A feasibility study into Total Electron Content prediction using Neural Networks

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

of

Rhodes University

by

John Bosco Habarulema

June 2007

Abstract

Global Positioning System (GPS) networks provide an opportunity to study the dynamics and continuous changes in the ionosphere by supplementing ionospheric measurements which are usually obtained by various techniques such as ionosondes, incoherent scatter radars and satellites. Total electron content (TEC) is one of the physical quantities that can be derived from GPS data, and provides an indication of ionospheric variability. This thesis presents a feasibility study for the development of a Neural Network (NN) based model for the prediction of South African GPS derived TEC. The South African GPS receiver network is operated and maintained by the Chief Directorate Surveys and Mapping (CDSM) in Cape Town, South Africa. Three South African locations were identified and used in the development of an input space and NN architecture for the model. The input space includes the day number (seasonal variation), hour (diurnal variation), sunspot number (measure of the solar activity), and magnetic index (measure of the magnetic activity). An attempt to study the effects of solar wind on TEC variability was carried out using the Advanced Composition Explorer (ACE) data and it is recommended that more study be done using low altitude satellite data. An analysis was done by comparing predicted NN TEC with TEC values from the IRI2001 version of the International Reference Ionosphere (IRI), validating GPS TEC with ionosonde TEC (ITEC) and assessing the performance of the NN model during equinoxes and solstices. Results show that NNs predict GPS TEC more accurately than the IRI at South African GPS locations, but that more good quality GPS data is required before a truly representative empirical GPS TEC model can be released.

Acknowledgment

It is my wish to extend my sincere gratitude and high appreciation to my supervisors Dr. Lee-Anne McKinnell and Dr Pierre Cilliers for their guidance towards the success of this study. I am very grateful for all the support that Dr Lee-Anne provided during my project. Thanks to the staff of Hermanus Magnetic Observatory (HMO) for their hospitality, logistical, technical and work support, encouragement during my study.
Thanks to the Chief Directorate Surveys and Mappings (CDSM), Cape Town, South Africa for the provision of GPS data.

The assistance of Dr. Elijah Oyeyemi and Mr. Patrick Sibanda is highly acknowledged. Thanks to Mr. Ben Opperman for his time in helping me to understand and obtain data.

Special thanks go to National Astrophysics and Space Science Programme (NASSP), University of Cape Town/Ford Foundation for my Honours and Masters degree scholarships. The support, care and guidance of the NASSP Coordinator, Professor Peter Dunsby and the Administrator, Mrs Penny Middlekoop is much appreciated.

Thanks to my family for the continuous support since I joined the academic field. A special thanks goes to Mr and Mrs David Rwarinda and their entire family for all the support. Thanks to all my friends and colleagues who assisted me in many ways. The space is not enough to mention all the names.

To my coursemates and friends, Oyapo and Zama, your academic contributions, formal and informal discussions kept me going. Thanks guys.

And greatest of all, I would like to eternally thank the provider of life for having brought me to this world and for keeping me alive for all this time. I accord all due respect to him.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This study aims to assess the suitability of Neural Networks (NNs) for predicting Total Electron Content (TEC) obtained from South Africa's GPS dual frequency receiver network. GPS signals are mostly affected by the ionosphere as they travel from the GPS satellite to the receiver on the ground. The ionosphere has been extensively studied using measurements from satellites, rockets, incoherent scatter radars and ionosondes. Ionospheric research was boosted by the possibility of studying the ionosphere using GPS measurements. The ionosphere is a region of the atmosphere (between an altitude of 50 km to 1000 km) that affects radio wave propagation. It is ionised due to ultraviolet (UV) radiations from the sun. A contribution from particle precipitation near the South Atlantic anomaly (SAA) regions (e.g South Africa) is expected. However, according to Sibanda (2006), the effects of particle precipitation over South Africa as a result of being close to SAA are insignificant. TEC along a ray path through the ionosphere can be considered as a measure of ionisation. TEC estimations using ionosondes (up to an altitude of 1000 km) have been supplemented by the TEC calculated from GPS measurements. Because of the altitude of GPS satellites (about 20 000 km), GPS measurements give information about the topside ionosphere and the contribution of the plasmasphere to the TEC.

1.1 Relevance of the study

TEC variations have significant effects on radio communications, applications involving navigational systems, GPS surveying and space weather. In order to un-

derstand these effects, there is a need to develop nowcasting and forecasting techniques. The ionosphere affects radio signals of all frequencies due to its dispersive character (Klobuchar, 1991). As a GPS signal travels through the ionosphere, a time delay is experienced which is directly proportional to TEC and inversely proportional to the square of its frequency. Radio waves propagating in the ionosphere are affected differently at different altitudes. A radio wave at an altitude of sufficient high plasma density in the ionosphere is reflected and this makes it possible to send signals from one place to another around the Earth (Chen, 1984). TEC has a highly non-linear variability and NNs have been found applicable to modelling which involves physical quantities that exhibit non-linear characteristics (Williscroft and Poole, 1996; Hernandez-Pajares *et al.*, 1997; Xenos *et al.*, 2003; Sarma and Mahdu, 2005; Leandro and Santos, 2007).

1.2 Overview

The ionospheric TEC is influenced by various parameters such as diurnal and seasonal variations and solar and magnetic activities from where the input space for the NN was derived. GPS data from a GPS network of dual frequency receivers was calculated using the Adjusted Spherical Harmonic Analysis (ASHA) model developed by Opperman, (2007). The GPS dual frequency receiver network is operated and maintained by the Chief Directorate Surveys and Mapping (CDSM) based in Cape Town, South Africa.

Diurnal and seasonal variations were quantified and represented by hour (HR) and day number (DN) respectively. The appropriate representation of magnetic activity was derived from the K-index data recorded at Hermanus Magnetic Observatory. The local K index was used since the study aimed at developing a local TEC prediction model, however, the planetary Kp index is also another alternative suitable for representing the geomagnetic activity. Solar activity data was represented by sunspot number of which the historic and current data is readily available for scientific research and other applications.

This study was carried out separately on each of three South African GPS receiver locations, namely the Square Kilometre Array (SKA) Hub location (30.71^{o}S,

21.39^{o}E), Sutherland (32.38^{o}S, 20.81^{o}E) and Springbok (29.67^{o}S, 17.88^{o}E). The choice of the first two locations was primarily due to available GPS data while Springbok was chosen due to its close proximity to the South African Ionosonde located at Louisvale (28.5^{o}S, 21.2^{o}E). Results from Springbok could then be validated with ionosonde measurements.

This thesis consists of six chapters.

In chapter 1 the aim of the project, source of GPS data and an overview are given.

Chapter 2 gives the theoretical background and explains the technique used for TEC modelling.

Chapter 3 provides detailed information and the necessary derivations used to extract TEC from GPS data.

Chapter 4 deals with the identification and determination of optimum parameters using the SKA Hub (30.71^{o}S 21.39^{o}E) data for TEC predictions.

In chapter 5 a broader application of the results obtained in chapter 4 by considering the remaining two GPS receiver stations, is discussed. Both local noon and full (all hours) models to predict TEC are discussed.

The discussions and conclusions are summarised in chapter 6.

Chapter 2

Theoretical background

2.1 Introduction

In this chapter the basic definition of Total Electron Content (TEC) is given. The technique used in TEC modelling (Neural Networks) is described in detail. A detailed discussion of the algorithm (standard back propagation) is presented. A brief discussion of the ionosphere's structure and the relationship between the peak electron density and the critical frequency of the F2 layer are given.

2.2 Radio wave propagation

The understanding of radio waves propagating through a refractive medium (e.g. the ionosphere) is correlated to its refractive index n (defined in terms of magnetic permeability and dielectric constant) (Chen, 1984) as

$$n = (\mu\varepsilon)^{1/2}$$

where
μ is a constant known as the magnetic permeability ($4\pi \times 10^{-7}$ Hm^{-1}).

The dielectric constant ε is defined as

$$\varepsilon = \frac{\text{the strength of the wave's electric field in a vacuum}}{\text{the strength of the wave's electric field in the ionosphere}}$$

The solution for the dielectric constant which is obtained by solving the equation of motion for the trapped electrons in the ionospheric plasma is a function of the plasma frequency, gyrofrequency and the frequency at which the wave propagates. A full discussion of this concept covering all cases for both ordinary and extraordinary wave propagation can be found in Chen, 1984. In this project, the discussion will focuss on TEC variability along a radio wave (GPS signal) in the ionosphere.

TEC in the ionosphere is a variable quantity since it is determined by several dynamic parameters related to solar radiations, solar wind, ionospheric storms, magnetic storms as well as magnetic and ionospheric substorms. TEC is defined as the number of electrons in a column of cross-sectional area 1 m^2 along a path of the signal through the ionosphere. It is expressed in Total Electron Content units (TECU), where 1 TECU=1×10^{16} el/m^2. TEC values can be obtained from Global Positioning System (GPS) data and ionosonde data. The TEC values from these two types of data are different due to the fact that GPS satellites are at an altitude of about 20 000 km while ionosonde TEC (ITEC) is only determined up to a height of about 1000 km. The TEC determined from GPS measurements is slant TEC (TEC along the line of sight) and can be converted to vertical TEC using a suitable mapping function (Opperman, 2007). The determination of GPS TEC is covered in chapter 3 in more detail. ITEC is determined from the electron density derived from ionosonde measurements as a combination of inverted bottomside ionogram and modelled topside (McKinnell *et al.*, 2007). Ionospheric researchers suggest that plasmaspheric contributions are responsible for the differences in the TEC values of the two techniques. The non-linear behaviour of the ionosphere makes it a difficult region to model.

The determination of TEC is important for the prediction of the ionospheric time delay in the applications involving GPS (Sarma and Mahdu, 2005). The ionosphere has a number of definitions. In this project it is defined as a region of the Earth's atmosphere (at altitudes in the range of about 70-1000 km) where radio propagation takes place. It is a constantly changing medium as a result of the ionising radiation that comes in the form of solar Extreme ultraviolet (EUV) and x-ray emissions. The variations of ionospheric parameters affect various scientific missions involving satellites, such as communication efficiency, radar and naviga-

tion systems (since they employ high frequency radio waves) and forecasting of weather conditions. Thus, in order to clearly explain what is happening in the ionosphere and give sufficient information about more events that are likely to occur with time, there is a need to understand the dynamics of the ionosphere. TEC plays a central role in most of the processes that occur in the ionosphere, and the objective of this work is to investigate the possibilities of its (TEC) prediction using Neural Networks.

The delay of electromagnetic signals travelling across the ionosphere is proportional to TEC and this is one of the sources of errors in GPS satellite network based observations (Leandro and Santos, 2007). One of the causes of the delay is ionospheric refraction on the GPS signal which is fully discussed in chapter 3. The ionosphere is a dispersive medium and therefore the refraction which takes place in the ionosphere is a function of frequency. A major factor that is responsible for the attenuation of signals in the ionosphere is the varying refractive index as the signals propagate through different layers. It is unity for radio waves (electromagnetic waves) in a vacuum. However, in the ionosphere the situation is quite different due to the presence of free electrons caused by various processes such as ionisation and radiative recombination of elements. This presence of electrons leads to the decrease of refractive index in the medium (ionosphere) to a value which is less than unity resulting in the increase of the phase velocity of propagation (Chen, 1984) according to the following equation:

$$n = \frac{c}{V_\phi} \tag{2.1}$$

where n is the refractive index, c is the speed of light (3.0×10^8m/s) and V_ϕ is the phase velocity.

This in turn reduces the group velocity of the GPS signal (radio wave) below the frequency of propagation in free space. This follows from the fact that for sufficiently high frequencies, the group velocity and the phase velocity have an inverse relation as follows:

$$V_g \propto \frac{1}{V_\phi} \tag{2.2}$$

where V_g is the group velocity.

Thus the propagation of GPS signals through the ionosphere is not at a constant velocity. It greatly varies as the refractive index changes along the path. In a simple way, TEC is calculated from dual frequency GPS measurements. Mathematically, the difference between propagation times at two frequencies is proportional to the TEC along the path of the signal between the ground based GPS receiver and the GPS satellite. The details of this concept are fully discussed in chapter 3.

2.3 Structure of the ionosphere

The ionosphere is not accurately predicted due to the continuous variability of its parameters. It is defined as the region of the atmosphere at a height between 70 km and 1000 km, although it begins to be detectable at a height of about 50 km (Langley, 2000; Reddy, 2002; McKinnell, 2002). The existence of the ionosphere depends on the UV radiation from the sun which ionises a fraction of the neutral atmosphere of the Earth. The ionosphere is divided into different layers depending on the ion composition, electron density and variability (Reddy, 2002). These are the D (50-90 km), E (90-140 km) and F (140-600 km) layers. The F layer manifests in F1 and F2 layers during the day. The ionosphere exhibits significant changes during nighttime, where the D and F1 layers disappear and the E layer becomes very weak.

Due to this multi-layered structure of the ionosphere, radio signals are affected differently as they propagate through the different layers. The D and E layers have a negligible effect on the GPS frequencies compared with a combination of the E and F1 layers which account for about 10% of the ionospheric time delay encountered in the GPS signals (Yizengaw and Essex, 2002). Maximum electron density occurs in the F layer more specifically in the F2 layer. Beyond 1000 km the ionosphere contains contributions from the plasmasphere or protonosphere, magnetosphere and also the interplanetary plasma (Langley, 2000; Reddy, 2002).

As mentioned before, there are many processes that contribute to continuous variability of the ionosphere, with the most dominant factor being the diurnal

variation.

2.3.1 Critical frequency of the F2 layer, foF2

The F2 layer of the ionosphere is the dominant influence on radio wave propagation. The critical frequency of the F2 layer (foF2) is one of the quantities that can be obtained from an ionosonde. An ionosonde is the traditional technique of measuring the ionospheric properties by vertically transmitting and receiving radio wave signals at frequencies in the range 5-30 MHz (McKinnell, 2002). The outcome from the ionosonde is the graph of frequency against virtual height (h') commonly known as the ionogram, where h' is the apparent height of the ionospheric layer derived from the time interval between the transmission and reception of signals (McKinnell, 2002). In otherwords, h' the height at which a radiowave would be reflected had it continued to travel at the speed of light after entering the ionosphere. Figure 2.1 shows an example of an ideal ionogram with different heights (h'E, h'F1 and h'F2) and critical frequencies (foE, foF1 and foF2) for E, F1 and F2 layers well illustrated. Radio waves may be considered as ordinary or extraordinary waves and f_XF2 is the critical frequency of the F2 layer for the extraordinary waves. The electron density profile n(h) is determined from a non-trivial conversion procedure (scaling) of virtual height profile (ionogram) to a real height profile. The ionosonde TEC (ITEC) is then determined from n(h).

The critical frequency foF2 is related to the maximum electron density in the ionosphere and is therefore important in the estimation of TEC. This helps in the prediction of time delay in GPS applications (Sarma and Mahdu, 2005). It is a function of latitude, longitude, local time, season, solar activity and magnetic activity (Williscroft and Poole, 1996). It is related to the maximum ionospheric electron density as follows:

$$\mathrm{n(max)} = 1.24 \times 10^{10} (foF_2)^2 \tag{2.3}$$

where
foF_2 is the critical frequency in MHz,
n(max) is the peak electron density in electrons m^{-3}.

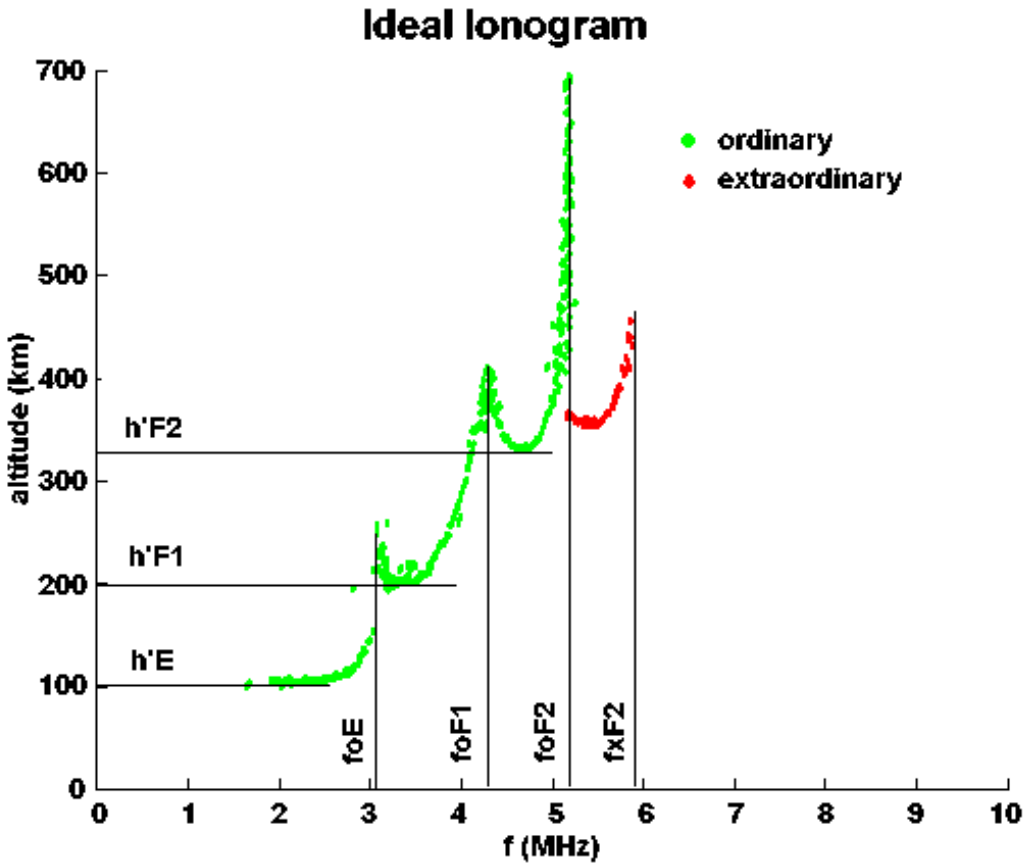


Figure 2.1: A typical ionogram. The green and red colours represent ordinary and extraordinary waves (Source: http://ecjones.org/physics.html).

The measure of the strength of the ionosphere is determined by the foF2 values. The ionosphere is said to be weaker when foF2 values are lower and stronger when foF2 values are higher.

Research has shown that factors which affect foF2, and hence ITEC are the same as those that influence GPS TEC variations in the ionosphere. There is a strong correlation between the TEC and the peak electron density (Kouris *et al.*, 2004; McKinnell *et al.*, 2007). In general, TEC depends on geographic latitude, longitude, local time, season, geomagnetic activity, solar activity and viewing direction in case of GPS TEC (Hofmann-Wellenhof *et al.*, 1992; Reddy, 2002). This concept is further explained in chapter 3.

2.4 TEC models

Currently, there are two existing algorithms in South Africa from where TEC can be infered. The Adjusted Spherical Harmonic Analysis (ASHA) algorithm calculates TEC from the GPS measurements (Opperman *et al.*, 2007). This is an analytical algorithm that relies on the availability of real and near real time data from the network of GPS receivers distributed throughout the country.

The second model is known as the South African Bottomside Ionospheric Model (SABIM) which was developed based on bottomside ionospheric data using the three available South African ionosondes located at Grahamstown (33.3°S,26.5°E), Louisvale (28.5°S,21.2°E) and Madimbo (24.4°S,30.9°E) (McKinnell, 2002). The validity of SABIM is within the confines of the bottomside ionosphere (a region below 350 km in the atmosphere) and can therefore predict TEC up to a maximum altitude of 350 km.

On the other hand, one of the most commonly used global models is the International Reference Ionosphere (IRI) which provides among other physical parameters, monthly average TEC as a function of height, location, local time and sunspot number for magnetically quiet conditions (Bilitza, 2001). The IRI is a standard empirical model that was developed based on all the available data sources (ionosondes, incoherent scatter radars, in-situ measurements on satellites and rockets, etc) and is updated annually. TEC from the IRI is obtained by numerical integration in 1 km steps from 50 km to 2000 km. However, due to the general scarcity of data from the Southern Hemisphere, the IRI does not give accurate predictions over Southern Africa.

Due to the non-linear behaviour of TEC, its nowcasting and forecasting requires statistical and computational modelling techniques. It has been established that the Neural Network technique is capable of generalising the variational trends of various non-linear parameters (e.g. Williscroft and Poole (1996); McKinnell (2002); Sarma and Mahdu (2005); Tulunay *et al.* (2006); Leandro and Santos (2007) and references there in).

2.5 Neural Networks

Neural Networks (NNs) are information processing systems which consist of simple processing elements known as neurons (Haykin, 1994; Bishop, 1995). NNs operate on a principle that is similar to that of the human brain and they are very important tools for non-linear approximation (Bishop, 1995; Barrile *et al.*, 2006). NNs have specific architectures (set up) which are determined by the objectives that they are trained to accomplish. In general, a NN architecture has one input layer with inputs, one hidden layer with neurons and one output layer with the outputs (Tulunay *et al.*, 2006). More about NN architecture is discussed in the following subsections.

NNs have been applied previously in modelling ionospheric TEC using GPS data (Hernandez-Pajares *et al.*, 1997), time dependent TEC predictions and forecasting using Faraday-rotation derived TEC (Xenos *et al.*, 2003), vertical TEC computation and modelling (Leandro and Santos, 2004, 2007) and forecasting of total electron content maps (Tulunay *et al.*, 2006). NNs have also been found applicable to ionospheric modelling involving ionospheric foF2 predictions with reference to the peak electron density in mid-latitude regions (Williscroft and Poole, 1996; Poole and McKinnell, 2000) and equatorial anomaly regions (Sarma and Mahdu, 2005).

For neuron processes to give rise to the output signal, neuron activation functions are used. It is also important to include a term (applied externally) known as the threshold or bias that does the work of increasing or decreasing the neuron input (Leandro and Santos, 2007), although biases and weights may also be considered in one physical quantity (Bishop, 1995).

Generally, a sigmoid function is used when applying feed forward networks with back propagation algorithm. This function may be either a binary or bipolar sigmoid function (Fausett, 1994) and is defined as follows:

$$f(x) = \begin{cases} \frac{1}{1+e^{-ax}} & 0 \leq x \leq 1 \quad \text{binary} \\ \\ \frac{2}{1+e^{-ax}} - 1 & -1 \leq x \leq 1 \quad \text{bipolar} \end{cases} \tag{2.4}$$

where a is the slope parameter of the sigmoid function and x is the input training vector.

A bipolar sigmoid function can also be represented by a tangent hyperbolic function as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.5}$$

where $\tanh(x)$ is the hyperbolic tangent function of x.

Throughout this work, a binary sigmoid function with slope parameter of one is used as a neuron activation function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.6}$$

where $f(x)$ is the binary sigmond function.

The above function is both differentiable and continuous which is a key consideration in choosing activation functions for back propagation algorithm. When presented with data, NNs are capable of learning a trend and generalising when similar data, but not necessarily identical, is presented.

2.5.1 Neural Network architecture

In general, a NN consists of an input layer, a hidden layer and an output layer. The hidden layers are situated between the input layer and the output layer. The number of nodes in the layers is a variable. The number of input nodes in the input layer is equivalent to the number of inputs presented to the NN. The number of hidden nodes also varies and there is no standard formula for their determinination within the hidden layer. The output nodes correspond to the number of outputs. It has been shown that including more than one hidden layer does not lead to much difference in the accuracy of results (Haykin, 1994), although more hidden

layers may make training easier in some cases (Fausett, 1994). The standard feed forward NN with back propagation algorithm was used. Different types of NNs and their applications are described fully in other sources (Haykin, 1994; Fausett, 1994). The input layer generates the input for the first hidden layer. The process is a cascade, thus the output signal of the last hidden layer automatically becomes the input signal of the output layer.

2.5.2 Training, validating and testing

Training is a crucial aspect in NN applications. Before training, data which contains the input and known output parameters is first processed into a form that is compatible with NNs. The data set is divided into the training, validating and testing sets. The training set is the one used for training the NN, the validating set determines the performance of the NN on patterns that are not trained during learning and the testing set is used to assess the final overall performance of the NN (Zell *et al.*, 1998). Training a NN is performed by an adjustment of weights in the data set (Leandro and Santos, 2004, 2007). The initialisation of weights was done randomly to fall within the range $-1 \leq w \leq +1$. The consequence of this adjustment is the tendency to decrease the difference between the predicted and the measured physical quantity under study (TEC in this case). This adjustment is done iteratively with the help of a training algorithm. The validation set helps to avoid over-training.

2.5.3 Single hidden layered networks

This study uses a single hidden layered network which has one input layer, one hidden layer and one output layer. Single layered networks perform well on unseen data (data not used in training) (Fausett, 1994; Bishop, 1995) and thus, the overall performance was assessed on the testing set in some cases. The input layer consists of input parameters, the hidden layer has a number of nodes where summation of random weights and transformation of the resulting weight (w) in the range $-1 \leq w \leq +1$ are performed (Williscroft and Poole, 1996; Zell *et al.*, 1998), and the output layer has an output (predicted TEC). In this study, an algorithm known as back propagation was used. Other training algorithms are detailed in Fausett (1994) and Bishop (1995).

2.5.4 Back propagation algorithm

There are three stages involved when applying back propagation algorithm (Fausett, 1994), namely

- Feed forward of the input training pattern
- Calculation and back propagation of the errors
- Adjustment of the randomly assigned weights

In the input layer each of the input nodes receives an input parameter or signal and transmits it to each hidden unit in the hidden layer. Each hidden node in the hidden layer calculates its activation using a suitable activation function as defined in equation (2.6) and forwards the signal to the output node in the output layer (full connection). The output unit does the computation of its activation leading to the result or response for each given input pattern (Fausett, 1994). As the training continues, there is a comparison between the output node's calculated activation and its target value in order to estimate the associated error for that training pattern with that output node. The estimation of error is done by using a gradient descent technique (Bishop, 1995), in which weights and biases are considered to fall within a single weight for a particular training pattern. From this estimated error, a factor is computed which distributes back the error to the hidden nodes and updates the random weights between the hidden and output layers (Fausett, 1994; Bishop, 1995).

Using the same procedure, a different factor is computed for each of the nodes in the hidden layer which updates the weights between the input and the hidden layers. The weights are adjusted simultaneously during training until the number of iterations specified in the NN have been completed.

It is important to note that weights are chosen randomly. In this study, the NN was initially instructed to randomly choose weights in the range $-1 \leq w \leq 1$. Weights adjustment and error estimation are influenced by the choice of the learning parameter (α). If it is too small, the reduction error is low and divergent oscillations occur if it is too large (Bishop, 1995; Zell *et al.*, 1998).

The flow chart in Figure 2.2 shows the complexity of a training algorithm during NN applications. It clearly shows all three phases applied in the back propagation algorithm. In this figure, biases and weights are updated iteratively by means of addition processes (Fausett, 1994) as follows:

Hidden unit: $v_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$
Output unit: $w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$.

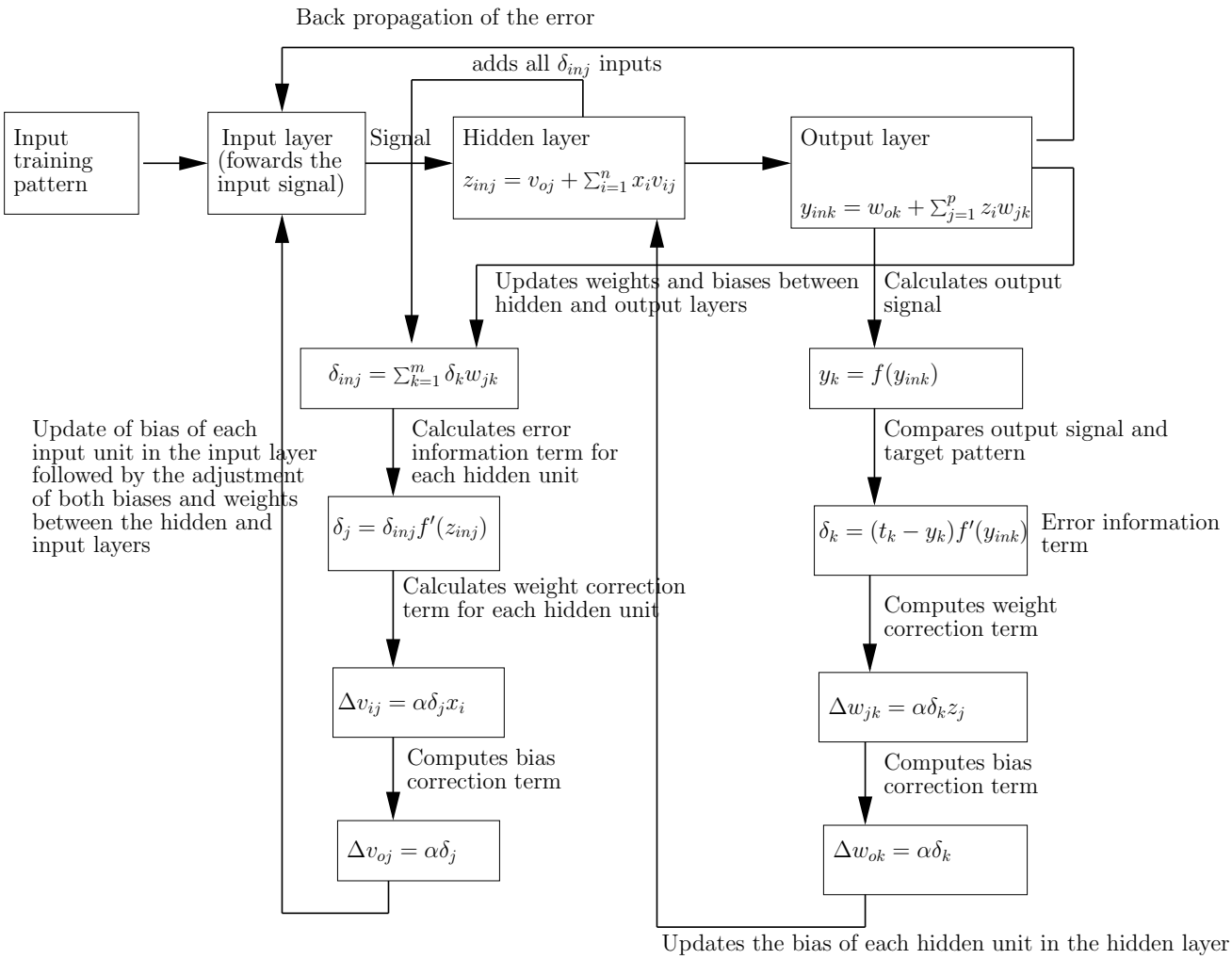


Figure 2.2: Back propagation algorithm. The output signals at each hidden layer and the output layer are calculated after applying the activation functions at the two layers respectively. This training algorithm was adopted from Fausett (1994), pp. 289-302.

The symbols used in the above flow chart are defined as follows:

α is the learning parameter ($0 \leq \alpha \leq 1$)
w_{ok} is the bias on the output unit k
x_i is the input/output signal of the input unit i
y_{ink} is the weight input signal at the output unit k
y_k is the output signal of the output unit k
t_k is the target signal of the output unit k
δ_k is the error information term of output unit k
δ_j is the error information term at hidden unit j
δ_{inj} is the sum of all delta input units at hidden unit j
v_{oj} is the bias on hidden unit j
z_{inj} is the net input at the hidden unit j
v_{ij} is the updated bias between input unit i and hidden unit j
w_{jk} is the updated weight between hidden unit j and output unit k
m is the number of outputs ($m = 1$ in this case)
n is the number of input units
p is the number of hidden units.

2.6 Summary

TEC has been defined as the integral number of electrons in a unit cross sectional area along a signal path from the GPS satellite to the receiver on the ground. The ionosphere with reference to GPS signal propagation has been discussed. The technique of NNs with a focus on the application of one training algorithm (back propagation) has been presented.

Chapter 3

Deriving TEC from GPS data

3.1 Introduction

In this chapter an introduction to the Global Positioning System (GPS) is given. The effects of the ionospheric variability on GPS measurements are discussed with derivations for the differences between phase or group geometric ranges and true geometric range in terms of TEC. A brief description of the algorithm used to calculate TEC values used in this study is presented.

3.2 The Global Positioning System (GPS)

The Global Positioning System (GPS) is a space-based navigation system consisting of a constellation of at least 24 satellites orbiting at a height of about 20 000 km from the Earth (Hofmann-Wellenhof *et al.*, 1992; Reddy, 2002). GPS satellite orbits are inclined at an angle of 55^o relative to the equator. At least four satellites are observable from any point on or near the Earth. The satellites are operated and maintained by the United States Department of Defense (USDoD) for accurate determination of position, velocity and time (NAVSTAR GPS, 1996). It was originally initiated for US military purposes, but civil use was later allowed and various applications were started, including surveying and scientific research.

The GPS consists of three main segments, namely space, control and user segments.

The space segment consists of 24 fully operational satellites in six orbital planes with four satellites in each plane.

The control segment consists of monitor and control stations which track satellites in view, thereby gathering ranging data that is processed at the master control station to ensure that the satellites remain in their orbits (NAVSTAR GPS, 1996). Using the ground antennas for transmission, the satellites' navigational messages are updated from the control stations.

The user segment comprises antennas and user receivers that give velocity, accurate positioning and precise timing information to the users. The radio navigation GPS receivers are capable of receiving, decoding and processing GPS satellite ranging codes and navigation data messages (NAVSTAR GPS, 1996). GPS receivers may be hand-held, installed in particular places (e.g. GPS receiver stations) or placed in cars, depending on the user requirements.

Signals from GPS satellites are transmitted at two L-band frequencies, L1 (1575.42 MHz) and L2 (1227.60 MHz). Both are derived from the fundamental frequency f_o (10.23 MHz), where L1=154f_o and L2=120f_o. These different frequencies on GPS satellites are meant to carry out different functions. The GPS navigation messages, Standard Positioning Service (SPS) code signals (coarse acquisition code, C/A-code) and the Precise code (P-code), mainly used for Precise Positioning Service (PPS) are carried by the L1 frequency. The L2 also carries the P-code used for the PPS. The P-code is modulated on both L1 and L2 carrier frequencies (Dana, 1997). The differences for both the carrier phase and pseudorange measurements at L1 and L2 give an estimate of the ionospheric delay from which TEC can be derived.

The PPS was originally restricted to unauthorised users by the activation of two features, namely Selective Availability (SA) and Anti-Spoofing (A-S). SA is the denial of full accuracy to unauthorised users by degrading the accuracy of GPS position, velocity and time, while A-S is a safety feature used to prevent false transmissions of the satellite data through the encryption of the P-code into the Y-code. The C/A-code is free of A-S, while SA was turned off in 2000. PPS

receivers have the advantage of using both the P(Y)-code (encrypted P-code) and C/A-code (NAVSTAR GPS, 1996).

The GPS navigation signal carries various kinds of information. The signal has modulated time-tagged data bits that allow travel time to be estimated as the difference between the time of origin from the GPS satellite and the time of arrival on the GPS receiver (Dana, 1997). Among the information contained in the GPS navigation message is the ionospheric model used in the GPS receiver to estimate the phase delay at any location and time through the ionosphere. Dual frequency P-code receivers are used to correct ionospheric range errors by combining pseudoranges observed on L1 and L2, while single frequency receivers with C/A-code provide reduced measurement accuracy and therefore a correction model for the ionospheric error is needed in order to obtain improved measurements (Klobuchar, 1991). The turning off of SA made the ionosphere the largest source of error in GPS positioning and navigation (Gao and Liu, 2002).

Under the theme "Modernisation of GPS" more frequencies are being added to the existing ones. L3 (1381.05 MHz) was added specifically for use by the Defense Support Program for purely military purposes and under study is L4 (1379.913 MHz) to improve ionospheric correction. Another satellite is scheduled to be launched in 2008 to provide GPS signal information at L5 (1176.45 MHz) which is intended to have less or no interference.

3.3 TEC and GPS measurements

The Global Positioning System (GPS) measurements can be used to give an estimate of the number of electrons (TEC) along a ray path between a GPS satellite and the receiver on the ground. TEC has been defined as the total number of electrons in the ionized plasma contained in a unit cross-sectional area (1 m^2) between a GPS satellite and a GPS receiver. GPS measurements are recorded by either single or dual frequency GPS receivers. In the estimation of TEC dual frequency receivers are preferred, because they are designed to eliminate ionospheric error which is the major source of error in GPS measurements. The GPS data is recorded in Receiver Independent Exchange (RINEX) format and is then con-

verted to GPS observable files using a suitable model. These GPS observables are either code pseudoranges (P) or carrier phase (Φ) measurements. The time delay between GPS signals at the two frequencies L1 and L2 is proportional to the total electron content along the signal's path.

TEC from carrier phase measurements has relatively less noise, although there is the presence of an unknown ambiguity in a sense that signals transmitted at L1 and L2 may not be easily differentiated from each other. On the other hand, TEC from code pseudoranges measurements is free of ambiguity, but with relatively much noise. Linearly combining both code pseudorange and carrier phase measurements give TEC a better degree of accuracy, although it is also not absolute TEC. Pseudorange and phase measured quantities are mathematically expressed (Gao and Liu, 2002; Barrile *et al.*, 2006) as follows:

$$P_1 = \rho + c(dt - dT) + d\rho + d_T + d_{IP_1} + b^S_{P_1} - b^R_{P_1} + \varepsilon(P_1) \tag{3.1}$$

$$P_2 = \rho + c(dt - dT) + d\rho + d_T + d_{IP_2} + b^S_{P_2} - b^R_{P_2} + \varepsilon(P_2) \tag{3.2}$$

$$\Phi_1 = \rho + c(dt - dT) + d\rho + d_T + \lambda_1 N_1 - d_{I\Phi_1} + b^S_{\Phi_1} - b^R_{\Phi_1} + \varepsilon(\Phi_1) \tag{3.3}$$

$$\Phi_2 = \rho + c(dt - dT) + d\rho + d_T + \lambda_2 N_2 - d_{I\Phi_2} + b^S_{\Phi_2} - b^R_{\Phi_2} + \varepsilon(\Phi_2) \tag{3.4}$$

where
P_1 and P_2 are the pseudorange measurements at L1 and L2,
Φ_1 and Φ_2 are the carrier phase measurements at L1 and L2,
ρ is the true geometric range (m) between the satellite and the receiver,
dρ is the orbital error (m),
c is the speed of light (3.0×10^8 m/s),
dt is the satellite clock error with respect to GPS time (s),
dT is the receiver clock error with respect to GPS time (s),
λ_1 and λ_2 are the wavelengths of the carrier signals at L1 and L2 frequencies,
N_1 and N_2 are the carrier phase integer ambiguities,
d_T is the tropospheric error,
d_{IP_1} and d_{IP_2} are the pseudorange ionospheric delays (m) at L1 and L2
$b^S_{P_1}$ and $b^S_{P_2}$ are the pseudorange satellite delays (m) at L1 and L2
$b^R_{P_1}$ and $b^R_{P_2}$ are the pseudorange receiver delays (m) at L1 and L2

$d_{I\Phi_1}$ and $d_{I\Phi_2}$ are the ionospheric carrier phase delays (m)
$b^S_{\Phi_1}$ and $b^S_{\Phi_2}$ are the carrier phase satellite delays or interfrequency biases (m)
$b^R_{\Phi_1}$ and $b^R_{\Phi_2}$ are the carrier phase receiver delays (m)
$\varepsilon(P_1)$ and $\varepsilon(P_2)$ are the pseudorange measurement noises which include multipath errors (m)
$\varepsilon(\Phi_1)$ and $\varepsilon(\Phi_2)$ are the carrier phase measurement noises with multipath errors included (m).

If there is no loss of lock, the unknown carrier phase ambiguities (N1 and N2) remain constant, but this is not a guarantee as sometimes there is rapidly changing TEC as a result of ionospheric scintillation. A detailed treatment of how ionospheric and tropospheric errors are eliminated can be found in Hofmann-Wellenhof *et al.*, 1992. Multipath errors occur when the signal arrives at the receiver through multiple paths and may be caused by any reflective surfaces near the antenna (Gao and Liu, 2002). Both satellite and receiver delays (pseudorange and carrier phase) are associated with the satellite and receiver hardware.

Ionospheric delay is obtained by differencing either pseudorange or carrier phase measurements:

$$P_2 - P_1 = d_{IP_2} - d_{IP_1} + (b^S_{P_2} - b^S_{P_1}) - (b^R_{P_2} - b^R_{P_1}) + \varepsilon(P_2) - \varepsilon(P_1) \tag{3.5}$$

$$P_2 - P_1 = d_{IP} + b^S_P - b^R_P + \varepsilon(P_2) - \varepsilon(P_1) \tag{3.6}$$

where
$d_{IP} = d_{IP_2} - d_{IP_1}$ is the ionospheric delay,
$b^S_P = b^S_{P_2} - b^S_{P_1}$ is the differential satellite delay,
$b^R_P = b^R_{P_2} - b^R_{P_1}$ is the differential receiver delay between L1 and L2 frequencies

Ionospheric delay d_{IP} contains information from which TEC is derived.
Subtracting the carrier phase observations gives

$$\Phi_2 - \Phi_1 = \lambda_2 N_2 - \lambda_1 N_1 - d_{I\Phi_1} - d_{I\Phi_2} + b^S_{\Phi_2} - b^S_{\Phi_1} + b^R_{\Phi_1} - b^R_{\Phi_2} + \varepsilon(\Phi_2) - \varepsilon(\Phi_1) \tag{3.7}$$

$$\Phi_2 - \Phi_1 = \lambda N - d_{I\Phi} + b^S_\Phi - b^R_\Phi + \varepsilon(\Phi_2) - \varepsilon(\Phi_1) \tag{3.8}$$

where
$\lambda N = \lambda_2 N_2 - \lambda_1 N_1$
$d_{I\Phi} = d_{I\Phi_2} - d_{I\Phi_1}$
$b^S_\Phi = b^S_{\Phi_2} - b^S_{\Phi_1}$
$b^R_\Phi = b^R_{\Phi_2} - b^R_{\Phi_1}$.

Equations (3.6) and (3.8) show that the satellite and receiver biases should be estimated simultaneously with the ionospheric delay parameters. Due to a relatively high noise level in the code pseudorange measurements, the carrier phase measurements are used for smoothing the code observations through a technique known as "carrier phase levelling" in order to achieve more precise vertical TEC estimates (Gao and Liu, 2002).

3.3.1 Ionospheric effects on GPS signal

Due to the dispersive nature of the ionosphere, the delay in GPS signals is inversely proportional to square of frequency and directly proportional to TEC (Reddy, 2002). Using dual frequency GPS observations, one can take the advantage of the ionosphere's dispersive character to compute TEC (Langley, 2000; Fedrizzi *et al.*, 2005). Ionospheric dispersion is almost similar to dispersion as applied in optics, which is the splitting of light into its constituent colours through a prism. When measurements of frequencies are taken at two well-spaced intervals and linearly combined, a large percentage of ionospheric effects is eliminated. This is the major reason why GPS signals are transmitted at two different carrier frequencies (Klobuchar, 1991).

Radio waves (GPS signals) are mostly affected by the ionosphere as they travel through the Earth's atmosphere. To effectively quantify the effects of propagation on a GPS signal travelling through the Earth's atmosphere, its refractive index should be known. This is important for the derivation of the total number of electrons (TEC) in the ionosphere along a GPS signal as it travels from the satellite to the receiver (Yizengaw and Essex, 2002). The propagation speed of a GPS signal in the ionosphere is characterised by its refractive index defined in chapter 2, equation (2.1). Thus, the refractive index is a fundamental quantity in the determination of TEC from GPS signals.

In a dispersive medium like the ionosphere, the phase velocity from which phase refractive index is derived, is a function of frequency (Langley, 2000). The phase refractive index of the ionosphere was derived by Edward Appleton and Douglas Hartree in early the 1930's as

$$n^2 = 1 - \frac{X}{1 - iZ - \left(\frac{Y_T^2}{2(1-X-iZ)}\right) \pm \left(\frac{Y_T^4}{4(1+iZ-X)^2} + Y_L^2\right)^{\frac{1}{2}}} \tag{3.9}$$

where X, Y and Z are dimensionless quantities defined by the following expressions

$$X = \omega_p^2/\omega^2$$
$$Y = \omega_{\mathbf{B}}/\omega$$
$$Z = \nu/\omega$$

and

$$Y_L = \omega_L/\omega = \frac{\omega_{\mathbf{B}} cos\theta}{\omega}$$
$$Y_T = \omega_T/\omega = \frac{\omega_{\mathbf{B}} sin\theta}{\omega}$$

where
ω_p is the angular plasma frequency,
$\omega_{\mathbf{B}}$ is the electron gyrofrequency, $\omega_{\mathbf{B}}$=$\mathbf{B}$e/m,
ω_L is the longitudinal component of $\omega_{\mathbf{B}}$,
ω_T is the transverse component of $\omega_{\mathbf{B}}$,
ν is the electron collision frequency,
θ is the angle between the direction of propagation of the wave and the geomagnetic field,
$\mathbf{B}$ is the magnetic field strength.

When frequencies are high and the electron density concentration is low, the collisions are few, thus the collision term may be ignored, i.e. $\nu \simeq 0$ and the Z term vanishes. For transverse or near transverse waves $\theta \simeq 90^o$ and $Y_L \simeq 0$ and the

Appleton-Hartree equation reduces to

$$n^2 = 1 - \frac{X}{1 - \left(\frac{Y_T^2}{2(1-X)}\right) \pm \left(\frac{Y_T^2}{2(1-X)}\right)} \tag{3.10}$$

Thus

$$n_+^2 = 1 - \frac{X(1-X)}{1-X-Y_T^2} \qquad \text{Extra-ordinary wave and}$$

$$n_-^2 = 1 - X \qquad \text{Ordinary wave.}$$

The ordinary wave is the one that is not affected by the magnetic field (Chen, 1984). Taking the refractive index n as complex (n=$\mu + i\chi$) and by ignoring the effects of magnetic field (Langley, 2000), both the ordinary and extraordinary waves are the same:

$$n_\pm^2 = (\mu + i\chi)^2 = 1 - X. \tag{3.11}$$

The real part of the above equation gives

$$\mu^2 = 1 - X$$
$$\mu = 1 - \frac{1}{2}X + \frac{\frac{1}{2}(\frac{1}{2}-1)}{2!}X^2 + ...$$

Neglecting higher powers of X

$$\mu \approx 1 - \frac{1}{2}X = 1 - \frac{\omega_p^2}{2\omega^2}. \tag{3.12}$$

Angular plasma frequency is given by

$$\omega_p^2 = \frac{N_e e^2}{m\epsilon_o}$$
$$\omega_p = 2\pi f_p, \quad \omega = 2\pi f$$

where
ω_p is the angular plasma frequency,
ω is the angular wave frequency,
N_e is the electron density (per m^3),
e is the charge of an electron (-1.602×10^{-19} C),
ϵ_o is the permittivity of free space (8.854×10^{-12} Fm^{-1}),
m is the electron rest mass (9.11×10^{-31} kg),
f_p is the plasma frequency (Hz),
f is the signal carrier frequency (Hz).

Thus

$$\mu \approx 1 - \frac{f_p^2}{2f^2} = 1 - \frac{N_e e^2}{8\pi^2 m \epsilon_o f^2} \tag{3.13}$$

Substituting the respective values for electron mass, electron charge and the permittivity of free space, the above phase refractive index expression reduces to

$$\mu_\phi = 1 - \frac{40.28 N_e}{f^2} \tag{3.14}$$

The following are the assumptions used in the Appleton-Hartree equation:

1. The number of ions and free electrons in the ionosphere are equal.

2. Since the mass of an ion is greater than the electron mass, electrons are more mobile and thus ions have a negligible effect on radio waves.

3. Ionospheric electrons oscillate at an angular plasma frequency.

A modulated GPS signal is considered to be a superposition of GPS signals of different frequencies which are all centred on the carrier frequency. For the case of the dispersive ionosphere, each of the signals makes up a group travelling at a different velocity from that of the carrier signal. The velocity at which the modulated signal propagates is known as the group velocity (Hofmann-Wellenhof *et al.*, 1992; Langley, 2000). The group refractive index can be written in terms

of the phase refractive index (Hofmann-Wellenhof *et al.*, 1992; Langley, 2000) as

$$\mu_g = \mu_\phi + f\frac{d\mu_\phi}{df} \tag{3.15}$$

Differentiating the phase refractive index gives

$$d\mu_\phi = \frac{80.56N_e}{f^3}df \tag{3.16}$$

Substituting equations (3.14) and (3.16) in equation (3.15), and simplifying gives

$$\mu_g = 1 + \frac{40.28N_e}{f^2} \tag{3.17}$$

Thus the phase and group velocities of the signal would be

$$V_\phi = \frac{c}{1 - \frac{40.28}{f^2}N_e} \tag{3.18}$$

$$V_g = \frac{c}{1 + \frac{40.28}{f^2}N_e} \tag{3.19}$$

where c is the speed of light.

From chapter 2 (equations (2.1) and (2.2)), the refractive indices can be re-written as

$$n_\phi = 1 - \frac{40.28}{f^2}N_e \tag{3.20}$$

$$n_g = 1 + \frac{40.28}{f^2}N_e \tag{3.21}$$

where n_ϕ and n_g are phase and group refractive indices respectively.

3.4 Refraction of GPS signals by the ionosphere

The refractive index of the ionosphere is a variable due to the constantly changing number of electrons in different layers. As the GPS signal propagates through the

ionosphere, it is bent as a result of Fermat's principle. Fermat's principle states that; of all paths that can be taken by an electromagnetic wave (e.g. a light wave, a radio wave or a GPS signal) travelling from a point to another, it will take the path which requires the least amount of time (Hofmann-Wellenhof *et al.*, 1992; Langley, 2000). Thus the determination of ionospheric refraction follows from the expression

$$ndX = cdt \tag{3.22}$$

where
n is the refractive index,
dX is the distance travelled by the signal,
dt is the time taken to travel a distance dX in the ionosphere.

The time taken by the GPS signal to move from the satellite to the receiver on the ground can be determined by integration over the entire path signal

$$t = \int_x \frac{n}{c} dX. \tag{3.23}$$

The measured geometric phase and group ranges are obtained by substituting the expressions for phase and group refractive indices in the expression

$$\rho = ct = c \int_x \frac{n}{c} dX \tag{3.24}$$

where ρ is the geometric range or distance.

Therefore

$$\rho_\phi = \int_x \left(1 - \frac{40.28 N_e}{f^2}\right) dX \qquad \text{Phase geometric range} \tag{3.25}$$

$$\rho_g = \int_x \left(1 + \frac{40.28 N_e}{f^2}\right) dX \qquad \text{Group geometric range.} \tag{3.26}$$

Re-writing the above equations results in

$$\rho_\phi = \int_x dX - \frac{40.28}{f^2} \int_x N_e dX \tag{3.27}$$

$$\rho_g = \int_x dX + \frac{40.28}{f^2} \int_x N_e dX \tag{3.28}$$

where
$\int_x dX = \rho$ (true geometric range),
$\int_x N_e$ dX = TEC (total electron content).

The differences between the phase or group and true geometric ranges give the ionospheric refraction

$$\rho_\phi - \rho = \Delta_\phi = -\frac{40.28}{f^2} T \tag{3.29}$$

$$\rho_g - \rho = \Delta_g = \frac{40.28}{f^2} T \tag{3.30}$$

where T in the above equations is TEC.

It can be deduced that the ionosphere reduces the carrier phase measurements (phase advance) and increases the pseudorange measurements (signal delay) by equal but opposite amounts (Langley, 2000; Gao and Liu, 2002). TEC in the above equations is slant TEC (along the signal path) and therefore to obtain vertical TEC (VTEC), a suitable mapping function is used which takes into account the satellite zenith distance (Hofmann-Wellenhof *et al.*, 1992), as follows:

$$\Delta_\phi = -\frac{1}{cos(z')} \frac{40.28}{f^2} T \tag{3.31}$$

$$\Delta_g = \frac{1}{cos(z')} \frac{40.28}{f^2} T \tag{3.32}$$

where z' is the zenith angle at the ionospheric piercing point.

The zenith angle at the ionospheric piercing point (IPP) can be expressed as (see Figure 3.1)

$$\sin(z') = \frac{R_e}{R_e + H} \sin(z) \tag{3.33}$$

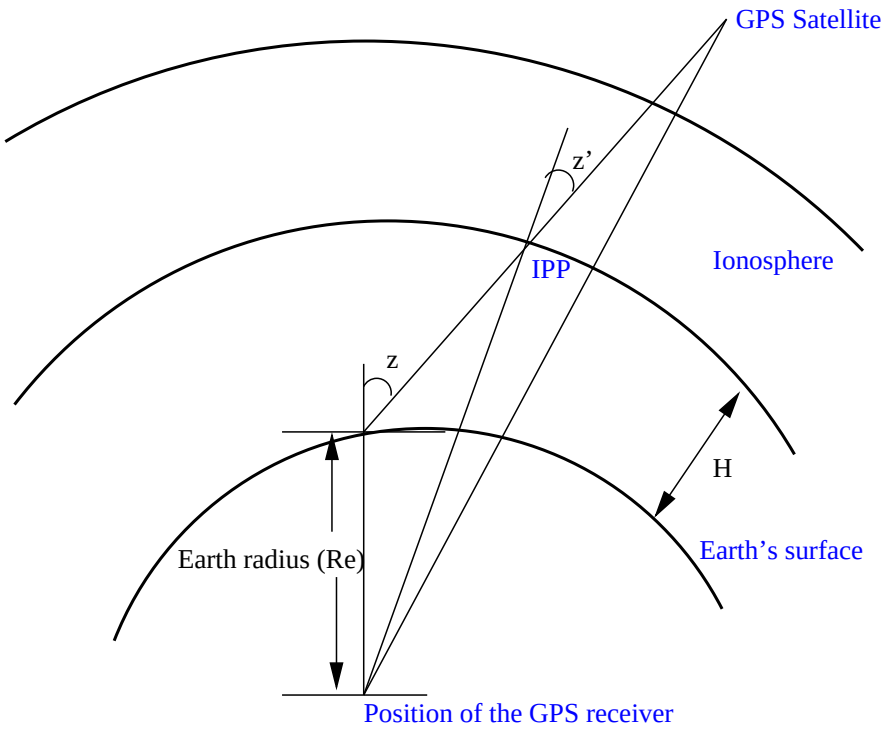


Figure 3.1: Geometrical representation of GPS signal delay due to the ionosphere. Adopted from Hofmann-Wellenhof *et al.,* 1992.

where
z is the satellite zenith angle at the point of observation (GPS receiver location),
R_e is the Earth equatorial radius (6378.134 km),
H is the assumed single layer model height of 350 km.

This is the mapping function which was used in derivation of Vertical TEC (VTEC) from slant TEC (STEC). The ionosphere is assumed to be a single layer at a height of 350 km. The IPP is the point of intersection between the GPS signal and the ionospheric shell (Hofmann-Wellenhof *et al.*, 1992; Yizengaw and Essex, 2002; Gao and Liu, 2002). In particular, the STEC is the ionospheric delay at the IPP above a certain observation point O on the ground, and the aim of the mapping function is to allow a geometric conversion of this slant ionospheric delay to vertical ionospheric delay at the same point O.

3.5 The Adjusted Spherical Harmonic Analysis (ASHA) model

The ASHA model is a GPS-based algorithm that was developed by Opperman, (2007) for regional ionospheric TEC mapping and prediction service over Southern Africa. It is based on the global Spherical Harmonic Model (SHM) used by Center for Orbit Determination in Europe (CODE), and contains an algorithm which allows for the derivation of TEC from the GPS receiver data (Opperman, 2007). In the ASHA model, the co-latitude (90^o-latitude of the IPP) of the GPS satellite is scaled to the range [0, $+90^o$] using the half angle θ according to the following expression

$$\phi' = \frac{\theta}{90^o}(\phi - \phi_o)$$

where
ϕ' is the scaled co-latitude,
ϕ_o is the centre co-latitude.

The model uses a sun-fixed longitude system which maps longitude in the range [0,360^o] over the whole day (24 hour period), thus taking into account time changes and regional time dependent longitude variation (Opperman *et al.*, 2007). It makes a provision of computing TEC for void locations (areas with no GPS receivers) using the surrounding nearby GPS receivers. In the ASHA model TEC estimation and modelling is done using the spherical harmonic expansion

$$TEC(\lambda, \phi) = \sum_{n=0}^{N} \sum_{m=0}^{n} P_{nm}\cos(\phi)\,[a_{nm}\sin(m\lambda) + b_{nm}\cos(m\lambda)] \tag{3.34}$$

where
λ is the IPP sun-fixed longitude,
ϕ is the IPP co-latitude,
P_{nm} are the normalised associated Legendre functions,
a_{nm} and b_{nm} are the desired spherical harmonic coefficients.

The ASHA model calculates slant TEC from pseudorange GPS measurements and uses it to eliminate the ambiguity offset in the calculated carrier phase TEC

by a process known as "levelling". This combination of measurements eliminates a large percentage of ionospheric error and gives TEC to a better accuracy.

3.5.1 The ASHA algorithm

The slant TEC (STEC) calculated from GPS observations is converted to vertical TEC (VTEC) by assuming the ionosphere to be a single layer. The mapping function used is given by

$$T_s = \frac{T_v}{cos(z')} \tag{3.35}$$

where
T_s and T_v are the slant and vertical TEC respectively,
z' is determined by using equation (3.33).

The general ionospheric delay is directly proportional to TEC and inversely proportional to the square of frequency as follows:

$$\Delta\tau = \left(\frac{40.28}{cf^2}\right) \times TEC = \alpha.T, \qquad \alpha = \frac{40.28}{cf^2}. \tag{3.36}$$

TEC in the above expression is STEC along the ray path of the signal through the ionospheric pierce point. The vertical ionospheric delay is given by

$$\Delta\tau = \left(\frac{40.28}{cf^2}\right)\left(\frac{T_v}{cos(z')}\right) + (b^R - b^S) = \alpha\beta.T_v + (b^R - b^S) \tag{3.37}$$

where
$\beta = 1/cos(z')$
b^R and b^S are the receiver and satellite interfrequency differential biases respectively.

Writing vertical TEC (VTEC) in terms of spherical harmonics,

$$T_v(\lambda, \phi) = \sum_{n=0}^{N} \sum_{m=0}^{n} P_{nm}\cos(\phi)\left[a_{nm}\sin(m\lambda) + b_{nm}\cos(m\lambda)\right].$$

VTEC can be expressed through its components as

$$T_v(\lambda,\phi) = \sum_{n=0}^{N}\sum_{m=0}^{n} P_{nm}\cos(\phi).a_{nm}\sin(m\lambda) + \sum_{n=0}^{N}\sum_{m=0}^{n} P_{nm}\cos(\phi).b_{nm}\cos(m\lambda). \tag{3.38}$$

Substituting equation (3.38) in equation (3.37), the ionospheric delay becomes

$$\Delta\tau = \alpha\beta\left(\sum_{n=0}^{N}\sum_{m=0}^{n} XY + \sum_{n=0}^{N}\sum_{m=0}^{n} XZ\right) + (b^R - b^S) \tag{3.39}$$

where

$$X = P_{nm}\cos(\phi), \quad Y = a_{nm}\sin(m\lambda) \ \textit{and} \ \ Z = b_{nm}\cos(m\lambda).$$

The Jacobian method is applied where partial derivatives of ionospheric delay with respect to the coefficients a_{nm} and b_{nm} and the biases b^R and b^S are taken thereby estimating them:

$$J(a_{nm},\ b_{nm},\ b^R,\ b^S) = \left(\frac{\partial\Delta\tau}{\partial a_{nm}} \quad \frac{\partial\Delta\tau}{\partial b_{nm}} \quad \frac{\partial\Delta\tau}{\partial b^R} \quad \frac{\partial\Delta\tau}{\partial b^S}\right). \tag{3.40}$$

The estimated coefficients are then used in the spherical harmonic expression to estimate the ionospheric delay. The full discussion about the calculation of receiver and satellite biases is beyond the scope of this study. A detailed treatment of the ASHA model can be found in Opperman (2007).

3.5.2 Calculation of TEC using the ASHA model

The ASHA model utilises data from a GPS network of geodetic grade dual frequency Ashtech GPS receivers operated by the Chief Directorate Surveys and Mapping (CDSM) based in Cape Town, South Africa. The above algorithm was implemented in MATLAB and used to calculate TEC values for three GPS receiver locations in South Africa. They are the Square Kilometre Array (SKA) Hub location (30.71^{o}S, 21.39^{o}E), Sutherland (32.38^{o}S, 20.81^{o}E) and Springbok (29.67^{o}S, 17.88^{o}E). TEC was computed at 60 s intervals over a period of five years

for both SKA Hub location and Sutherland, and three years for Springbok.

3.5.3 Final GPS TEC consideration

All GPS receiver stations used in this study fall within the mid-latitude region. For mid-latitude sites, nighttime vertical TEC values are of order 10 TECU with the corresponding day values of the order 100 TECU (Langley, 2000). This range changes significantly for high latitude and equatorial regions due to different ionospheric compositions and additional processes e.g. the equatorial anomaly may influence the rate of ionisation in equatorial regions. Comparison of GPS TEC and ITEC at Grahamstown (33.3^{o}S, 26.5^{o}E), South Africa, which is in a mid-latitude region, during a low solar activity period (2005, March to June inclusive) revealed that maximum GPS TEC values are less than 40 TECU (McKinnell *et al.*, 2007). TEC varies greatly during equinoxes (Kouris *et al.*, 1999). Diurnal variation of TEC during equinoxes and solstices during 2000 at Sutherland (32.38^{o}S, 20.81^{o}E) (chapter 5, Figure 5.5) confirm this. Although 2000 was a year of high solar activity, maximum TEC values were about 100 TECU. For these reasons maximum TEC values used in this project do not exceed 100 TECU.

Studies of TEC variations during the disturbed days (due to the magnetic storm of 22 September 1999) over mid-latitude sites such as Hobart (42.8^{o}S, 147.4^{o}E) and Macquarie (54.5^{o}S, 158.9^{o}E) in Australia when TEC values were depleted (Yizengaw and Essex, 2002), show that on average, TEC values at 10h00 UT are at least greater than 10 TECU. However, a comprehensive investigation about TEC changes during geomagnetic disturbed conditions has not been done during this study and therefore the effects of magnetic storms on TEC are beyond the scope of this thesis. The validation of the GPS TEC values done in chapter 5 (Figure 5.13) using ITEC during a low solar activity year (2004) indicates that ITEC values at 10h00 UT are also greater than 10 TECU. Thus, for the initial input space of this project, TEC values in the range 10$\leq$ TEC $\leq$100 TECU were considered for the local midday period, where 1 TECU=1×10^{16} electrons.m^{-2}.

For the extraction of the midday (10h00 UT) TEC values from the data set produced using the ASHA algorithm implemented in MATLAB, all values less than 10 TECU and greater than 100 TECU were ignored. When considering the hourly

TEC values, TEC values less or equal to zero were neglected and the highest value of TEC considered was 100 TECU. Other outliers were removed manually. It has been observed that maximum TEC values occur between local midday and about 14h30 local time. The highest values do not normally exceed 100 TECU for the data used in this study. Details of diurnal TEC variations are found in chapter 5.

3.6 Summary

The ASHA model used for the estimation of TEC from GPS measurements has been briefly discussed. Details on the ASHA model can be found in Opperman (2007). The derivation of ionospheric delay which contains TEC information has been fully covered. A full discussion and derivation of both ionospheric phase and group refractive indices with relation to attenuation of electromagnetic waves as they propagate through the ionosphere has been presented. Finally, this chapter dealt with the rationale of choosing limits for the TEC values in the mid-latitude regions. The preliminary low mid-day minimum value was set to be 10 TECU while the maximum is 100 TECU. This limit was based on comparison of mid-day values for disturbed and quiet days. However, the confirmation of these values requires more investigation.

Chapter 4

Determination of input space

4.1 Introduction

In this chapter all input parameters used in training and testing the Neural Networks (NNs) are discussed. Their effects on TEC distribution are explained with some illustrations. Data which was used in South Africa's bid to host the Square Kilometre Array (SKA) radio telescope was used. Thus, the GPS receiver station in this chapter will be referred to as the SKA Hub and is located at latitude 30.71^{o}S and longitude 21.39^{o}E. An attempt to study the effects of solar wind on TEC predictions is discussed.

4.2 NN software

The software used in the development of this model is the Stuttgart Neural Network Simulator (SNNS) which was developed by the Institute for Parallel and Distributed High Performance Systems (IPVR), University of Tübingen and the Wilhelm-Schickard-Institute for Computer Science in Germany. SNNS is a simulator for NNs which creates an efficient and flexible simulation environment for research involving the application of NNs (Zell *et al.*, 1998). The software distributed by the University of Tübingen is freely available on the internet (**http://www-ra.informatik.uni-tuebingen.de/SNNS/**).

4.3 Parameter identification

As described in chapter 2, the critical frequency of the F2 layer (foF2) has been modelled using NNs and it was found that local time, solar activity, magnetic activity and the geographical position of the point of observation are some of the main factors that influence foF2 prediction (Williscroft and Poole, 1996; Poole and McKinnell, 2000; Sarma and Mahdu, 2005). Maximum ionisation in the ionosphere takes place within the F2 layer. During the process of ionisation, electrons and protons are produced, their number depending on the variations and dynamic processes taking place within the Earth's upper atmosphere. The quantity foF2 can be used in TEC predictions (Sarma and Mahdu, 2005). It has been shown that there is a strong correlation between foF2 and TEC which can be expressed in terms of slab thickness (breadth of the ionosphere) and electron density (Kouris *et al.*, 2004), as follows:

$$\tau = \frac{TEC}{N_{max}} \qquad \text{and} \qquad TEC = 1.24 \times 10^{-6} \tau (foF2)^2 \tag{4.1}$$

where
τ is the slab thickness in m,
TEC is in TECU,
$foF2$ is in MHz,
N_{max} is the ionospheric maximum electron density in m^{-3} which is related to foF2 by equation (2.3).

Other parameters known to influence GPS TEC are solar activity, geographical position of the receiver, line of sight (the elevation angle of the GPS satellite), diurnal and seasonal variations and magnetic activity (Hofmann-Wellenhof *et al.*, 1992; Reddy, 2002). Ionosonde TEC (ITEC) and GPS TEC are influenced by the same parameters with exception of line of sight for the former. The difference between GPS TEC and ITEC is the altitude at which TEC values are estimated. GPS TEC values are expected to be higher since the integration of the number of electrons is done for an altitude of $\sim$ 20 000 km from the Earth, while ITEC is only determined up to a maximum altitude of about 1000 km using measured bottomside and modelled topside ionospheric measurements. Illustrations showing the differences between GPS TEC and ITEC are shown in chapter 5.

TEC is globally determined by EUV radiation, although some electrons come from the magnetosphere and enter the ionosphere in the auroral regions (Langley, 2000). Diurnal variation is the most dominant due to the incident solar radiation. During the night TEC decreases because of the absence of solar radiation and the recombination of free electrons and ions. The major contributor of TEC during nighttime hours is the uppermost region of the ionosphere which is responsible for about 50% of the overall night TEC. TEC also varies seasonally which is effectively represented by day number (DN). Other inputs are solar activity and magnetic activity. In ionospheric studies the indices used to represent solar activity are sunspot number (SSN) and solar radio flux at 10.7 cm (F10.7) wavelength. For regions which exhibit strong solar variations below the peak of the F layer, F10.7 is a better index to use in ionospheric modelling, while SSN has been found to be a better index above the F layer where there is a strong link between the rate of change of various dynamic processes and the influences from the solar system (Bilitza, 2001). Thus, in this study the SSN was adopted as a representation of solar activity since GPS TEC is obtained by integration of the number of free electrons along a signal's path from a GPS satellite (located at about 20 000 km which is far above the peak of the F layer) to the receiver on the ground.

4.4 Input space

The input space was developed from the parameters that are known to affect GPS TEC such as seasonal variation, solar activity, magnetic activity, diurnal variation and geographic position of the GPS receiver. The consideration of a particular single station at a particular time (10h00 UT) meant that only the first three parameters were included in the input space. The seasonal and diurnal variations are represented by the day number (DN) and hour (HR), and were each split into two components to allow data continuity (Poole and McKinnell, 2000) as follows:

$$\mathrm{DNS} = \sin\left(\frac{2\pi \times DN}{365.25}\right) \quad \mathrm{DNC} = \cos\left(\frac{2\pi \times DN}{365.25}\right) \tag{4.2}$$

$$\mathrm{HRS} = \sin\left(\frac{2\pi \times HR}{24}\right) \quad \mathrm{HRC} = \cos\left(\frac{2\pi \times HR}{24}\right) \tag{4.3}$$

where
DNS, DNC, HRS and HRC are the sine and cosine components of DN and HR.

In the above equations normalisation of the DN to 365.25 takes into account the fact that part of the data used was for leap years. There is a requirement to determine the appropriate representation of solar and magnetic activities. The magnetic index was derived from the K index recorded by the Hermanus Magnetic Observatory (HMO) in South Africa. The magnetic activity parameter (denoted by K) is the measure of the geomagnetic field variations over Hermanus, South Africa as a result of currents flowing in the ionosphere and the magnetosphere. The solar activity is represented by the sunspot number (SSN) for which the historic and present data is readily available. Due to tremendous variations in the day-to-day SSN, monthly running averages are computed based on individual sunspot numbers and sunspot groups to account for long term trends (Langley, 2000). The calculation for daily SSN is done according to the Wolf sunspot number R as

$$R = k(10g + s) \tag{4.4}$$

where
k is the variable scaling factor which takes into account the hardware and software biases for telescopes, observing conditions and technical biases caused by the observers,
g is the number of sunspot groups,
s is the total number of individual spots visible on the sun in all respective groups.

The observatory factor k is unity for Zürich Observatory and varies ($k \leq 1$) for other observatories to obtain roughly the same R.

4.4.1 Determination of optimum parameters

A number of NNs were trained in order to find the suitable parameters to represent sunspot number (SSN) and magnetic activity. For this purpose, the SKA Hub (30.71^oS, 21.39^oE) data for a period of five years (2000-2004) was used and the TEC values extracted at 10h00 UT (midday South African Standard Time, SAST=UT+2). The training set consisted of 1503 data points each with four pa-

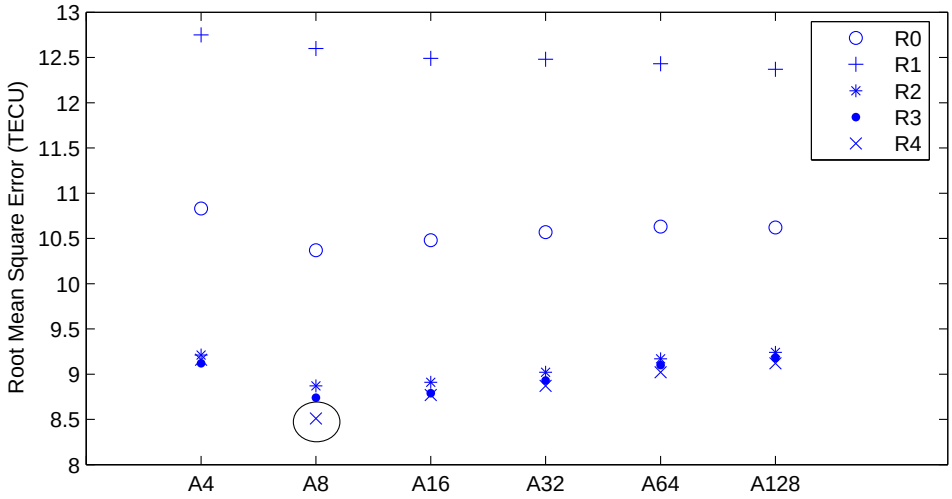
rameters, DNS, DNC, R (solar activity) and A (magnetic activity). Training was done using data for R0, R1, R2, ... , R12 with each of A4, A8, A16, ... , A128. R0 is the daily sunspot number and R1 is the 1-month running mean of the daily sunspot number. R2, R3, ..., R12 take on similar appropriate definitions. The A-index is obtained from the 3-hourly values of the K-index. Since for each value of K-index recorded, there is a corresponding a_k value, A4 is defined as the value corresponding to the mean (average) of the previous 4 a_k indices. A8, ..., A128 are defined in the same manner. Data used was divided into training and testing/validating pattern sets. Mean square errors for training and testing patterns were monitored, and training allowed to proceed as long as the errors decreased. When the errors started increasing, training was terminated as it is believed that the network is no longer generalising due to the memorisation of the training pattern (Fausett, 1994).

The method used for determining the optimum NN in this thesis is the Root Mean Square Error (RMSE) which is defined as

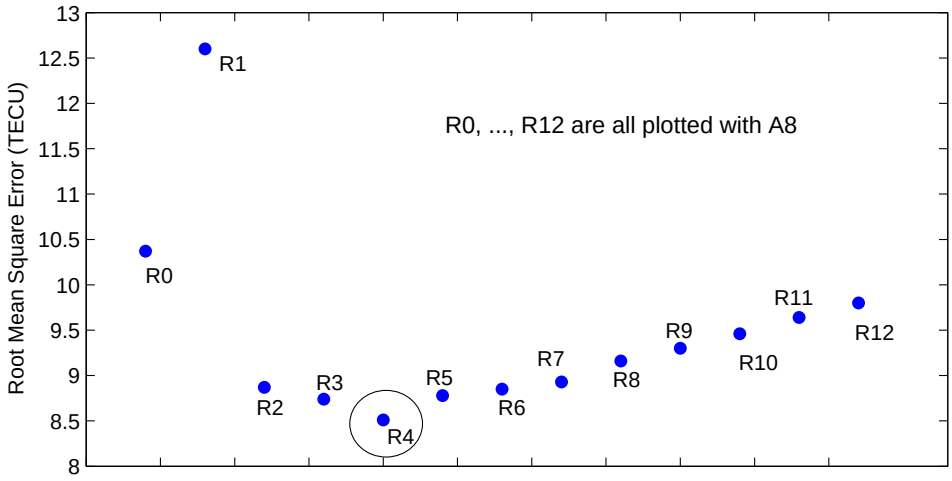
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_{pred} - T_{meas})^2} \tag{4.5}$$

where
N is the number of data points,
T_{pred} is TEC predicted by a model,
T_{meas} is TEC estimated from GPS observations using the ASHA model.

A combination of input parameters which gave the least RMSE was considered to be the best for TEC predictions. The RMSE method has been successively used as a means to determine optimum parameters during ionospheric foF2 predictions (Williscroft and Poole, 1996; McKinnell, 2002) and the criterion is that minimum RMSE gives accurate predictions. Figure 4.1 illustrates graphically the RMSE values obtained from each of the NNs trained. It is clear from this figure that the NN with R4 and A8 provide the optimum solution.



(a) Computed Root Mean Square Errors (RMSEs) between GPS TEC and NN TEC (TECU) values during the determination of the appropriate representation for the magnetic activity. The previous eight 3-hourly magnetic A index values (A8) gave the least RMSE and was hence taken as the optimum.



(b) A plot of RMSEs between GPS TEC and NN TEC (TECU) values for R0, R1,..., R12 with the running mean of the previous eight 3-hourly magnetic A index values (A8). R0 is the daily SSN, R1 is 1-month running mean of daily SSN and R2, R3, ..., R12 are defined in a similar way.

Figure 4.1: RMSE values for the NNs trained with different combinations of input parameters. Circled points represent the combination R4A8 that gave the optimum solution.

4.5 NN architecture

As described in chapter 2, NNs have different sets of architecture and design. This mainly depends on the complexity of the problem at hand, the field of study and the nature of the data. The architecture that gave the least RMSE during the determination of optimum NN using the SKA Hub (30.71^{o}S, 21.39^{o}E) data, was single layered with 7 hidden nodes in the hidden layer as shown in Figure 4.2. The

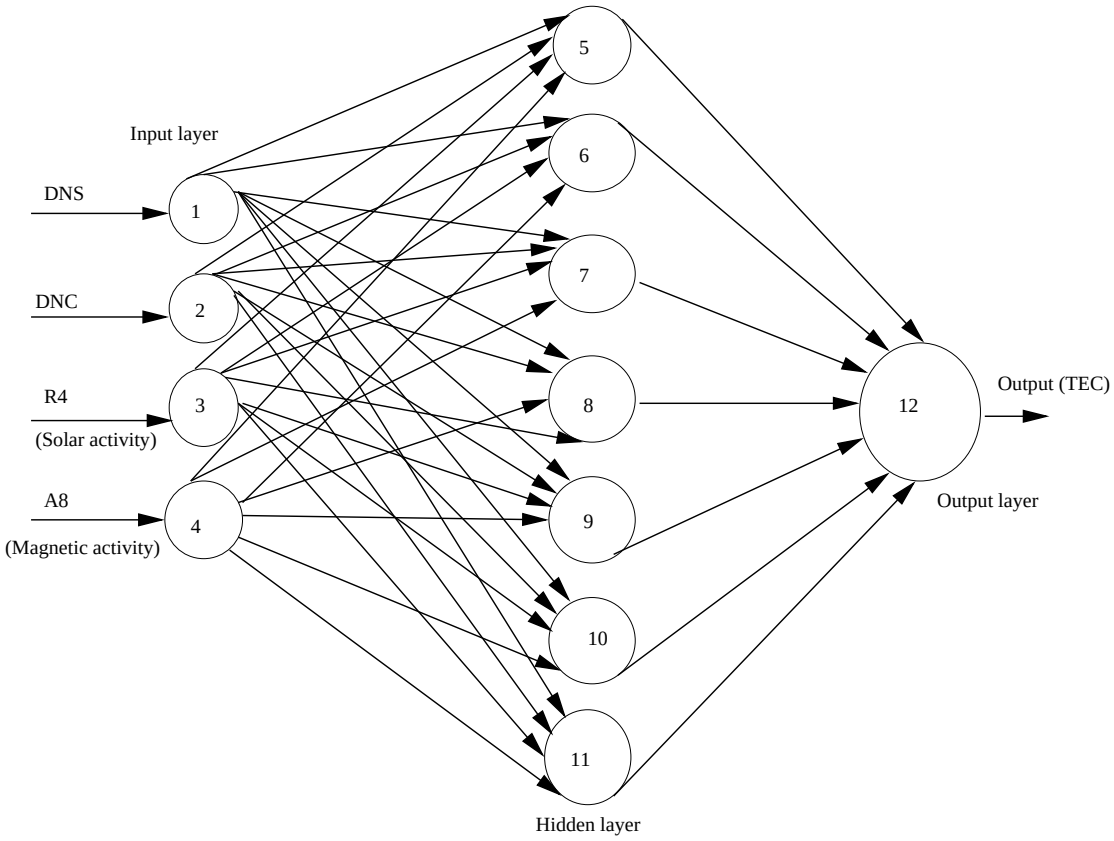


Figure 4.2: NN architecture that gave the optimum solution (least RMSE).

number of nodes in the input layer increases with the number of inputs. For example, the inclusion of the solar wind parameter (explained later in this chapter) increased the input layer nodes from four to five.

The procedure for determining the optimum architecture was the addition of one hidden node at a time, training the NN, testing it with data and finally computing the RMSE between the GPS TEC and the NN predicted TEC values.

4.6 Data analysis

An analysis was done by computing both the absolute and relative errors. Further analysis was also done by graphically comparing the predicted TEC values from the NN and IRI-2001 models with GPS TEC values for the SKA Hub.

4.6.1 Absolute error

In this context, absolute error is defined as the magnitude of the difference between the TEC as predicted by the NN model and the TEC computed from the GPS measurements using the dual-frequency algorithm (Leandro and Santos, 2007). It is expressed as

$$|E_{abs}| = |T_{pred} - T_{meas}| \tag{4.6}$$

where
E_{abs} is the absolute error,
T_{pred} and T_{meas} are TEC values predicted by the NN model and calculated from the GPS measurements respectively.

4.6.2 Relative error

The relative error is a function of the normalized differences (DTEC) between the NN predicted and the GPS TEC values. DTEC is calculated using the following equation:

$$T_{diff} = \frac{|T_{pred} - T_{meas}|}{T_{meas}} \tag{4.7}$$

where T_{diff} is the DTEC.

Relative error expressed as a percentage is written as

$$|E_{rel}| = \left(\frac{|E_{abs}|}{T_{meas}}\right) \times 100\% \tag{4.8}$$

where E_{rel} is the relative error.

The absolute and relative errors provide an indication of the accuracy with which the models predict TEC, with the aim of minimising these errors. If these abso-

lute errors are minimal, then the TEC values (predicted and observed) are close to each other and the model is effective (Leandro and Santos, 2007). Both the predicted and GPS observed TEC values are expressed in TEC units (TECU). The relative correction values which give an idea about the efficiency of the model are calculated as follows:

$$|Rel_{corr}| = (100 - |E_{rel}|)\% \tag{4.9}$$

where Rel_{corr} is the relative correction.

Figure 4.3 shows a plot of average absolute and relative errors calculated us-

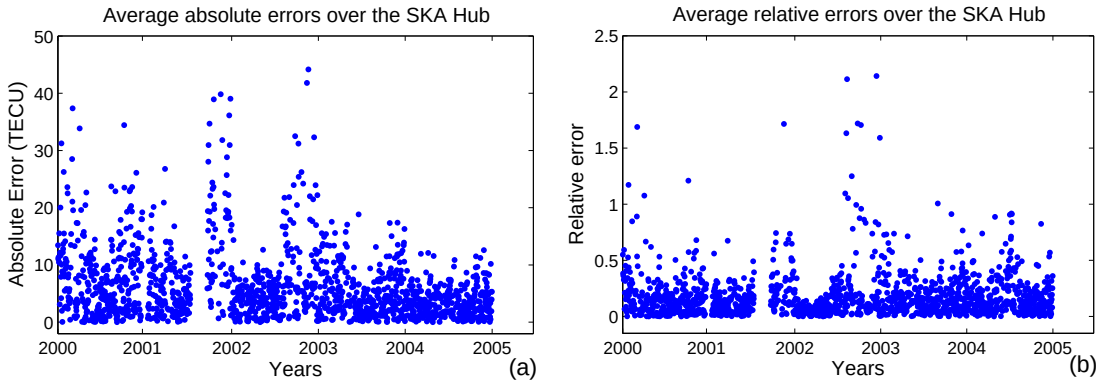


Figure 4.3: Average absolute and relative error variations for five years over the SKA Hub (30.71^{o}S, 21.39^{o}E).

ing equations (4.6) and (4.7) for GPS TEC and NN predicted TEC values. From Figure 4.3 (a), it can be seen that average absolute errors are higher for periods of high solar activity (years 2000 and 2001) and less for low solar activity periods (especially in 2004). The main reason for this is the higher TEC values and higher variability within the TEC parameters due to high solar activity (Leandro and Santos, 2007). On the other hand, average relative errors are high during periods of low solar activity compared to high solar activity periods as shown in Figure 4.3 (b), because of lower TEC values for the former. The low average absolute errors (as defined in equation (4.6)) between the predicted and measured TEC values implies that the developed NN model performs better during low solar activity periods as compared to high solar activity periods (Leandro and Santos, 2007). Figure 4.4 (a) clearly shows the solar cycle variations from which the years of high

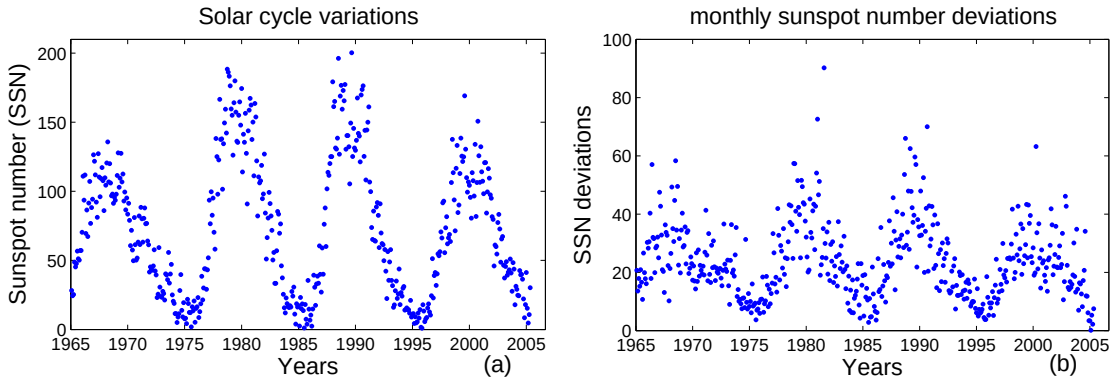


Figure 4.4: Monthly averages of (a) sunspot numbers and (b) their standard deviations from January 1965 to April 2006. Data obtained from http://solarscience.msfc.nasa.gov/.

and low solar activity periods can be identified.

4.6.3 Comparison of NN and IRI models over the SKA Hub (30.71^o**S,** 21.39^o**E**)

The International Reference Ionosphere (IRI) is the international standard for the specification of ionospheric densities and temperatures (Bilitza, 2001). It provides the monthly average values of electron density, electron content, electron and ion temperature, ion composition as a function of height, location, local time and sunspot number for magnetically quiet conditions (Bilitza, 2001; McKinnell, 2002; Bhuyan and Borah, 2007). It is a standard empirical model that was developed based on all the available data sources Worldwide (ionosondes, incoherent scatter radars, in-situ measurements on satellites and rockets, etc) and is updated annually. Although the IRI utilises data Worldwide and remains one of the best global models, Regional TEC models may perform better in some regions, especially in the Southern Hemisphere. TEC in the Southern Hemisphere has not been modelled in detail compared to the work done in the Northern Hemisphere, partly due to an historic paucity of data available from the Southern Hemisphere (McKinnell, 2002).

The IRI-2001 model was used to get the required SKA Hub TEC data for the days with noon GPS TEC data over a period of five years (2000-2004). TEC from

the IRI is obtained by numerical integration in 1 km steps from 50 km to 2000 km. All the days with missing GPS TEC data were carefully identified in IRI data and removed for effective comparisons. The IRI model uses a 12-month running mean of the sunspot number (Rz_{12}) observed at the Zürich Observatory, in its estimations (Bilitza, 2001).

Using the input space that provided the optimum parameters (DNS, DNC, R4, A8), local midday TEC values were predicted for the SKA Hub (30.71^{o}S, 21.39^{o}E) and compared with the measured data and equivalent IRI predictions in Figure 4.5 (a).

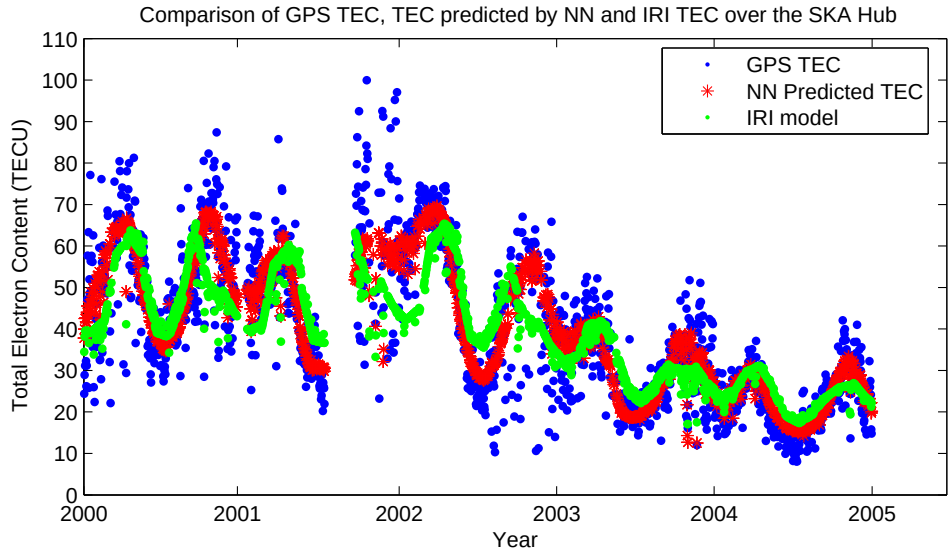
Figure 4.5 (b) shows differences in TEC (ΔTEC) between GPS TEC, IRI TEC and NN TEC over a five year period at 10h00 UT. The average Δ TEC is defined as

$$\Delta T_{av} = \frac{1}{N}\sum_{i=1}^{N}(T_G - T_{j=1,2}) \tag{4.10}$$

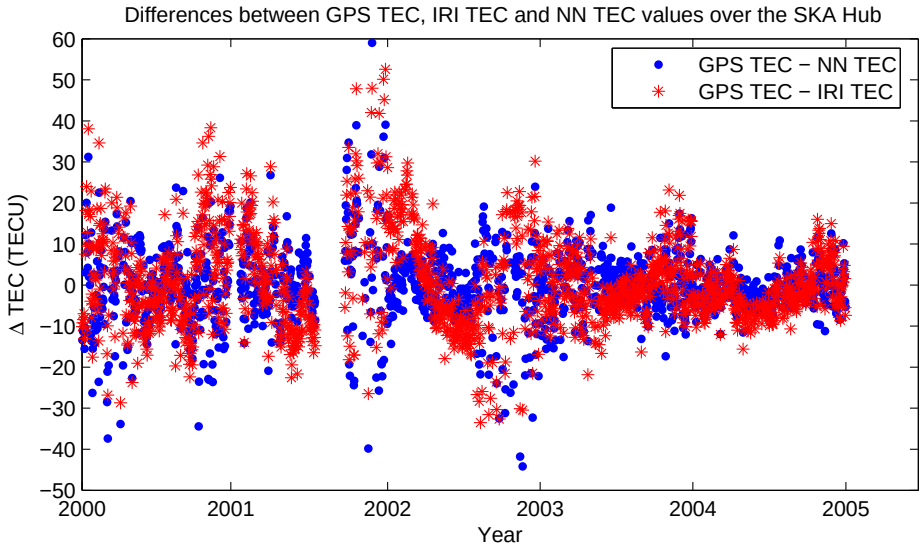
where
ΔT_{av} is the average difference between the GPS TEC (T_G) and NN TEC ($T_{j=1}$) or IRI TEC ($T_{j=2}$),
$i = 1, 2, ..., N$,
N=1503 is the total number of data points.

The average ΔTEC as defined in equation (4.10) is $\sim$ -0.32 TECU and 0.69 TECU for NN TEC and IRI TEC respectively with the most ΔTEC values varying between -30 TECU and 30 TECU. The least average bias value (ΔT_{av}) between the GPS TEC and NN TEC provides an idea about the NN model accuracy of predicting GPS TEC compared to the IRI model.

Table 4.1 shows the computed RMSEs between the GPS TEC and TEC values as predicted by both NN and IRI models for each year. From this Table, the average RMSEs are $\simeq$ 10.84 TECU and 8.51 TECU for the IRI model and NN respectively. This gives an average NN improvement of about 27% compared to the IRI model over a period of five years. An average relative error of about 18% means that the NN can predict at least 82% of the GPS TEC on average. Figure 4.5 shows that



(a) A plot for the variation of GPS TEC, IRI and NN predicted TEC (TECU) values.



(b) A plot for the differences between GPS TEC, IRI and NN predicted TEC (TECU) values.

Figure 4.5: Comparisons between GPS TEC, IRI and predicted NN TEC (TECU) values, and their differences for the available data (2000-2004) for the SKA Hub (30.71^{o}S, 21.39^{o}E) at 10h00 UT.

Table 4.1: Comparison between IRI and NN model errors for the years 2000-2004 with absolute and relative errors as defined in equations (4.6) and (4.7), computed per year

Year	IRI error (TECU)	NN error (TECU)	Absolute error (TECU)	Relative error (NN)	Relative correction(%)
2000	12.30	10.28	7.83	0.174	82.6
2001	15.59	12.30	6.31	0.174	82.6
2002	13.29	9.09	6.33	0.209	79.1
2003	7.27	6.46	4.99	0.173	82.7
2004	5.73	4.40	3.55	0.187	81.3

there were large variations of TEC during the years 2000 and 2001. This is because these years were solar maximum years, 2001 being the last solar maximum year in cycle 23 (see Figure 4.4(a)). There is an increase in the number of electrons as a result of increased ionisation rate due to the highly active sun during the high solar activity periods.

The number of days with missing GPS TEC noon values were 38, 131, 70, 36 and 49 days for years 2000, 2001, 2002, 2003 and 2004 respectively. Percentage-wise this amounts to:

$$\text{MD} = \frac{324}{1827} \times 100 = 17.73\%$$

MD represents missing data.

Thus, 2001 had a significant number of missing data that could have contributed to the high RMSE compared to the rest of the years. The smaller TEC variations are observed in 2004, which was close to solar minimum.

4.7 Attempt to include solar wind in the input space

Solar wind is a stream of energetic particles (in the form of a plasma) that is continuously ejected from the sun and reaches the Earth's atmosphere at a speed typically in the range 300 km/s to 700 km/s. Its composition is mainly electrons

and protons (about 95%). TEC is expected to have some degree of variation with the solar wind due to the fact that during solar minimum periods, solar radiation variations are minimal, but there is continuous solar wind flowing from the sun (Jakowski *et al.*, 1999). In this study attempts to include the solar wind were considered by separately including solar wind bulk velocity, V_{sw} (km/s) and proton number density, P_{nd} (per cm^3) in the input space for the NN. This expanded the database to include another input parameter. The input parameters used in the NN were DNS, DNC, R4, A8 and V_{sw} or P_{nd}.

4.7.1 Solar wind bulk velocity and proton number density

Solar wind data (both for V_{sw} and P_{nd}) at 10h00 UT for years 2000-2004 was downloaded from the Coordinated Data Analysis Web
(http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/).

This data was obtained from Advanced Composition Explorer (ACE) satellite. The ACE satellite is located at the L1 point (a point where the gravitational pulls on the satellite from the Sun and the Earth cancel out), which is at a distance of more than $200R_E$ (R_E is the Earth's radius in km) from the Earth's surface, making it relatively more steady with respect to both the Sun and the Earth.

The solar wind bulk velocity is computed by combining the three components of velocity in x, y and z directions using the following equation:

$$V_{sw} = \sqrt{V_x^2 + V_y^2 + V_z^2} \tag{4.11}$$

where V_{sw} is the solar wind bulk velocity.

Since the number of electrons in the solar wind is roughly equal to the number of protons, the proton number density data was used instead of electron number density (which was not available during this study). There existed days without data for solar wind velocity and proton number density, but with calculated GPS TEC data. This discontinuity in the ACE data was catered for, by replacing these days with the monthly median values of days for which data was available according to

the following equations:

$$SW_m = \frac{1}{N_d} \sum_{i=1}^{N_d} X(SW) \tag{4.12}$$

$$PD_m = \frac{1}{N_d} \sum_{i=1}^{N_d} X(PD) \tag{4.13}$$

where
SW_m and PD_m represent missing solar wind bulk velocity and proton number density values,
i=1, 2,..., N_d, with N_d being the total number of days for a given month,
$X(SW)$ and $X(PD)$ are the daily solar wind bulk velocity and proton number density values at 10h00 UT for the remainder of the days of the month.

4.7.2 Results and discussion

The NN converged after completion of epochs using 3000 and 2700 iterations when the solar wind bulk velocity and proton number density were included respectively. An epoch is defined as the presentation of all 1503 training patterns comprising of all input parameters to the NN.

Table 4.2 shows the RMSEs obtained by including V_{sw} or P_{nd} separately in the input space. It is observed that increasing the number of hidden nodes tends to reduce the RMSE (in some cases) between the derived GPS TEC and the NN predicted TEC values. However, taking into account that the NN may start memorising the trend of the data and also the complexity of the network (Sarma and Mahdu, 2005), this fact may be ignored.

This difference between the number of iterations when solar wind bulk velocity and proton number density are included separately as inputs to the NN, yet with the same data points, may probably be attributed to the magnitudes of their respective values. Figure 4.6 shows both solar wind bulk velocity (V_{sw}) and proton number density (P_{nd}) values plotted for the available SKA Hub (30.71^{o}S, 21.39^{o}E) data, and it is clear from this figure that $V_{sw} > P_{nd}$ in magnitude.

Table 4.2: RMSEs for solar wind bulk velocity and proton number density with varying hidden nodes.

Hidden nodes	RMSE (TECU) for V_{sw}	RMSE (TECU) for P_{nd}
7	8.57	8.77
8	8.61	8.83
9	8.57	8.70
10	8.53	8.61
11	8.58	8.67
12	8.57	8.74
13	8.51	8.61

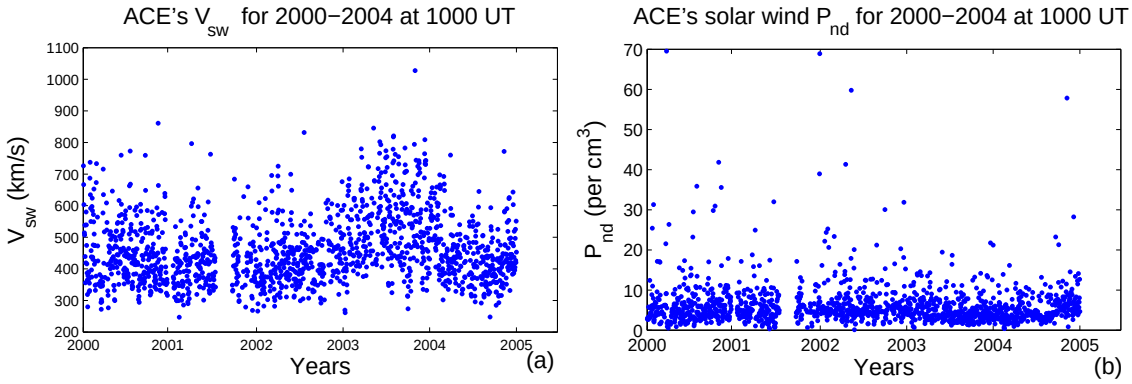


Figure 4.6: Variation of ACE's (a) solar wind bulk velocity and (b) proton number density for the available SKA Hub (30.71^{o}S, 21.39^{o}E) data at 10h00 UT.

The results obtained show no improvement achieved due to the inclusion of solar wind as an input parameter. The previous optimum NN is realised in the case of considering solar wind bulk velocity and when the hidden nodes are 13 where the RMSE is 8.51 TECU as obtained before, during the determination of optimum parameters. However, the effects of interplanetary magnetic field (IMF) have not been investigated due to the time resolution of the GPS data used in this study. This is worth persuing in future since the southward turning of the z-component of IMF, Bz is one of the indicators of magnetic storms which may inturn lead to an enhancement or depletion of TEC (Yizengaw *et al.*, 2005). Before concluding that solar wind is one of the parameters that influence TEC variations, it should be subjected to further study using other sources of data. Thus, for the remaining part of this thesis, solar wind data was not used in NN training, validating and

testing processes.

4.8 Summary

In this chapter the optimum parameters for TEC prediction have been determined as 4-months running mean of daily sunspot number (R4) and the running mean of the previous eight 3-hourly magnetic A index values (A8). The optimum input space also includes seasonal variations. Solar wind may have an effect on TEC, but a more comprehensive study needs to be done using data, preferably from low earth orbit (LEO) or medium earth orbit (MEO) satellites, to make a general conclusion. The rationale of suggesting LEO or MEO satellite data is based on the fact that their altitudes are low compared to ACE satellite. In addition, most of the ionospheric processes that influence TEC variations such as ionization and recombination of ions occur at quite low altitudes (below 2000 km). This is supported by the findings that the contribution of the plasmasphere to TEC is lower compared to the one for bottomside and topside ionosphere (Bilitza *et al.*, 2006; Mosert *et al.*, 2007). More physical changes, such as solar wind dynamic pressure, ion flux, IMF Bz (or related indices) and electron number density with regard to TEC predictions should also be investigated. The input space used in determining the feasibility of predicting TEC using NNs, throughout the remainder of this thesis, consists of DN, HR, R4 and A8.

Chapter 5

Single station models

5.1 Introduction

In this chapter the procedure for developing NN models for TEC at selected GPS receiver stations is presented. Two South African GPS stations were considered: Sutherland (32.38^{o}S, 20.81^{o}E), which is an International GPS Service (IGS) station managed by the Hartbeesthoek Radio Astronomy Observatory (HartRAO) and Springbok (29.67^{o}S, 17.88^{o}E), a station in the network of GPS receiver stations of the Chief Directorate Surveys and Mappings (CDSM). A NN model was constructed for Sutherland at 10h00 UT which was then extended to an hourly model and separately, an hourly model for Springbok was also constructed. Sutherland was chosen due to data availability, while the choice of Springbok was due to its close proximity to the South African ionosonde located at Louisvale (28.5^{o}S, 21.2^{o}E). Results from Springbok could then be validated with Louisvale ionosonde measurements.

5.2 Sutherland (32.38^{o}S, 20.81^{o}E)

5.2.1 10h00 UT model

Having determined the optimum parameters as R4 and A8, a similar procedure (see chapter 4), using the same number of inputs and architectural format, was used to develop a NN model to predict TEC for Sutherland, South Africa GPS receiver station at 10h00 UT (local midday, SAST). Data for a 5 year period

(2000-2004) was available for this station, with four years (2000, 2001, 2003, 2004) making up the training set and the year 2002 used for testing . There were 939 data points in the training set, with the inputs for each data point comprising DNS, DNC, R4 and A8, as defined in chapter 4. A number of NNs were trained with different architectures. The root mean square errors (RMSEs), as defined in equation (4.5), were computed using the testing data from year 2002 to determine the optimum NN. The architecture which gave the least RMSE had 4 input nodes, 7 hidden nodes in the hidden layer and one output node. The GPS TEC values as computed by the ASHA algorithm were compared with the corresponding TEC values of the NN and IRI and are shown in Figure 5.1. It was observed that

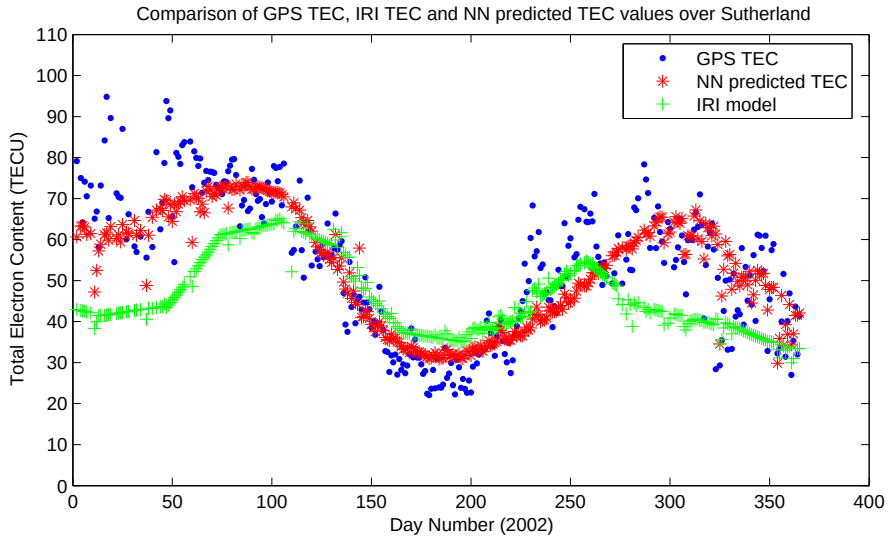


Figure 5.1: GPS TEC, NN predicted TEC and IRI TEC for year 2002 for Sutherland (32.38^{o}S, 20.81^{o}E) at 10h00 UT.

for 2002 (the testing set) the NN model performed more accurately than the IRI model throughout the whole year. The RMSEs between the GPS TEC and TEC as predicted by the NN and IRI models for the year 2002 were 8.81 TECU and 16.06 TECU respectively.

Figure 5.1 shows TEC variations over Sutherland in 2002 at 10h00 UT. It should, however be noted that the year 2002, while reserved for testing, fell in the middle of the training set period. It was therefore expected that the NN model would perform well when predicting the 2002 TEC values, since NNs are known to inter-

polate well within the input space.

The optimum NN was also tested with data that had been used in training (year 2000) and a similar comparison to that in Figure 5.1 was done and is shown in Figure 5.2. The RMSE values between measured and predicted TEC values for

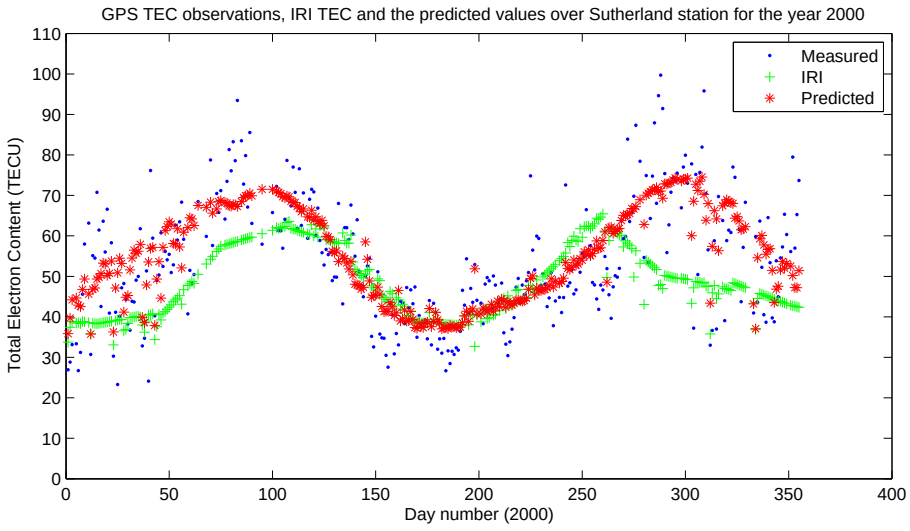


Figure 5.2: Perfomance of the NN Model during the data period used in training (year 2000).

2000 were 10.24 TECU and 14.59 TECU for the NN and IRI models respectively. The RMSE obtained when year 2000 was used for testing is higher than that for 2002, but still more accurate than the corresponding value for the IRI model. High RMSE for 2000 is associated with large TEC variations as a result of higher solar activity in 2000 than in 2002 as shown in Figure 4.4.

5.2.2 Hourly model

An hourly model for GPS TEC over Sutherland (32.38^{o}S, 20.81^{o}E) was constructed by increasing the input space to include the hour (HR) parameter and by expanding the database to include all hours. This time, data for 2000 was used for testing and the rest of the data was used for the training and validating processes. The total number of inputs were 6 (DNS, DNC, HRS, HRC, R4 and A8) and the architecture of the network was single-layered with nine hidden nodes in the hidden layer as shown in Figure 5.4. Data presented to the network consisted of 23 095

data points covering all inputs between 2001 and 2004. Figure 5.3 is a histogram showing the data points that were used in NN training and validating processes.

An analysis of the model's perfomance during the four different seasons (Au-

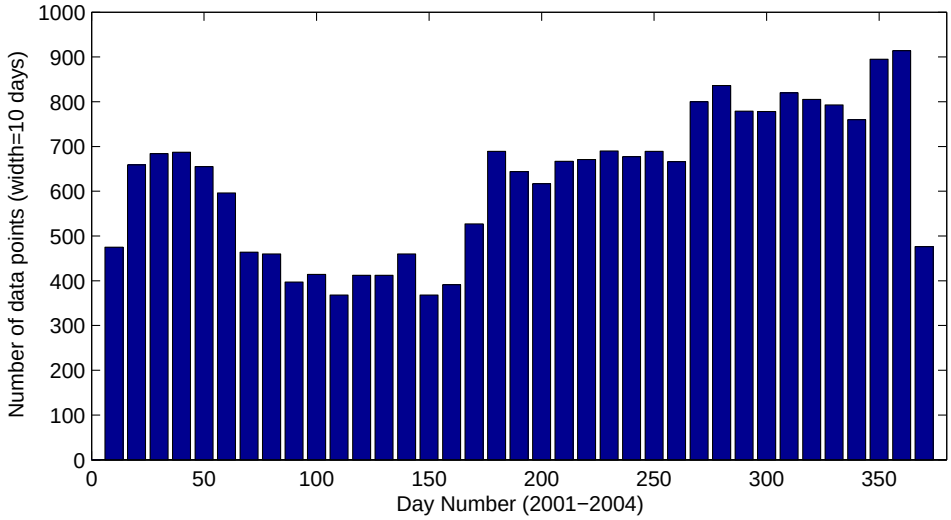


Figure 5.3: A histogram showing the number of data points used in NN training for the hourly model at Sutherland (32.38^{o}S, 20.81^{o}E).

tumn, Winter, Spring and Summer) was undertaken by predicting TEC values for days of equinoxes and solstices during the test year (2000). The results are plotted in Figure 5.5. Due to unavailability of data for December, testing for summer solstice was done on 8^{th} January, 2000.

Figure 5.5 (a)-(d) shows the comparison between the predicted and measured TEC values. The RMSEs as illustrated graphically in Figure 5.6 are 12.01 TECU, 15.40 TECU, 1.68 TECU and 2.26 TECU for Figure 5.5 (a), (b), (c) and (d) respectively.

The hourly model was used to test the NN prediction level for local midday TEC values and the results are plotted in Figure 5.8. In all cases (equinoxes and solstices) a gradual morning rise and evening decrease of TEC was observed. Maximum TEC occurs between 10h00 UT and 12h00 UT. Figure 5.5 (b) shows that the NN overpredicts during the Spring. This was observed for almost all of

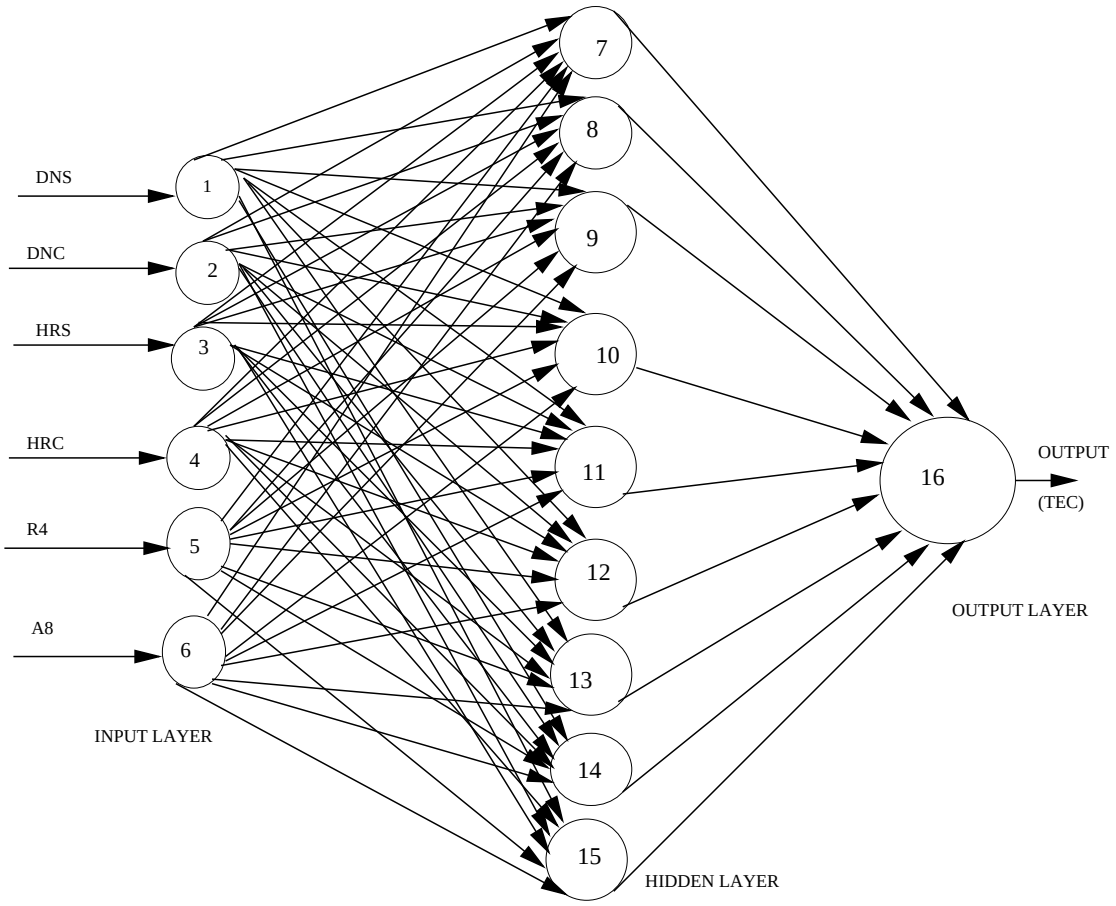


Figure 5.4: Optimum NN architecture when input parameters increased to 6.

September. This is attributed to insufficient data used in training at the high solar activity periods.

Figure 5.9 illustrates the data available during the years 2000 and 2001, by plotting the solar activity variable R4 for these years at times when complete data sets of TEC and all input parameters were available. Note the large data gap in 2001 (Figure 5.9-(b)). Both 2000 and 2001 were periods of high solar activity, therefore for the spring prediction in 2000, all parameters were not effectively represented. Thus, the NN experienced a difficulty in learning the trend of the data and making an informed generalisation. Poor prediction in September may also be due to the fact that TEC varies strongly during the equinoxes (Kouris *et al.*, 1999).

When comparing results obtained from the NN hourly model (Figure 5.8) and

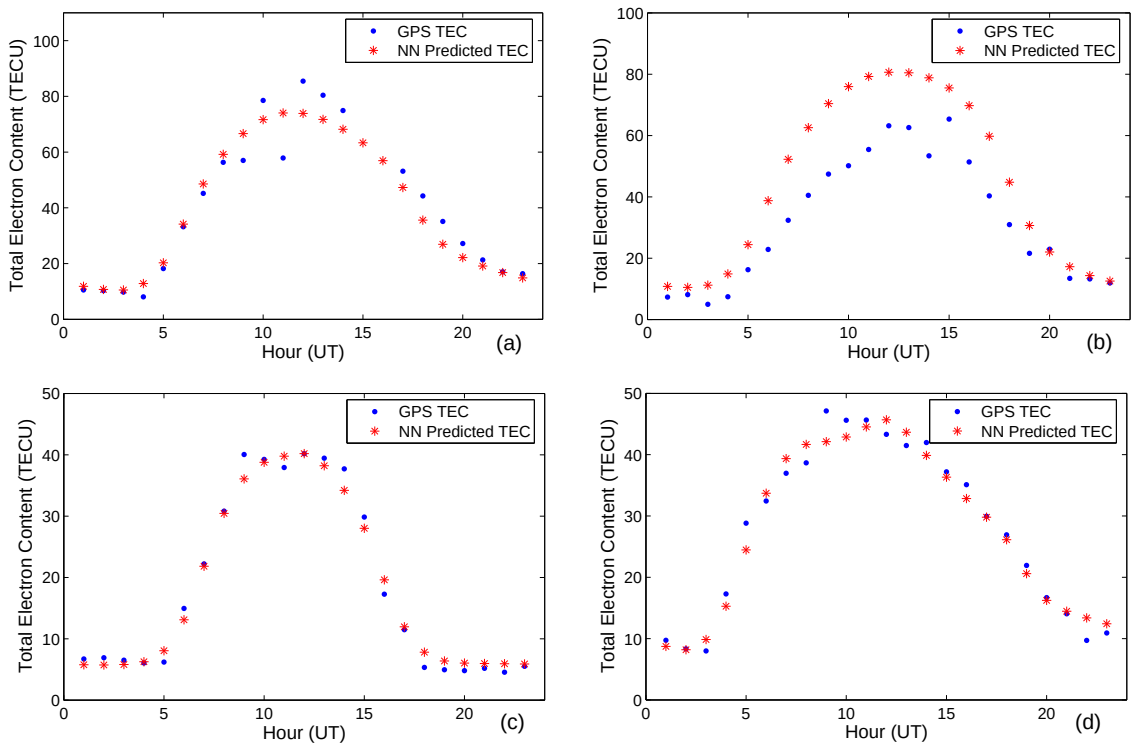


Figure 5.5: TEC variations for Sutherland during equinoxes: (a) March 22, (b) September 22 and solstices: (c) June 22, (d) January 8, all in 2000.

the NN 10h00 UT model (Figure 5.2), the former has a RMSE of 11.72 TECU and the RMSE for the 10h00 UT model is 1.48 TECU less than that for the hourly model at 10h00 UT. This may be due to two reasons. The first reason is that during the development of the 10h00 UT model, data for the year 2000 was used in training and later tested for, while for the hourly model, the data for 2000 was reserved solely for testing. Secondly, there are two additional inputs present in the hourly model namely DNS and DNC (sine and cosine components of hour) which are not used in the 10h00 UT prediction model, and therefore the input space includes an additional dimension. The few data points available for the year 2001 (Figure 5.9 (b)) used in NN training for the hourly model probably contributed to the difference in RMSE between the two NN models. It should be noted that prediction in the hourly model is done for a high solar activity year (2000) for which the input space to the NN had limited data (only 2001 with significant missing data), representing the considered parameters that affect TEC variability.

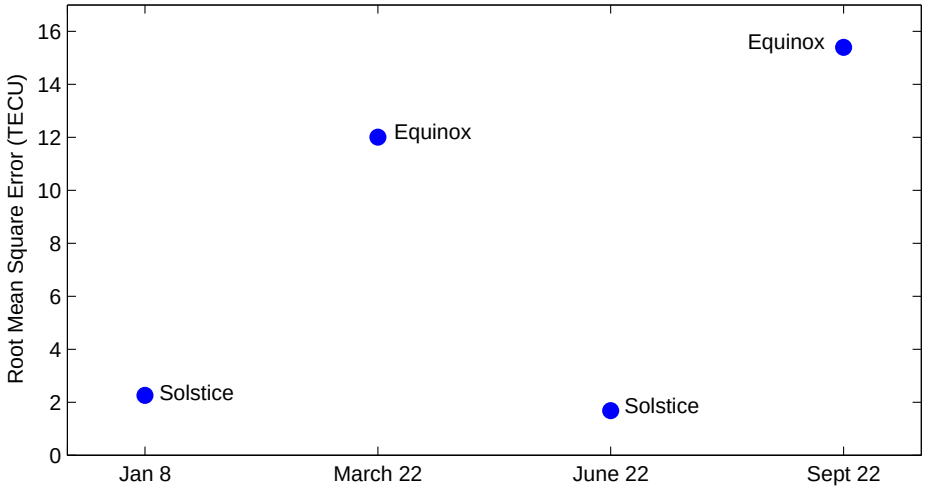


Figure 5.6: RMSEs for days representing solstices and equinoxes in 2000 for Sutherland (32.38^{o}S, 20.81^{o}E).

Figure 5.7 shows the scatter plot for the NN predicted TEC and measured GPS TEC for days representing equinoxes and solstices over Sutherland (32.38^{o}S, 20.81^{o}E), with lines of best fit inserted for all cases. Large TEC variations are observed in Figure 5.7-(a) and (b). For a perfect match (prediction) the slope should be unity. The slope is nearer to unity in Figure 5.7 (c) and (d), which is a confirmation of the results plotted in Figure 5.5 namely, that better prediction results are obtained at solstices than at equinoxes.

From Figure 5.3 it is clear that there were few data points during March equinox that could have caused the large TEC variations between the predicted and measured TEC in Figure 5.7 (a). The small number of data points correspond to the limited number of available input parameters which influence TEC during this period.

Another reason for low prediction accuracy could be the occurance of certain physical activities such as sudden ionospheric changes due to travelling ionospheric disturbances, magnetic storms as well as magnetic and ionospheric substorms. For example, a study of GPS TEC changes over mid-latitude sites showed that TEC values were significantly depleted during the geomagnetic storm that occured on

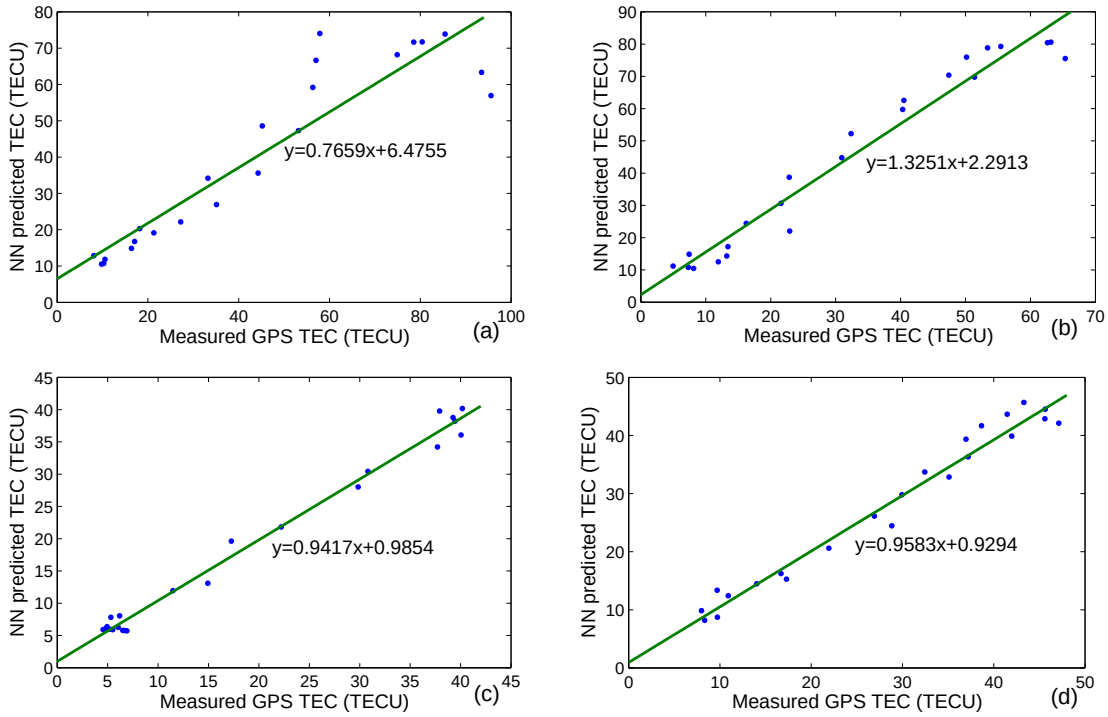


Figure 5.7: GPS TEC versus NN predicted TEC variations for Sutherland (32.38^{o}S, 20.81^{o}E) during equinoxes: (a) March 22, (b) September 22 and solstices: (c) June 22, (d) January 8, all in 2000.

Table 5.1: Average mean and standard deviation values (in TECU) over each of the four days tested which represent the equinoxes and solstices in 2000 for Sutherland (32.38^{o}S, 20.81^{o}E).

Day (2000)	Mean GPSTEC	Std Dev. GPSTEC	Mean NN	Std Dev. NN	Average E_{rel}	Relative correction(%)
March 22	44.76	28.56	40.76	23.71	0.158	84.2
June 22	18.42	14.32	18.34	13.56	0.129	87.1
Sept 22	32.31	20.05	45.10	27.15	0.442	55.8
Jan 8	28.17	13.47	27.92	13.09	0.083	91.7

September 22, 1999 (Yizengaw and Essex, 2002). None of these physical activities was investigated in this study during TEC predictions using NNs. Further study should be conducted to determine the level of effect of these activities on TEC prediction results.

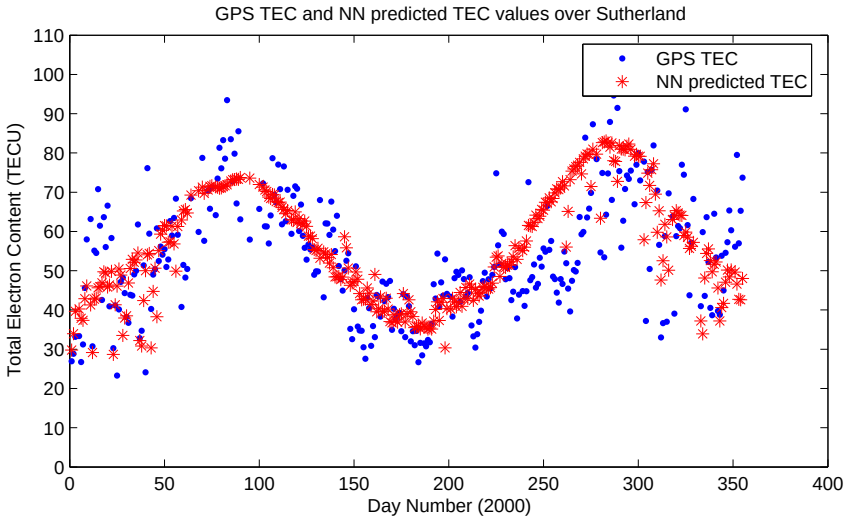


Figure 5.8: Noon GPS TEC and NN predicted values from the hourly model for Sutherland (32.38^{o}S, 20.81^{o}E) during the year 2000 at 10h00 UT.

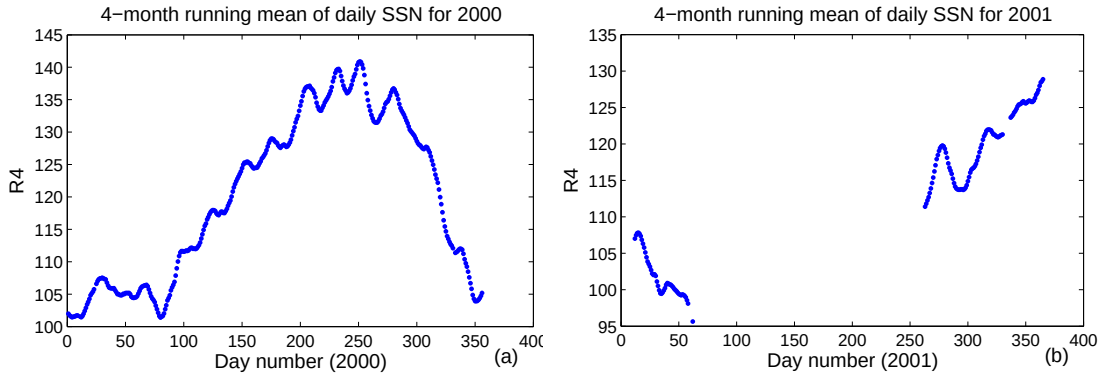


Figure 5.9: Four months running mean of daily sunspot number for (a) 2000 and (b) 2001 data recorded at Sutherland (32.38^{o}S, 20.81^{o}E).

There could also be other parameters that affect GPS TEC which were not considered as inputs to the NN. NNs learn the trend and non-linear characteristics of data when provided with a set of known input and output parameters. A generalisation is made and then applied to another data set with a similar parameters for prediction.

Table 5.1 shows a summary of statistics for GPS TEC and NN TEC at equinox and solstice days in 2000 over Sutherland (32.38^{o}S, 20.81^{o}E). In this table, the average relative error (E_{rel}) and relative correction (Rel_{corr}) values are computed

using their definitions in equations (4.7) and (4.9). The average large relative error (relative correction during Spring of about 56%) in Table 5.1 is a confirmation of the less accurate prediction level of the developed NN hourly model for periods with insufficient data available for training.

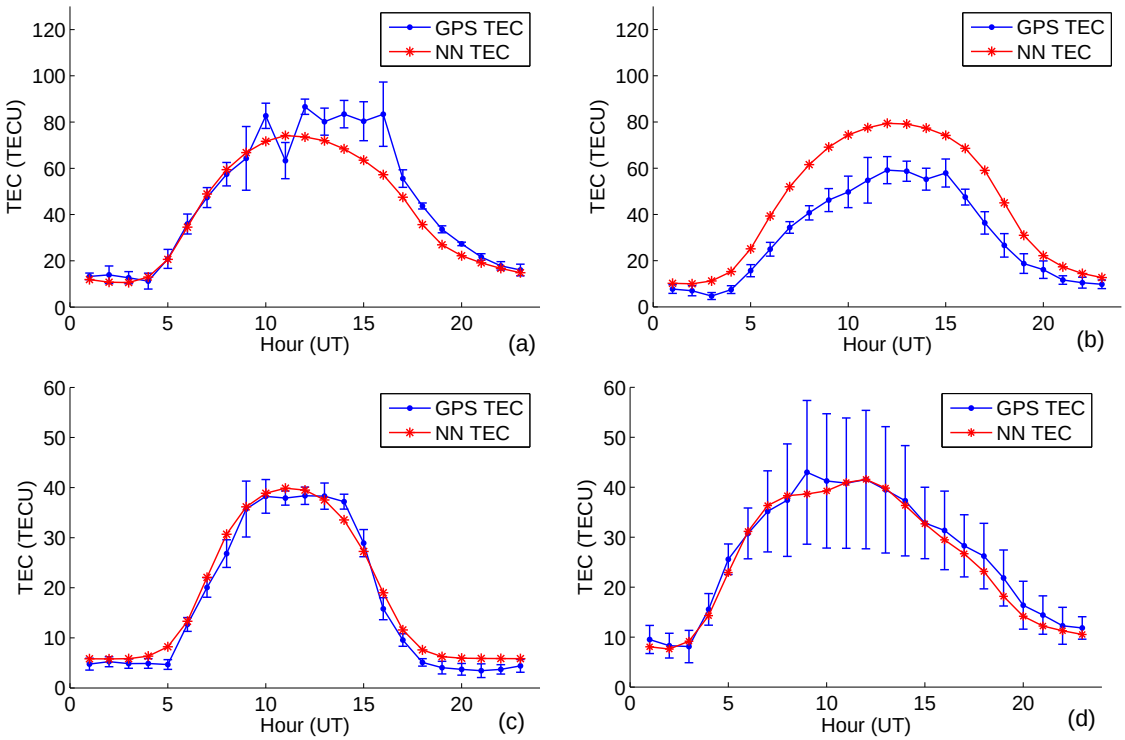


Figure 5.10: GPS TEC and NN TEC variations for Sutherland (32.38^{o}S, 20.81^{o}E) averaged over 7 days (3 days before and after days representing equinoxes and solstices) during: (a) Autumn, (b) Spring, (c) Winter and (d) Summer, all in 2000.

Averaging TEC variations over a week (nearest 3 days before and after the days representing equinoxes and solstices) also shows that the NN hourly model over predicts GPS TEC during the spring equinox (Figure 5.10(b)).

5.3 Springbok (29.67^{o}S, 17.88^{o}E)

A procedure similar to the one used to create Sutherland's hourly model was applied to data from a GPS receiver station located at Springbok, South Africa (29.67^{o}S, 17.88^{o}E). There were three years of available data (2002, 2003 and 2004). Data used for training and validating purposes consisted of 16 942 data points as depicted in Figure 5.12. The last two months (November and December) of 2004 were reserved for testing, which means that the training set included all data from January 2002 to October 2004 inclusive. In 2004, equinoxes and solstices occured as follows:

- Autumn or March equinox; 20^{th} March
- Winter or June solstice; 21^{st} June
- Spring or September equinox; 22^{nd} September
- Summer or December solstice; 21^{st} December

Due to unavailability of data for March equinox (days 20^{th} and 21^{st}) and December solstice, the days tested for in Autumn and Summer are 22^{nd} March and 19^{th} December respectively. The same architecture that proved to be optimum for the Sutherland hourly model was also optimum for the Springbok hourly model. It should be noted that this observation was made after considering the two stations independently. Predictions for daily GPS TEC values at 10h00 UT in 2004 were done and Figure 5.11(a) shows results over the entire year while November and December predictions are separately plotted in Figure 5.11(b). For the data in Figure 5.11(a) the RMSE is 6.56 TECU and the average relative error is 0.161 which implies that the NN is capable of predicting about 84 % of the measured GPS TEC on average.

5.3.1 Comparison of GPS TEC, NN predicted TEC and ITEC

The GPS receiver located at Springbok was the closest receiver to Louisvale, South Africa ionosonde station (28.5^{o}S, 21.2^{o}E) with sufficient data available for testing the feasibility of NN modelling. Therefore GPS TEC values recorded at Springbok were compared to ionosonde TEC (ITEC) values from Louisvale. The data

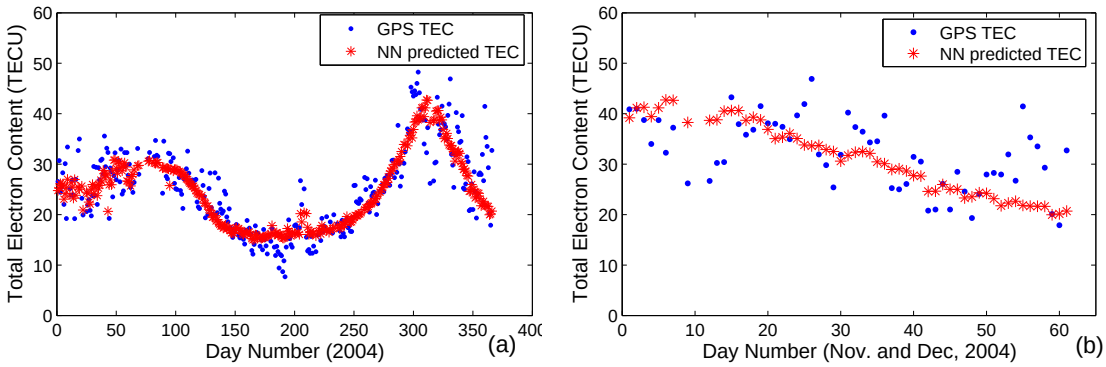


Figure 5.11: GPS TEC and NN predicted values during (a) 2004 and (b) November and December 2004 for Springbok (29.67^{o}S, 17.88^{o}E) at 10h00 UT.

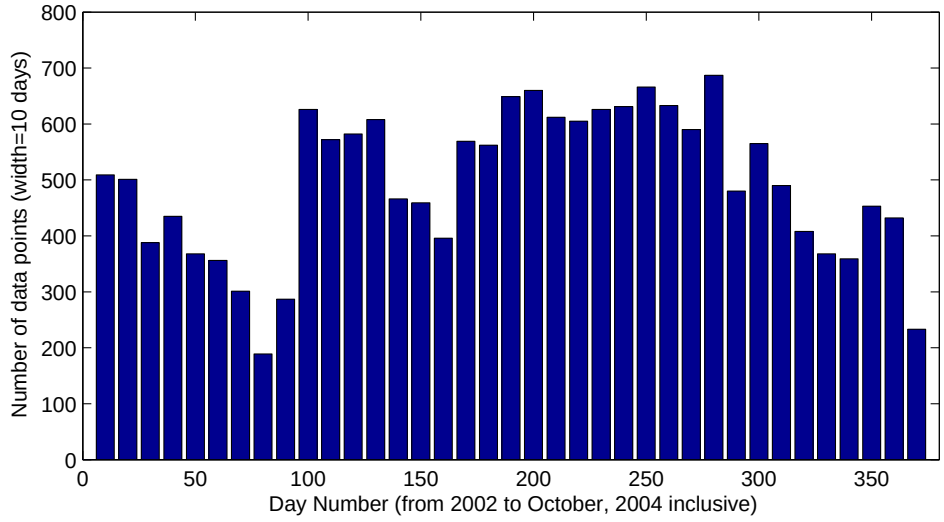


Figure 5.12: A histogram showing the number of data points used to train the hourly model's NN for Springbok (29.67^{o}S, 17.88^{o}E).

from the Louisvale digital pulse sounder (DPS) is automatically scaled using the University of Massachusetts Lowell Center for Atmospheric Research (UMLCAR) software known as ARTIST4 which gives the output in Standard Archiving Output (SAO) format. The Louisvale ionograms for equinox and solstice days in 2004 were manually edited using SAO explorer (Reinisch *et al.*, 2004), from which ITEC values were extracted that represent TEC to an altitude of 1000 km. While the bottomside ionosonde profile is measured, the topside is modelled. More details about the verification and editing of ionograms obtained from digital pulse sounders can

be found in Reinisch *et al.*, 2004. Though the GPS receiver station is not located exactly at the ionosonde station, the results show the level of variability of TEC obtained using the two methods. As expected, ITEC is lower than GPS TEC due to the difference in height at which TEC is determined. The comparison of GPS TEC and ITEC at a South African station has been done by McKinnell *et al.*, 2007, and this study confirms their findings.

The measured and predicted TEC values are graphically represented in Fig-

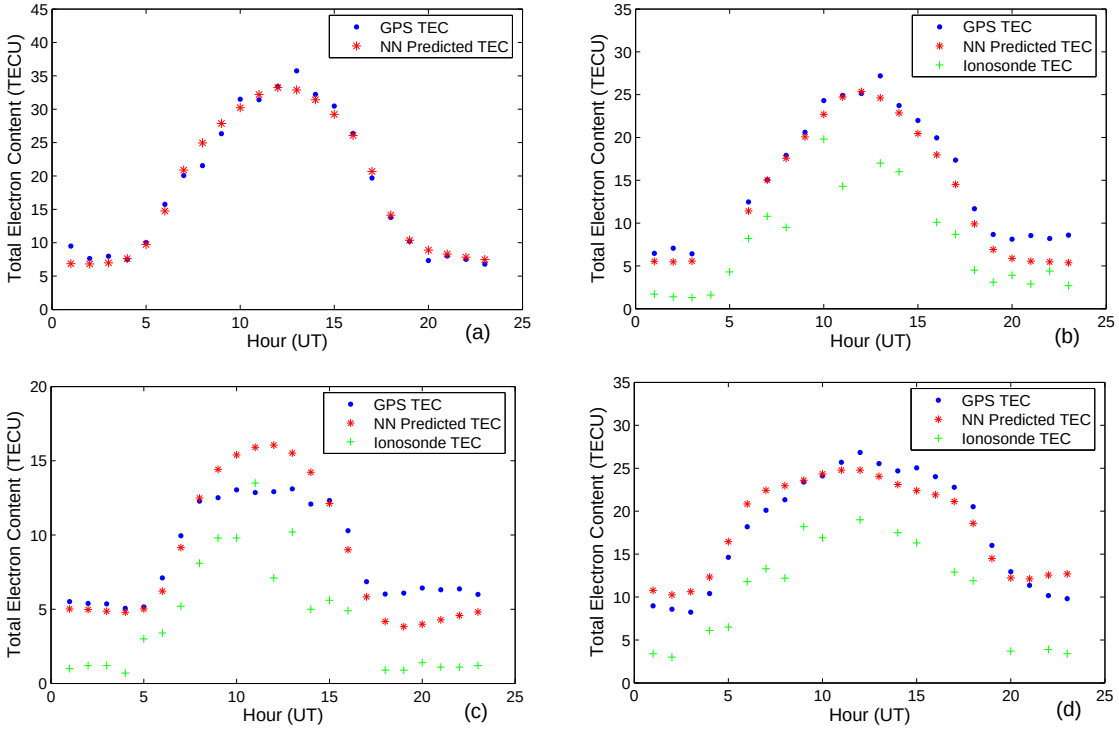


Figure 5.13: TEC variations for Springbok during equinoxes: (a) March 22, (b) September 22 and solstices: (c) June 21, (d) December 19, all in 2004.

ure 5.13 with ITEC values plotted for Spring equinox (b) and solstices (c) and (d). The variability of the ITEC values during Autumn equinox is not shown in Figure 5.13 (a), because observations recorded at Louisvale during March ended on 17^{th}. In Figure 5.13(a) - (d) TEC values are lower compared to TEC in Figure 5.5(a) - (d). The reason does not only relate to the two different stations considered, but is also linked to the levels of solar activity during both 2000 and 2004.

In Figure 5.13 the NN is trained on GPS TEC and predicts GPS TEC well. The difference between GPS TEC and ITEC illustrates the importance of training the NN on the best TEC values available. Thus, if a better representation of "true" TEC other than GPS TEC could be found with the same or more abundant data, a NN should do as well.

5.4 Summary and discussion

A NN single hour model to predict TEC at 10h00 UT for Sutherland (32.38^{o}S, 20.81^{o}E) was developed with the inputs used representing seasonal variation, magnetic and solar activities. This model was extended to an hourly model by taking into account all hours of the day. This led to diurnal variation being included in the input space. The one hour model was compared to the IRI model and it was found that the IRI underpredicts TEC during the summer and overpredicts TEC during winter months.

Data analysis based on equinoxes and solstices revealed that the NN model over predicts GPS TEC at spring equinox, and that there exists large variations between predicted and measured TEC during both Autumn and Spring equinoxes. The reason for this was traced to a lack of sufficient data during the period of high solar activity since the prediction was done for a high solar activity year. TEC variations are also generally larger during equinoxes (Kouris *et al.*, 1999). There was a lack of data during the year 2001 which was the last high solar activity year in solar cycle 23. A large database is one of the most important requirements in applications involving NN modelling and therefore NNs may not perform well without enough data.

A separate hourly model was developed for Springbok (29.67^{o}S, 17.88^{o}E) to validate the suitability of NNs in predicting TEC hourly values. Results obtained in Figure 5.13 confirmed that the prediction is possible when there is a sufficient amount of data. To further validate the GPS TEC values, GPS TEC at Springbok during equinoxes and solstices was compared to ITEC at Louisvale (28.5^{o}S, 21.2^{o}E). ITEC is less than GPS TEC due to the presence of plasmaspheric contribution in the GPS TEC.

By computing relative errors it has been shown that NNs are capable of predicting at least 80% of the GPS TEC on average during periods where there is sufficient data available for NN modelling. The prediction of GPS TEC at 10h00 UT by the NN from both the 10h00 UT model and the hourly model at Sutherland (32.38^{o}S, 20.81^{o}E) seem to agree. The difference, as quantified through the RMSE method, can be mainly attributed to the additional parameters DNS and DNC (diurnal variation) and the few data points representing high solar activity periods during NN training in the hourly model. Data for both solar activity years (2000 and 2001) was used in training during the construction of the 10h00 UT model, while only data for 2001 was included in the training set for the case of the hourly model.

Chapter 6

Discussion and conclusions

This chapter summarises the results discussed in the previous chapters and gives an insight into future work on the modelling of TEC predictions. The aim of this study was to investigate whether NNs can be used to predict TEC obtained from GPS satellite measurements. It is the first attempt to model TEC with NNs using South African derived data and the preliminary results obtained show that NNs are indeed suitable for TEC modelling with the following input parameters: day number (DN), hour (HR), a 4-month running mean of daily sunspot number (R4) and the running mean of the previous eight 3-hourly magnetic A index values (A8).

6.1 Discussion

The ionosphere is a difficult region to model due to the continuous variability of its parameters. TEC in the ionosphere has a non-linear variability since it is determined by dynamic processes related to solar radiations, solar wind changes, ionospheric storms and magnetic storms. NNs are important tools for non-linear approximations (Bishop, 1995; Barrile *et al.*, 2006) and have been used before in modelling involving GPS TEC observations (Hernandez-Pajares *et al.*, 1997; Xenos *et al.*, 2003; Leandro and Santos, 2004, 2007).

The ASHA model was developed for a regional ionospheric TEC mapping and prediction service over Southern Africa and contains an algorithm which allows the derivation of TEC from the GPS receiver data (Opperman, 2007). In this study GPS TEC values as derived by the ASHA model were not compared to

any other model constructed to derive GPS TEC. Thus, the GPS TEC values as computed by the ASHA algorithm, were taken as the "true" TEC values and compared to the corresponding TEC values of the NN and IRI models. NNs predict GPS TEC values well and therefore, if a better representation of "true" TEC other than GPS TEC could be found with the same abundance of data, one would expect NNs to perform as well. In future, prediction will be done using GPS TEC values computed by different models.

Comparison of NN and IRI-2001 models indicate that NNs provide more accurate prediction results than the IRI model over South African GPS station locations. This was observed for both the SKA Hub location (30.71^{o}S, 21.39^{o}E) and Sutherland (32.38^{o}S, 20.81^{o}E) at 10h00 UT. In predicting TEC using NNs, the solar activity was represented by a 4-month running mean of daily SSN (R4) in contrast with the IRI which uses the 12-month running mean of SSN (R12). Also the IRI model gives the estimate TEC values without taking into account the magnetic activity effects on TEC variations which were represented in NN modelling by the previous eight 3-hourly magnetic A index values (A8) derived from data recorded at Hermanus Magnetic Observatory (34.42^{o}S, 19.22^{o}E). NN TEC prediction using R12 was done over a South African location SKA Hub (30.71^{o}S, 21.39^{o}E), and not found appropriate due to the fact that the RMSE between the GPS TEC and NN predicted TEC was high (Figure 4.1). The rest of the parameters have similar representations except that day number and hour were each separated into two cyclic components to allow data continuity in the NN model. This is because data for December 31 and January 1 are relatively far apart and thus the NN experiences difficulty in considering them as adjacent data points (Williscroft and Poole, 1996).

There is a general agreement between GPS TEC and NN predicted TEC values for Sutherland and Springbok stations when considering the equinoxes and solstices. The differences relating to an overprediction for Sutherland and Springbok in Spring and Winter respectively, correlate with the data set itself. It has been shown that there were fewer data points used in training the NN for these periods (Figures 5.3 and 5.12). For Sutherland, it was found that data set used in NN training during the high solar activity period (2001) was very limited (Figure 5.9 (b)) and this could have been the major reason for the overprediction of TEC in

Spring by the NN, since the testing year (2000) was also a high solar activity year. TEC values also have greater variability during equinoxes (Kouris *et al.*, 1999). Less accurate prediction results could also be due to other factors not investigated in this study. Various physical activities take place within the atmosphere such as travelling ionospheric disturbances (TIDs), magnetic storms, ionospheric storms as well as magnetic and ionospheric substorms. The effects of these physical phenomena on TEC variations have not been considered and are worth investigating so that one can confirm the TEC prediction level.

In all three cases considered, the achieved average NN prediction accuracy is at least 80% for periods where substantial data was available for NN training. This percentage was exceeded in the 10h00 UT model for testing year (2002) at Sutherland (32.38^{o}S, 20.81^{o}E). This is because 2002 fell in the middle of the training set and the obtained result was expected since NNs are known to interpolate well within the input space (Fausett, 1994). There were five years of data available (2000 - 2004) for the SKA Hub (30.71^{o}S, 21.39^{o}E) and Sutherland (32.38^{o}S, 20.81^{o}E), while Springbok (29.67^{o}S, 17.88^{o}E) had only three years (2002, 2003 and 2004) of data. The years 2000 and 2001 were near solar maximum while 2003 and 2004 were near solar minimum (Figure 4.4). An improvement in results is expected when data for at least 11 years (one solar cycle) is used so that both high and low solar activity periods are taken into account.

There is no known standard formula for determining the number of hidden nodes to be included in the hidden layer. While determining the optimum parameters for solar activity and magnetic activity using the SKA Hub data at 10h00 UT, the optimum NN consisted of four input nodes, seven hidden nodes and one output node. The same architecture proved to be optimum during the development of 10h00 UT model for Sutherland. There were four inputs to the NN: DNS, DNC, R4 and A8. In the hourly models for both Sutherland and Springbok, the number of inputs increased to six by including the two components of hour (HRS and HRC). The architecture that gave the least RMSE and hence considered optimum, had nine nodes in the hidden layer. One could conclude that the number of hidden

nodes H_N can be approximated by:

$$H_N \approx n + 3 \tag{6.1}$$

where n is the number of inputs to the NN, as shown in Table 6.1. However, more extensive study is needed to prove this.

Table 6.1: Summary of NN architecture specifications. A binary sigmoid function was used in all cases as the neuron activation function

GPS station	Input nodes	Hidden nodes	Output nodes
SKA hub (10h00 UT model)	4	7	1
Sutherland; (a) 10h00 UT model	4	7	1
(b) Hourly model	6	9	1
Springbok (Hourly model)	6	9	1

Comparison of GPS TEC at Springbok and ITEC at the Louisvale ionosonde station was done for validation. GPS TEC is greater than ITEC due to plasmaspheric contributions present in the former. GPS TEC and ITEC comparisons at a South African station, Grahamstown (33.3^{o}S, 26.5^{o}E) were previously done (McKinnell *et al.*, 2007), and this study confirms the results. It however, remains a challenge to accurately validate GPS TEC since there is no known way of obtaining absolute TEC. In the estimation of ITEC, the bottomside ionosphere is measured while the topside is modelled. Therefore, part of a model is used to validate another model. Furthermore, the data from an ionosonde only provides a measure of the electron density distribution directly above the geographical location of the ionosonde.

An attempt to include the solar wind as an input parameter was done as described in chapter 4, section 4.7. The results obtained do not show an improvement, neither do they introduce significant increments in the RMSEs especially when solar wind bulk velocity is considered. This makes it difficult to reach a conclusion about the level of TEC variability with changing solar wind parameters. It would therefore be prudent to recommend that solar wind parameters be studied in detail with regard to their influence on TEC predictions, especially the solar wind velocity and the electron number density. Other physical parameters such as solar wind dynamic pressure, ion flux, temperature, etc should be investigated

to determine whether they have any effects on TEC variability. Data from Low Earth or Medium Earth Orbit satellites may be advantageous since the altitudes of these satellites are quite low compared to the altitude of the Advanced Composition Explorer satellite (about 1.28 million km). LEO satellites orbit at an altitude of about 250 - 2000 km while MEO satellites orbit above LEO but below geostationary satellites the altitude of which is about 36 000 km above the Earth's surface.

6.2 Summary of Results

The study started by identifying parameters that influence TEC variations, followed by the determination of appropriate representations for parameters of which the level of influence towards TEC fluctuations was unknown. The following is a summary of the results obtained:

- In addition to seasonal and diurnal variations represented by day number (DN) and hour (HR) respectively, the optimum parameters for solar activity and magnetic activity to predict TEC were determined as a 4-month running mean of the daily sunspot number (denoted as R4) and the running mean of the previous eight 3-hourly magnetic A index values (A8).

- A NN 10h00 UT model was developed for Sutherland. Separate hourly models were also constructed and discussed for both Sutherland and Springbok GPS stations.

- NNs were found suitable for GPS TEC modelling using data from the South African GPS receiver stations considered. It was observed that the number of hidden nodes in the hidden layer increases with the number of inputs to the NN.

- More data is required for at least one solar cycle to effectively represent both high and low solar activity periods during the NN training and validating processes.

- The choice of TEC limits is an aspect that requires to be investigated further. This is because TEC changes significantly during disturbed conditions such

as ionospheric and magnetic storms. There could have been a possibility of underestimating or even overestimating the maximum TEC value at mid-day

Thus, the well established factors that influence TEC are seasonal and diurnal variations, magnetic activity and solar activity.

6.3 Future work

In this feasibility investigation, the study was restricted to three locations within South Africa. Future work will include investigating other locations within South Africa, particularly near the other ionosonde stations, and developing one TEC model for the whole of South Africa. Currently, the network of dual frequency GPS receivers over South Africa contains 39 locations, with plans for extensions in the future. The developed model can be expanded to cover a wider geographical area based on the achieved results. To develop a truly representative model for the TEC parameter over South Africa, data for at least 11 years (one solar cycle) is needed from at least 9 receivers distributed uniformly across the country. However, the model can be developed using data presently available in a such way as to allow for future updates. One of the uses for a TEC model over South Africa would be in the prediction of TEC maps for space weather purposes where real-time TEC data is required and may not be available. Therefore, the primary present intention is to create an accurate, well-represented database of GPS TEC values for South Africa, and to use this database to develop an accurate NN based model for the prediction of this parameter.

References

Barrile V., Cacciola M. and Cotroneo F., "Multipath Reduction of GPS Measures through Heuristic Techniques of Compensation", *Progress In Electromagnetic Research Symposium 2006, Cambridge, USA*, pp. 528–532, 2006.

Bhuyan P. and Borah R.R., "TEC derived from GPS Network in India and comparison with models", *Advances in Space Research*, **39**, pp. 830–840, 2007.

Bilitza D., "International Reference Ionosphere 2000", *Radio Science*, **36**(2), pp. 261–275, 2001.

Bilitza D., Reinisch B.W., Radicella S.M., Pulinets S., Gulyaeva T. and Triskova L., "Improvements of the International Reference Ionosphere model for topside electron density profile", *Radio Science*, **41**, pp. RS5S15, doi 10:1029/2005RS00370, 2006.

Bishop C.M., *Neural Networks for Pattern Recognition* , Oxford University Press Inc.,New York, 1995.

Chen F., *Introduction to Plasma Physics and Controlled Fusion*, Plenum Press, New York, 1984.

Dana P., "Global Positioning System Overview", University of Texas, USA, http://www.ncgia.ucsb.edu/giscc/units/u017/u017.html, posted August 28, 1997.

Fausett L., *Fundamentals of Neural Networks; Architectures, Algorithms and Applications*, Prentice-Hall, Inc. New Jersey 07632, 1994.

Fedrizzi M., de Paula E.R., Langley R.B., Komjathy A., Batista I.S. and Kantor I.J., "Study of the March 31, 2001 magnetic storm effects on the ionosphere using GPS data", *Advances in Space Research*, **36**(3), pp. 534–545, 2005.

Gao Y. and Liu Z., "Precise Ionospheric Modelling Using Regional GPS Network Data", *Journal of Global Positioning System*, **1**(1), pp. 18–24, 2002.

Haykin S., *Neural Networks, A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.

Hernandez-Pajares M., Juan J. and Sanz J., "Neural network modelling of the ionospheric electron content at global scale using GPS", *Radio Science*, **32**, pp. 1081–1090, 1997.

Hofmann-Wellenhof B., Lichtenegger H. and Collins J., *Global Positioning System Theory and Practice* , Springer-Verlag Wien New York, 1992.

Jakowski N., Hocke K., Schluter S. and Heise S., "Space weather effects detected by GPS based TEC monitoring", *Proceedings, Workshop on Space Weather, WPP-155, ESTEC, Noordwijk*, pp. 241–244, 1999.

Klobuchar J., "Ionospheric Effects on GPS", *Reprinted from GPS World*, 1991.

Kouris S.S., Fotiadis D.N. and Zolesi B., "Specifications of the F-region variations for quiet and disturbed conditions", *Physics and Chemistry of the Earth*, **24**(4), pp. 321–327, 1999.

Kouris S.S., Xenos T.D., Polimeris K.V. and Stergiou D., "TEC and foF2 variations: preliminary investigations", *Annals of Geophysics*, **47**(4), pp. 1325–1332, 2004.

Langley R.B., "GPS, the Ionosphere, and the Solar Maximum", *GPS World*, **11**(7), pp. 44–49, 2000.

Leandro R.F. and Santos M., "Regional Computation of TEC using a Neural Network Model", Poster presented at the Joint Assembly of CGU, AGU, SEG and EEGS, Montreal, 2004.

Leandro R.F. and Santos M.C., "A neural network approach for regional vertical total electron content modelling", *Studia Geophysica et Geodaetica*, **51**(2), pp. 279–292, 2007.

McKinnell L.A., *A Neural Network based Ionospheric model for the bottomside electron density profile over Grahamstown South Africa*, PhD thesis of Rhodes University, Grahamstown, 2002.

McKinnell L.A., Opperman B. and Cilliers P.J., "GPS TEC and Ionosonde TEC over Grahamstown, South Africa: First Comparisons", *Advances in Space Research*, **39**, pp. 816–820, 2007.

Mosert M., Gende M., Brunini C., Ezquer R. and Altadill D., "Comparison of IRI TEC predictions with GPS and digisonde measurements at Ebro", *Advances in Space Research*, **39**, pp. 841–847, 2007.

NAVSTAR GPS, 1996, *Navigation Signal Timing and Ranging (NAVSTAR) GPS user equipment introduction (Public Release Version)*, http://www.navcen.uscg.gov/pubs/, 1996.

Opperman B., *Reconstructing Ionospheric TEC over South Africa using signals from a Regional GPS network*, Submitted as a PhD thesis to Rhodes University, Grahamstown, 2007.

Opperman B.D.L., Cilliers P.J., McKinnell L.A. and Haggard R., "Development of a Regional GPS-based Ionospheric TEC model for South Africa", *Advances in Space Research*, **39**, pp. 808–815, 2007.

Poole A.W. and McKinnell L.A., "On the predictability of foF2 using neural networks", *Radio Science*, **1**, pp. 225–234, 2000.

Reddy C., "Study of space weather effects using GPS", *Proceedings 2002, GPS in Atmospheric Sciences*, India International Centre, New Delhi, India, 2002.

Reinisch B.W., Galkin I.A., Khmyrov G., Kozlov A. and Kitrosser D.F., "Automated collection and dissemination of ionospheric data from the digisonde network", *Advances in Radio Science*, **2**, pp. 241–247, 2004.

Sarma A. and Mahdu T., "Modelling of foF2 using neural networks at an equatorial anomaly station", *Current Science*, **89**(7), pp. 1245–1247, 2005.

Sibanda P., *Particle precipitation effects on the South Africa ionosphere*, Master's thesis, Rhodes University, Grahamstown, South Africa, 2006.

Tulunay E., Senalp E.T., Radicella S.M. and Tulanay Y., "Forecasting total electron content maps by neural network technique", *Radio Science*, **41**, doi:10.1029/2005RS003285, 2006.

Williscroft L.A. and Poole A.W., "Neural Networks, foF2, sunspot number and magnetic activity", *Geophysical Research Letters*, **23**(24), pp. 3659–3662, 1996.

Xenos T.D., Kouris S.S. and Casimiro A., "Time-dependent prediction degredation assessment of neural-networks-based TEC forecasting models", *Nonlinear Processes in Geophysics*, **10**, pp. 585–587, 2003.

Yizengaw E. and Essex E.A., "Use of GPS Signals to Study Total Electron Content of the Ionosphere during Geomagnetic storm on 22 September 1999", *Workshop on Applications of Radio Science (WARS)*, Blue Mountains, Australia, 2002.

Yizengaw E., Moldwin M., Dyson P. and Immel T., "Southern Hemisphere ionosphere and plasmasphere response to the interplanetary shock event of 29-31 October 2003", *Geophysical Research, A09S30, doi:10.1029/2004JA010920*, **110**, 2005.

Zell A., Mamier G.M. and Vogt N., *Stuttgart Neural Network Simulator (SNNS), User Manual ,Version 4.2*, Universities of Stuttgart and Tubingen, 1998.