

Long Term Electrical Load Forecasting: An Empirical Study across Techniques and Domains

Sujit Kumar Panda*, Sachi Nandan Mohanty and Alok Kumar Jagadev

School of Computer Engineering, KIIT University, Bhubaneswar – 751024, Odisha, India;
mail2sulin@gmail.com, snmohantyfc@kiit.ac.in, alok.jagadevfc@kiit.ac.in

Abstract

Objective: The rapid development of human population, buildings, industries and technology has caused electric consumption to grow rapidly. So it is crucial to have accurate load forecasts for resource planning, rate cases, designing rate structures, financial planning etc. Long-term load forecasting & consumption is used for planning of power system expansion for developing and developed countries or states. In the past years, research mostly focused on short term forecasting instead of than long term. **Methods:** So here more emphasis is given to the methodologies used for long-term forecasting. Parametric methods like trend analysis, end use & econometric technique and computational intelligence based methods like artificial neural network; fuzzy logic model; wavelet networks and genetic algorithm/programming are explained in detail. **Findings:** The results from the comparative study show that the ANN model is a superior technique for long term electric load forecasting due to its ability to give satisfactory results where the availability of past data is low. It also gives lower error than the other models discussed. **Application/Improvement:** The findings will help the researchers to go for hybrid models for improving the error in long term forecasting. It can be applied for finding out the appropriate methodologies of long term forecasting of cities, states, countries or islands with respect to different economic scenarios.

Keywords: ANN, Econometric, Electrical Load, End Use, Fuzzy Logic, Genetic Algorithm, Long Term Forecasting, Trend Analysis, Wavelet

1. Introduction

Economic growth of a country is directly proportional to the electric consumption. In order to supply the energy demand constrained on the worldwide limited energy sources, it is essential to provide an accurate electrical load forecasting.¹ Underestimation or overestimation in the electrical load prediction can introduce various challenges to the system operators. Problems in power quality or reliability and insufficient provided reserves stem from the energy demand underestimations. On the other hand, load forecast overestimation results in unnecessary establishments and spinning reserves, inefficient energy distributions, increasing the operation costs and a huge financial loss. Therefore, precise load forecast is very important.

Electrical load forecasting is of three types Long-term, midterm and short term considering the length of the forecasting time.² If the forecasting is done for the next hour or next week or for next month then it comes under short term forecasting. A lot of work has been done in this area and the most used method is Artificial Neural Networks (ANN).³ Midterm and Long term forecasting usually covers long time horizons. It deals with forecasts for one year, ten years and sometimes up to thirty years. It gives the valuable input for capacity expansion, capital investment, revenue analysis and budgeting. Long term electric load forecasting depends on uncertainties of the future thus making it extremely challenging.⁴

From the literature survey it is quite clear that, the long term load forecasting is not yet explored much. It is also shows that computational intelligence and parametric

*Author for correspondence

based techniques are mostly used. All the methods used in this field such as three parametric viz. trend analysis, end use & econometric technique and four computational intelligence based methods viz. artificial neural networks, fuzzy logic model, wavelet networks and genetic algorithm/programming are discussed in detail in this paper.

2. Long Term Load Forecasting using Trend Analysis

Trend Analysis extends past data of electricity demand to the future, the forecast is carried out by inserting time value into the regression equation. The regression equation is determined from time series data using the least squares method and then project the future for a forecast. Trend analysis is a special case of regression analysis. The general equation for a trend line is given by:

$$F = a + bt + u \quad (1)$$

Where, F – forecasted Value (dependent variable)

t – Time (independent variable)

a – intercept on Y – Axis

b – Slope of the line

u – Error component

Following are the pre-requisite for using the trend analysis for forecasting:

- Data pattern should be close to linear.
- Sufficient correlation should be there between the parameter time and time series data.

The correlation coefficient, R, measures the strength and direction of linear relationships between two variables. A correlation near zero indicates poor linear relationship and a correlation near one indicates a strong linear relationship between the two variables. If the data pattern is non linear then logarithmic, polynomial, power, exponential functions are used to carry out trend analysis.

The trend analysis is generally simple to use. There are standard package such as SPSS, Minitab available which can be used for carrying out trend analysis. Trend analysis are not applicable to long term forecasting like electrical forecasting because of the fact that there are several trends (due to different growth rates) when we take into account a longer period say 30 years.

3. End - Use Technique

End - use modeling approaches are characterized by their primary reliance on physical analysis of the energy

consuming equipment. Models are developed as functions of the capacity and efficiency of equipment, rates of saturation (ownerships – market penetration) of the equipment, frequency of utilization during peak demand and weather. Engineering principles are typically used to develop a simplified mathematical representation which relates energy utilization to the thermal properties of environment and the electro-mechanical characteristics of the energy using equipment.⁵ End - use models focus on various uses of electricity in residential, commercial, agricultural, public lighting, industrial and miscellaneous (research establishment, Port Trusts, electric traction, public water works) sectors.^{6,7} The electricity demand is derived from the customer demand of various categories/sectors listed above. Such models usually begin by specifying reasonably homogeneous uses for which energy is ultimately required, such as heating water, cooling buildings and cooking food. The model then describes, via mathematical equations and accounting identities, the types of energy using equipment that businesses and households have and how much energy is used by each type of equipment to satisfy the predetermined levels of end use energy demanded. By summing up the units of equipment times the average energy used by each class of equipment, total energy demand by fuel type is arrived. This method predicts energy consumption. However, the peak load can also be calculated, if sector wise energy consumption and corresponding load factor is available. The system load factor can be calculated using Equation (2).⁸

$$\text{Load Factor} = \frac{\text{Average load demand}}{\text{peak lead demand}} \dots \dots \dots (2)$$

$$\text{Annual Load Factor}(\%) = \frac{\text{Annual Energy requirement(in MU)} \times 10^5}{(\text{peak lead in MW}) \times 8760 \frac{\text{hours}}{\text{year}}}$$

3.1 End – Use Model for Demand Forecasting and Load Shape Analysis

The load forecast model by end-use method can be used for demand forecasting and load shape analysis. It can be divided into two major components: economic/demographic module and power module.⁹

The economic/demographic module yields consistent forecasts of factors essential for load forecasting such as households, housing (households pertains to the equipments in the house while housing pertains to number of consumers), income, employment and industrial output. The module contains a demographic sector that models

the labour supply and an economic or employment sector that models the labour demand. Each sector attempts to capture the basic elements that shape the future dimensions of a region's population and employment structure. The two sectors are tied together by a dynamic feedback loop, which continuously examines the labour supply and demand balance.

In the power module, the projections of the various economic and demographic parameters are combined with consumption estimates and patterns of electricity usage to produce projections of annual or monthly consumption, peak demand and daily load shape. The basic concept underlying the power module's structure involves breaking down numerous end-use categories for electricity and treating them in approximately the same way.

The power module is further divided into conventional utility customer categories: residential, commercial and industrial. To ensure consistency between annual or monthly energy consumption and hourly demand projections, the power module forecasts hourly demands for typical days in the year, then obtains annual, monthly projections by summing the corresponding hourly demand values. The following basic end - use forecast equation 3 is followed in the article.⁹

$$D_{ij} = N_i * C_i * F_{ij} \dots\dots\dots (3)$$

Where

D_{ij} = Demand at hour j by end - use component i
 N_i = Number of use components of type i
 C_i = Connected load per use component of type i
 F_{ij} = Fraction of the connected load of use component i Which is operating at hour j

The energy consumption over any period is calculated by summing the corresponding demand values.

Thus,

$$E_i = \sum D_{ij} \dots\dots\dots (4)$$

Where

E_i = Energy consumption by end - use component i, for a

Specific period

D_{ij} = Demand at hour j by end - use component i

The equation 4 states that the total demand by a given end-use is a product of total connected load and fraction of these, operating at any given time.

Repeating these calculations for each hour of the day and for all end - uses yields the daily load profile for the system. The given equations are strictly applicable to the

residential sector and require some modification for the industrial and commercial sectors.

The model structure allows the utility planner to examine contribution to system hourly demand of the major end - uses of electric power for extreme and typical days. From its detailed structure, the planner is able to explore various strategies of load management and contingency planning and to examine numerous meaningful "what if" questions⁵. The "what if" analysis are becoming more and more important as utility planners are faced with an increasing number of new factors and quickly changing planning environment.

3.2 Development – Focused End Use (DEFENDUS) Model

The DEFENDUS scenario for energy demand and supply focuses on people based development through the promotion of energy services, identifying technological opportunities for better utilization of energy through a scrutiny of the end uses of energy and adhering to a least cost approach to the mix of energy supplies.¹⁰ The DEFENDUS scenario involves the illumination of all homes in Karnataka, an emphasis on employment generating industry, the energization of all irrigation pump sets up to the limit imposed by the ground water potential and the establishment of decentralized rural energy centers in villages. It also gives energy and power requirements in the year 2000 which are only about 38 per cent and 42 per cent respectively of the Long Range Plan for Power Projects (LRPPP) demand.¹⁰

3.2.1 Methodology for Scenario Construction – DEFENDUS Scenario

The construction of a DEFENDUS scenario for electricity is carried out in following steps:

Step (1) Estimation of 'true' demand in the base year B, takes into account category wise consumption, generation, energy requirement (including proposed power cut) as shown in Figure 1.

Step (2) An estimation of energy requirement in the terminal year B + n (corresponding to the time horizon) through the development of a DEFENDUS scenario for energy demand is shown in Figure 2, it has two components:

- Specific growth rates of the connections in different consumer categories.

- Efficiency improvements and electricity substitution so that the consumption norms for different consumer categories are altered.

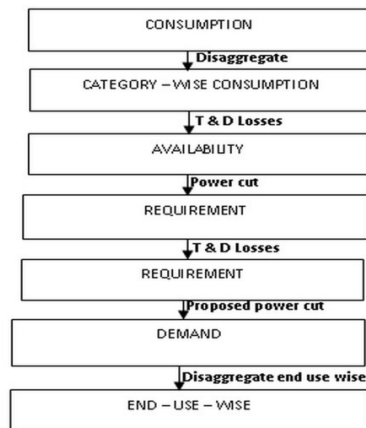


Figure 1: ESTIMATION OF DEMAND IN BASE YEAR B
Source: Amulya Kumar N. Reddy *et. al.*⁶

Figure 1. Estimation of demand in base year b.

Source: Amulya Kumar N. Reddy *et. al.*⁶

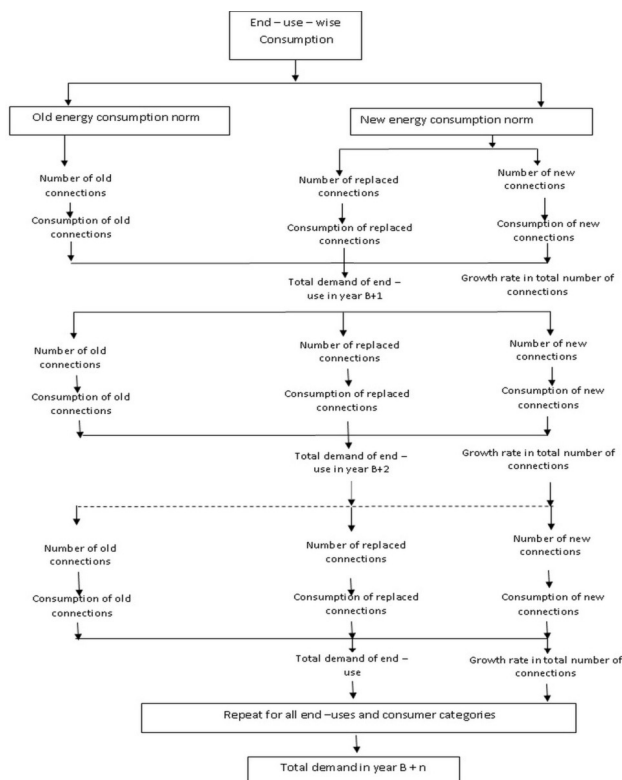


Figure 2: End – Use wise consumption
Source: Amulya Kumar N. Reddy *et. al.* 1991⁶

Figure 2. End – Use Wise Consumption.

Source: Amulya Kumar N. Reddy *et. al.* 1991⁶

Step (3) Development of a ‘Frozen Efficiency’ (FE) scenario for energy demand based on the assumption such that there will be no efficiency improvements during the entire period spanned by the DEFENDUS scenario.

Step (4) Estimation of the potential contribution that technologies of saving or generating energy can make to meet the energy requirement, this involves, estimating of the energy that could be saved in efficiency improvement and electricity substitution and estimating the generated energy for a generation technology.

Step (5) Ranking of all the energy technologies relevant to the region (including the technologies for efficiency improvement and electricity substitution, decentralized generation and centralized generation) according to increasing unit cost of energy (Rs/kWh).

Step (6) Identifying a mix of technologies to meet the energy requirements by starting with a cheapest technology and moving to the next costlier technology when the potential of the previous technology is exhausted.

3.3 End Use Approach followed by Government of India for the 17th Energy Power Survey

The power surveys, demand forecast in our country is made usually under two time frames (i) detailed forecast on a micro basis for a period of 5 – 7 years, (ii) projections of perspective demand over a longer time frame on a macro basis. Generally, the former is used as a basis of extrapolation for the later.⁷ Being the latest EPS, 17th EPS is used for the study.

Historically, the present day methodology of forecast, which has come to be known as “Partial End Use” technique, is an energy based forecast. It was adopted for the first time in the 8th Power Survey and fully developed by the advent of the 10th Power Survey. (Forecast of Peak Demand used to be made directly up to the 7th Power Survey). This methodology is a combination of End Use technique for the sectors, where sufficient data for the past is available and the programme for the future is well defined viz. major industrial and non – industrial loads with a demand of 1 MW and above, programme of Energization of Pump sets, Lift Irrigation Schemes, Major Public Water Works and Track – electrification Programme of Railways; while trend analysis in the case of others (non – industrial bulk consumers like Research Establishments, Port Trusts, Military Engineering Services). The results presented in the Survey reports are

the outcome of micro exercises carried out for each State/UT/System and depend much on the reliability of the data available.

4. Econometric Technique

Econometric methods rely on the statistical historical time series analysis and cross section data to develop a model of peak load or energy demand as a function of behavioral and structural variables such as economic activities, population and weather. Economic theory is usually used to develop a mathematical model and its parameters are estimated by some type of regression analysis.

The method of econometric forecasting is used for long range forecasting of electric energy demand. These models relate the demand for electric energy to economic and demographic variables such as population, income, and electricity rates.¹¹

Econometric load forecasts of demand are based on a model whose parameters are estimated from historical data. A linear or log linear model is usually postulated. In the linear model, the variable to be forecasted is a linear function of independent variable, the input variables. In general, a linear dependence of the endogenous variable upon past (lagged) values of itself and of the exogenous variables may also be accommodated by the linear model. In the log – linear model, a linear relationship is assumed between the logarithm of the endogenous variable and logarithms of the exogenous variables. Log – linear models are the ones most commonly used for modeling the demand for electric energy.

Either the linear or log – linear models may be expressed by the following equation:¹¹

$$y(i) = \sum_{j=1}^k a_j * y(i-j) + \sum_{j=1}^m b_j * u_j(i) + c + v(i) \dots \dots \dots (5)$$

Where;

y (i) is either the annual energy sales or the logarithm of annual energy sales for year i

u_j (i) is the jth exogenous variable, or logarithm of the jth variables for year i.

v (i) is a random deviation for year i

a_j, b_j, c are model coefficients.

The model coefficients are estimated using historical data. They are not exact for two reasons. First, there are only a finite number of data points, and thus because of the random component v (i) in the model, they can never

be exactly estimated. Secondly, the model is not structurally perfect because there are additional exogenous variables which have influence on demand, but are not included in the model.

The construction of econometric models is usually accomplished on a trial and error basis. Exogenous variables, which would logically influence demand, are selected, a linear regression is performed and statistical tests are made. Among the tests ordinarily performed are the F – test on the ratio of the variance explained by the regression to the unexplained variance, the t – test on the model coefficients to determine their significance and the Durbin – Watson test to test for serial correlation of the residual errors.¹¹ Generally, there may be more than one acceptable model. It is entirely possible to develop econometric models, which fit the historical data well but which are very sensitive to errors in the model coefficients and to errors in the forecasted exogenous variables. Such models lead to unacceptably large errors in the forecasted demand. Some of the models developed for the load forecasting are listed in the sequel.

The forecasting of electricity consumption was carried out in Cyprus up to year 2030 considering climate change.¹² The estimation was carried out with single – equation auto-regressive distributed lag models (ARDL). US Energy Information Administration, European Commission or the International Energy Agency do the energy forecasting by taking GDP, population and fuel prices as exogenous and energy use is determined with single – equation models.¹² In regression analysis involving time series data, if the regression model includes not only the current but also the lagged (past) values of explanatory variables then the model is known as distributed lag model. If it has lagged values of the dependent variables among its explanatory variables then it is called an autoregressive model.^{13,14} Thus,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \dots \dots \dots (6)$$

Represents a distributed lag model, whereas

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t \dots \dots \dots (7)$$

Is an example of autoregressive model, where

Y_t	=	The dependent variable at t
X_t	=	Explanatory (independent)
variable at t		
U_t	=	error term.
T is time, α , β & γ are coefficients		

4.1 Classical Econometric Model

It is proposed and used in power utilities for many years. Using logarithmic functions, a model has previously been proposed to represent a relationship between energy and socio-economic variables.¹⁵

$$\log S_t = c \cdot \log S_{t-1} + d \cdot \log \left(\frac{GSP_t}{GSP_{t-1}} \right) + e \cdot \left(\log \left(\frac{GSP_t}{GSP_{t-1}} \right) \right)^2 + f \cdot \log \left(\frac{P_t}{P_{t-1}} \right) + g \cdot \log \left(\frac{POP_t}{POP_{t-1}} \right) \dots \dots (8)$$

Where, S_t = energy consumption in year t

S_{t-1} = energy consumption in year t – 1

GSP_t = GSP in year t

GSP_{t-1} = GSP in year t – 1

P_t = real tariff in year t

P_{t-1} = real tariff in year t – 1

POP_t = population in year t

POP_{t-1} = population in year t – 1

The parameters in the relationship in (15) are found by minimizing the following total error:

$$E = 1/2 \sum_{t=1}^M (y_t - \hat{y}_t)^2 \quad (9)$$

Here M : past years selected for model synthesis, y_t is the model output in year t, which is $\log S_t$, $\hat{y}_t$ is the specified value for year t, which is formed from $\log$ [actual energy in year t], w_t is the weighting coefficient for year t, and c, d, e, f, g are model parameters to be determined. One of the major barriers for use of deterministic model is availability and accuracy of data.

5. Computational Intelligence based Technique

The above discussed techniques have inherent inaccuracies and numerical instability.^{16,17} The non – stationary property of the load prediction process, coupled with complex relationship between weather variables and the electric load render such conventional techniques less effective as these methods assume simple linear relationships during the prediction process. The problems encountered in conventional techniques are overcome with the help of Artificial Intelligence techniques. In this section, four artificial intelligence based techniques viz. Artificial Neural Network, Fuzzy Logic, Genetic Algorithm/Programming and Wavelet network for long term load forecasting are discussed.

5.1 Artificial Neural Network (ANN)

In recent years, much research is conducted on application of computational intelligence techniques to load forecasting problems. ANN has received extensive attention and is used extensively in forecasting.¹⁸ There are several classes of NN, classified according to their learning mechanism. However, broadly they can be identified as three classes of networks. Following steps have been suggested in the presented article for ANN modeling for forecasting:¹⁹

5.1.1 Network Architecture

ANN consists of nodes in one or more layers such as one input layer, one output layer and one or many hidden layers. For implementing one has to decide:

5.1.1.1 Input Nodes

The number of input nodes must be determined as it relates to the number of variables in the input vector used for forecasting. Normally, the number of inputs remains transparent and not so difficult to choose. In case of time series forecasting the number of input nodes relates to the number of lagged observations which is used to discover the underlying pattern for forecasting.

5.1.1.2 Hidden Layers and its Nodes

Hidden layers and its nodes are very important for many applications of NN. The hidden layer nodes allow the networks to detect feature, to perform non linear mapping between output and input variables and to capture the data pattern. If hidden nodes are not present then the result will be equivalent to the output of statistical forecasting models.

5.1.1.3 Output Nodes

It is not so difficult to decide as it can be directly found from the problem under study. In case of time series problems, the output nodes often correspond to the forecasting horizon. There are two types of forecasting: one – step – ahead (which uses one output node) and multi – step – ahead forecasting. There are two ways of making multi – step forecasts the first is called the iterative forecasting as used in the Box – Jenkins model in which the forecast values are iteratively used as inputs for the next forecasts.

5.1.1.4 Interconnection of Nodes

The network architecture also depends on the interconnections of nodes. The connection between nodes in a network fundamentally determines the network's behavior. For most forecasting as well as other applications, the networks are fully connected in that all nodes in one layer are only fully connected to all nodes in the next higher layer except for the output layer.

5.1.2 Activation Function

It is also called transfer function. It also determines the relationship in between inputs and outputs of a network. A degree of non linearity that is valuable for most of the applications is generally caused by the activation function. These include:

1. The sigmoid function:

$$f(x) = \frac{1}{1 + \exp(-x)} \quad \dots \dots \dots (10)$$

2. The tangent (hyperbolic - tanh) function:

$$f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad \dots \dots \dots (11)$$

3. sine or cosine function:

$$f(x) = \sin(x) \text{ or } f(x) = \cos(x) \quad \dots \dots \dots (12)$$

4. Linear function:

$$f(x) = x \quad \dots \dots \dots (13)$$

Logsigmoid is the mostly used.

5.1.3 Training Algorithm

It is an unconstrained nonlinear minimization problem. The weights are iteratively modified to keep the overall mean and/or total squared error within desired and actual output values for all the output nodes over all the input patterns. There is perhaps, no algorithm currently available to guarantee the global optimal solution for a general nonlinear optimization problem in a reasonable amount of time. As such, all optimization algorithms in practice inevitably suffer from the local optima problems; however, optimization method can be used to give the "best" local optima if the true global solution is not available.

The back propagation algorithm is mostly used by all and a steepest gradient descent method. For the gradient descent algorithm, a step size must be specified, which is the learning rate. The learning rate is crucial for back

propagation learning algorithm, since it determines the magnitude and rate of weight change.

The standard back propagation technique with appropriate momentum rate is adopted by most researchers. Generally through experimentation, the best values of the momentum and learning rate are chosen simultaneously. As the learning rate and the momentum can take on any value between 0-1, it is actually impossible to do an exhaustive search to find the best combination of these training parameters. Only selected values are considered by the researchers. In the article, nine combinations are tried with three learning rates (0.1, 0.5, 0.9) and three momentum values (0.1, 0.5, 0.9).²⁰

The training parameters play a critical role in the performance of ANNs.²¹ Using different learning parameters, they re-test the performance of ANNs for many time series which have been previously reported to have worse results with ANNs compared to Box Jenkins methods. They find that for each of these time series there is an ANN with appropriate learning parameters, which performs significantly better.

5.1.4 Data Normalization

Nonlinear activation functions such as the logistic function typically have the squashing role in restricting or squashing the possible output from a node to, typically, (0, 1) or (-1, 1). Data normalization is often performed before the training process begins. As mentioned earlier, when nonlinear transfer functions are used at the output nodes, the desired output values must be transformed to the range of the actual outputs of the network. Even if a linear o/p transfer function is used, it may still be advantageous to standardize the outputs and inputs to avoid computational problems, to meet algorithm requirement and to facilitate network learning.

Four methods for input normalization are summarized in the presented article, what follows:²²

- Along channel normalization: A channel is defined as a set of elements in the same position over all input vectors in the training or test set, i.e. each channel can be thought of as an "independent" input variable. The along channel normalization is performed column by column if the input vectors are put into a matrix. In other words, it normalizes each input variable individually.

- Across channel normalization: This type of normalization is performed for each input vector independently, i.e. normalization is across all the elements in a data pattern.
- Mixed channel normalization: As the name suggests, this method uses some kind of combinations of along and across normalization.
- External normalization: All the training data are normalized into a specific range.

The choice of the above methods usually depends on the composition of the input vector. For a time series forecasting problem, the external normalization is often the only appropriate normalization procedure as the input & output variables are usually formed by considering time lags in the data.

5.1.5 Training and Test Sample

The training samples are used for model development and the model's forecasting ability is evaluated by test samples. For small data sets the same test set can be used for testing and validation. The performance of ANNs may be affected by the way the training and test sets are selected.

It is critical to have both the training and test sets representative of the population or underlying mechanism. Inappropriate separation of the training and test sets will affect the selection of optimal ANN structure and the evaluation of ANN forecasting performance.¹⁹

Another closely related factor is the sample size. No definite rule exists for the requirement of the sample size for a given problem. In general, as in any statistical approach, the sample size is closely related to the required accuracy of the problem. The larger the sample size, the more accurate the results will be.

5.1.6 Performance Measures

Prediction accuracy is the main criteria for measure of performance of ANN forecasters. An accuracy measure is defined by the forecasting error (Actual value – forecasting result). Different methods are available for accuracy measure.¹⁸ The most used measures are:

$$\text{The mean absolute deviation (MAD)} = (\sum |e_t|) / N \quad \dots \dots \dots (14)$$

$$\text{The sum of squared error (SSE)} = \sum |e_t|^2 \quad \dots \dots \dots (15)$$

$$\text{The mean square error (MSE)} = (\sum |e_t|^2) / N \quad \dots \dots \dots (16)$$

$$\text{The root mean square error (RMSE)} = \sqrt{MSE} \quad \dots \dots \dots (17)$$

$$\text{The mean absolute percentage error (MAPE)} = \frac{\sum e_t (100)}{N y_t} \quad \dots \dots \dots (18)$$

Where e_t : individual forecast error

y_t : actual value

N: Number of error terms.

The above measures such as MAD, SSE, MSE and RMSE are standard statistical measures, while MAPE is a relative measure. MSE or SSE has two drawbacks, first it refers to fitting a model to historical data and such fitting does not necessarily imply good forecasting. Secondly, the measure of accuracy is related to different methods use different procedures in fitting phase, eg. decomposition methods include the trend – cycle in the forecasting phase as though it were known, regression methods minimize the MSE by giving equal weights to all observations and Box – Jenkins methods minimize the MSE of a non linear optimization procedure. Even the MAPE itself does not give a good basis of comparison as to the gains in accuracy made by applying a specific forecasting method.²³ Because of the limitations associated with each individual measure, one may use multiple performance measures in a particular problem. However, one method judged to be the best along one dimension is not necessarily the best in terms of other dimensions.

5.1.7 Multi Layered Feed Forward Back Propagation Network

Feed – Forward Back Propagation (FFBP) is one of the most widely used neural network paradigms, which is applied successfully in application studies. FFBP can be applied to any problem that requires pattern mapping.²⁴ Given an input pattern, the network produces an associated output pattern. Its learning and update procedure is intuitively appealing, as it is based on a relatively simple concept.

A schematic diagram of a BP network is shown in Figure 3. It employs three or more layers, an input layer, an output layer and at least one hidden layer. The computational procedure of this network is as below:²⁵

$$Y_j = f(\sum w_{ij} X_i) \quad \dots \dots \dots (19)$$

Where Y_j is the output node j , $f(\cdot)$ the transfer function, w_{ij} the connection weight between node j and node i in the lower layer and X_i the input signal from the node i in the lower layer.

The BP algorithm is a gradient descent algorithm. It tries to improve the performance of the neural network

by reducing the total error by changing the weights along its gradient. The BP algorithm minimizes the square errors, which can be calculated by:

$$E = 1/2 \sum_p \sum_j (O_{pj} - Y_{pj})^2 \quad (20)$$

Where E: square error, p: index of pattern, j: index of output vector, Y is the network output and O: actual (target) output.

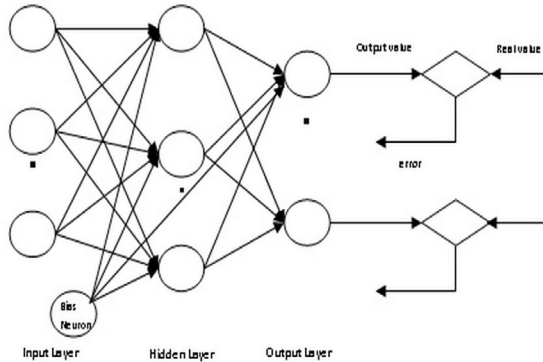


Figure 3: Typical back propagation network
Source: Che-Chiang Hsu and Chia-Yon Chen, 2003¹⁹

Figure 3. Typical back propagation network.
Source: Che-Chiang Hsu and Chia-Yon Chen, 2003¹⁹

For a three layered network the following steps are followed:

- *Select input variables and define output variables.*
- *Determine number of layer(s) and the number of neurons in hidden layers.* Any number of hidden layers can be taken. But, to choose most favorable number of hidden neurons is one of the challenging aspects in designing multi layer network feed forward network. In Hecht – Nielson approach is followed to finalize number of hidden neuron for a single hidden layer network i.e. $2(N_i + 1)$, where N_i represents number of input neurons.²⁶
- *Learning (or training) from historical data.* By learning process a NN modifies its weights with respect to external inputs for minimizing the global error. This equation which specifies this change is called learning rule.
- *Testing.* When a NN is trained well after learning, the neural networks are processed via a test set containing historical data that the network

has never seen. If the testing results are in an acceptable range, the network can be considered as fully trained. The next step i.e. forecasting can then be performed.

- *Post Processing.* The neural network output need de-scaling to generate the desired forecasted loads.
- *Error Analysis.* In forecasting, observation of error is important due to variation of the characteristics of load.

Overall, there may be a limit to what ANNs can learn from the data and make predictions. This limitation of ANNs may come from their non – parametric property. Since they are data – driven and model – free, ANNs are quite general but can suffer high variance in the estimation, that is, they may be too dependent on the particular samples observed.¹⁹

5.2 Fuzzy Logic Technique

The following approach is followed for long term load forecasting using fuzzy logic:

5.2.1 Transformation of Input Variables

The input variables are number of customers and type of customers, temperatures, energy consumption and statistical indices like consumption norms, consumer growth rate, growth rate. The values of the variables with normal growth are transformed as the difference of the corresponding variables and for exponentially growing variables as the relative difference. By using difference or relative difference is useful because the values are in a limited width and their future values can be predicted easily. Thus, instead of value X_{jk} of the j th variable X_j during the k th year,

Either difference d_{jk}

$$d_{jk} = X_{jk} - X_j(k-1) \quad (21)$$

Or the relative difference

$$r_{jk} = d_{jk} / X_j(k-1) \quad (22)$$

Can be taken and they are denoted by p_j

$$p_{jk} = r_{jk} \text{ or } d_{jk} : k = 1, \dots, Y \quad (23)$$

Where Y represents the years for which data is available.

If any input attributes of a year are missing, that year is omitted. Figure 4 and 5 show transformation of the input variable with normal & exponential growth rate.

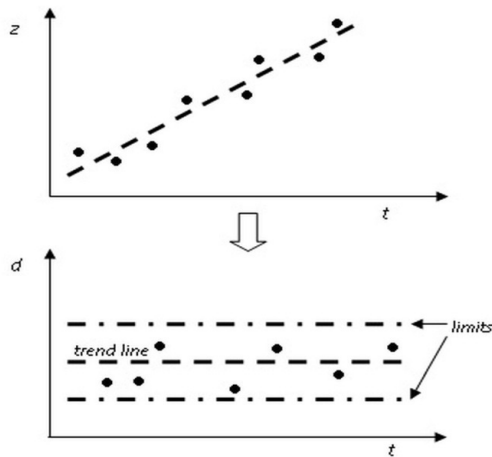


Figure 4. Transformation of input variables with normal growth.

Source: Charalambos N. Elias and Nikos D. Hatziaargyriou [11](#)

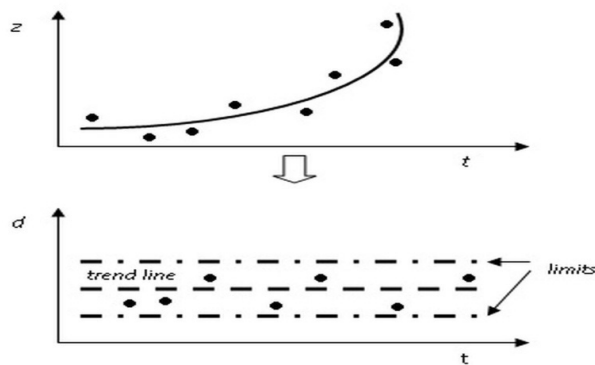


Figure 5. Transformation of input variables with exponential growth.

5.2.2 Fuzzification and Rule Base

It is done by triangular membership functions. The membership functions t is chosen in odd numbers of and the triangle's base width are also selected, for optimizing the performance of fuzzy based forecasting system. For each triangle the center of middle triangle c_j of a variable p_j and the initial value of the base width b_{ji} are given by the following expressions:

$$c_{ij} = \frac{\sum_{k=1}^Y p_{jk}}{Y} \dots \dots \dots (24)$$

$$b_{ji} = 2 \cdot ((\max_{k=1, \dots, Y} p_{jk} - \min_{k=1, \dots, Y} p_{jk}) / ((t_j - 1)^*)) \dots \dots \dots (25)$$

Where t_j : number of membership functions of p_j . While the center of middle triangle stays constant and the base

width is modified by $\pm a\%$ with step $s\%$. Thus, for each variable the number of possible triangles h is,

$$h = 1 + 2 \cdot \left(\frac{a}{s}\right) \dots \dots \dots (26)$$

So, for n number of input variables there will be HN number of combinations. Following the fuzzification and the rule base are computed for all alternative values of t_j and b_{ji} for all the variables.

5.2.3 Deduction Mechanism and Results Validation

All combinations of the rules are classified and the fuzzy output is obtained from the training years for every combination of input rules. The output is found by all weighted average of fuzzy values, rather than by the value that frequently occurs. Let's take an example and assume that each fuzzy output value corresponds to weight, -2 , -1 , 0 , 1 and 2 for "Very Negative", "NEgative", "ZEro", "PoSitive" and "Big Positive". In a certain rule the output may appear with the following frequencies:

$$VN(1) \quad NE(3) \quad ZE(2) \quad PS(2) \quad BP(2)$$

Based on the maximum frequency, the fuzzy value "NE" should have been chosen as output, whereas based on the weighted average

$$([1 \cdot (-2) + 3 \cdot (-1) + 2 \cdot 0 + 2 \cdot 1 + 2 \cdot 2] / ([1 + 3 + 2 + 2 + 2])) = 0.1 \dots \dots \dots (27)$$

The system chooses the closet fuzzy value "ZE". If a rule does not occur at all, then the fuzzy value "ZE" is selected as output. If a rule occurs only once, the output weight is divided by two, so that it does not affect forecasting more. Then taking the input data of evaluation set, the system sets up the left part of the rules for the corresponding years of this set. With the deduction mechanism and COA method, the difference between the forecasted value, which will be needed during the present year and the amount of energy of the last year, is calculated. At last, for each combination the mean absolute percentage error (MAPE) is found for the evaluation set and used as a criterion of comparison along with the results of the different combinations.

5.2.4 Input Variables' Selection Optimization

Examination of all combinations of the N input variables would need the execution of 2^N times of the fuzzy model. So, considering the optimization of membership function with the width of the triangle's base, the final combinations will be

$$\text{combinations} = \sum_{v=1}^N \binom{N}{v} \cdot h^v : t_j = \text{const}, v_j \quad (28)$$

As the initial preprocessing it gives nearly 15 variables, it becomes very important that the preprocessing is replicated with a correlation index between input – output and a correlation between the input variables, so that it decreases the number of combinations. If the correlation index between p_j and output y is more than a pre – specific value cor_1 , the transformed input variable p_j is kept for next processing; otherwise it is not considered any more. Then, a cross correlation analysis is performed for the retained inputs. If the correlation index among any two terms is less than a pre – specified value cor_2 then both terms are taken; otherwise, only the term with the highest correlation as compared to output y is kept. So the input variables decrease from N to n and so as the combinations.

5.2.5 Final Prediction

It is possible for combinations of n input variables, each input's fuzzy value is determined, which matches up with the data of each training year. So, rules are created for every year, whose number varies between $2^0 - 2^n$. It is repeated to check all combinations and the combination with minimum MAPE in the forecast of the evaluation set is selected. This combination can be used for midterm energy forecasting. The combination of base widths and input data is given. The forecast energy for each year is calculated from the difference between the forecasted value of this year and its previous year.

5.2.6 Standard Deviation

In order to calculate the standard deviation analytically the respective expression for only one triangular membership function is first found. Based on the basic principles of mechanics in the case of more than one triangular membership functions the mean value equals the abscissa of the gravity center and the standard deviation is equivalent to the moment of the inertia with respect to the axis, which crosses the gravity center and is parallel to the axis $m(x)$ of the ordinates.

The outline of the procedure to build the fuzzy model is presented in Figure 6.

It is seen from various literature that fuzzy logic is used for long term electrical load forecasting for a maximum period of five years.

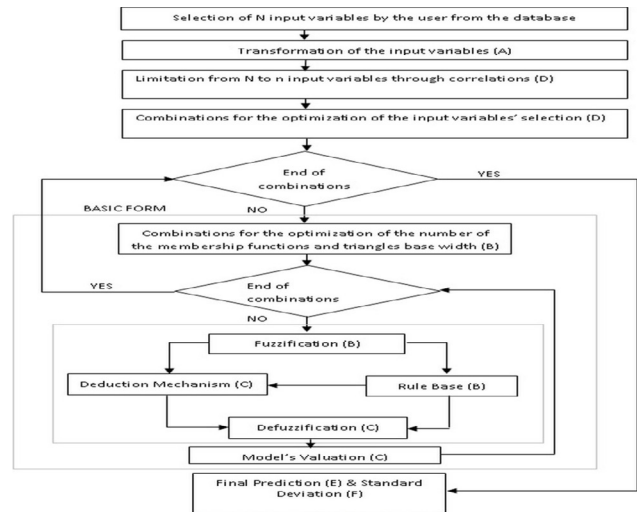


Figure 6. Flowchat to develop fuzzy model.

Source: Charalambos N. Elias and Nikos D. Hatziaargyeiou¹¹

5.3 Wavelet Network Technique for Long Term Electrical Load Forecasting

Wavelet theory provides powerful and flexible tool to decompose load data into different frequency components, making it possible to analyze the characteristics of each component and improve forecasting accuracy. Wavelet packet analysis is an extension of wavelet analysis and gives better frequency resolution.

Following approach is to be followed for using wavelet network for long term electrical load forecasting.²⁷

5.3.1 Finalization of Structure of the Wavelet Network

The wavelet network has a structure very similar to a multilayer neural network as shown in Figure 7.²⁷

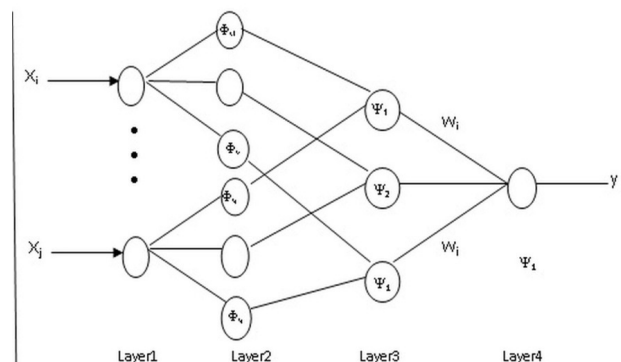


Figure 7. Schematic of the wavelet network.

Source: Khoa *et. al*²¹

This architecture shown in Figure 7 can approximate any desired function $y(x_1, x_2, \dots, x_{N_i})$ through 4 layers. Layer1 includes N_i input variables $x_1, x_2, \dots, x_{N_i}$. Layer2 consists of wavelet functions in the field of multi-resolution analysis that has been applied to image and signal processing.²⁷

$$\phi_{t,s}(\tau) = \phi\left(\frac{x-t}{s}\right) \quad (42)$$

$$\text{where } \tau = \frac{x-t}{s}$$

Φ is a set of daughter wavelets generated by dilation s and translation t from a mother wavelet Φ , the Polywog 2 is chosen as a mother wavelet in the presented article.²⁷

$$\phi(\tau) = (\tau^3 - 3\tau)e^{\frac{\tau^2}{2}} \quad (43)$$

Each input x_i is decomposed by its set of $\phi_{t,s}^2$. Layer 3, in the case of a problem with N_i inputs, multidimensional wavelets must be considered. The simplest way is the product of N_i mono-dimensional wavelets of each input.

$$\psi_i = \prod_{j=1}^{N_i} \phi_j(\tau) \quad (44)$$

If network has one input $N_i=1$ then $\Psi_i = \Phi_i$. Layer 4 is an input layer that sums k dimensions of multi resolution:

$$y = \sum_{i=1}^k w_i \psi_i(\phi) + \bar{y} \quad (45)$$

$\bar{y}$ is introduced in order to make it easier to approximate function with nonzero average, because wavelet $\Psi(x)$ has zero mean.

5.3.2 Initialization of the Network Parameters

This process consists in the evaluation of the parameters t_{ij} , s_{ij} , w_j and $\bar{y}$. For $i=1 \dots N_i$ is the number of inputs, for $j=1 \dots k$ is the number of multi-resolution dimensions. To initialize $\bar{y}$, the estimation of mean function $y(x_1, x_2, \dots, x_{N_i})$ is to be carried out. All WI are simply set to zero.²⁷ For each input x_j , t_j and s_j is set from the domain $[a, b]$ of x_j by selecting a point p between a and b and then:

$$t_j = p \text{ and } s_j = \zeta(b - a)$$

where $\zeta > 0$ and is a properly selected constant (typical value of ζ is 0.5). The point p divided $[a, b]$ into two

parts and the same procedure is recursively repeated to initialize t_2, s_2 and so on until N_j .

5.3.3 Training

As usual, the training is based on the minimization of the following quadratic cost function:

$$E = \frac{1}{2} \sum_{n=1}^N (y_n - d_n)^2 = \frac{1}{2} \sum_{n=1}^N e_n^2 \quad (47)$$

where y_n and d_n are the target output and the actual output for N patterns.

5.3.4 Forecasting

Depending on the utility study, the historical database used in the development of the long term loads consists of quarterly and/or annual data. The historical data base is called the estimation set. This estimation set is further divided into a training set and a testing set. Information in the training set is used directly for determining parameters of the forecast model, where the information in the testing set is used indirectly for testing again the errors of models in subsequent years. If the errors are small enough then these models can be implemented for more number of years.

5.4 Genetic Algorithm/Genetic Programming Technique for Long Term Electrical Load Forecasting

Genetic Algorithm (GA) method is based on the mechanism of natural selection and natural genetics, which combines the notion of survival of the fittest, random and yet structured, search and parallel evaluation of the points in the search space. GA is successfully applied to various electrical power system problems including long term load forecasting.²⁸ Genetic Programming (GP) is considered as a special case of GA, where each individual is a computer program (not just raw data). GP explores the algorithmic search space and evolves computer programs to perform a definite task.

In applying genetic programming to a problem, there are five major preparatory steps. These five steps involve determining:

- Predefined set of terminals,
- Predefined set of primitive functions,
- A fitness measure (called a "fitness function" to specify what needs to be done)

- The parameters for controlling the run, such as population size, reproduction operators, probabilities of the operator and so on, and
- The method for designating a result and the criterion for terminating a run.

Genetic programming is an iterative method. The algorithm used by genetic programming in finding solutions is summarized in Figure 8.

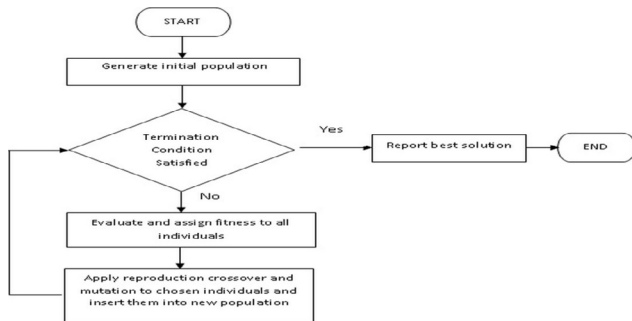


Figure 8. A Simplified form of the algorithm used by genetic programming.

Source: Kortham Karabulut *et al.*²²

The first step in the genetic programming, an initial population is generated of either by taking random compositions of functions and terminals or by taking a predefined strategy.

In the next step, the termination condition is checked. If the termination condition is reached, the process is ended and best results attained so far are reported. If the termination condition is not reached, the following steps are repeated:

- Each individual in the population is evaluated and assigned a fitness value using the fitness function.
- A new population is created by applying the following operations. The operations are applied to individuals taken from the population with a probability with respect to fitness:
 - i. Darwinian Reproduction: Reproduce an existing individual by simply copying it into the new population.
 - ii. Crossover: Create two new individuals by genetically recombining randomly chosen parts of two existing individuals using the crossover operation.
 - iii. Mutation: Create one new individual from one existing individual by mutating a randomly chosen part of the individual.

- The individual that is identified by the method of result designation is taken as the result for the run. This result may be a solution (or an appropriate solution) to the problem.

There are two primary reproduction operators. Crossover operation is one of them where two solutions are combined or joined to make a new solutions or offspring. A random point is chosen in both of the parents, and the nodes below the crossover points are exchanged between parents. Figure 9 shows an example of crossover operation.

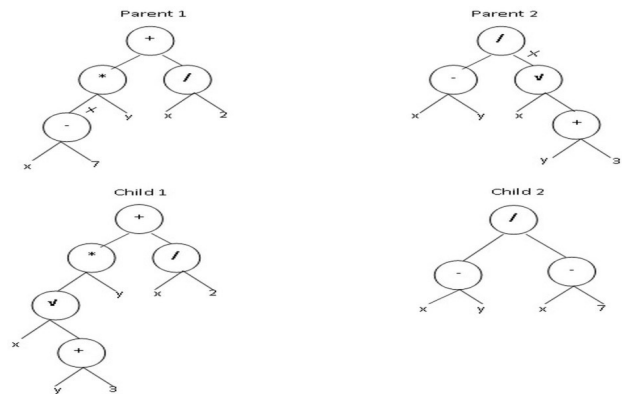


Figure 9. An example of crossover operation (x denotes the crossover point).

Source: Karabulut *et. al.*²²

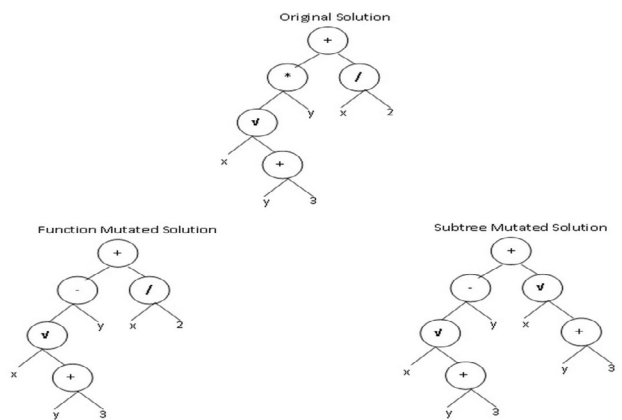


Figure 10. An example of mutation operation.

Source: Kortham Karabulut *et al.*²²

Mutation operation is one of the important features of GP (genetic programming). From two kinds of mutation operation one is either a randomly chosen function or a terminal, which can be replaced with a new function or terminal and the other is a randomly chosen sub tree that

Table 1. Summary of AI techniques used for load forecasting in the surveyed literatures

S. No.	Reference No.	Author Name	Brief
1	13	Bahman Kermanshahi, Hiroshi Iwamiya	ANN technique is used to forecast peak load for 18 years. Two scenarios viz. Business As Usual and Structure Reform were considered. The model has shown its ability to forecast the future load within 3%.
2	20	H.S. Hippert, C.E. Pedreira and R.C. Souza	Review and evaluation of various literature pertaining to NN based forecasting system is carried out from 1991 to 1999 and issues in designing NN based forecasting system is discussed.
3	31	Ladan Ghods and Mohsen Kalantar	It compares various techniques of long term electric load demand forecasting. It brings out the shortcoming of parametric method in handling non linear problems, while suggest usage of ANN for long term load forecasting. It also suggests usage of genetic algorithm to find future load. While Wavelet technique can be used to estimate peak and valley of load behaviour better than fourier series.
4	40	Pituk Bunnoon, Kusumal Chalermyanont, and Chusak Limsakul	It compares the long term load forecasting using ANN technique by taking whole country and comparing with the forecast value by dividing country into various regions. The forecasting is done for one year period and MAPE in the second case is 1.47 as against 4.75
5	41	George J. Tsekouras, Nikos D. Hatziaargyriou and Evangelos N. Dialynas	The paper uses adaptive ANN method for mid term forecasting in three years (2001 – 2003) using adaptive ANN model. Error is found between 0.01 % to – 1.06 % ANN model provides better or atleast equivalent results and it is not sensitive to small variance of the input variance.
6	42	T.Q.D. Khoa, L.M. Phoung, P.T.T. Binh and N.T.H. Lien	This paper deals with application of wavelet & NN and compare the energy sales forecast for 3 years. Energy sales forecast using functional link network (FLN), multilayer neural net (MLNN) and wavelet network is compared and it concludes that the accuracy of wavelet network is similar to extended FLNN and MLNN.
7	43	Korhan Karabulut, Ahmet Alkan, Ahmet S. Yilmaz	The comparison of three models has been made, polynomial model, power equation model and genetic propagation model. The calculated Sum of Square Error is the least using Genetic Progression. The forecast has been done for five years.
8	44	M.A. Farhat	Estimation of industrial load has been made for 24 months in the industrial city of 10 Ramadan, Egypt using neural network technique and fuzzy interference network The test result show the proposed method to provide considerable improvement of the forecasting accuracy.

can be replaced by another sub tree. Figure 10 shows an example of each kind of mutation.

In order to carry out the forecasting using GP, GPLAB Tool box is used. GPLAB is free, highly configurable and extendable genetic programming toolbox supporting up to date features of the recent genetic programming research.

6. Conclusion

In this paper an effort is made to highlight theoretical aspects of three parametric techniques and four artificial intelligence based methods used in long term electric load forecasting. Under parametric methods, trend analysis, end use technique and econometric technique are dis-

cussed while under artificial intelligence method, ANN, fuzzy logic, wavelet and genetic algorithm techniques used in long term load forecasting is discussed.

In end – use technique John H. Broehl model, DEFENDUS model and the approach followed by EPS followed by Govt. of India is discussed. In econometric, Auto Regressive Lag Model and Classical Econometric model is discussed. In ANN based technique, recurrent network, multi layered feed forward network, multi layered feed forward back propagation network, single layer feed forward network and ANN modeling in forecasting are discussed. Basic principles of other artificial intelligence methods such as fuzzy logic, wavelet network and genetic algorithm/programming and their applications in long term load forecasting are also discussed. More studies are required to find the applicability of these methods to Islands, by taking as a case study.

All the surveyed literatures pertaining to Artificial Intelligence technique is complied and presented in Table 1. Peak load forecasting for up to 2020 is carried out for Japan under two growth scenarios by using ANN. Three years load forecasting is carried out using Adaptive ANN model. It also shows that for long term energy forecasting, ANN technique is used in most of the cases compared to other Computational Intelligence techniques like fuzzy logic, genetic algorithm and wavelet network. The comparison of time series model and ANN model on forecasting of long term electric load shows that the ANN gives results which are closed to the actual data. So it can be concluded that ANNs can be used for long-term load forecasting with minimum errors. It also shows the ANN model provides accurate results using a minimum amount of historical data rather than other forecasting techniques.

7. References

1. Boylan JE, Goodwin PM. Reproducibility in forecasting research. *International Journal of Forecasting*. 2015; 3(1):79–90. Crossref
2. Okolobah V, Ismail Z. An Empirical Mode Decomposition Approach to Peak Load Demand Forecasting. *Indian Journal of Science and Technology*. 2013; 6(9):5201–7.
3. Clements KA, Wood AJ. Error Analysis of Econometric load forecasting models. *IEEE Transactions on Power Apparatus and Systems*. 1979; 98(2):393–9. Crossref
4. Zachariadis T. Forecast of electricity consumption in Cyprus up to the year 2030: the potential impact of climate change. *Energy Policy*. 2010; 2(1):744–50. Crossref
5. Sanghanil A, Bhatt N, Chauhan NC. A Review of Soft Computing Techniques for Time Series Forecasting. *Indian Journal of Science and Technology*. 2016; 9(S1):1–6.
6. Samuel IA, Emmanuel A, Odigwe IA, Firstlady CFN. A Comparative Study of Regression Analysis and Artificial Neural Network Methods for Medium-Term Load Forecasting. *Indian Journal of Science and Technology*. 2017; 10(10):1–7. Crossref
7. Elias CN, Hatziaargyriou ND. An annual midterm energy forecasting model using fuzzy logic. *IEEE Transaction on Power Systems*. 2009; 24(1):469–78. Crossref
8. Bansal RC. Overview and literature survey of artificial neural networks applications to power systems. *Institute of Engineering (India). Journal-El*. 2006; 86(1):282–96.
9. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*. 1998; 1(1):1435–62. Crossref
10. Azoff EM. *Neural network time series forecasting of financial markets*. John Wiley and Sons: Chichester; 1994.
11. Makridakis S, Wheelwright SC, McGee VE. *Forecasting: Methods and Applications*. 2nd Edition. John Wiley: New York; 1983. Crossref, Crossref, Crossref
12. Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. *Electrical Power and Energy Systems*. 2002; 2(1):789–97. Crossref
13. Hsu CC, Chen Yon. Regional load forecasting in Taiwan–applications of artificial neural networks. *Energy Conversion and Management*. 2003; 44(1):1941–9. Crossref
14. Ghosh S, Kishore NK. Modeling PD inception voltage of epoxy resin post insulators using an adaptive neural network. *IEEE Transaction on Dielectrics and Electrical Insulation*. 1999; 6(1):131–4. Crossref
15. Khoa TQD, Phoung LM, Binh PTT, Lien NTH. Application of wavelet and neural network to long term load forecasting. *IEEE Power on International Conference on Power System Technology*; Singapore; 2004. p. 840–4. Crossref
16. Karabulut K, Alkan A, Yilmaz AS. Long term energy consumption forecasting using genetic programming Association for Scientific Research. *Mathematical and Computational Applications*. 2008; 13(2):71–80. Crossref
17. Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. *Energy*. [cited 2017 Jun 3]. Available from: Crossref.
18. Boylan JE, Goodwin PM. Reproducibility in forecasting research. *International Journal of Forecasting*. 2015; 31(1):79–90. Crossref
19. Mustapha M, Mustafa MW, Khalid SN, Abubakar I, Abdilahi AM. Correlation and Wavelet-based Short-Term Load Forecasting using Anfis. *Indian Journal of Science and Technology*. 2016; 9(46):1–8. Crossref

20. Gujarati DN, Sangeetha. Basic Econometrics. 4th Edition. Tata McGraw-Hill: New Delhi; 2008.
21. Bunn DW, Farmer ED. Comparative Models for Electrical Load Forecasting. John Wiley & Sons. 1985; 2(4):501–5.
22. Expansion Planning for Electrical Generating Systems: A Guidebook. Vienna: International Atomic Energy Agency; 1984.
23. Ghods L, Kalantar M. Methods for long term electric load demand forecasting: a comprehensive investigation. IEEE International Conference on Industrial Technology; 2008. p. 1–4. Crossref
24. Feinberg EA, Genethliou D. Load Forecasting in Applied Mathematics for Restructured Electric Power Systems. Optimization Control and Computational Intelligence; Springer, New York; 2005. p. 269–85. Crossref
25. Background paper for the 1st meeting of 17th Electric Power Survey Committee scheduled on 2004 Jan 14.
26. Sharda R, Patil RB. Connectionist approach to time series prediction: an empirical test. Journal of Intelligent Manufacturing. 1992; 3(1):317–23. Crossref
27. Tang Z, Fishwick PA. Feed forward neural nets as models for time series forecasting. ORSA Journal on Computing. 1993; 5(4):374–85. Crossref
28. Reddy AKN, Gladys DS, Balachandra P, D'Sa A. A development – focused end use oriented electricity scenario for Karnataka. Economic and Political Weekly. 1991; 26(14):891–910.