THE FORECASTING OF TRANSMISSION NETWORK LOADS

by

DANIEL FREDERIK PAYNE

submitted in part fulfilment of the requirements
for the degree of

MASTER OF SCIENCE

in the subject

OPERATIONS RESEARCH

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF CA VAN DER MERWE

JOINT SUPERVISOR: PROF C ROON

NOVEMBER 1997

Acknowledgements

I would like to acknowledge the following people, all of whom contributed to this endeavour in their different ways:

First and foremost, I would like to thank my wife, Henta, for her unwavering support whenever I needed it most, and for all the sacrifices she made during my early years of study.

Secondly, all my Unisa lecturers, who helped to lay the foundation of this dissertation.

My colleagues, for their support, helping me to understand the practical side of the load forecast; and especially Tracy Hill, who drew all the sketches and diagrams, with outstanding results.

Professor CA van der Merwe, my supervisor, and Professor C Roon, my joint supervisor, for their inputs, which enabled me to find a practical load forecasting model.

Finally, an acknowledgement I cannot express adequately in words, to our Creator, who made this milestone possible.

Preface

The forecasting of Eskom transmission electrical network demands is a complex task. The lack of historical data on some of the network components complicates this task even further. In this dissertation a model is suggested which will address all the requirements of the transmission system expansion engineers in terms of future loads and market trends. Suggestions are made with respect to ways of overcoming the lack of historical data, especially on the point loads, which is a key factor in modelling the electrical networks. A brief overview of the transmission electrical network layout is included to provide a better understanding of what is required from such a forecast. Lastly, some theory on multiple regression, neural networks and qualitative forecasting techniques is included, which will be of value for further model developments.

Key words

Electrical load forecast, demand load forecast, load forecast model, demand load forecast model, point load forecast for electrical network modelling, electrical network parameter forecast, MW forecast, transmission electrical network forecast, transmission electrical network demand forecast, neural networks, multiple regression

Acknowledgements

Preface

Key words

CHAPTER 1

Introduction

CHAPTER 2

Regression

CHAPTER 3

Neural networks

1

CHAPTER

INTRODUCTION

1.1 Purpose of the dissertation

The purpose of this dissertation is to develop an electrical demand load forecasting model and suggest possible forecasting techniques for later improvements. This model has to meet the requirements of Eskom's transmission system expansion engineers and be aligned with Eskom's business process plan (see section 6.2).

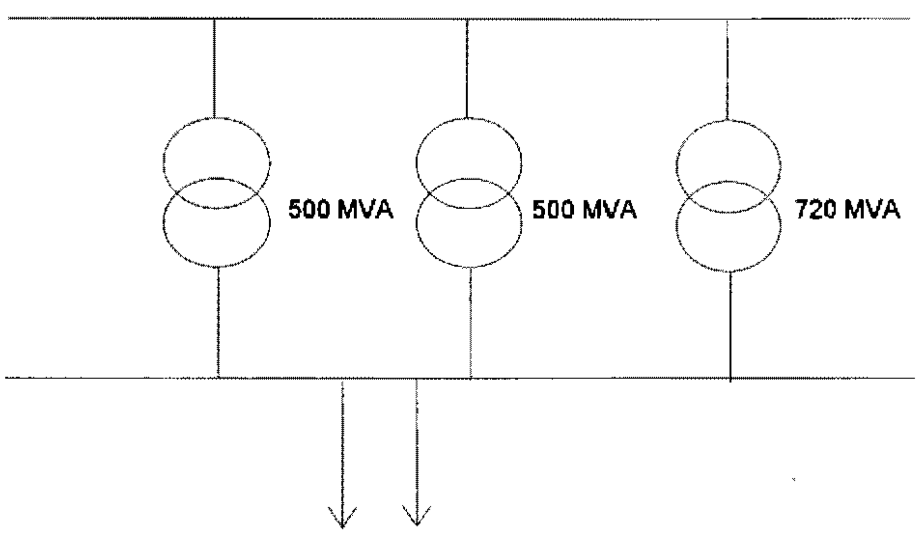


Fig 1.1: Firm capacity

The forecasting model is required to improve the accuracy of the demand load forecast, which has a horizon of 10 to 15 years. The forecast results are used in electrical network studies and to determine which transmission substations will exceed their firm capacities under normal operating conditions. Firm capacity is the maximum megavolt amperes

(MVA) a substation can supply when the transformer with the highest rating is out of commission. Fig 1.1 depicts three installed transformers, with a firm substation capacity of 1 000 MVA.

1.2 Network studies

In network studies (load flow studies) the total load on the system for a given condition is simulated by a set of point loads. Each point load consists of two values: a megawatt (MW) and a megavolt ampere reactive (Mvar) value. An example of this is a condition in which the Eskom system is at its peak demand, also called time of system peak (TOSP). Some point load demands are as low as 50% of their maximum demand at TOSP. Eskom's transmission electrical networks are geographically grouped into 25 customer load networks and the peak demands for the customer load networks have, so far, not occurred at the same time. It is therefore good practice to calculate a set of point loads when each customer load network is at its peak demand and also one for TOSP; thus a total of 26 conditions.

An electrical network consists of three principal components: the generating stations, the transmission system and the distribution systems. The transmission lines are the connecting links between the generating stations and the distribution systems. The connection points between the transmission and distribution systems are simulated by a set of point loads.

For load flow studies, each power station's nett sent out MW (section 6.8), is modelled at its maximum capacity. Solving the base-case files, the engineers scale these maximum capacities down to obtain a more realistic generation pattern and to minimise the flow (MW) on the slack busbar. The slack busbar is used in network studies, to either react as a load (if the nett sent out MW is more than the MW for the loads, including system losses), or otherwise react as a power station. When the hourly total power station nett sent out figures (MW) are available, a more realistic generation pattern can be modelled instead of modelling the maximum capacity. This generation pattern has to be verified with System Operation scheduling the daily power station nett sent out (see section 6.8).

In the past the distribution networks have been only Eskom supply points and the generating stations only Eskom power stations. However, with Eskom's new trading policy with regard to neighbouring states, the generating stations include power from neighbouring states and the distribution systems include supplies to neighbouring states.

Diversification is the scaling of the maximum point load demands to determine, for a given condition, a new set of loads. There is, however, no historical data available on the point loads, which makes the diversification of point loads very difficult.

Eskom has more than a million account numbers. Some of those account numbers contain only energy consumption megawatt-hour (MWh) figures. To add up all these account numbers' demands to obtain a given point load demand would be too laborious. A more practical solution would be to follow a macro-approach in forecasting the maximum demands for the point loads and the substations.

1.3 Techniques

Neural networks and multiple regression have been investigated as possible techniques to forecast these maximum demands. But these approaches have a number of shortcomings. One of the main problems is the forecasting horizon of ten years or more. Neural networks and multiple regression are not long-term forecasting techniques. Therefore, for a period longer than five years, long-term forecasting techniques should be considered.

Developments in electrical supply technology are also an important aspect to consider. For example, an electric motorcar service network throughout the Southern African continent might look far-fetched at present, but in ten years it may be a viable option to replace the ageing combustion technology. The associated new electrical car manufacturing industry could also have an impact on the required electrical loads. To determine the resultant future trends in electrical requirements from specific developments and the economy in general, long-term holistic forecasting techniques are needed.

The mathematical modelling of experts' views regarding key growth issues (increases and decreases in demands) in the South African electrical industry has to be researched.

Theory on decision-making, pattern recognition, identification of underlying causal relationships, linear programming to restrict output (dependent) variable(s) between lower and upper limits, etc are also important to consider as possible theory for later improvements.

1.4 Data

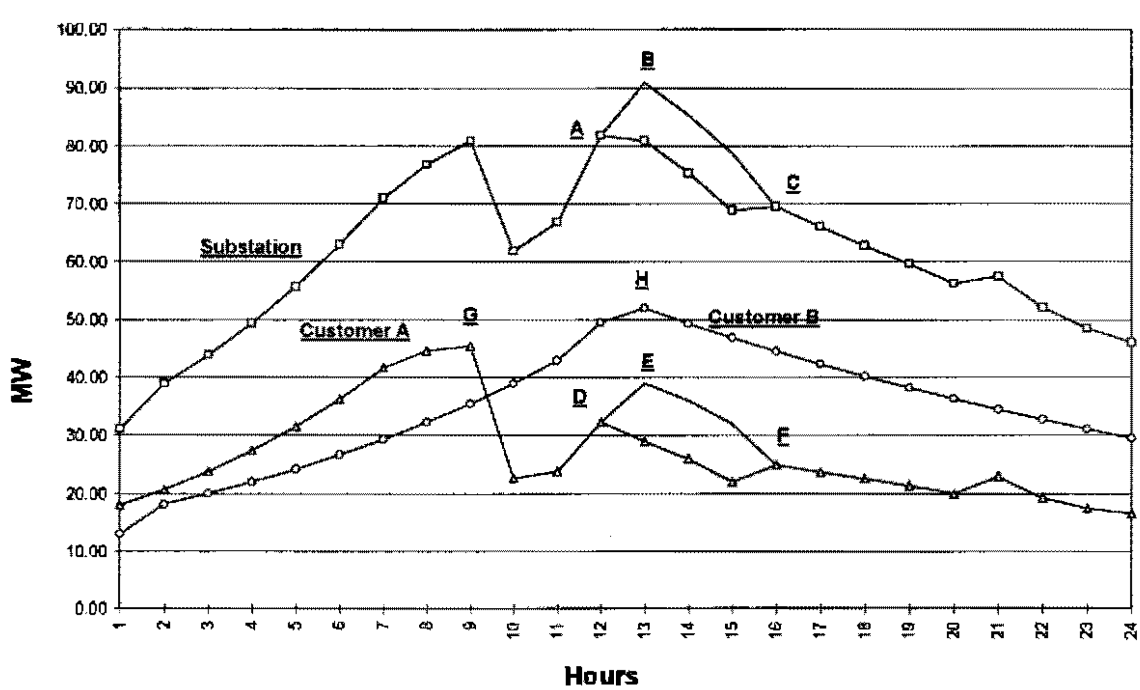
The forecasting model, and especially the model's inputs, depend on the type of data available in terms of historical and future values. For a more detailed discussion on the type of historical data available, refer to section 6.11. In this section only a few comments will be made.

As mentioned in section 1.2, Network studies, no historical point load demand data is available. A metering project has been initiated to record those demands in the future, and in accordance with the project terminology those point loads will be called "stats" points.

For some transmission substations not directly connected to the distribution networks, no historical data is recorded. Those substation demand forecasts are currently done by subjective methods based on knowledge of load flows and the network.

Another important factor is the lack of system operating data such as data on network switching, load shifting, etc. In some cases a customer is supplied by more than one Eskom substation, and it is possible to do switching on the customer's networks to shift loads between these Eskom substations.

For example, if a substation feeds two customers, and one of the customers (customer A) also takes supply from a second substation, then the following may happen. In graph 1.1 the maximum demands for the substation and customers A and B under normal network operating conditions are respectively points A, G and H. If the load for customer A should increase by 10 MW from 13:00 to 16:00 (points D, E and F), it would not change the maximum demand of customer A, but the substation's maximum demand would increase (points A, B and C).



Graph 1.1: Hourly maximum demands (substation and customers)

Although a customer (for example customer A) may have more than one point of delivery (POD) - sometimes more than 20 - the customer can still have only one account number. These PODs can be supplied from more than one Eskom substation (see Fig 1.2). The monthly maximum demand shown on the account is the monthly maximum demand of the hourly (or half-hourly) summated POD demands, and the maximum demand indicated on the account is usually less than the sum of the individual POD maximum demands. In some cases the customers' monthly demands do not vary significantly due to load shifts - only the demands of the individual substations may show higher demand readings. When other customers are also fed from these substations, the determination of the actual demand becomes fuzzy.

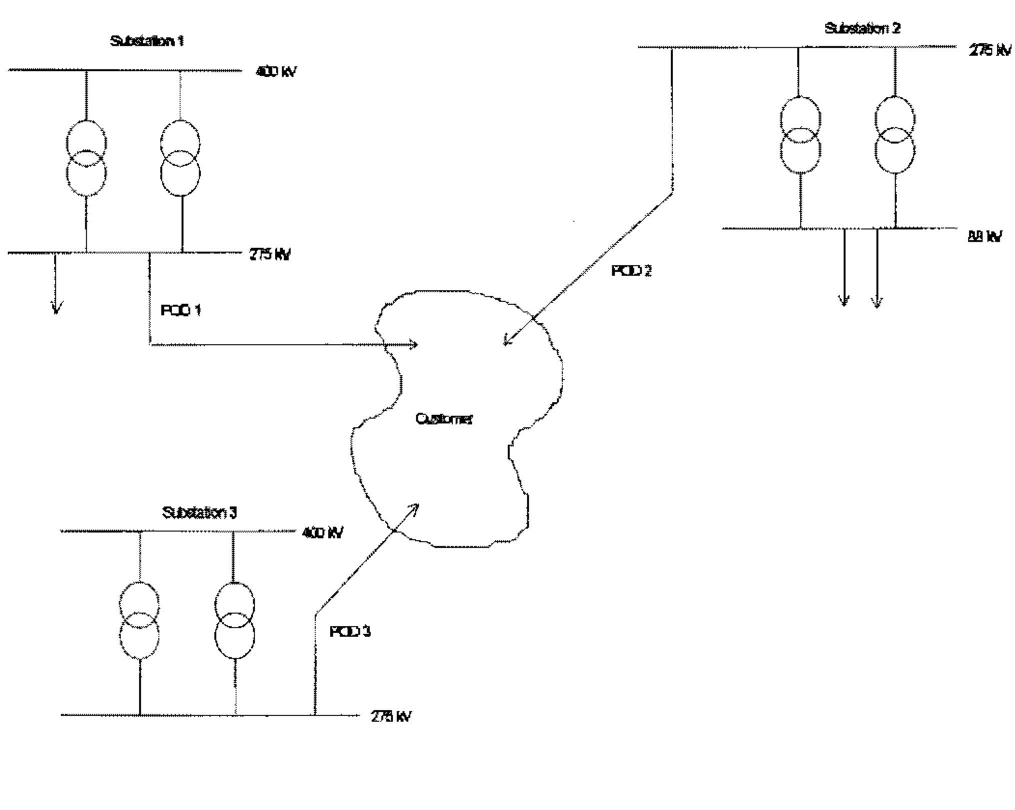


Fig 1.2: Points of delivery

The forecast POD demands are annual maximum figures, which makes it extremely difficult to forecast a substation's annual demand. This is because the summation of the POD demands has to be diversified to determine the substation demands (see section 6.13) and no data exists to determine the diversification factors. A further complication arises in that these substation annual maximum demands are only available from 1993 onwards (hourly readings only from August 1993). Thus only four values are available at that time to determine the relationship between the maximum POD demands and a maximum substation demand.

1.5 The model

In Chapter 5 the substation demands are predicted, using the customers' diversified maximum demands. This method produced unsatisfactory results. The transmission substation demands have been defined as:

$$D = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \ldots + \beta_n P_n \qquad \text{(Eq 1.1)}$$

where

n = the number of accounts allocated to the substation

- D is the demand for the substation in kW
- β_0 is the residual values (actual minus forecast)
- β_j is the diversity factor for the j th account allocated to the substation
- P_j is the j th account annual maximum demand in kW, allocated to the substation

It appears to be an easy **multiple regression** problem, but it is not that simple. The regression coefficients (diversity factors) change over time due to changing weather conditions and patterns, developments in the electrical supply technology, changing demand patterns, etc. Secondly, some accounts have more than one POD, feeding from more than one transmission substation (ring feeders - see section 6.5), and because of the different network operating conditions the demands from the different substations also change. Some of the PODs are not measured in kW, but in either kWh or kVA. To convert the kWh or the kVA readings into kW, load factors and power factors have to be used. These factors also change over the years.

Another constraint is the lack of data, but, despite the data problems, a model has been created which produces the required forecasting results. The model takes into account the variation of diversity factors, the effects of the differing operating conditions, etc. The lack of historical data on the point loads and the higher transmission substations has also been overcome (see Chapter 6).

The model operates on the principle of balancing known values with other known values and then using the balanced terms to produce an acceptable value for the missing data, for example the point loads.

1.6 Engineering requirements

The engineers require not only a set of point loads for network load studies but also a generation pattern. The current program uses MW values and power factors to determine a set of point loads.

The power factors are modelled to determine the Mvar required to supply the MW demands at the point load. The relationship between MW, MVA, Mvar and power factor is best explained by a triangle, ie the Pythagoras' theorem (see Fig 1.3).

In the determination of the point loads, different power factors are used for the respective point loads, but the same set of power factors is used for all the years. This is not acceptable. A forecast of substation power factors is required for each of the years constituting the forecasting horizon, and for each condition. Although historical data is not available, some effort should be made to improve power factor forecasts.

All available generating stations are modelled at their maximum capacity (MW and Mvar). Those capacities are diversified according to the modelled set of point loads, the generating cost, etc. This is called a **generation pattern**.

No historical Mvar figures or data are available. For the Eskom generating stations, however, more than five years of hourly MW figures are available.

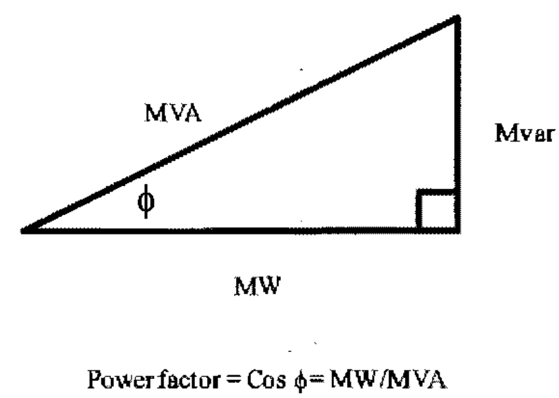


Fig 1.3: Relationship between MW, MVA, Mvar and power factor

The transmission system expansion engineers' requirements can thus be summarised as the following:

1. The annual maximum demand for all transmission substations directly connected to the distribution networks.
2. A set of point load demands for all specified conditions (MW and power factors).
3. For each point load, the lower and upper Mvar limits and the corresponding MW at the time of the lower and upper limits respectively, for all specified conditions.
4. The generation pattern and var equipment status for all specified conditions.
5. The annual demands for all transmission substations for all specified conditions.
6. The annual maximum demands for all transmission transformers and lines.
7. New loads and changes in existing loads, ie increases or decreases,

having the effect of a step function on existing load patterns.

8. The effect of network topology changes on the transfer of loads on the transmission electrical system.
9. The expected energy not served due to possible network failures.
10. Transmission losses.

1.7 Overview of the other chapters

In the following chapters some theory will be discussed, ie regression theory (Chapter 2), neural network theory (Chapter 3) and qualitative techniques (Chapter 4). Then, in Chapter 5, the results obtained from substation forecasting and why those results are not satisfactory will be discussed. Lastly the proposed forecasting model (Chapter 6) will be looked at.

CHAPTER 2

REGRESSION

To generate forecasts of future events, information has to be available concerning historical events. The historical data is analysed and, if accepted, it will be used as a basis to predict future events.

Quantitative forecasting methods involve the analysis of historical data in an attempt to predict future values of a variable (area demands) that is of interest. Quantitative forecasting methods can be grouped into two kinds - time series analysis and causal techniques.

The use of causal forecasting models involves the identification of other variables that are related to the variable to be predicted. Once these related variables have been identified, a statistical model describing the relationship between these variables and the variable to be forecast is developed. The statistical relationship derived is then used to forecast the variable of interest. The area demands may be related to a number of political and economic indicators, etc. In that case, the area demands would be referred to as the dependent variable, whereas the other variables would be referred to as the independent variables. The next step is to estimate, statistically, the functional relationship between the dependent variable and the independent variable(s). Having determined this, and also that the relationship is statistically significant, the predicted future values of the independent variables can be used to predict the future values of the dependent variable.

Both for an area load forecast and for a substation load forecast, more than one independent variable is required to describe the functional relationship between the dependent and the independent variables. Therefore only the least squares procedure as used in multiple linear regression is explained. If only one customer is fed from a substation, then the customer demand equals the substation demand.

Before discussing the least squares method, it is important to understand what is meant by linear and nonlinear regression. A linear model is defined as a model that is linear in the parameters, even if the regressor variables (independent variables) are nonlinear. For example, if the relationship between the dependent variable and the independent variables is of a polynomial nature, then the model is still linear. Nonlinear relationships will be discussed in section 2.32.

2.1 The least squares estimation

The method of least squares will be used to determine the statistically functional relationship between the dependent and the independent variable(s); it has been used more extensively than any other estimation procedure for building regression models [1]. The method minimises the residual (actual values minus forecast values) sum of squares when the functional relationship between the dependent and the independent variable(s) is determined.

The linear model can be written as:

$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon \qquad [2]$$

where $\mathbf{X}$ is the design matrix and of dimension n x p. The number of observations equals n, the number of independent variables is k, and p (number of model parameters) equals (k + 1).

The error vector ε and the response vector $\mathbf{y}$ are random variables. The model error is assumed to be uncorrelated from observation to observation, with mean zero and constant variance. The independent variables are not random.

The terms of the linear model $\mathbf{y} = \mathbf{X}\,\beta + \varepsilon$, can also be written as:

$$\mathbf{y} = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{vmatrix} \qquad \mathbf{X} = \begin{vmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ & \vdots & \vdots & & \vdots \\ & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{vmatrix} \qquad \beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_1 \\ \vdots \\ \beta_k \end{vmatrix}$$

and

$$\varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{vmatrix}$$

The column vector β (the regression coefficients vector) is of dimension p x 1, or (k + 1) x 1.

The least squares estimates of the parameters are given by the matrix equation

$$\beta = \mathbf{(X'X)^{-1}\,X'y}$$

where $\mathbf{X'}$ is the transpose of the matrix $\mathbf{X}$ and $\mathbf{(X'X)^{-1}}$ is the inverse of the matrix $\mathbf{X'X}$.

The question is: will these coefficients determine a functional relationship between the dependent and the independent variables which is statistically significant? The most significant relationship will not always give the best results in practice [3].

In the following sections some important aspects of regression analysis will be discussed.

2.2 Regression coefficients' signs

It is important to check whether the regression coefficients have the correct signs [4]. Sometimes, due to multicollinearity, the coefficient is of the opposite sign. This problem will be discussed in more detail in section 2. 28.

2.3 The $(X'X)^{-1}$ matrix

The regression coefficients are determined by the least squares estimation, (section 2.1) and are given as:

$$\beta = (X'X)^{-1} X'y$$

The matrix $(X'X)^{-1}$ is also used to determine the variance-covariance matrix of β (see section 2.6). The matrix $(X'X)$ has to be nonsingular, otherwise no inverse can be obtained. When the independent variables are highly correlated (see section 2.28) singularity becomes a problem in determining the inverse matrix.

2.4 Sum of squares

The sum of squares is widely used in statistical inference [5]. For regression applications the sum of squares is as follows:

$$SS_{Total} = SS_{Reg} + SS_{Res}$$

where

$$SS_{Total} = \Sigma (y_i - \bar{y})^2$$

$$SS_{Reg} = \Sigma (\hat{y}_i - \bar{y})^2$$

$$SS_{Res} = \Sigma (y_i - \hat{y}_i)^2$$

and

y_i is the i th observation

$\hat{y}_i$ is the i th fitted value

$\bar{y}$ is the average value

2.5 Model error variance

In practice an estimate of the model error variance (σ^2) is required. The estimate is used to calculate the coefficient standard errors for hypothesis testing. The residuals (actual values minus fitted values) are the observed errors of fit and are also the empirical counterparts of the model errors which are not observed. Thus it is reasonable that the sample variance of the residuals' should provide an estimator of σ^2. If the residuals' sum of squares is divided by (n - p), an unbiased estimator is produced [6].

The quantity s^2 is often called the error mean square.

$$s^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n - p)$$

$$= SS_{Res} / (n - p)$$

and

(n - p) is the degree of freedom

2.6 Variance of the regression coefficients

The calculation of the variance-covariance matrix is subject to the following assumptions [7]:

1. The model error variance is homogeneous (constant variance) with mean zero.
2. The model errors are uncorrelated from observation to observation.
3. The model errors are assumed to be normally distributed.
4. The independent variables are not random and are measured with negligible error.

The variance-covariance matrix of β is given by

$$Var(\beta) = s^2 (X'X)^{-1} \qquad [7]$$

2.7 Analysis of variance (ANOVA)

Given that the errors have an $N(0 ; \sigma^2)$ distribution and σ^2 is constant, the analysis of variance is given as follows [8].

The F statistic is the ratio of the two mean squares (regression/residual). The numerator refers to the variance that is explained by the regression; and the denominator refers to the variance of what is not explained by the regression, namely the errors.

In the case of simple regression, the F-test is, in fact, the same as testing the significance of the slope [9]. In the case of multiple regression, the F-test is testing the overall significance of the regression model [10].

As a rule of thumb, the F value has to be above 3.23 (say 5), for the regression model to be of overall significance [11].

Source	Sum of squares (SS)	df (degrees of freedom)	Mean square (MS)	F value
Regression (explained)	SS_{Reg}	p - 1	$SS_{Reg}/(p - 1)$	$F = MS_{Reg} / s^2$
Residual (unexplained)	SS_{Res}	n - p	s^2	
Total	SS_{Total}	n - 1		

2.8 Coefficient of determination (R^2)

The coefficient of determination is the proportion of the explained variation as part of the total variation [12]. The R^2 can be no smaller than 0 and no larger than 1 in any regression situation.

As a rule of thumb the R^2 values have to be larger than 0.7 [12]. However, if the values are smaller than 0.7, another independent variable may be included in the model, but there are certain pitfalls, such as overfitting (see section 2.17).

$$R^2 = SS_{Reg} / SS_{Total}$$

or

$$R^2 = 1 - SS_{Res} / SS_{Total}$$

and

$$SS_{Total} = \Sigma (y_i - \bar{y})^2$$

$$SS_{Reg} = \Sigma (\hat{y}_i - \bar{y})^2$$

$$SS_{Res} = \Sigma (y_i - \hat{y}_i)^2$$

2.9 The t-test

The t-test is used to test whether a regression coefficient is statistically significantly different from zero or not [13]. Thus the hypothesis for each parameter is:

$$H_0 : \beta = 0$$

$$H_1 : \beta \# 0$$

and the degrees of freedom are (n - p).

Once again, as a rule of thumb, if the absolute t-statistic value is larger than 2, then the hypothesis can be rejected and the parameter is statistically significantly different from zero [14].

For β_0

$$t_0 = \beta_0 / (\mathrm{Var}(\beta_0))^{1/2}$$

and for β_i

$$t_i = \beta_i / (\mathrm{Var}(\beta_i))^{1/2} \quad (i = 1, 2, \ldots, n)$$

The t statistic is given as:

$$t = \text{coefficient / standard error of coefficient}$$

2.10 Regression coefficient intervals

Under normal theory assumptions the confidence intervals for the coefficients are respectively [15]:

$$\beta_0 \pm t_{\alpha/2\ n-p}\ (s^2 / \mathrm{Var}(\beta_0))^{1/2}$$

$$\text{and}\quad \beta_i \pm t_{\alpha/2\ n-p}\ (s^2 / \mathrm{Var}(\beta_i))^{1/2} \qquad (i = 1, 2, \ldots, n)$$

2.11 Confidence intervals on mean response and prediction intervals

The assumption of normality of the errors is again very important. The confidence intervals on the mean response are used for $\mathbf{x} = \mathbf{x_0}$ when $\mathbf{x_0}$ is not a new observation; otherwise, if $\mathbf{x_0}$ is a new observation, then the confidence intervals on prediction interval are used. The estimated response for $\mathbf{x} = \mathbf{x_0}$ is given as $y(\mathbf{x_0})$ [16].

Mean response:

$$y(\mathbf{x_0}) \pm t_{\alpha/2\ n-p}\ s\ (\mathbf{x_0}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0})^{1/2}$$

Prediction interval:

$$y(\mathbf{x_0}) \pm t_{\alpha/2\ n-p}\ s\ (1 + \mathbf{x_0}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0})^{1/2}$$

2.12 Criteria for a best model

Myers described three criteria, which can be summarised as follows [17]:

1. Learn something about the system from which the data is taken. This may be nothing more than the knowledge of a "sign" of a coefficient or the slope of a growth rate.
2. Learn which regression coefficients are important and which are not.
3. Select, from a group of possible candidate models, the best one for prediction.

The following paragraph by Myers summarises the situation very well [17]: **"We cannot ignore input from experts in the scientific discipline involved. (Statistics are rarely a substitute for sound scientific knowledge and reasoning.) Statistical procedures are vehicles that lead us to conclusions; but scientific logic paves the road along the way. However, a good scientist must remember that to arrive at an adequate prediction equation, balance must be achieved that takes into account what the data can support. There are times when inadequacies in the data and random noise may not allow the true structure to come through. For these reasons, a proper marriage must exist between the experienced statistician and the learned expert in the discipline involved."**

2.13 Standard criteria for comparing models

Three possible standard criteria which could be used to compare candidate models are [18]:

1. Coefficient of determination, which is a measure of the model's capability to fit the present data (see section 2.8).
2. The estimate of error variance (see section 2.5). Normally the candidate model with the smallest s^2 value is chosen.
3. Adjusted R^2. This is a statistic which guards against overfitting, ie the inclusion of an excessive number of model terms. If a new term is added, then the coefficient of determination will increase (or at least not decrease).

The adjusted R^2 is given as:

$$\text{Adj. } R^2 = 1 - \frac{SS\ Reg/\ (n - p)}{SS\ Total/\ (n - 1)} \quad [18]$$

In general the ordinary residuals do not provide information on how the regression model will predict. These residuals are measures of quality of fit and do not assess quality of future prediction. In the next two sections some measures will be given to assess the model's quality of prediction, followed by some suggestions on underfitting and overfitting, and then by some procedures for variable selection.

2.14 Data splitting

The actual data is split into two samples: sample one (the fitting sample) and sample two (the validation sample) [19]. A model is built and tested to check whether the relationship between the dependent and the independent variables is statistically significant. If the relationship is statistically significant, the data from the validation sample must be used to calculate predicted values and compare these values with the actual values.

2.15 PRESS statistic

In the case of the substation forecasts, not enough data is available to split the data into samples (fitting and validation samples). Myers suggested a form of validation, very much in the spirit of data splitting, ie the PRESS (Prediction sum of squares) statistic [20]. Consider a set of data in which the first observation is withheld from the sample and the remaining n - 1 observations are used to estimate the coefficients. The first observation is then replaced and the second observation withheld, whereafter the coefficients are estimated again. Each observation is removed in turn and the coefficients are estimated again, n times in total.

The individual PRESS residuals give separate measures of the stability of the regression and they can assist the forecaster to remove data points or observations which have a significant influence on the outcome of the regression.

$$\mathrm{PRESS} = \sum_{i=1}^{n} \left(e_i / (1 - h_{ii})\right)^2$$

where

e_i = actual value - fitted value for the i th observation

h_{ii} = i th diagonal element of the H matrix

The H matrix, also called the HAT Matrix, is defined as :

$$\mathbf{H = X(X'X)^{-1}\,X'} \qquad [21]$$

2.16 Impact of underfitting

If the model is underfitted (underspecified), important variables have been ignored [22]. The variation accounted for by the ignored variables is then deposited in the residual sum of squares. Also, s^2 equals the residual sum of squares divided by (n - p). Thus, if a new variable is added, given that the residual sum of squares stays constant, s^2 will increase. If the model is underfitted, the residual sum of squares will most likely be reduced, and thus also s^2.

2.17 Impact of overfitting

A model is overfitted (overspecified) when variables have been added which contribute little or nothing, but only produce variances that are larger than those of simpler models [23].

2.18 The C_p statistic

This statistic is used to check whether the model is underfitted or overfitted. A reasonable value for C_p is p (a value suggesting that the model contains no estimated bias) [24].

$$C_p = p + \frac{(s^2 - \underline{\sigma}^2)/(n-p)}{\underline{\sigma}^2}$$

$$\underline{\sigma}^2 = \frac{s^2(n-2)}{n}$$

2.19 Variable selection procedures

Three methods will be discussed, ie forward selection, backward elimination and stepwise regression [25].

FORWARD SELECTION

In the case of forward selection, the initial model contains only a constant term. The procedure selects for entry the variable that produces the largest R^2 of any single regressor. Otherwise one can, say, choose the regressor with the largest partial F for entry. The above process continues until the regressor with the largest partial F for entry does not exceed a value pre-selected for the F values to enter.

BACKWARD ELIMINATION

This procedure begins by including all the variables. The first to be removed is the regressor, which results in the smallest decrease in R^2 (thus the smallest partial F statistic). The procedure is continued until the candidate regressor for removal has a partial F value which exceeds the preselected F values required to leave.

STEPWISE REGRESSION

Stepwise regression provides an important modification of forward selection. At each stage of selection, all regressors currently in the model are evaluated through the partial F-test. A preselected F value for a regressor to leave is used. Thus, at all stages following entry, a variable must continue to perform or it will be eliminated. The procedure terminates when no additional regressors can enter or be eliminated, depending on preselected entering and leaving F values.

2.20 Sequential F-test on individual parameters

Given additional regressor variable(s) has (have) to be included, ie $\mathbf{X}_1$. The partial F-test is done to determine whether the extra sum of squares regression is sufficient to warrant inclusion of β_1 in the model [26].

Suppose $\mathbf{X}$ and β are subdivided and $\mathbf{X}_1$ is n x p_1 ($p_1 + p_2 = p$) and β_1 contains the p_1 scalar elements associated with the columns of $\mathbf{X}_1$., then

$$\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2] \qquad \beta = \begin{vmatrix} \beta_1 \\ \dots \\ \beta_2 \end{vmatrix}$$

The reduction in the residual sum of squares by introducing $\mathbf{X}_1\beta_1$ into a model containing $\mathbf{X}_2\beta_2$ is given as

$$\mathbf{R}(\beta_1 \mid \beta_2) = \mathbf{R}(\beta_1 , \beta_2) - \mathbf{R}(\beta_2)$$

or

$$\mathbf{R}(\beta_1 \mid \beta_2) = \mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2']\mathbf{y}$$

The term $\mathbf{R}(\beta_1 , \beta_2)$ represents the regression explained by all the model terms, including the constant. The term $\mathbf{R}(\beta_2)$ represents the regression sum of squares explained by the model involving only $\mathbf{X}_2\beta_2$.

The partial **F** statistic is then given as

$$F = \frac{\mathbf{R}(\beta_1 \mid \beta_2)/ \mathbf{p}_1}{s^2}$$

Thus if F is larger than $\mathbf{F}_{p1,n-p}$ (the test statistic), then β_1 is significantly different from zero and can be included.

2.21 Analysis of residuals

The residuals are often used to detect and assess the degree of discrepancy between the model assumed and the data observed. In this section the residuals will be analysed to ensure that the model's performance is acceptable not only for fitting but also for prediction. A model with a good fitting is not necessarily a good model for use in prediction.

The residuals are also called the errors. An error is the difference between the actual measured value for the dependent variable and the model output for a given observation. As mentioned previously, these residuals should be normally distributed with a mean of zero and a constant variance. Further, the residuals should be uncorrelated between observations.

The variance-covariance matrix of the residuals can be determined by using the HAT matrix, ie:

$$\mathrm{Var}\,(\mathbf{e}) = \sigma^2\,(\mathbf{I} - \mathbf{H}) \quad [27]$$

and

$$H = X(X'X)^{-1}\,X'$$

2.22 Plotting the residuals

Plotting the residuals can provide some valuable information. Fig 2.1 shows that there is no obvious pattern in the residuals, meaning that the model may be acceptable.

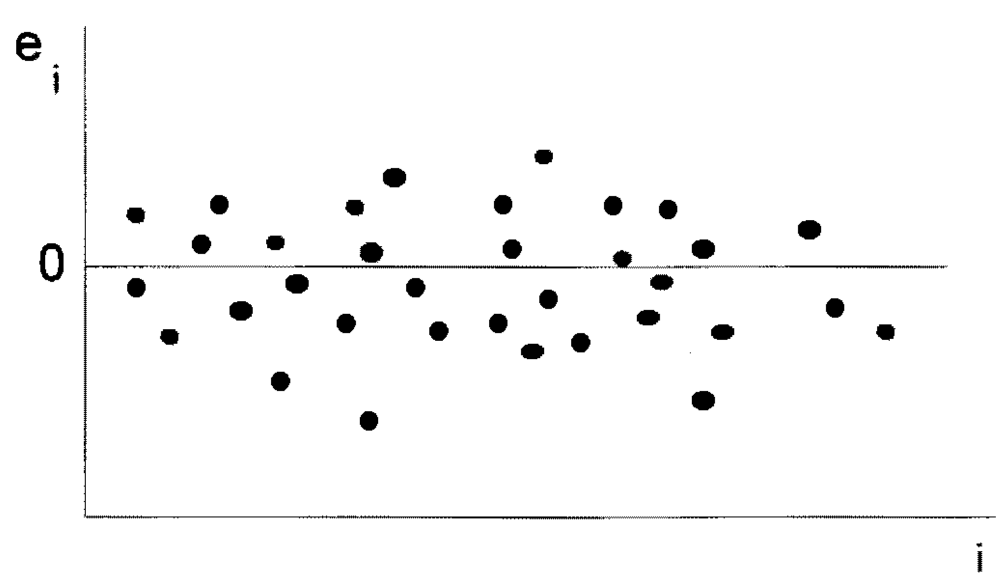


Fig 2.1: No obvious pattern

Figs 2.2 and 2.3 illustrate the two possible cases of autocorrelation [28]. In Fig 2.2 the case of negative autocorrelation is apparent. Negative autocorrelation exists when a negative error is followed by a positive error, then another negative error, and so on.

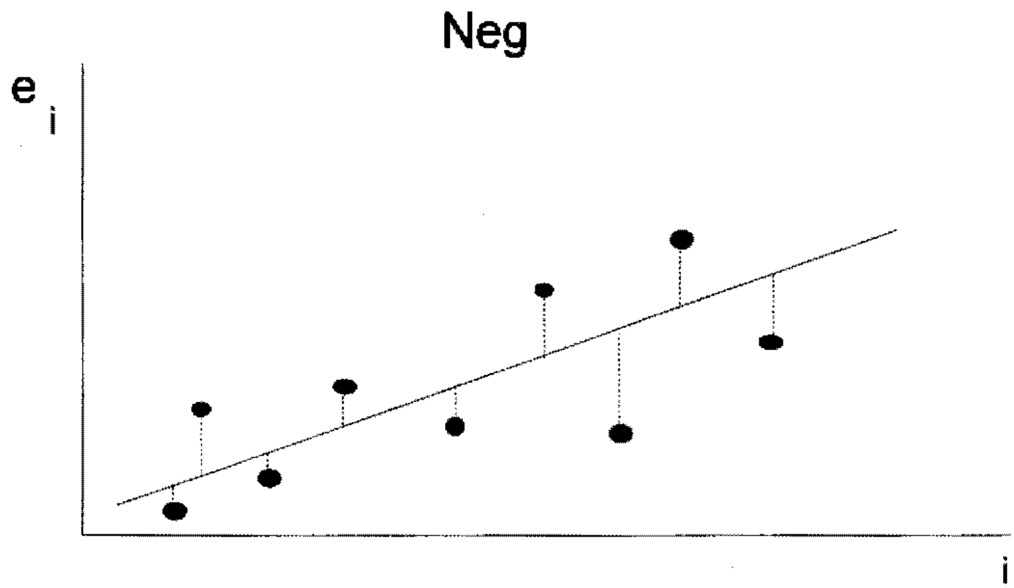


Fig 2.2: Negative autocorrelation

Fig 2.3 indicates the existence of positive autocorrelation. In positive autocorrelation, positive errors tend to be followed by other positive errors, while negative errors are followed by other negative errors.

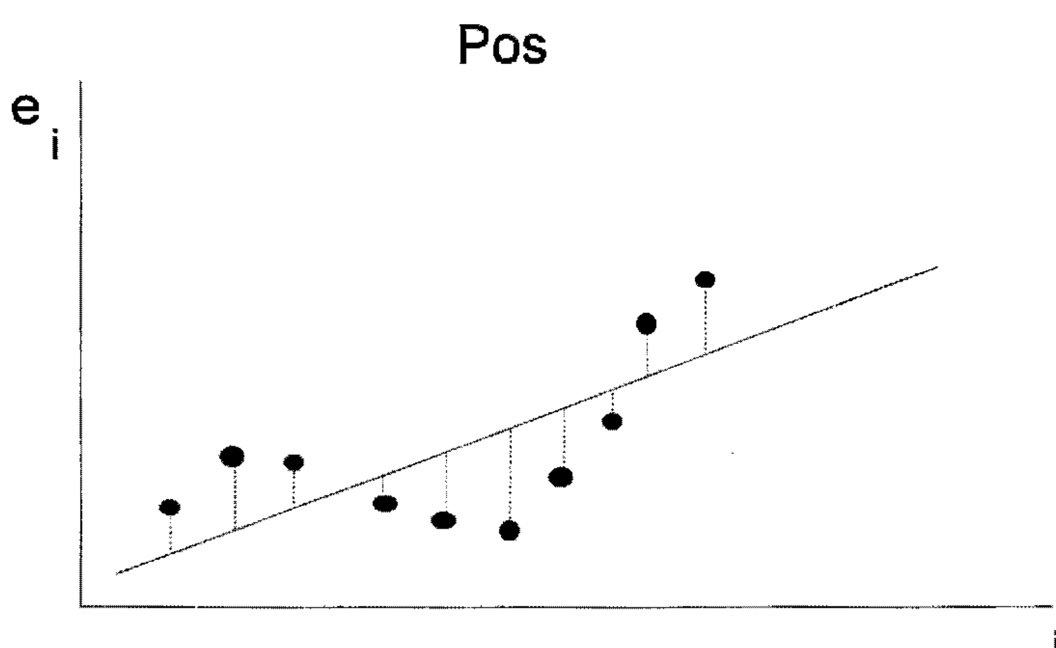


Fig 2.3: Positive autocorrelation

The funnel effect shown in Fig 2.4 indicates the existence of heterogeneous variance, also called heteroscedasticity [29]. In section 2.30 hetero-scedasticity is explained in more detail.

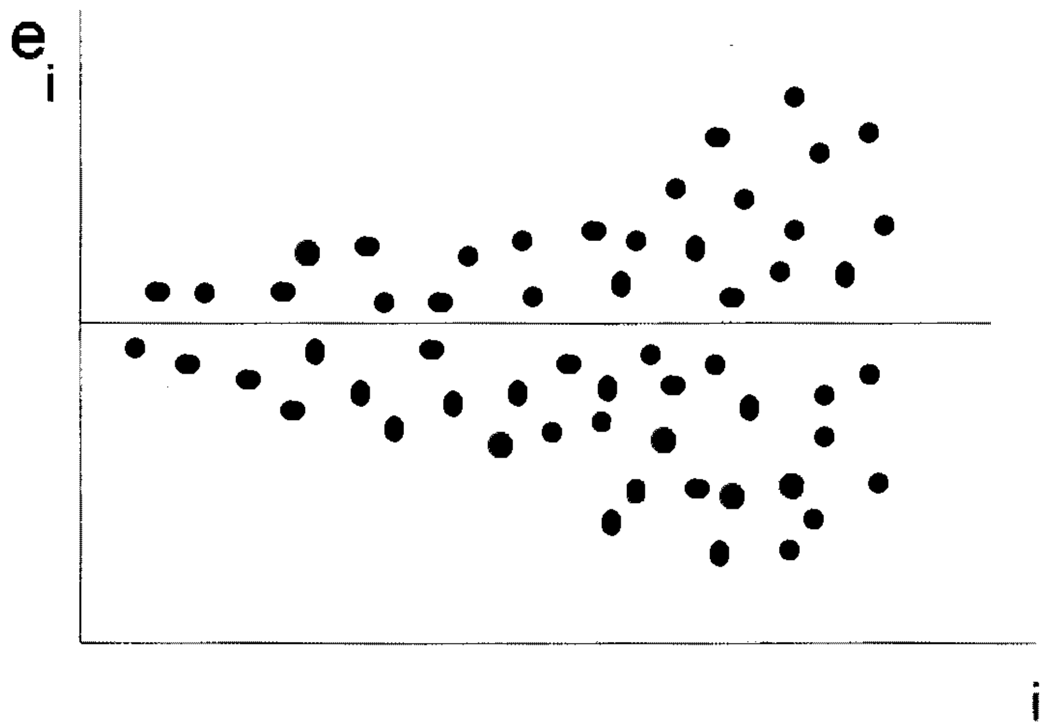


Fig 2.4: The funnel effect (heteroscedasticity)

2.23 Normality detection in residuals

The t-tests, F-tests and confidence limits on parameters require the assumption of normality. Further motivation for detection of violation of the normality assumption lies in the fact that the method of least squares possesses more desirable properties when the errors are normal. Minor departures from normality have little impact on the regression results.

The true model errors, eg ε_i, when plotted against their respective expected values, eg $\mu_{(i)}$, should produce a plot that is reasonably close to a straight line through the origin, since $\mu = 0$ for the ε_i. The normal probability plot is a plot of residuals against their expected values. If the residuals are suitably standardised, the straight line should go through the origin with a slope of 1. One possible approach is as follows:

Plot $(y_i - \hat{y}_i)/s$ against $z\ ((i - 0.375)/(n + 0.25))$ [30], where $(y_i - \hat{y}_i)$ is the i th smallest residual (see Fig 2.5) [30].

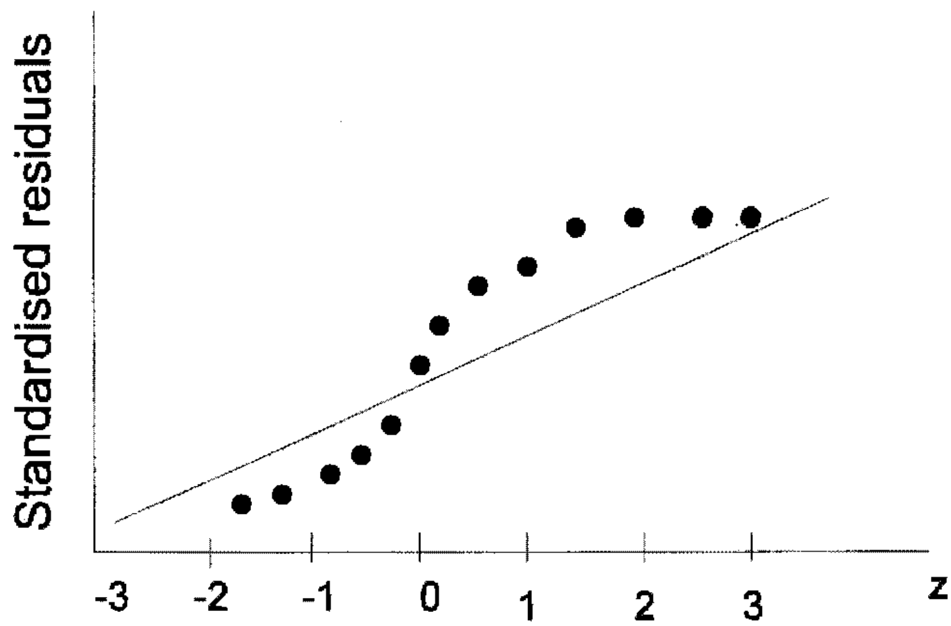


Fig 2.5: Normal probability plot

2.24 Studentised residuals

If there is a large variation in the HAT diagonals among the data points, there will be large differences in the variances of the residuals. A HAT diagonal near unity defines a data point that is remote from the data centre. Standardised residuals have zero mean and unit variance. It is simpler to develop a crude yardstick to measure their size than it would be in the case of the ordinary residual. The advantage of using the studentised residual over the use of the ordinary residual stems from the fact that the standardisation eliminates the effect of the location of the data point in the regressor space [31].

The studentised residuals (r_i) are given as

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}$$

Let us look at some results given in Myers to compare the differences between the residuals, the studentised residuals and the diagonal elements of the HAT matrix.

The results (Table 3.1) for installations 23 and 24 represent examples of cases in which the use of studentised residuals is clearly more informative than that of ordinary residuals. In the case of installation 23 the ordinary residual (-160.276) is certainly not large in comparison with the other residuals. The HAT diagonal, however, is 0.9885, characterising the data point as being remote from the data centre.

The resultant studentised residual of -3.278 strongly suggests that, under ideal conditions, the residual represents an unusual departure from zero, given the location of the point in the regressor space.

Installation	y_i	y_i	e_i	r_i	h_{ii}
1	180.23	209.985	-29.755	-.0.076	0.2573
2	182.61	213.796	-31.186	-0.075	0.1609
3	164.38	360.486	-196.106	-0.470	0.1614
4	284.55	360.106	-75.556	-0.181	0.1631
5	199.92	380.703	-180.783	-0.430	0.1475
6	267.38	510.373	-242.993	-0.582	0.1589
7	999.09	685.167	313.923	0.763	0.1829
8	1103.24	1279.299	-176.059	-0.483	0.3591
9	944.21	815.466	128.744	0.334	0.2808
10	931.84	891.846	39.994	0.094	0.1295
11	2268.06	1632.137	635.923	1.493	0.1241
12	1489.50	1305.177	184.323	0.453	0.2024
13	1891.7	1973.416	-81.716	-0.187	0.0802
14	1387.82	1397.786	-9.966	-0.023	0.0969
15	3559.92	4225.131	-665.211	-2.197	0.5576
16	3115.29	3134.895	-19.605	-0.056	0.4024
17	2227.76	2698.738	-470.978	-1.302	0.3682
18	4804.24	4385.778	418.462	1.236	0.4465
19	2628.32	2190.326	437.994	1.007	0.0868
20	1880.84	2750.910	-870.070	-2.401	0.3663

Table 3.1

Installation	y_i	$\hat{y}_i$	e_i	r_i	h_{ii}
21	3036.63	2210.134	826.496	1.883	0.0704
22	5539.98	5863.847	-323.894	-1.536	0.7854
23	**3534.49**	**3694.766**	**-160.276**	**-3.278**	**0.9885**
24	**8266.77**	**7853.505**	**413.265**	**2.580**	**0.8762**
25	1845.89	1710.861	135.029	0.441	0.5467

Table 3.1 (cont)

2.25 Outliers

A suitable statistic to detect outliers in the data is the R-Student statistic [32]. The degrees of freedom are (n - p - 1), and in Myers, Table C4, figures are provided to determine the test statistic.

The R-Student statistic is formulated as:

$$t_i = \frac{y_i - \hat{y}_i}{s_{-i}\sqrt{(1 - h_{ii})}}$$

where

$(y_i - \hat{y}_i)$ is the i th residual and

$$s_{-i} = \sqrt{\frac{(n-p)\, s^2 - e^2_i/(1 - h_{ii})}{n - p - 1}}$$

If t_i is smaller than the test statistic, the null hypothesis can be accepted; that means the i th point is not an outlier. In the example given in section 2.24, the R-Student results (Table 3.2) are as follows.

Given that 25 residuals are being tested, the proper 0.05 level critical value for the R-student (8 model parameters) is given as 3.69 (see Myers, Table

C.4, Appendix C). As a result, the only data point that can be classified as an outlier is that associated with installation 23, in which the R-Student value is -5.2423.

Installation	e_i	r_i	R-Student	h_{ii}
1	-29.755	-.0.076	-0.0736	0.2573
2	-31.186	-0.075	-0.0726	0.1609
3	-196.106	-0.470	-0.4594	0.1614
4	-75.556	-0.181	-0.1762	0.1631
5	-180.783	-0.430	-0.4196	0.1475
6	-242.993	-0.582	-0.5704	0.1589
7	313.923	0.763	0.7532	0.1829
8	-176.059	-0.483	-0.4720	0.3591
9	128.744	0.334	0.3246	0.2808
10	39.994	0.094	0.0914	0.1295
11	635.923	1.493	1.5537	0.1241
12	184.323	0.453	0.4426	0.2024
13	-81.716	-0.187	-0.1818	0.0802
14	-9.966	-0.023	-0.0224	0.0969
15	-665.211	-2.197	-2.5192	0.5576
16	-19.605	-0.056	-0.0541	0.4024
17	-470.978	-1.302	-1.3310	0.3682
18	418.462	1.236	1.2566	0.4465
19	437.994	1.007	1.0074	0.0868
20	-870.070	-2.401	-2.8657	0.3663
21	826.496	1.883	2.0538	0.0704
22	-323.894	-1.536	-1.6057	0.7854
23	**-160.276**	**-3.278**	**-5.2423**	**0.9885**
24	413.265	2.580	3.2093	0.8762
25	135.029	0.441	0.4299	0.5467

Table 3.2

2.26 Dummy variables

If we have a single quantitative regressor variable x_1 in a situation where there are two categories, for example summer and winter, and the regression models for the two categories have different intercepts but the same slope, then the following model is applicable [33]:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 z_t + \varepsilon_t$$

with $z_t = 0$ if in the first category

$z_t = 1$ if in the second category

The variable z, with a value of either zero or one, is sometimes called a dummy variable, or according to Myers a categorical variable.

If the four quarters of a year have to be modelled due to seasonality, then three dummy variables can be used as follows :

Quarter	z_1	z_2	z_3
One	0	0	0
Two	1	0	0
Three	0	1	0
Four	0	0	1

2.27 Interaction

In section 2.26 the use of a dummy variable (categorical variable) was discussed. If the slopes of the regression lines (for the two categories) are also different, then another term has to be added to the model, ie [34]:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 z_t + \beta_3 x_{1t} z_t + \varepsilon_t$$

This occurs when there is interaction between x_1 and the categorical variable z.

2.28 Multicollinearity

What is multicollinearity?

This is a condition with multiple regression models when two or more regressor variables are highly correlated. The term *multicollinearity* is self-explanatory; multi implying many, and collinear implying linear dependencies [35].

Consequences of multicollinearity

Multicollinearity increases the standard errors of the regressor coefficients. When high multicollinearity is present, confidence intervals for the coefficients tend to be very wide and t-tests for significance testing tend to be very small. Sometimes, although the t-tests tend to be very small, the F-test for the full model is significant. Also, with high multicollinearity the signs for the parameters could be wrong; they could, for example, be negative when they should be positive. In substation load forecasts this would theoretically mean that, when a customer increases its electricity demand, the substation supplying that customer will decrease its demand. Although this cannot actually happen, it is a good example of the consequences of multicollinearity.

If multicollinearity is a serious problem, and a large set of regressor variables are removed, the remaining coefficients may change by large amounts, and perhaps also the sign. This instability of the regressor variables is unsatisfactory.

Detecting multicollinearity

A simple but effective method is to determine the correlation between the independent variables. If the correlation coefficients of the independent variables are larger than 0.7 it can be regarded as an indication of multicollinearity between those variables [36].

Secondly, Myers mentioned the use of a variance inflation factor [37]:

$$VIF = 1/((1 - R^2_i)$$

R^2_i is the coefficient of multiple determination of the regression produced by the regressor variable x_i against the other regressor variables, x_j ($j \neq i$). Thus, the higher the multiple correlation in the artificial regression, the lower the precision in the estimate of the coefficient b_i. With no multiple correlation $R^2_i = 0$, which means that the VIF = 1.

The VIFs represent a considerably more productive approach for detection correlation between the independent variables than simple correlation values. Though no rule of thumb on numerical values is foolproof, it is generally believed that if any VIF exceeds 10, there is a reason for at least some concern.

Dealing with multicollinearity

One method of combating multicollinearity is **ridge regression** [38].

The ridge regression estimator of the coefficient β is given by

$$\beta = (X' X + k D)^{-1} X' y$$

where D is an n x n diagonal matrix

and

k is the shrinkage parameter [39]

The data is first normalised (for example, for regressor variable x_{1t} : (x_{1t} - average(x_1))/s_1) to reduce the value of k. Then k is increased until stability is indicated in all coefficients [40].

A second method is **principal component regression** [41]. With this method a least squares estimation is performed on a set of artificial variables called principal components of the correlation matrix. Principal components are orthogonal to each other, therefore it is quite easy to attribute a specific amount of variance to each. This technique was not tested because no statistical package with this technique was available.

Another method is to drop the regressor variables causing the problem of multicollinearity. But in the demand forecast it is important to include as many customers' demands as possible. The reason for this is that customers' demands are used for the regressor variables and when one is dropped, the future growth of the substation can be incorrectly predicted because growth differs from one customer to another. Where possible, customers from the same sector (gold mines, ferrochrome, etc) are therefore grouped together and then used as a new regressor variable.

2.29 Autocorrelation

One of the assumptions of the ordinary least squares regression model is that the error terms are independent and normally distributed, with a mean of zero and a constant variance. If this is true, we do not expect to find any regular pattern in the error terms. When a significant time pattern that violates the independence assumption is found in the error terms, autocorrelation is indicated.

When autocorrelation exists, problems can develop in using and interpreting the OLS regression function. The existence of autocorrelation does not bias the regression coefficients that are estimated, but it does make the estimates of the standard errors smaller than the true standard errors. This means that the t ratios calculated for each coefficient will be overstated, which in turn may lead to the rejection of the null hypothesis, which should in fact not be rejected. Thus the coefficient is not statistically significantly different from zero, but due to autocorrelation the t ratio indicates that the coefficient is statistically significantly different from zero and the regressor variable will, therefore, erroneously be included.

In addition, the existence of autocorrelation causes the R^2 and F statistics to be unreliable in evaluating the overall significance of the regression function.

Three possible methods to deal with autocorrelation are suggested:

1. Include another independent variable.
2. Include the square of one of the existing independent variables in the regression function as a new variable.
3. Include an independent variable that is the lagged value of the dependent variable.

Durbin-Watson test

The Durbin-Watson (DW) statistic is a test to determine the existence of autocorrelation statistically [42]. The DW statistic has been included here simply because it is another method to analyse the residuals. If there are no patterns in the errors, autocorrelation does not exist.

The DW statistic will always be in a range of 0 to 4. As a rule of thumb, a value close to 2 (eg between 1.75 and 2.25) indicates that there is no autocorrelation [43]. As the degree of negative autocorrelation increases, the value of the DW statistic approaches 4. If positive autocorrelation exists, the value of the DW statistic approaches 0. To be more precise in evaluating the significance of the DW statistic, use the table below according to the following five tests. The values for d_l (lower limit) and d_u (upper limit) are determined from Appendix B, depending on the values of k and N (the number of observations).

As indicated in section 2.1, k is the number of regressors and p = k + 1.

The Durbin-Watson statistic is:

$$DW = \sum (e_t - e_{t-1})^2 / \sum e_t^2$$

Test	Value of calculated DW	Conclusion
1.	$0 < DW < d_l$	Positive autocorrelation exists.
2.	$d_l < DW < d_u$	Result is indeterminate.
3.	$d_u < DW < (4 - d_u)$	No autocorrelation exists.

Test	Value of calculated DW	Conclusion
3.	$d_u < DW < (4 - d_u)$	No autocorrelation exists.
4.	$(4 - d_u) < DW < (4 - d_l)$	Result is indeterminate.
5.	$(4 - d_l) < DW < 4$	Negative autocorrelation exists.

2.30 Heteroscedasticity

The assumption of homogeneous error variance is often violated in practical situations. If the error variance is not constant from one observation to another, a strategy called weighted least squares can be considered.

As indicated in section 2.1

$$\boldsymbol{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

and the variance-covariance matrix of the residuals is

$$\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}$$

and I is of dimension n x n.

If the homogeneous error variance assumption does not apply, then, as in section 2.21, the variance-covariance matrix of the residuals can be determined by using the HAT matrix, ie

$$\text{Var}(\mathbf{e}) = \sigma^2 (\mathbf{I} - \mathbf{H}) = \mathbf{V} \qquad [44, 45]$$

and

$$H = X(X'X)^{-1} X'$$

then the regression coefficients are determined as follows:

$$\boldsymbol{\beta}^* = (\mathbf{X'VX})^{-1}\mathbf{X'V^{-1}y} \quad [46]$$

The estimator β^* is unbiased, ie $E(\beta^*) = \beta$ and the estimators in β^* achieve minimum variance of all unbiased estimators under the condition $\varepsilon \sim N(\mathbf{0},\mathbf{V})$.

The failure of the homogeneous variance assumption is very natural and to be expected [47]. The source of the problem is that the error variance is often not independent of the mean of E(y). As the regressor values and responses in the data become larger, variance around the regression tends to grow. Often a transformation of the response is suggested to stabilise the error variance, but that is dependent on how the error variance fluctuates with the mean response.

A variance-stabilising transformation that is quite often used is the natural log transformation, ie log (y) [47]. This is an appropriate transformation when the error standard deviation is proportional to E(y). Another transformation is the inverse transformation, ie 1/ y. This transformation may be appropriate in the rather extreme situation in which the error standard deviation appears to be the quadratic function of the mean response. That is, σ is proportional to $[E(y)]^2$.

In his book entitled Basic Econometrics, Gujarati [48] suggested that heteroscedasticity may be reduced by using log transformations. He also discussed the Park test to detect heteroscedasticity. The test is formulated as:

$$e_i^2 = \alpha + \beta X_i + v_i \qquad [48]$$

If β proved to be statistically significant, it would suggest that heteroscedasticity is present in the data; if insignificant homoscedasticity may be assumed.

2.31 Influence diagnostics

From section 2.25 (Outliers) it is evident that the conditions detected in the outlier diagnostics are errors in the y direction, that is a residual which is larger than expected. Myers mentioned the concept of a regression data point which is a high-influence observation [49]. In fact all high-influence observations are not due to errors in the y direction; they can also be due to extremes in the x direction, ie a disproportionate distance away from the data centroid in the x's, even though it is a proper observation.

The R-Student statistic and the HAT diagonal values may be used as guidelines to detect these outliers and high-influence observations. The R-Student statistic was discussed in section 2.25.

In the case of the HAT diagonal, we can make use of the fact that [50]

$$\sum_{i=1}^{n} h_{ii} = p$$

where p is the number of model parameters. Thus the average h_{ii} is p/n provides a norm. Certainly any $h_{ii} > 2p/n$ has a potential to exert a strong influence on the results.

The R-Student statistic and HAT diagonal values reveal which individual observations have the potential to exert excessive influence. The following three statistics can be used to determine the extent of the influence.

DFFITS

The DFFITS diagnostic is used to gain some insight as to what influence observation i has on the predicted value or fitted value $\dot{y}_i$ [51]. Reasonable values for $(\text{DFFITS})_i$ are between -2 and 2.

$$(\text{DFFITS})_i = (\text{R-Student})_i \, (h_{ii} / (1 - h_{ii}))^{½}$$

DFBETAS

A large value in the diagnostic $(\text{DFBETAS})_{ji}$ indicates that the i th observation has a considerable impact on the j th regression coefficient [52]. Reasonable values for $(\text{DFBETAS})_{ji}$ are between -2 and 2. Again the diagnostic presents the combination of leverage measures and the impact of errors in the y direction.

Let matrix **R** be

$$\mathbf{R} = (\mathbf{X'X})^{-1}\,\mathbf{X'}$$

with the (q,s) element denoted by $r_{q,s}$ and the j th row of **R** by $\mathbf{r}'_j$. The diagnostic is

$$(\text{DFBETAS})_{ji} = (\text{R-Student})_i \,(1/(1 - h_{ii})^{1/2})\,(r_{j,i} / (\mathbf{r}'_j\, \mathbf{r}_j)^{1/2})$$

Cook's D statistics

As in the $(\text{DFFITS})_i$ and $(\text{DFBETAS})_{j,i}$, D_i is related to the residuals and data point leverage measures. D_i is computed as [53]

$$D_i = (r_i^{\,2}/ p)\,(h_{ii}/ (1 - h_{ii})$$

where r_i is the i th studentised residual. D_i becomes large for either of, or both, the following

1. A poor fit (large $r_i^{\,2}$) at the i th point
2. High leverage (h_{ii} close to 1)

How large should Cook's D be? The forecaster should, through experience, determine his or her own rule of thumb, depending on the situation.

2.32 Nonlinear regression

If linear modelling is not successful, nonlinear modelling can be considered [54]. But many nonlinear models fall into categories that are designed for specific situations. Some of these models are shown below.

Logistic growth model	$y = \dfrac{\alpha}{1 + \beta \exp(-kx)} + \varepsilon$	α = the value y approaches as x grows larger β and k must be positive
Gompertz growth model	$y = \alpha \exp[-\beta\ e^{-kx}] + \varepsilon$	Again the parameter α is the limiting growth

Richards growth model	$y = \dfrac{\alpha}{[1 + \beta e^{-kx}]^{1/\delta}} + \varepsilon$	
The Weibull growth model	$y = \alpha - \beta \exp[-\gamma x^{\delta}] + \varepsilon$	In this case the growth at $x = 0$ is $\alpha - \beta$, while again the growth approaches its maximum $y = \alpha$ as $x \to \infty$
The Mitcherlich law	$y = \alpha - \beta\gamma^{x} + \varepsilon$	

2.33 Transformations improving fit and prediction

Sometimes it may be useful to transform some of the variables to achieve better results. Some of the most often used transformations are [55]:

Parabola	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$
Hyperbola (inverse)	$1/y = \beta_0 + \beta_1 1/x + \varepsilon$
Natural log on y	$\log(y) = \beta_0 + \beta_1 x + \varepsilon$
Natural log on y and x	$\log(y) = \beta_0 + \beta_1 \log(x) + \varepsilon$
Inverse exponential	$\log(y) = \beta_0 + \beta_1 (1/x) + \varepsilon$
Polynomial	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$
Square root	$y = \beta_0 + \beta_1 (x)^{0.5} + \varepsilon$

2.34 Suggestions for regression software packages

For a regression software package to be effective in the field of demand load forecasts, apart from the obvious calculations and tests of the regression coefficients, the following options have to be available.

1. R^2 and the adjusted R^2
2. Regression coefficient intervals
3. Confidence intervals on mean response and prediction
4. Variable selection procedures
5. Sequential F-tests on individual parameters
6. Transformations of variables (eg to cater for interaction)
7. PRESS statistic
8. C_p statistic
9. Graphical analysis of residuals
10. Normality plots
11. Studentised residual r_i
12. R-Student statistic
13. Methods to detect and combat multicollinearity (ridge regression and principal component regression)
14. Methods to detect (Durbin-Watson test) and combat autocorrelation
15. Methods to detect and combat heteroscedasticity (weighted least squares)
16. Influence diagnostics (DFFITS, DFBETAS, Cook's D, etc)

2.35 Multicollinearity and autocorrelation influences on test results

As stated in sections 2.28 and 2.29 multicollinearity and autocorrelation may effect the restults of the statistical tests associated with regression analysis. These effects are summarised below as follows.

	Multicollinearity	**Autocorrelation**
t-tests	Small	Overstated
R^2		Unreliable
F-test	Significant	Unreliable

	Multicollinearity	Autocorrelation
Instability of regressor variables	Possible	
Opposite regressor variable signs	Possible	

2.36 Application in demand load forecasts

In section 1.3, the mathematical modelling of experts' views regarding key growth issues (increases and decreases in demands) in the South African electrical industry was discussed. These results can be used to predict area or electrical sector future loads which can play an important role in the balancing section (section 6.15.11). Multiple regression can be investigated to predict future area and sector loads. In section 4.2 a typical application wil be discussed, in which Dr Choi predicted the peak demands.

3

CHAPTER

NEURAL NETWORKS

3.1 Introduction

Artificial neural networks are sometimes considered as simplified models of the human brain. Some authors feel that this is misleading, because the human brain is too complex and is not well understood. But there are nevertheless a number of similarities which can be examined.

Human nerve cells, called neurons, consist of three parts: the cell body, the dendrites and the axon.

The body (soma) is the large, relatively round central body in which almost all the logical functions of the neuron are performed. It carries out the biochemical transformations required to synthesise the enzymes and other molecules necessary to the life of the neuron.

Each neuron has a hair-like structure of dendrites (inputs) around it. The dendrites are the principal receptors of the neuron and serve to connect its incoming signals.

The axon (output) is the outgoing connection for signals emitted by the neuron.

Synapses are specialised contacts on a neuron which are the termination points for the axons from other neurons. Synapses play the role of interfaces connecting some axons of the neurons to the spines of the input dendrites

[56,57,58].

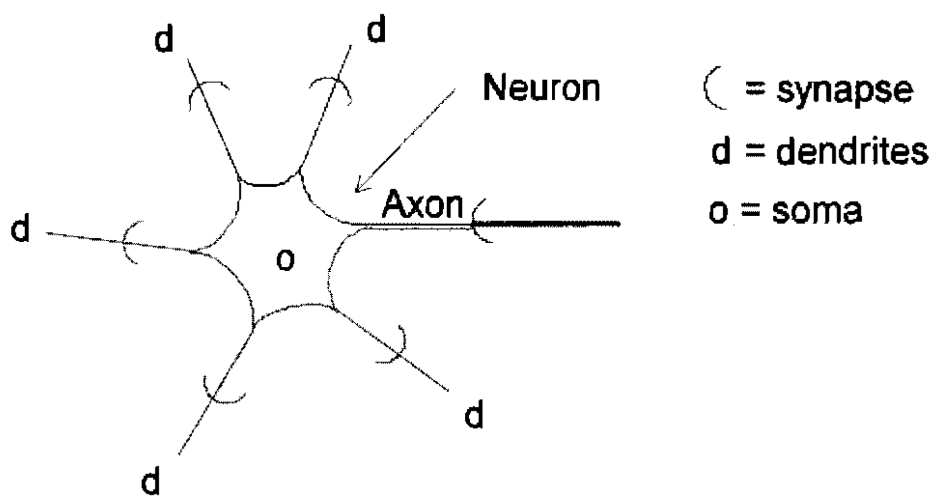


Fig 3.1: Components of a neuron

A discussion follows on the development of artificial neural networks over the last decades, followed by some terminology commonly used with regard to artificial neural networks (hereafter called neural nets). Next, the discussion will deal with some neural nets which may be used for area load forecasts, power generation patterns, pattern recognition, and identifying different network operations. The discussion ends with some remarks on more advanced neural nets that could possibly be applied to the estimation of missing data and ways to improve the efficiency of the model in cases of network changes.

3.2 Neural network developments over the last decades

3.2.1 McCulloch-Pitts neuron

In 1943 a neurobiologist, Warren McCulloch, and a statistician, Walter Pitts, designed artificial neurons and linked them together, which is generally regarded as the first neural network. This model compares a weighted sum of all the inputs with the value of a threshold. If the sum exceeds the threshold value, then the neuron output is set at 1, otherwise the output is 0. With an output of 1, it is also said that the neuron "fires". Mc Culloch-Pitts neurons are used most widely as logic circuits [Anderson and Rosenfeld, 1988] [59].

3.2.2 Hebb learning

The earliest and simplest learning rule for a neural net is generally known as the Hebb rule, suggested by Donald Hebb in 1949. He was, at the time, a psychologist at McGill University. He proposed that learning occur by modification of the synapse strengths (weights) in a manner such that if two interconnected neurons are both "on" at the same time, then the weight between those neurons should be increased. Some refinements were made to allow computer simulations [Rochester, Holland, Haibt and Duda, 1956]. This idea is closely related to the correlation matrix learning developed by Kohonen (1972) and Anderson (1972) [59].

In biological terms, this means that a neural pathway is strengthened every time activation on each side of the synapse is correlated.

3.2.3 Perceptron

In 1957 Frank Rosenblatt of Cornell University, USA, published the first major research project on neural computing, ie The Development of an Element Called a Perceptron. Rosenblatt's perceptron gave rise to a considerable amount of research in neural computing.

Frank Rosenblatt (1958, 1959, 1960) introduced and developed a large class of artificial neural nets. The single-layer perceptron consists of an input layer (input signals x_i). These input signals are multiplied by a set of adaptable weights. The processed inputs and weights are subject to a hard limiter, resulting in a quantised binary output. The perceptron learning rule uses an iterative weight adjustment that is more powerful than the Hebb rule.

The early successes with perceptrons led to enthusiastic claims. A paper by Minsky and Papert in 1969 on the limitations of the perceptron dampened this enthusiasm, however, which was the beginning of the *quiet years of neural nets.*

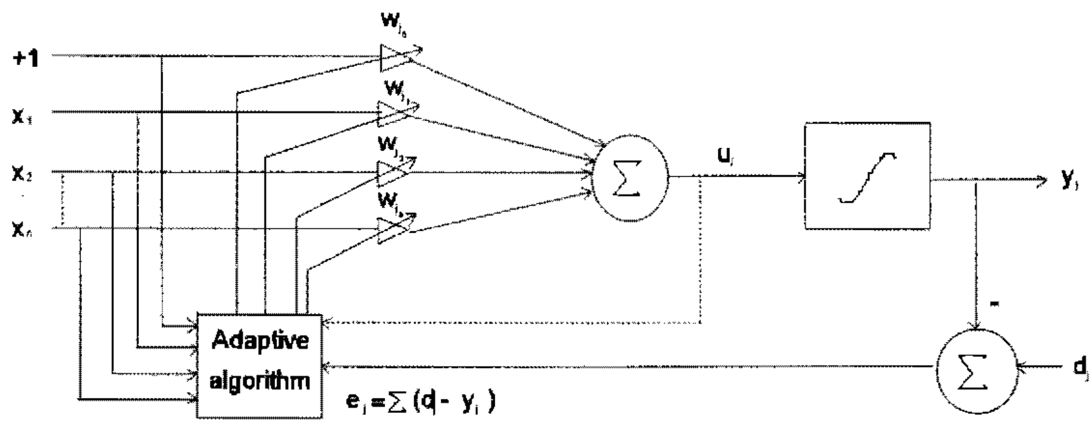


Fig 3.2: Single-layer perceptron with hard-limiter

One of Minsky and Papert's discouraging results shows that a single-layer perceptron cannot simulate a simple exclusive-or function. This function accepts two inputs that can be only zero or one. It produces an output of one only if either input (but not both) is one [60, 61]]

These results are summarised as follows:

X input	Y input	Output
0	0	0
1	0	1
0	1	1
1	1	0

3.2.4 Adaline

One of the simplest models of the artificial neuron with learning capabilities is the adaptive linear neuron, called Adaline, which was developed by Bernard Widrow and his student Hoff in 1960 [60, 62].

The Adaline and a two-layer variant, the Madaline (multiple Adaline), were used for a variety of applications including speech recognition, character recognition, weather prediction and adaptive control. Later the Adaline was modified to produce a continuous rather than a discrete output.

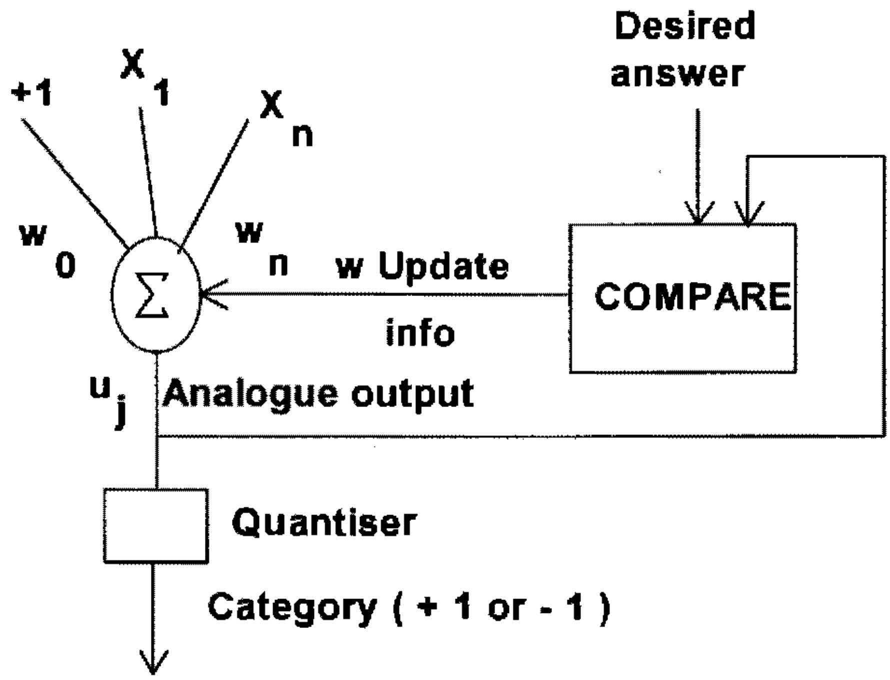


Fig 3.3: Adaline

Later the Widrow-Hoff delta rule or least mean squares (LMS) rule was used to adjust the synaptic weights [63].

Adalines are often used as basic building blocks in artificial neural networks.

Summary of the Adaline [64]:
(Note: The training pair **s** is bipolar).

$$y\_in = b + \sum_{i=1}^{n} x_i w_i$$

$$x_i = s_i$$

$$b = b^{Old} + \beta (t - y\_in)$$

$$w_i = w_i^{Old} + \beta (t - y\_in) x_i$$

where

y_in is the analogue output

x_i is the input vector pattern i

w_i is the weight vector i

t is the target value

β is the learning constant ($0 < \beta < 1$)

b is the bias

3.2.5 Kohonen

His more recent work, in 1982, was the development of self-organising feature maps (SOM) that use a topological structure for the cluster units. Some typical applications are pattern and speech recognition. *The travelling salesman problem* is a typical problem that can be solved with the Kohonen net. See Fig 3.11 for a simplified Kohonen net [65].

3.2.6 Counter-propagation network

Robert Hecht-Nielson developed the counter-propagation network (1987a, 1987b, 1988). Compared with back-propagation, it can reduce training hundredfold. Counter-propagation provides solutions for those applications that cannot tolerate long training sessions. Counter-propagation is a combination of two well-known algorithms: the self-organising Kohonen map (1988) and the Grossberg (1969, 1971, 1982) outstar. The Grossberg outstar produces a desired excitation pattern to other neurons whenever it fires.

The outputs of the Grossberg layer are then the vector Y, equal to:

$Y = K V$

where

- Y is the Grossberg layer output vector
- K is the Kohonen layer output vector
- V is the Grossberg layer weight matrix

Methods such as counter-propagation, which combine network paradigms in building-block fashion, may produce networks more similar to the brain's architecture than any homogeneous structure. It does seem that the brain cascades various specialised modules to produce the desired computation.

If the Kohonen layer is operated such that only one neuron's output is at one and those of all the others are at zero, only one element of the K vector is non-zero. The only action of each neuron in the Grossberg layer is to output the value of the weight that connects it to the single non-zero Kohonen neuron [66].

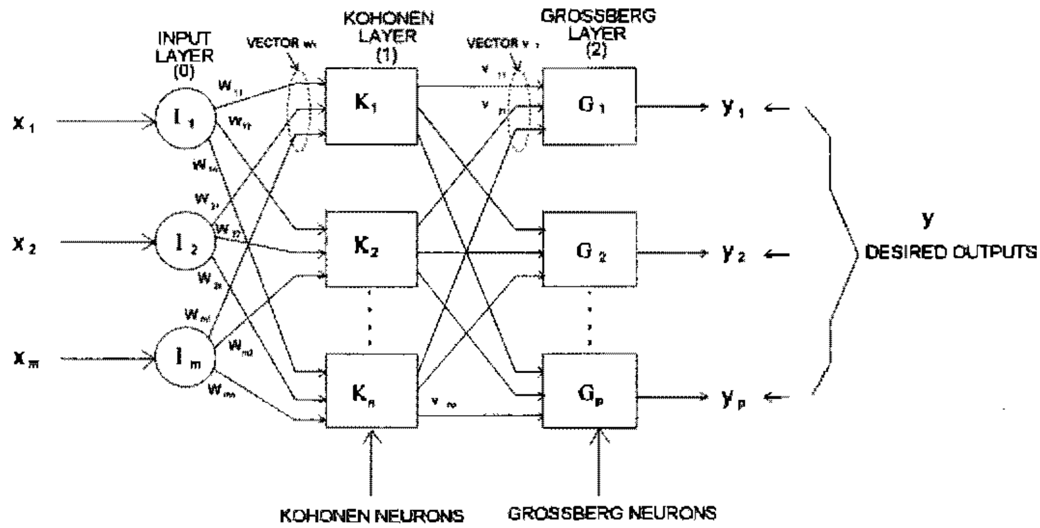


Fig 3.4: Feed-forward counter-propagation network

3.2.7 Hopfield

In 1982 John Hopfield produced a paper entitled Neural Networks and Physical Systems with Emergent Collective Computational Abilities. This paper is considered by many to have spurred the new interest in neural nets.

The approach of Hopfield is relatively novel, although it does refer to many previous results. According to him, the nervous system attempts to find stable states. Neighbouring states tend to approach a stable state, enabling errors to be corrected and providing the ability to fill in information which is missing [67].

The Hopfield net is discussed in more detail in section 3.7.

3.2.8 Back-propagation algorithm

Back-propagation has an interesting history. In 1985 Parker had already started with work on back-propagation. In 1986 Rumelhart, Hinton and Williams presented a clear description of the back-propagation algorithm. Shortly after this, Werbos (1974) was found to have described the algorithm still earlier.

The invention of the back-propagation algorithm has played a large part in the new interest in artificial neural nets. It contains a systematic method for training multilayer neural nets. Despite its limitations, back-propagation has expanded the applications of neural nets to a much wider range of problems [68].

3.3 Terminology frequently used

The following terminology is frequently used in dealing with neural nets.

3.3.1 Supervised training

Supervised training is accomplished by presenting a sequence of training vectors, or patterns, each with an associated target output vector. The weights are adjusted according to a learning algorithm in order to achieve the target output vector as closely as possible [69, 70].

3.3.2 Unsupervised training

Unsupervised training is, for example, used with self-organising neural nets. A sequence of input vectors is provided, but no target vectors are specified. The net modifies the weights so that the most similar input vectors are assigned to the same output unit [71].

3.3.3 Fixed-weight net

Fixed weights are used for constrained optimisation problems. The Boltzmann machine (without learning) and the continuous Hopfield net can be used for these types of problems. When these nets are designed, the weights are set to represent the constraints and the quantity to be maximised or minimised [71].

3.3.4 Associative memories

Human memory operates in an associative manner. If it hears only a portion of a known song, it can then produce the rest of the song. A recurrent net operates in the same manner. If an input vector has, for example, half of the data and is presented to the net, the net is sometimes capable of producing the correct output vector [72].

3.3.5 Bidirectional associative memories

The bidirectional associative memory (BAM) is hetero-associative; that is, it accepts an input vector on one set of neurons and produces a related, but different, output vector on another set. The BAM produces correct outputs despite corrupted inputs. Also, adaptive versions can abstract, extracting the ideal from a set of noisy examples [73].

3.3.6 Adaptive resonance theory

The adaptive resonance theory (ART) has the ability to learn new patterns while preventing the modification of patterns that were learned previously. The mathematics behind ART are complicated and many people have found the theory difficult to understand [74].

3.3.7 Layer pruning

In practice only one or two hidden layers are used in feed-forward multilayer perceptron nets. The net can be trained by starting with both hidden layers. If one of the layers is removed, for example to look at the viability of reducing the sensitivity of the net, the process is called layer pruning [75].

3.3.8 Neuron pruning

The number of neurons to be used in hidden layers is not known in advance. One possible approach is to construct a neural net with an excessive number of neurons in each hidden layer. When, during the training process, two neurons in the same hidden layer convey the same information, one of the two neurons must be removed. Neurons in the hidden layer(s) whose outputs are approximately constant for all training examples must also be removed. This process of removing neurons is called neuron pruning. See also section 3.6.8.

3.3.9 Generalisation

Generalisation is the ability of the net to be insensitive to variations in the input data during training and to recognise the pattern in the data despite noise and distortion [76].

3.3.10 Overtraining

Overtraining is almost the opposite of generalisation. If the net is over-trained, then small variations in the input data can result in the net's output varying significantly from the actual value.

3.4 Neural network architectures

The architectures of artificial neural nets can be divided roughly into three large categories, ie [77]:

3.4.1 Feed-forward networks

Each neuron may receive an input from the external environment and/ or other neurons, but no feedback is formed. Once trained, the output to a given input will be the same. There are no stability problems and the output(s) are often simplified to be a nonlinear mapping of the inputs.

3.4.2 Feedback (recurrent) networks

Recurrent nets have connections from their outputs back to their inputs, either modified or not modified.

3.4.3 Cellular networks

The neurons communicate directly with other neurons only in their immediate neighbourhood, by means of mutual lateral interconnections.

3.5 Basic features of neural networks

1. Neural networks are trained rather than programmed to perform a given task.
2. The state or output of one neuron affects the potential of all the neurons to which it is connected.
3. Each connection link has an associated weight, which, in a typical

neural net, is multiplied by the transmitted signal.

4. The connection weights (synaptic weights) for most of the neural networks are modified during the training phase.
5. The neurons contain typically nonlinear activation functions.
6. Neural networks are generally characterised by high robustness to noisy input data [78, 79].

3.6 Area load forecast

The use of causal forecasting models (for area load forecasts) is suggested in Chapter 6. These causal forecasting models involve the identification of other variables that are related to the variable (area load) to be predicted. Once these related variables have been identified, a neural net describing the relationship between these variables and the variable to be forecast can be developed. The relationship derived is then used to calculate the area load forecast. The area loads might be related to a number of customers' demands, economic and population growth indicators, etc. Thus the area demands are referred to as the dependent variable (output), while the other variables are referred to as the independent variables (input).

The use of a multilayer perceptron using the standard back-propagation algorithm is suggested for area load forecasts. The identification of independent variables and the development of such causal forecasting models are not covered by the scope of this dissertation. Only the theory for these types of neural nets will be discussed.

Cichocki and Unbehauen's publication entitled Neural Networks for Optimization and Signal Processing was studied for the theory on the multilayer perceptron [81].

Multilayer perceptrons consist of three types of layers: an input layer, one or more hidden layers, and an output layer. In practice only one, or maybe two, hidden layers are used. Let us first consider a single-layer perceptron, before discussing the standard back-propagation algorithm and multilayer perceptron.

3.6.1 Single-layer perceptron

A single-layer perceptron consists of a processor or an input layer, the activator and an output layer.

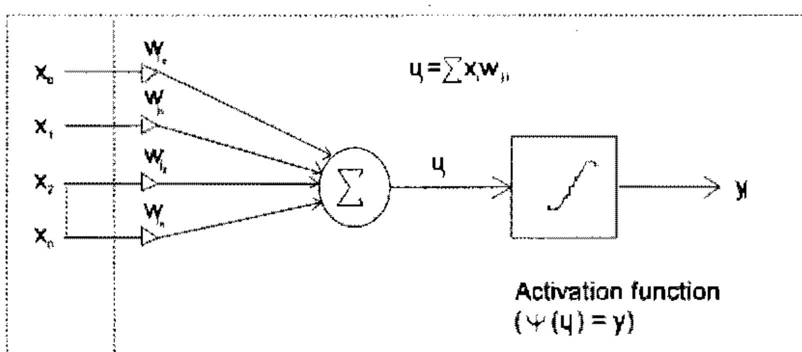


Fig 3.5: Single-layer perceptron

Some authors give different definitions of the layers, but the notation we will follow is that layer 0 is the input layer and layer 1 is the output layer. Thus we can say that this perceptron is a single-layer feed-forward net. Fig 3.5 has been simplified and redrawn as Fig 3.6.

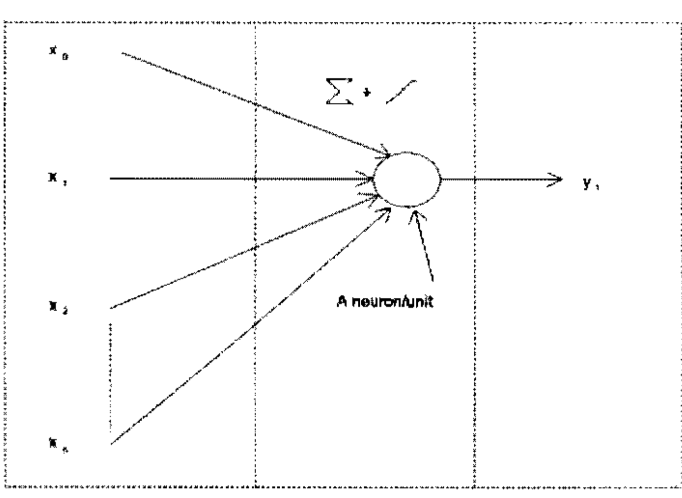


Fig 3.6: Single-layer perceptron (simplified)

3.6.2 Input layer

The input layer has a number of numerical inputs. There are no rules on, or limitations to, the number of inputs. The input x_0 is a bias with a value equal to one. To include a bias is optional - it can even be included further on in the net; depending totally on the structure of the net.

3.6.3 Data scaling

Depending on the magnitudes of the inputs and the outputs, it may be necessary to scale these inputs and outputs. It is possible that an input can be dominant, with the result that the data pattern is not well captured by the neural net, if the magnitude of that input is much larger than that of the other inputs.

The following scaling is often used:

$$\text{New}\ (X_i) = \text{Old}\ (X_i)/\text{Maximum}\ (X)$$

Scaling is not required on binary inputs.

3.6.4 Synaptic weights

Each input is multiplied by its own synaptic weight before summation with all the other weighted inputs. These weights start with very small values, generally randomly allocated between a lower and an upper value. During the training process the weights are adjusted to produce a desired output for a given input.

As previously mentioned, the initial values chosen for the weights should be rather small, because if they are too large the activation function may saturate from the beginning. The network will thus become stuck in a very flat plateau or a local minimum near the starting point.

The upper and lower values can be chosen between (0.5/units which are fed forward to the unit) and (-0.5/units which are fed forward to the unit).

3.6.5 Internal potential

The internal potential is the summated value of all the weighted input values, ie:

$$u_j = \sum_{i}^{n} w_{ji}\, x_i + \Theta_j$$

with $w_{j0} = \Theta_j$ and $x_0 = 1$

3.6.6 The activation functions

A number of activation functions can be used, again depending on the application. If the activation function is the identity function, then this net is simply a multiple regression model. A list of possible activation functions follows:

Function name	Function	Derivative of the function
Identity	$f(u) = u$	$f'(u) = 1$
Hyperbolic tangent	$f(u) = \tanh(\gamma u) = \frac{(1 - \exp(-2\gamma u))}{(1 + \exp(-2\gamma u))}$	$f'(u) = \gamma (1 - y^2)$
Unipolar	$f(u) = \frac{1}{(1 + \exp(-\gamma u))}$	$f'(u) = \gamma y (1 - y)$
Exponential	$f(u) = \exp(u)$	$f'(u) = \exp(u)$
Inverse exponential	$f(u) = 1/\exp(u)$	$f'(u) = -1/\exp(u)$
Logarithm	$f(u) = \log(u)$	$f'(u) = 1/u$

γ is the slope of the activation function.

3.6.7 Adding neurons to the output

If a second neuron, but with a different output value, is required, a second neuron is added to the output layer and all the inputs, each with its own

weight, are connected to the new neuron. The connections from the inputs to the first neuron remain unaltered.
Generally the number of outputs should not be more than the number of inputs, but neural nets with more outputs than inputs have also been trained with satisfactory results.

It is advisable to use the same activation functions for all layers, but different activation functions may also be used for different layers.

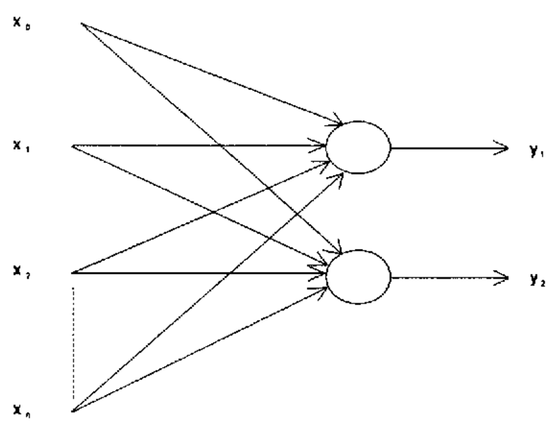


Fig 3.7: Single-layer perceptron with two outputs

3.6.8 Hidden layers

No rules exist as to how many hidden layers have to be used. We can start with one, then try a second one, but in practice we do not normally use more than two hidden layers. It is better to look at the input data and try to increase or decrease the magnitudes of the input values or add or delete some inputs. Another option is to increase or decrease the number of neurons in the hidden layers.

A guideline for determining the number of neurons per hidden layer is that it depends on the application of the neural net, as explained in the following example.

One hidden layer $m = ½ (i + o)$
Two hidden layers $n = ½ (m + o)$

where:

i is the number of input neurons

m is the number of neurons for first hidden layer

n is the number of neurons for second hidden layer

o is the number of output neurons

3.6.9 Error function

The aim of learning (training the neural net) is to minimise the instantaneous squared error of the j th output signal (y_j)

$$E_j = \frac{1}{2}(d_j - y_j)^2 = \frac{1}{2} e_j^2$$

by modifying the synaptic weights. We need to determine how to increase or decrease the synaptic weights in order to decrease the local error function E_j. This can be done by using nonlinear optimisation techniques (see section 3.6.10).

The error function ($E_j = \frac{1}{2}(d_j - y_j)^2 = \frac{1}{2} e_j^2$) has to be modified when more than one output signal is required or batch training is used instead of on-line training.

In on-line training only one sample (pattern) is applied to the net during the training process. When the pattern is applied, the error is calculated and the weights are updated. Batch training is similar, but all p patterns are first applied before the weights are updated.

Thus the error function can be summarised as follows:

Training	**One output**	**More than one output**
On-line	$E_j = \frac{1}{2}(d_j - y_j)^2$	$E_j = \sum_j \frac{1}{2}(d_j - y_j)^2$
Batch	$E_j = \sum_p \frac{1}{2}(d_{jp} - y_{jp})^2$	$E_j = \sum_p \sum_j \frac{1}{2}(d_{jp} - y_{jp})^2$

dj_p is the desired output for the net output y_j and pattern p.

3.6.10 Nonlinear optimisation techniques

The following drawing visually explains the working of the delta rule. The delta rule moves the weight vector such that its projection onto the minimum error parabola moves down the negative gradient of the bowl.

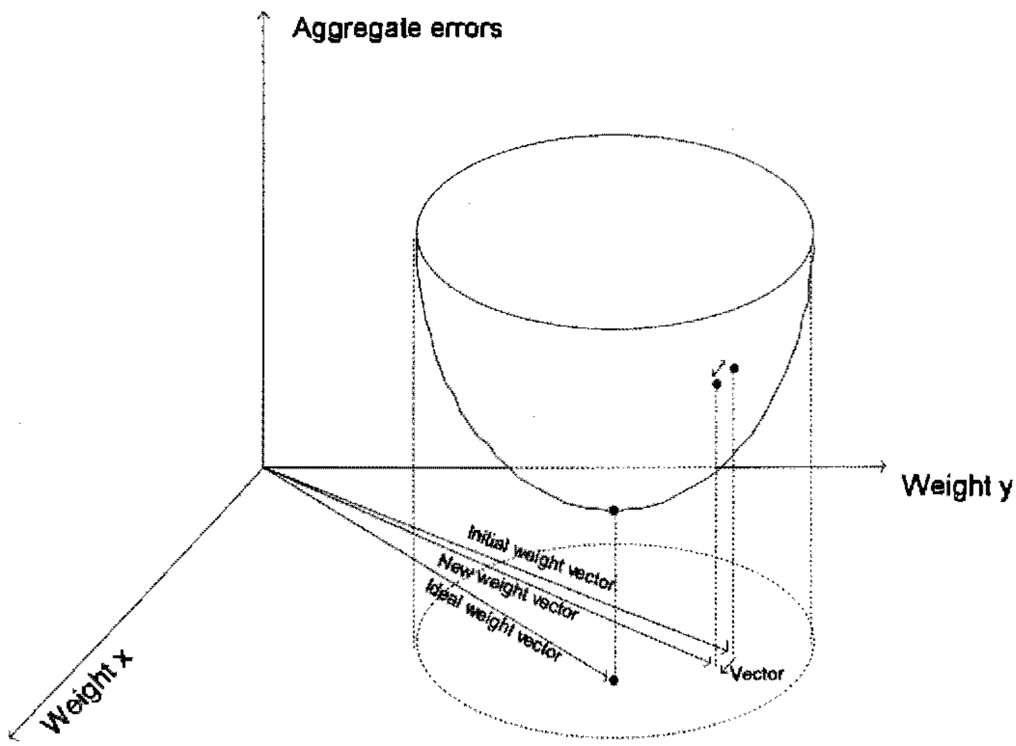


Fig 3.8: Global error surface

Four basic techniques can be used to optimise the error function:

1. Steepest-descent
2. Newton's method
3. Quasi-Newton
4. Conjugate gradient

The steepest-descent algorithm is very easy to implement, but may exhibit extremely slow convergence in some cases. Newton's method sometimes gives singularity problems in calculating the inverse Hessian matrix but has the advantage of extremely rapid convergence when the starting point is sufficiently close to the global minimum point. The quasi-Newton technique is today the most efficient and sophisticated algorithm, but it requires a lot of storage space. The conjugate gradient method requires less space than the quasi-Newton, but requires exact determination of the learning rate and the parameter ensures that the sequence of vectors satisfies a mutual conjugacy condition.

When the instantaneous squared error of the output signals is minimised, the aggregated error (Fig 3.8) is moving in the direction of the "ideal weight vector", thus to a local minimum.

3.6.11 Back-propagation algorithm for the multilayer perceptron

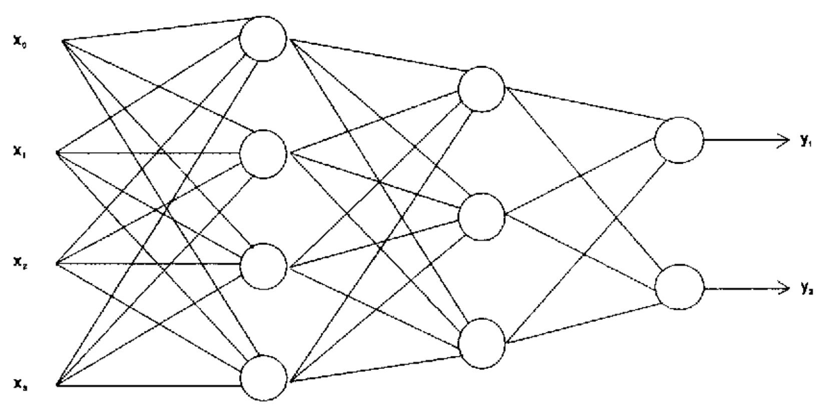


Fig 3.9: Multilayer perceptron with two outputs

As stated in section 3.6.9, the error function for a single-layer perceptron is

$$E_j = ½ (d_j - y_j)^2 = ½ e_j^2$$

and the instantaneous squared error of the output signals can be minimised by modifying the synaptic weights w_{ji}. The steepest-descent gradient rule can be used to increase or decrease the weights to decrease the local error function E_j. Thus the steepest-descent gradient rule can be expressed as:

$d\, w_{ji}/dt = -\mu\, \partial E_j/\partial w_{ji}$ (μ is the learning rate) (1)

Using the chain rule, $d\, w_{ji}/dt$ becomes

$d\, w_{ji}/dt = -\mu(\partial E_j/\partial e_j) (\partial e_j/\partial w_{ji})$ (2)

and

$d\,w_{ji}/dt = -\mu(\partial E_j/\partial e_j) (\partial e_j/\partial u_j) (\partial u_j/\partial w_{ji})$ (3)

The error function is given as:

$E_j = ½\, e_j^2$, therefore

$\partial E_j/\partial e_j = e_j$ (4)

The error is the difference between the actual output and the desired output, ie

$e_j = d_j - y_j$ and

$y_j = \psi (u_j)$, the activation function output

Thus

$e_j = d_j - y_j = d_j - \psi (u_j)$ and

$\partial e_j/\partial u_j = - \psi' (u_j)$ (5)

The internal potential is the summated value of all weighted input values, ie

$u_j = \sum w_{ji}\, x_i + \Theta_j$ and

$\partial u_j/\partial w_{ji} = x_i$ (6)

Replace (4), (5) and (6) in (3)

$d\,w_{ji}/dt = \mu\, e_j\, \psi' (u_j)\, x_i$ (7)

If we choose a hyperbolic tangent activation function, then

$\psi' (u_j) = \gamma_j (1- y_j^{\,2})$ (γ_j is the slope or steepness) and

$d\,w_{ji}/dt = \mu\, e_j\, \gamma_j (1- y_j^{\,2})\, x_i$ (8)

Replace $\delta_j = e_j\, \psi' (u_j) = e_j\, \gamma_j (1- y_j^{\,2})$, then (8) becomes

$d\,w_{ji}/dt = \mu\, \delta_j\, x_i$ (9)

For multilayer perceptrons the change of weights is calculated in a similar way.

Let us start with the weights for the output layer, and given that there are two hidden layers

$[d\,w_{ji}/dt]^{[3]} = \mu\, \delta_j^{[3]}\, o_i^{\,[2]}$ where $o_i^{\,[2]}$ is the i th output from layer two (10)

and $\delta_j^{[3]} = e_{jp}\, (\psi_j^{[3]})'$

The change of weights for the hidden layers is similar, except for the δ_j.

For the second hidden layer

$$[d\, w_{ji}/dt]^{[2]} = \mu\, \delta_j^{[2]}\, o_i^{[1]} \quad \text{with}$$

$$\delta_j^{[2]} = \psi'\,(u_j)^{[2]} \sum_{i}^{n_3} \delta_j^{[3]}\, w_{ij}^{[3]}$$

and n_3 is the number of output neurons.

For the first hidden layer

$$[d\, w_{ji}/dt]^{[1]} = \mu\, \delta_j^{[1]}\, x_i \qquad \text{with}$$

$$\delta_j^{[1]} = \psi'\,(u_j)^{[1]} \sum_{i}^{n_2} \delta_j^{[2]}\, w_{ij}^{[2]}$$

and n_2 is the number of neurons in the second hidden layer.

In practice, for batch training, the back-propagation learning algorithm usually takes a more sophisticated form. Besides the learning rate factor, the algorithm has two other factors, ie the momentum and the decay factors.

The momentum factor improves the convergence rate and the steady-state performance of the algorithm. The correct setting of the momentum coefficient α ($0 \leq \alpha < 1$) helps the algorithm to get out of local minimums.

The decay factor γ (with typical values of 10^{-3} to 10^{-5}) prevents the algorithm from generating very large weights.

Let p be the number of samples (patterns) available to train the net; k the number of epochs (iterations) for which the algorithm has to be trained; and η the learning factor. The formula is

$$\triangle\, w_{ji}^{[s]}\,(k) = \eta\, / n_{s-1} \sum_{p\, \varepsilon\, \text{pattern set}} \delta_{jp}^{[s]}\, o_{ip}^{[s-1]} + \triangle\, w_{ji}^{[s]}\,(k-1) - \gamma\, w_{ji}^{[s]}\,(k)$$

η is the learning factor and n_{s-1} is the number of processing units in the (s - 1)-th layer, where s = 1, 2, 3.

3.7 Power generation pattern

The generation pattern algorithm is quite difficult to automate. Firstly the outputs should not exceed their maximum values (available capacity); some linear programming constraints for neighbourhood search have to be modelled. Previous generation trends have to be evaluated, together with the required system peak and the individual areas' corresponding peaks. Furthermore, the generation costs should also be added as a constraint. Thus it is a typical linear programming problem, having a cost function and a number of constraints (or, in terms of neural nets, a constrained optimisation problem).

Neural nets have been used to solve such problems. An energy function has to be created for the appropriate neural net, ie converting the constraints into penalty functions and adding the results to the cost function, thereby creating a new function called an energy function.

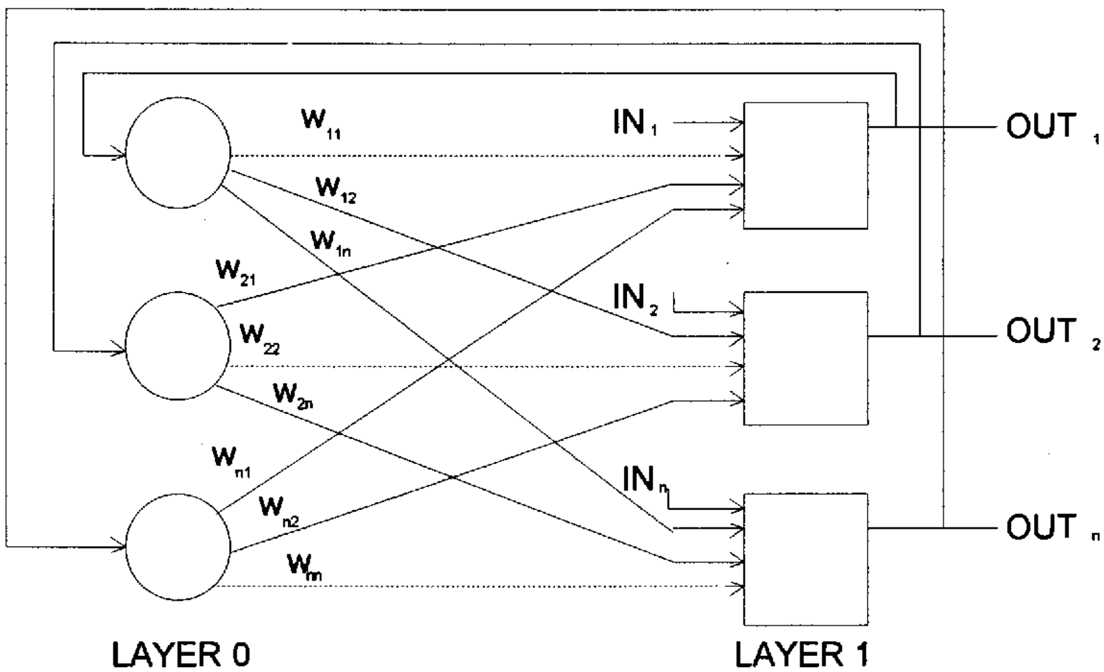
Recurrent neural nets such as Hopfield nets have been used. The following theory on Hopfield nets might be helpful.

3.7.1 Hopfield neural net

The Hopfield nets are recurrent nets, ie nets that have feedback paths from their outputs back to their inputs. The response of such networks is dynamic; that is, after applying a new input, the output is calculated and fed back to modify the input. The output is then recalculated and the process is repeated. For a stable network, successive iterations should produce smaller and smaller output changes until the outputs eventually become constant.

For many networks the process never ends, and such networks are said to be unstable. As a matter of interest, according to P Wasserman unstable networks have interesting properties and have been studied as examples of chaotic systems [81].

The weight matrix W (n x n matrix) is stable when all diagonal elements are zeros ($w_{ii} = 0$) and all other elements adhere to $w_{ij} = w_{ji}$.



Dotted lines indicate weights of zero

Fig 3.10: Hopfield net

In Hopfield's early work in 1982, the activation function was a simple threshold. The output of such a neuron is one if the weighted sum of the outputs of the other neurons is greater than a threshold T_j; otherwise it is zero. Later, in 1984, Hopfield showed other cases in which the activation function was continuous.

Hopfield nets have been used to solve the travelling salesman problem (TSP). As a linear programming problem, the TSP consists of a cost function and a number of constraints. These constraints are required to avoid subtours and to ensure that all cities are visited but any one city is visited only once. The tour will start at one of the cities and end at the same city. The constraints have been converted to penalty functions and added to the cost function, ie the energy function for the Hopfield net [82].

3.8 Pattern recognition nets.

In most cases it is difficult to identify system operating conditions that are not normal. If the electrical loads are grouped, ie the respective years' loads as vectors and the number of input and output electrical loads to the forecasting model as elements of a given vector, then certain neural nets can be used to cluster those various vectors. One such neural net is the Kohonen net.

3.8.1 Kohonen net

The principle of the Kohonen net is "winner takes all'. For example, when we have m n-dimensional vectors, how can the Kohonen net be used to cluster these m vectors?

The number of inputs into the Kohonen net is equal to the vectors' dimension to be clustered, which is n in our example. Each input, through its own synaptic weight, is connected to a "topographical feature map". This map is like a grid, usually two-dimensional, but sometimes a three-dimensional grid is used. In our example we have m vectors, which means, at most, m clusters. Therefore, in the case of a two-dimensional grid, a p x p space has to be used, with p larger than m, depending on the problem to be studied.

Each input is connected to each of the p x p connections (D (j)) on the feature map. Each connection has its own weight, thus a total number of weights of p x p x n. Before training the net on the m number of vectors, all the synaptic weights have to be initialised. This can be done by randomly allocating a value to each weight between certain specified limits.

When the first vector is applied as an input to the net, the node on the feature map with the minimum Euclidean distance (between the weights and the corresponding inputs) is selected as the winner. Its weights are adjusted and a number of other nodes in its neighbourhood, depending on the neighbourhood parameter, are also adjusted.

Figure 3.11 shows the Kohonen net with its n inputs and two-dimensional p x p feature map.

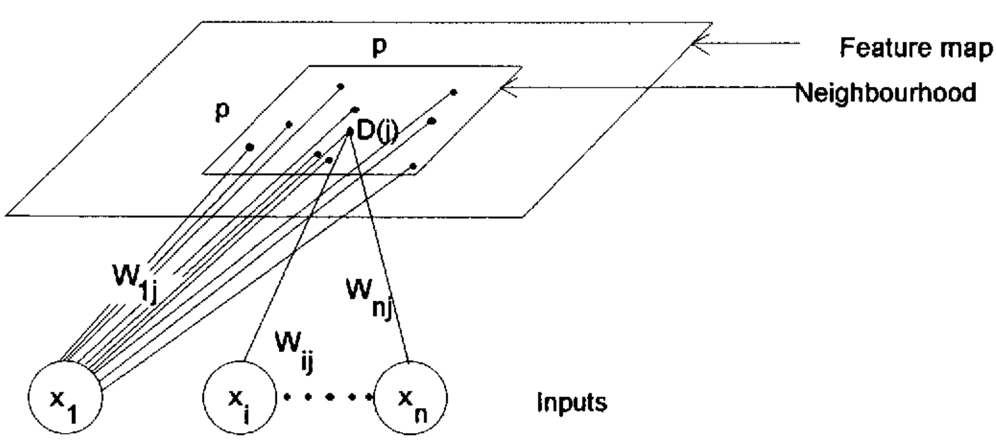
Define the node on the grid as D(j) and the synaptic weights as w_{ij} with $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p^2$.

The winning node is the node with the minimum value for $\sum_i (w_{ij} - x_i)^2$.

The update of the weight for the winning node and for the weights in the neighbourhood is as follows:

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha\,[x_i - w_{ij}(\text{old})]$$

where α (normally: $0 < \alpha \leq 1$) is the learning rate which is updated (decreasing) during the training process (unsupervised).
[24, 25, 26]



Note: D(j) is a connection on the feature map

Fig 3.11: Kohonen net

It would be worthwhile to study Kohonen and Hopfield nets to identify transmission network patterns, especially those of unrecorded events. If such patterns can be identified, better results may be obtained in forecasting the substations' load demands.

3.9 Missing data

In Chapter 6 the point loads are obtained by balancing the electrical loads, multiplied by scaling factors, with the lower transmission substations. Given that the project for the point load metering (stats metering points according to the project terminology) has been completed, it would be worthwhile to train a neural net and retrospectively verify the estimated point loads obtained by the forecasting model. If satisfactory results can be obtained, many more years' data will be available to balance the forecasting model.

A typical net that could be considered is a bidirectional associative memory (BAM) net. The BAM is hetero-associative; that is, it accepts an input vector on one set of neurons and produces a related but different output vector on another set. BAM is capable of generalisation, producing correct outputs despite corrupted inputs.

3.10 Network changes

The ART nets, discussed in section 3.3.6, are a possible solution in cases where electrical network changes would change the load allocation in the forecasting model. As previously mentioned, an ART net can be learned by preventing modification to the existing learned patterns. More detailed research is required for possible implementation in the forecasting model, therefore it is not within the scope of this dissertation.

3.11 Conclusion

As shown in the results set out in Chapter 5, it follows that neural nets cannot be properly trained to forecast the substation demands. This is because only 40 data points have been used to train the nets on. Depending on the number of inputs, at least 10 years' data is required, but that is not available.

If appropriate inputs are available, neural nets could possibly be used for area demand forecasts. Neural nets could also be considered for constrained linear programming applications, for example the generation nett sent out, which should not exceed certain upper limits; or for the diversity factors, which have to be constrained between upper and lower limits.

In Chapter 6 it is stated that if iterative algorithms could be developed, the forecasting model would need very little human intervention in the forecasting process. From the theory discussed so far, especially the theory on perceptrons, it may be possible to develop those iterative algorithms. Also, a number of summations and multiplications similar to neural nets are being used in the forecasting model. Sections 3.6, 3.7, 3.8, 3.9 and 3.10 show that neural nets have some advantages that may be considered as possible techniques to improve the forecasting model and reduce human (error-prone) intervention.

In future neural net studies will therefore be conducted to develop the iterative algorithms, design a neural net to balance the electrical loads and reduce human intervention to a minimum.

4

CHAPTER

QUALITATIVE FORECASTING

4.1 Introduction

Possible future events which may have an impact on substation demands should be identified before they occur, otherwise it may be too late to expand the electrical network(s) accordingly. The forecaster should therefore explore the future trends and developments to identify emerging factors of influence. For example, before a new zinc smelter can be erected in an area, electrical network studies have to be done to ensure that the electrical networks are able to supply the load. In some cases network strengthening may be required, which could take up to three years or even longer if new line servitudes are required. Thus there is a need for long-term forecasting techniques identifying possible growth in different areas. Sometimes only possible electricity demand growths will be identified, but not any specific types of loads.

Some possible long-term forecasting techniques will be discussed in the following paragraphs. References have, however, been included for more in-depth studies.

4.2 Possible area demand forecast model

In 1977 Dr Keewhan Choi used a regression model to make forecasts for a large electric utility company in the United States, containing annual system peak demand for the years 1977 to 1990 [83]. The company provides electricity to an entire state. The forecasts were used to apply for increased service rates and for the construction of a new power-generating plant.

Using the following regression (econometric) model, the company forecast the annual system peak demand for the years 1977 to 1990 from the annual peak demand data for the years 1960 to 1976:

$$\log D = \log A + B \log (GSP) + C \log (TPY)$$

where D is the "weather-normalised" peak demand, GSP the gross state product and (TPY) the total personal income in the state for the previous year. The parameters A, B and C were estimated from the past data on "weather-normalised" peak demand, gross state product and total personal income by the ordinary least squares method.

For South Africa, the summated larger customers' demands can be used instead of the area maximum demand. To increase either the gross state product or the total personal state income in an area, new developments are required in the area, and this in turn means the possibility of increased or new electricity demands.

4.3 S-curves

Many researchers have shown that the population growth tends to follow a pattern similar to that of the growth of biological organisms [84]. The same conclusion was reached by Gompertz and Von Bertalanffy [84]. The S-type curve can be used in predicting population growth patterns and their point of stability by using analogies which compare historical patterns with existing situations in order to forecast future progress and developments. The concept of analogy between biological and other kinds of growth has recently been applied to such phenomena as the growth pattern of particular technologies, transportation speeds, the life cycle of individual products, and the growth of government spending [84].

Growth analogies are similar to S-shaped curves, except that no prediction of parameters is required, because in growth analogies the curve is considered known and the purpose is to fit the data to the curve rather than vice versa. This application allows the curve to be used in predicting time estimates for the occurrence of various events in the future. With growth curves, as with S-curves, saturation will eventually be reached and identification of when that will occur is very important in determining the appropriate growth analogy. Whether such saturation is more descriptive of the future than an increasing rate of growth, like that depicted by exponential curves, is a matter that is still open to debate.

S-type curves were used successfully to predict the introduction of new commercial-type aeroplanes [85]. The same approach can be explored to establish, for example, whether population growth and total income in the area is of such a nature that new developments are required to provide jobs and to improve total personal income.

4.4 Scenario development and the Delphi technique

Two other exploratory methods of technology forecasting are scenario development methods and the Delphi approach. Makridakis describes these methods as follows. Scenario writing takes a well-defined set of assumptions, then develops an imaginative conception of what the future would be like if these assumptions were true [86]. In this sense, scenarios are not future predictions in themselves. Rather, they represent a number of possible alternatives, each one based on certain assumptions and conditions. It is then up to the decision-maker to assess the validity of the assumptions in deciding which scenario is most likely to become a reality.

The objective of the Delphi approach is to obtain a reliable consensus of opinion from a group of experts that can be used as a future forecast, while at the same time minimising the undesirable aspects of group interaction [87].

Two of the main developers of the Delphi approach, Helmer and Rescher, have described the Delphi method as follows [88]:

"The Delphi technique eliminates committee activity altogether, thus further reducing the influence of certain psychological factors, such as specious persuasion, the unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion. This technique replaces direct debate by a carefully designed program of sequential individual interrogations (best conducted by questionnaires), interspersed with information and opinion feedback derived by computer consensus from the earlier parts of the program. Some of the questions directed at respondents may, for instance, inquire into the 'reasons' for previously expressed opinions and a collection of such reasons may then be presented to each respondent in the group, together with an invitation to reconsider and possibly revise his or her earlier estimates."

Much of the work on scenario writing was done by Kahn of the Hudson Institute. Kahn (1964) and Kahn et al (1976) developed a number of alternative scenarios for the world [89]. In one scenario he assumed that an arms control agreement between the United States and Russia would be reached and that China would follow only a defensive policy rather than an offensive policy. Based on these and other assumptions, he developed a scenario describing a future political-social environment, following a predictable sequence of developments, constraints and ideologies. From this set of assumptions, Kahn predicted a number of scenarios. One was that Russia would lose control over the world communist movement.

A number of evaluative studies have been made to summarise experiences with the Delphi technique and some of its advantages and disadvantages. One of the most thorough of these is that prepared by Sackman (1975) [90]. He not only attempts to evaluate the Delphi approach, but also seeks to describe some of its variations and some of the alternative procedures that have been found to best overcome the shortcomings of the method originally developed. Another work by, Linstone and Turoff (1975), also added to the descriptive literature on the Delphi approach and expanded discussion of its practical applications.

These two methods can be used to assess more accurately the diversity factors and demands for given conditions to forecast future substation demands.

4.5 Cross-impact matrix

A technological method of forecasting closely related to both the Delphi method and the use of scenario development is that of cross-impact matrices. A number of papers have reported applications of this methodology [91]. A cross-impact matrix describes two types of data for a set of possible future developments. The first type estimates the probability that each development will occur within a specified time period in the future. The second estimates the probability that the occurrence of any one of the potential developments will have an effect on the likelihood of occurrence of each of the others. The data for such a matrix can generally be obtained by using either subjective assessment procedures or a method such as the Delphi approach.

The aim of cross-impact analysis is to refine the probabilities relating to the occurrence of individual future developments and their interaction with other developments to the point where these probabilities can be used either as a basis for planning or as a basis for developing scenarios that can subsequently be used in planning.

Cross-impact analysis can thus be an important tool for the forecaster to determine the probabilities of occurrence of the diversity factors and the expected diversity value for a given probability. This can also be applied to the POD future demands and probability of occurrence. This may enable the forecaster to identify future trends and incorporate these trends into the proposed forecasting model.

4.6 Curve fitting

As a technological forecasting method, curve fitting can also be considered [92]. In the case of technology forecasting the time horizon is generally much longer, only a limited number of data points are available, some rather tenuous assumptions must be made, and an interpretation of results is required. However, in some instances this approach can be most helpful to the forecaster.

4.7 Relevance trees

The origin of relevance trees dates back to the development of decision theory and the construction of decision trees aimed at aiding the decision-maker in selecting the best strategy out of a number of alternatives [93]. Relevance trees help to identify the long-range developments that are most important for the accomplishment of specific objectives.

4.8 Catastrophe theory

Most forecasting methodologies look at average values as the basis for prediction. They generally assume that randomness consists of a large number of unimportant factors that can take on either positive or negative values. The result of these numerous random elements is a unimodal distribution. Such random elements are generally ignored because their expected values are assumed to equal zero.

One of the main rationales for the catastrophe theory is that such assumptions about randomness are inappropriate in a number of situations [94]. The catastrophe theory assumes that bimodality (or, in general, multimodality) very often exists in the real world. In such instances the outcome that is observed has as much chance of moving toward one modal point as toward another, so that on average it is very unlikely that it will fall halfway between those modes. In fact, the average outcome is extremely unlikely to occur, because the average is not one of the modes.

The Zeeman (1976) report on the use of the catastrophe theory is one of the most complete descriptions on this subject [94]. Using a number of graphics and concepts from topology, Zeeman gives several different examples of phenomena that exhibit bimodality or multimodality.

The next three sections are taken from the SBL seminar: Forecasting with Microcomputers.

4.9 Technology forecast (most important techniques)

The most important techniques used in this field can be summarised as follows [95]:

Expert opinion	Interview Delphi Nominal group technique
Extrapolation of trends	Parameters Analogies Leading indicators S-curve chains Substitution curves
Normative techniques	Opinion surveys Morphological matrices Implication wheels Relevance trees
Surveillance	Scanning, monitoring, tracking
Dynamic modelling	
Cross-impact analysis	Cross-impact analysis is a means of analysing the impacts that a series of

possible events have on one another.

Scenarios

4.10 Stages of technological progress

Eight stages have been identified for technological progress [96]:

1. **Origin**
2. **Proposal of the concept**
3. **Verification**
4. **Laboratory demonstration**
5. **Field trial**
6. **Commercial introduction**
7. **General acceptance**
8. **Proliferation** (also applied in other fields)

4.11 Surveillance

Surveillance can be defined in terms of three activities, ie scanning, monitoring and tracking [97]:

1.	**Scanning**	Scanning is used to identify, at an early stage, developments in the social, economic, political and ecological environments. It is a general, wide-ranging and ear-on-the-ground activity.
2.	**Monitoring**	Monitoring focuses on a specific range of signals of interest as identified by scanning, and aims to establish where the developments they indicated are going. It is best done within the more formal analysis and planning function of an organisation.

3.	**Tracking**	Tracking is the very specific and intense activity aimed at finding out the exact rate, direction and impact of the few really important developments. Tracking should be linked to the formal strategy formulation function in the organisation.

4.12 Spreadsheet: tracking events

The scanning of newspapers; the monitoring of signals of interest as identified by scanning; and the tracking of certain specific activities are an important aspect of the electrical demand forecast. Therefore this temporary arrangement on spreadsheets is done to keep track of important events. The system may look very simple, but it is effective and gives results which add value to the electrical demand forecast.

The columns and rows are respectively marked alphabetically and numerically. All subjects on which information is available are listed in column B, rows 2 to 9.

The articles are numbered sequentially, and the number allocated to an article can appear next to one or more subjects (see, for example, article 134). To keep track of the number to be allocated next, use the following statement in a reference cell on a separate sheet: "=max (range article numbers are stored (for example C2:L9)) + 1".

The underlined digits in row 1 and column B serve as counters for each subject. For example, the subject "Power stations" has only one article, ie number 22. The article number is given in column D; the underlined number in D1 is 1; add that to the highest number in column B, next to the subject (Power stations), which is 0 - thus there is only one article. Similarly, "Gold mines" has a number 2 in E1; add that to B9, and the number of articles will equal 12.

For example, if an article or newspaper contains some valuable information on a gold mine in Mpumalanga, it can be stored as follows: First look for an indicator which will give a new record number (= 134) for the next article. Then the new article number (134) is added next to the "Mpumalanga development" and "Gold mines" subjects. Once this number has been added, the indicator value will increase to 135.

If some information is required on Aids, it will be evident from the table that only article 2 contains information on Aids. The articles are filed sequentially in files, with dividers after every 25 articles and with 150 articles per file.

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Subjects**		0	1	2	3	4	5	6	7	8	9
2	**National Party**	0		1	24							
3	**Mpumalanga development**	0		3	4	6	7	12	134			
4	**Aids**	0		2								
5	**Power stations**	0		22								
6	**Midrand**	0		5								
7	**Cement**	0		8	9							
8	**Gold mines**	0		13	14	15	17	18	19	23	45	56
9		10	99	101	134							

4.13 Conclusion

In conclusion, long-term forecasting techniques can be seen as complementary to the proposed forecasting model discussed in Chapter 6. These techniques can play an important role in the area forecasts. Also, possible future technology developments, as well as social, ecological, political and economic changes and their impact on the electrical network demands, can be identified for future network expansion(s). Technology forecasting can prevent unpleasant surprises which could cause substation demands to exceed their firm capacities. Technology forecasting is also a valuable tool in the decision-making process.

5

CHAPTER

SUBSTATION FORECAST

5.1 Introduction

The independent variables used to predict the substation maximum demands (MW) were the account numbers' monthly maximum demands. The monthly maximum substation demands are only available from January 1993, leaving 40 data points for building a regression model or to train a neural net, and another 4 points for validation. Ten substations' forecast results have been included for discussion.

In some cases **15** independent variables are used for regression and as a rule of thumb more than ***15 times 10*** (ie **150**) data points have to be used. Neural nets require even more data, especially where the variances change significantly from one observation to another (see the forecast results).

The substation maximum demand does not necessarily occur in the same month as the account number demands. The peak demands of the various account numbers also occur in different months.

Another shortcoming is the forecast POD demands, which are annual maximum figures; thus a model has to be built not on monthly demands, but on annual maximum demands. This will limit the number of data points to **only 4.**

In one of the examples an account number has three PODs, each POD being fed from a different transmission substation, but fortunately each of those three substations is feeding only that particular POD. Therefore it would be better to use the substation's hourly demand readings to predict the annual substation demands. In Table 5.1 the account number demands and the three substation demands are given. The forecast results of one of the substations (number 9) have been included for discussion later on. The Tot Subs column gives the summation of the three monthly substation maximum demands, eg for Jan 93: 499.60 + 204.00 + 239.00 = 942.60.

	Acc Number	Sub 11	Sub 12	Sub 9	Tot Subs
Jan 93	871.17	499.60	204.00	239.00	942.60
Feb 93	844.99	461.60	277.60	244.40	983.60
Mar 93	886.35	459.20	318.00	231.00	1008.20
Apr 93	870.02	497.20	282.40	237.40	1017.00
May 93	1023.66	556.00	283.00	218.60	1057.60
Jun 93	1222.54	622.00	285.00	221.00	1128.00
Jul 93	1082.32	502.00	302.40	248.40	1052.80
Aug 93	1012.89	436.40	276.00	203.60	916.00
Sep 93	895.08	431.20	238.40	224.40	894.00
Oct 93	811.27	413.60	226.28	222.20	862.08
Nov 93	773.55	339.60	198.82	206.00	744.42
Dec 93	681.08	295.00	197.00	200.00	692.00
Jan 94	629.71	301.60	211.24	183.40	696.24
Feb 94	669.45	328.00	217.78	167.40	713.18
Mar 94	701.93	351.80	223.01	158.30	733.11
Apr 94	655.49	326.40	205.36	117.00	648.76
May 94	720.59	334.20	251.79	153.70	739.69
Jun 94	877.49	415.80	289.07	184.40	889.27
Jul 94	1045.66	495.20	339.43	193.40	1028.03
Aug 94	857.08	399.00	291.68	167.95	858.63
Sep 94	804.51	378.00	296.26	126.78	801.04
Oct 94	759.23	388.20	302.80	135.90	826.90
Nov 94	736.23	319.20	296.82	117.10	733.12
Dec 94	759.05	281.00	295.61	169.70	746.31
Jan 95	765.70	342.00	302.80	124.40	769.20
Feb 95	665.76	306.80	277.30	106.30	690.40

Table 5.1: MW demands

	Acc Number	Sub 11	Sub 12	Sub 9	Tot Subs
Mar 95	699.56	337.80	302.15	124.30	764.25
Apr 95	714.37	308.40	296.26	136.80	741.46
May 95	864.57	463.60	336.81	157.90	958.31
Jun 95	989.05	472.80	345.31	172.00	990.11
Jul 95	1023.98	492.00	321.77	179.50	993.27
Aug 95	807.31	442.80	273.37	180.45	896.62
Sep 95	779.86	433.20	268.79	148.95	850.94
Oct 95	730.49	442.40	247.87	131.90	822.17
Nov 95	760.62	386.20	229.55	126.50	744.25
Dec 95	843.76	400.20	282.53	150.40	833.13
Jan 96	825.16	427.20	285.14	171.03	883.37
Feb 96	853.81	465.60	275.33	241.35	982.28
Mar 96	783.06	429.60	281.87	177.90	889.37
Apr 96	759.99	409.80	333.54	192.45	935.79
May 96	920.79	510.00	371.40	245.10	1126.50
Jun 96	1019.93	461.80	388.20	266.55	1116.55
Jul 96	1120.29	534.60	406.80	321.60	1263.00
Aug 96	970.50	440.40	338.16	274.23	1052.79

Table 5.1: MW demands (cont)

Table 5.2 gives the results of the three substation demands, as well as the Tot Subs demand divided by the account number demand for each month.

	Sub 11/ Acc	Sub 12/ Acc	Sub 9/ Acc	Tot Subs/ Acc
Jan 93	0.57	0.23	0.27	1.08
Feb 93	0.55	0.33	0.29	1.16
Mar 93	0.52	0.36	0.26	1.14
Apr 93	0.57	0.32	0.27	1.17
May 93	0.54	0.28	0.21	1.03
Jun 93	0.51	0.23	0.18	0.92
Jul 93	0.46	0.28	0.23	0.97
Aug 93	0.43	0.27	0.20	0.90
Sep 93	0.48	0.27	0.25	1.00
Oct 93	0.51	0.28	0.27	1.06
Nov 93	0.44	0.26	0.27	0.96
Dec 93	0.43	0.29	0.29	1.02

Table 5.2: MW demands

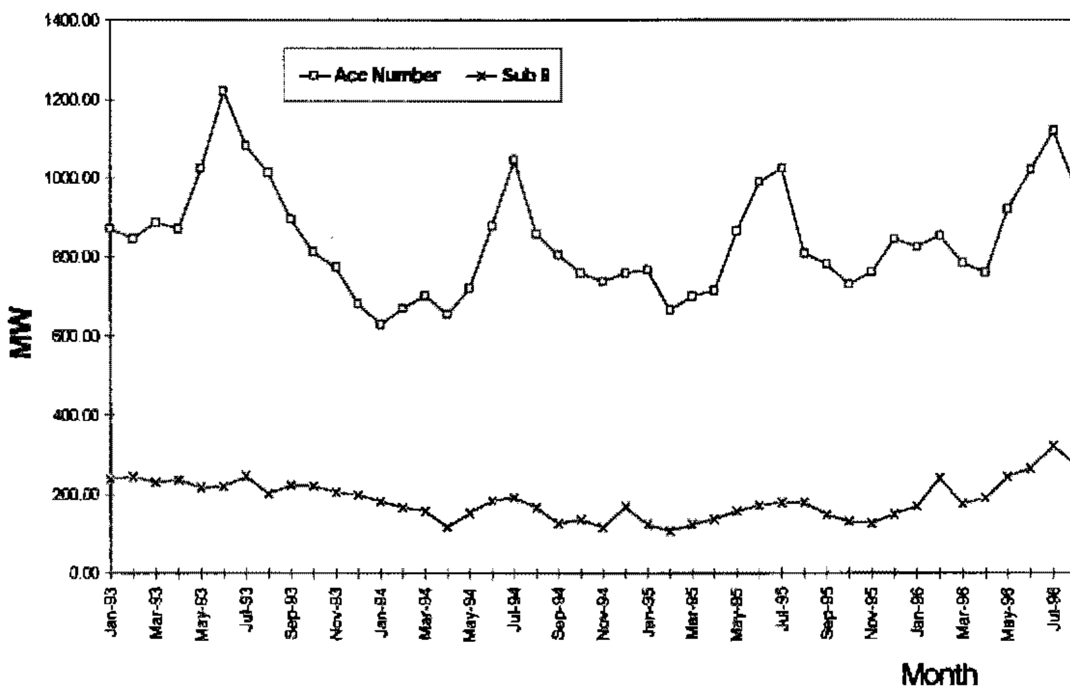
	Sub 11/ Acc	Sub 12/ Acc	Sub 9/ Acc	Tot Subs/ Acc
Jan 94	0.48	0.34	0.29	1.11
Feb 94	0.49	0.33	0.25	1.07
Mar 94	0.50	0.32	0.23	1.04
Apr 94	0.50	0.31	0.18	0.99
May 94	0.46	0.35	0.21	1.03
Jun 94	0.47	0.33	0.21	1.01
Jul 94	0.47	0.32	0.18	0.98
Aug 94	0.47	0.34	0.20	1.00
Sep 94	0.47	0.37	0.16	1.00
Oct 94	0.51	0.40	0.18	1.09
Nov 94	0.43	0.40	0.16	1.00
Dec 94	0.37	0.39	0.22	0.98
Jan 95	0.45	0.40	0.16	1.00
Feb 95	0.46	0.42	0.16	1.04
Mar 95	0.48	0.43	0.18	1.09
Apr 95	0.43	0.41	0.19	1.04
May 95	0.54	0.39	0.18	1.11
Jun 95	0.48	0.35	0.17	1.00
Jul 95	0.48	0.31	0.18	0.97
Aug 95	0.55	0.34	0.22	1.11
Sep 95	0.56	0.34	0.19	1.09
Oct 95	0.61	0.34	0.18	1.13
Nov 95	0.51	0.30	0.17	0.98
Dec 95	0.47	0.33	0.18	0.99
Jan 96	0.52	0.35	0.21	1.07
Feb 96	0.55	0.32	0.28	1.15
Mar 96	0.55	0.36	0.23	1.14
Apr 96	0.54	0.44	0.25	1.23
May 96	0.55	0.40	0.27	1.22
Jun 96	0.45	0.38	0.26	1.09
Jul 96	0.48	0.36	0.29	1.13
Aug 96	0.45	0.35	0.28	1.08

Table 5.2: MW demands (Cont)

This example proves that it is not possible to use the diversified demands to predict the future demands on substation 9. The maximum, minimum and average values for each of the four columns in Table 5.2 are given in Table 5.3. The ratio (Substation/Account number demand) for substation 9 varies between 0.29 and 0.16, with an average of 0.22. This is also proved by Graph 5.1. The R squared for substation 9 is 0.32 (Table 5.14), which also proves that it is not possible to use the diversified demands to predict the future demands on substation 9.

	Sub 11/ Acc	Sub 12/ Acc	Sub 9/ Acc	Tot Subs/ Acc
Maximum	0.61	0.44	0.29	1.23
Minimum	0.37	0.23	0.16	0.90
Average	0.49	0.34	0.22	1.05

Table 5.3: Maximum, minimum and average values



Graph 5.1: Account number vs substation 9

5.2 Autocorrelation

The Durbin-Watson test results prove that autocorrelation exists in most cases. This confirms that more independent variables have to be included. If those variables are available, it will only worsen the problem with the number of samples available.

Table 5.4 summarises the Durbin-Watson test results for the those substations.

Substation	Autocorrelation results
1	Positive autocorrelation
2	Positive autocorrelation
3	Positive autocorrelation
4	Results are indeterminate
5	Positive autocorrelation
6	No autocorrelation
7	Positive autocorrelation
8	Positive Autocorrelation
9	**Positive autocorrelation**
10	Results are indeterminate

Table 5.4: Autocorrelation results

5.3 Multicollinearity

Multicollinearity is also a concern - in cases where two or more regressor variables are used, some of the regressor variables are negative. This would mean that if the customer increased its demand, the demand on the substation would decrease. Obviously that cannot be true.

Substation	Number of regressors	Multicollinearity results
1	6	Yes
2	6	Yes
3	2	No
4	4	Yes
5	1	---
6	16	Yes

Table 5.5: Multicollinearity results

Substation	Number of regressors	Multicollinearity results
7	1	---
8	3	Yes
9	1	---
10	7	Yes

Table 5.5: Multicollinearity results (Cont)

5.4 Heteroscedasticity

No heteroscedasticity tests have been done, but when looking at the validation part of almost all the graphs it is clear that the predictions are moving away from the actual values. This is the funnel effect described in Chapter 2 (page 29). This is a violation of the assumption of homogeneity of the error variances. The result of heteroscedasticity is that at the forecasting horizon, depending on the degree of heteroscedasticity, the predicted values could be very high or even very low relative to the actual values at that time. Depending on the actual observation values, a negative prediction value could occur.

5.5 Test statistics

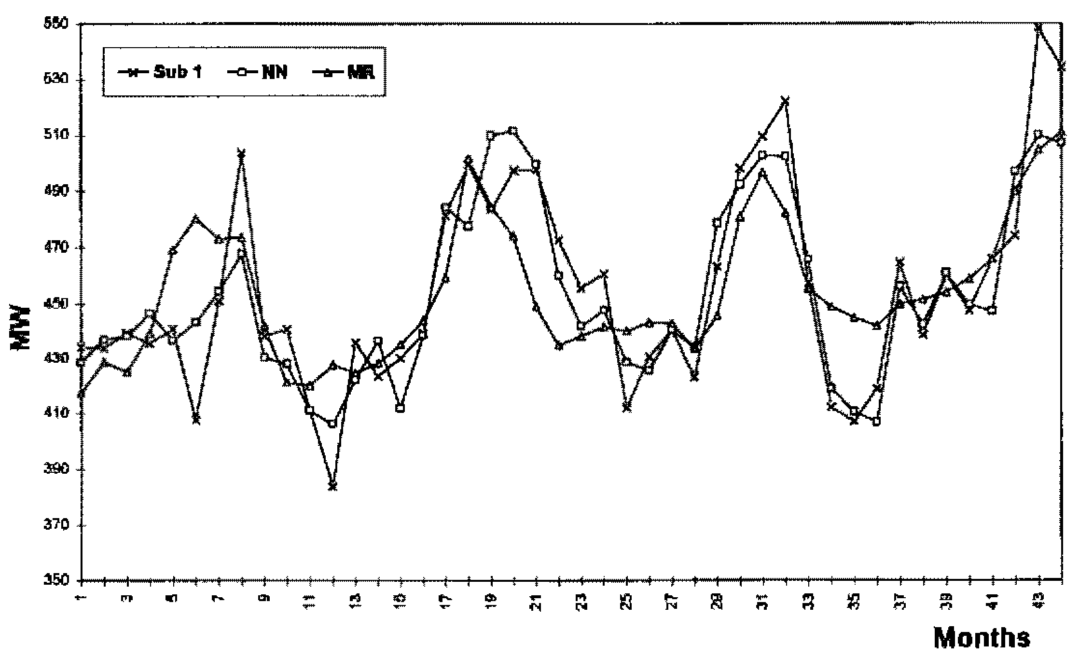
The forecast test statistics (spreadsheet results) for all ten substations are given (Tables 5.6 to 5.15). The outputs from the neural net and multiple regression model for each substation have been plotted (see Graphs 5.2 to 5.11).

The multicollinearity problem is best described by the substation 1 statistics. The value for the F-test is nearly 5, but the absolute values for the t-tests are very low.

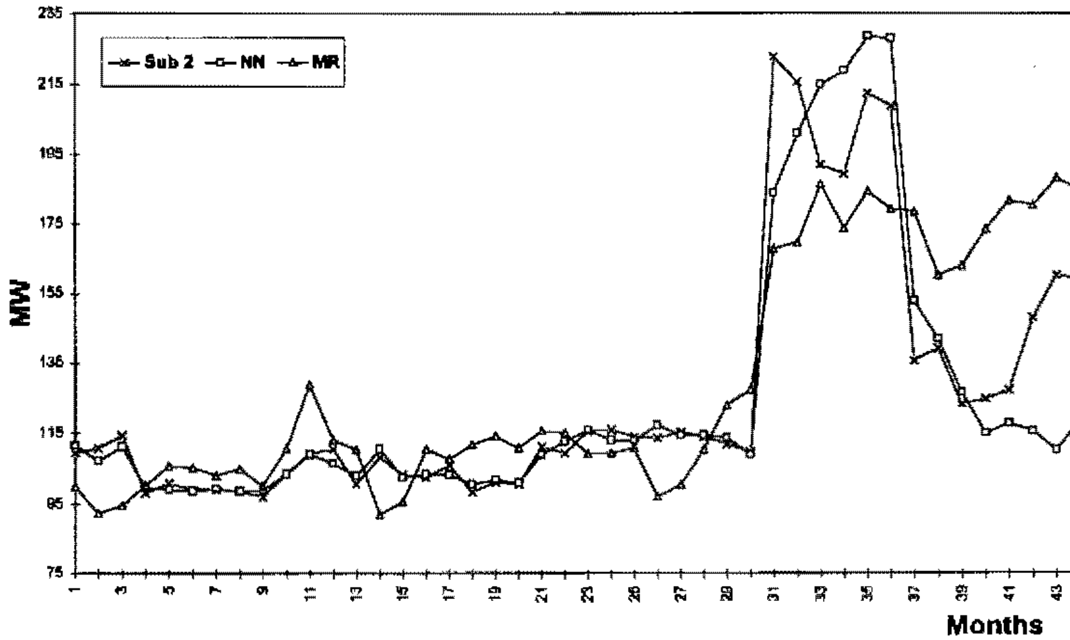
The test results for substation 9, prove the existence of autocorrelation, which does not bias the regression coefficients that are estimated, but does make the estimates of the standard errors smaller than the true standard errors.

This means that the t ratio calculated for each coefficient will be overstated, which in turn may lead to the rejection of the null hypothesis, which should in fact not be rejected. Thus the coefficient is not statistically significantly different from zero, but due to autocorrelation the t ratio indicates that the coefficient is statistically significantly different from zero and the regressor variable will therefore erroneously be included.

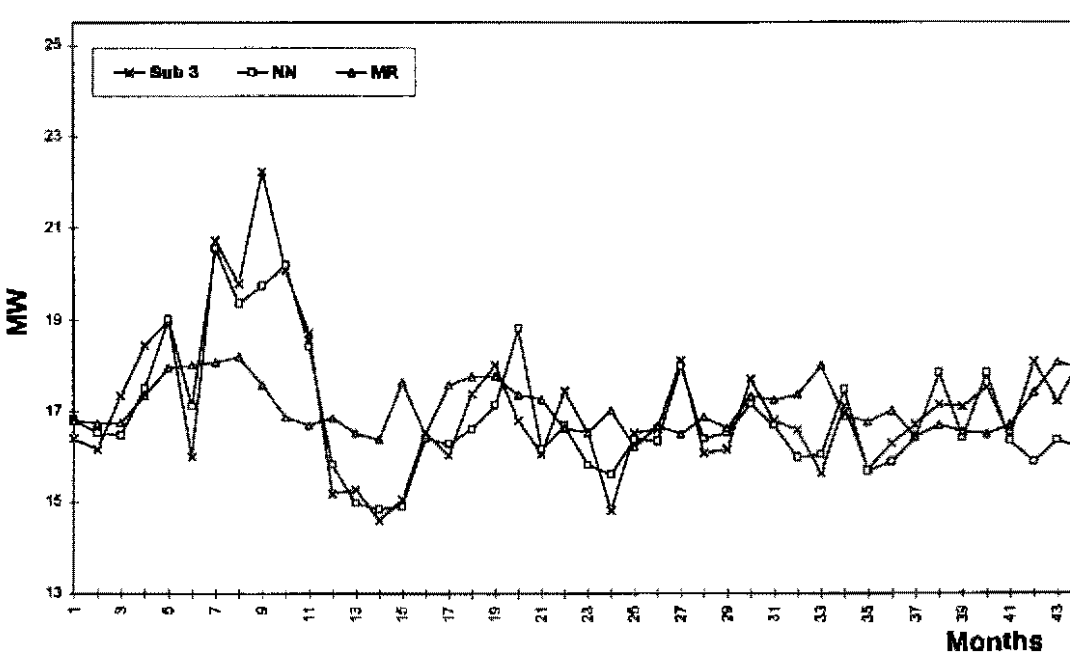
In addition, the existence of autocorrelation causes the R^2 and F statistics to be unreliable in evaluating the overall significance of the regression function.



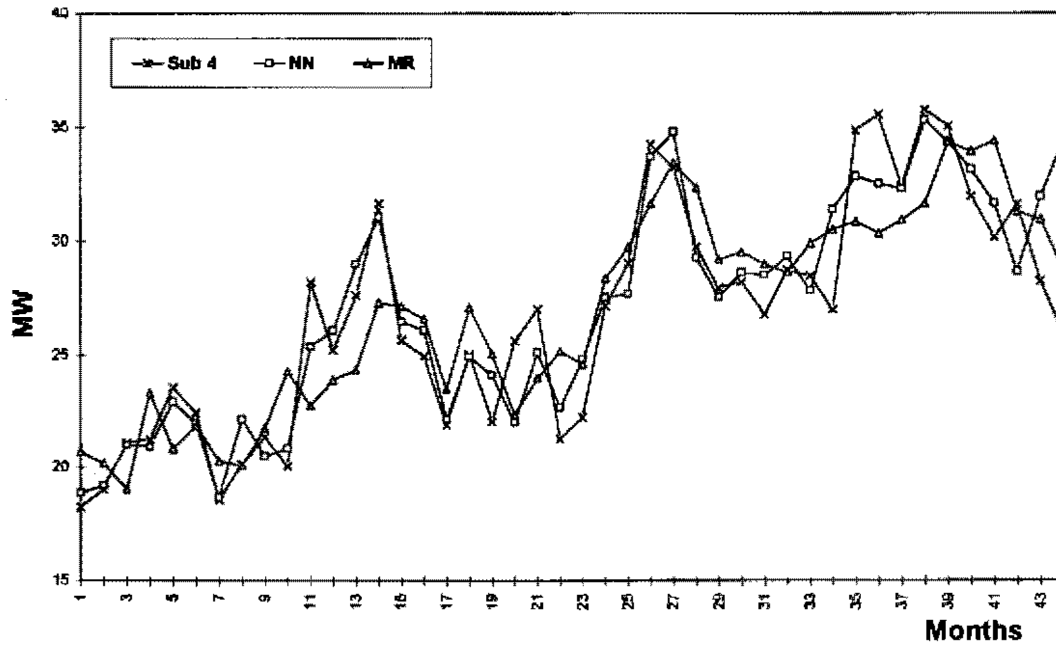
Graph 5.2: Substation 1 (results)



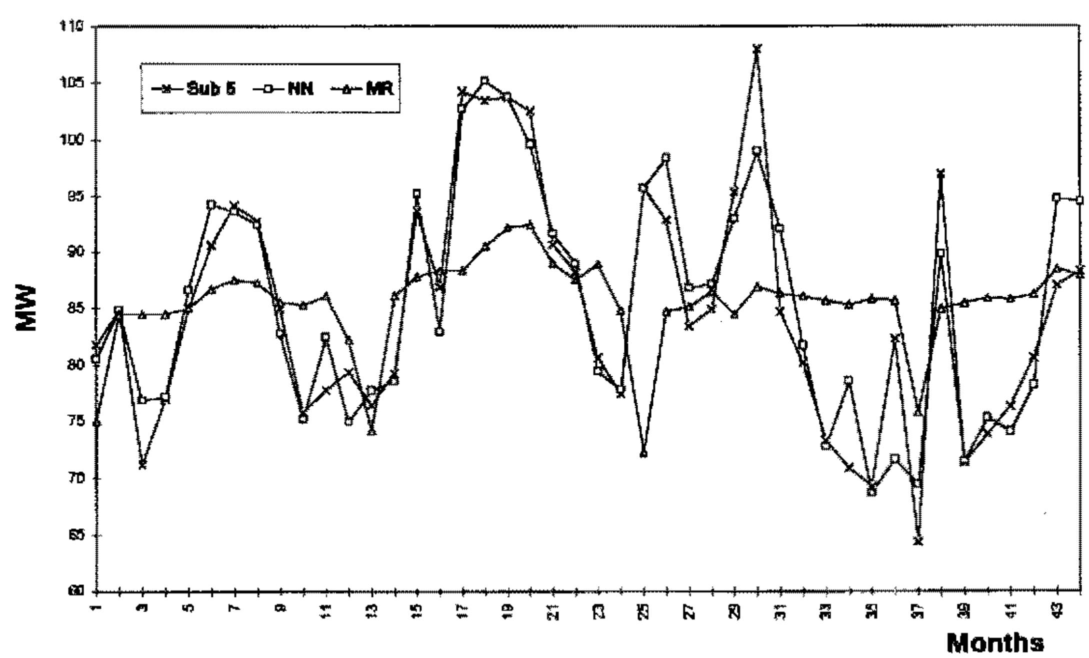
Graph 5.3: Substation 2 (results)



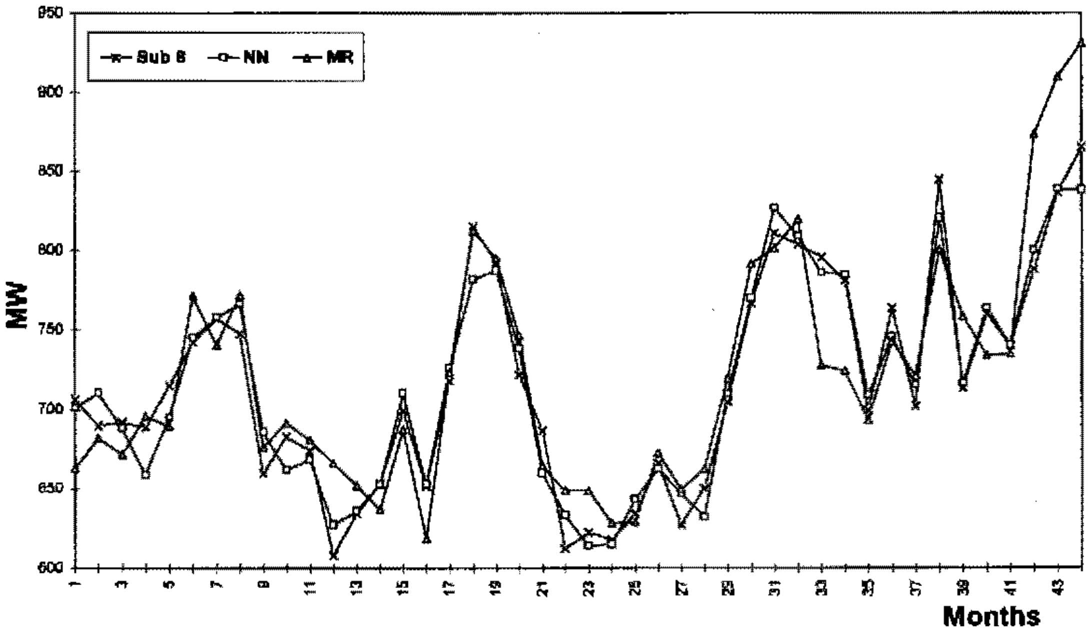
Graph 5.4: Substation 3 (results)



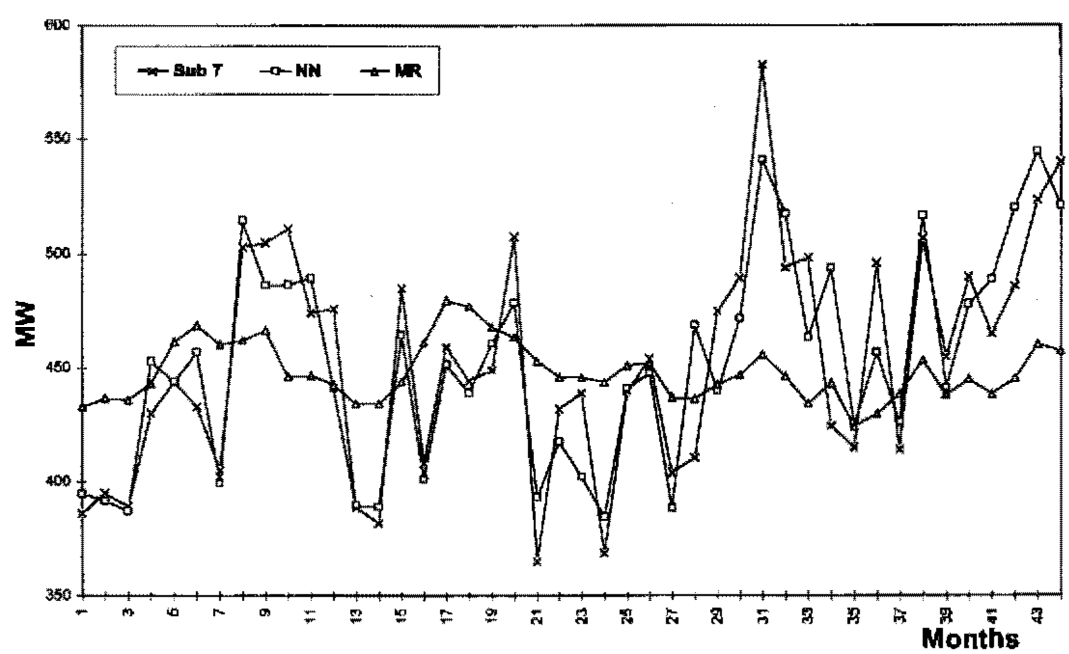
Graph 5.5: Substation 4 (results)



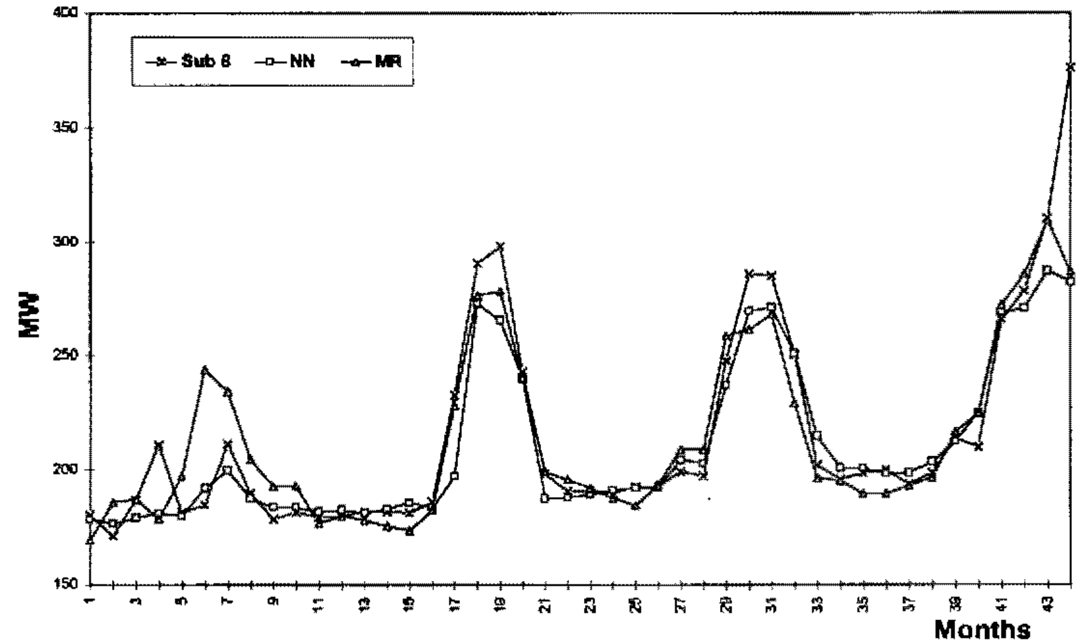
Graph 5.6: Substation 5 (results)



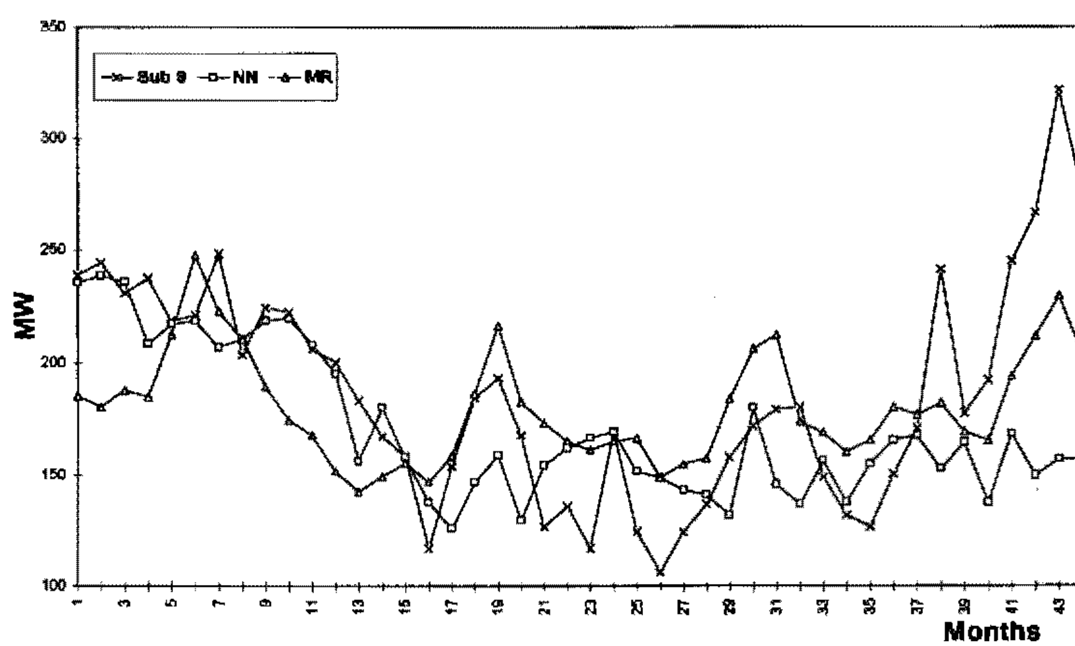
Graph 5.7; Substation 6 (results)



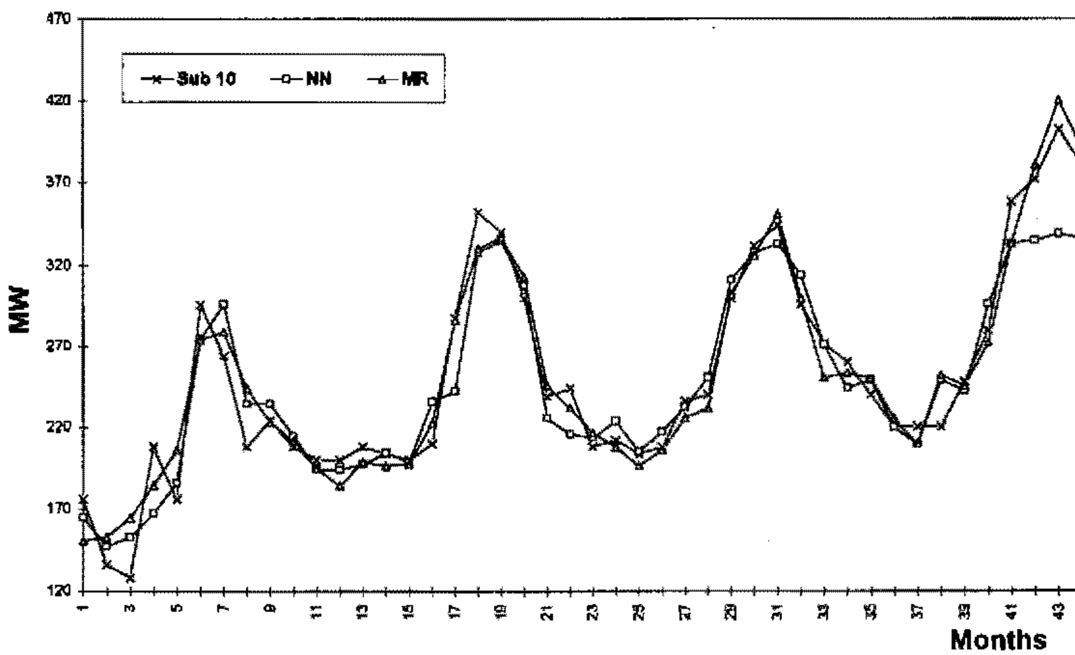
Graph 5.8: Substation 7 (results)



Graph 5.9: Substation 8 (results)



Graph 5.10: Substation 9 (results)



Graph 5.11: Substation 10 (results)

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.66
R square	0.44
Adjusted R square	0.33
Standard error	26.62
Observations	40.00

ANOVA

	df	SS	MS	F
Regression	6.00	18133.28	3022.21	4.27
Residual	33.00	23377.76	708.42	
Total	39.00	41511.04		

	Coefficients	Standard error	t stat	P value	Lower 95%	Upper 95%
Intercept	129.94	117.06	1.11	0.27	-108.21	368.09
X variable 1	0.05	0.09	0.56	0.58	-0.14	0.24
X variable 2	-0.74	3.86	-0.19	0.85	-8.59	7.12
X variable 3	-14.59	15.19	-0.96	0.34	-45.49	16.31
X variable 4	2.19	3.41	0.64	0.53	-4.75	9.12
X variable 5	4.52	5.53	0.82	0.42	-6.74	15.77
X variable 6	2.46	5.45	0.45	0.66	-8.63	13.55

Table 5.6: Substation 1 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.83
R square	0.68
Adjusted R square	0.62
Standard error	22.50
Observations	40.00

ANOVA

	df	SS	MS	F
Regression	6.00	35758.98	5959.83	11.77
Residual	33.00	16713.39	506.47	
Total	39.00	52472.37		

	Coefficients	Standard error	t stat	P value	Lower 95%	Upper 95%
Intercept	106.65	57.54	1.85	0.07	-10.41	223.70
X variable 1	1.44	1.08	1.33	0.19	-0.76	3.64
X variable 2	-3.83	6.61	-0.58	0.57	-17.27	9.62
X variable 3	8.78	17.48	0.50	0.62	-26.78	44.33
X variable 4	-26.82	16.09	-1.67	0.10	-59.56	5.91
X variable 5	2.99	3.82	0.78	0.44	-4.77	10.76
X variable 6	2.22	0.34	6.58	0.00	1.53	2.91

Table 5.7: Substation 2 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.34
R square	0.12
Adjusted R square	0.07
Standard error	1.56
Observations	40.00

ANOVA

	df	SS	MS	F
Regression	2.00	12.08	6.04	2.47
Residual	37.00	90.47	2.45	
Total	39.00	102.55		

	Coefficients	Standard error	t stat	P value	Lower 95%	Upper 95%
Intercept	11.48	2.85	4.03	0.00	5.71	17.25
X variable 1	2.20	1.01	2.18	0.04	0.15	4.24
X variable 2	0.23	0.26	0.88	0.39	-0.30	0.75

Table 5.8: Substation 3 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.86
R square	0.73
Adjusted R square	0.68
Standard error	2.87
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	6.00	743.90	123.98	15.03
Residual	33.00	272.15	8.25	
Total	39.00	1016.05		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	25.02	9.16	2.73	0.01	6.39	43.65
X variable 1	-0.55	1.63	-0.34	0.74	-3.87	2.78
X variable 2	2.67	0.88	3.02	0.00	0.87	4.46
X variable 3	-11.71	3.97	-2.95	0.01	-19.78	-3.64
X variable 4	-1.42	0.78	-1.81	0.08	-3.01	0.18
X variable 5	4.21	1.98	2.13	0.04	0.19	8.23
X variable 6	11.10	6.94	1.60	0.12	-3.03	25.23

Table 5.9: Substation 4 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.39
R square	0.15
Adjusted R square	0.13
Standard error	10.12
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	1.00	701.19	701.19	6.85
Residual	38.00	3892.13	102.42	
Total	39.00	4593.32		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	50.12	13.53	3.70	0.00	22.73	77.51
X variable 1	0.69	0.26	2.62	0.01	0.16	1.23

Table 5.10: Substation 5 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.90
R square	0.81
Adjusted R square	0.66
Standard error	35.55
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	16.00	124936.85	7808.55	6.18
Residual	23.00	29060.74	1263.51	
Total	39.00	153997.59		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	-621.94	455.49	-1.37	0.19	-1564.19	320.30
X variable 1	2.95	2.03	1.45	0.16	-1.25	7.15
X variable 2	-40.97	58.02	-0.71	0.49	-161.00	79.06
X variable 3	-4.79	3.97	-1.21	0.24	-13.01	3.42
X variable 4	1.48	1.92	0.77	0.45	-2.49	5.44
X variable 5	1.09	6.90	0.16	0.88	-13.18	15.37
X variable 6	2.36	4.69	0.50	0.62	-7.34	12.05
X variable 7	-20.61	17.17	-1.20	0.24	-56.13	14.92
X variable 8	19.23	11.52	1.67	0.11	-4.60	43.05
X variable 9	1.67	1.38	1.21	0.24	-1.19	4.53
X variable 10	-1.44	7.84	-0.18	0.86	-17.66	14.78
X variable 11	-6.61	11.13	-0.59	0.56	-29.64	16.41
X variable 12	7.45	2.90	2.57	0.02	1.45	13.45
X variable 13	-3.53	3.36	-1.05	0.30	-10.49	3.43
X variable 14	-21.80	12.91	-1.69	0.10	-48.52	4.91
X variable 15	15.73	12.10	1.30	0.21	-9.30	40.77
X variable 16	7.23	4.45	1.62	0.12	-1.98	16.43

Table 5.11: Substation 6 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.27
R square	0.07
Adjusted R square	0.05
Standard error	47.23
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	1.00	6675.08	6675.08	2.99
Residual	38.00	84771.82	2230.84	
Total	39.00	91446.90		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	334.33	66.25	5.05	0.00	200.21	468.45
X variable 1	0.18	0.11	1.73	0.09	-0.03	0.40

Table 5.12: Substation 7 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.89
R square	0.79
Adjusted R square	0.78
Standard error	15.94
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	3.00	35021.95	11673.98	45.93
Residual	36.00	9150.75	254.19	
Total	39.00	44172.70		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	-89.75	54.98	-1.63	0.11	-201.25	21.75
X variable 1	9.29	3.88	2.40	0.02	1.43	17.15
X variable 2	2.76	0.77	3.58	0.00	1.20	4.32
X variable 3	0.47	0.33	1.43	0.16	-0.20	1.15

Table 5.13: Substation 8 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.57
R square	0.32
Adjusted R square	0.30
Standard error	34.36
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	1.00	21153.04	21153.04	17.91
Residual	38.00	44870.55	1180.80	
Total	39.00	66023.59		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	30.46	35.12	0.87	0.39	-40.65	101.56
X variable 1	0.18	0.04	4.23	0.00	0.09	0.26

Table 5.14: Substation 9 statistics

SUMMARY OUTPUT

Regression statistics	
Multiple R	0.96
R square	0.91
Adjusted R square	0.90
Standard error	17.14
Observations	40.00

ANOVA

	df	*SS*	*MS*	*F*
Regression	7.00	100062.24	14294.61	48.67
Residual	32.00	9397.66	293.68	
Total	39.00	109459.90		

	Coefficients	*Standard error*	*t stat*	*P value*	*Lower 95%*	*Upper 95%*
Intercept	-22.68	98.32	-0.23	0.82	-222.95	177.60
X variable 1	7.69	5.88	1.31	0.20	-4.29	19.66
X variable 2	3.90	0.69	5.62	0.00	2.49	5.31
X variable 3	-6.41	7.44	-0.86	0.40	-21.57	8.75
X variable 4	-0.25	1.05	-0.24	0.81	-2.39	1.88
X variable 5	14.95	7.84	1.91	0.07	-1.02	30.92
X variable 6	1.94	1.07	1.82	0.08	-0.23	4.12
X variable 7	-0.13	0.38	-0.34	0.73	-0.90	0.64

Table 5.15: Substation 10 statistics

5.6 Neural net results

The neural net results obtained from the ten substations indicate that the training of neural nets was not done correctly.

There are two main reasons for this. Firstly, not enough data is available. In almost all the cases the neural nets are overtrained (see section 3.3.10). Thus any small variation in new (test) input data can result in a significant variation in the net's output compared with the actual value, as shown especially by Graphs 5.3, 5.7, 5.10 and 5.11. Generalisation, as discussed in section 3.3.9, could not be achieved. With generalisation a net is trained that is insensitive to variations in the input data during training and that recognises a pattern in the data despite noise and distortion.

Secondly, the approach of using customer-diversified demands (inputs) to forecast substation demands is not recommended. The inputs do not provide adequate information to the neural net during training to enable it to recognise the pattern in the data.

5.7 Regression results

The regression results obtained from the ten substations are very similar to the results obtained by neural nets. The results are unsatisfactory due to insufficient data, multicollinearity, autocorrelation and heteroscedasticity.

5.8 Learned from experience

Through the experience gained by this exercise, a number of very important lessons were learned.

Firstly, in a field as complex as the flow of MW in electrical networks, the forecasting process should be drawn as flow diagrams (see Chapter 6). Further, the various relationships should be defined mathematically.

The outputs required and what data is required in order to produce those outputs should be establish clearly. If some of the data is not available, methods must be included in the model to overcome this problem.

Only when this has been established, can appropriate techniques be considered to predict future demands.

5.9 Conclusion

This approach is not suggested for the forecasting of substation demands. This conclusion is based on the fact that not enough observations are available and there is insufficient data (variables) - particularly variables to model the effect of network operations. This is required for cases where only the substation demands change, and not the maximum monthly demand of the customer (see section 1.4).

CHAPTER 6

FORECASTING MODEL

6.1 Introduction

As was mentioned in Chapter 1, a forecasting model is needed to improve the accuracy of the demand load forecast. The forecasting horizon is 10 to 15 years, which means that long-term forecasting techniques have to be used. More accurate point load demands are required for electrical network studies. If the accuracy of transmission substation demand forecasts is improved, it will be possible to determine with more certainty which substations will exceed their firm capacities under normal operating conditions, and when. The generation pattern is also an important issue, therefore appropriate algorithms are required to determine future generation patterns.

The forecasting model has to be aligned with Eskom's suggested business process (see Map out Future Environment, Appendix A, for more details). The input data and control parameters for the business process have to be the inputs for the forecasting model, and the forecasting model has to provide the same outputs as the business process.

To construct such a model, some electrical network knowledge is required. After the business process has been discussed, some time will be spent discussing the electrical networks to define them in terms of a transportation model. The model has been developed on that basis, ie that of a transportation model.

6.2 The Eskom business process

Eskom's business process (Map out Future Environment) comprises a number of subprocesses, one of them being Map Inputs and Controls into Engineering Requirements (see Appendix A). In the rest of Chapter 6 this process (Map Inputs and Controls into Engineering Requirements) will be referred to only as the 'process'. This process in turn contains four subprocesses and provides three outputs (see bottom right-hand corner of the figure in Appendix A).

The process has four control inputs (from the top) plus four data inputs (from the left) and a number of inputs from supporting systems (from the bottom). The inputs from the supporting systems are databases storing historical data.

The four control inputs are Customer Needs; Electricity Supply Technology; Long-Term Load Forecast; and Plant Condition and Network Information. The four data inputs are Market Information; Customer Future Demands; Generation Capacity Requirement; and Negotiated Outage Plan. The outputs (also the outputs for the bottom right subprocess Analyse Demand Growth, Appendix A) are Market Trends; Projected Load Growth; and PSS/E Loading Parameters. For more details on the supporting system inputs, see section 6.11.

One of the key issues in load forecasting is to understand the needs of our customers; for which of those needs electricity is required; and how those needs will change over time. Thus the customer, his needs, and the means of satisfying those needs through an energy source, constitute an important control parameter in the load forecasting process.

New developments in electricity supply technology as well as environmental pressures may influence customers to use electrical equipment rather than, for example, petrol, diesel or gas-driven equipment to satisfy their needs, therefore it is important as a control parameter.

The long-term load forecast constraint has to be used as a strategic and tactical input into the model.

The Plant Condition and Network Information control input as well as the Negotiated Outage Plan input determines a topology model which controls the allocation of the expected loads on the system. This will take into account all network changes (see item 8 under the transmission system expansion engineers' requirements, section 1.6, Chapter 1).

The Market information input has to provide the views of experts regarding key issues in the South African electrical industry. This is information on the electrical market sectors, used to identify new developments, long-term planning scenarios, etc. The sectors are, for example, ferrochrome production, gold mines, large municipalities, etc. Also important are the inputs on political key issues, social and welfare issues, economic growth, new or revised environmental policies, etc. Weather conditions and changing weather patterns are also important inputs into the load forecast model.

The Customer Future Demands input contains information on all existing and new customers' expected future demand (MW), which is the information required for the expected future POD values.

The last input should contain data on the availability and capability of all possible generating stations, including not only Eskom power stations, but also power imported from neighbouring states.

The outputs have to provide, for each condition, a set of PSS/E load parameters (point loads). PSS/E-19 is the abbreviation for power system simulator, version E-19. These results are required in order to expand the transmission networks and increase, where necessary, the substations' firm capacities to meet future market trends and load growth. The outputs should also provide information on market trends and load growth projections.

The upgrading of a substation's firm capacity can take as long as five years, while generating capacity increases can take up to seven years.

6.3 Electrical loads

An electrical load can be described as an electrical supply to a customer. A customer may have one or more electrical loads, feeding into the same geographical area. An electrical load may also be referred to as a POD (see Fig 1.2) furnishing a customer with electricity.

Such electrical loads are fed from electrical busbars and the busbars are connected to transmission substations by means of overhead power lines or underground cables.

6.4 Point loads

A point load is used to simulate the electricity flow between the transmission electrical networks and the distribution or generation or neighbouring states' electrical networks respectively, ie where electrical power flows from the transmission networks into the distribution or generation or neighbouring states' networks. Thus a point load is the grouping of one or more electrical loads.

It was not possible to simulate in PSS/E all the transmission and distribution electrical networks; the PSS/E files are simply too large and problems occurred with the processing of the files (finding a realistic representation of the actual electrical load flow). Especially radial type distribution networks are grouped as one point load. A radial network is an electrical layout that has only one link between the substation and the load. If a load has links with more than one substation, those networks are called ring networks.

If historical data exists on all the point loads, the links between some electrical busbars and the PODs' demands are not required. This is only required to verify the demand values on the busbars (see section 6.15). The links between the electrical busbars and the PODs may no longer be required once the metering project mentioned in Chapter 1 has been completed and links between the point loads and the "stats" points have been successfully completed. Where possible, a number of PODs are allocated to an electrical busbar - if possible up to 80% or more of the expected busbar demand. In some cases electrical busbars are feeding a number of domestic and commercial electrical loads (measured in kWh). If available, the summated kWh values are used to balance the electrical busbar demands.

Point loads representing traction loads (for electrically powered trains) are difficult to simulate, as the load is moving through a number of point loads and the total demand on the electrical system is not the summation of the various point loads, but the electrical load of the train. In such cases it is good practice to discuss the matter with the transmission system expansion engineers, who will identify the point load(s) where it is best to simulate the electrical load.

6.5 Point load configurations

The grouping of transmission and distribution into point loads is represented by Figs 6.1, 6.2, 6.3 and 6.4. Those transmission substations are the lower transmission substations (see section 6.6).

The demand (kW) on the lower transmission substation, for a given condition, can be defined as

$$T_X(L)_{kt} = \Big(\sum_{i \,\varepsilon\, A_{kt}} S_{ikt}\ BL_{it}\Big) + L_{kt} \qquad \text{(Eq 6.1)}$$

where

$T_X(L)_{kt}$ is the demand for the k th substation (kW), for a given year t

S_{ikt} is the scaling factor for the busbar load, BL_{it}, allocated to the k th substation, for a given year t

BL_{it} is the i th annual maximum busbar load, for a given year t

A_{kt} is an index set, allocating certain busbar loads to the k th substation, for a given year t

L_{kt} is the supply losses for the k th substation, ie making provision for line losses to supply the point loads, for a given year t

Note : No subindex has been included for the different conditions.

The scaling factor is defined as

$$S_{ikt} = Div_{it}\, p_{ikt} \qquad \text{(Eq 6.2)}$$

where

S_{ikt} is the scaling factor for the busbar load, BL_{it}, allocated to the k th substation, for a given year t

Div_{it} is the diversity factor for the busbar load, BL_{it}, for a given year t

p_{ikt} is the supply factor for the busbar load, BL_{it}, allocated to the k th substation, for a given year t

To determine the j th point load demand, a number of scaled busbar loads have to be added, ie:

$$PL_{jt} = \sum_{i \varepsilon B_{jt}} Div_{it}\, BL_{it} \qquad \text{(Eq 6.3)}$$

where

PL_{jt} is the j th point load demand, for a given year t

Div_{it} is the diversity factor for the busbar load, BL_{it}, for a given year t

BL_{it} is the i th annual maximum busbar load, for a given year t

B_{jt} is an index set, allocating certain maximum busbar loads to the j th point load, for a given year t

The main reason for defining the lower transmission substations and the point loads according to equations 6.1, 6.2 and 6.3 is to separate the important "influences". Research and forecasting are then possible in respect of each of these factors individually. This will give a better understanding of these factors and their possible future values.

The diversity factors are dependent on time, different weather patterns, alternative customer network operations, tariff structures, etc. Thus, through research and the use of appropriate statistical techniques, better assessments of those diversity factors can be obtained.

The supply factors are dependent on the network topology; the operation of the networks, including the generation pattern and the switching of the var equipment; and the loads on the electrical busbars.

The supply losses are dependent on the loads on the electrical busbars and the network topology supplying the loads, especially the type of conductors installed.

In normal situations a substation k only supplies the point loads according to the index set A_k; otherwise the situation is abnormal. In one example the maximum demand is 265 MW for normal situations, whereas demands of 387 MW have been recorded for abnormal situations.

In Fig 6.1 the demands for the electrical load (EL), the substation, and the point load (PL) are equal. In this case the loss, diversity and supply factors equal 1.

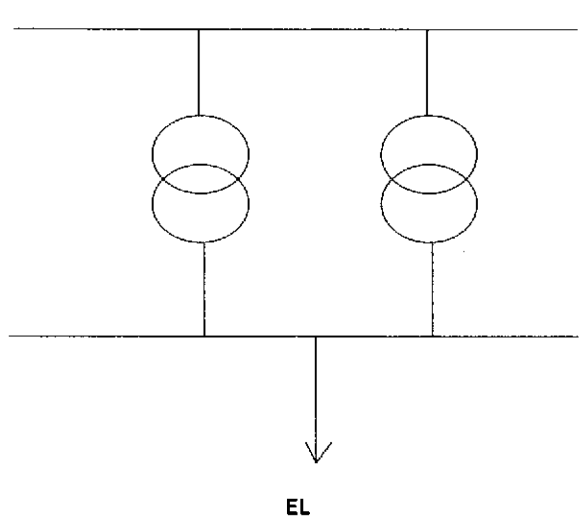


Fig 6.1: Single electrical load (single substation) layout

In Fig 6.2 more than one electrical load is fed from the substation, but the load is still modelled as one point load. Thus equations 6.1, 6.2 and 6.3 are still applicable to determine the transmission substation demands.

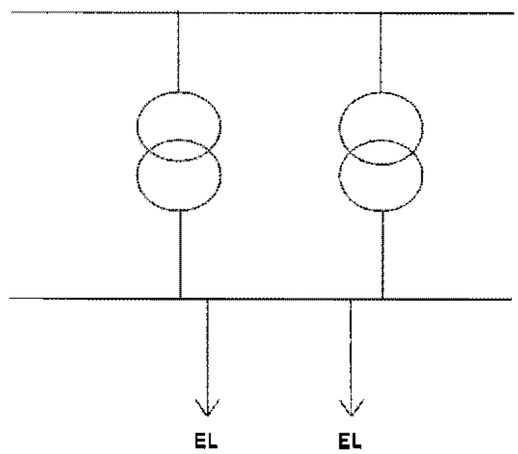


Fig 6.2: Multi-electrical load (single substation) layout

If the annual maximum demands for the electrical loads are available, then the demand for the point load is given as:

$$PL = (EL_1 + EL_2 + \ldots + EL_n).\ \text{diversity factor} \qquad \text{(Eq 6.4)}$$

where

n is the number of electrical loads fed from this point load

The layout in Fig 6.3 is more complex - two transmission substations are interconnected and there are a number of point loads fed by the two substations. It is also possible that a single point load (or a busbar) can be fed from both substations. In that case the supply factor will be less than one.

In this layout all five busbars are feeding one or more electrical loads. To model the loads, five point loads can be used.

When a line connecting two substations is in commission most of the time, we say that the line is normally closed (N/C); as opposed to the line being normally open (N/O). Sometimes a substation only exceeds its firm supply when an N/O line is closed for maintenance or for contingency reasons.

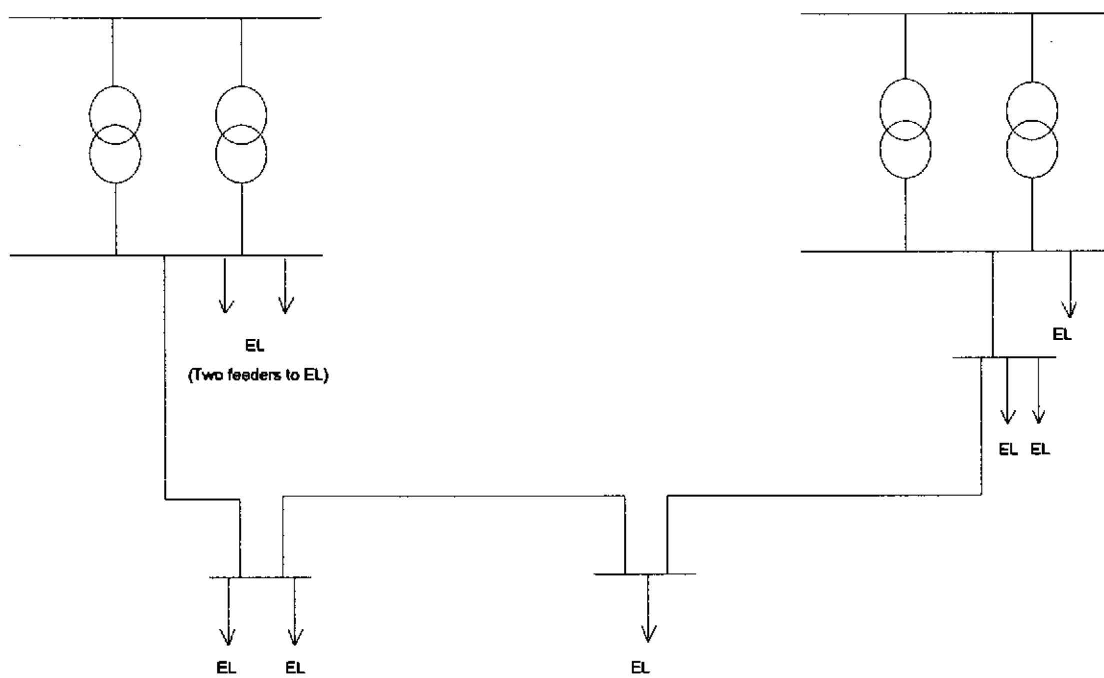


Fig 6.3: Multi-electrical load (double substation) layout

In the next layout (Fig 6.4) there are more than two transmission substations which are interconnected.

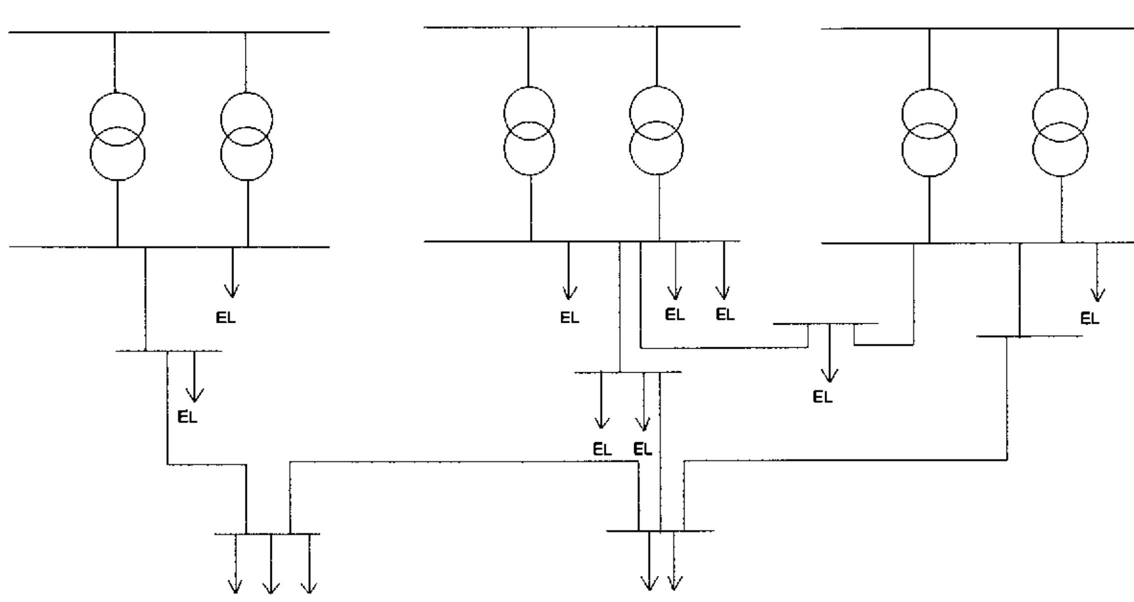


Fig 6.4: Multi-electrical load (multiple substation) layout

The connection of the distribution networks connecting the transmission substations with the electrical load busbars can best be described as a spider-web. A good knowledge of the electrical networks can provide some information about boundaries to determine the index sets for the transmission substations.

The lack of historical data on the point loads is a major constraint in the production of accurate point load demands, which are required to model PSS/E. To improve their accuracy, two approaches - a bottom-up and a top-down approach - have been developed (see section 6.15).

6.6 Transmission substations

The transmission substations have been divided into two groups - those connected directly to the distribution networks are called **low-level nodes** ($T_X(L)$) and the others, which are not directly connected, are called the **high-level nodes** ($T_X(H)$).

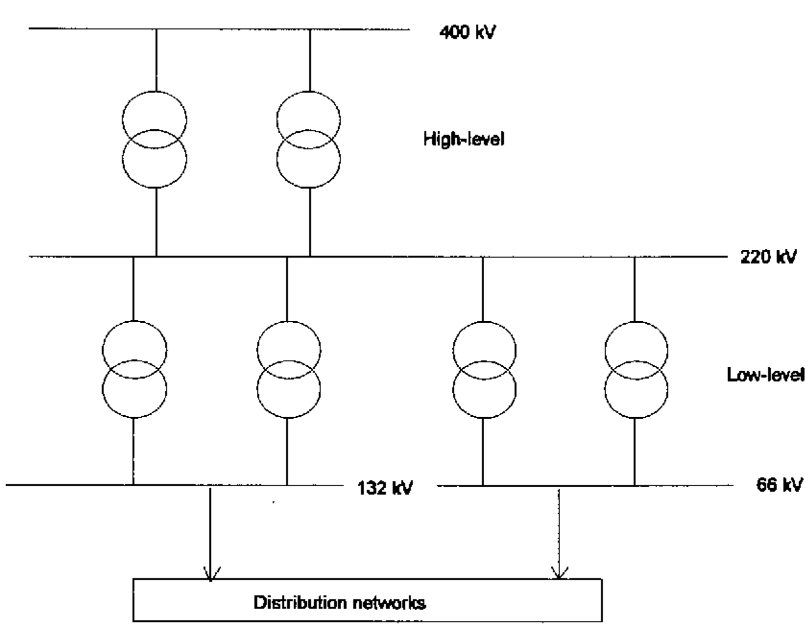


Fig 6.5: A substation with high-level and low-level nodes

A substation normally has two voltage levels (for example 275 kV and 88 kV).

Such a substation then has only one voltage transformation (275/88 kV in this case). If a substation has four voltage levels and three different voltage transformations (Fig 6.5), then the grouping in terms of high-level and low-level nodes will be as follows: the voltage transformation from 400 kV to 220 kV is a high-level node and the other two voltage transformations (220/132 kV and 220/66 kV) are two different low-level nodes.

6.7 Export/import feeders

Feeders connecting Eskom transmission networks with those of neighbouring states can either **export** (away from the substation) or **import** (towards the substation) MW (see Fig 6.6). The exported or imported MW respectively represent an outflow and an input into the data flow diagram. It is obvious that such feeders can either export or import, and in any one-hour metering interval both import and export figures can be recorded.

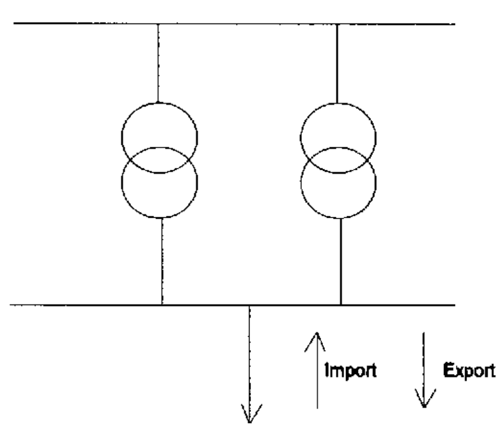


Fig 6.6: A substation feeder either exports or imports

6.8 Generation pattern

Not all MW generated by a generator is supplied to the transmission networks. A certain amount of MW is required for own use (auxiliary supplies). Those auxiliary supplies are fed from the unit boards. Thus the *nett sent* out to the transmission networks is the total MW generated by the power station minus all the MW required for the auxiliary supplies fed from the unit boards. The *nett sent* out is called a *generation pattern*.

A typical power station arrangement is shown in Fig 6.7.

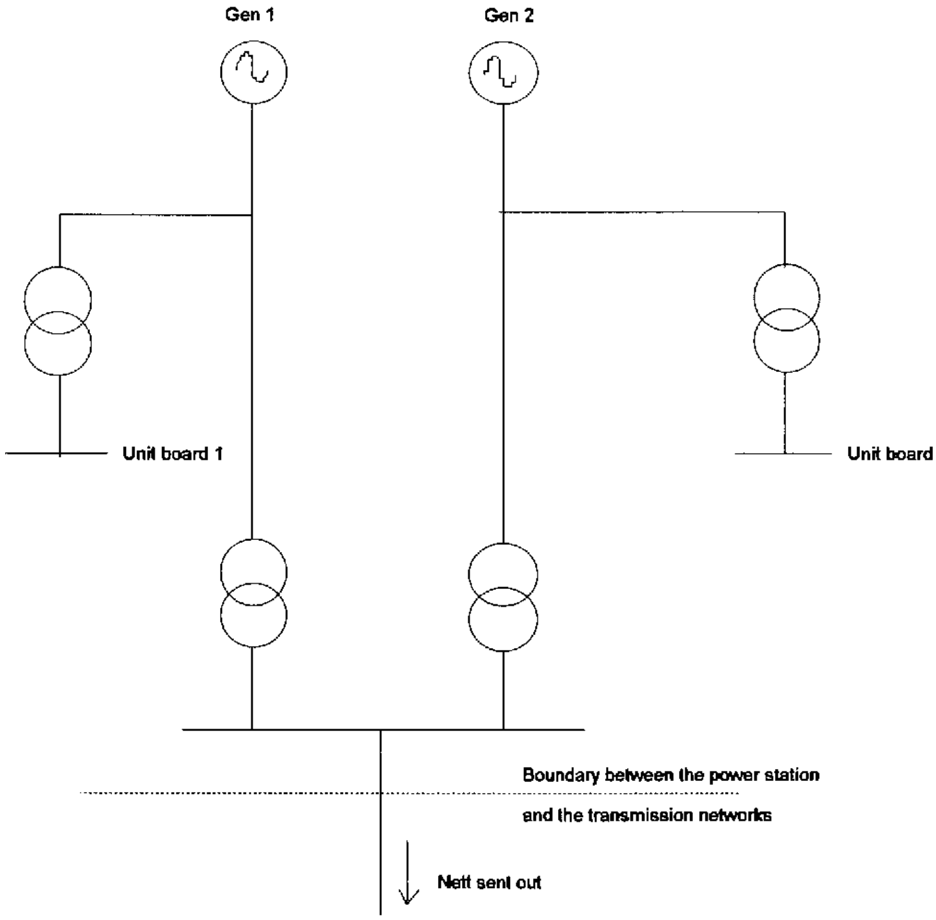


Fig 6.7: A typical power station arrangement

6.9 The transportation model diagram

A transportation model exists of the source, intermediate and demand nodes. For our purposes two levels of intermediate nodes are used.

The source nodes are the nett sent out from Eskom power stations (section 6.8) and the imports from neighbouring states (section 6.7).

The first and second levels of intermediate nodes are the high-level and low-level transmission substations respectively (section 6.6). There are also interconnections between the nodes for both these levels.

The demand nodes are the point loads as discussed in sections 6.4 and 6.5.

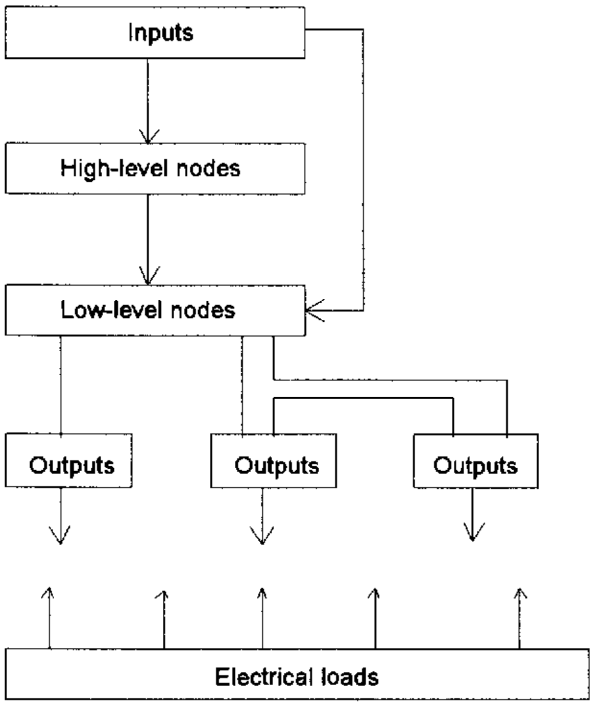


Fig 6.8: Transportation model diagram

6.10 Electrical terms

This section will give the reader an understanding of the relationship between the different electrical terms. For more in-depth studies, consult electrical textbooks.

kW = kVA x power factor (Cos Φ see Chapter 1)

kWh = kW x hours (If an electrical load runs constantly at 12 kW for 10 hours, then the kWh = 120)

Note: If the load is not constant, the kWh is the area underneath the curve when the kW is plotted over time (X-axes)

$$\text{Load factor} = \frac{\text{kWh}}{\text{kW}_{max} \times \text{hours}}$$

where kW_{max} is the maximum kilowatts measured in a given time period kWh is the total kilowatt hours in the time period

hours is the total number of hours in the time period

Also MW = 1 000 kW

MVA = 1 000 kVA

6.11 Data available

The data for the electrical loads (customer loads) is kept in the Eskom billing database. Depending on the tariff and the notified maximum demand (NMD), ie the maximum supply a customer can take from the electrical system, Eskom customers are grouped into two categories. The one category is called large power users (LPU) and the other small power users (SPU). For SPU customers only total monthly kWh figures are available. For some LPU customers demand (kW or kVA) and total monthly kWh figures are available, but for others only total monthly kWh figures are available. The reason for this is that only kWh values are required for billing purposes, and that only kWh metering has therefore been installed. The monthly figures are available from at least 1993 onwards.

On most of the point loads (probably more than 80%), no historical data exists for the past three years, except data from metering sheets at some substations.

The lower-level transmission substations (also those of neighbouring states) have hourly MW readings (total for the substation) available from the beginning of August 1993. Since the introduction of the new metering project, not only total substation figures are available, but also figures for individual transformers and/or feeders.

The hourly nett sent figures (MW) for Eskom power stations are available for more than ten years.

There is no historical data for the higher-level transmission substations.

Snapshots (instantaneous measurements) are recorded, but are only kept for three days. The possibility of recording those figures for longer periods - perhaps for five years - is being investigated.

6.12 Condition

The term *condition* describes a certain event for which a set of point loads and a generation pattern are required. An event may, for example, be the total system reaching its peak (TOSP); or another event, perhaps for the same TOSP but with a different generation pattern, etc.

6.13 Diversification rule

This rule is very important, especially in the verification of point load demands against the demands recorded for the substation. A diversity factor is described as the sum of individual maximum demands of customers divided by the maximum load on the system [97].

If a substation supplies more than one point load and if the maximum demands (measured in kW), both for the point loads and for the substation, did not occur at the same point in the time period, then the summation of the maximum demands for the point loads divided by the substation maximum load (excluding line losses) will be larger than one.

6.14 Typical electrical network

Fig 6.9 gives a typical electrical network layout, ie one with generation, higher and lower transmission substations, point loads and electrical loads.

The generation is the nett sent out, as explained in section 6.8. The generation points are connected either directly to the lower transmission substations ($T_x(L)$) or via higher transmission substations ($T_x(H)$).

The substation on the left is shown as a radial point load configuration and the other two as a ring point load configuration.

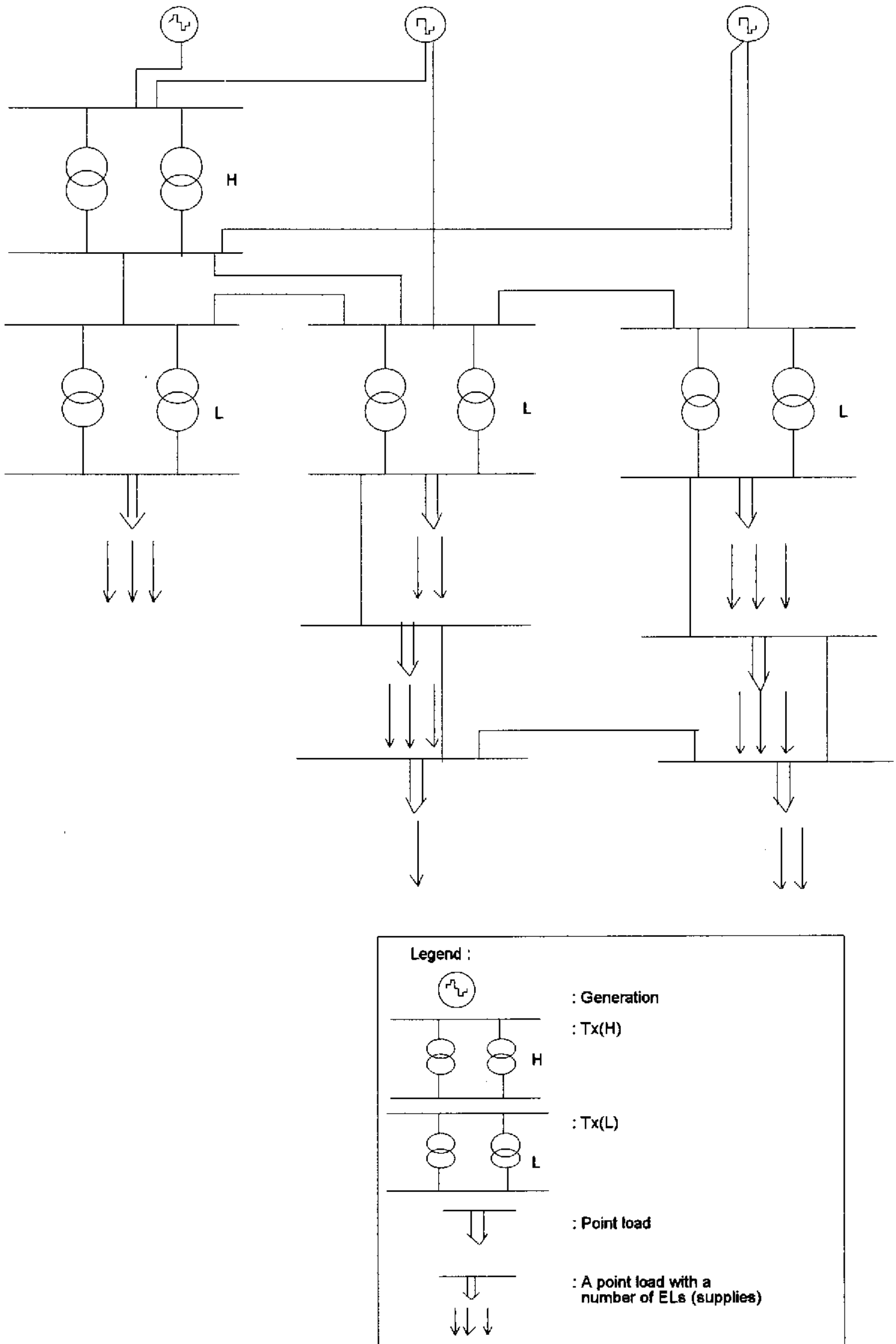


Fig 6.9: A typical electrical network

Each point load can have either one or more electrical loads attached to it. Some point loads are the equivalent (combination) of an electrical network (see section 6.4).

6.15 The forecasting model

It is suggested that a program be developed to perform the outputs as required by the model. The programming should be such that, from a main menu, subprograms are selected; and once those subprograms have completed their tasks, they return to the main menu.

6.15.1 Linking`

This section consists of three subsections which provide all the links required to build the electrical networks into a transportation model. The first subsection links the lower transmission substations with the electrical busbars, the second links the electrical busbars with the PODs, and the last one links the transmission networks, including the generating stations. This should include all historical links, ie for all the years for which historical data exists and until the end of the forecasting horizon.

The first linking is required to determine, from the distributors' forecasts, the lower-level transmission substation forecasts.

The second linking is required in cases where no historical data exists on electrical busbars, then the PODs' maximum demands are used to verify the busbar demand (see section 6.15.7). The number of PODs linked to a busbar will depend mainly on the importance of the busbar load and the maximum demand on the busbar relative to the maximum lower transmission substation demand. When the busbar supplies rural areas, or traction loads or very small municipalities, the linking is mostly not worth considering. As explained in section 1.2, this is simply too laborious.

The third linking is required to produce generation patterns and forecasts on the higher transmission substations, feeders and transformers.

The linking architecture will depend mainly on the algorithms to be developed in the different sections.

6.15.2 Grouping

The transmission system expansion engineers do not require the total distribution networks to be modelled in PSS/E, therefore some electrical busbar loads have to be grouped into one load: a so-called point load. Because of network topology changes, different busbar groupings to the same point load, over the different years of the forecasting horizon, have to be possible.

This section should also be able to group a number of busbar loads into areas or sections in order to compare the distributors' forecasts with other experts' forecasts for a given area or section. Those experts may be from the mining industry, water supply, population growth, financial or other sectors.

6.15.3 Revising generation information

This section updates the generation cost (if included as a parameter in determining the generation pattern) and the available generation capacity, ie the nett sent out for all Eskom power stations and imports from neighbouring states. Once updated, new generation patterns can be determined for a given condition and a specified network topology (see section 6.15.13).

6.15.4 Revising var equipment information

This information is required in section 6.15.13 to determine the higher transmission substation demands. Three different types of var compensation equipment have to be catered for, ie shunt reactors (absorbing Mvar), shunt capacitors (injecting Mvar) and static var compensators (either absorbing or injecting Mvar). Static var compensator is abbreviated to SVC.

The switching of the var equipment (status) is an important aspect to consider when future loads for the higher transmission substations are being determined.

6.15.5 Updating verified data

The data to be updated includes some billing data to verify the busbar loads, as well as the hourly demands for the lower transmission substations, the power stations and the neighbouring states. The demands for the point loads and the higher transmission substations will also be updated once they become available.

Before updating the data some checks are required to rectify obvious errors.

Because the demands mentioned above are stored, only the verified demands for normal situations, are stored and updated.

6.15.6 Revising completion dates

To meet future demands, either or both of the distribution and transmission network topologies have to be changed. It is therefore important to cater for such network topology changes in the model by making future linking dependent on network completion dates.

6.15.7 Verification

The most important output required from the forecast model is a set of point loads for a given condition, covering the forecasting horizon of ten years (only for the point loads) and reflecting all trends and movements in the electrical markets.

As already mentioned, no recorded historical demand figures exist on those point loads. The distributors use snapshots (instantaneous demand values) during the winter periods to determine the annual maximum demands. In some cases those maximum demands were either too low or too high.

Two rules are applied to verify the maximum demands on the point loads and the substations' demands by using the hourly data on the lower transmission substations and POD demand figures.

The first rule (bottom-up) is: all the annual maximum demands for the electrical loads linked to the busbar are summated and divided by the annual maximum busbar demand. The ratio is defined as

$$\text{Ratio } 1_{it} = \sum_{h \varepsilon C_i} EL_{ht} / BL_{it} \qquad \text{(Eq 6.5)}$$

where

Ratio 1_{it} is ratio 1 for the electrical busbar, BL_{it}, for a given year t

BL_{it} is the i th annual maximum busbar load, for a given year t

EL_{ht} is the h th annual maximum electrical load, for a given year t

C_i is an index set, allocating certain electrical loads to the h th electrical busbar

t = 1 for 1993

2 for 1994

3 for 1995

4 for 1996, etc

i = 1, . . ., number of busbar loads

and should not show any significant increases or decreases between years, except for new loads or loads decommissioned.

This first rule is shown diagrammatically in Fig 6.10. The demands for a number of electrical loads (ELs) are summated and divided by the busbar load (BL) demand. Ratio 1 is tested; if not acceptable (ie outside the defined range), then adjustments are made to the busbar demand.

If there are other electrical loads - usually very small loads - also being fed from the busbar, then it is difficult to determine whether ratio1 is too low or too high. In such cases the user can check only for obvious differences between years. If the verification for ratio 2 has been completed, the model user may run another ratio 1 verification on certain busbar demands.

Ratio 1 is only applied where one or more electrical loads have been linked to a busbar.

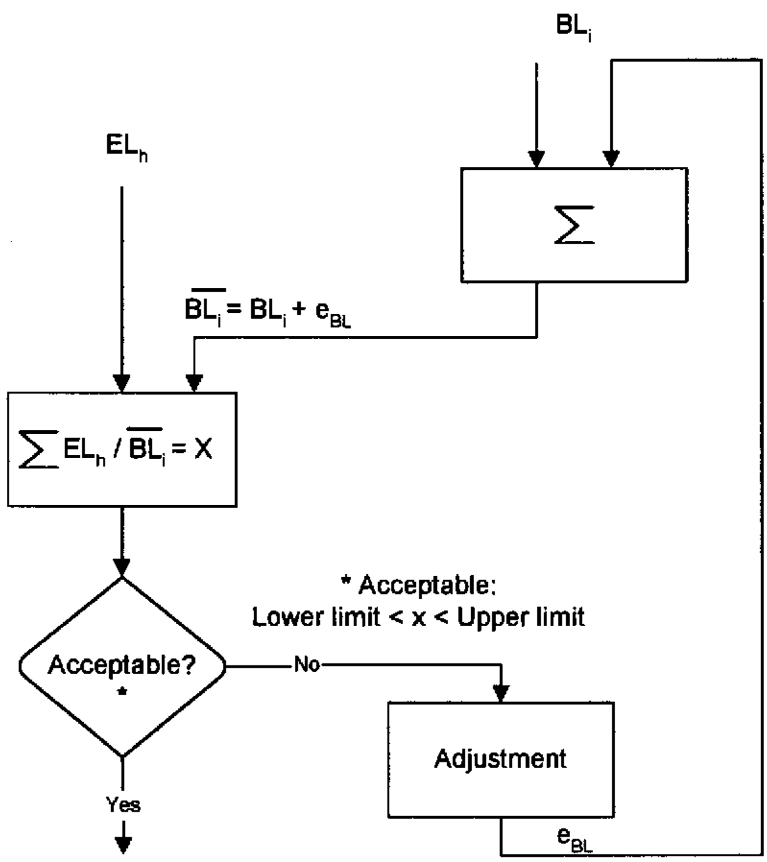


Fig 6.10: Verification section (ratio 1)

The second rule (top-down) is based on the diversification rule. The sum of the annual maximum monthly busbar demands allocated to a lower transmission substation is divided by the annual maximum monthly substation demand.

Ratio 2 is actually the diversification factor between the substation and the busbars. It is advisable to test the annual ratio values between lower and upper limits, as shown in Fig 6.11.

$$\text{Ratio } 2_{kt} = \sum_{i \varepsilon A_k} BL_{it} / T_x(\text{Low})_{kt} \qquad \text{(Eq 6.6)}$$

where

Ratio 2_{kt} is ratio 2 for the k th lower-level transmission substation, for a given year t

T_x (Low) $_{kt}$ is the demand for the k th substation (kW), for a given year t

BL_{it} is the i th annual maximum busbar load, for a given year t

A_k is an index set, allocating certain busbar loads to the k th substation, for a given year t

t = 1 for 1993

2 for 1994

3 for 1995

4 for 1996, etc

k = 1, . . ., the number of lower transmission substations

As previously mentioned, one or more busbars can be allocated to a transmission substation and a busbar can be allocated to more than one substation.

Ratio 2 is tested between its upper and lower limits. If ratio 2 is found to be outside these limits, then adjustments are made to either the substation demand or the busbar demands.

If all the busbar demands are correct and only the substation demand is too high, then ratio 2 will be too low. This happens when the substation supplies other areas not normally supplied from that substation. Graphs are used to determine the additional MW to be subtracted. More advanced techniques are required to smooth such additional demands from the data.

However, if the substation demand is high (not normal) and the busbar demands are also incorrect (some too high and others too low), then the adjustments to either or both of the substation and busbar demands become complicated. In some cases a busbar supplies a POD which is one of a number of PODs supplying an area, making verification even more complicated.

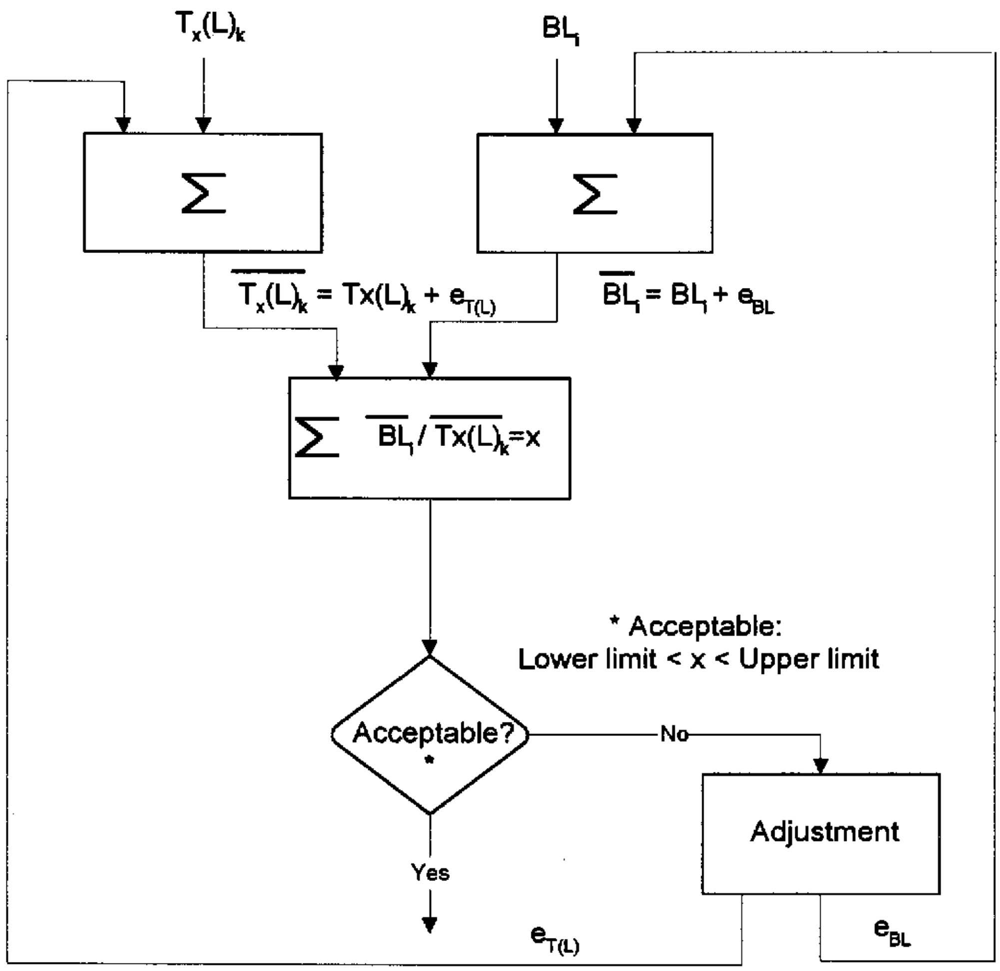


Fig 6.11: Verification section (ratio 2)

The upper and lower limits depend on the supply losses and scaling factors. For example, if a very large point load is fed a considerable distance from the substation, and it is the only load, then ratio 2 will never be 1 - it could be 0.97 or even less. For point loads with small supply losses, ratios of 1.5 or even higher may be obtained. As a rule of thumb, ratio 2 should be larger than 0.5 but smaller than 3.5.

6.15.8 Balancing

Once ratios 1 and 2 are acceptable, the balancing section balances a set of busbar demands to a lower transmission substation maximum demand, according to equation 6.1.

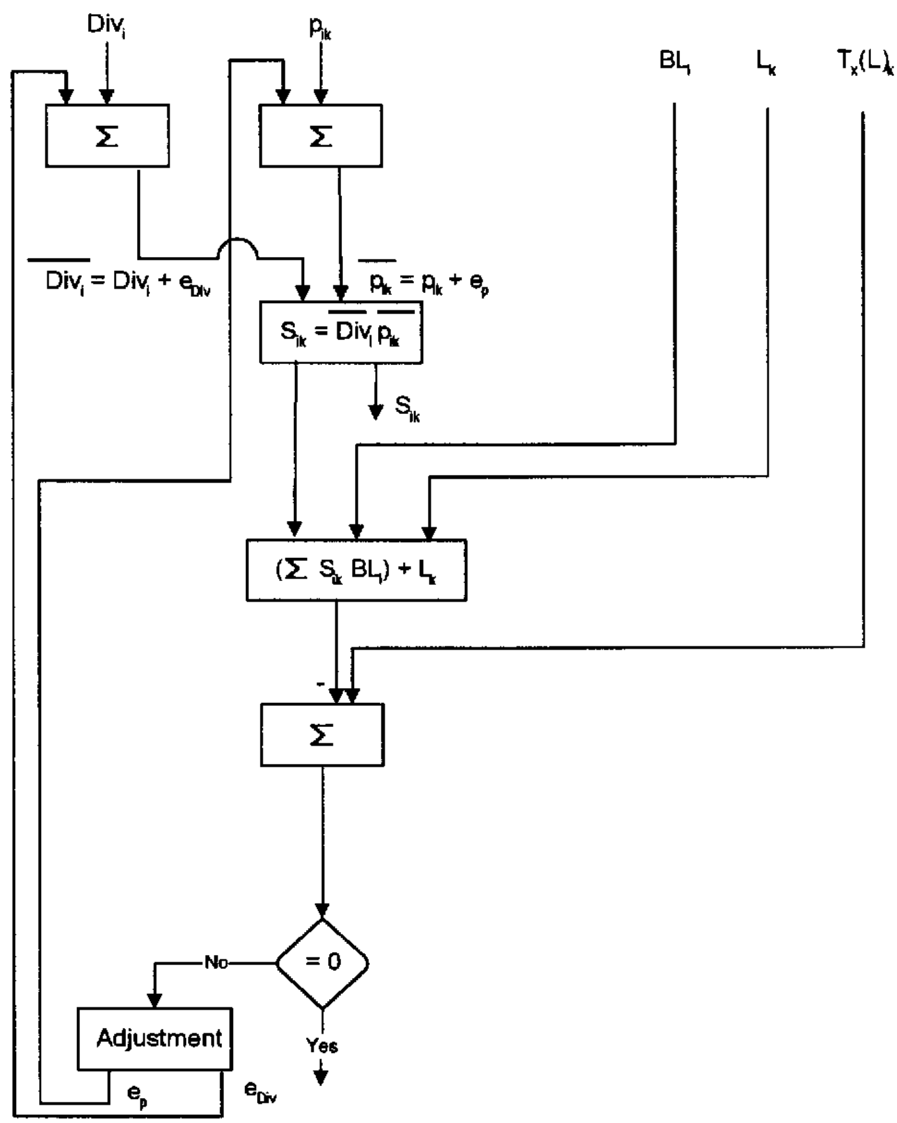


Fig 6.12: Balancing section

It is good practice to check the busbar's diversity and supply factors after each adjustment, to determine whether the factors are still between the allowable upper and lower limits. However, the number of iterations has to be limited to prevent the balancing from continuing indefinitely.

6.15.9 Distributors' forecasts

The forecasts received from the distributors are the annual expected maximum busbar demands. Although those forecasts are verified against area demand forecasts, the distributors' forecasts are a major and important key factor in producing accurate results in the forecasting model.

When revised forecasts are received, checks have to be carried out to determine any major differences between the latest and the previous forecasts.

6.15.10 Factor forecasts

To determine the point loads and lower transmission substation demands, each busbar maximum demand has to be multiplied by a scaling factor. As previously mentioned, a scaling factor consists of a diversity factor and a supply factor, and once the scaled maximum busbar demands have been summated, the summated value has to be added to the supply losses to obtain the transmission substation demand.

The lower transmission substations' demands for a given condition are determined by multiplying the maximum substation demand by a diversity factor. These diversity factor forecasts also have to be included in this section.

Research is required to model future values more accurately, which will in turn improve the load forecast's accuracy.

6.15.11 Area forecasts

Similar to the approach discussed in section 4.2, area load forecasts have to be developed to provide a more accurate basis for balancing the electrical loads. Currently the total of the summated demands for the point loads and the MW due to system losses are balanced with the expected system peak.

Area load forecasts have to convert the views of experts into MW with a given probability of occurrence. If possible, more than one area load forecast is required for a given area, reflecting different experts' expectations of future load growths.

The option of applying long-term forecasting techniques is also required in order to scan, monitor and track future events. Thus a better assessment of new demands or significant changes in existing demands in the future will be obtained, together with probability of occurrence.

6.15.12 Maximum $T_x(L)$ forecasts

The maximum lower transmission substation demand is determined by equation 6.1.

Another factor has been added to the scaling factor, to include the point load in or exclude it from the substation due to network changes. The factor (t), which is used to either exclude the point load from, or include it in, the substation demand, is a binary digit, thus t is either 1 (included) or 0 (excluded).

The scaling factor is defined as

$$S_{ikt} = Div_{it} \, p_{ikt} \, t_{ikt} \qquad \text{(Eq 6.7)}$$

where

- S_{ikt} is the scaling factor for the busbar load, BL_{it}, allocated to the k th substation, for a given year t
- Div_{it} is the diversity factor for the busbar load, BL_{it}, for a given year t
- p_{ikt} is the supply factor for the busbar load, BL_{it}, allocated to the k th substation, for a given year t
- t_{ikt} is a binary digit to exclude (t=0) or to include (t=1) the busbar

The latest information on the completion dates (section 6.15.6) and the forecasts of the factors (6.15.10) are used to determine the scaling factors.

6.15.13 Condition forecasts

This option determines, for a given condition, a set of point loads, lower transmission substation demands, a generation pattern, and the demands on the higher transmission substations.

The lower transmission substation demands are determined by diversifying the maximum lower transmission substation demands with appropriate diversity factors.

Once the busbar maximum demands have been balanced with the lower transmission demands (diversified for the given condition), and the busbar groupings into point loads have been completed, the point load demands can be calculated. In cases where a point load is fed from more than one substation, checks should be built in to ensure that all values have been added. If not, the program should be stopped immediately on receiving a fault message.

The generation pattern is still determined separately from the forecasting model. A more advanced algorithm has to be developed to produce a generation pattern taking into account the percentage losses on the transmission networks, the network topology, the cost of generation, the generation capacity available, the loads on the lower transmission substations and the surplus capacity required on the slack busbar.

For the moment, the generation pattern which is most likely to occur is adapted and adjusted. The adjustments are based on the demand changes on the lower transmission substations, the changes in the network topology, and the generation capacity available.

The generation pattern is simulated on PSS/E and the direction of the MW flow is recorded. The new generation pattern is adjusted and, where possible, an attempt is made to achieve the same flow of the MW. In the event of new or mothballed power station coming into operation, it will not be possible to follow the same pattern.

For such events a "shortest path" approach is followed, taking into account the principles of MW flows in electrical networks.

A possible solution in producing a generation pattern is the application of the theory known as the plant location problem [97]: There are m plants (generating stations or imports from neighbouring states) that produce a single commodity for n customers (lower transmission substations or point loads). Each plant can produce a certain maximum number of units (MW) and each customer requires a certain number of units (MW). There is a fixed cost for each unit in operation and a per-unit cost to ship it from a plant to a customer. This is, exactly, the problem of producing a generation pattern. A branch and bound algorithm is recommended to solve this type of problem.

This branch and bound algorithm solves the LP problem by inspection, which produces a good solution within a reasonable time compared with the standard LP algorithms which explore all possible solutions.

The key factors which have an influence on the demands of the higher transmission substations are the generation pattern, the status of the var equipment, the lower transmission substations' demands and the topology of the networks.

As previously mentioned, hourly MW demand values are available from 1990 for the generation nett sent out figures (generation pattern), and from August 1993 for the lower transmission substations. Information is also available from which the network topology from 1993 can be determined. However, no historical demand values are available for the higher transmission substations.

Forecasts of the higher transmission substations' demands are temporarily processed by means of PSS/E. All loads on the lower transmission substations are updated directly from files created in spreadsheets and the results from PSS/E are imported directly into spreadsheets.

There are, however, three concerns regarding this approach. The first concern is that all the lower transmission substations' demands cannot be modelled at their maximum values, thus the generation pattern has to be increased to cater for the additional demands. If some of the generation stations are at their maximum generation available and other generation stations' outputs have to be increased, incorrect MW flows can be obtained through some of the higher transmission substations.

The second concern is that snapshot values are used to verify the PSS/E results.

If the network topology at the time of the snapshot does not correspond with the PSS/E modelled network, incorrect results can be obtained. To check the network topology against the modelled network is a cumbersome process and not a time-effective approach.

A typical example is that in which a snapshot value of 230 MW was recorded for a specific line, but load flows on PSS/E showed only 53 MW for the same line. The difference in loads was due to the fact that a power station had been out of commission at the time the snapshot was recorded.

Thirdly, using PSS/E to produce forecasts of the higher transmission substations' demands is time-consuming, especially when results are required for a number of scenarios, as spreadsheets have to be created and information then has to be imported into the spreadsheets.

Algorithms for matrix algebra problems may provide an alternative solution for the higher transmission substation forecasts [98}. These algorithms can be incorporated into the forecasting model, because load flow studies are based on matrix algebra algorithms.

6.15.14 Forecasts of expected energy not served

As explained in section 6.13, if the hourly demands (MW) are plotted over time, the area underneath the curve indicates the energy consumption in MWh. If the substation demands exceed the firm capacity, the expansion engineers plot the substation demands for the last twelve months to determine the expected energy at risk (EEAR), ie all the energy above the MW line determined by the demand in MW in terms of firm capacity multiplied by a risk factor. Usually risk factors, of 1.2 are used.

The EEAR is multiplied by a probability of failure to determine the expected energy not supplied (EENS). The EENS (in MWh) is multiplied by a cost factor (R/MWh). Only if this value exceeds the cost of a new transformer, will the next transformer be installed. The cost factor depends on the type(s) of load(s) supplied by the substation.

This section should provide the expansion engineers with the future EEAR values and the corresponding probability of failures which will enable them to determine whether another transformer can be justified and should thus be ordered and installed.

6.15.15 Point load power factor forecasts

This section is very important in terms of the planning to install var compensation equipment (see section 16.15.4) on the transmission networks. If the point load power factors are wrong, no correct var flows can be obtained through the transmission substations. Incorrect var flows may also affect the results of the higher transmission load forecasts.

Although no historical data exists on point load power factors, research is needed to assess more accurately the power factors and possible changes over the years due to the different types of loads to be supplied.

6.15.16 Set of point loads for PSS/E

This option writes the MW and MVar values for a given year and a given condition to a file, which is then used to update PSS/E.

6.15.17 Feeder forecasts

Forecasts are required for all feeders from Eskom power stations as well as from the higher and the lower transmission substations. The feeders from the power stations comprise only feeders to the higher and lower transmission substations. The feeders from the higher transmission substations comprise feeders to the lower transmission substations and the interconnections with other higher transmission substations. The lower transmission substations will obviously only feed into the distribution networks and the interconnections with other lower transmission substations.

The forecasts will not only be done for the three levels, but also per given condition and for each year within the forecasting horizon.

6.15.18 Transformer forecasts

This option is similar to the option discussed in 6.15.14, except that the forecasts required are for all transformers.

6.15.19 Variance reports

This option will compare the actual data captured as mentioned in 6.15.5 with the corresponding forecasting data. All errors (actual data minus forecasting data) indicating that absolute values exceed predefined limits have to be highlighted.

The option of full reports (reporting all errors) or reports in which only those errors that exceed their limits are reported, has to be available. The user should also have the option of selecting reports only on certain sections of the data flow model, such as point loads, lower transmission substations, etc.

6.15.20 Graphical display of linking and grouping

This option is self-explanatory. This section provides a graphical presentation of the linking and grouping options to meet the users' requirements.

6.15.21 Reports to printer

This option is self-explanatory. The number of reports will depend on the users' requirements.

6.15.22 Exit from main menu

This option is self-explanatory. Once the user has completed his tasks, an option is required enabling him to close all data files and to stop the program.

6.16 Further developments

The most important development required is to develop a program to make the model more user-friendly. There are also a number of algorithms to be developed.

Algorithms are required for the verification and balancing sections. This may include techniques to "smooth" the effects of load shifts on the transmission substation data.

More accurate diversity and supply factor forecasts are also required.

It is also important to develop area and section forecasts, together with the exploration of qualitative forecasting techniques to scan, track and monitor future events which are of vital importance to the demand load forecasts.

Appropriate techniques to "smooth" the effects of load shifts on lower transmission substation data should be studied in order to obtain data only for normal situations.

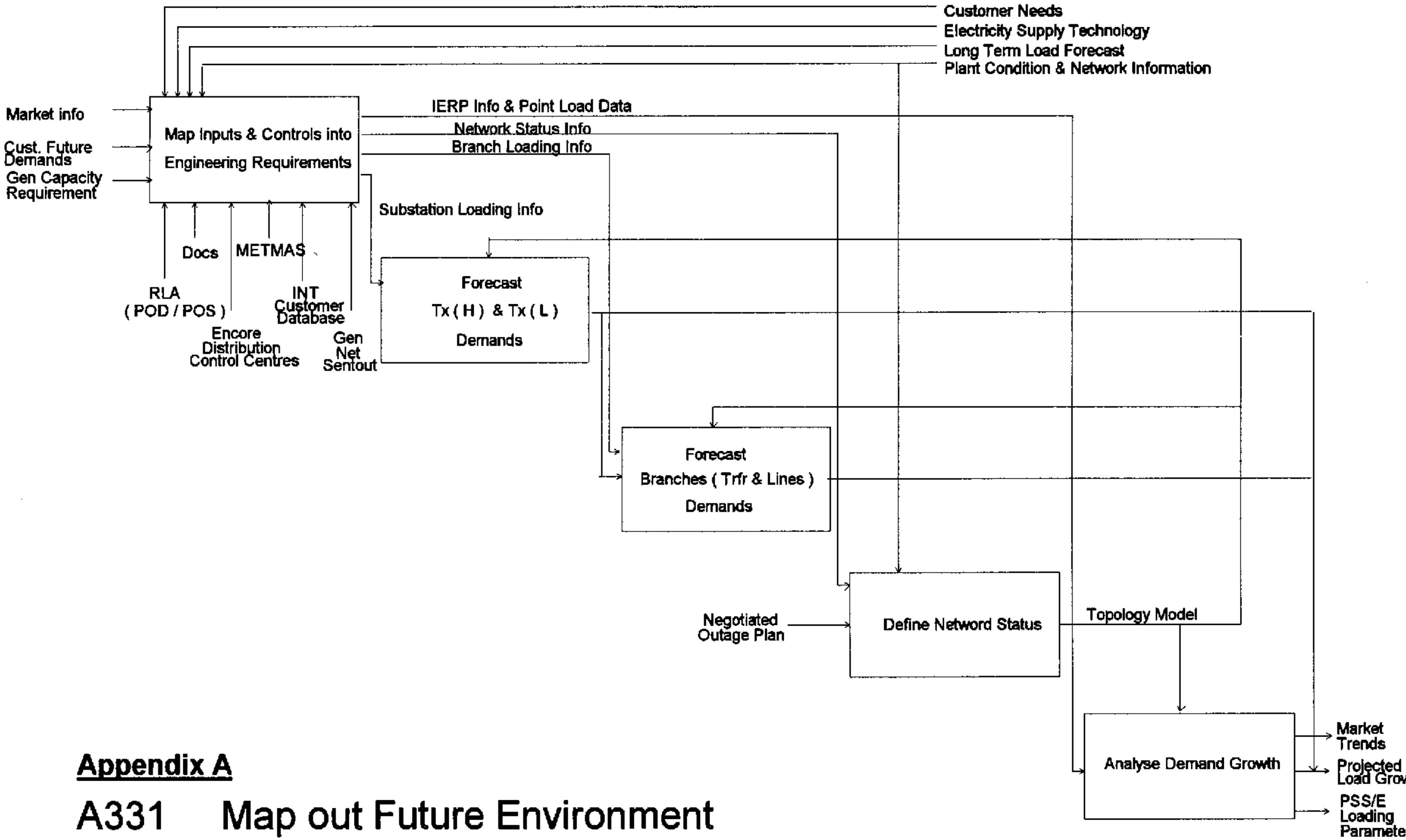
The studying of techniques to provide more realistic generation patterns and higher transmission substation demands also needs attention. In the case of generation patterns, the source nodes should not exceed the maximum available MW of the generation station being modelled. The necessary constraints should also be included in order to include the effect of electrical loads and the network topology. Similar constraints must also be included for the higher transmission substation forecast.

6.17 Conclusion

Despite the fact that verification and balancing are still done deterministically and the distributors' forecasts are only compared with the maximum Eskom system peaks, the results are already much more satisfactory.

Once the proposed program has been developed and all the required algorithms are in place, the model will address all the requirements of the transmission system expansion engineers.

The problem of step transmission substation load increases or decreases causing discontinuities has been overcome by summating the electrical loads according to index sets in order to obtain the substation loads. This also solves the problem of new transmission substations taking loads away from other substations, which also results in step loads on the existing substations. The determination of a set of point loads is also done more accurately by summating some electrical loads according to index sets in order to obtain the point load demands.



Appendix A

A331 Map out Future Environment

APPENDIX B

The Durbin-Watson Statistic

	k=1		k=2		k=3		k=4		k=5	
N	**dl**	**du**	**dl**	**du**	**dl**	**du**	**dl**	**du**	**dl**	**du**
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

k = number of independent variables
N = the number of observations used in the regression

Source: J Durbin and G S Watson "Testing for Serial Correlation in Least Squares Regression, " Biometrika 38 (June 1951) : 73

Index

Glossary

Activation function

A function that transforms the nett input to a neuron into its firing.

Algorithm

A computational step-by-step procedure for setting the weights.

Architecture

Arrangement of nodes and pattern of connection links between them in a neural network.

Artificial neurons

Artificial neurons, also called neuron cells, processing elements or simply nodes or units attempt to simulate the structure and function of biological neurons.

Back-propagation

A learning algorithm for a multilayer neural net based on minimising the squared error.

Binary

0 or 1

Bipolar

-1 or 1

Coefficient of determination (R^2)

The coefficient of determination is the proportion of the explained variation as part of the total variation.

Diversification

Diversification is the scaling of the maximum point load demands to determine a new set of loads for a given condition.

Electrical load

An electrical load can be described as an electrical supply to a customer.

Epoch

One iteration through a neural network.

F statistic

The Fstatistic is the ratio of the two mean squares (regression/residual). The numerator refers to the variance that is explained by the regression; and the denominator refers to the variance of what is not explained by the regression, namely the errors.

Firm capacity

Firm capacity is the maximum megavolt amperes (MVA) a substation can supply when the transformer with the highest rating is out of commission.

Heteroscedasticity

This condition exists when the errors do not have a constant variance across an entire range of values.

Impact of overfitting (regression)

A model is overfitted (overspecified) when variables have been added which contribute little or nothing, but only produce variances that are larger than those of simpler models.

Impact of underfitting (regression)

If the model is underfitted (underspecified), important variables have been ignored.

Interaction

Interaction occurs when two independent variables, both measured at the interval level, are thought to interact in influencing Y such that the slope of the relationship between each independent variable and E(Y) is linearly related to the value of the other independent variable.

Learning algorithms

Procedure for modifying the weights on the connection links in a neural net.

Least squares estimation

This approach estimates the parameter values in an equation and minimises the squares of the errors that result from fitting that particular model.

Multicollinearity

This is a condition with multiple regression models when two or more regressor variables are highly correlated.

Negative autocorrelation

Negative autocorrelation exists when a negative error is followed by a positive error, then another negative error, and so on.

Pattern

Information processed by a neural network; a pattern is represented by a vector with discrete or continuous valued components.

Point of delivery

A point of delivery is the supply from an Eskom substation furnishing a customer with electricity.

Point load

A point load is used to simulate the electricity flow between the transmission electrical networks and the distribution networks or generation networks or neighbouring states' electrical networks, ie where electrical power flows from the transmission networks into the distribution networks or generation networks or the neighbouring states' networks. Thus a point load is the grouping of one or more electrical loads.

Positive autocorrelation

Positive autocorrelation exists when positive errors tend to be followed by other positive errors, while negative errors are followed by other negative errors.

Slack busbar

The slack busbar is used in electrical network studies, to either react as a load (if the nett sent out MW is more than the MW for the loads, including, system losses) or as a power station.

The t-test

The t-test is designed to test whether a regression coefficient is statistically significantly different from zero or not.

Transfer function

See activation function.

Transportation model

A transportation model exists of source, intermediate and demand nodes.

Travelling salesman problem

A classic constrained optimisation problem in which a salesman is required to visit each of a group of cities once before returning to the starting city, finding the shortest route and avoiding subtours.

References

[1] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 12 - 14

[2] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 85 - 86

[3] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 170

[4] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 125

[5] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 22

[6] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 89

[7] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 91 - 92

[8] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 27

[9] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 27, 31

[10] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", pp. 154 - 155

[11] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 155

[12] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 37 - 38

[13] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 29 - 31

[14] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 128

[15] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 32

[16] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 112

[17] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 165

[18] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 166

[19] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 169

[20] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 170 - 177
[21] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 134
[22] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 178
[23] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 179
[24] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 180
[25] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 185 - 186
[26] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 96 - 98
[27] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 210
[28] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 129
[29] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 212
[30] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 61 - 65
[31] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 217 - 221
[32] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 221 - 222
[33] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 136 - 138
[34] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 144 - 145
[35] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 123
[36] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 157
[37] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 127
[38] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 392 - 396
[39] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, p. 394

[40] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 124

[41] Raymond. H Myers 1990 "Classical and Modern Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 411 - 416

[42] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", pp. 129 - 131

[43] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 130

[44] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 277

[45] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 210

[46] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 278

[47] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 286

[48] Damodar N. Gujarati 1988 "Basic Econometrics" second edition, McGraw-Hill Book Company, pp. 329 - 330

[49] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 250

[50] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 254

[51] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 258

[52] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, p. 259

[53] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 260 - 262

[54] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 436 - 439

[55] Raymond. H Myers 1990 "Classical and Modem Regression with Applications" second edition, PWS-KENT Publishing Company, pp. 293 - 315

[56] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p.. 5 - 7

[57] Eric Davalo and Patrick Naïm, translated by A Rawsthorne 1993 "Neural Networks", The MacMillan Press Ltd, pp. 3 - 9

[58] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, p.p. 38 - 41

[59] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 22

[60] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 23

[61] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, pp. 29 - 32

[62] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, pp. 47 - 48

[63] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, pp. 48 - 53

[64] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, pp. 80 - 83

[65] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 24

[66] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, pp. 61 - 75, 216

[67] Eric Davalo and Patrick Naïm, translated by A Rawsthorne 1993 "Neural Networks", The MacMillan Press Ltd, p. 63

[68] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 25

[69] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, pp. 15 - 16

[70] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, p. 89

[71] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 16

[72] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, pp. 98 - 99

[73] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, pp. 113 - 114

[74] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, pp. 127 - 128

[75] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, p. 136

[76] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, p. 132

[77] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, pp. 63 - 66

[78] A Cichocki and R Unbehauen August 1994 "Neural Networks for Optimization and Signal Processing" first edition, John Wiley and Sons, p. 64

[79] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, p. 3

[80] Laurene Faussett 1994 "Fundamentals of Neural Networks" first edition, Prentice-Hall, pp. 38 - 36
[81] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, p. 93
[82] Philip D. Wasserman 1989 "Neural Computing - Theory and Practice" first edition, Van Nostrand Reinhold New York, p. 108
[83] Proceedings of the International Conference held at Nottingham University, edited by O D. Anderson, March 1980 "Forecasting Public Utilities", Nort-Holland Publishing Company - Amsterdam, pp. 197 - 198
[84] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 661
[85] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", pp. 60 - 61
[86] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, pp. 651 - 652
[87] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, pp. 652 - 655
[88] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 653
[89] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 651
[90] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 655
[91] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 655
[92] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p. 658
[93] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, p., 673
[94] Makridakis, Wheelwright and McGee 1993 "Forecasting - Methods and Applications" second edition, John Wiley and Sons, pp. 667 - 668
[95] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", pp. 11 - 12
[96] School of Business Leadership (SBL) 25 - 27 May 1992 "Forecasting with the Aid of Micro Computers", p. 21