR

We now present the BCP, a new type of predictor, which outperforms all three predictors presented in the previous section. The BCP is based on probabilistic concepts; namely conditional probability and Bayes' rule. The original idea appears in [18, 19]; see also [23]. We have used the original probabilistic formulation, as well as nonprobabilistic generalizations in the context of Time Series *Classification*. This is an application of the so called *Predictive Modular Neural Networks (PREMONNs)* [22, 23] For a further discussion see Section 5. In the rest of this section, when we consider a load ime series y_t and its forecasts y_t^k , we are referring to a fixed hour of the day, say 1am, 7 pm and so on. The arguments presented are exactly the same for any hour considered.

3.1 RECURSIVE APPLICATION OF BAYES' RULE

Suppose that the load time series is in fact produced by one of the three models listed in Section 3: long-past LR, short-past LR or ANN. By this we mean that

(1) $$y_t = y_t^k(y_{t-1}, y_{t-2}, \ldots, y_1) + e_t^k, \qquad k = 1,2,3$$

where y_t is the actual load, y_t^k is the forecast (y_t^1 is the long past LR forecast, y_t^2 is the short past LR forecast and y_t^3 is the ANN forecast) e_t^k is the respective forecast error and t is current time. *However, eq.*(1) *will actually hold true only for one value of k* (1, 2 or 3). We do not know which of the three is the correct or "true" model. We express this uncertainty by introducing a variable Z, which can take the values 1, 2 or 3.

The *conditional posterior probability* p_t^k (for k=1,2,3, t =1,2, ...) is defined by

(2) $$p_t^k = \mathrm{Prob}(Z = k \mid y_{t-1}, y_{t-2}, \ldots, y_1)$$

and the *prior probability* p_0^k (for k=1,2,3) is defined by

(3) $$p_0^k = \mathrm{Prob}(Z = k \text{ at time } t = 0).$$

Conditioning on the observed loads $y_{t-1}, y_{t-2}, \ldots, y_1$ expresses the fact that when new load data become available they can be used to test which model most closely conforms to the data. We will show how this test is performed presently. Let us also remark that at t=0, when no load data have been observed, we can choose p_0^k according to our prior knowledge (e.g. set p_0^1 =.05, p_0^2 =.05, p_0^3 =.90, if we strongly believe in the ANN model), or we can choose $p_0^1 = p_0^2 = p_0^3$ =0.333, in case we have no prior knowledge. Note that the p_0^k's must add up to one, since they are probabilities. Given p_0^k, we proceed to recursively compute p_t^k, for t =1, 2, ... , starting with Bayes' rule

$$p_t^k = \frac{\mathrm{Prob}(y_t, Z = k \mid y_{t-1} \ldots y_1)}{\sum_{l=1}^{3} \mathrm{Prob}(y_t, Z = l \mid y_{t-1} \ldots y_1)} \tag{4}$$

and the fact that

$$\mathrm{Prob}(y_t, Z = k \mid y_{t-1}, y_{t-2}, \ldots, y_1) = \mathrm{Prob}(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1, Z = k) \cdot p_{t-1}^k \tag{5}$$

Substituting (5) into (4) we get (for k = 1, 2, 3)

$$p_t^k = \frac{p_{t-1}^k \cdot \mathrm{Prob}(y_t^k \mid y_{t-1} \ldots y_1, Z = k)}{\sum_{l=1}^{3} p_{t-1}^l \cdot \mathrm{Prob}(y_t^l \mid y_{t-1} \ldots y_1, Z = l)} \tag{6}$$

Eq.(6) gives a recursive method to compute p_t^k from p_{t-1}^k, provided we know $\mathrm{Prob}(y_t^k \mid y_{t-1} \ldots y_1, Z = k)$. Let us now determine this probability. With eq.(1) we have already assumed that the forecast error is given by $e_t^k = y_t - y_t^k$. Assume further that e_t^k is a Gaussian white noise time series. In that case we have

$$\mathrm{Prob}(y_t^k \mid y_{t-1} \ldots y_1, Z = k) = \mathrm{Prob}(e_t^k = y_t - y_t^k \mid y_{t-1} \ldots y_1, Z = k) = e^{-\left(\frac{y_t - y_t^k}{\sqrt{2\pi}\sigma}\right)^2} \tag{7}$$

which finally yields

$$p_t^k = \frac{p_{t-1}^k \cdot e^{-\left(\frac{y_t - y_t^k}{\sqrt{2\pi}\sigma}\right)^2}}{\sum_{l=1}^{3} p_{t-1}^l \cdot e^{-\left(\frac{y_t - y_t^l}{\sqrt{2\pi}\sigma}\right)^2}} \tag{8}$$

Eq. (8) and the regression equations for y_t^1, y_t^2, y_t^3, constitute the algorithm for recursive computation of the Bayesian posterior probabilities. The validity of the algorithm depends on our assumptions, namely that (a) the load time series is produced by one of the three models of Section 3 and (b) the forecast error is Gaussian white noise. When these assumptions hold, p_t^k expresses the probability that model k actually generates the observed load data; this probability is conditional, dependent on observations up to time t.

There is an alternative, nonprobabilistic interpretation of the p_t^k 's. Note that in eq.(8), models which have large forecast errors (large $\left(y_t - y_t^k\right)^2$) are heavily penalized resulting in decreased p_t^k. So one can consider the BCP algorithm as a heuristic credit assignment scheme: the model that best forecasts the observed load data is the one with highest p_t^k (and so with highest conditional probability, under the Bayesian interpretation).

In the final analysis, the validity of our assumptions and the BCP method will be judged on how efficient the p_t^k 's are in forecasting the load time series. In the next section we present a forecast method that makes use of the p_t^k 's as well as the pure predictors.

3.2 DERIVATION OF THE BCP

There are at least two ways to use the p_t^k 's for load forecasting. One could at every time step use the forecast of the model with maximum p_t^k ; since this model is most likely to have produced the load time series, it must on the average have smaller forecast error. However there is another way to use the p_t^k 's, which yields the BCP.

Start with following well known fact of probability theory [24]: of all the predictors of y_t that depend on past $y_{t-1}, y_{t-2}, \ldots, y_1$ values, the predictor with minimum mean square error is the conditional mean

(9) $$y_t^* = E\left(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1\right).$$

It is also a standard probabilistic result [24] that

(10) $$E\left(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1\right) = E\left[E\left(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1, Z\right) \mid y_{t-1}, y_{t-2}, \ldots, y_1\right].$$

Now, *assuming that Z can only take the values* 1, 2, 3, we combine (9) and (10) to obtain

(11) $$y_t^* = p_t^1 y_t^1 + p_t^2 y_t^2 + p_t^3 y_t^3.$$

It should now be obvious why we call this a “combined” predictor, as opposed to the “pure” predictors of Section 2. Eq.(11) together with eq.(8) and the regression equations for y_t^1, y_t^2, y_t^3, give a complete description of the BCP. Theoretically, this predictor is superior to any “pure” predictor, *if* the several assumptions presented above hold. An alternative, heuristic interpretation of eq.(11) is the following: we combine the three possible forecasts into a weighted sum, where p_t^k, the weight given to forecast *k*, depends on its past predictive performance.

Aside from interpretations, the practical expedience of the BCP will be judged by its performance on a practical STLF task. Such an application is presented in the next section.

4. EXPERIMENTS

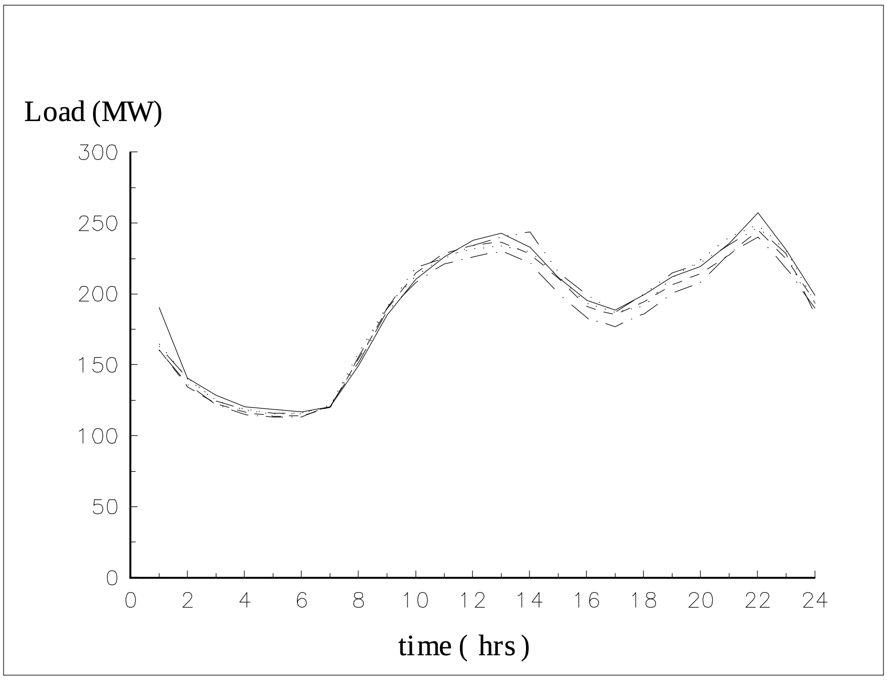
In this section we compare the new BCP to the three “pure” predictors, on the task of STLF for the power system of Crete island. Actually we consider 24 BCP's and 24×3=72 "pure" predictors, one predictor of each type corresponding to every hour of the day. The three pure predictors of each hour have been trained as explained in Section 2. The BCP's needs no training per se; they simply combine the forecasts of the respective “pure” predictors by application of eq. (8) and (11). Results presented correspond to the period from April to June 1994. In Table I we present average errors for the four types of predictors used, and for the 24 hours of the day. The reader can observe that the BCP outperforms all “pure” predictors. In Fig.1 a comparative plot of the errors obtained is presented.

Table I. Hourly average errors for June to September 1994.

	errors (%)			
hours	long past LR	short past LR	ANN	BCP
1	2.89	1.92	2.26	1.96
2	2.22	1.72	2.09	1.63
3	2.14	1.93	2.50	1.69
4	2.55	2.38	2.49	2.17
5	2.71	2.23	2.44	2.31
6	2.55	2.31	2.41	2.16

7	2.47	2.15	2.16	2.01
8	3.11	3.09	2.72	2.38
9	2.57	2.85	2.17	2.07
10	2.72	2.95	2.53	2.36
11	2.53	2.86	2.72	2.32
12	2.44	2.87	2.91	2.43
13	2.24	3.07	2.85	2.16
14	2.29	3.19	2.51	2.10
15	2.26	2.77	2.36	1.95
16	2.29	2.93	2.44	2.12
17	2.35	3.30	2.38	2.13
18	1.82	2.96	2.41	2.06
19	1.98	2.97	2.51	2.16
20	2.30	3.17	2.52	2.34
21	1.83	2.98	2.76	1.99
22	1.95	2.58	2.40	1.93
23	1.70	1.84	1.92	1.56
24	2.20	1.96	2.13	1.78
total	2.34	2.62	2.44	2.07

Fig.1 A typical daily load curve, pure and combined rforecasts



In Table II average total errors for training and test data are given for various values of the total number of regression coefficients. The rows in bold letters correspond to the "pure" predictors actually used in the BCP combination. The reader can see that an increase in the number of regression coefficients yields improved training errors but test errors remain the same or even increase. This is an instance of overfitting. On the other hand, the BCP also uses an increased number of coefficients, namely the sum of the numbers of coefficients of the three predictors. In our case this would be 1200+2448+1200=4848. While we have not tried to train any pure predictor with 4848 free regression coefficients, extrapolating from Table II, one expects the test error to be actually larger than that of any pure predictor with fewer coefficients. However, BCP increases the number of coefficients in a judicious and strcuctured way, resulting in the marked decrease of test error to 2.07. Similarly, training time scales very efficiently for the BCP. It actually is 1.38+0.86+2.35=4.59 secs (the total time for training the three pure predictors). Using the data in Table II and interpolating linearly, one would expect training time to be around 6.79 secs. In fact, linear interpolation is probably too optimistic. For the LR predictors, it is known that matrix inversion time scales cubically with the size of the problem; as for the ANN predictor, increasing the size of the network may result in failure of the training procedure (e.g. attainment of local minima).

Finally, it is quite instructive to observe the evolution of the posterior probabilities of the three pure predictors for two different hours. In Fig. 2a we plot the evolution of the posteriors for the hour 1pm and in Fig.2b for the hour 1am. The reader will see that in Fig.2a the highest probability is generally assigned to the LP LR predictor, even though over short time intervals one of the other two predictors may outperform it. Similarly, in Fig.2b the highest probability is generally assigned to the SP LR predictor, even though over short time intervals one of the other two predictors may outperform it. These results are consistent with the general test errors of Table I; the additional information presented in Fig.2 is that a predictor that generally performs poorly, may still outperform its competitors over short time intervals; in such cases the BCP will take this improved performance into account,as evidenced by the adaptively changing posterior probabilities. This explains why the BCP is generally *better than the best pure predictor.*

Table II. Error Dependence on Number of Parameters

Predictor Type	Total Nr. of Parameters	Train Error (%)	Test Error (%)	Training Time (secs)
LP LR	864	2.41	2.44	1.05
LP LR	1032	2.34	2.37	1.22
LP LR	**1200**	**2.30**	**2.34**	**1.38**
LP LR	1368	2.27	2.36	1.60
LP LR	1536	2.26	2.38	1.85
LP LR	1704	2.25	2.39	2.09
SP LR	1224	2.61	2.72	0.12
SP LR	1248	2.53	2.69	0.13
SP LR	1800	2.43	2.68	0.34
SP LR	**2424**	**2.36**	**2.62**	**0.86**

SP LR	2448	2.33	2.73	1.13
SP LR	3000	2.32	2.72	1.45
ANN	**1200**	**2.50**	**2.44**	**2.35**
ANN	1200	2.00	3.76	8.92

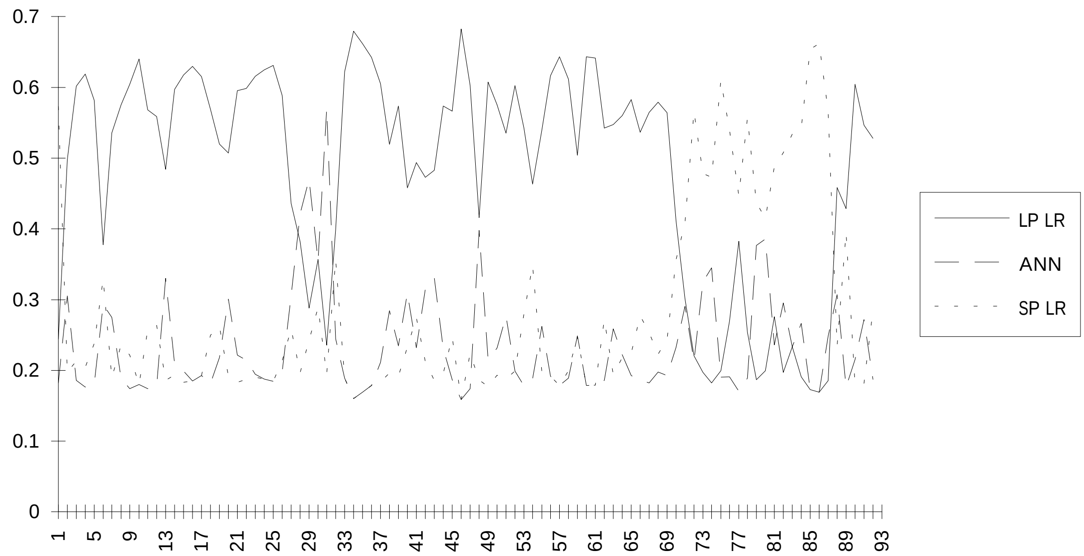


Fig.2a

Evolution of posterior probabilities for the predictors of 1pm load, over the period July 1st, 1994 to September 30th, 1994.

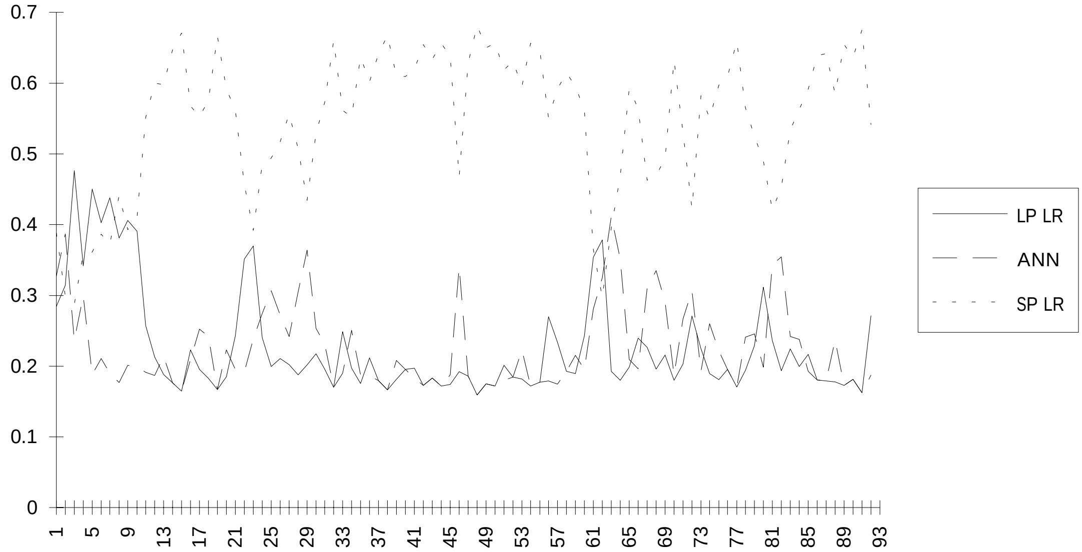


Fig.2b

Evolution of posterior probabilities for the predictors of 1am load, over the period July 1st, 1994 to September 30th, 1994.

5. CONCLUSIONS

We have applied a new predictor, BCP, to the problem of STLF. This predictor is based on conditional probability concepts that enable us to *adaptively* combine different types of predictors and extract from each one of them the best possible performance over a given time period. Hence BCP enables to pick the best features of each predictor used. From a somewhat different point of view, BCP is a judicious way to combine a large number of regression coefficients avoiding overfitting problems. Finally, yet another point of view discards the probabilistic interpretation in a favor of a heuristic one: predictors with inferior predictive power are penalized and their forecasts are underweighted in a weighted sum combined predictor; the weights are updated adaptively. At any rate, our experiments indicate that the BCP method outperforms all conventional, "pure" prediction methods in the test problem we have considered.

Finally, it is worth mentioning that BCP is a special case of a more general class of time series predictors/classifiers, the so called PREMONNs [22, 23]. Several simplifications of the general PREMONN theory have been made in this paper. For instance, in the general case it is not necessary to use three "pure" predictors; any finite number K can be used with the same methodology. Also, predictors need not be only of the LR or ANN types; fuzzy, hybrid etc. types of predictors can be used. The probabilistic combination method can also be discarded; a number of nonprobabilistic combination mathods are available.

REFERENCES

[1] A. D. Papalexopoulos and T. C. Hesterberg, "A Regression-Based Approach to Short Term System Load Forecasting," *IEEE Trans. on Power Systems*, Vol. 5, No. 4, pp. 1535-1547,1990.

[2] S. Vemuri, W. L. Huang and D. J. Nelson, "On-Line Algorithms for Forecasting Hourly Loads of an Electric Utility," *IEEE Trans. Power App. & Syst.*, Vol. PAS-100, No. 8, pp. 3775-3784, 1981.

[3] W.R. Christiaanse, "Short term load forecasting using general exponential smoothing," *IEEE Trans. Power App. & Syst.*, Vol. PAS-90, No. 2, pp. 900-911, 1971.

[4] G. Gross and F.D. Galiana, "Short term load forecasting,"*Proc. IEEE*, Vol. 75, No. 12, pp. 1558-1573, 1987.

[5] S. Ranman and R. Bhatnagar, "An expert system based algorithm for short term load forecast," *IEEE Trans, PWRS*, Vol. 3, pp.392-399, 1988.

[6] K.L. Ho, Y.Y. Hsu, C.F. Chen, T.E. Lee, C.C. Liang, T.S. Lai, and K.K. Chen, "Short term load forecasting of Taiwan power system using a knowledge-based expert system," Paper 90 WM 259-2 PWRS, presented at the IEEE/PES 1990 Winter Meeting.

[7] D. Niebur et al., ‘Artificial neural networks for Power Systems,’ CIGRE TF38.06.06 Report, ELECTRA, April 1995, pp. 77-101.

[8] D.C. Park, M.A. El-Sharkawi, R. J. Marks, L.E. Atlas, and M.J. Damborg, "Electric Load Forecasting Using an Artificial Naural Network," *IEEE Trans. on Power Systems*, Vol. 6, No. 2, pp. 442-449, 1991.

[9] T.M. Peng, N.F. Hubele, and G. G. Karady, "Advancement in the Application of Neural Networks for Short Term Load Forecasting," *IEEE Trans. on Power Systems*, Vol. 7, No. 1, pp. 250-258, 1992.

[10] S.T. Chen, D.C. Yu, and A.R. Moghaddamjo, "Weather Sensitive Short-Term Load Forecasting Using Nonfully Connected Artificial Neural Network," *IEEE Trans. on Power Systems*, Vol. 7, No. 3, pp. 1098-1105, 1992.

[11] K.L. HO, Y. Y. Hsu, and C. C. Yang, "Short Term Load Forecasting Using a Multilayer Neural Network with an Adaptive Learning Algorithm," *IEEE Trans. on Power Systems*, Vol. 7, No. 1, pp. 141-149, 1992.

[12] C.N. Lu, H.T. Wu and S. Vemuri, "Neural Network Based Short Term Load Forecasting," *IEEE Trans. on Power Systems*, Vol. 8, No. 1, pp. 336-342, 1993.

[13] A.D. Papalexopoulos, S. How and T.M. Peng: "An Implementation of a Neural Network based Load Forecasting Model for the EMS", Paper 94 WM 209-7 PWRS presented at the *IEEE/PES 1994 Winter Meeting.*

[14] O. Mohammed, D. Park, R. Merchant, T. Dinh, C. Tong, A. Azeem, J. Farah and C. Drake: "Practical Experiences with an Adaptive Neural Network Short Term Load Forecasting System", Paper 94 WM 210-5 PWRS presented at the *IEEE/PES 1994 Winter Meeting*.

[15] A. Bakirtzis, V. Petridis, S. Kiartzis, M. Alexiadis and A. Maissis, “A Neural Network Short Term Load Forecasting Model for the Greek Power System”, accepted for presentation in *IEEE/PES 1995 Summer Meeting*.

[16] D. Srinivasan, C.S. Chang and A.C. Liew, “Demand Forecasting Using Fuzzy Neural Computation, with special emphasis on weekend and public holiday forecasting”, paper 95 WM 158-6-PWRS presented in *IEEE/PES 1995 Winter Meeting*.

[17] A. Bakirtzis, J. Theocharis, S. Kiartzis and K. Satsios, “Short Term Load Forecasting Using Fuzzy Neural Networks”, paper 95 WM 155-2-PWRS presented in *IEEE/PES 1995 Winter Meeting*.

[18] C.G. Hilborn and D.G. Lainiotis, “Optimal Estimation in the Presence of Unknown Parameters”, *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 5, No. 1, pp. 38-43, 1992.

[19] D.G. Lainiotis, “Adaptive Estimation and Structure Identification”, *IEEE Trans. on Automatic Control*, Vol. ?, No. ?, pp. ??, 1971.

[20] D.G. Lainiotis and K. Plataniotis, “Adaptive Dynamic Neural Network Estimators", in *Proceedings of IEEE Int. Conf. on Neural Networks*, pp.4736-4745, Orlando, 1994.

[21] B. Widrow and M.A. Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline and Back Propagation", *Proc. IEEE*, Vol. 78, No. 9, pp.1415-1442, 1990.

[22] V. Petridis and A. Kehagias, "Theoretical Analysis of Bayesian Partition Algorithms for Modular Neural Networks", Internal Report, Div. of Electronics and Comp. Eng., Dept. of Elec. Eng., Aristotle Univ. of Thessaloniki, 1994.

[23] V. Petridis and A. Kehagias, "A Recurrent Network Implementation of Time Series Classification", submitted in *Neural Computation*.

[24] P. Billingsley, *Probability and Measure*, Wiley, 1986.