INTERNATIONAL JOURNAL OF
ENERGY AND ENVIRONMENT

Volume 2, Issue 1, 2011 pp.123-150
Journal homepage: www.IJEE.IEEFoundation.org

An overview of load demand and price forecasting methodologies

George Kourtis, Ioannis Hadjipaschalis, Andreas Poullikkas

Electricity Authority of Cyprus, P.O. Box 24506, 1399 Nicosia, Cyprus.

Abstract
In this work, an overview of the various methodologies developed in recent years for short, mid and long term load and price forecasting is carried out. In the analysis the advantages and disadvantages of each method are introduced, together with the factors that influencing the different types of forecasting. Unless the effects of these factors are well taken into consideration errors can occur in the forecasting results and that results in increasing operational costs. The analysis indicates that the best suited method for all types of forecasting is artificial neural network, which outperforms better with nonlinear functions and on weekend days or national holidays. If are not to be distinguished from week day data, weekend and national holidays data a good alternative would be an autoregressive integrated moving average based method.

Keywords: load forecasting, price forecasting, unit commitment, artificial neural networks.

1. Introduction
Demand prediction is an important aspect in the development of any model for electricity planning. The form of the demand depends on the type of planning and accuracy that is required. It can be represented as an annual energy demand (GWh), a peak load demand (MW), or load duration curves like daily, weekly or annual. The load forecasts help in determining which devices to operate in a given period, so as to minimize costs (or maximize profit) and secure demand even when local failures may occur in the system, [1].
Load forecasting can be short term, mid term and long term. Long-term forecasts are required for resource planning, utility expansion and staff hiring. Medium-term forecasts are used for purchasing fuel and revising electricity tariffs. Short-term load forecasting is important for scheduling functions, such as generator unit commitment, hydro-thermal coordination, short-term maintenance, fuel allocation, power interchange, transaction evaluation, as well as network analysis functions, such as dispatcher power flow and optimal power flow, [2].
In short term load forecasting, the prediction times are of the order of hours. The time boundaries are from the next hour, or possibly half-hour up to 168h. The basic quantity of interest is the hourly integrated total system load. In addition, it is also concerned with the daily peak system load, the values of system load at certain times of the day, the hourly or half hourly values of system energy and the daily and weekly system energy. The principal objective of short term load forecasting is to provide the load predictions for the basic generation scheduling functions, assessing the security of the power system at any time point and timely unit dispatch information.
The primary application of the short term load forecasting function is to drive the scheduling functions that determine the most economic commitment of generation sources consistent with reliability

requirements, operational constraints and policies, and physical, environmental and equipment limitations. A second application of short term load forecasting is for predictive assessment of the power system security. The third application is to provide system dispatchers with timely information (the most recent load forecast), with the latest weather prediction and random behavior taken into account, [3].
Short-term demand forecasting plays a role in the process of regulation. A precise estimate of demand is important for the purpose of setting tariffs. A detailed consumer category-wise consumption forecast helps in the determination of a just and reasonable tariff structure wherein no consumer pay less than the cost incurred by the utility for supplying the power. Also, the utility can then plan the power purchase requirements so as to meet the demand while maintaining the merit order dispatch to achieve optimization in the use of their resources. A time of day tariff structure to manage peaks and troughs in electricity demand, an hour-by-hour load shape forecast has become an essential prerequisite. Accurate load forecasts help the utilities to operate at the highest possible efficiency, [4].
The system load is the sum of all the individual demands at all the nodes of the power system. However, the demand or usage pattern of an individual load (device) or customer is quite random and highly unpredictable. Also, there is a very broad diversity of individual usage patterns in typical utility. These factors make it impossible to predict the system demand levels by extrapolating the estimated individual usage patterns. Fortunately, however, the totality of individual loads results in a distinct consumption pattern which can be statistically predicted, [3].
Every electric utility should be able to have an idea about the amount of required power in order to prepare for maximum electric load demand ahead on time with a long-term load forecasting. However, neither the accurate amount of needed power nor the preparation for such amounts of power is as easy as it looks, because (a) long-term load forecasting is always inaccurate, (b) peak demand is very much dependant on temperature (at peak period), (c) some of the necessary data for long-term forecasting including weather condition and economic data are not available, (d) it is very difficult to store electric power with the present technology and (e) it takes several years and requires a great amount of investment to construct new power generation stations and transmission facilities, [5].
The fact that electricity cannot be stored means that amounts sold, amounts bought and amounts consumed must all coincide each and every time. There thus must be a number of mechanisms and devices to link the commercial side of the business (buying and selling) and its technical side (generating and consuming electricity), so that they are compatible. This is achieved by accurate electricity price forecasting which can help electricity generators regulate their generation capacity to remain within their financial profit margins, [6].
In most competitive electricity markets the hourly price series presents the following characteristics: high frequency, non-constant mean and variance (non-stationary series), multiple seasonality (corresponding to a daily and weekly periodicity, respectively), calendar effect (such as weekends and holidays), high volatility and high percentage of unusual prices (mainly in periods of high demand) due to unexpected or uncontrolled events in the electricity markets. This unstable behavior makes accurate price forecasting difficult to achieve, [7].
Proper electricity price forecast can help to build up cost effective risk management plans for the participating companies in the electricity market. If the electricity market price can be predicted properly, the generation companies and the load service entities as the main market participating entities can reduce their risks and maximize their outcomes further, [7].
The purpose of this work is to provide an overview of the various methodologies developed in recent years for short, mid and long term load and price forecasting. In section 2, the methods for short term load forecasting are described, together with the factors that influencing short term load forecasting. Section 3, concern mid term load forecasting and section 4, long term load forecasting. In section 5 the various available methods for electricity price forecasting are presented. The conclusions are summarized in section 6.

2. Short term load forecasting methods

Many short term forecasting techniques have been developed, ranging from very simple extrapolation methods to more complex time-series techniques, extensive accounting frameworks and optimization methods or even hybrid models that use a combination of these for purposes of prediction. The forecaster may use a combination of techniques that give him aggregate annual forecasts and those that predict hour-by- hour demand for electricity in individual sectors. This helps greatly in tariff setting and designing demand-side management programs, [4].

2.1 Peak load models
In these models only the daily or weekly peak load is modeled, usually as a function of the weather. Time does not play a role in such models which are typically of the form

$$P = B + F(W) \quad (1)$$

where P is the peak load. The base load B is an average weather insensitive load component. The weather variables W can include the temperature at the peak load time or a combination of predicted and historical temperatures. Humidity, light intensity, wind speed and precipitation have also been considered in such models. The function of the weather $F(.)$ can be linear or nonlinear.
The advantages of a peak load models are its structural simplicity and its relatively low data requirements to initialize and to update. The parameters of these models are estimated through linear or nonlinear regression. The disadvantages of such models are that they do not define the time at which the peak occurs, nor do they provide any information about the shape of the load curve. Since the models are essentially static, dynamic phenomena such as correlation across the periods cannot be forecasted, [3].

2.2 Load shape models
Such models describe the load as a discrete time series over the forecast interval. The load sampling time interval is typically one hour or one half hour, while the quantity measured is generally the energy consumed over the sampling interval in MWh. Basically, there exist two types of load-shape models: time-of-day and dynamic models. Combinations of these two basic types are also possible, [3].

2.2.1 Time-of-day models
The time-of-day model defines the load at each discrete sampling time of the forecast period of duration T by a time series. In its simplest form, the time-of-day model stores T load values based on previously observed load behavior. Some utilities today still use the previous week's actual load pattern as a model to predict the present week's load. Alternatively, a set of curves is stored for typical weeks of the year and for typical weather conditions, such as wet, dry, cloudy, or windy days, which are heuristically combined with the most recent weekly load pattern to develop the forecast.
A more common time-of-day model takes the form

$$z(t) = \sum_{i=1}^{N} a_i f_i(t) + v(t), \quad t \in \tau \quad (2)$$

where the load at time t, $z(t)$, is considered to be the sum of a finite number of explicit time functions $f_i(t)$, usually sinusoids with a period of 24h or 168h, depending on the forecasting lead time. The coefficients a_i are treaded as slowly time-varying constants, while $v(t)$ represents the modeling error, assumed to be white noise. The model assumed to be valid over a range of time interval τ covering the recent past, the present and a future time period covering the maximum lead time.
The advantages of these models are that they are structurally quite simple and that most parameters can be updated very simply through linear regression or linear exponential smoothing. The nature of these schemes is such that recursive algorithms requiring a relatively low computational effort can be devised to update the parameters, as well as the forecast, as new load data is measured. The disadvantage is that time-of-day models do not accurately represent the stochastically correlated nature of the load process, or its relation to weather variables.
There is a second class of time-of-day models, which is based on spectral decomposition. It has the advantage that the time functions chosen to represent the load time series are optimal in the sense that they can more closely approximate its autocorrelation function that is its second-order probabilistic behavior. The disadvantage is that it is also susceptible to error under conditions of sudden and large weather variations, since these effects are not explicitly modeled, [3].

2.2.2 Autoregressive moving average dynamic model
Dynamic load models recognize the fact that the load is not only a function of the time of day, but also of its most recent behavior, as well as that of weather and random inputs. The autoregressive moving average (ARMA) model depends primarily on the time of day and on the normal weather pattern for the

particular day and on weather pattern deviations from normal and random correlation effects, which influence an additive residual load. The additive nature of the residual load is justified by the fact that such effects are usually small compared to the time-of-day component. Nonlinear models describing the interaction of the periodic and residual component also exist, but are less common. ARMA models are made up of different types of models such as Box-Jenkins, time series, transfer function, stochastic and autoregressive integrated moving average (ARIMA).
The ARMA type model takes the general form

$$z(t) = y_p(t) + y(t) \tag{3}$$

where $y_p(t)$ is a component which depends primarily on the time of day and on the normal weather pattern for the particular day. This component can be represented by a periodic time function of the type given by (3). The term $y(t)$ is an additive load residual term describing influences due to weather pattern deviations from normal and random correlation effects. The additive nature of the residual load is justified by the fact that such effects are usually small compared to the time-of-day component. Nonlinear components describing the interaction of the periodic and residual components also exist, but are less common. The residual term $y(t)$ can be modeled by an ARMA process of the form

$$y(t) = \sum_{i=1}^{n} a_i \, y(t-i) + \sum_{k=1}^{n_u} \sum_{j_k}^{m_k} b_{j_k} u_k(t-j_k) + \sum_{h=1}^{H} c_h \, w(t-h) \tag{4}$$

where $u_k(t)$, $k=1, 2,\ldots,n_u$ represent the n_u weather dependent inputs. The impact of the weather-dependent variables is considered to be significant. These inputs are functions of the deviations from the normal levels for a given hour of the day of quantities such as temperature, humidity, light intensity and precipitation. The inputs $u_k(t)$ may also represent deviations of weather effects measured in different areas of the system. The process $w(t)$ is a zero-mean white random process representing the uncertain effects and random load behavior. The parameters a_i, b_{jk} and c_h, as well as the model parameters n, n_u, m_k and H are assumed to be constant but unknown parameters to be identified by fitting the simulated model data to observed load and weather data.
Pre-filtering is a process in which the load data are pre-filtered so as to eliminate the periodic component as an explicit time series. The pre-filtering is basically done by defining a new load process of the form

$$z'(t) = z(t) - z(t - t_p) \tag{5}$$

where t_p is the period of the time-of-day component (usually 24h or 168 h). The resulting process is, therefore free of periodic terms and satisfies an ARMA equation similar to that of (5). This now has the advantage that more standard techniques can be applied to the identification of the parameters of the resulting ARMA model. The disadvantage of pre-filtering lies in the fact that such a scheme is basically equivalent to differentiating a process which almost certainly contains measurement and modeling errors. The result is a potential amplification of measurement errors leading to corresponding modeling inaccuracies.
Explicit modeling of the time-of-day component, on the other hand, does not require pre-filtering and is, therefore, not subject to this type of pre-filtering errors. However for such models, a nonlinear parameter estimation scheme must be used to identify the model parameters. This results in a slight increase in computational effort in the parameter estimation step.
Only some ARMA models include weather as an input. Those that do not include weather, automatically update some parameters to take into account the effect of meteorological variations on the load. This approach is not satisfactory during rapidly changing climatic conditions under which the assumption that the load process is stationary is no longer satisfied. Some of the available ARMA models describe meteorological effects by additional explicit inputs, while others rely on a more heuristic approach where the load process is corrected for temperature influences before applying an ARMA model to the corrected load. The most important weather input is based on the temperature deviations and is usually expressed as a nonlinear function of the differences between the actual and the normal temperatures. Such functions take into account the varying effects of temperature on load during the different seasons,

dead bands and other nonlinearities. Certain nonlinear effects are, however, well known. Thus in the summer, most systems experience higher loads due to increasing temperature, with the inverse phenomenon taking place in the winter. Therefore it is not the absolute value of the weather variable that affects the load, but its deviation from some normal level for that particular hour of the day and for that specific day of the year. The time-of-day or periodic component will take care of the long-term seasonal effect of weather on the power consumption.

The identification of the parameters of an ARMA model is generally more computationally intensive than those of the time-of-day models; however, this extra effort is needed to obtain a more robust model that incorporates dynamic, weather and random effects. In the long run, less parameter tuning is required and better forecasting performance is obtained.

In general, the updating of model parameters is not a very computationally demanding task, even in cases which require an iterative solution of a nonlinear estimation problem. In models where parameters may be estimated using linear regression, the parameter updating may be performed recursively on-line as new load and weather data are acquired. Such frequent parameter updating is unnecessary, unless the model is very simple, such as a pure time-of-day model, which requires continuous updating. For more elaborate model types, such as ARMA, the model structure and its parameters remain unchanged over a period of a few days. The updating of these parameters on an hourly basis may, in fact, be undesirable, particularly during periods of anomalous load behavior. In such instances, the model parameters should definitely not be updated. For ARMA models, daily parameter updating is probably sufficient. In this case, the data from the previous 24h, after cleaning their anomalous behavior, are added to the data set and the oldest 24h of data are removed. Daily parameter updating is not a critical task and can be done at a time when the computer is least busy, [3].

There are various special cases of ARMA models, such as double seasonal ARIMA modeling, regression method with Principal Component Analysis (PCA), non-parametric regression method and wavelet-armax-winters (WAW) method.

2.2.3 State-space dynamic models

Dynamic load models recognize the fact that the load is not only a function of the time of day, but also of its most recent behavior, as well as that of weather and random inputs. An ARMA model can be converted into a state-space model and vice versa, so that conceptually there are no fundamental differences between the two types of models. In these models, the load at time t, $z(t)$, is generally given by

$$z(t) = \mathbf{c}^T \mathbf{x}(t) \tag{6}$$

where

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{w}(t) \tag{7}$$

Here the state vector at time t is denoted by $\mathbf{x}(t)$, the vector of weather variable-based input is $\mathbf{u}(t)$, while the vector of random white noise inputs is $\mathbf{w}(t)$. The matrices $\mathbf{A}$, $\mathbf{B}$ and the vector c are assumed constant. There exist a number of variations of this basic state-space model. In some cases, the states $x_i(t)$, $i=1, 2, \ldots N_s$, may represent the periodic load component for a certain day of the week at a given hour, or a parameter of this model, or a combination of load and weather-dependent inputs.

One difference between the state-space and ARMA models lies in the fact that the available techniques for state-space models assume that the parameters defining the periodic component of load are random processes. In essence, this allows one to make use of some a priori information about their values which may help in the parameter estimation step via Bayesian techniques. One possible area where state-space methods may prove advantageous is the development of bus load forecasting, where the bus loads exhibit a high degree of correlation, [3].

2.2.4 Time series methods

This method is defined to be an ordered set of data values of a certain variable. These models are essentially econometric models, where the only explanatory variables used are lagged values of the variable to be explained and predicted. The intuition underlying time-series processes is that the future behavior of variables is related to its past values, both actual and predicted, with some built-in adjustment to take care of how past realization deviated from the one expected. Thus, the essential prerequisite for a time-series forecasting technique is data for the last 20 to 30 time periods. The general equation of these models is

$$y(t) = F\{\bar{x}(t), \bar{x}(t-dt), ..., \bar{x}(t-ndt), y(t-dt), ..., y(t-ndt)\} \quad (8)$$

where *y(t), y(t-dt), y(t-ndt)* represent the load at *t, t-dt* and *t-ndt, dt* being the time period considered, represents a vector of other factors such as the day, the hour or the temperature. The difference between econometric models based on time series data and time series models lies in the explanatory variables used. In an econometric model, the explanatory variables (such as incomes, prices, population) are used as causal factors while in the case of time series models only lagged (or previous) values of the same variable are used in the prediction. The most valuable applications of time series come from developing short term forecasts, for example monthly models of demand for three years or less. Another advantage of time-series models is their structural simplicity. They do not require collection of data on multiple variables. Observations on the variable under study are completely sufficient.
A disadvantage of these models however, is that they do not describe a cause and effect relationship. Thus a time series does not provide insights into why changes occurred in the variable. Often in the analysis of time series data, there are technical problems wherein more than one of the variables is highly correlated with another (multi-co linearity) or with its own past values (auto-correlation). Time series may be made more accurate, but this requires large amounts of historical good quality data, [4].

2.3 Trend method
It may serve as a useful cross check in the case of short-term forecasts. This method falls under the category of the non-causal models of demand forecasting that do not explain how the values of the variable being projected are determined. Here the variable to be predicted is expressed purely as a function of time, which is obtained as the function that best explains the available data.
It has the advantage of simplicity and ease of use. However, the main disadvantage of this approach lies in the fact that it ignores possible interaction of the variable under study with other economic factors. For example the role of incomes, prices, population growth and urbanization, policy changes, are all ignored by the method. The underlying notion of this analysis is that time is the factor determining the value of the variable under study, or, the pattern of the variable in the past will continue on the future. Therefore it does not offer any scope to internalize the changes in factors such as the effects of government policy (pricing or others) underlying institutional structure, regulatory regimes, demographic trends, aggregate and per capita growth in incomes and technological developments, [4].

2.4 End-use method
The end use approach attempts to capture the impact of energy usage patterns of various devices and systems. This approach focuses on various applications in the residential, commercial, agricultural and industrial sectors of the economy. The end-use method is based on the premise that energy is required for the service that it delivers and not as a final good. The following relation defines the end use methodology for a sector

$$E = S \times N \times P \times H \quad (9)$$

where E is the energy consumption of an appliance in kWh; S is the penetration level in terms of number of such appliances per customer; N is the number of customers; P is the power required by the appliance in kW and H is the number of hours of appliance use. This, when summed over different end-uses in a sector, gives the aggregate energy demand. This method takes into account improvements in efficiency of energy use, utilization rates, inter-fuel substitution etc, in a sector, as these are captured in the power required by an appliance, P. In the process the approach implicitly captures the price income and other economic and policy effects as well.
The advantage of end-use approach is that it is most effective when new technologies and fuels have to be introduced and when there is lack of adequate time-series data on trends in consumption and other variables. However the approach demands a high level of detail on each of the end-uses. This method

may lead to a mechanical forecasting of demands, without adequate regard for behavioral responses of consumers. It also does not give regard to the variations in the consumption patterns due to demographic, socio economic or cultural factors, [4].

2.5 Combining econometric and time-series models

By this combination we achieve greater precision in the forecasts. This has the advantage of establishing causal relationships as an econometric model along with the dependency relationship. The functional form of the model is arrived at after a trial and error process. A model is built using the available data, truncating the last few observations. The procedure for testing the model entails making predictions for the last few time periods for which actual data are available and were truncated. The functional form where the forecasts have least deviations from the data available is chosen, [4].

2.6 Integration of econometric and end-use approach

This would allow integration of physical and behavioral factors in a common framework, while the econometric relationships would internalize the influence of price income and policy effects, the end-use approach will provide an accounting plane for aggregating end-use and sectored energy demands projected into the future. The integrated approach will provide a better grasp of many diverse influences that shape the demand for energy into the future, [4].

2.7 Exponential smoothing for double seasonality

Exponential smoothing has a widespread use in automated applications, such as inventory control. The application requires an extension of the standard Holt-Winters exponential smoothing formulation to accommodate the two seasonal cycles in the electricity demand series. This involves the introduction of an additional sectional index. An important point to note regarding the double seasonal exponential smoothing approach is that there is no model specification involved. This gives the method strong appeal in terms of simplicity and robustness. It involves a small number of parameters, which make it easy to implement and attractive for online demand forecasting, as it has a very good forecasting performance. The formulation for double seasonality is given in the following expressions

$$S_t = a(y_t/(D_{t-s_1}W_{t-s_2})) + (1-a)(S_{t-1}+T_{t-1}) \tag{10}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{11}$$

$$D_t = \delta(y_t / S_t W_{t-s_2})) + (1-\delta) D_{t-s_1} \tag{12}$$

$$W_t = \omega(y_t / S_t W_{t-s_1})) + (1-\omega) W_{t-s_2} \tag{13}$$

$$\hat{y}_t(k) = (S_t + k T_t) D_{t-s_1+k} W_{t-s_2+k} + \phi^k (y_t - ((S_{t-1}+T_{t-1}) D_{t-s_1} W_{t-s_2})) \tag{14}$$

S_t and T_t are the smoothed level and trend; D_t and W_t are the seasonal indices for the intraday and intraweek seasonal cycles, respectively; a, γ, δ and ω are the smoothing parameters and $\hat{y}_t(k)$ is the k step-ahead forecast made from forecast origin t. The term involving the parameter φ, in the forecast function expression ( ), is a simple adjustment for first-order autocorrelation. All the parameters in the method, a, γ, δ, ω and φ, are estimated in a single procedure by minimizing the sum of squared one step-ahead in sample errors. The initial smoothed values for the level, trend and seasonal components are estimated by averaging the early observation, [1].

2.8 Artificial neural networks

Artificial neural networks (ANNs) are mathematical models based upon the functioning of the human brain and are composed of three different layers (input, hidden and output layers), each of which are composed of a certain number of neurons. Two characteristics of ANNs are: the ability to approximate practically any function (even non-linear ones) and the opportunity for piece-wise approximations of the functions. From a mathematical point of view, ANNs can approximate the best function to a set of data, which is especially important when the functions are complex. Moreover, ANNs are non-linear by nature, which means that they can not only correctly estimate non-linear functions, but also extract non-linear elements from the data. An ANN with one or more hidden layers can separate the space in different areas and build different functions for each of them. This means that ANNs have the capacity to build non-linear piece-wise models. Estimation with ANNs can be automatized and model review is not necessary, given the fact that they learn automatically. ANNs have been used to deal with the non-linearity in time series forecast with good results. The main advantage of ANN in forecasting problems is that they are capable of inferring hidden relationship in data, [6, 7].

For short term load forecasting there are two approaches. In one the neural network is constructed for each type of day or subset of hours. In the other one a single neural network is developed, but within the input variables information concerning the specific type of day or hour is required. The neural network consists of a set of k inputs, which are connected to an output. The output is demand and the inputs are lag demand.

Forecasting with ANN involves two steps: training and simulation. Training of feed-forward networks is normally performed in a supervised manner. One assumes that a training set is available, given by the historical data, containing both inputs and the corresponding desired outputs, which is presented to the network. The adequate selection of inputs for neural network training is highly influential to the success of training. In the learning process, a neural network constructs an input-output mapping, adjusting the weights and biases at each iteration based on the minimization of some error measure between the output produced and the desired output. Thus, learning entails an optimization process. The error minimization process is repeated until an acceptable criterion for convergence is reached. The knowledge acquired by the neural network through learning process is tested by applying new data that it has never seen before, called the testing set. The network should be able to generalize and have an accurate output for this unseen data, [1, 7].

The main difficulty of the neural networks training method is the non-linear optimization problem. In such problems the chance to find the best solution using typical, gradient-based learning techniques starting from a large, multi-layered network is minimal, often encountering multiple local minima in the cost function. Therefore, in general, there may be no algorithm capable of finding the optimal set of parameters which has computation time that is bounded by a polynomial in the input dimension. This means that the optimization algorithm may converge to different local minima when starting from different initial guesses for the parameters, [8].

ANN's are able to approximate any continuous function at an arbitrary accuracy, provided the number of hidden neurons is sufficient. However, this ability has a downside that such close approximation can become an approximation to the noise. As a consequence, the model yields solutions that generalize poorly when new data are presented. This problem is called over-fitting and may come about because the ANN is too complex (i.e. it possesses too many parameters). The complexity determines the generalization capability (measured by the generalization or test error) of the model since an ANN that is either too simple or too complex will give poor predictions. There are mainly two approaches to controlling ANN complexity, namely architecture selection and regularization techniques.

Architecture selection controls the complexity by varying the number of ANN's parameters (called weights and biases). One of the simplest ways involves the use of networks with a single hidden layer, in which the number of free parameters is controlled by adjusting the number of hidden units. Other approaches consist in growing or pruning the network structure during the training process. The approach taken by the pruning methods is to start with a relatively large network and gradually remove either connections or complete hidden units, [9].

Neural networks suffer from a number of limitations, including difficulty in determining optimum network topology and training parameters. Such design parameters include the number and size of the hidden layers, the type of neuron transfer functions for the various layers, the training rate and momentum coefficient, and training stopping criteria to avoid over-fitting and ensure adequate generalization with new data. Another limitation is the black box nature of neural network models, giving poor explanation facilities and providing little insight into the modeled relationship and the relative significance of various inputs, [10].

The technique of regularization encourages smoother network mappings by favoring small values for the ANN parameters. Indeed, small values for the weights decrease the tendency of the model to over-fit. One of the simplest forms of regularizer is called weight decay and a regularization coefficient (if it is too small, then over-fitting occurs, if it is too large then the model cannot fit the data well) allows controlling the degree of regularization. However, each of these techniques requires tuning of a parameter (i.e. regularization coefficient, number of hidden units, pruning parameter) in order to maximize the generalization performance of the ANN. Classically, the setting of this control parameter is done by using the so called cross-validation (CV) techniques. Indeed, CV provides an estimation of the generalization error and therefore, offers a possibility to select the best architecture or the optimal regularization coefficient. Unfortunately, CV presents several disadvantages.

First, the CV technique needs a separate data set (so fewer data for the training set), named validation set, in order to evaluate the variation of the generalization error (validation error) as the number of hidden neurons or the value of the regularization coefficient is changed. The optimal number of hidden nodes or the optimal value of the regularization coefficient corresponds to the minimum validation error. Secondly, because intrinsic noise exists in real datasets and because of the limited amount of data, one has to repeat the CV experiment multiple times using different divisions of the data into training and validation sets. CV may become computationally demanding and tedious and regarding for instance the regularization technique, a small range of weight decay coefficients is usually tested. Another critical issue is determination of the relevant input variables. Indeed, too many input variables of which some are irrelevant to estimation of the output, could hamper the model, [9].

The neural network load forecasting literature contains many different approaches and designs, but there is no consensus as to the preferred form. Neural networks will perform poorly if the time series is non stationary. It has a poor performance in forecasting, but this is not the case in all forms of neural networks, because a different specification might have performed better. One reason for the poor performance is the use of forecasted values as inputs for the multi-step predictions. Other reason is that non-linear models can spread errors very dramatically. Also, if the data is not separated into weekday and weekend observations (datasets) and separate models are used for each dataset, it would give again poor forecasting performance. Another reason is that by using less data, results in downgrading the importance of the non-linearity of the model, [1].

2.8.1 Multi-layer feed-forward perception (MLP) with one hidden layer ANN

The system load depends on a number of factors such as, a set of standardized load shapes for each type of day that has been identified as occurring throughout the year, a weather sensitive part of the load, a special event part, which is the occurrence of an unusual or special event leading to a significant deviation from the typical load behavior and a random part, which is an unexplained component.

In the present competitive electricity markets, system load may also be significantly affected by prices. Since electricity must be produced and consumed instantaneously and there are only limited transfer capabilities of transmission systems, electricity prices vary depending on place and time, presenting relatively high variations as compared to other commodities. Hence, price should be factored in the relationship between system load and its influencing factors i.e.

$$L = f(day, weather, special, price, random) \tag{15}$$

The function of those factors is highly nonlinear and is difficult to represent explicitly, and hence forecasting the system load accurately with traditional statistical methods is a rather complex problem.

The Neuron model is made up of a number of simple and highly interconnected Processing Elements (PE), called neuron. Its mathematical model is expressed as

$$O_j = f_j \sum_k (w_{jk}\, x_k) \tag{16}$$

where O_j is the output of neuron; f_j is a transfer function, which is differentiable and non-decreasing, usually represented using a sigmoid function, such as a logistic sigmoid, a tangent sigmoid; w_{jk} is an adjustable weight that represents the connection strength; x_k is the input of a neuron. It has three fully connected layers, an input layer, one hidden layer and an output layer. Signal propagation is allowed only from the input layer to the hidden layer and from the hidden layer to the output layer. Input variables come from historical data corresponding to the factors that affect the load. The outputs are the desired forecasting results, one for each hour of the day. The number of inputs, the number of hidden nodes, transfer functions, scaling schemes and training methods affect the forecasting performance and hence

ISSN 2076-2895 (Print), ISSN 2076-2909 (Online)

need to be chosen carefully. The training of this model basically consists of determining the network parameters such as weights and others, which allow achieving the desired objective based on the available training sets.

2.8.2 Bayesian neural network approach

The Bayesian approach considers a probability density function (pdf) over the weight space. This pdf represents the degrees of belief taken by the different values of the weight vector. This pdf is set initially to some prior distribution once the data have been observed through the use of Bayes' theorem. So instead of the single best set of weights computed by the classical approach of maximum likelihood (through minimization of an error function), Bayesian methods yield a complete distribution can then be used, for instance, to infer predictions of the network for new values of the input variables. The prior pdf is expressed as a Gaussian distribution with a large variance

$$p(\mathbf{W}/a) = \frac{1}{Z_w(a)} e^{(-a \cdot E_w)} \tag{17}$$

where a represents the inverse of the variance on the set of weights and $Z_w(a)$ represents the normalization constant of the pdf and is given by

$$Z_w(a) = \int e^{(-a \cdot E_w)} d\mathbf{w} = (\frac{2\pi}{a})^{m/2} \tag{18}$$

The Bayesian approach to modeling offers significant advantages over the classical ANN learning process. Among others, one can cite automatic tuning of the regularization coefficient using all the available data and selection of the most important input variables through a specific technique called automatic relevance determination (ARD). In addition, reliabilities in the forecast are taken into account as the method computes an error bar on the model output. The latter takes into account in a natural way two contributions: one arises from the intrinsic noise in the data and one arising from uncertainties in the parameter values given by the width of the posterior pdf. The Bayesian method offers a means to select the optimal ANN model (by performing a model comparison). By using a Bayesian approach to control the model complexity the over-fitting problem can be solved. The method is able to deal with model complexity (and therefore with the problem of over-fitting) through the use of the evidence framework and model selection.

2.8.3 Particle Swarm Optimization (PSO) training technique on MLP ANN

The PSO algorithm is a new adaptive algorithm based on a social-psychological metaphor that may be used to find optimal (or near optimal) solutions to numerical and qualitative problems. Most particle swarms are based on two socio-metric principles. Particles fly through the solution space and are influenced by both the best particle in the particle population and the best solution that a current particle has discovered so far. The best particle in the population is typically denoted as "global best", while the best position that has been visited by the current particle is donated as "local best". The "global best" individual conceptually connects all members of the population to one another. That is, each particle is influenced by the very best performance of any member in the entire population. The "local best" individual is conceptually seen as the ability for particles to remember past personal success. The particle swarm optimization makes use of a velocity vector to update the current position of each particle in the swarm. The position of each particle is updated based on the social behavior that a population of individuals adapts to its environment by returning to promising regions that were previously discovered.

PSO is a non-gradient based optimization and search algorithm belonging to probabilistic search algorithms. These algorithms generally mimic some natural phenomena, for example evolutionary algorithms (EA) and simulated annealing. EA (i.e. genetic algorithms (GA), genetic programming (GP), evolutionary programming (EP), evolutionary strategies (ES) and differential evolution (DE)) model the evolution of the species, based on Darwin's principle of survival of the fittest and competition to produce better adapted generations in their problem solution space while simulated annealing is based on the statistical mechanics and models of atoms during an annealing process. The PSO in contrast to the EA, relies on cooperation rather than competition. Moreover good solutions in the problem set are shared

with their less-fit brethren so that the entire population improves and poorly performing members are not killed off as in GA. Although these probabilistic search algorithms generally require many more function evaluations to find an optimal solution, as compared to gradient-based algorithms (i.e. simple gradient descent, conjugate gradient descent), they are promising approaches due to their effectiveness in searching very large spaces and the ability to perform global search for best forecasting model. Moreover, they are generally easy to program, can efficiently make use of large numbers of processors, do not require continuity in the problem definition, and generally are better suited for finding a global, or near global solution. PSO helps in shedding some of the neuron weights by reducing their values by global searching.
Let the ith particle of the swarm be represented by the D-dimensional vector $X_i= (x_{i1}, x_{i2},...., x_{iD})$ and the best particle in the swarm, i.e. the particle with the smallest function value, be denoted by the index g. the best previous position (the position giving the best function value) of the ith particle is recorded and represented as $P_i= (p_{i1}, p_{i2},..., p_{iD})$, and the position change (velocity) of the particle is $V_i= (v_{i1}, v_{i2},..., v_{iD})$. The particles are manipulated according to the equations

$$v_{id} = wv_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \tag{19}$$

$$x_{id} = x_{id} + v_{id} \tag{20}$$

where $d=1, 2, ...D; i=1, 2, ...N$ and N is the size of population; w is the inertia weight; c_1 and c_2 are two positive constants; r_1 and r_2 are two random values in the range [0, 1]. The performance of each particle is measured according to a predefined fitness function, which is problem-dependent. In PSO algorithm the inertia weight is employed to control the impact of the previous history of velocities on the current velocity. In this way, the inertia weight regulates the trade-off between the global (wide-ranging) and local (nearby) exploration abilities of the swarm. A large inertia weight facilitates global exploration (searching new areas); while a small one tends to facilitate local exploration, i.e. fine-tuning the current search area. A suitable value of inertia weight usually provides balance between global and local exploration abilities and consequently a reduction in the number of iterations required to locate the optimum solution. A general rule of thumb suggests that it is better to initially set the inertia to a large value, in order to make better global exploration of the search space and gradually decrease it to get more refined solutions, thus a time decreasing inertia weight value is used, [8].

2.9 Naïve benchmark
The random walk is the most widely used and simplest naïve benchmark method used in forecasting studies. However, for multi-step-ahead forecasting of seasonal data, the method is likely to perform poorly. A more natural benchmark forecast is provided by the seasonal version of the random walk, which takes as a forecast the observed value for the corresponding period in the most recent occurrence of the seasonal cycle. For an hourly data the forecast function is

$$\hat{y}_t(k) = y_{t+k-168} \tag{21}$$

where y_t is demand in period t and k is the forecast lead time ($k \leq 168$), [1].

2.10 Pattern recognition methodology
The set of load curves for each customer is organized into well-defined and separated classes, in order to successfully describe the respective electricity customer's behavior. This allows the selection of adequate tariffs or the successful application of demand side management programs. The main steps of pattern recognition methodology are data and features selection, data pre-processing and main application of pattern recognition methods. The results of the developed methodology can be used for the proper selection of an adequate tariff for the customer or the recommendation of a tariff from the supplier, for the settlement of the customer's bills in the case of energy and power bought from more than one supplier, for the feasibility studies of the energy efficiency and demand side management measures, which are proper for the customer, for the customer's short-term and mid-term load forecasting and for the selection of the representative chronological load diagram of the customer by choosing the type of

typical day (such as the most populated day, the day with the peak demand load or with the maximum demand energy) which is going to be used for the customers' classification by the suppliers, [11].

2.11 Factors influencing short term load forecasting

The system load behavior is influenced by a number of factors. These factors are economic, time, weather and random effects. Economic factors such as the service area demographics, levels of industrial activity, changes in the farming sector, the nature and level of penetration/saturation of the appliance population, developments in the regulatory climate and more generally, economic trends have significant impacts on the system load growth/decline trend. In addition utility-initiated programs, such as changes in rate design and demand management programs also influence the load. It is important to account for these factors in the updating of forecasting models from one year to the next or possibly from one season to another. The economic factors are not, however, explicitly represented in the short term load forecasting models because of the longer time scales associated with them.

Time factors such as seasonal effects, weekly-daily cycle and legal and religious holidays play an important role in influencing load patterns. Certain changes in the load pattern occur gradually in response to seasonal variations such as the number of day-light hours and the changes in temperature. On the other hand, there are seasonal events, which bring about abrupt but important structural modifications in the electricity consumption pattern. They are the shifts to and from Daylight Savings Time, changes in the rate structure (time-of-day of seasonal demand), start of the school year and significant reductions of activities during vacation periods (Christmas-New Year period). The existence of statutory and religious holidays has the general effect of significantly lowering the load values to levels well below 'normal'.

Meteorological conditions are responsible for significant variations in the load pattern. This is because most utilities have large components of weather-sensitive load, such as those due to space heating, air conditioning and agricultural irrigation. In many systems temperature is the most important weather variable in terms of its effects on the load. Past temperatures also affect the load profile. Humidity is a factor that may affect the system load in a manner similar to temperature particularly in hot and humid areas. Thunderstorms also have a strong effect on the load due to the change in temperature that they induce. Other factors that impact on load behavior are wind speed, precipitation and cloud cover/light intensity.

Finally a power system is continuously subject to random disturbances reflecting the fact that the system load is composed of a large number of very small disturbances. There are also certain events such as widespread strikes, shut-down of industrial facilities and special television programs whose effect in the load is uncertain, [3].

3. Mid term load forecasting methods

A methodology that best suits mid term load forecasting is the application of two statistical models, one providing daily and other monthly demand predictions up to 12 months ahead, utilizing primitive (relative humidity) and derived (heating and cooling degree-days) meteorological parameters.

3.1 Autoregressive models

Autoregressive methods reduce serial correlation. According to econometric regression theory, if the residuals are not independent, the estimate of the coefficients may be unstable. It is therefore obvious that the autocorrelation problem can lead to misleading results. One of the common methods to reduce the serial correlation observed in econometric regression is the incorporation of an autoregressive structure in the error term.

There are two multiple regression models that have been developed linking electricity demand directly to climatic conditions as well as seasonal activity patterns. The two models differ on the interval (day or month) used to march forward in time. The models are shown to provide high accuracy forecasts for a number of months ahead, which extend well over a 1-year period, assuming that reasonable weather forecasts are available for this period. In the daily autoregressive model the most important weather parameters that affect electricity consumption are the temperature of the day that electricity demand is projected, the temperature of the two previous days and the relative humidity. In the monthly autoregressive model the predictive power of the model is better compared to that of the daily model because of the greater effectiveness of the autoregressive contribution that stems from the much smaller number (12 than 365) of steps ahead required. However, it should be noted that it models the influence of unusual or extreme weather on electricity consumption. Both models showed that the electricity demand

ISSN 2076-2895 (Print), ISSN 2076-2909 (Online)

compared to the previous year would be increased in the case of extreme meteorological conditions from the case of a moderate meteorological year that is close to the historical average year. The equation of the daily model is

$$\begin{aligned}\log(DE_t) = c + a\,t + k\,HDD_t + l\,CDD_t + k_1\,HDD_{t-1} + l_1\,CDD_{t-1} + \\ k_2\,HDD_{t-2} + l_2\,CDD_{t-2} + n\,HUM_t + \sum_{i=2}^{7} d_i\,D_{it} + \sum_{j=2}^{12} f_j\,M_{jt} + \\ p\,CH_t + s\,CH_{t-1} + q\,PP_t + r\,VH_t = e_t\end{aligned} \quad (22)$$

where $c, a, k, l, k_1, l_1, k_2, l_2, n, d_i$ (i from 2 to 7), f_j (j from 2 to 12), p, s, q and r are the coefficients to be estimated from the regression analysis and e_t is the residual term. HDD_t and CDD_t are the heating and cooling degree-days of the corresponding day t that electricity demand is assumed,

$$HDD_t = \max\,(T_{ref} - T_t, 0) \quad (23)$$

$$CDD_t = \max\,(T_t - T_{ref}, 0) \quad (24)$$

where the weighted average temperature for the day t is T_t and T_{ref} is a reference temperature that should be adequately selected to separate the heat and cold branches of the demand-temperature relationship. HDD_{t-1}, HDD_{t-2}, CDD_{t-1} and CDD_{t-2} are the heating and cooling degree-days of the two previous days of the reference day t. HUM_t is the average relative humidity for the reference day t. D_{it} are six dummy variables aiming at representing the significant daily variability of the electricity demand. CH_t, CH_{t-1}, PP_t and VH_t are four dummy variables that cover anomalous events in electricity demand related to holidays or days near a holiday. M_{jt} are eleven dummy variables in the model account for the monthly seasonality of electricity demand not related to the weather conditions.

The equation of the monthly model is

$$\log(ME_t) = c + a\,t + k\,MHDD_t + l\,MCDD_t + \sum_{j=2}^{12} f_j\,M_{jt} + p\,H_t + e_t \quad (25)$$

where c, a, k, l, f_j (j=2,…,12) and p are the coefficients to be estimated from the regression analysis and e_t is the residual term. $MHDD_t$ and $MCDD_t$ are the total number of monthly heating and cooling degree-days estimated for the month t.

$$MHDD_t = \max(\sum_{i=1}^{m} HDD_{t,i} - \sum_{i=1}^{m} CDD_{t,i}, 0) \quad (26)$$

$$MCDD_t = \max(\sum_{i=1}^{m} CDD_{t,i} - \sum_{i=1}^{m} HDD_{t,i}, 0) \quad (27)$$

where m is the number of days of the month t. M_{jt} are eleven dummy variables, which are aimed at representing the monthly seasonality of electricity demand and are not related to the weather conditions. H_t represents the number of non-working days (Sundays, Saturdays and holidays) during the month t, [12].

3.2 Abductive networks

Abductive network models select effective inputs and can be simpler than neural network models, which can improve interpretability, give better insight into the modeled system and allow the derivation of analytical model relationships. Optimum subsets of model inputs selected automatically through abductive modeling can be used to reduce data dimensionality and therefore improve the forecasting performance of neural network models. Dimensionality reduction is particularly important for reducing model over-fitting and improving generalization in applications with a small number of training examples, which is the case in many medium and long-term demand forecasting applications. Below are presented two special cases of Abductive Networks.

3.2.1 Abductory inductive mechanism (AIM)

AIM is a supervised inductive machine-learning tool for automatically synthesizing abductive network models from a database of inputs and outputs representing a training set of solved examples. The tool can automatically synthesize adequate models that embody the inherent structure of complex and highly nonlinear systems. The automation of model synthesis not only lessens the burden on the analyst but also safeguards the model generated from being influenced by human biases and misjudgments.

AIM uses the predicted squared error (PSE) criterion for selection and stopping to avoid model over-fitting, thus eliminating the problem of determining when to stop training in neural networks. AIM expresses the *PSE* error as

$$PSE = FSE + CPM\,(2K/n)\,\sigma_p^2 \tag{28}$$

where *FSE* is the fitting error in the training data, *CPM* is a complexity penalty multiplier selected by the user, K is the number of model coefficients, n is the number of samples in the training set and σ_p^2 is a prior estimate for the variance of the error obtained with the unknown model. This estimate does not depend on the model being evaluated and is usually taken as half the variance of the dependent variable y. AIM builds networks consisting of various types of polynomial functional elements such as a white element which consists of a constant plus the linear weighted sum of all outputs of the previous layer, i.e.

$$White\,Output = w_0 + w_1 x_1 + w_2 x_2 + ... + w_n x_n \tag{29}$$

where $x_1, x_2....x_n$ are the inputs to the element and $w_1, w_2....w_n$ are the element weights. Single, double and triple elements which implement a third-degree polynomial expression with all possible cross-terms for one, two and three inputs, respectively;

$$Double\,Output = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + w_6 x_1^3 + w_7 x_2^3 \tag{30}$$

The network size, element types, connectivity and coefficients for the optimum model are automatically determined using well-proven optimization criteria, thus reducing the need for user intervention compared to neural networks. This simplifies model development and reduces the learning/development time and effort. The model takes the form of layered feed-forward abductive networks of functional elements (nodes).

3.2.2 Group method of data handling learning algorithm (GMDH)

As AIM, the tool can automatically synthesize adequate models that embody the inherent structure of complex and highly nonlinear systems. The GMDH approach is a formalized paradigm for iterated (multi-phase) polynomial regression capable of producing a high-degree polynomial model in effective predictors. The process is evolutionary in nature, using initially simple (myopic) regression relationships to derive more accurate representations in the next iteration. To prevent exponential growth and limit model complexity, the algorithm selects only relationships having good predicting powers within each phase. Iteration is stopped when the new generation regression equations start to have poorer prediction performance than those of the previous generation, at which point the model starts to become overspecialized and therefore unlikely to perform well with new data. The algorithm has three main elements: representation, selection and stopping. It applies abduction heuristics for making decisions concerning some or all of these three aspects.

To illustrate these steps for the classical GMDH approach, consider an estimation of n_e observations (rows) and $m+1$ columns for m independent variables ($x_1, x_2....x_m$) and one dependent variable y. in the first iteration is assumed that the predictors are the actual input variables. The initial rough prediction equations are derived by taking each pair of input variables (x_i, x_j; I, j=1, 2….m) together with the output y and computing the quadratic regression polynomial

$$y = A + B\,x_i + C\,x_j + D\,x_i^2 + E\,x_j^2 + F\,x_i\,x_j \qquad (31)$$

Each of the resulting $m(m+1)/2$ polynomials is evaluated using data for the pair of x variables used to generate it, thus producing new estimation variables (z_1, z_2....$z_{m(m-1)/2}$) which would be expected to describe y better than the original variables the resulting variables are screened according to some selection criterion and only those having good predicting power are kept. The original GDMH algorithm employs and additional and independent selection set of n_s observations for this purpose and uses the regularity selection criterion based on the root mean squared error r_k over the data set, where

$$r_k^2 = \sum_{l=1}^{n_s} (y_l - z_{kl})^2 / \sum_{l=1}^{n_s} y_l^2, \quad k = 1,2,....,m(m-1)/2 \qquad (32)$$

The method offers the advantages of improved forecasting performance, faster model development requiring little or no user intervention, faster convergence during model synthesis without the problems of getting stuck in local minima, automatic selection of relevant input variables and automatic configuration of model structures. With the resulting model represented as a hierarchy of polynomial expressions, analytical model relationships can be derived. Such relationships provide insight into the modeled phenomena, highlight contributions of various inputs and allow comparison with previously used empirical or statistical models. The technique automatically avoids over-fitting by using a proven regularization criterion based on penalizing model complexity without requiring a dedicated validation data set during training, as in the case with many neural network paradigms, [10].

3.3 Factors influencing mid term load forecasting

Mid-term (i.e. monthly and yearly) electricity demand forecasting in power systems is a complicated task because it is affected directly or indirectly by various factors primarily associated with the economy and the weather. The weather elements which influence electricity demand, consists of temperature, humidity, wind and precipitation in a decreasing order of importance. The variation of electricity demand with temperature is non-linear, increasing both for decreasing and increasing temperatures (reflecting mainly the use of electric heating appliances in winter and air conditioners in summer). The power system seems to be less sensitive to temperature fluctuations in winter, since a fall in average temperature of 1oC would result in an increase in electricity consumption but less than that of the case of 1^oC increase in average temperature in the summer. This is mainly attributed to the fact that final consumers can use a variety of energy sources for heating (e.g. diesel oil, natural gas, electricity) and practically only electricity for cooling. In mid-term load forecasting the errors that occur from autocorrelation (misleading results) are reduced by the use of a smaller number of steps ahead in the autoregressive model, [12].

4. Long term load forecasting methods

The ideal method of long-term forecasting would be one that could find non-linear relations between load and various economic and other factors and is adaptable to changes. A methodology that best suits these requirements is the application of Artificial Neural Networks (ANN).

4.1 Recurrent neural networks

A more direct method of including temporal information is the recurrent neural networks (RNN). RNN contains feedback connections, which are enable to encode temporal context internally. It has the ability to learn patterns from the past records and also to generalize and project the future load patterns for an unseen data. It can describe the evolution of dynamical systems. Different types of RNNs have been developed such as Jordan RNN which has feedback connections from output to input and Elman RNN which has feedback connections from its hidden layer neurons back to its inputs.

Additional neurons in the input layer, which accept these feedback connections, are called state or context neurons. Their role is to get inputs from the upper layer and after processing, send their outputs to the hidden layer together with other plan inputs. As the period of target forecast loads becomes longer, the forecast errors might increase relatively. The context neurons are used only to memorize some past states of the hidden neurons and so the outputs of the networks depend on an aggregate of the previous states and the current input. These context neurons are considered to function as one step time delays. In Jordan RNN, each context neuron on the input side not only receives input from the corresponding output neuron, but also has a feed-back from its own immediate value. It should be noted that feed-forward networks are faster than feed-back nets, because you can set a solution with only one pass. In addition, feed-forward nets are guaranteed to reach stability. On the other hand, feed-back networks must iterate over many cycles, until the system stabilizes. Even then, the system may not be stable. The capacity of networks with feed-back is more limited, too. RNN can do things that other networks cannot do. Feed-back loops permit trainability and adaptability. Linear networks are limited; recurrence provides necessary non-linearity, [5].

4.2 Back-propagation neural networks

Back-propagation neural networks (BP) is one of the most widely used neural network paradigms. The BP training algorithm is an iterative gradient descent algorithm. BP can be applied to any problem that requires pattern mapping. Given an input pattern, the network produces an associated output pattern. Its learning and update procedure is intuitively appealing, because it is based on a relative simple concept: the network is supplied with both a set of patterns to be learned and the desired system response for each pattern. If the network gives the wrong answer, then the weights are corrected so that the error is lessened and as a result future responses of the network are more likely to be correct.

The advantages of using such a network center on some of their properties, too. Firstly, they automatically generalize their knowledge enabling them to recognize patterns, which they have seen. Secondly, they are robust enough to recognize patterns, which have been obscured by noise. Lastly, once they have been trained on the initial set of patterns, their recognition of similar patterns is accomplished very quickly. BP training has also two more key advantages for applications, which other network paradigms do not process. The first advantage is that BP training is mathematically designed to minimize the mean squared aggregate error between the actual output of a multilayer feed-forward perception and the desired output and updates the weights by moving them along the gradient-descendent direction. This process is made up of two passes through the network layers, one forward and one backward. In the forward pass, the input is applied to the sensory nodes and its effect propagates through the network layers, the response then appears at the output nodes. This output is compared against a desired value, to produce an error signal which then propagates backward through the network. In the forward pass, the synaptic weights of the network are fixed, but during the backward pass they are adjusted so that the network output moves closer towards the desired response. The other advantage is that it is a supervised training technique. This means that the network designer can dictate the exact results he wants the network to achieve and the network's performance can always be measured against those results. Supervised training is predictable and easy to use.

The neural network predictors trained by BP are prone to over-fitting of the model because of the large number of parameters that must be estimated. This is due to two reasons: (1) because the model is over-trained; (2) or because it is too complex, [5, 8].

4.3 Econometric approach

This method is usually preferred for long term forecasts. This approach combines economic theory with statistical methods to produce a system of equations for forecasting energy demand. The dependant variable in our case, the demand for electricity, is expressed as a function of various economic factors. These variables could be population, income per capita or value added or output (in industry or commercial sectors), price of power, price of alternative fuels (that could be used as substitutes), proxies for penetration of appliances and equipment (capture technology effect in case of industries etc). Thus, one would have

$$ED = f(Y, P_i, P_j, POP, T) \tag{33}$$

where, ED is the electricity demand; Y is the output or income; P_i is the own price; P_j is the price of related fuels; POP is the population and T is the technology. The econometric methods require a consistent set of information over a reasonably long duration. This requirement forms a pre-requisite for establishing both short term and long term relationships between the variables involved. Thus, for instance, if one was interested in knowing the price elasticity of demand, it is hard to arrive at any meaningful estimates, given the long period of administered tariffs and supply bottlenecks. However the price effect will have an important role to play in the years to come. In such a case one may have to broaden the set of explanatory variables apart from relying on more rigorous econometric techniques to get around the problem. Another criticism of this method is that during the process of forecasting it is incorrect to assume a particular growth rate for the explanatory variables. Further the approach fails to incorporate or capture in any way the role of certain policy measures/economic shocks that might otherwise result in a change in the behavior of the variable being explained. This would have to be build into the model maybe in the form of structural changes, [4].

4.4 Fuel share model

It is a variant of econometric models and it considers a two-step approach for estimating energy demands. First the total energy consumption by a sector is estimated, which is then used in the determination of fuel shares, defined as ratios of individual fuels consumed to the total energy consumption by the sector.

A drawback of this method is its failure to recognize the interdependence between prices and quantity. The estimating equations assume that fuel prices are determined independently on both total energy consumption and the distribution of consumption by fuels. The sequential estimation procedure also assumes that total energy consumption is independent of fuel shares. Thus all fuel supplies must be perfectly elastic and price elasticity is meaningful only if total energy demand remains fixed in response to a change in relative prices. Also, the aggregate energy quantities and prices are weighted averages of individual fuels expressed in common heat units, which is acceptable only if all fuels are substitutable in different applications. The weights are unaffected by relative prices making the aggregation procedure inconsistent with the premise that fuel shares shift in response to relative fuel prices.

All the models built using this approach have predicted that electricity demand is highly price-responsive. Also energy prices are important in determining both total energy consumption and the fuel choice while income is more important in determining total energy demand, than the fuel choice. Thus the expectations based on demand theory that relative fuel prices, not income affect fuel choices, while both determine the energy consumption levels, [4].

4.5 Co-integration

This method is a variant of the time-series approach and it attempts to overcome some of the limitations of the simple econometric forecasts wherein a growth rate to the explanatory economic factors is prescribed. The overall pattern or relationship between any set of variables is likely to persist into the future as well. It is observed that some economic variables tend to behave in a similar fashion in the long run. In such a case, it is often found that these factors have significant causal effects on each other. The long run (common) equation capturing the relationship between the variables involved is called the co-integrating vector. Various software packages have now been developed that can establish such relationships with a fair degree of simplicity and then utilize them for arriving at future projections. We use a system of equations to build the model, compared to just one equation in the case of a simple econometric or time-series model. In addition, we also include an additional term called the error correction term, to account for the long run effects, while the short run effects are captured by the co-integrating vector.

The advantage of this technique is that one does not need to prescribe the growth rates of any of the variables that are co-integrated with one another. The system of equations internally generates the forecasted values of the variables involved, based on the long run pattern established in the past. In addition, introducing shocks into the system could capture the effect of policy implications. The major disadvantage of this approach is the need for a consistent time-series spanning at least 30-time-period as a prerequisite, [4].

4.6 Factors influencing long term load forecasting

Unlike short-term forecasting, long-term forecasting is mainly affected by economical factors rather than weather conditions. These economical factors are complex factors that have non-linear characteristics and good results may not be obtained using traditional methods. The economical factors that influence

the long-term forecasting and are used as inputs to the networks are gross national product, gross domestic product, population, number of households, number of air-conditioners, amount of CO_2 pollution, index of industrial production (IIP), oil price, amount of energy consumption, electricity price, average temperature and the maximum electric power of the previous year.
The level of influences of selected inputs on the output is determined by the contribution factor, which is the sum of the absolute values of the weights leading from one particular variable. The higher the number, the more the variable is contributing to the prediction. The contribution factor can be used to decide, which variable can be removed in order to simplify the network and make the training faster.
The forecast errors that occur in this forecasting might increase as the period of target forecast loads becomes longer, [5].

5. Electricity price forecasting methods

Electricity price forecasting is quite challenging. It has been approached using techniques such as time series models, unit commitment models, neural networks, structural models, machine learning models and hybrid models. The selection of the best forecasting technique depends on factors such as product (spot/forward price), term (day-ahead, month to month), and market design (single, multiple settlement system).
Up to now, different techniques consist of statistical and non-statistical methods for future price forecasting in which former are more frequently used than the latter. Time series and artificial intelligence methods are two main categories of the statistical methods, [13].

5.1 Unit commitment modeling and approaches

Unlike other commodities, electricity cannot be stored. Therefore, a delicate balance must be maintained between generation and customer load demands; 24 hours a day, 7 days a week, 8760 hours a year. Customer load demands in electric power systems are subject to changes because human activities follow daily, weekly and monthly cycles. The load demands are generally higher during the daytime and early evenings when industrial loads are high, lights are on and so forth, and lower during late evenings and early mornings when most of the population is asleep. It is, therefore, required to commit enough generating units to meet the load demands in electric power systems. If we commit enough generating units to meet the peak load demand and keep these units ON at all times, we provide a brute force solution to generating unit schedule. However, turning units OFF when they are not needed can possibly save a great deal of money. We need to satisfy load demands while operating the power system economically.
Unit commitment is the process of determining optimal schedule of generating units over a set of study period subject to unit and system operating constraints.
The past years has seen a dramatic change in the manner in which the power industry is organized. It has moved from a formerly vertically integrated and highly regulated industry to one that has been horizontally integrated in which the generation, transmission and distribution are unbundled. The design of electricity markets has been the subject of massive debate since proposed market models have ranged from highly centralized and controlled to decentralized and relatively bilateral models.
The purpose of this note is to present the two different unit commitment approaches currently used by transmission system operators and power producers for electricity pricing. These are known as the security constrained unit commitment approach and the price-based unit commitment approach.

5.1.1 The security constrained unit commitment approach

Prior to deregulation, in a vertically integrated utility environment, a utility system operator, who had the knowledge of system components, constraints and operating costs of generating units, was the one who made decisions on unit commitment for minimizing the utility's generation cost. This approach is known as the security constrained unit commitment (SCUC) and determines the generating unit schedules in a utility for minimizing the operating cost and satisfying the prevailing constraints, such as load balance, system spinning reserve, ramp rate limits, fuel constraints, multiple emission requirements and minimum up and down time limits, over a set of time periods. Three elements are included in the SCUC approach such as satisfying load demand, maximizing security and minimizing cost. Satisfying load demand is a hard constraint and an obligation for SCUC. Maximizing security could often be satisfied by supplying sufficient spinning reserve at less congested regions, which could easily be accessible by loads. Cost

minimization is realized by committing less expensive units while satisfying the corresponding constraints and dispatching the committed units economically.
The SCUC is formulated as an optimization problem, which minimizes power system operating cost, c, using the functional

$$c = \min \sum_i \sum_t \left[I_{(i,t)} C_i \left(P_{(i,t)} \right) + S_{(i,t)} \right] \tag{34}$$

where $I_{(i,t)}$ is the commitment status of unit i at time t (0=OFF, 1=ON), $C_i(\,)$ is a quadratic cost function of unit i, $P_{(i,t)}$ is the generation of unit i at time t and $S_{(i,t)}$ is the start-up cost of unit i at time t. Functional (34) is subject to the following system and unit constraints:

(a) System demand constraints (system real power balance):

$$P_{D(t)} = \sum_i I_{(i,t)} P_{(i,t)} \tag{35}$$

where $P_{D(t)}$ is the system total real power load demand at time period t.

(b) System spinning reserve constraints:

$$R_{S(t)} \le \sum_i r_{s(i,t)} I_{(i,t)} \tag{36}$$

where $R_{S(t)}$ is the spinning reserve requirement at time t and

$$r_{s(i,t)} = \min \left[10 MSR_i , P_{g,\max(i)} - P_{(i,t)} \right] \tag{37}$$

where MSR_i is the maximum sustain ramp rate of unit i (in MW/min) and $P_{g,max(i)}$ is the maximum capacity of unit i.

(c) System operating reserve constraints:

$$R_{O(t)} \le \sum_i r_{o(i,t)} I_{(i,t)} \tag{38}$$

where $R_{O(t)}$ is the operating reserve requirement at time t and

$$r_{o(i,t)} = \begin{cases} q_i \text{, if unit } i \text{ is OFF} \\ r_{s(i,t)} \text{, if unit } i \text{ is ON} \end{cases} \tag{39}$$

where q_i is the quick start capability of unit i.

(d) Unit capacity constraints:

$$P_{g,\min(i)} \le P_{(i,t)} \le P_{g,\max(i)} \tag{40}$$

where $P_{g,min(i)}$ is the minimum capacity of unit i.

(e) Unit start-up / shut-down times constraints:

$$\left[X_{(i,t-1)}^{ON} - T_{(i,t)}^{ON}\right]\left[I_{(i,t-1)} - I_{(i,t)}\right] \geq 0$$
$$\left[X_{(i,t-1)}^{OFF} - T_{(i,t)}^{OFF}\right]\left[I_{(i,t-1)} - I_{(i,t)}\right] \geq 0 \qquad (41)$$

where $X_{(i,t)}$ is the time duration for which unit i has been ON or OFF at time t and $T_{(i)}$ is the minimum ON or OFF time of unit i.

(f) Unit ramp rate limits constraints:

$$P_{(i,t)} - P_{(i,t-1)} \leq UR_{(i,t)}, \text{ as unit } i \text{ ramps up}$$
$$P_{(i,t-1)} - P_{(i,t)} \leq DR_{(i,t)}, \text{ as unit } i \text{ ramps down} \qquad (42)$$

(g) System and unit fuel constraints:

$$F_{\min(i)} \leq \sum_i C_{fi}\left[P_{(i,t)} I_{(i,t)}\right] + S_{f(i,t)} \leq F_{\max(i)} \qquad (43)$$

where $F_{min(i)}$ is the minimum total fuel consumption of unit i, $F_{max(i)}$ is the maximum total fuel consumption of unit i, $C_{fi(\,)}$ is the fuel consumption function of unit i and $S_{f(i,t)}$ is the start-up fuel of unit i at time t. Also,

$$F_{\min(FT)} \leq \sum_{i \in FT}\sum_t C_{fi}\left[P_{(i,t)} I_{(i,t)}\right] + S_{f(i,t)} \leq F_{\max(FT)} \qquad (44)$$

where $F_{min(FT)}$ is the minimum total fuel consumption for fuel type FT and $F_{max(FT)}$ is the maximum total fuel consumption for fuel type FT.

(h) System emissions constraints:

$$\sum_i\sum_t C_{ei}\left[P_{(i,t)} I_{(i,t)}\right] + S_{e(i,t)} \leq E_{\max} \qquad (45)$$

where $C_{ei(\,)}$ is the emission function of unit i, $S_{e(i,t)}$ is the start-up emission of unit i at time t and E_{max} is the maximum total emission allowance.

(i) Transmission flow limits from bus k to bus m

$$-P_{km}^{\max} \leq P_{km(t)} = f\left[\mathbf{B}_{(t)}, \varphi_{(t)}\right] \leq P_{km}^{\max} \qquad (46)$$

where $\boldsymbol{B}_{(t)}$ is the real power vector, and $\varphi_{(t)}$ is the phase shifter control vector.

5.1.2 The price-based unit commitment approach

In some deregulated markets the unit commitment is the responsibility of the individual power producers in order to maximize their own profit. Such decision to commit generating units is associated with financial risks. This unit commitment has a different objective than that of SCUC and is referred to as price-based unit commitment (PBUC) to emphasize the importance of price signal. In PBUC, satisfying load is no longer an obligation and the objective would be to maximize the profit while security is now unbundled from energy and priced as ancillary service. In this approach, the signal that would enforce a unit's ON or OFF status would be the price, including the fuel purchase price, energy sale price, ancillary service sale price, etc. Each energy supplier would be responsible for its own decision on what and how to bid on energy to supply load and reserves markets. In this approach, bidders bear all the risks of their decisions for committing their units and bidding strategies. Power producers run their own price-based unit commitment where the objective is to maximize individual power generator profit, regardless of the system wide profit.

The objective of PBUC is to maximize the profit (i.e., revenue minus cost) subject to all prevailing constraints. For unit i at time t, the profit is given as

$$\begin{aligned} F_{(i,t)} = & \big[\rho_{gm(t)}\big(P_{(i,t)} - P_{o(i,t)}\big) + \rho_{rm(t)}R_{(i,t)} + \rho_{nm(t)}N_{(i,t)} - C_i\big(P_{(i,t)} + R_{(i,t)} + N_{(i,t)}\big) \\ & - S_{(i,t)} + f_i\big(P_{o(i,t)}\big)\big]\, I_{(i,t)} && \text{Part A} \\ & + \big[\rho_{nm(t)}N_{(i,t)} - \rho_{gm(t)}B_{(i,t)} - C_i\big(N_{(i,t)}\big) + f_i\big(P_{o(i,t)}\big)\big]\big(1 - I_{(i,t)}\big) && \text{Part B} \end{aligned} \tag{47}$$

where $F_{(i,t)}$ is the profit of unit i at time t, $\rho_{gm(t)}$ is the forecasted market price for energy at time t, $P_{o(i,t)}$ is the bilateral contract of unit i at time t, $\rho_{rm(t)}$ is the forecasted market price for spinning reserve at time t, $R_{(i,t)}$ is the spinning reserve of unit i at time t, $\rho_{nm(t)}$ is the forecasted market for non-spinning reserve at time t, $N_{(i,t)}$ is the non-spinning reserve of unit i at time t, $f_i(P_{o(i,t)})$ is the profit from bilateral contract of unit i at time t and $B_{(i,t)}$ is the power purchase of unit i at time t.

The first part of (47) represents the profit when the unit is ON. Profit is defined as the revenue from the sales of energy and ancillary services minus production costs. The profit from bilateral contracts can also be included, though it is assumed to be constant. The second part of (47) represents the profit when the unit is OFF. Here, profit represents the revenue from non-spinning reserve sales minus production costs as well as the cost of any energy purchases.

The PBUC problem is formulated as an optimization problem, which maximizes power producer profit using the functional

$$\max \sum_i \sum_t F_{(i,t)} \tag{48}$$

subject to the following system and unit constraints:

(a) System energy and reserve constraints:

$$P_{\min(t)} \le \sum_i P_{(i,t)} I_{(i,t)} \le P_{\max(t)} \tag{49}$$

$$R_{\min(t)} \le \sum_i R_{(i,t)} I_{(i,t)} \le R_{\max(t)} \tag{50}$$

$$N_{\min(t)} \le \sum_i N_{(i,t)} \le N_{\max(t)} \tag{51}$$

where $P_{min(t)}$ is the minimum total generation at time t, $P_{max(t)}$ is the maximum total generation at time t, $R_{min(t)}$ is the minimum total spinning reserve at time t, $R_{max(t)}$ is the maximum total spinning reserve at time t, $N_{min(t)}$ is the minimum total non-spinning reserve at time t and $N_{max(t)}$ is the maximum total non-spinning reserve at time t.

(b) System emissions constraints:

$$\sum_i \sum_t C_{ei}\big[\big(P_{(i,t)} + R_{(i,t)}\big) I_{(i,t)} + N_{(i,t)}\big] + S_{e(i,t)} \le E_{\max} \tag{52}$$

(c) System fuel constraints:

$$F_{\min(FT)} \leq \sum_{i \in FT} \sum_{t} C_{fi} \left[\left(P_{(i,t)} + R_{(i,t)} \right) I_{(i,t)} + N_{(i,t)} \right] + S_{f(i,t)} \leq F_{\max(FT)} \tag{53}$$

(d) Unit generation constraints:

$$P_{(i,t)} I_{(i,t)} + B_{(i,t)} \geq P_{o(i,t)} \tag{54}$$

$$0 \leq B_{(i,t)} \geq P_{o(i,t)} \tag{55}$$

$$P_{g,\min(i)} \leq P_{(i,t)} I_{(i,t)} + R_{(i,t)} I_{(i,t)} + N_{(i,t)} \leq P_{g,\max(i)} \tag{56}$$

$$R_{(i,t)} I_{(i,t)} \leq r_{s(i,t)} I_{(i,t)} \tag{57}$$

$$N_{(i,t)} \leq n_{o(i,t)} \tag{58}$$

where,

$$n_{o(i,t)} = \begin{cases} q_i \text{, if unit } i \text{ is OFF} \\ n_{s(i,t)} \text{, if unit } i \text{ is ON} \end{cases} \tag{59}$$

and

$$r_{s(i,t)} = \min \left[10 MSR_i , P_{g,\max(i)} - P_{(i,t)} \right] \tag{60}$$

$$n_{s(i,t)} = \min \left[10 MSR_i , P_{g,\max(i)} - P_{(i,t)} - R(i,t) \right] \tag{61}$$

(e) Unit start-up / shut-down times constraints

$$\begin{aligned} &\left[X_{(i,t-1)}^{ON} - T_{(i,t)}^{ON} \right] \left[I_{(i,t-1)} - I_{(i,t)} \right] \geq 0 \\ &\left[X_{(i,t-1)}^{OFF} - T_{(i,t)}^{OFF} \right] \left[I_{(i,t-1)} - I_{(i,t)} \right] \geq 0 \end{aligned} \tag{62}$$

(f) Unit ramp rate limits constraints

$$\begin{aligned} &P_{(i,t)} - P_{(i,t-1)} \leq UR_{(i,t)} \text{, as unit } i \text{ ramps up} \\ &P_{(i,t-1)} - P_{(i,t)} \leq DR_{(i,t)} \text{, as unit } i \text{ ramps down} \end{aligned} \tag{63}$$

(g) Unit fuel constraints:

$$F_{\min(i)} \leq \sum_{i} C_{fi} \left[\left(P_{(i,t)} + R_{(i,t)} \right) I_{(i,t)} + N_{(i,t)} \right] + S_{f(i,t)} \leq F_{\max(i)} \tag{64}$$

5.1.3 Discussion on unit commitment approaches

The past years has seen a dramatic change in the manner in which the power industry is organized. It has moved from a formerly vertically integrated and highly regulated industry to one that has been horizontally integrated in which the generation, transmission and distribution are unbundled. Market structures differ according to their participants, the amount of information that participants share with the transmission system operator (TSO), and the role of TSO in facilitating or controlling the individual markets. The design of these markets has been the subject of massive debate. Proposed market models have ranged from highly centralized and controlled PoolCo model to decentralized and relatively bilateral models. In some of the deregulated markets such as New York Market, the TSO will commit generating units in much the same way as system operators did in the vertically integrated structure using the SCUC.

In the SCUC approach, suppliers submit their bids to supply the forecasted daily inelastic demand. The TSO uses the bid-in costs submitted by power producers for each generating unit (multi-part bids that reflect start-up costs, minimum generation costs, running costs, etc.) to minimize the cost of operation, determine which units will be dispatched in how many hours and calculate the corresponding market clearing prices while the system security is retained.

In some other deregulated markets, such as New England Power Pool, California Market, New Zealand, and Australia, the unit commitment is the responsibility of the individual power producers in order to maximize their own profit. In the PBUC, satisfying load demand is no longer an obligation and the objective would be to maximize the profit while security is now unbundled from energy and priced as ancillary service. In this approach, the signal that would enforce a unit's ON or OFF status would be the price, including the fuel purchase price, energy sale price, ancillary service sale price, etc. Each power producer would be responsible for its own decision on what and how to bid on energy to supply load and reserves markets. In this approach, bidders bear all the risks of their decisions for committing their units and bidding strategies. Power producers run their own price-based unit commitment where the objective is to maximize individual power producer profit, regardless of the system wide profit.

In the PBUC approach, power producers submit single-part bids to the TSO for minimizing the risk of not knowing the number of hours that they would be dispatched or the market clearing prices that would be paid on during dispatched hours. The TSO uses power producers' single-part bids, aggregates them and determines the market clearing prices. Here, the TSO's objective will be to maintain the system security, while the power producers' objective will be to maximize its own profit.

In comparing SCUC with PBUC, it may assume that maximizing the profit is essentially the same as minimizing the cost. However, profit is not the negative of cost, rather it is defined as the revenue minus cost. That is, profit would not only depend on cost, but on revenue as well. If the incremental revenue is larger than the incremental cost, we may generate more energy for attaining more profit. On the contrary, if the incremental revenue is smaller than the incremental cost, it may be less attractive to sell energy. In an extreme case, if the objective in the PBUC would be to minimize the cost, a power producer might not opt to generate since it would have no obligation to serve a load with zero cost.

In the SCUC, demand forecast advised the system operators of the amount of power to be generated. However, in the PBUC, bilateral contracts will make part of the system demand known a priori and the remaining will be forecasted. The most distinct feature of PBUC approach is that all market information are reflected in market price. Although, the system load is not a hard constraint in PBUC, load forecasting would be required for market price forecasting. Likewise, security would not be a consideration in formulating PBUC, however, the TSO's criteria for maintaining security would impact the market price. It may be difficult for individual participants in an energy market to foresee other market participants' bidding strategies.

5.2 Takagi-Sugeno-Kang (TSK) fuzzy inference system

It is an ARMAX regression model with fuzzy logic based parameters for short-term forecasting (day ahead). The ARMAX's parameters are fuzzified using TSK fuzzy interference system. Two sets of price forecasts are computed: (1) forecasts assuming perfect forecasts of explanatory variables (e.g. load, gas price) are available; and (2) forecasts with naïve projections of explanatory variables: tomorrow's value equals to today's value. The overall output F_k of the TSK fuzzy interference system for each observation k in the training set is computed by taking the weighted average of the ARMAX models associated with each unique combination of membership functions (fuzzy region)

$$F_k = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{m=1}^{3} \mu_{p_i}(P_k)\,\mu_{L_j}(L_k)\,\mu_{I_m}(I_k)\,F(k)_{i,j,m}}{\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{m=1}^{3} \mu_{p_i}(P_k)\,\mu_{L_j}(L_k)\,\mu_{I_m}(I_k)} \tag{65}$$

Fuzzy logic simulates the way humans think as opposed to the way that computers typically work. In contrast to computers, humans get imprecise and incomplete information from the environment and through common-sense rules transform that information into an action. So although humans think in fuzzy, non crisp ways, their final actions are crisp. The process of translating the results of fuzzy reasoning to a crisp, non-fuzzy action is called defuzzification. Fuzzy logic is a problem solving system that provides a way to arrive at a definite conclusion (crisp) based on imprecise and incomplete information (fuzzy). In fuzzy logic, a fuzzy subset T of a set S is defined by a membership function which gives the degree of membership of each element of S belonging to T. A membership function defines how each input is mapped to a degree of membership between 0 and 1 to a given fuzzy input subset T. The primary difference between fuzzy logic and probability theory arises from the fact that membership functions are subjective and non-random.
Fuzzy reasoning is simply a set of if-then statements that take the inputs as condition and outputs as consequence. The fuzzy rules use the input membership values as weighting factors to determine their influence on the fuzzy output sets of the final output conclusion. A fuzzy inference system implements a non-linear mapping from an input space to output space by using a set of fuzzy if-then rules. The TSK provides a systematic approach for generating fuzzy rules from a given input-output data.
The TSK calculates subsets of explanatory variable coefficients. The universe is divided into clusters and each cluster contains a subset of coefficients. The electricity price is calculated by matching the input (explanatory variable values) with the most similar cluster and then applying the coefficient specific to that cluster to forecast the electricity price. The TSK model captures the dynamic and complex relationship between demand and supply because it uses coefficients calculated specific to each cluster.
Root mean square error (RMSE) is used to measure the predicting power of the model,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{66}$$

where N is the number of instances in the testing set; y_i is the actual value and $\hat{y}_i$ is the predicted value for the *ith* sample, [13].

5.3 Artificial neural networks

The training method for artificial neural nets is based on making a previous selection for the multilayer perception (MLP) training samples, using an ART-type neural network. The most common learning algorithm is the back propagation algorithm, in which the input is passed layer through layer until the final output is calculated and it is compared to real output to find the error. The error is then propagated back to the input adjusting the weights and biases in each layer. The standard back propagation learning algorithm is a steepest descent algorithm that minimizes the sum of square errors.
However, the standard back propagation learning algorithm is not efficient numerically and tends to converge slowly. An algorithm that trains a neural network 10-100 times faster than the usual back propagation algorithm is the Leverberg-Marquardt algorithm which is an approximation to Newton's method,

$$\Delta x = [J^T(x)J(x) + \mu I]^{-1} J^T(x)e(x) \tag{67}$$

where J is the Jacobian matrix, e is the error matrix and I is the unity matrix. The parameter μ is multiplied by some factor *(β)* whenever a step would result in an increased *V(x)* (a cost function which should be minimized). When a step reduces *V(x)*, μ is divided by β. When μ is large the algorithm becomes steepest descent and convergence is improved, but the learning time is increased; while for small μ the algorithm becomes Gauss-Newton and learning time is reduced, but error goal is also decreased.
The method can be classified into three steps: data preparation and sensitivity analysis, ANN training and simulation and accuracy assessment.
In ANN input preparation and sensitivity analysis, different sets of lagged prices have been proposed as input features for the price forecasting in different markets. Two different methods are used to calculate

24 hourly price forecasts for each day subject to analysis, 24 "one-step ahead" forecasts; for each case, real data up to the time period prior to forecast are used,

$$s_t = NN(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4},....) \tag{68}$$

$$s_{t+1} = NN(x_t, x_{t-1}, x_{t-2}, x_{t-3},....) \tag{69}$$

$$s_{t+23} = NN(x_{t+22}, x_{t+21}, x_{t+20},....) \tag{70}$$

and "24-step ahead", which have been calculated in an iterative way; obtained forecasts are used to calculate the next ones,

$$s_t = NN(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4},....) \tag{71}$$

$$s_{t+1} = NN(s_t, x_{t-1}, x_{t-2}, x_{t-3},....) \tag{72}$$

$$s_{t+23} = NN(s_{t+22}, s_{t+21}, s_{t+20},....) \tag{73}$$

where x_t represents the real value for period t, s_t represents the forecast value for period t and NN is a previously trained network. The problem with ANN is when too many observations exist.

This abundance of data makes the ANN training phase very slow, taking hours or even days, which contradicts the initial purpose of making forecasts for the next 24h of a given day. This problem led to a training method in which it is possible to calculate forecasts in a few minutes. This method uses an ART neural net to select training examples. This kind of net has been widely used with great success to solve complex classification problems. This process culminates in the selection of a few samples that are very similar to each other and to the one to be forecast. This makes the ANN training process much faster. The ART neural net uses the correlation coefficient definition in a sensitivity analysis on historical electricity prices and loads and selects lagged prices and loads. The correlation coefficient is a measure of how well trends in a special value follow trends in past actual values. The correlation coefficient is a number between -1 and 1. If there is no relationship between a special value and the past values, the correlation coefficient is 0 or very low (the past values are not suitable to use as input for ANN). As the strength of the relationship between a special value and past values increases, so does the correlation coefficient. A perfect fit gives a coefficient of 1. Thus, the higher the correlation coefficient the larger the merit order to use as ANN input. The system demands at different hours include different characteristics. During the peak load periods there are generally large locational marginal prices (LMPs) and sometimes protruding LMPs exist due to line flow congestion. During the off-peak load periods the LMP values are generally low without distinct abnormality. Therefore, by the use of fuzzy c-mean (FCM) procedure each day is clustered into three clusters with peak, medium-peak (normal) and off-peak loads.

Before the training period is selected we must select the architecture of the network first. In order to find the optimal network architecture, several combinations must be evaluated. These configurations include networks with different number of hidden layers, different number of neurons in the hidden layer, different types of transfer functions in the hidden and output layers and different types of training algorithms. The architecture that is usually preferred is the multilayer with one hidden layer. In this architecture neurons are organized in the form of layers. It includes an input layer of source nodes that communicates with a hidden layer, whose computation nodes are denominated hidden neurons. The outputs of the hidden layer translate the input to the output layer of neurons. Therefore this network works in a different manner. The neural network considered is fully connected in the sense that every node is connected to every other node belonging to the adjacent forward layer. Moreover, the information flow progresses from the input to the output layers through the hidden layers. To have a good comparison among the different parameters that are used as ANN input, first a basic forecasting architecture for electricity price forecasting with historical electricity price and demand are constructed and then the other parameters are forced to the model.

If functional relationships of a signal slowly vary with time, then a long history of the signal can be considered as the training period, which results in a large number of training samples. Moreover, obtained results from one training phase such as weights of an ANN, can be used for many forecasts. However, functional relationships of the LMP vary much more rapidly than the hourly load. On the other hand, if the training period is selected to be very short, then the ANN cannot derive all functional relationships of the LMP due to the small number of training samples. The last 48 days have been proposed as the training period for the electricity price forecasting. The model then forecasts the LMP of the next hour, which is unseen for the ANN. When LMP of an hour is forecasted, it is used to forecast the next hour LMP, and this cycle is repeated until the LMP of the whole forecast horizon is predicted. The training phase of the ANN is executed for each weekday cluster separately. To assess the prediction accuracy of the proposed models, different statistical measures are used. For all weeks two types of prediction error measures are computed: the one is the mean average error (MAE) and the other one is the mean absolute percentage error (MAPE). The daily MAE and MAPE are computed as follow

$$MAE_{day} = \frac{1}{24}\sum_{h=1}^{24}\left|P_h^{Actual} - P_h^{Forecast}\right| \tag{74}$$

$$MAPE_{day} = \frac{100}{24}\sum_{h=1}^{24}\frac{\left|P_h^{Actual} - P_h^{Forecast}\right|}{\overline{P}_h^{Actual,24h}} \tag{75}$$

where

$$\overline{P}_h^{Actual,24h} = \frac{1}{24}\sum_{h=1}^{24}P_h^{Actual} \tag{76}$$

where MAE_{day} is the daily mean average error, P_h^{Actual} is the actual price in hour h, $P_h^{Forecast}$ is the forecasted price for that hour and $\overline{P}_h^{Actual,24h}$ is the average actual price for the day to avoid the adverse effect of prices close to zero. Analogous to daily error, the peak, normal and off-peak *MAPE* for workdays and the weakly error can also be computed.

Having accurate and fast forecasts by using this method could be a key factor for any market agents' industrial strategies. Both energy producers and energy purchases could benefit from having such forecasts. This method can be applied to forecast time series composed of too many data and when fast forecasts are required. ANN can be used to track the non-linear behavior of the LMP signal. The disadvantage of this method is that it is less accurate in summer days due to price volatility and spikes. To face this problem, some pro-processing actions can be employed, such as limitations of the magnitude of spikes or exclusion of days with price spikes from the training data. These actions can improve training and testing performances, [6, 7].

5.4 Factors influencing electricity price forecasting

Electricity prices are not only determined by consumer patterns but also by variables like weather, generation outages and network constraints that may be hard to predict. Additionally, electricity prices are affected by lack of transparency and liquidity which makes difficult for traders to make informed decisions and therefore increasing the role of psychological factors. The locational marginal prices (LMP) are usually more volatile than hourly loads and so the LMP forecasting is more complex than the short term load forecasting. This is due to the fact that LMP is dependent on the hourly loads and some other stochastic signals, such as equipment outages and fuel prices. Thus, uncertainty of hourly loads and the above stochastic signals are combined resulting in a higher level of uncertainty in the LMP. Although price curves generally experience similar changes as load curves in a transaction day, the former can be affected by more uncertainty factors. This causes higher prediction error of the LMP than the hourly load in the similar conditions, [7].

6. Conclusion

Demand and price forecasting is very important for electricity planning. A precise estimate of demand is important for the purpose of setting tariffs. Also, the utility can plan the power purchase requirements so as to meet the demand while maintaining the merit order dispatch to achieve optimization in the use of their resources. With long-term load forecasting the utility should be able to have an idea about the amount of required power in order to prepare for maximum electric load demand ahead on time. Furthermore, with price forecasting, the generation companies can reduce their risks and maximize their outcomes further. Demand and price forecasting are influenced by a number of factors. Unless the effects of these factors are well taken into consideration errors can occur in the forecast results. Forecast error in load predictions results in increased operational costs. Under-prediction of load results in a failure to provide the necessary reserves, which in turn translates to higher costs due to the use of the expensive peaking units. Over-prediction of load on the other hand, involves the start up of too many units resulting in an unnecessary increase in reserves and hence operating costs, [3].

One should keep in mind however, that the actual 24-h prediction error will depend strongly on the input data, i.e. the type of load, weather parameters such as temperature, humidity, wind speed, etc. If it is a mix of residential, industrial and commercial components, peak load tends to be driven by many factors and therefore gives less accurate results than base load which is be driven by a consistent set of factors. Also prediction error depends on load's geographical location and distribution, as well as the season of the year and the day, if it is a working day or a weekend or a holiday, [3].

In more general terms, two principal factors, the length of the lead time and the uncertainty in the explanatory variables, act to limit the accuracy of forecasting models. As the lead time increases, the accuracy of the forecast deteriorates. Also, the greater the number of explanatory variables in the model, the more uncertainty is introduced in the forecast. This is particularly true when the forecast explanatory variables have a large uncertainty of their own. Furthermore, one should be aware that different forecast weather variables have different forecast accuracies. For example, it is considerably easier to forecast temperature than it is to forecast the amount of precipitation. In more specific, for utilities or market operators that are responsible for scheduling and dispatching generation for the balance of the day and next day after one day ahead, the forecast accuracy is subject to weather forecast error. Consequently, the use of independent variables that are difficult to forecast should be avoided so as to foreclose the possibility of generating forecasts with inherently large errors. Also the type of the method and its architecture has an effect on the accuracy. For the ANN the architecture of the network plays a significant role in the accuracy of the forecasting, [3].

For utilities that are responsible for developing one to three year ahead volumetric sales forecasts by class of service (e.g. Residential, Commercial, Industrial), the larger the customer base the more accurate the forecasts. The key sources of forecast errors are under/over forecasting customer growth and actual weather conditions differing significantly to average or normal weather conditions which are used for the forecast. For utilities or system operators responsible for transmission /distribution/generation capital investment, the forecast accuracy values depend on the length of the forecast horizon. Longer term forecasts (i.e., 5 to 20 years or longer) depend heavily on accurate customer growth forecasts.

For energy retailers and utilities that are responsible for procuring short-term power to meet the energy needs of their portfolio of clients the forecast accuracy range depends heavily on the composition of the customer portfolio. In general, the larger the portfolio (i.e. the greater the number of customers served) the more accurate the forecasts become. This is due to the fact that as bigger the number of the portfolio size is, the bigger the number of the customers' electricity demand that is taking into consideration as an input in the forecasting methods and the predictions of the load and price are more accurate than the case of a small portfolio size. However this has a drawback because as more input data are introduced the processing of this data becomes more complicated and the cost of the forecasting tool is increased.

There are many methods for load and price forecasting. Each method has its advantages and disadvantages. The best suited method for all types of forecasting is Artificial Neural Network, which outperforms better with nonlinear functions and on weekend days or national holidays. However the training period should be chosen carefully. If are not to be distinguished from week day data, weekend and national holidays data a good alternative would be an ARIMA based model.

References

[1] Taylor J., De Meneces L., McSharry P. A comparison of Univariate Methods for Forecasting Electricity Demand Up to a Day Ahead. International Journal of Forecasting. 2006, 22, 1-16.

[2] Chen H., Vidagovic B., Mavris D. Multiscale Forecasting Method using ARMAX Models. Statistics at Georgia Tech, 30, 2004.
[3] Gross G., Galiana F. Short-Term Load Forecasting. Proceedings of the IEEE, 75, 1987.
[4] Demand Forecasting for Electricity, www.teriin.org/pub/papers/ft30.pdf.
[5] Kermanshahi B., Iwamiya H. Up to year 2020 load forecasting using neural nets. Electrical Power and Energy Systems. 2002, 24, 789-797.
[6] Pino R., Parreno J., Gomez A., Priore P. Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. Engineering Applications of Artificial Intelligence. 2008, 21, 53-62.
[7] Vahidinasab V., Jadid S., Kazemi A. Day-ahead price forecasting in restructed power systems using artificial neural networks. Electric Power System Research. 2008, 78, 1332-1342.
[8] El-Telbany M., El-Karmi F. Short-term forecasting of Jordanian electricity demand using particle swarm optimization. Electric Power System Research. 2008, 78, 425-433.
[9] Lauret P. et al. Bayesian neural network approach to short time load forecasting. Energy Conversion and Management. 2008, 49, 1156-1166.
[10] Abdel-Aal R.E. Univariate modelling and forecasting of monthly energy demand time series using abductive and neural networks. Computers & Industrial Engineering. 2008, 54, 903-917.
[11] Tsekouras G. et al. A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. Electric Power Systems Research. 2008, 78, 1494-1510.
[12] Mirasgedis S. et al. Models for mid-term electricity demand forecasting incorporating weather influences. Energy. 2006, 31, 208-227.
[13] Arciniegas A., Rueda I. Forecasting short-term power prices in the Ontario Electricity Market (OEM) with a fuzzy logic based inference system. Utilities Policy. 2008, 16, 39-48.

George Kourtis holds a Diploma in Electrical Engineering from the Polytechnic School of Aristoteleio University of Thessaloniki (AUTH). During 2007 he has worked as a technician for LOGICOM LTD, where he was responsible for the maintenance of 3G Mobile stations. He is a postgraduate student at the University of Cyprus where he is doing his Master Degree on Electrical Engineering. He joined Electricity Authority of Cyprus (EAC) Research and Development Department in 2008, where he works on European funded research programs as an electrical engineer.

Ioannis Hadjipaschalis holds a B.Eng (Electrical Engineering), from the University of Sheffield, an M.Sc from the London School of Economics and an MBA. During 1996-2000 he worked at the Electricity Authority of Cyprus as Electrical Engineer, while during the period 2001-2005 he worked at ACNielsen Market Research. He joined EAC Research and Technological Development team in 2006, where he has been working on European funded research involving RES, Distributed Generation (DG), Electricity distribution networks, Hydrogen and CO_2 capture and storage (CCS) technologies.

Andreas Poullikkas holds a B.Eng. degree in mechanical engineering, an M.Phil. degree in nuclear safety and turbomachinery, and a Ph.D. degree in numerical analysis from Loughborough University of Technology, U.K. He is a Chartered Scientist (CSci), Chartered Physicist (CPhys) and Member of The Institute of Physics (MInstP). His present employment is with the Electricity Authority of Cyprus where he holds the post of Assistant Manager of Research and Development; he is also, a Visiting Fellow at the University of Cyprus. In his professional career he has worked for academic institutions such as a Visiting Fellow at the Harvard School of Public Health, USA. He has over 20 years experience on research and development projects related to the numerical solution of partial differential equations, the mathematical analysis of fluid flows, the hydraulic design of turbomachines, the nuclear power safety, the electric load forecasting and the power economics. He is the author of various peer reviewed publications in scientific journals, book chapters and conference proceedings. He is the author of the postgraduate textbook: Introduction to Power Generation Technologies (ISBN: 978-1-60876-472-3). He is, also, a referee for various international journals, serves as a reviewer for the evaluation of research proposals related to the field of energy and a coordinator of various funded research projects. He is a member of various national and European committees related to energy policy issues. He is the developer of various algorithms and software for the technical, economic and environmental analysis of power generation technologies, desalination technologies and renewable energy systems. E-mail address: apoullik@eac.com.cy