Short Term Load Forecasting in Greek Intercontinental Power System using ANNs: a Study for Input Variables

G.J. TSEKOURAS[1], F.D. KANELLOS[2], V.T. KONTARGYRI[2], C.D. TSIREKIS[2],
I.S. KARANASIOU[2], CH. N. ELIAS[2], A.D. SALIS[2], P.A. KONTAXIS[2], N. E. MASTORAKIS[1]
[1]Department of Electrical & Computer Science, Hellenic Naval Academy, Terma Hatzikyriakou, Piraeus
[2]School of Electrical and Computer Engineering, National Technical University of Athens, 9 Heroon Polytechniou Street, Zografou, Athens
GREECE
Email: tsekouras_george_j@yahoo.gr, kanellos@mail.ntua.gr, vkont@central.ntua.gr, ktsirekis@desmie.gr, ikaran@esd.ece.ntua.gr, xelias@hlk.forthnet.gr, anastasios.salis@gmail.com, pkont@mail.ntua.gr, mastor@wseas.org

Abstract: - The scopus of this paper is to compare the performance of different structures of Artificial Neural Networks (ANNs) regarding the input variables used for short-term forecasting of the next day load in intercontinental Greek power system. The input variables can be: (a) historical loads, (b) weather related temperatures, (c) hour and day indicators, in two ways: (i) selfsame, (ii) compressed using the Principal Components Analysis (PCA). The training algorithm is the scaled conjugate gradient algorithm, for which a calibration process is conducted regarding the crucial parameters values, such as the number of neurons, the kind of activation functions, etc. The performance of each structure is evaluated by the Mean Absolute Percentage Error (MAPE) between the experimental and estimated values of the hourly load demand of the next day for the evaluation set in order to specify the optimal ANN. Finally the load demand for the next day of the test set (with the historical data of the current year) is estimated using the best ANN structure, so that the verification of behaviour of ANN load prediction techniques was demonstrated.

Key-Words: - artificial neural networks, short-term load forecasting, ANN training scaled conjugate gradient algorithm, input variables

1 Introduction

In a liberated electrical energy market, short-term load forecasting (for the next few hours to a week) is very crucial problem, because its accuracy affects other operational issues of power systems, such as unit commitment [1], scheduling of spinning reserve [2], available transfer capability [3], system stability [3], application of load demand programs [4], etc. Higher reliability and lower operational costs for power systems is achieved by precise load forecasting. During last decade several forecasting methods have been implemented with different levels of success, such as ARMAX models [5], regression [6], ANNs [7], fuzzy logic [8], expert systems, etc. Especially, in Greece, ANNs have been used successfully either for the intercontinental power system [9-12], or autonomous big islands [9, 13-14]. The proposed ANNs techniques belong to either classical [10-12] or specialized ones [13] or they are based on ANNs combined with fuzzy logic algorithms [14].

In this paper the effects of the proper selection of the input variables for the ANN is examined. The basic structure of the ANN proposed by Kiartzis et al. [9-10] for the inputs and outputs neurons for the Greek intercontinental power system is used, while the training algorithm is the scaled conjugate gradient [16] according to the respective results by [12]. The input variables can be: (a) historical hourly loads of the last one, two or three days, (b) weather related temperatures of the respective days in direct form (temperatures recorded every three hours) and in indirect form (maximum value, minimum value, temperature dispersion from comfortable living conditions temperature, etc), (c) hour and day indicators, given by sinusoidal or binary form. The input variables can form the respective input vectors either directly, or through compression techniques, such as Principal Components Analysis (PCA) [17-18]. The main goals are:

- the modulation of the internal neural network structure (number of neurons of hidden layer, kind of activation functions, etc.) for each different case of input variables with respect to the smallest *Mean Absolute Percentage Error* (MAPE) of the evaluation set,
- the comparison of the respective cases in terms of MAPE and computational time and
- the suggestion of basic directions for the selection of the input variables for this case study.

The respective results are based on actual hour load data of the Greek intercontinental power system for years 1997-2000.

2 Proposed ANN Methodology for Short-term Load Forecasting

The short-term load forecasting is achieved by applying an ANN methodology through the proper selection of the parameters for scaled conjugate gradient algorithm. This methodology has the following basic steps and its flow chart is shown in Figure 1.

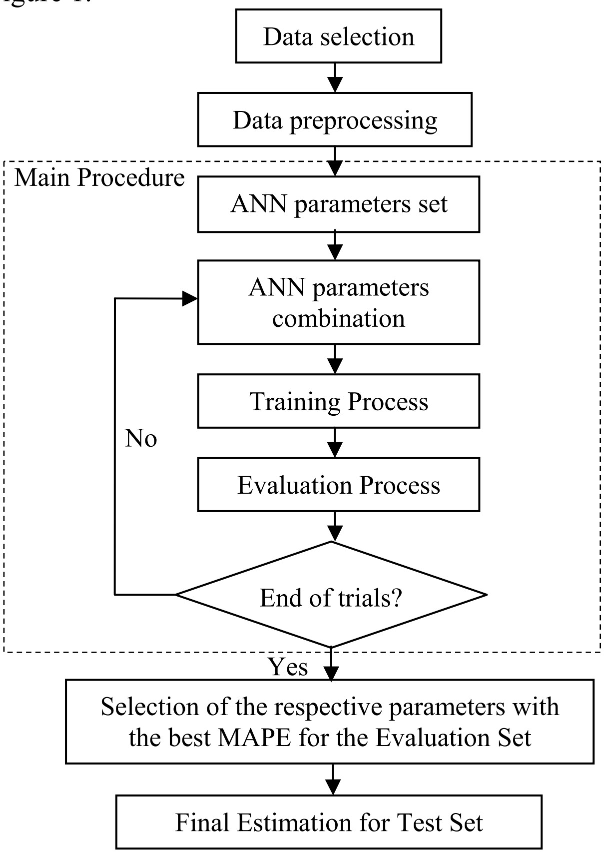

Fig. 1. Flowchart of the ANN methodology for the proper selection of ANN parameters for short-term load forecasting for different kind of input variables

(a) *Data selection*: In this step the input variables for load forecasting of *d*-th day are formed. The basic scenario (1st scenario) is the following according to Kiartzis et al [10] and Tsekouras et al [11]:

- the hourly actual loads of the two previous days: *L*(d-1,1),…, *L*(d-1,24), *L*(d-2,1),…, *L*(d-2,24) (in MW),
- the maximum mean temperature per three hours and the minimum mean temperature per three hours for Athens for the current and the previous day max_*temp*$_{\text{Ath}}$(d), min_*temp*$_{\text{Ath}}$(d), max_*temp*$_{\text{Ath}}$(d-1), min_*temp*$_{\text{Ath}}$(d-1) respectively (^{o}C),
- the maximum mean temperature per three hours and the minimum mean temperature per three hours for Thessalonica for the current and the previous day max_*temp*$_{\text{Th}}$(d), min_*temp*$_{\text{Th}}$(d), max_*temp*$_{\text{Th}}$(d-1), min_*temp*$_{\text{Th}}$(d-1) respectively (^{o}C),
- the temperature difference between the maximum mean temperature per three hours of the current day and the respective one of the previous day for Athens dif_*temp*$_{\text{Ath}}$ and Thessalonica dif_*temp*$_{\text{Th}}$ respectively:

dif_*temp*$_{\text{Ath}}$= max_*temp*$_{\text{Ath}}$(d)- max_*temp*$_{\text{Ath}}$(d-1) (1)

dif_*temp*$_{\text{Th}}$= max_*temp*$_{\text{Th}}$(d)- max_*temp*$_{\text{Th}}$(d-1) (2)

- the temperature dispersion from comfortable living conditions temperature for Athens, for the current and the previous day $T^2_{Ath}(d)$, $T^2_{Ath}(d-1)$, respectively, where:

$$T^2_{dispersion} = \begin{cases} \left(T_c - T\right)^2, & T < T_c \\ 0, & T_c < T < T_h \\ \left(T - T_h\right)^2, & T_h < T \end{cases} \quad (3)$$

where T_c=18 ^{o}C, T_h=25 ^{o}C.

- the temperature dispersion from comfortable living conditions temperature for Thessalonica for the current and the previous day $T^2_{Th}(d)$, $T^2_{Th}(d-1)$, respectively,
- seven digit binary numbers, which express the kind of the week day, e.g. Monday corresponds to 1000000, Tuesday to 0100000, etc,
- two sinusoidal functions ($\cos(2\pi d/T)$, $\sin(2\pi d/T)$), which express the seasonal behavior of the current day, where *T* is the number of the days of the current year.

So, each input vector comprises 71 elements.

Other scenarios are the following:

o 2nd scenario: It is the same with the 1st scenario except of using two sinusoidal functions ($\cos(2\pi d/7)$, $\sin(2\pi d/7)$) instead of the seven digit binary numbers for the kind of the week day. In this case the input vector comprises 66 elements.

o 3rd scenario: It is the same with the 2nd scenario with the difference of using three hours temperatures of the day under prediction and the previous one for Athens and Thessalonica instead of representative temperature functions. In this case the input vector comprises 84 elements.

o 4th scenario: It is the same with the 1st scenario using the hourly actual loads only of the last previous day. In this case the input vector comprises 47 elements.

o 5th scenario: It is the same with the 1st scenario using the hourly actual loads of the three previous days. In this case the input vector comprises 95 elements.

o6th scenario: It is the same with the 2nd scenario, but Principal Components Analysis is additionally used for the total number of the input variables decreasing them from 66 to 6 for 99% description percentage of data dispersion according to Kaiser's criterion [17].

o7th scenario: It is the same with the 2nd scenario, but Principal Components Analysis is additionally used for the total number of the input variables except the variables of the kind of the week day and of the season of the year decreasing their number from 62 to 15 for 99% description percentage of data dispersion according to Kaiser's criterion [17]. The total number of the input variables (with the respective ones of the kind of the week day and of the season of the year) is 19.

In all cases the output variables are the 24 hourly actual load demand of the current day $\hat{L}$ (d,1),…, $\hat{L}$ (d,24).

(b) *Data preprocessing*: Data are examined, in order to modify or delete the values that are obviously wrong (noise suppression). Due to the great non linearity of the problem, non linear activation functions are preferably used. In that case, saturation problems may occur. These problems can be attributed to the use of sigmoid activation functions that present non-linear behavior outside the region [-1, 1]. In order to avoid saturation problems, the input and the output values are normalized by the following expression:

$$\hat{x} = a + \frac{b-a}{x_{\max} - x_{\min}}\left(x - x_{\min}\right) \quad (4)$$

where $\hat{x}$ is the normalized value for variable x, $x_{\min}$ and $x_{\max}$ are the lower and the upper values of variable x, a and b are the respective values of the normalized variable.

(c) *Main procedure*: The ANN is trained using the scaled conjugate gradient algorithm (SCGA), whose basic steps are as follows [16]:

i. The first direction search $\vec{p}_0$ is initialized by the following equation:

$$\vec{p}_0 = -\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_0} \quad (5)$$

where G is the mean error function. The vector of the weights and biases $\vec{w}_0$ is properly chosen. The rest parameters (σ, λ_0, $\bar{\lambda}_0$, *flag*) may have the following values:

$$0 < \sigma \le 10^{-4} \qquad 0 < \lambda_0 \le 10^{-6} \qquad \bar{\lambda}_0 = 0 \qquad flag = 1$$

ii. If *flag* is 1, then the second order information are calculated:

$$\sigma_k = \sigma / \left\|\vec{p}_k\right\| \quad (6)$$

$$\vec{s}_k = \left(\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_k+\sigma_k\cdot\vec{p}_k} - \nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_k}\right)\Big/\sigma_k \quad (7)$$

$$\delta_k = \vec{p}_k^{\,T} \cdot \vec{s}_k \quad (8)$$

iii. The parameter δ_k is scaled:

$$\delta_k = \delta_k + \left(\lambda_k - \bar{\lambda}_k\right)\cdot\left\|\vec{p}_k\right\|^2 \quad (9)$$

iv. If $\delta_k \le 0$, then the Hessian matrix is made positive by:

$$\bar{\lambda}_k = 2\left(\lambda_k - \delta_k / \left\|\vec{p}_k\right\|^2\right) \quad (10)$$

$$\delta_k = -\delta_k + \lambda_k \cdot \left\|\vec{p}_k\right\|^2 \quad (11)$$

$$\lambda_k = \bar{\lambda}_k \quad (12)$$

v. The step size is calculated:

$$\mu_k = -\vec{p}_k^{\,T} \cdot \nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_k} \quad (13)$$

$$a_k = \mu_k / \delta_k \quad (14)$$

vi. The comparison parameter, Δ_k , is calculated:

$$\Delta_k = 2\cdot\delta_k\cdot\left(G(\vec{w})\Big|_{\vec{w}=\vec{w}_k} - G(\vec{w})\Big|_{\vec{w}=\vec{w}_k+a_k\cdot\vec{p}_k}\right)\Big/\mu_k^2 \quad (15)$$

vii. If $\Delta_k \ge 0$, then a successful reduction in error can be achieved:

$$\Delta\vec{w}_k = a_k \cdot \vec{p}_k \quad (16)$$

$$\vec{r}_{k+1} = -\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_{k+1}} \quad (17)$$

$$\bar{\lambda}_k = 0 \quad (18)$$

$$flag = 1 \quad (19)$$

If the increasing number of iterations is multiple of the population N_w of the weights and biases, then the algorithm will be restarted:

$$\vec{p}_{k+1} = -\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_{k+1}} \quad (20)$$

else:

$$\beta_{k+1} = \frac{\left\|\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_{k+1}}\right\|^2 - \nabla G(\vec{w})\Big|^T_{\vec{w}=\vec{w}_{k+1}} \cdot \nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_k}}{\mu_k} \quad (21)$$

$$\vec{p}_{k+1} = -\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_{k+1}} + \beta_{k+1}\cdot\vec{p}_k \quad (22)$$

If $\Delta_k \ge 0.75$, then $\lambda_k = 0.25\cdot\lambda_k$, else $\bar{\lambda}_k = \lambda_k$, $flag = 0$.

viii. If $\Delta_k < 0.25$, then

$$\lambda_k = \lambda_k + \delta_k\left(1 - \Delta_k\right)\Big/\left\|\vec{p}_k\right\|^2 \quad (23)$$

ix. If $\nabla G(\vec{w})\Big|_{\vec{w}=\vec{w}_{k+1}} \ne \vec{0}$, then $k = k+1$ and the step (ii) is repeated, else the training process has been completed.

The basic disadvantage of the SCGA algorithm is the calculation complexity per iteration, which is equal to $O\left(6N_w^2\right)$ instead of $O\left(3N_w^2\right)$ of the basic steepest descent method. Its basic advantage is that the error function decreases monotonically, because an increase in error is not allowed. If the error is constant for one or two iterations, the Hessian matrix has not been positive definite and λ_k has been increased.

The respective parameters of the neural network are selected through a set of trials. Specifically for each ANN parameter (such as the neurons of the hidden layer) the training algorithm is separately executed for the respective range of values (i.e. 20 to 70 neurons with step 1) based on the error function (sum of the square of errors for all neurons per epoch) for training set and the regions with satisfactory results (minimum *MAPE* for evaluation set) are identified. Following, the training algorithm is repeatedly executed, while all parameters are simultaneously adjusted to their respective regions, so that the combination with the smallest *MAPE* for the evaluation set is selected. It is noted that the *MAPE* index between the measured and the estimated values of hourly load demand for the days of the evaluation set is given by:

$$MAPE_{ev} = 100\% \cdot \frac{1}{m_{ev}} \cdot \sum_{d=1}^{m_{ev}} \sum_{i=1}^{24} \frac{\left|\hat{L}(d,i) - L(d,i)\right|}{L(d,i)} \quad (24)$$

where $L(d,i)$ is the measured value of load demand for the i-th hour of d-th day for the evaluation set, $\hat{L}(d,i)$ the respective estimated value, m_{ev} the population of the evaluation set. This index is a practical measure, which reflects the approximation of the actual load demand independently from its units.

(d) *Final estimation for the test set*: The actual load demand (in MW) for the days of the test set is finally estimated by using the respective ANN parameters of the current scenario of input variables.

It is mentioned that, according to Kolmogorov's theorem [19], an ANN can solve a problem by using one hidden layer, comprising the proper number of neurons. Under these circumstances one hidden layer is used, but the number of neurons must be properly selected. The parameters that can be adjusted to optimize the ANN are:

- the number of neurons N_n of the hidden layer,
- the type of the activation functions (hyperbolic tangent, logistic, linear),
- the parameters (a, b) of the activation functions, i.e. $\phi(x) = \tan(a \cdot x + b)$ for hyperbolic tangent,
- the maximum number of epochs (*max_epochs*),
- the parameters σ and λ_0.

It is also noted that during the training process for each ANN three stopping criteria are used [20]:

- weights stabilization (smaller than $\ell imit_1$),
- the respective error function not to be decreased (the variation between two epochs should be smaller than $\ell imit_2$) or
- the excitation of the maximum number of epochs (bigger than *max_epochs*).

Afterwards, the results of each scenario of input variables with the respective optimized parameters are compared, in order to choose the one leading to the smallest *MAPE* index within an acceptable computational time.

3 Analytical Application for Proposed Methodology for the Basic Scenario

Following, the aforementioned method is applied for short-term load forecasting in Greek intercontinental power system. The training and the evaluation sets consist of the 90% and 10% of the normal days (no holidays) of the years 1997-1999 respectively, while the test set consists of the normal days of the year 2000. The input vector $\vec{x}_{in}(n)$ is formed comprising 71 input variables according to the 1st scenario, where the load and the temperature data are normalized. The output vector $\vec{t}(n)$ is formed by the normalized 24 output actual load demand of the day under prediction.

The range of variation in the crucial ANN parameters is mentioned below:

- the number of the neurons of the hidden layer, which ranges from 20 to 70 with incremental step of 1 neuron,
- the parameters σ and λ_0 can be 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6}, 10^{-7}, $5\cdot10^{-8}$ respectively,
- the type and the parameters of the activation functions of the hidden and the output layers. The type can be *hyperbolic tangent*, *linear or logistic*, while the parameters a_1, a_2 get values from the set {0.1,0.2,…, 0.5} and b_1, b_2 from the set {0.0, ±0.1,±0.2}.

The parameters of the stopping criteria are defined after a few trials as max_epochs=5000, $\ell imit_1$=10^{-5}, $\ell imit_2$=10^{-5}.

The development of the abovementioned method in Visual Fortran 6.0 gives the capability to realize all possible combinations of the values of the crucial parameters. In this study the respective combinations account to 2,581,875, which practically can not be examined and this is the reason for the development of the proposed calibration process.

Firstly, the number of neurons varies from 20 to 70, while the remaining parameters are assigned with fixed values ($\sigma = 10^{-5}, \lambda_0 = 5\times10^{-8}$, activation functions in both layers: hyperbolic tangent, a_1=a_2=0.25, b_1=b_2=0.0). In Fig. 2 the *MAPE* indexes for the training, the evaluation and the test set are presented. The *MAPE* indexes of the evaluation and the test set keep step with the respective one of the training set, even if the *MAPE* index for test set is bigger than the respective one of the evaluation set and the *MAPE* index of evaluation set is bigger than the respective one of the training set. The *MAPE* index for the evaluation set has small values for all neurons (the smallest is for 43 and 52 neurons), which proves the stability of this training algorithm.

As a second step it is observed that better results are obtained using in both layers the hyperbolic tangent as an activation function with parameters a_1=0.5, a_2=0.25 and $b_1 = b_2$=0.0. The results are quite similar to the following regions for the parameters: 0.1<a_1<0.9, 0.1<a_2<0.5 και b_1≈b_2≈0. It is mentioned that in Table 1 the results for different activation functions are registered.

The selection of the parameters σ and λ_0 has not any significant participation in the final MAPE results.

The final calibration of the ANN model is realized for 43 to 53 neurons, $\sigma = 10^{-5}$, $\lambda_0 = 5\times10^{-8}$, and activation functions in both layers: hyperbolic tangent with parameters a_1=0.45-0.50-0.55, a_2=0.20-0.25-0.30, $b_1 = b_2$=0.

Finally, the best result for the *MAPE* index of the evaluation set is 1.487% and it is obtained for an ANN with 52 neurons in the hidden layer, a_1=0.5, a_2=0.25 and $b_1 = b_2$=0 using hyperbolic tangent activation function in both layers.

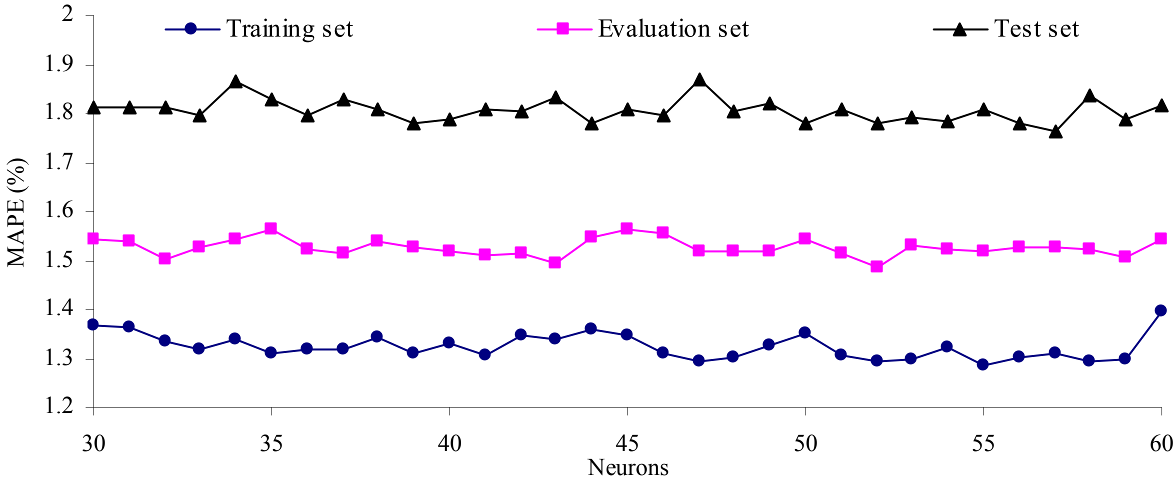

Fig. 2. *MAPE*(%) index for the all sets, 1st scenario, neurons: 20-70, $\sigma = 10^{-5}, \lambda_0 = 5\times10^{-8}$, activation functions in both layers: hyperbolic tangent, a_1=a_2=0.25, b_1=b_2=0.0

TABLE 1
MAPE(%) OF (A) TRAINING SET, (B) EVALUATION SET, (C) TEST SET FOR DIFFERENT ACTIVATION FUNCTIONS FOR NEURONS: 52, $\sigma = 10^{-5}, \lambda_0 = 5\times10^{-8}$, $a_1 = a_2$=0.25, $b_1 = b_2$=0.0

	Activation function of hidden layer								
	Hyperbolic sigmoid			Hyperbolic tangent			Linear		
Activation function of output layer	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
Hyperbolic sigmoid	1.260	1.512	1.821	1.294	1.517	1.814	1.629	1.755	1.978
Hyperbolic tangent	1.401	1.525	1.745	1.324	*1.509*	1.769	1.532	1.742	1.951
Linear	1.465	1.581	1.760	1.344	1.539	1.733	1.688	1.820	1.983

4 Application for Different Scenarios of Input Variables

The proposed methodology of section 2 is applied to six different scenarios of input variables for the short-term load forecasting in Greek intercontinental power system (which have been described in section 2). The training, the evaluation and the test sets are the same with the respective ones of section 3. The ANN's parameters are properly selected and the respective results are presented in Table 2.

The relation of the respective computational time (with the proper parameters calibration) for the seven basic scenarios is: 1÷0.98÷1.1÷0.8÷1.3÷0.4÷ 0.55.

The best results of *MAPE* for evaluation set are obtained with the basic scenario, which have been formed by Kiartzis et al [10] (only the training algorithm is different). The second and the third scenarios lead to similar results (the alternative way of formulation for week day and the direct use of three hours temperatures). It is worth mentioned that the best results of *MAPE* for the test set are given by the third scenario with similar results obtained by the first and the second scenarios. The subtraction or the addition of hourly loads of one day (alternatively the use of hourly loads of one or three previous days) leads to worse results. The use of compression technique decreases the computational time significantly, but it is crucial for the input variables of week and of seasonal behavior not to be suppressed (otherwise the performance of the *MAPE* for the evaluation set is decreased by 20%).

Finally, the first scenario is proposed to be used.

TABLE 2
MAPE(%) OF TRAINING, EVALUATION & TEST SETS FOR 7 DIFFERENT SCENARIOS OF INPUT VARIABLES WITH THE RESPECTIVE PROPERLY CALIBRATED PARAMETERS

No. of scenario	MAPE(%) of training set	MAPE(%) of evaluation set	MAPE(%) of test set	Neurons – Range of examined neurons	Activation functions
1	1.294	*1.487*	1.781	52 (20-70)	f_1=tanh(*0.50*x), f_o=tanh(*0.25*x)
2	1.315	1.516	1.776	48 (20-70)	f_1=tanh(*0.50*x), f_o=tanh(*0.25*x)
3	1.293	1.504	1.717	39 (20-80)	f_1=tanh(*0.50*x), f_o=tanh(*0.25*x)
4	1.574	1.674	2.224	21 (20-70)	f_1=tanh(*0.30*x), f_o=1/(1+exp(*-0.20*x))
5	1.524	1.804	1.808	76 (30-90)	f_1=tanh(*0.25*x), f_o=1/(1+exp(*-0.25*x))
6	1.901	1.960	2.570	21 (15-60)	f_1=tanh(*0.30*x), f_o=1/(1+exp(*-0.30*x))
7	1.458	1.583	1.900	28 (15-60)	f_1=tanh(*0.50*x), f_o=tanh(*0.25*x)

5 Conclusions

This paper compares the performance of seven different ANN structures regarding the input variables for short-term load forecasting in Greek intercontinental power system. The basic structure of the input variables is determined by Kiartzis et al [10]. The other scenarios for input variables involve historical hourly loads of different number of previous days, weather related temperatures of the respective days in direct form (temperatures of every three hours) and in indirect form (maximum value, minimum value, temperature dispersion from comfortable living conditions temperature, etc), hour and day indicators in sinusoidal or binary form. Alternatively Principal Component Analysis (PCA) can be used for data compression. The training algorithm used is the scaled conjugate gradient one according to the results of [11]. The rest parameters, such as the number of neurons of the hidden layer, activation functions, weighting factors, etc. are determined by the proposed calibration methodology of section 2 through an extensive search. The performance of each scenario is measured by the Mean Absolute Percentage Error (*MAPE*) of the evaluation set. Finally, the basic scenario of input variables is proposed, because of the small *MAPE* obtained. If small computational time is required, then PCA should be performed only on load and temperatures variables.

The proposed methodology can be improved (i) by using different kinds of outputs, (ii) by estimating the optimization process and (iii) by determining the confidence intervals of the under prediction chronological load curves.

Acknowledgements

The authors would like to express their sincere gratitude to C. Anastasopoulos, D. Voumboulakis and P. Eustathiou from PPC for the supply of all the necessary data for this application.

References:

[1] B.F. Hobbs, S. Jitprapaikulsarn, D.J. Maratukulam. Analysis of the value for unit

commitment of improved load forecasts. *IEEE Tran. on Power Systems*, Vol. 14, No.4, 1999, pp. 1342-48.

[2] H.B. Gooi, D.P. Mendes, K.R.W. Bell, D.S. Kirschen. Optimal scheduling of spinning reserve. *IEEE Tran. on Power Systems*, Vol. 14, No.4, 1999, pp. 1485-1492.

[3] M.H. Gravener, C. Nwankpa. Available transfer capability and first order sensitivity. *IEEE Tran. on Power Systems*, Vol. 14, No.2, 1999, pp. 512-518.

[4] G. J. Tsekouras, C.A. Anastasopoulos, F.D. Kanellos, V.T. Kontargyri, I.S. Karanasiou, A.D. Salis, N.E. Mastorakis. A Demand Side Management Program of Vanadium Redox Energy Storage System for an Interconnected Power System. *WSEAS International Conference on Energy Planning, Energy Saving, Environmental Education*, (EPESE '08), Corfu Island, Greece, October 26-28, 2008.

[5] H.T. Yang, C.M. Huang, C.L. Huang. Identification of ARMAX model for short term load forecasting: an evolutionary programming approach. *IEEE Trans. on Power Systems,* Vol. 11, No. 1, 1996, pp. 403 – 408.

[6] T. Haida, S. Muto. Regression based peak load forecasting using a transformation technique, *IEEE Trans. on Power Systems,* Vol. 9, No. 4, 1994, pp. 1788 – 1794.

[7] H.S. Hippert, C.E. Pedreira, R.C. Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. on Power Systems,* Vol. 16, No. 1, 2001, pp. 44-55.

[8] P.A.Mastorocostas, J.B.Theocharis, A.G. Bakirtzis. Fuzzy modeling for short-term load forecasting using the orthogonal least squares method. *IEEE Trans. on Power Systems,* Vol. 14, No. 1, 1999, pp. 29–36.

[9] S.J.Kiartzis. *Artificial Intelligence Applications on Short-term Load Forecasting*. Dissertation, Thessaloniki, A.U.Th., 1998 (in greek).

[10] A.G. Bakirtzis, V. Petridis, S.J. Kiartzis, 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, 1996, pp. 858-863.

[11] G. J. Tsekouras, F.D. Kanellos, V.T. Kontargyri, C.D. Tsirekis, I.S. Karanasiou, Ch.N. Elias, A.D. Salis, N.E. Mastorakis. A comparison of Artificial Neural Networks algorithms for short term load forecasting in Greek intercontinental power system. *WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing*, (CSECS '08), Puerto De La Cruz, Canary Islands, Spain, December 15-17, 2008.

[12] S.J. Kiartzis, C.E. Zournas, J.M. Theocharis, A.G. Bakirtzis, V. Petridis. Short term load forecasting in an autonomous power system using artificial neural networks. *IEEE Trans. on Power Systems*, Vol. 12, No.4, 1997, pp. 1591-1596.

[13] K. Kalaitzakis, G.S. Stavrakakis, E.M. Anagnostakis. Short-term load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, Vol. 63, 2002, pp. 185-196.

[14] A.G. Bakirtzis, J.B. Theocharis, S.J.Kiartzis, K.J.Satsios. Short term load forecasting using fuzzy neural networks. *IEEE Trans. on Power Systems*, Vol. 10, No. 3, 1995, pp. 1518-1524.

[15] G.H. Asimakopoulou, V.T. Kontargyri, G. J. Tsekouras, F.H. Asimakopoulou, I.F. Gonos, I.A. Stathopulos. Artificial Neural Network Optimization Methodology for the Estimation of the Critical Flashover Voltage on Insulators. *IET Science, Measurement & Technology*, Vol. 3, Issue 1, January 2009, pp. 90-104.

[16] M.F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, Vol. 6, 1993, pp. 525-533.

[17] L.M. Saini, M.K. Soni. Artificial neural network-based peak load forecasting using conjugate gradient methods. *IEEE Trans. on Power Systems*, Vol. 17, No 3, 2002, pp.907-912.

[18] D. Hand, H. Manilla, P. Smyth. *Principles of data mining*. 1st edition, The M.I.T. Press, Cambridge, Massachusetts, London, England, 2001.

[19] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1994.

[20] G. J. Tsekouras, J. Koukoulis, M.-A. Nikolinakou, N.E. Mastorakis. Prediction of face settlement during tunneling excavation using artificial neural network. *WSEAS International Conference on Engineering Mechanics, Structures, Engineering Geology* (EMESEG '08), Heraklion, Crete Island, Greece, July 22-25, 2008.