Proceedings of the IV Brazilian Conference on Neural Networks - IV Congresso Brasileiro de Redes Neurais
pp. 007-012, July 20-22, 1999 - ITA, São José dos Campos - SP - Brazil

Neural Network Based Short-Term Electric Load Forecasting with Confidence Intervals

Luciano S. Moulin[1], Alexandre P. Alves da Silva[1]
[1]Escola Federal de Engenharia de Itajubá / Instituto de Engenharia Elétrica
Av. BPS, 1303 - Itajubá, MG, 37500-000, Brasil
E-mails: lsm@iee.efei.br, alex@iee.efei.br

Abstract

Through traditional statistical models, like ARMA and Multilinear Regression, confidence intervals can be computed for the short-term electric load forecasting, assuming that the forecast errors follow a normal probability distribution. In this paper, the 1-24 steps ahead load forecasts are obtained through Multi-Layer Perceptrons trained by the back-propagation algorithm. Three techniques for the computation of confidence intervals for this neural network based short-term load forecasting are presented: (i) Error Output, (ii) Resampling and (iii) Multilinear Regression adapted to neural networks. A comparison of the three techniques is performed through simulations of on-line forecasting.

1. Introduction

With power systems growth and the increase in their complexity, many factors have become influential in the electric power generation and consumption (load management, energy exchange, spot pricing, independent power producers, non-conventional energy, etc.); therefore, the forecasting process has become even more complex, and more accurate forecasts are needed. The relationship between the load and its exogenous factors is complex and non-linear, making it quite difficult to model through conventional techniques, such as time series and linear regression analysis. Besides not giving the required precision, most of the traditional techniques are not robust enough. They fail to produce accurate forecasts when quick weather changes occur. Other problems include noise immunity, portability and maintenance [1].

Neural Networks (NNs) have succeeded in several power system problems, such as: planning; control; analysis; protection; design; load forecasting; security analysis; and fault diagnosis. The last three ones are the most popular [2]. The NN ability in mapping complex non-linear relationships is responsible for the growing number of their applications to the Short-term Load Forecasting (STLF) [3-5]. Several utilities over the world have been applying NNs for load forecasting in an experimental or operational basis [1, 2, 4].

Despite its success in that application, up to now the uncertainty in the NN based forecasts and the estimation of confidence intervals (CIs) have been ignored. Box & Jenkins have studied the subject of CIs in ARMA models [6, 7]. CIs can also be estimated for forecasts through Multilinear Regression [7]. However, those models are based on normality and independence assumptions which do not always hold in practical applications.

One should not produce a forecast of any kind without an idea of its reliability. However, there are many difficulties in computing those indices for non-linear models. The NN literature is almost devoid of information on that subject [8, 9]. Confidence intervals should be as narrow as possible, while encompassing a number of true values that justifies its reliability.

In Section 2, three techniques for the computation of CIs for the NN based STLF are presented; in Section 3, their implementation and the NN structure are detailed; in Sections 4 and 5, those techniques are compared through simulations of on-line forecasting.

2. Techniques for CI estimation

Among the three presented techniques for CI estimation, only the Multilinear Regression (MR) assumes that the forecasting errors follow normal probability distribution. The Resampling (RE) and Error Output (EO) techniques do not make that assumption.

The three methods are implemented through the same structure of a three-layer, fully connected Perceptron, trained by the back-propagation algorithm.

2.1. Error Output

In this technique, a NN with two outputs has been trained with output patterns corresponding to the hourly load and to the hourly load forecast error, respectively. Thus, confidence intervals are inherent to the forecasting process. That idea assumes that it is possible to capture patterns possibly present in the forecasting errors, as well as in the electric load.

In the NN training, the patterns for the error output neuron are computed at each training epoch. Each time the input training patterns are passed to the output, the load output errors are computed. Therefore, at each epoch a different training pattern is used for the error output.

As the training process converges for a set of weights with low load forecasting errors, it is expected that the error forecasts present low errors too, once the training patterns for the error output become more stable along the iterations. Otherwise, the training process would diverge.

During the training process, the absolute percent error of the load output neuron is taken as training pattern for the error output neuron, because they seem to be easier to be learned than the relative error. After the NN training, the error output is added and subtracted to the load forecast in order to create a symmetric CI.

In this technique, the confidence degree is no pre-defined. It is computed by counting the successful estimations of the CI widths, for the test set.

2.2. Resampling

This is a technique for the computation of confidence intervals [10] for the NN based STLF which does not assume any probability distribution for the load forecasting errors.

For each forecasting lead time, it is assumed that there is a resampling set with n errors obtained from a test set, each of them corresponding to the difference between the forecast load and the actual load value. Considering that the resampling set is representative of the actual data to be found in future forecasts, it can be inferred that the magnitude of the errors will be preserved in the future. Those errors are collected through a resampling window which moves one step each time in the resampling set. At each time, one error value is taken for each forecasting lead time. In Figure 1, a recursive forecasting process is shown with 3 lagged inputs, 1-4 steps ahead forecasts, with one-step ahead NN training. The upper dotted line shows the resampling window.

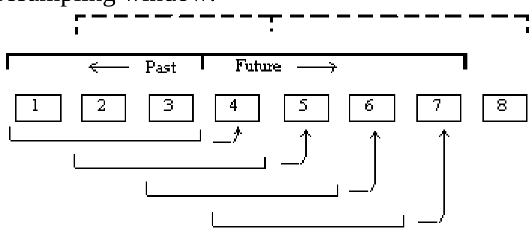

Fig. 1: Example of the resampling process

If that procedure is applied to the 1-24 steps ahead recursive forecasting (i.e., load forecasts feed the NN) with lagged inputs, the number of error measures for each forecasting horizon is given by:

$$n = \text{seriesl} - \text{maxlag} - \text{dist} + 1 \qquad (1)$$

where

seriesl → total length of the load series used to obtain the errors;

maxlag → maximum input lag used by the NN; and

dist → maximum forecasting horizon .

Sorting the n errors in ascending order (considering the signs), they can be represented by $e_{(1)}, e_{(2)},..., e_{(n)}$. The cumulative sampling distribution of the errors can be defined as the following:

$$S_n(e) = \begin{cases} 0, & e < e_{(1)} \\ r / n, & e_{(r)} \le e < e_{(r+1)} \\ 1, & e_{(n)} \le e \end{cases} \qquad (2)$$

that is to say, $S_n(e)$ is a fraction of the collection of errors containing those ones smaller or equal to e.

If n is taken large enough such that $S_n(e)$ is close to F(e), the true cumulative probability distribution, confidence intervals can be computed for forecasts by using the limits of that truncated collection, according to the desired confidence degree. These intervals are computed so they are symmetrical in probability.

The number of cases to discard in each tail is np, where p is the probability in each tail. However, np is generally a fractional number, so np is conservatively truncated, and (np - 1) is taken as the number of cases to discard in each tail.

E_p is denoted as an independent value of F, such that there is a probability p that an error is smaller or equal to E_p. This indicates that E_p is the lower confidence limit for future forecasts errors. Similarly, E_{1-p} is the upper limit.

The value $n.S_n(E_p)$ represents the estimate of how many elements in the collection of errors are smaller or equal to E_p. Considering that the errors are independent of each other, then $m = n.S_n(E_p)$ follows a binomial distribution:

$$B(m,n,p) = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \qquad (3)$$

independently of the distribution F.

B(m,n,p) represents the probability that exactly m, among n cases randomly sampled, are smaller than E_p. In fact, if B(m,n,p) is computed for m = 0, 1,..., n, it can be shown that the largest probability is obtained when m = np.

Some conditions should be satisfied in the computation of the confidence intervals:

- the resampling set reasonably represents the actual population;
- the error samples are independent and have the same, although unknown, probability distributions.

2.3. Multilinear Regression adapted to NNs

In this technique, if linear activation functions are used in the output neurons, a multilinear regression model (MR) [7] can be implemented, as shown in (4) and in Figure 2; the inputs are taken as the outputs of the hidden neurons, and the regression coefficients are taken as the connection weights of the output neurons.

$$y = b_1 + w_1.x_1 + w_2.x_2 + w_3.x_3 \tag{4}$$

The computation of the CIs is accomplished through the estimate of the forecasting variance:

$$(T - q)s^2 = \sum_{t=1}^{T} (y_t - \tilde{y}_t)^2 \tag{5}$$

where

s^2 → load series variance estimate.
T → number of elements in the training set;
q → number of neurons in the hidden layer plus one;
y_t → output pattern used in the NN training; and
$\tilde{y}_t$ → NN output (load forecast).

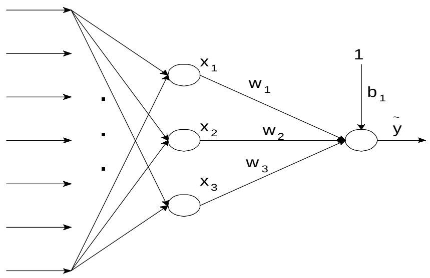

Fig. 2: Redefining the MR independent variables

In that way, with the variance estimate (5), with the desired confidence degree and with the NN inputs for the desired forecast, the corresponding confidence intervals can be computed, considering that

$$\frac{\tilde{y}_\tau - y_\tau}{s\sqrt{1 + \mathbf{x}_\tau' A^{-1} \mathbf{x}_\tau}} \tag{6}$$

follows a normal distribution, where

τ → time instant of the current load forecast;
$\tilde{y}_\tau$ → load forecast;
y_τ → true load value;
s^2 → y_τ's variance estimate;
$\mathbf{x}_\tau$ → column vector made up of the hidden neurons outputs and the bias of the output neuron (b_1 in (4)); and

$$A = \mathbf{x}_\tau . \mathbf{x}'_\tau \tag{7}$$

That is,

$$\tilde{y}_\tau - z_{\alpha/2}\, s\sqrt{1 + \mathbf{x}_\tau' A^{-1} \mathbf{x}_\tau} \le y_\tau \le \tilde{y}_\tau + z_{\alpha/2}\, s\sqrt{1 + \mathbf{x}_\tau' A^{-1} \mathbf{x}_\tau} \tag{8}$$

where

α → total area of the unit normal distribution tails, such that 1 - α equals the desired confidence degree; and
$z_{\alpha/2}$ → value of the standard statistics such that the probability of the unit normal distribution between $\pm\ z_{\alpha/2}$ is 1 - α.

3. NN structure and implementation

The selected NNs have seven inputs, three neurons in the hidden layer and two neurons in the output layer, with hyperbolic activation function in the hidden layer and linear function in the output layer.

The NN inputs correspond to the lagged values of the hourly load series, at 1h, 2h, 24h, 168h and 192h, and two inputs $HS(k) = sen(2\pi k / 24)$ and $HC(k) = \cos(2\pi k / 24)$ coding the hour of the day. The outputs correspond to the load forecast and to the load forecast error at hour k.

A seven-week window has been used for training and testing, with data grouping according to the day of the week. For each day of the week, a NN has been trained, applying the back-propagation algorithm with cross validation. The partition in training and test sets are determined randomly for each 100 epochs.

After the seven NNs are trained for the EO method, the same NNs are used for the RE and MR methods, each of which requires only the load output values of the NNs. Thus, the same forecast values are used to compare the three techniques.

During the NN training and test there is no special treatment for holidays. Special days have been excluded from the training set. A load series from a utility at Rio de Janeiro was used with load values from December 1994, January 1995 and February 1995. The load series has been scaled to values between 0 and 1.

After the one-step ahead training, the 1-24 steps ahead recursive load forecasts and the estimation of CIs were made for the 24 hours right after the training/test set. After that, the training/test set and the validation forecasts were moved one hour ahead to perform the next on-line forecast, with adaptive NN training. This procedure is repeated along all the validation period, that is from 01/27/95 through 02/23/95, with 649 values of on-line forecasts for each one of the 24 lead times of forecast. Results obtained for other seasons of 1995 and 1996 have been similar.

The confidence degrees of the EO method are estimated by using test subsets to obtain the rate between: (a) the number of actual load values lying inside the confidence limits and (b) the number of samples in the test subsets. This is done for each forecasting lead time in a recursive manner. For a reasonable comparison, those estimates are used as reference confidence degrees by the other two methods.

During the validation period the actual load values lying inside the confidence limits have been counted in order to check the confidence degrees of the three methods, and the percent confidence limits have been computed, for each forecasting lead time.

During the validation period, for those actual load values lying inside the confidence intervals estimated by the three methods, the rate between the absolute forecast error and the CI width, for each forecasting lead time, has also been computed:

$$RCI(k) = \frac{|LOAD(k) - LF(k)|}{|CL(k) - LF(k)|} \tag{9}$$

where

RCI(k) → relative confidence interval at hour k;
LOAD(k) → actual load value at hour k (MW);
LF(k) → load forecast at hour k (MW); and
CL(k) → lower limit (MW), if (LOAD(k) - LF(k)) < 0, or upper limit (MW), if (LOAD(k) - LF (k)) > 0.

RCI ranges from 0 to 1 and indicates the relation between the confidence limits and the forecasting errors. If its value is close to 1, it means that the confidence interval is just wide enough to encompass the actual load value.

Finally, in the MR technique, for the computation of s and A (equations (5) and (7)), the training set of the last training epoch is used.

4. Tests

Initially, three simulations were carried out using the last 1, 2 and 3 weeks of the test set to estimate the EO confidence degree (used as reference to compute the RE and MR confidence limits) and to build the resampling set for the RE technique, for each on-line forecast. However, the reference confidence degrees (CDs) are too small (in general, below 50%); therefore correction factors (ε's) are added to the NN error output, so that the confidence degrees can be enlarged. In order to obtain approximately the same CDs on different forecasting lead times, the EO technique only works with those correction factors. Different ε's have to be used for different forecasting horizons.

With the correction factors, the simulations with 1, 2 and 3 weeks periods have produced very similar results. For the MR technique, this period is irrelevant anyway. The one-step ahead mean absolute percent training error for the load output is 1,54% and the one-step ahead mean absolute training error for the error output is 0,94%. The one-step ahead test errors are 1,63% and 1,01%, respectively. Tables 1 to 4 show validation results using the 1 week period. The following evaluation indices are considered:

MAPE → 1-24 steps ahead mean absolute percent error for the load output during validation;
MAE → 1-24 steps ahead mean absolute error for the error output (%) during validation;
CD → 1-24 steps ahead mean reference confidence degrees;
EO-CD→ 1-24 steps ahead validation confidence degrees for the EO technique;
RE-CD→ 1-24 steps ahead validation confidence degrees for the RE technique; and
MR-CD→ 1-24 steps ahead validation confidence degrees for the MR technique (%).

As shown in Table 1, MR-CDs are close to CDs in the first hour, however it is too low on the others. EO-CDs and RE-CDs are very similar on the 24 hour forecasting.

Table 1: Reference and validation confidence degrees

HOUR	MAPE (%)	MAE (%)	CD (%)	EO-CD (%)	RE-CD (%)	MR-CD (%)
1	1.96	1.18	84.2	73.5	74.7	76.9
2	3.00	1.93	83.1	70.4	71.3	58.6
3	3.48	2.43	83.4	68.6	69.7	52.5
4	3.76	2.66	83.2	69.7	68.4	49.9
5	3.91	2.83	82.1	66.6	65.2	47.9
6	4.03	2.89	81.5	64.1	64.4	48.1
7	4.08	2.96	81.2	65.0	65.3	47.0
8	4.08	2.99	80.9	65.5	65.2	47.5
9	4.10	2.98	80.8	66.0	63.5	46.8
10	4.11	2.97	80.8	65.3	63.6	45.9
11	4.10	2.99	80.9	65.0	63.2	46.4
12	4.09	2.97	80.9	64.1	65.5	47.2
13	4.13	2.99	80.9	64.1	63.9	46.2
14	4.11	2.99	80.9	64.7	64.1	45.3
15	4.13	3.00	80.9	63.8	64.6	46.8
16	4.09	2.97	80.9	63.9	64.3	46.7
17	4.08	2.94	80.9	65.8	65.6	48.1
18	4.10	2.95	80.8	64.9	65.5	46.7
19	4.05	2.96	80.8	63.8	64.9	48.2
20	4.07	2.93	80.8	64.7	66.4	48.1
21	4.09	2.95	80.7	65.2	66.9	48.1
22	4.07	2.98	80.7	64.4	63.9	45.2
23	4.06	2.95	80.6	63.9	66.0	47.9
24	4.12	3.01	80.6	64.9	64.1	47.3
MEAN	3.91	2.81	81.3	65.7	65.8	49.1

Table 2: Relative confidence intervals and percent limits

HOUR	NP	EO-RCI	RE-RCI	MR-RCI	ε (%)	EO-PL (%)	RE-PL1 (%)	RE-PL2 (%)	MR-PL (%)
1	448	0.41	0.39	0.37	±1.20	±1.60	-3.14	2.65	±3.10
2	371	0.36	0.35	0.46	±2.30	±1.59	-4.45	3.53	±3.01
3	332	0.31	0.32	0.45	±2.80	±1.58	-5.16	3.92	±3.08
4	313	0.30	0.31	0.46	±3.00	±1.58	-5.41	4.05	±3.15
5	295	0.29	0.29	0.44	±3.00	±1.58	-5.52	3.98	±3.17
6	297	0.30	0.30	0.47	±3.00	±1.58	-5.57	3.94	±3.24
7	294	0.30	0.30	0.46	±3.00	±1.58	-5.63	3.90	±3.33
8	294	0.29	0.29	0.44	±3.00	±1.59	-5.67	3.87	±3.48
9	288	0.29	0.29	0.47	±3.00	±1.59	-5.69	3.85	±3.61
10	281	0.29	0.29	0.44	±3.00	±1.59	-5.67	3.84	±3.76
11	283	0.29	0.29	0.45	±3.00	±1.59	-5.66	3.83	±3.89
12	291	0.30	0.31	0.45	±3.00	±1.59	-5.65	3.82	±4.02
13	281	0.30	0.32	0.45	±3.00	±1.59	-5.65	3.81	±4.16
14	274	0.28	0.29	0.42	±3.00	±1.59	-5.66	3.81	±4.30
15	282	0.30	0.30	0.44	±3.00	±1.58	-5.66	3.80	±4.45
16	284	0.30	0.30	0.42	±3.00	±1.58	-5.66	3.80	±4.56
17	287	0.30	0.31	0.44	±3.00	±1.57	-5.66	3.79	±4.66
18	282	0.30	0.31	0.43	±3.00	±1.57	-5.66	3.79	±4.77
19	285	0.28	0.29	0.40	±3.00	±1.56	-5.66	3.80	±4.88
20	286	0.30	0.31	0.41	±3.00	±1.56	-5.66	3.81	±4.96
21	290	0.31	0.30	0.44	±3.00	±1.55	-5.64	3.81	±5.06
22	273	0.28	0.28	0.38	±3.00	±1.54	-5.63	3.81	±5.14
23	293	0.30	0.30	0.41	±3.00	±1.54	-5.62	3.81	±5.24
24	278	0.29	0.29	0.40	±3.00	±1.55	-5.61	3.81	±5.31
MEAN	299.	0.30	0.31	0.43	±2.89	±1.58	-5.46	3.78	±4.10

The relative confidence intervals, Eq. (9), and the percent confidence limits for the validation data are shown in Table 2. The following indices are considered:

NP → number of forecasts used to compute the mean RCI;
EO-RCI → 1-24 steps ahead mean RCIs of the EO technique;
RE-RCI → 1-24 steps ahead mean RCIs of the RE technique;
MR-RCI→ 1-24 steps ahead mean RCIs of the MR technique;
EO-PL → 1-24 steps ahead mean EO percent limits;
RE-PL1 → 1-24 steps ahead RE mean lower percent limits;

RE-PL2 → 1-24 steps ahead RE mean upper percent limit; and

RM-PL → 1-24 steps ahead RM mean percent limits.

It can be seen from Tables 1 and 2 that the MR technique produces very low validation confidence degrees, and higher relative confidence intervals than the others. Again, the EO and RE results are very similar to each other.

The following simulation has been carried out by using actual forecasting errors to estimate the reference confidence degrees and to build the resampling set for the RE method. The one-step ahead NN training and test errors, and the validation MAPE and MAE have been the same as in the previous simulation. Since the actual forecasting errors are needed, validation indices could only be computed after 1 week of on-line simulation. Notice that for the same ε's defined in Table 2, the corresponding CDs have gotten to values lower than 80% as a result of using actual forecasting errors instead of test errors. That has made EO-CDs and RE-CDs much closer to CDs. MR-CD resulted in lower values, as in the previous simulations.

Table 3 - Reference and validation confidence degrees using actual errors

HOUR	CD (%)	EO-CD (%)	RE-CD (%)	MR-CD (%)
1	73.8	73.0	75.1	67.6
2	70.9	71.0	68.8	47.4
3	68.9	67.8	67.6	41.2
4	69.8	68.5	69.0	37.6
5	66.1	64.5	64.9	35.6
6	63.2	61.8	60.5	33.7
7	64.0	62.7	61.5	32.4
8	64.3	62.9	63.0	34.1
9	65.4	63.7	64.2	34.1
10	64.0	62.0	63.4	33.1
11	62.8	61.6	61.8	30.4
12	61.8	60.6	62.0	31.2
13	62.7	61.4	61.1	29.9
14	63.2	62.0	62.2	32.2
15	62.3	60.6	62.8	30.2
16	63.1	61.4	63.2	32.6
17	64.4	63.7	64.9	33.7
18	63.6	62.5	64.0	33.3
19	63.0	61.0	63.6	33.5
20	64.3	62.0	64.2	34.7
21	63.8	62.5	64.9	34.1
22	63.1	62.2	63.6	34.9
23	62.4	61.2	63.2	35.1
24	63.8	62.5	64.9	34.3
MEAN	64.8	63.5	64.3	35.7

Table 4 presents the relative confidence intervals and the mean percent confidence limits for the three methods. Again, MR-RCIs resulted in greater values than EO-RCIs and RE-RCIs. It means that the relative MR confidence intervals are narrower than the EO and RE ones, although their confidence is very low. When actual forecasting errors are used for the EO and RE techniques, their validation confidence degrees become more reliable than when test errors are used. This shows the difficulty to produce good generalization for the confidence interval estimation. The relative confidence intervals for the EO and RE techniques are comparable.

In terms of computational costs for each adaptive training, if previous weights can be used as initialization for the NN training, 7 NNs with load and error output neurons (used in the EO technique) take 24sec in a Pentium 233 MHz PC. On the other hand, 7 NNs with one load output neuron (used in the RE and MR techniques) take 18sec to train. However, if random weight initialization is used for the NN training, 7 NNs with load and error output neurons take 9min20sec, while 7 NNs with one load output neuron take 5min50sec. The one week resampling process takes 10sec for each on-line forecasting (1-24 steps ahead). Therefore, the computational costs are comparable when previous connection weights are used for retraining. In the case of random initialization, the computational cost for the EO technique is almost twice the others. The estimation time for A and s, Eqs. (5) and (7), in the MR technique is negligible.

Table 4 - Relative confidence intervals and percent limits using actual errors

HOUR	NP	EO-RCI	RE-RCI	MR-RCI	ε (%)	EO-PL (%)	RE-PL1 (%)	RE-PL2 (%)	MR-PL (%)
1	310	0.381	0.365	0.430	±1 20	±1 52	-2.89	2.59	±2 35
2	222	0.287	0.277	0.482	±2.30	±1.52	-4.09	3.65	±2.22
3	197	0.242	0.252	0.472	±2.80	±1.52	-4.67	3.90	±2.20
4	179	0.245	0.235	0.495	±3.00	±1.52	-4.92	4.56	±2.31
5	168	0.221	0.230	0.482	±3.00	±1.51	-4.79	4.16	±2.22
6	157	0.218	0.229	0.515	±3.00	±1.51	-4.73	4.04	±2.19
7	151	0.213	0.216	0.479	±3.00	±1.52	-4.75	4.12	±2.32
8	159	0.209	0.216	0.462	±3.00	±1.52	-4.71	4.18	±2.47
9	159	0.237	0.235	0.524	±3.00	±1.52	-4.81	4.47	±2.62
10	153	0.220	0.217	0.493	±3.00	±1.52	-4.63	4.45	±2.72
11	139	0.202	0.196	0.451	±3.00	±1.52	-4.57	4.32	±2.75
12	142	0.228	0.236	0.498	±3.00	±1.53	-4.35	4.28	±2.81
13	136	0.220	0.217	0.495	±3.00	±1.53	-4.65	4.29	±2.97
14	144	0.216	0.213	0.482	±3.00	±1.52	-4.66	4.26	±3.09
15	136	0.211	0.212	0.465	±3.00	±1.52	-4.54	4.11	±3.17
16	149	0.221	0.221	0.472	±3.00	±1.52	-4.76	4.00	±3.32
17	152	0.220	0.210	0.467	±3.00	±1.52	-4.75	4.34	±3.47
18	152	0.236	0.235	0.480	±3.00	±1.51	-4.58	4.09	±3.54
19	152	0.215	0.220	0.441	±3.00	±1.50	-4.48	4.07	±3.59
20	157	0.242	0.236	0.469	±3.00	±1.50	-4.65	4.11	±3.77
21	155	0.235	0.224	0.460	±3.00	±1.50	-4.59	4.29	±3.79
22	161	0.229	0.233	0.428	±3.00	±1.49	-4.67	4.05	±3.74
23	160	0.238	0.233	0.467	±3.00	±1.49	-4.41	4.08	±3.89
24	156	0.215	0.209	0.413	±3.00	±1.49	-4.66	4.33	±4.05
MEAN	164.	0.23	0.23	0.47	±2.89	±1.51	-4.55	4.11	±2.98

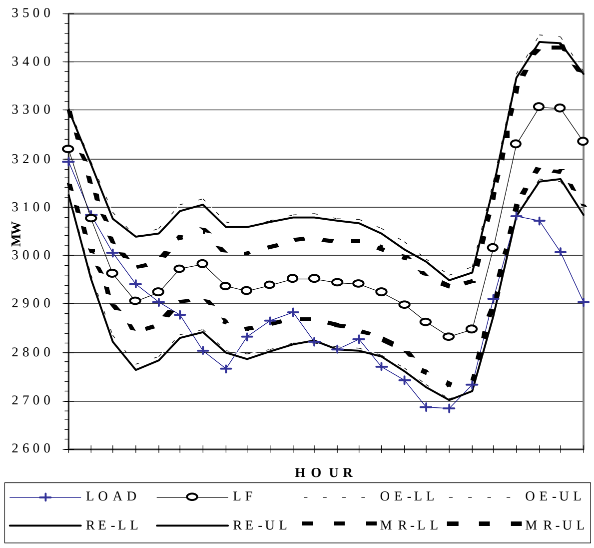

Fig. 3: 1-24 hours load forecasts, Feb.11.1995

In Figure 3, a forecasting example is shown for Feb. 11th, 1995, Saturday, with 4,45% MAPE, up to 24 steps ahead. Table 5 shows the corresponding RCIs for each technique, whenever the forecast lies inside the confidence interval.

Table 5 - Example widths

HOUR	EO-RCI	RE-RCI	MR-RCI
1	0.265	0.236	0.264
2	0.051	0.052	0.073
3	0.319	0.367	0.520
4	0.244	0.263	0.397
5	0.153	0.137	0.248
6	0.670	0.537	-
7	-	-	-
8	-	-	-
9	0.785	0.609	-
10	0.615	0.472	0.900
11	0.577	0.439	0.840
12	-	0.841	-
13	-	0.887	-
14	0.951	0.738	-
15	-	0.999	-
16	-	-	-
17	-	-	-
18	-	0.999	-
19	0.803	0.775	-
20	0.690	0.664	-
21	0.979	0.875	-
22	-	-	-
23	-	-	-
24	-	-	-

Table 6 shows the cross correlation between the validation MAPE and each of the mean percent limits. High positive correlations indicate that the confidence intervals predict well the ups and downs of the forecasting errors. In Table 6, the following indices are considered:

EO-CL → correlation between MAPE and |EO-PL|;
RE-CL1 → correlation between MAPE and |RE-PL1|;
RE-CL2 → correlation between MAPE and |RE-PL2|;
MR-CL → correlation between MAPE and |MR-PL|.

Case 1 refers to the values from Tables 1 and 2, and Case 2 refers to the values from Tables 3 and 4.

Table 6: Cross-correlation indices

	EO-CL	RE-CL1	RE-CL2	MR-CL
Case 1	-0.356	0.994	0.855	0.498
Case 2	-0.182	0.867	0.885	0.428

5. Conclusions

In this paper, it is shown that the performance of the RE method strongly relies on the similarity between the resampling data and the current data. In the EO and MR methods, direct influence of the current data is guaranteed by the NN inputs.

The MR technique assumes that the forecasting errors follow a normal distribution, which is not always true. Even if the NN inputs follow normal distributions, after passing through non-linear functions in the hidden layer the assumption does not hold anymore. The MR technique has very low validation confidence degrees, which indicates that its normality assumption is too costly.

In the EO technique, the CI computation is inherent to the NN training, which demands few additional computations, besides the own load forecast. On the other hand, its NN architecture is more complex, demanding more training time.

The EO and RE techniques present some similar indices, but the EO cannot cope with the 1-24 steps ahead forecasting; it should predict the increasing errors of the recursive load forecasting, which is not confirmed by the correlation indices. For that reason, RE confidence intervals are more reliable.

Recursive forecasting could be avoided by using NNs with 24 outputs, one for each forecast hour. This was not tried due to the difficulty presented by this kind of architecture to adapt to changes in load dynamics and to the high computational costs.

Larger NN structures have been tried, which have not improved the errors of the NN error output. In fact, choosing the NN architecture is an inherent difficulty when applying the EO technique. Non-parametric NN models should be investigated.

6. Acknowledgments

This work was supported by CNPq under grant No. 300054/91-2. The authors also thank the financial support from FINEP/RECOPE (project 0626/96-SAGE).

7. References

[1] A. Khotanzad, R. Hwang, A. Abaye, D. Maratukulam. "An Adaptive Modular Artificial Neural Hourly Load Forecaster and its Implementation at Electric Utilities". IEEE Trans. on Power Systems, Vol. 10, No. 3, pp. 1716-1722, August 1995.

[2] H. Mori. "State-of-the-Art Overview on Artificial Neural Networks in Power Systems". A Tutorial Course on Artificial Neural Networks with Applications to Power Systems. IEEE Power Engineering Society, Edited by M. El-Sharkawi and D. Niebur, Chapter 6, pp. 51-70. 1996.

[3] K. Liu, S. Subbarayan, R.R. Shoults, M.T. Manry, C. Kwan, F.L. Lewis, J. Naccarino. "Comparison of Very Short-Term Load Forecasting Techniques". IEEE Trans. on Power Systems, Vol. 11, No. 2, pp. 877-882, May 1996.

[4] A.G. Bakirtzis, V. Petridis, S.J. Klartzis, M.C. Alexiadis, A.H. Maissis. "A Neural Network Short-term Load Forecasting Model for the Greek Power System". IEEE Trans. on Power Systems, Vol. 11, No. 2, pp. 858-863, May 1996.

[5] O. Mohammed, D. Park, R. Merchant, T. Dinh, C. Tong, A. Azeem, J. Farah, C. Drake. "Practical Experiences with an Adaptive Neural Network Short-Term Load Forecasting System". IEEE Trans. on Power Systems, Vol. 10, No. 1, pp. 254-265, February 1995.

[6] G.E. Box, G.M. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, 1976.

[7] T.W. Anderson. *The Statistical Analysis of Time Series*. John Wiley & Sons, 1970.

[8] D. Husmeier, J. G. Taylor, "Predicting Conditional Probability Densities of Stationary Stochastic Time Series". Neural Networks, Elsevier Science Ltd., Vol. 10, No. 3, pp. 479-497, 1997.

[9] A. S. Weigend, D. A. Nix, "Predictions with Confidence Intervals (Local Error Bars)". Proceedings ICONIP 94, M-W, Kim & S-Y, Lee, Korea Advanced Institute of Technology, pp. 847-852, 1994.

[10] T. Masters. *Neural, Novel and Hybrid Algorithms for Time Series Prediction*. John Wiley & Sons, 1995.