Short-Term Electrical Load Forecasting Using a Fuzzy ARTMAP Neural Network

Stefan E. Skarman, Michael Georgiopoulos, and Avelino J. Gonzalez

Department of Electrical and Computer Engineering, College of Engineering,
University of Central Florida, Orlando, FL, 32816

ABSTRACT

Accurate electrical load forecasting is a necessary part of resource management for power generating companies. The better the hourly load forecast, the more closely the power generating assets of the company can be configured to minimize the cost. Automation of this process is a profitable goal and neural networks have shown promising results in achieving this goal. The most often used neural network to solve the electric load forecasting problem is the back-propagation neural network architecture. Although the performance of the back-propagation neural network architecture has been encouraging, it is worth noting that it suffers from the slow convergence problem and the difficulty of interpreting the answers that the architecture provides. A neural network architecture that does not suffer from the above mentioned drawbacks is the Fuzzy ARTMAP neural network, developed by Carpenter, Grossberg, and their colleagues at Boston University. In this work we applied the Fuzzy ARTMAP neural network to the electric load forecasting problem. We performed numerous experiments to forecast the electrical load. The experiments showed that the Fuzzy ARTMAP architecture yields as accurate electrical load forecasts as a back-propagation neural network with training time a small fraction of the training time required by the back-propagation neural network.

1. INTRODUCTION AND BACKGROUND

Short-term load forecasting is an essential function in planning the operation of an electrical power system. The better the forecast, the more closely the power generation assets of the company can be configured to minimize the cost. If the forecast load is higher than the actual load, power generating units will be needlessly activated. If, on the other hand, the forecast is too low, the deficit has to be made up through the operation of "peakers," small standby power units that can be brought online relatively quickly. The drawback with peakers is that they are expensive to run compared to large generating units. However, large generating units take more than a day to go online from a cold state. Therefore, they cannot be used to respond to sudden load changes in real time.

Florida Power Corporation (FPC), the power company whose load forecasting requirements and data were used as the basis for this study, and other power generating utilities generally employ a human to perform the electrical load forecasting function. The forecaster typically predicts the load at least once every work day. Since the forecaster typically does not work during the weekends, the forecast for Saturday, Sunday, and Monday are made on Fridays. Forecasters typically receive a five-day weather forecast every day. A weather forecast becomes more inaccurate the further in advance it is made because of the difficulty in making such predictions. The load forecaster faces the same problem, but the results become even worse because of the inaccuracy of the weather forecasts that have to be used to make the load forecast. The forecaster typically has at his/her disposal several years of weather and load data available in printout form. The old values are used to find days of the same type (e.g., weekdays, weekends, holidays) and with similar weather conditions as the day whose load is to be forecast. The historical load profiles, the forecast weather parameters, and the type of day are then used in the determination of the forecast, based on their relative importance.

The main objective of short-term electrical load forecasting is to predict the hourly loads one day or even up to one week ahead of time. This is necessary for the operational planning of a power system. Since the electrical load demand is a function of weather variables and human social activities, time series models [3,4], regression models [5,6], and expert systems [7] have been proposed to solve this problem. Time series models employ historical load data for extrapolation to obtain future hourly loads. These models assume that the load trend is stationary, and regard any abnormal data as bad data. Regression models analyze the relationship between load, weather, and social activities. The disadvantage with regression models is that they require complex modeling techniques and heavy computations to produce accurate results. Expert systems use an expert forecaster's experience and heuristics to capture the relationship between social activities, weather, and future loads. However, to capture the expert's experience and knowledge, transforming the knowledge into IF-THEN rules that can be used in a computer program is not an easy task. The problem with the described models is that they lack the desired accuracy and they have difficulty producing accurate forecasts in the case of rapid weather changes [8]. Other reported problems [9] are that a model developed for one utility cannot easily be modified for use at another utility and that installation of the model at a new site is often a complex and time-consuming procedure.

During the last several years, artificial neural networks (ANN's) have been applied to the Short-Term Load Forecasting (STLF) problem with considerable success. A problem reported by Gerber [2] is that a very long training time is required for training of the most commonly used ANN architecture (the Back-Prop. NN). The excessive Back-Prop training time prevented Gerber [2] from doing extensive experimentation with the available data. The power generating companies need a system that can make a forecast up to seven days ahead of time. However, this would require even longer training time than what is needed to train a system to make an hourly forecast 24 hours ahead of time.

The hypothesis of this paper was to apply the Fuzzy ARTMAP neural network to the load forecasting problem. The short training time of the Fuzzy ARTMAP NN allowed us to conduct more experiments in order to identify an "optimum" input parameter set, and led eventually to more accurate forecasts. Another problem with the Back-Prop. NN is the difficulty to interpret its answers. The operation of the Fuzzy ARTMAP architecture is better understood and the answers derived by the architecture can be logically explained. Therefore, it should be possible to determine why certain days are more difficult to forecast than others, and hence what to do to improve the forecast accuracy for those days.

In section 2, the Fuzzy ARTMAP NN algorithm is described. Section 3 describes the experiments and presents the results achieved in this research. Section 4 summarizes and describes the conclusions drawn from this work.

2. FUZZY ARTMAP NEURAL NETWORK

The Fuzzy ARTMAP neural network is an architecture that can learn arbitrary mappings from analog or digital inputs of any dimensionality to analog or digital outputs of any dimensionality. "The architecture applies incremental supervised learning of recognition categories and multidimensional maps in response to arbitrary sequences of analog or binary input vectors, which may represent fuzzy or crisp set of features." [10].

The Fuzzy ARTMAP NN consists of two Fuzzy ART modules designated as Fuzzy ART_a and Fuzzy ART_b, as well as an inter Fuzzy ART module, as shown in Figure 1. Inputs are presented at the Fuzzy ART_a module, while the outputs are presented at the Fuzzy ART_b module. The Inter-Fuzzy ART module includes a MAP field, whose purpose is to determine whether the mapping between the presented input and output is the desired one.

The F_0^a layer receives the input vector a and produces the complement encoded input vector I. The complement encoded vector I is created as:

$$I = (a, a^c) = (a_1, \ldots, a_{Ma}, a_1^c, \ldots, a_{Ma}^c) \tag{1}$$

$$a_i^c = 1 - a_i,\ 1 \leq i \leq M_a, \text{ where } M_a \text{ is the dimensionality of the input vector a} \tag{2}$$

Hence, the F_0^a layer acts as a preprocessor to the Fuzzy ART_a architecture. The vector I is subsequently applied at the F_1^a layer. In most cases the complement encoding is not necessary for the output vector O. F_2^a and F_2^b are the layers where compressed representations of the input patterns and the output patterns are established respectively.

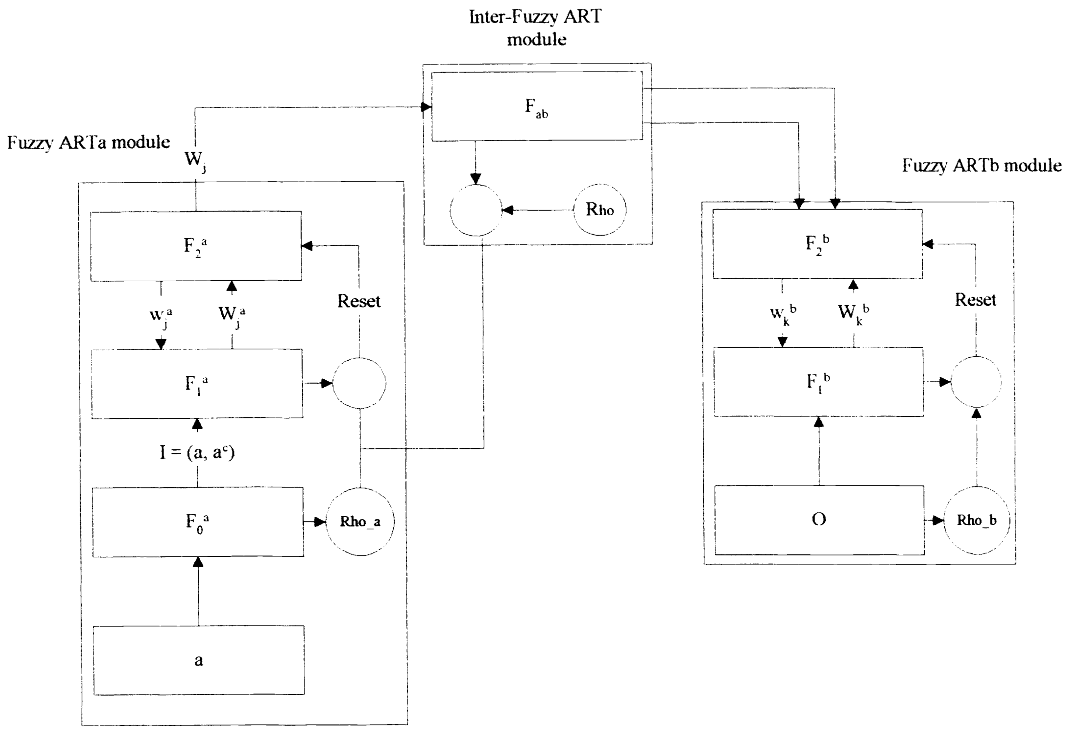

Figure 1 - Block diagram of the Fuzzy ARTMAP architecture.

Index i is used to designate the nodes in the F_1^a layer, index j is used to designate nodes in the F_2^a layer, index k is used to designate nodes in the Fuzzy MAP layer (F_{ab}) and the F_2^b layer, index l is used to designate nodes in the F_1^b layer. Hence $W_j^a = (W_{1j}^a, \ldots, W_{(2Ma)j}^a)$ is the vector of bottom-up weights emanating from node i in the F_1^a layer to node j in the F_2^a layer. The vector of top-down weights emanating from node j to node i is denoted as $w_j^a = (w_{j1}^a, \ldots, w_{j(2Ma)}^a)$. Also, $W_k^b = (W_{1k}^b, \ldots, W_{(Mb)k}^b)$ is the vector of bottom-up weights converging to node k in F_2^b. Furthermore, $w_k^b = (w_{k1}^b, \ldots, w_{k(M^b)}^b)$ is the vector of top-down weights converging to node l from node k in F_2^b. Finally, $w_j = (w_{j1}, \ldots w_{jNb})$ is the vector of weights emanating from node j in F_2^a and converging to the nodes of the MAP layer F_{ab}.

A vector whose components are the top-down weights emanating from a single node in F_2^a/F_2^b is called a template of the Fuzzy $ART_{a/b}$ module.

An uncommitted template is defined to be the vector of top-down weights associated with a node in F_2^a/F_2^b which has not yet been chosen to represent an input/output pattern. Each of the components of an uncommitted template is equal to one.

The vigilance parameters, $\rho_{a/b}$ (designated as Rho_a/b in Figure 1) constraints the maximum size of the compressed representations of the input/output patterns. The range of the vigilance parameter is the interval [0,1]. Small values of the vigilance parameter results in coarse clustering of the patterns, while large values lead to fine clustering. The choice parameters $\beta_{a/b}$ affect the choice of node. The range of the choice parameter is the interval $(0, \infty)$. Realistically, the value should not exceed a value that is proportional to the numbers M_a and M_b respectively.

3. RESULTS

3.1 Network parameter evaluation

The Fuzzy ARTMAP network performance depends on the values of the parameters $\beta_{a/b}$ and $\rho_{a/b}$. In order to identify an "optimal" set of network parameter values for the problem at hand, we conducted a number of experiments.

The network parameter evaluation was performed using a one day ahead, one hour forecast, predicting Wednesdays load at 7:00 am. The training examples used for parameter evaluation consist of all normal Wednesdays (Wednesdays that are not holidays) from 1993 and 1994, up to 10/12/93 and 10/12/94 respectively. Wednesdays after 10/12 are not used for training due to the fact that 1995 data (used for testing) includes data only up to 10/12/95. The test examples used consist of all normal Wednesdays from 1995, up to 10/12/95. The training examples correspond to 94 days and the test examples correspond to 40 days.

A basic training set (the same training set as used by Gerber [2]) was used for the network parameter evaluation. This training set included 27 temperatures and 6 loads. The 27 temperatures were spread over a three day period; the day to forecast, the current day (the day before the day to forecast), and the previous day (the day two days before the day to forecast). For each day, the hour to forecast and the two preceding hours were used as inputs. The temperature data is available for Orlando, St. Petersburg, and Tallahassee. This yields 27 temperature inputs (3 cities * 3 hours* 3 days = 27). The six loads used for the input were spread over a two day period, the current day, and the previous day. For each day the hour to forecast and the two preceding hours were used for the load inputs. This yields six load inputs (2 days*3 hours = 6).

For each experiment, the ρ_a was chosen as 0, 0.3, 0.6, or 0.9, while the ρ_b was chosen as 0.95, 0.99, 0.995, 1.0, and $\beta_a = \beta_b$ used values of 0.01, 1, 5, 10. Table 1 shows the network parameter evaluation results by choosing a specific ρ_a value and obtaining the network performance for all possible combinations of the other network parameters. The MAPE (Mean Absolute Percentage Error) value is calculated as the average APE (Absolute Percentage Error) value for all test Wednesdays, where

$$APE = \frac{|Actual\ load - Predicted\ load|}{Actual\ load} \times 100 \tag{3}$$

The best result achieved was a MAPE = 7.4% for $\rho_a = 0.0$, $\beta_a = \beta_b = 10$, and $\rho_b = 0.95$. These parameter values created a fairly small network with 17 nodes in F_2^a and 7 nodes in F_2^b. The corresponding result obtained by Gerber [2] was a MAPE = 6.8%, using the same training examples and the same test set. However, the training time required to train the Fuzzy ARTMAP neural network was very short (a few minutes) compared to hours of training for the Back-Propagation neural network used by Gerber.

The order of training pattern presentation to Fuzzy ARTMAP affects its performance. Therefore by training 10 networks with the same training set, but with different orders of pattern presentation and taking the predicted load to be the average of the outputs from the 10 networks, we expected a performance improvement.

Table 2 show the results for the ten network approach. The best result was a MAPE = 7.3%. This result is only 0.1% points better than the one network approach. The minor performance improvement, the increased training time (factor 10), and increased weight storage (factor 10) of the ten network approach allows us to conclude that for this experiment the ten network approach is inferior to the one network approach.

Table 1 - MAPE using 1 neural network to forecast the load for Wednesdays at 7:00 am.

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.0	9.3	9.7	8.7
1	7.4	9.5	9.5	8.7
5	7.6	8.2	9.6	8.7
10	7.4	7.9	8.7	8.7

Table 1.1.MAPE. rho_a = 0.0

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.0	9.3	9.7	8.7
1	7.4	9.5	9.5	8.7
5	7.6	8.2	9.6	8.7
10	7.4	7.9	8.7	8.7

Table 1.2. MAPE. rho_a = 0.3

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.0	9.3	9.7	8.7
1	7.4	9.5	9.5	8.7
5	7.6	8.2	9.6	8.7
10	7.4	7.9	8.7	8.7

Table 1.3. MAPE. rho_a = 0.6

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	9.5	8.8	9.6	8.7
1	9.7	8.7	9.5	8.7
5	9.2	7.8	9.4	8.7
10	8.9	7.6	8.7	8.7

Table 1.4. MAPE. rho_a = 0.9

Table 2 - MAPE using 10 neural network to forecast the load for Wednesdays at 7:00 am.

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.1	8.6	9.1	8.7
1	7.7	8.5	9.1	8.7
5	7.5	8.2	9.0	8.7
10	7.3	8.1	8.6	8.7

Table 2.1, MAPE, rho_a = 0.0

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.1	8.6	9.1	8.7
1	7.7	8.5	9.1	8.7
5	7.5	8.2	9.0	8.7
10	7.3	8.1	8.6	8.7

Table 2.2. MAPE. rho_a=0.3

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	8.1	8.6	9.1	8.7
1	7.7	8.5	9.1	8.7
5	7.5	8.2	9.0	8.7
10	7.3	8.1	8.6	8.7

Table 2.3. MAPE. rho_a=0.6

Rho_b	0.95	0.99	0.995	1
Beta				
0.01	10.5	8.2	9.0	8.7
1	10.3	8.1	9.0	8.7
5	10.5	8.1	8.8	8.7
10	11.2	8.1	8.4	8.7

Table 2.4. MAPE. rho_a = 0.9

3.2 24 hour forecast for Wednesdays

The forecast was made one day ahead of time for all 24 hours of Wednesdays in 1995, up to 10/12/95, for the same reasons mentioned in section 3.1. Every hour is predicted with one and ten separate networks, using the parameter values that yielded the best performance during the network parameter evaluation phase ($\rho_a = 0.0$, $\rho_b = 0.95$, $\beta_a = \beta_b = 10$). The basic training set using 27 temperatures and 6 loads was used for training and testing. The network was trained using data from 1993 and 1994 (Wednesdays only, 94 days) and tested using data from 1995 (Wednesdays only, 40 days).

Tables 3 and 4 show the performance when one and ten networks were used to forecast each hour, respectively. MAPE denotes the average performance for all test days. MAX is the maximum error for all test days. MIN is the minimum error for all testing days. For this experiment there is a significant improvement by using 10 networks, the MAPE decreases by 1.2% points, however, the MAPE value is still high.

Table 3 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, 24 hour forecast (Wednesdays) using 1 NN.

MAPE = 7.7%
MAX = 25.4%
MIN = 3.0%
Std dev = 4.5%
MAPE (Days above 10% excluded (7 days)) = 6.0%

Table 4 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, 24 hour forecast (Wednesdays) using 10 NNs.

MAPE = 6.5%
MAX = 25.5%
MIN = 1.9%
Std dev = 4.5%
MAPE (Days above 10% excluded (7 days)) = 5.2%

3.3 Grouping of days

According to Hsu and Yang [11], several days of the week have similar load profiles and can therefore be lumped into the same group. According to Hsu and Yang, Mondays and days after holidays have a similar load profile, Tuesdays through Fridays have a similar load profile, Saturdays and days before holidays have a similar load profile, and finally Sundays and holidays have a similar load profile. One possible method to decrease the relatively high MAPE achieved so far, is to increase the number of training examples. Therefore, the method of grouping days together was tried.

The training/testing set described in previous sections were used for this group of experiments. The network parameter set derived in section 3.1 was used. Two sets of experiments were made:

1. Identical to an experiment evaluated by Gerber [2], where the forecast was made one day ahead of time, predicting the load at 7:00 am.
2. Load forecast one day ahead of time ,predicting all 24 hours, where each hour is predicted with one and ten networks, respectively.

Table 5 shows the performance for the one day ahead forecast, predicting the load at 7:00 am. The MAPE for the forecast of Wednesdays at 7:00 am used in the grouping of Tuesdays trough Fridays is equal to 5.2%, the corresponding MAPE when Wednesdays alone were used for training is 7.4% (Table 1). The grouping of days significantly improves the performance, the MAPE decreased by 2.2% points and the standard deviation decreased by 3.1% points. The corresponding MAPE reported by Gerber [2] is 4.4%, 0.8% points better than the result achieved here. The reason that the Back-Propagation neural network yields better results than the Fuzzy ARTMAP algorithm is most likely due to the fact that the Back-Prop. NN provides as answers to test inputs interpolations of the answers to training inputs. The Fuzzy ARTMAP, after training, provides as answers to test inputs only the answers provided to training inputs in the training phase.

Tables 6 and 7 show the result for the 24 hour forecast. The MAPE for the 24 hour forecast of Wednesdays used in the grouping of Tuesdays trough Fridays is equal to 6.3% (Table 6), the corresponding MAPE when Wednesdays alone were used for training is 7.7% (Table 3). The result for the Wednesdays used in the grouping of days is 1.4% points better than when only those Wednesdays were used for training. The result for the 10 network approach is a MAPE = 5.8%, 0.5% points better than the one network approach. Once again, it is demonstrated that the performance when several networks are used to forecast each hour is better than the performance when a single network is used to forecast each hour.

The grouping of days had a positive effect on the average MAPE for Wednesdays. Since the results for the other groups of days (see [1]) are close to the result for the Tuesdays-Fridays group, one can conclude that an increased number of training examples has a positive effect on the performance of the network.

Table 5 - MAPE, Max MAPE, Min MAPE, and Standard deviation of MAPE for one day ahead forecast (forecast for Tuesdays-Fridays at 7:00am).

MAPE = 5.4%
MAX = 28.8%
MIN = 0.05%
Std dev = 4.9%
MAPE (Days above 10% excluded) = 3.9%
MAPE (Wednesday's only) = 5.2%

Table 6 - MAPE, Max MAPE, Min MAPE, and Standard deviation of MAPE for one day ahead 24 hour forecast using 1NN(forecast for Tuesdays-Fridays).

MAPE = 6.3%
MAX = 20.1%
MIN = 2.9
Std dev = 2.8%
MAPE (Wednesday's only) = 6.3%

Table 7 - MAPE, Max MAPE, Min MAPE, and Standard deviation of MAPE for one day ahead 24 hour forecast using 10 NNs (forecast for Tuesdays-Fridays).

MAPE = 5.8%
MAX = 20.1%
MIN = 2.3
Std dev = 2.8%
MAPE (Wednesday's only) = 6.4%

3.4 Seasonal influence

The load profile for days belonging to a certain group changes as the season changes. To take the seasonal effect into consideration, a 24 hour forecast was produced using networks trained with data from June, July, August of 1993, and 1994 and tested with data from June, July, and August of 1995. A 24 hour forecast was also performed using networks that are trained with data from January, and February of 1993, and 1994 and tested with data from January, and February of 1995. The basic training set used included 27 temperatures and 6 loads. The same network parameter set that yielded the best results during the parameter evaluation section was used. Both training and testing were performed on the Tuesdays to Fridays group (100 training examples and 43 testing examples for the June-August group, 64 training examples and 28 testing examples for the January-February group).

As shown in Tables 8 and 9 (the left columns show the performance with seasonal influence and the right columns shows the performance without seasonal influence). There is a small improvement by using seasonal data for training, since the average MAPE decreased by 0.3% points and 0.1% points respectively. An important advantage is that the maximum error for most days is lower when the seasonal effect is taken into consideration. For certain days the maximum error decreased significantly (e.g. from 21.6% to 8.3% for 6/20/95) for the June-August period. The January-February period shows similar behavior, for certain days the maximum error decreased significantly (e.g. from 28% to 10% for 2/7/95).

Using seasonal data for training has the advantage that for certain days it decreased the maximum APE. This implies that the seasonal training is a worthwhile procedure, despite the fact that the average MAPE does not decrease significantly.

Table 8 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, 24 hour forecast for Tuesdays-Fridays, with and without seasonal influence (June, July, August).

MAPE	6.3	6.6
MAX	19.0	19.2
MIN	2.9	2.9
Std dev	3.1	3.0

Table 9 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, 24 hour forecast for Tuesdays-Fridays, with and without seasonal influence (January, February).

MAPE	7.5	7.6
MAX	19.6	20.1
MIN	3.3	3.3
Std dev	4.1	3.9

3.5 Daily forecast using 1 neural network

In previous sections, one or ten separate neural networks have been used to predict the load for every single hour of the day. In this section, one network is used to predict the load for all 24 hours of the next day. The one day ahead, 24 hour forecast was performed using networks trained with data from June, July, and August of 1993, and 1994 and tested with data from June, July, and August of 1995 (100 training examples and 43 testing examples). Also, a one day ahead, 24 hour forecast using networks trained with data from January, and February of 1993, and 1994, and tested with data from January, and February of 1995 was performed (64 training examples and 28 testing examples). The training/testing uses yesterdays actual temperatures and loads, plus the forecast temperatures for the day to forecast. This result in 168 input parameters (24 temperatures for 3 cities for 2 days = 144 inputs + 24 loads = 168). Both training and testing was performed on the Tuesdays to Fridays group.

The left column of Table 10 shows the performance for the June-August group. The results are reasonable with MAPE = 6.3%. The January-February group yield slightly worse results with MAPE = 7.6% (left column of Table 11). In both cases, the shape of the load curves was correctly predicted for all the test dates. However, both predictions underestimate, in most instances, the actual load. Therefore, an effort to correct the underestimation was undertaken.

The load consumption has increased from year to year over the time period covered in this research. Since the forecast value in almost every test case was found to be lower than the actual value, the simple approach to increase the forecast value by a certain percentage was evaluated in this section. Experiments have showed that by increasing the forecast load by 4% yielded the greatest performance improvement. The average performance improved by 1.2% points for the June-August group (compare the left column of Table 10 to the right column of Table 10) and by 1.1% points for the January-February group (compare the left column of Table 11 to the right column of Table 11). As shown by Table 10 the MAPE decreases but the maximum and the minimum APE value increases for the June-August group. The reason for this is because the forecast load for some days is already higher than the actual load before the increment of the forecast is applied. However, this type of days are uncommon compared to the number of days that have a forecast value that is lower than the actual value. On the other hand, for the January-February group (Table 11) the MAPE, the Max, and the Min decreases when the forecast value is increased. The reason for this is due to the fact that the forecast value is more often lower than the actual value for the January-February group than for the June-August group.

Table 10 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, daily forecast, using 1NN without and with incremented prediction (June-August).

MAPE	6.3	5.1
MAX	14.8	19.3
MIN	1.5	1.8
Std dev	3.4	4.4

Table 11 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for one day ahead, daily forecast, using 1NN without and with incremented prediction (January-February).

MAPE	7.6	6.5
MAX	21.6	20.5
MIN	2.2	1.9
Std dev	5.2	5.2

3.6 Input parameter set evaluation

The input parameter set used in previous sections has been a basic training set including a few temperatures and loads as inputs. The reason the basic input parameter set has been used was to facilitate fair comparisons with the results obtained by Gerber [2]. However, since a different neural network architecture is used in this research, a different input parameter set might improve the performance. The input parameter sets evaluated in this section was:

- Basic: 27 temperatures (temperatures for the hour to forecast plus the two preceding hours for yesterday, today and the day to forecast for Orlando, St. Petersburg, and Tallahassee) and 6 loads (electrical load for the hour to forecast plus the two preceding hours for yesterday and today), are used in this data set.
- 6h: 54 temperatures (temperatures for the hour to forecast plus the five preceding hours for yesterday, today and the day to forecast for Orlando, St. Petersburg, and Tallahassee) and 12 loads (electrical load for the hour to forecast plus the five preceding hours for yesterday and today), are used in this data set.
- 12h: 36 temperatures (temperatures for the hour to forecast plus the eleven preceding hours) and 11 loads (electrical load for the eleven hours preceding the hour to forecast), are used in this data set.
- Average: 9 temperatures (temperatures for the hour to forecast plus the two preceding hours for yesterday, today and the day to forecast; the temperature is set to be the average temperature of Orlando, St. Petersburg, and Tallahassee) and 6 loads (electrical load for the hour to forecast plus the two preceding hours for yesterday and today), are used in this data set.

The training/testing examples used were from the Tuesday-Thursday group (100 training examples and 43 testing examples for the June-August group, and 64 training examples and 28 testing examples for the January-February group).

Table 12 shows the results for $\rho_b = 0.995$. The top half of Table 12 shows the results for the January-February group and the bottom half shows the results for the June-August group. Similarly, Table 13 shows the results for $\rho_b = 0.95$.

The basic input parameter set yields an APE of 5.1% with a standard deviation of 3.4% for $\rho_b = 0.995$ for the June-August group. Both the 6h and the 12h input parameters set yield better results. This makes sense, since more information is used in the input pattern. In particular the 12h input parameters set, yields the most promising results. This also makes sense since data from the hour to forecast plus the eleven preceding hours are used for the input, for both load and temperature. The greatest improvement is noticed for the January-February group where the 12h input parameters set is 4.2% points better than the 6h input parameter set and 4.5% points better than the basic input parameter set (Table 12). The comparison between the 12h input parameter set and the other input sets is not totally fair because during this test actual values were used for both temperatures and loads in the 12h input parameters set. However, during a real execution of the system this would not be possible if the forecast is to be made at least one day ahead of time. Therefore, the 12h input parameters set should be evaluated using forecast values.

The use of average temperature shows a small performance improvement compared to the basic input parameters set for the June-August group. For the January-February group, however, the result for the input parameter set with average temperature is the worst with APE = 13.2% and standard deviation = 10.9%. The reason for this behavior is due to the fact that there are large temperature differences between the three cities during the winter months. On the other hand, during the summer months, the temperatures in the three cities are not significantly different from each other, and hence the average temperature idea works. When the average temperature idea is used to forecast the load for the winter months, a lot of temperature information is lost and therefore, a worse forecast is produced.

Table 12 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for four parameter sets, $\rho_b = 0.995$.

	Basic	6h	12h	Average
MAPE	11.6	11.3	7.1	13.2
MAX	32.7	26.7	15.6	38.6
MIN	0.2	0.2	0.4	0.2
Std dev	9.2	7.7	4.8	10.9

MAPE	5.1	4.6	3.8	4.9
MAX	13.6	15.8	21.8	17.7
MIN	0.0	0.0	0.1	0.4
Std dev	3.4	3.1	3.4	3.4

Table 13 - MAPE, Max MAPE, Min MAPE, and standard deviation of MAPE for four parameter sets, $\rho_b = 0.95$.

	Basic	6h	12h	Average
MAPE	8.9	7.6	6.2	9.3
MAX	21.6	20.0	15.1	25.7
MIN	0.0	0.0	0.0	0.0
Std dev	7.2	6.5	4.1	7.1

MAPE	4.6	4.8	4.1	4.8
MAX	18.8	23.8	23.8	23.8
MIN	0.1	0.1	0.1	0.1
Std dev	3.8	4.5	4.4	4.5

4. REVIEW AND CONCLUSIONS

In an earlier work, Gerber [2] used the Back-Prop. NN to forecast the electrical load for the Florida Power data. Gerber reported excessive training time to train the Back-Prop. NN. Therefore, we applied the Fuzzy ARTMAP NN to the electrical load forecasting problem and conducted identical experiments as the ones carried out by Gerber. The forecasting results achieved with the Fuzzy ARTMAP NN were slightly worse than what Gerber achieved with the Back-Prop. NN. The reason for this is that the Back-Prop. NN has a better interpolating ability than Fuzzy ARTMAP. However, the training time required by the Fuzzy ARTMAP NN is only a small fraction of what the Back-Prop. NN requires to solve the same problem.

Experiments have showed that by using 10 networks to forecast each hour and use the average output from the 10 networks as the forecast value, the network's performance is significantly improved compared to a single network's performance. Especially the fluctuations in the forecast load profile experienced with the one network approach is significantly decreased. The grouping of days had a positive effect on the network's performance. By grouping days with similar load profiles together, and thereby increasing the number of training examples, the performance improved by as much as 2.2 % points. The season had an effect on the extremely high APE values that occurred for certain days. When the season was taken into consideration, the APE value for some hours decreased with as much as 18% points. Since historical data are used for training and the fact that the load demand increases over the covered time period, it is expected that the forecast load should be lower than the actual load. This observation is especially true when one network is used to forecast all 24 hours of the next day. The input parameter set has a significant effect on the network's performance. Experiments showed a 6.1% points difference between the best and the worst input parameter set. More research has to be done to find Fuzzy ARTMAP's optimum input parameter set for the short term electrical load forecasting problem.

ACKNOWLEDGEMENTS

The authors are grateful to the Florida Power Corporation's Energy Control Center, St. Petersburg, Florida for the data provided for this research.

REFERENCES

[1] Stefan E. Skarman, "Short-Term Electrical Load Forecasting Using a Fuzzy ARTMAP Neural Network", Master's thesis, University of Central Florida, Orlando, 1997.

[2] William J. Gerber, "Parametric analysis of electrical load forecasting using artificial neural networks," Master's Thesis, University of Central Florida, Nov. 1996.

[3] W. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Transactions on Power Apparatus and Systems*. vol. PAS-90, April, 1971, pp. 900-910.

[4] D. W. Bunn and E. D. Farmer, *Comparative models for electrical load forecasting*. New York: John Wiley & Son, 1985.

[5] C. Asbury, "Weather load model for electrical demand energy forecasting," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-94, no. 4, 1975, pp. 1111-1116.

[6] A. D. Papalexopoulos and T. C. Hesterberg, "A regression based approach to short-term system load forecasting," *IEEE Transactions on Power Systems*, vol. 5, Nov. 1990, pp. 1535-1544.

[7] S. Rahman and R. Bhatnagar, "An expert system based approach for short-term load forecast," *IEEE Transactions on Power Systems*, vol. 3, No. 2, May 1988, pp. 392-399.

[8] A. Khtoanzad, R-C. Hwang, A. Abaye, and D. Maratukulam, "An adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities," *IEEE-PES Winter Meeting 1995*.

[9] Shin-Tzo Chen, David C. Yu, and A. R. Moghaddamjo, "Wheather sensitive short-term load forecasting using non-fully connected artificial neural networks," *IEEE Transactions on Power Systems*, vol. 7, no. 3, Aug. 1992.

[10] G. A Carpenter, S. Grossberg, N. Markuzon, H. Reynolds, and D. B Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, Sept. 1992, pp. 698-712.

[11] Y-Y. Hsu, and C-C. Yang, "Electrical load forecasting," Alan F. Murray, Applications of Neural Networks, Kluwer Academic Publishers, 1995, pp. 157-189.