AN INVESTIGATION INTO IMPROVED IONOSPHERIC F1 LAYER PREDICTIONS OVER GRAHAMSTOWN, SOUTH AFRICA.

A thesis submitted in fulfilment of the

requirements for the degree of

MASTER OF SCIENCE

of

RHODES UNIVERSITY

by

Linda Jacobs

January 2005

ABSTRACT

This thesis describes an analysis of the F1 layer data obtained from the Grahamstown (33.32°S, 26.50°E), South Africa ionospheric station and the use of this data in improving a Neural Network (NN) based model of the F1 layer of the ionosphere. An application for real-time ray tracing through the South African ionosphere was identified, and for this application real-time evaluation of the electron density profile is essential. Raw real-time virtual height data are provided by a Lowell Digisonde (DPS), which employs the automatic scaling software, ARTIST whose output includes the virtual-to-real height data conversion. Experience has shown that there are times when the ray tracing performance is degraded because of difficulties surrounding the real-time characterization of the F1 region by ARTIST. Therefore available DPS data from the archives of the Grahamstown station were re-scaled manually in order to establish the extent of the problem and the times and conditions under which most inaccuracies occur. The re-scaled data were used to update the F1 contribution of an existing NN based ionospheric model, the LAM model, which predicts the values of the parameters required to produce an electron density profile. This thesis describes the development of three separate NNs required to predict the ionospheric characteristics and coefficients that are required to describe the F1 layer profile. Inputs to the NNs include day number, hour and measures of solar and magnetic activity. Outputs include the value of the critical frequency of the F1 layer, foF1, the real height of reflection at the peak, hmF1, as well as information on the state of the F1 layer. All data from the Grahamstown station from 1973 to 2003 was used to train these NNs. Tests show that the predictive ability of the LAM model has been improved by incorporating the re-scaled data.

DECLARATION

The contents of Chapter 2 were presented at the International Reference Ionosphere (IRI) Workshop in October 2003, and at the local South African Institute of Physics (SAIP) conference held in Bloemfontein in June 2004
The contents of Chapter 2 were also used as the subject of a paper, co-authored by Prof. A.W.V. Poole and appear in *Jacobs et al. [2004]*.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Prof. Poole for all his help and patience. Special thanks to Lee-Anne McKinnell for her input, encouragement and support throughout the course of this thesis.

I would like to acknowledge GrinTek Ewation and the Hermann Ohlthaver Institute for Aeronomy (HOIA) for providing financial support for this project for 2003 and 2004.

I also acknowledge the use of magnetic data obtained from the Hermanus Magnetic Observatory (HMO).

Finally I would like to thank Emmanuel Lamprecht for all his help in solving my numerous computer problems.

In memory of

Santiago and Inéz

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 4

Chapter 1
INTRODUCTION

In this thesis I present an analysis of the automatically scaled F1 layer data, explicitly foF1, collected at the Grahamstown (33.3°S, 26.5°E), South Africa ionospheric station. In order to perform this analysis, automatically scaled ionograms collected using a Digital Portable Sounder (DPS) system had to be manually re-scaled. The re-scaled data has been used to update an existing F1 Neural Network (NN) based ionospheric model. Improvements to the F1 contribution to the LAM model *(McKinnell, [2002])* are demonstrated.

1.1 The Bottomside Ionosphere

The upper atmosphere is ionised by energetic radiations from the sun, resulting in free electrons and ions, which recombine to establish a balance between the electron-ion production and loss. The processes of ion production, recombination and transformation by radiation are known as photochemistry. The bottomside ionosphere covers the region from 90 – 300km and is generally categorized into three layers: E, F1 and F2. The E region is from 90 – 110km, above which lies the F1 layer extending to 160km, leading to the F2 layer which reaches its peak at about 300km above the surface of the Earth. Below the E region is the lower ionosphere, or D region, and above 300km is known as the topside ionosphere. Figure 1-1 shows a diagram of a model electron density profile of the bottomside ionosphere with each layer clearly defined. Also shown is the corresponding idealized daytime ionogram with the critical frequency of each layer indicated. Electron density is related to frequency by the equation:

$$N/m^{-3} = 1.24 \times 10^{10} (f/MHz)^2 \qquad (1\text{-}1)$$

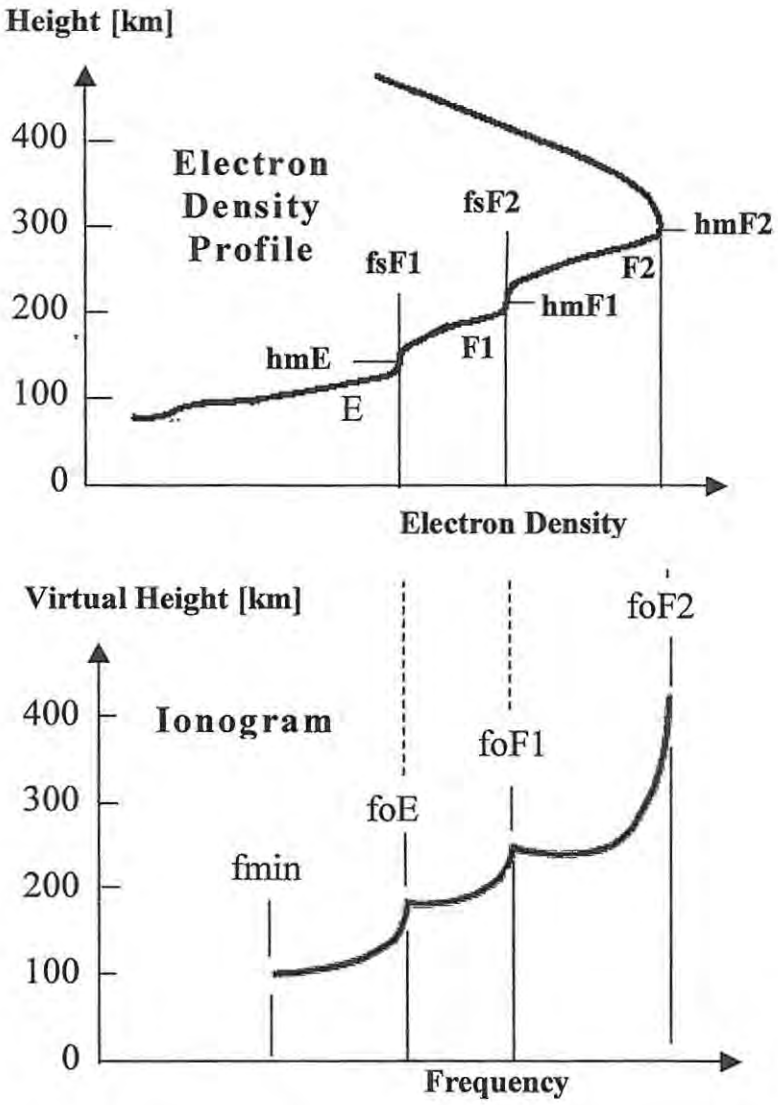


Figure 1-1: A model electron density profile with the corresponding ionogram illustrating the different regions of the bottomside ionosphere with the critical frequency of each layer indicated.

Different processes are responsible for the formation of the separate layers. The F1 layer is thought to form as a result of the processes that contribute to the formation of the E region overlapping with the processes contributing to the formation of the F2 region. Details on the physical processes that give rise to the formation of the F1 layer can be found in *McNamara [1991]* and *Rishbeth and Garriott [1969]*.

The first step in explaining the formation of any layer is to examine the formation of the ionosphere. The rate of change of electron density (N) at any altitude is given by the continuity equation *(Hargreaves, [1979])*:

$$\partial N/\partial t = \text{rate of production (q)} - \text{recombination rate (L)} - \text{loss by movements (div}(Nv)) \quad (1\text{-}2)$$

where v is the mean drift velocity of the electrons. Equilibrium is reached when $\partial N/\partial t = 0$.

Chapman theory works out the form of an ionospheric layer and its evolution with time. The Chapman production function *(Hargreaves, [1979])* gives the rate of production of ions at a height h, as:

$$q/q_m = \exp\left[1 - (h - h_m)/H - \exp\left[-(h - h_m)/H\right]\right] \quad (1\text{-}3)$$

where

H is the scale height for a monatomic gas of molecular mass m, and is defined (from the gas law) as:

$$H = kT/mg \quad (1\text{-}4)$$

T is the temperature in Kelvin

g is the gravitational constant

k is Boltzmann's constant

h_m is the height at which maximum production, q_m, occurs.

Electron loss in most of the ionosphere occurs not directly but by a two-stage process:

$$X^+ + A_2 \rightarrow AX^+ + A \qquad \text{(i)} \quad \text{Rate} = \beta[X^+]$$

$$AX^+ + e^- \rightarrow A + X \qquad \text{(ii)} \quad \text{Rate} = \alpha[AX^+]\,N \quad (1\text{-}5)$$

where

β is an attachment coefficient and is a function of altitude, and

α is the recombination coefficient.

X^+ and AX^+ are atomic ions and molecular ions, respectively, with X, A and A_2 being the corresponding neutral species.

The rate of recombination is to a degree dependent on the composition of the ionosphere and the partial pressures of each species. In the E-region the relative proportions of molecular species such as O_2 and N_2 is high and therefore the overall rate is controlled by process (ii). Higher up in the ionosphere where the F2 region is formed, the proportion of ionised species increases and the rate of recombination is controlled by the rate of process (i). Chapman theory is sufficient for explaining the shape of the E region, but in the F2 region other loss mechanisms occur, such as those arising from ambipolar diffusion, horizontal winds and electromagnetic drifts. The F1 region exists at an altitude where the transition from process (i) to process (ii) occurs. The electron density in the F1 region is given by *(Rishbeth and Garriott, [1969])*:

$$N = q/2\beta \left[1 + (1 + 4\beta^2 / \alpha q)^{\frac{1}{2}}\right] \quad \textbf{(1-6)}$$

The shape of this layer depends to some extent on the value $4\beta^2 / \alpha q$. For larger values of $4\beta^2 / \alpha q$ the layer becomes more pronounced and typically occurs during daytime hours around noon when the value of β is large. Therefore, as the value of β varies, the prominence of the F1 layer fluctuates and it has been observed that the F1 layer can exist in one of three states during the course of a day. For the purposes of this thesis, I have defined the different F1 states as follows:

(i) N – when no F1 layer is present.

(ii) F – when the F1 layer is fully formed and the critical frequency foF1 of the F1 layer can be determined.

(iii) L – when the F1 layer is in a transition state between the N and F states. An L state, otherwise known as L-condition, is indicative of the presence of an F1 layer where the peak and shape are not defined enough for measurements to be possible.

Further information on ionospheric properties and propagation can be found in *McNamara [1991]* and *Davies [1990]*.

1.2 South African Ionospheric Data

The first ionospheric station in South Africa was established in 1946 near the Johannesburg city centre. The single sweep pulse ionosonde used at the Johannesburg station was developed by Dr Wadley at the Telecommunications Research Laboratory (later the NITR) of the Council for Scientific and Industrial Research (CSIR). After several years this station was moved to Frankenwald, Johannesburg. A second station with similar equipment was established near Cape Town in 1948. In May 1971 the station near Cape Town was relocated to the Magnetic Observatory at Hermanus. During 1981 the ionosonde at Frankenwald was replaced by a KEL 42 ionosonde and in 1982 the station at Hermanus was similarly equipped. The Johannesburg and Hermanus stations continued in operation until March 1993.

The NITR also established an ionospheric station on Marion Island, in the South Indian Ocean, during the IGY of 1957/1958. The station was temporarily re-opened in 1972 using a locally developed transistorised ionosonde, and operated with variable reliability until May 1980.

During 1962 a COSSOR ionosonde was installed at the Sanae Base in Antarctica. This ionosonde was used until April 1975, when a BR Comms. Chirpsounder was installed. This made it possible to make oblique soundings between Sanae and Grahamstown. These oblique soundings were phased out during 1987 *(Baker, [1998])*.

There are currently three ionospheric stations operating in South Africa, namely Grahamstown (33.3°S, 26.5°E), Louisvale (28.5°S, 21.2°E) and Madimbo (22.4°S, 30.9°E), the locations of which are shown in figure 1-2.

The Grahamstown station has been collecting data since 1973. Prior to 1996 this data was collected using a Barry Research Vertical Chirp Sounder (Verti) and the ionograms were manually scaled *(Wakai et al., [1985])*. The database includes information on the important ionospheric characteristics and instances of L-condition but no real height information was retained. In April 1996 a DPS system, which operates on a pulse sounding technique *(UMLCAR, [1996])*, was installed at the Grahamstown station and the ionograms are automatically scaled using software developed by the University of Massachusetts Lowell Center for Atmospheric Research (UMLCAR), called ARTIST. More information on ARTIST and the scaling procedure can be found in *Huang and Reinisch [1996]* and *Reinisch and Huang [1983]*.

The ARTIST scaling software records the ionospheric characteristics and also gives a description of the electron density profile. This description is either a listing of real heights with the corresponding electron densities, or a set of Chebyshev coefficients and peak parameters. However, ARTIST was not designed to determine instances of L-condition. The other two stations, Louisvale and Madimbo, were built in 1999 and also operate with DPS systems. All three stations collect and archive the data on a half-hourly basis. GrinTek Ewation, a company that designs and builds direction finding (DF) systems, makes use of the data from all the stations for real-time ray tracing purposes. However, often this data is not available in real-time, and therefore a model was needed to predict the behaviour of the ionosphere when no real-time information is accessible.

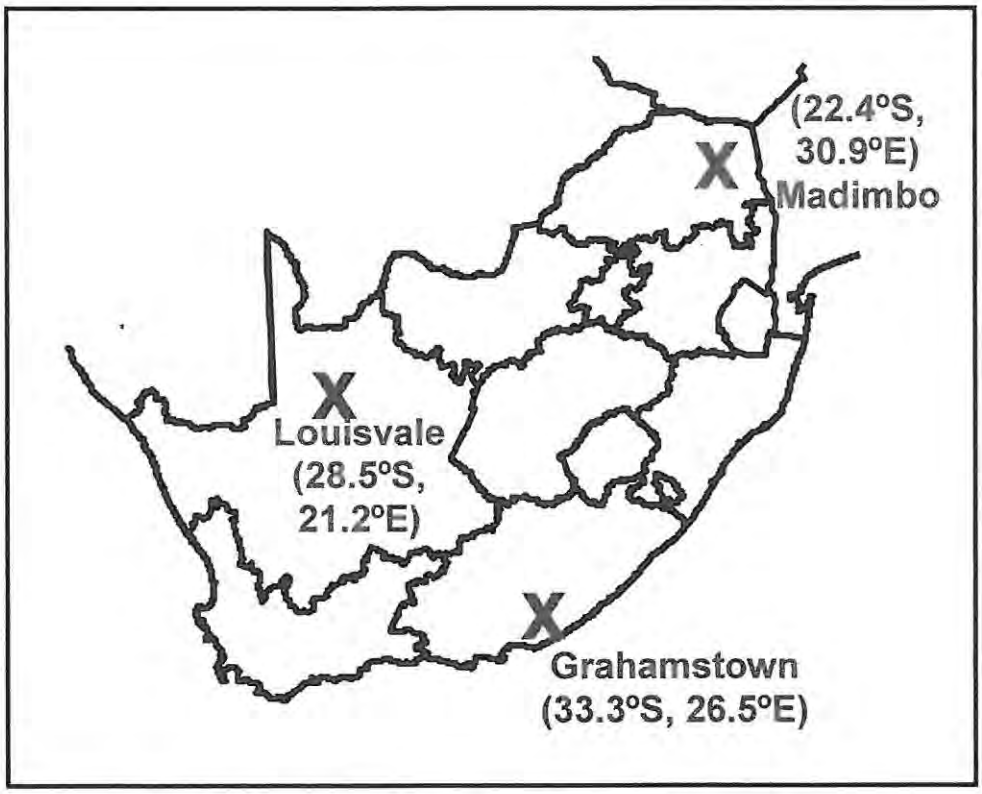


Figure 1-2: A map showing the locations of the three ionospheric stations in South Africa. From *McKinnell [2002].*

1.3 Existing Ionospheric Models

There have been many attempts in the past to produce a global model for the bottomside ionosphere. However, these global models often do not perform well at predicting ionospheric parameters over Southern Africa due to an historic lack of available ionospheric data at these latitudes. Two models are discussed in this chapter: The International Reference Ionosphere (IRI) *(Bilitza [1990], Bilitza [1997], Bilitza [2001])*, which currently is the best available global ionospheric model, and the LAM model *(McKinnell, [2002])*, developed for use in South Africa. It should also be mentioned that multi-segmented quasi-parabolic (MQP) ionospheric models were developed in South Africa for analytic near real-time ray tracing use with single station location (SSL) high frequency direction finding (HFDF) systems *(Baker and Burden [1991], Chen et al [1991])*. SSL HFDF systems and the process of

ray tracing is further discussed in Chapter 1.4. In the MQP model the ionosphere is regarded as geocentric, with the electron density distribution described in terms of distance from the centre of the earth. The electron density distribution is given by *(Baker and Burden, [1991])*:

$$N(r) = a \pm b\,(r-r_m)^2\,(r_r/r)^2 \qquad (1\text{-}7)$$

where r is the range measured from the centre of the earth, r_m is the height of the maximum or minimum and r_r is some convenient reference range. N(r) is the electron density at range r, a is the maximum (or minimum) electron density and $b = a/y^2$ where y is the layer semi-thickness.

The electron density profile is obtained by constructing quasi- and inverse quasi-parabolic segments for each region. By applying continuity requirements for the electron density distribution and its range derivatives it is possible to derive expressions for the ranges at which transitions from a lower lying inverted parabolic segment to a higher normal parabolic layer occurs.

1.3.1 The International Reference Ionosphere

The International Reference Ionosphere (IRI), founded by a joint working group of the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI), was established to provide an empirical standard model of the ionosphere. The IRI provides monthly averages of the electron density, electron and ion temperatures, and ion composition for magnetically quiet conditions between 50km and 1000km.

The IRI has adopted the relationship established by *Ducharme et al. [1971, 1973]* for the variation of the critical frequency of the F1 layer, foF1, with solar zenith angle χ, solar activity (represented by the 12 month running mean sunspot number, R12) and magnetic dip latitude (Ψ). Within the IRI this parameter is modelled as:

$$\mathbf{foF1 = f_s \cos^n \chi} \quad \textbf{(1-8)}$$

where

$$\mathbf{f_s = f_0 \, (f_{100} - f_0) \, R12/100} \quad \textbf{(1-9)}$$

$$\mathbf{f_0 = 4.35 + 0.058 \, |\Psi| - 0.00012 \, \Psi^2} \quad \textbf{(1-10)}$$

$$\mathbf{f_{100} = 5.348 + 0.011 \, |\Psi| - 0.00023 \, \Psi^2} \quad \textbf{(1-11)}$$

$$\mathbf{n = 0.093 + 0.0046 \, |\Psi| - 0.000054 \, \Psi^2 + 0.0003 \, R12} \quad \textbf{(1-12)}$$

The IRI model also provides a criterion for the occurrence probability of the F1 layer. An F1 layer is only assumed to exist when the solar zenith angle is smaller than χ_s, defined as:

$$\mathbf{\chi_s = \chi_0 \, (\chi_{100} - \chi_0) \, R12/100} \quad \textbf{(1-13)}$$

where

$$\mathbf{\chi_0 = 49.85 + 0.35 \, |\Psi|} \quad \textbf{(1-14)}$$

$$\mathbf{\chi_{100} = 38.96 + 0.51 \, |\Psi|} \quad \textbf{(1-15)}$$

Furthermore, the IRI omits the F1 layer at night and in winter. When it has been determined that an F1 layer is probable, the electron density profile is given by *(Ramakrishnan and Rawer, [1972])*:

$$\mathbf{N(h)/NmF2 = \exp(-x^{B_1})/\cosh(x) + C_1 \, ((hmF1 - h)/B_0)^{1/2}} \quad \textbf{(1-16)}$$

where

$$\mathbf{x = (hmF2 - h)/B_0} \quad \textbf{(1-17)}$$

NmF2 and hmF2 are the electron density and height of the F2 peak. The B_1 parameter, which adjusts the shape of the bottomside electron density profile, is mostly taken to be 3. The F1 peak height hmF1 is found as the height at which the bottomside IRI profile reaches the F1 peak density and is affected by the bottomside thickness parameter, B_0. For lower latitudes Gulyaeva's model is used to determine this thickness parameter, B_0 *(Bilitza and Rawer, [1990])*. C_1 is an Epstein function dependant on the modified dip latitude, further details of which can be found in *Bilitza [1990]*. In the

case where no F1 layer is determined probable, the parabolic component of equation 1-16 is ignored when determining the electron density profile.

Although the IRI model is the global standard used when modelling the ionosphere and incorporates data obtained from numerous sources including worldwide ionosonde stations, rocket measurements and satellite data, it still fails to accurately model the F1 layer over South Africa. The LAM model was therefore created for use in the Southern African region.

1.3.2 The LAM Model

The LAM model, a Neural Network (NN) based ionospheric model, was designed to realistically predict the entire bottomside electron density profile over Grahamstown. A NN is a computer program that can be trained with any number of inputs to predict one or more known, measured outputs. The NN learns to identify the relationship between the input vectors and the output. NNs are explained in more detail in Chapter 3. The F1 layer component of the LAM model is explained in detail in this section since this thesis is based on the F1 contribution to this model. The LAM model F1 layer is derived by combining the predictions from three separate NNs as follows:

- A probability NN that predicts the state of the F1 layer at any given day, hour and solar activity. The probability, P(N), P(F) and P(L), a value between 0 and 1, is given for each of the states N, F and L defined on page 4.
- An F1 peak prediction NN that determines the value of the critical frequency of the F1 layer, foF1 and the real height, hmF1, at this frequency when an F1 layer is predicted to be present by the probability NN.
- A profile NN that determines the shape of the layer.

Only the first two model contributions will be dealt with in this thesis.

The probability NN requires instances of L-condition and therefore this NN was trained using only chirp sounder data spanning the period 1973 to 1996. The F1 peak prediction NN was trained with all the available data, both Verti and DPS, from 1973 to 2000. However, since real height information was needed to train the hmF1 prediction NN, only five years of data (1996 to 2000) collected with the DPS system could be used.

In order to predict the parameters of the F1 region profile several steps are implemented. Firstly the F1 probability NN determines the state of the F1 layer, N, L or F as defined at the end of section 1.1. If an F state is predicted as probable the F1 peak NN determines the value of the critical frequency, foF1 and the real height of reflection, hmF1. If, however, an L-condition is concluded as most probable, the L-condition algorithm *(McKinnell, [2002])* is applied. This algorithm ensures that, although no measurement of foF1 would have been possible on an ionogram, an F layer electron density profile is obtained which, if converted into its equivalent ionogram, shows evidence of an L-condition. There are two known patterns of L-condition occurrence:

(i) L-condition occurs for a few hours around sunrise and sunset.

(ii) L-condition occurs throughout daytime hours.

The L-condition algorithm was designed to deal with both of these situations. Once the L-condition has been determined as most probable for a particular input vector, the period of time over which the L-condition extends for that day is determined by working backwards and forwards from the input time in intervals of 0.1 hour until a time is reached for which the L-condition occurrence is no longer probable.

A weighting function (wf) is then calculated as follows:

$$Y = \frac{P(L)}{P(L) + P(F)} \qquad (1\text{-}18)$$

$$wf = \frac{Y_x - Y_{min}}{Y_{max} - Y_{min}} \qquad (1\text{-}19)$$

P(F) and P(L) are outputs predicted from the probability NN. Y is calculated for each of the three hours: min, max and x, where x is the hour of interest. In case (i) above, the hours of min and max are the hours of the start and end of L-condition occurrence. In the case of (ii), the L occurrence period is divided in two, and each half dealt with separately. In this case, the middle point is taken as the maximum value and either end as the minimum.

For the L-condition hour, two profiles are predicted. One as if no F1 layer existed and the other as if a definite F1 layer existed.

Thus for each frequency, two different heights are found: h_N (no F1 layer) and h_F (definite F1 layer). The height for the L-condition hour, h_X, is then calculated as:

$$h_X = h_N + (h_x - h_N) \times wf \qquad (1\text{-}20)$$

Furthermore, the LAM model implements a smoothing algorithm at the F1-F2 boundary to minimize the discontinuity in the electron density profile that arises from the UMLCAR model's use of Chebyshev polynomials to produce the profile. Different Chebyshev polynomials are used to produce the E and F regions and although this Chebyshev polynomial method ensures that the peak height of the F1 layer, hmF1, is equal to the starting height of the F2 layer, hsF2, the slope of the profile, dh/df, tends to infinity at the critical frequency of the F1 layer, foF1. The smoothing algorithm takes the form of a 3rd order polynomial fitted to the electron density profile at two points. More details of this smoothing algorithm are found in *McKinnell [2002]* and *McKinnell and Poole [2004]*.

A full description of the process that the LAM model follows when predicting a profile for a particular set of inputs is shown diagrammatically in figure 1-3. The user inputs to the NN are explained in Chapter 3. GrinTek Ewation has made use of the LAM model in DF systems to improve the accuracy of the ray tracing results when no real-time ionospheric data is available.

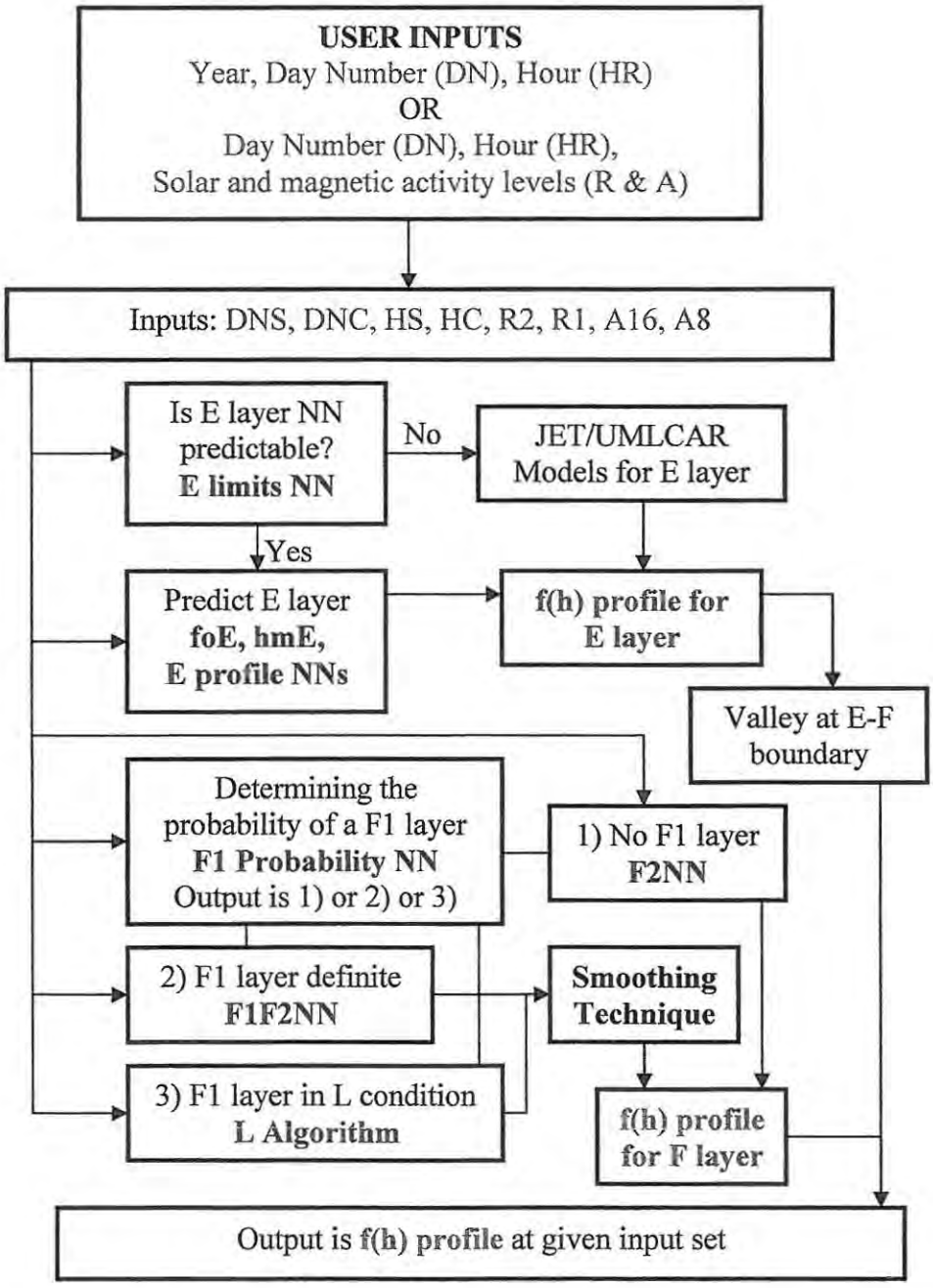


Figure 1-3: A block diagram depicting the process that the LAM model follows when determining a profile. From *McKinnell [2002].*

1.4 Example of an Application

An application for real-time ray tracing through the South African ionosphere was identified, and for this application real-time evaluation of the electron density profile is essential. Raw real-time virtual height data are provided by a Lowell Digisonde (DPS), which employs the automatic scaling software, ARTIST whose output includes the virtual-to-real height data conversion. Ray tracing is a technique employed by direction finding (DF) systems to geographically locate the source of high-frequency (HF) transmissions by determining the ground range between the transmitter and the receiver by following the path that the radio wave takes through the ionosphere from the transmitter to the receiver. Conventional HFDF systems rely on line of bearing (LOB) observations from two or more sites to triangulate the position of the transmitter site. However, single station location (SSL) HFDF systems are interferometric systems whereby the direction of arrival of the incoming wave is deduced from the phase measurements made on a number of spaced antennas. From these phase measurements the horizontal (azimuth) and vertical (elevation) components of the angle of arrival can be obtained. By using a larger aperture for the receiving antenna array accuracy can be greatly increased. Range estimates can then be obtained using the elevation angle and the radio frequency of the transmitted signal in conjunction with an ionospheric model. The method requires either numerical or analytical integration (used with the MQP model described in Chapter 1.3) depending on the model used. Ray tracing is SSL in reverse, since the path is followed from the receiver back to the transmitter *(McNamara, [1991])*.

The process of ray tracing relies on the fact that radio waves are reflected by the ionosphere, and the path that these radio waves follow can be determined with knowledge of the elevation angle and the radio frequency of the signal,

and a description of the ionosphere in the form of an electron density profile or suitable ionospheric parameters. Therefore an intimate knowledge of the behaviour of the ionosphere in determining the path that the radio waves follow is paramount. An application of the use of ray tracing in DF systems is given in *Coetzee [2004]*. It is obvious that for this DF technique accurately scaled ionospheric parameters are essential for accurate ray tracing. GrinTek Ewation has reported that the DF method gives ground ranges that can be inaccurate by up to 30%. The cause for these errors is thought to be inaccurate F1 layer scaling by ARTIST as it has been shown *(McKinnell, [2002])* that small changes in the value of foF1 can produce large deviations in the electron density profile, which in turn produce large variations in the ground range determination. It is for this reason that an assessment of the accuracy of the automatically scaled ionograms employed in calculating the ground range was needed. However, the ray tracing application is not covered by the work in this thesis.

1.5 Contributions of this Thesis

In developing the LAM model it was felt that ARTIST data in the F1 region was inadequate, and therefore in order to check for accuracy the available DPS data from 1996 to 2003 was manually re-scaled. Instances of L-condition status were also recorded as ARTIST fails to do so. An analysis of the ARTIST data is presented in the next chapter. The re-scaled F1 layer data was incorporated into the LAM model in an attempt to improve and update the F1 layer predictions. Details and results from this effort appear in Chapters 3 and 4.

The aim of this thesis is to present an assessment on the F1 layer over Grahamstown such that future modelling attempts can include remedial steps to address known issues.

Chapter 2
DATA ANALYSIS

2.1 Introduction

This chapter describes an analysis of automatically scaled F1 layer data recorded over Grahamstown, South Africa (33.3°S, 26.5°E). The data from this station is automatically scaled using the ARTIST scaling software. Poor real-time ray tracing results have indicated possible shortcomings in the ability of ARTIST to characterise the F1 region. Experience with modelling the F1 region over Grahamstown has also shown that there are times when ARTIST produces inaccurate F1 characterisation. The purpose of this investigation is to establish the extent of the problem, and the times and conditions under which most errors occur, with a view to formulating remedial alternative strategies, such as predictive modelling.

Furthermore, the re-scaled data is required in order to update an existing Neural Network based F1 layer model, for which instances of L-condition are essential, and is explained in Chapter 3. Results from the analysis presented in this chapter appear in *Jacobs et al. [2004]*.

2.2 ARTIST

ARTIST, software developed by UMLCAR for use with DPS systems, scales the ionogram and extracts the vertical ordinary echo trace from which the electron density profile is calculated. This inversion technique is the basis for the UMLCAR model, and determines ionospheric parameters including foF1 and hmF1 describing the peak of the F1 layer. For autoscaling to be successful, amplitude information as well as polarization flagging, and the distinction between vertical and off-vertical echoes are

needed, and are provided by the DPS system. The large variations in height and shape of the F region traces, as well as the occurrence of multiple echoes and spread F, make it impossible to use a simple curve-fitting procedure for the entire F trace. Therefore the F trace is determined by means of a frequency-by-frequency search process *(Reinisch and Huang, [1983])*. The frequency-by-frequency search process defines a rough trace close to the leading edge of the pulse and is able to successfully trace the F1-F2 transitions, but also produces some undesirable virtual height discontinuities when man-made emissions interfere with the echo pulses at certain frequencies. Careful smoothing is applied, which maintains the cusp at the F1-F2 transition.

2.3 The Analysis

At any time, the F1 layer can be assigned one of three states: N, when no F1 layer is present; F, when an F1 layer is present and the critical frequency foF1, can be determined; and L, when the F1 layer is present but is shown as a ledge rather than a cusp on the ionogram and therefore foF1 cannot be determined. The L-condition, which occurs frequently over Grahamstown, exists in a transition state as the F1 region moves between the N state during nighttime hours and the F state during the day. ARTIST makes no allowance for the L-condition state even though there is evidence of significant departures in the real height profile from the no-F1-layer case.

In order to perform the analysis, all DPS ionograms had to be manually re-scaled and the conditions under which ARTIST might be unreliable assessed. The ionograms were generated from ARTIST SAO format files *(Galkin, [1998])*, which were produced from the original raw data files. SAO format is the international standard format for recording ionospheric data.

Ionograms generated from eight years of Grahamstown DPS data (1996 to 2003) were examined and value of the critical frequency of the F1 layer, foF1, as well as instances of L-condition, were recorded.

Five different types of errors in ARTIST's ability to scale the F1 layer were identified, as follows:

1. ARTIST reports a value for foF1 when an L-condition is present (figure 2-1).
2. ARTIST determines that no F1 layer is present when an L-condition is present (figure 2-2).
3. When no F1 layer is present ARTIST reports a value for foF1 (figure 2-3).
4. When the F state is present ARTIST fails to identify foF1 and regards the F1 layer as being in the N state (figure 2-4).
5. ARTIST correctly identifies an F state but reports an incorrect value of foF1 (figure 2-5). For all of the ionograms that fell into this category the difference between the automatically scaled foF1 value and the manually scaled foF1 value was greater than 0.5 MHz.

The first two errors are the most common as they occur during times when the F1 layer is in the L state and it is well known that ARTIST does not record instances of L-condition. Errors of type (5) above are mostly caused by "noise" on the ionogram, which hinders the frequency-by-frequency search process implemented by ARTIST. Also, if a spurious cusp of foF1 at a lower frequency is present, ARTIST will generally identify this cusp as being foF1. Furthermore, as shown in figure 2-5, ARTIST will also occasionally identify foE as being foF1. In the type (5) error category, ARTIST underestimates the value of foF1 93% of the time.

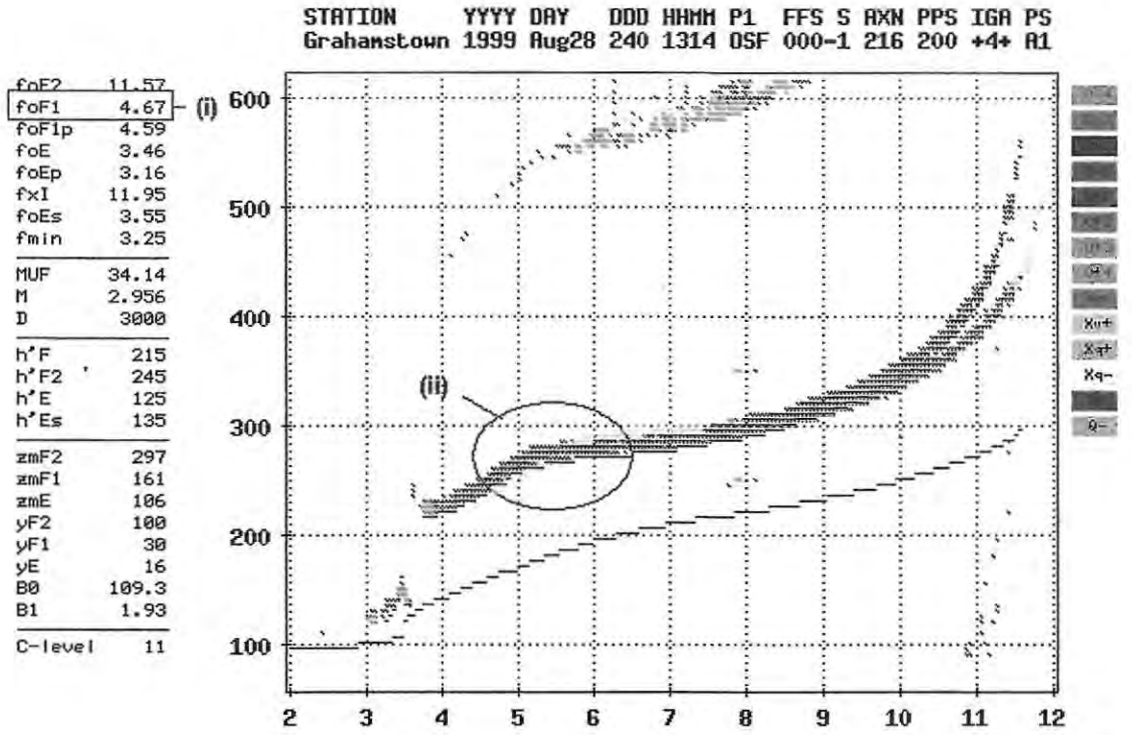


Figure 2-1: An ionogram showing the F1 layer being in the L-condition state. (i) ARTIST identifies a value for foF1 even though (ii) a ledge rather than a cusp is seen and the critical frequency of the F1 layer cannot be determined.

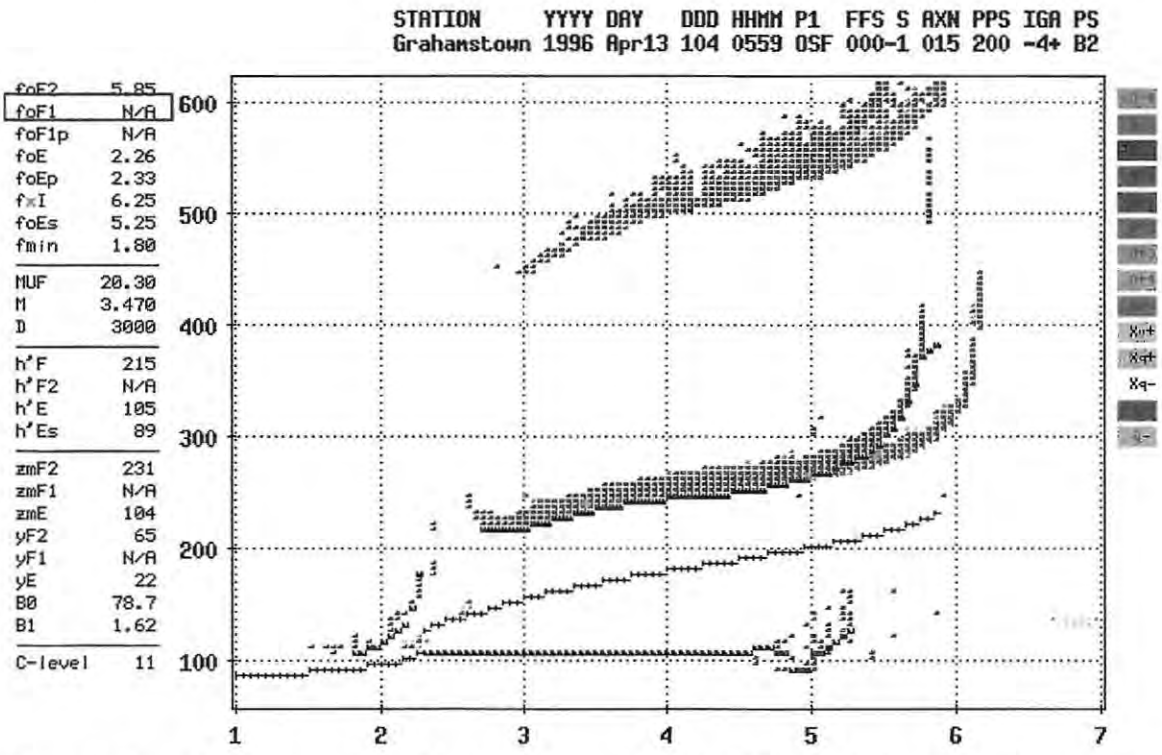


Figure 2-2: As in figure 2-1, this ionogram shows the F1 layer being in the L-condition state. ARTIST, correctly, finds no value of foF1, indicated by the red box on the ionogram.

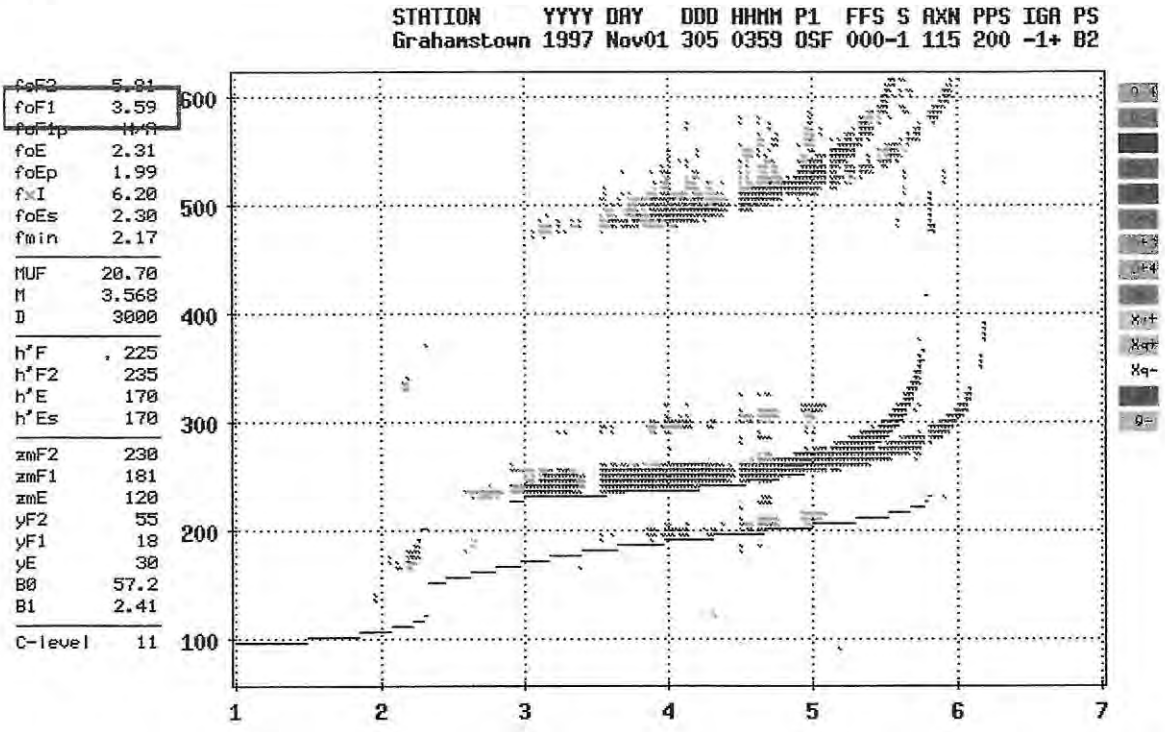


Figure 2-3: An ionogram illustrating an instance when no F1 layer is present yet ARTIST presents a value for foF1, indicated by the red box on the ionogram.

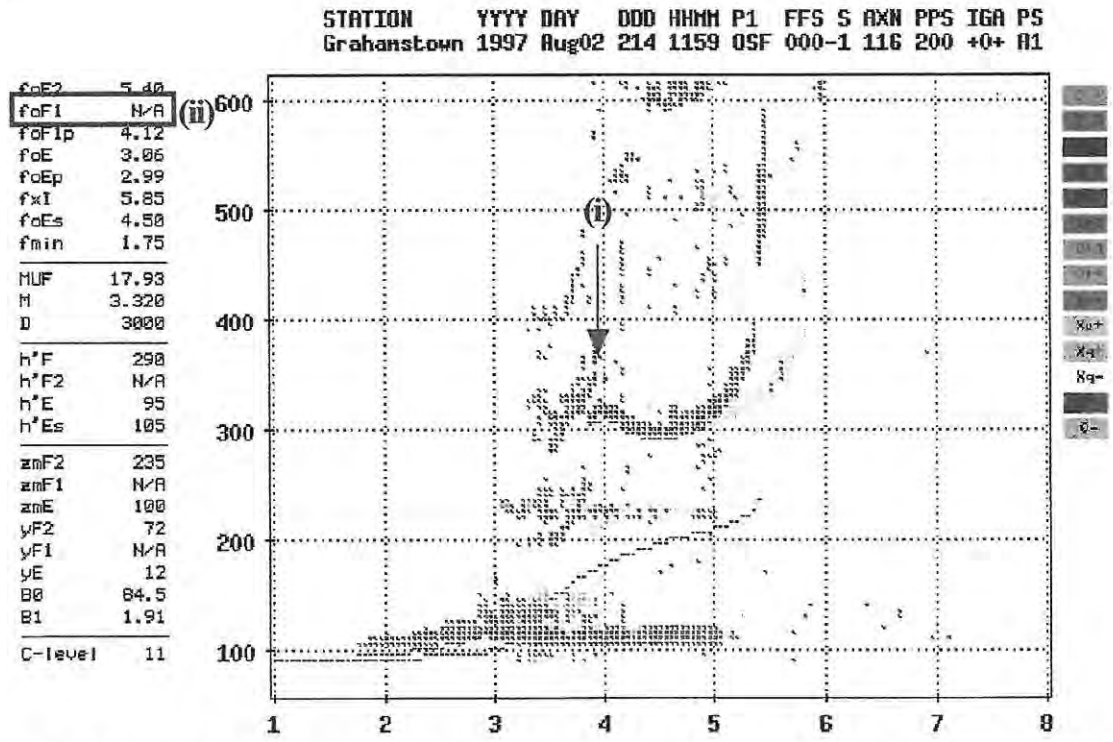


Figure 2-4: This ionogram shows the F1 layer to be in the F state with (i) a definite cusp being present at 3.9 MHz. However (ii) ARTIST fails to identify the value of foF1 and reports that no F1 layer is present.

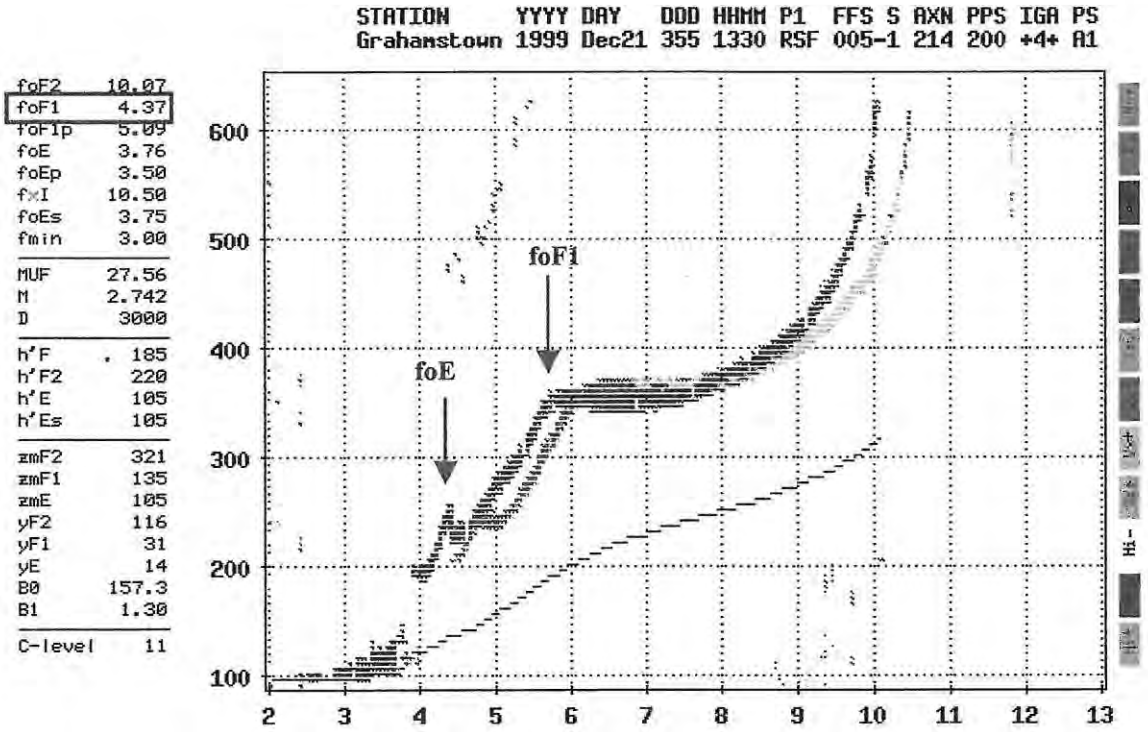


Figure 2-5: An ionogram showing the F1 layer in the F state. The correct value of foF1 should be 5.8 MHz as indicated on the ionogram. However, ARTIST finds a value of 4.37 MHz, highlighted by the red box on the ionogram. This value is due to ARTIST incorrectly identifying foE as being foF1.

Figure 2-6 shows the distribution of the ARTIST values and the manually re-scaled values of type (5) above at 12h00 SAST spanning the period 1996 to 2003. From the graph it can be seen that there is an abnormally high occurrence of type (5) error around the start of the year 2002. This could be due to the fact that during this period the solar activity was particularly high, resulting in the formation of spurious cusps of foF1 at lower frequencies.

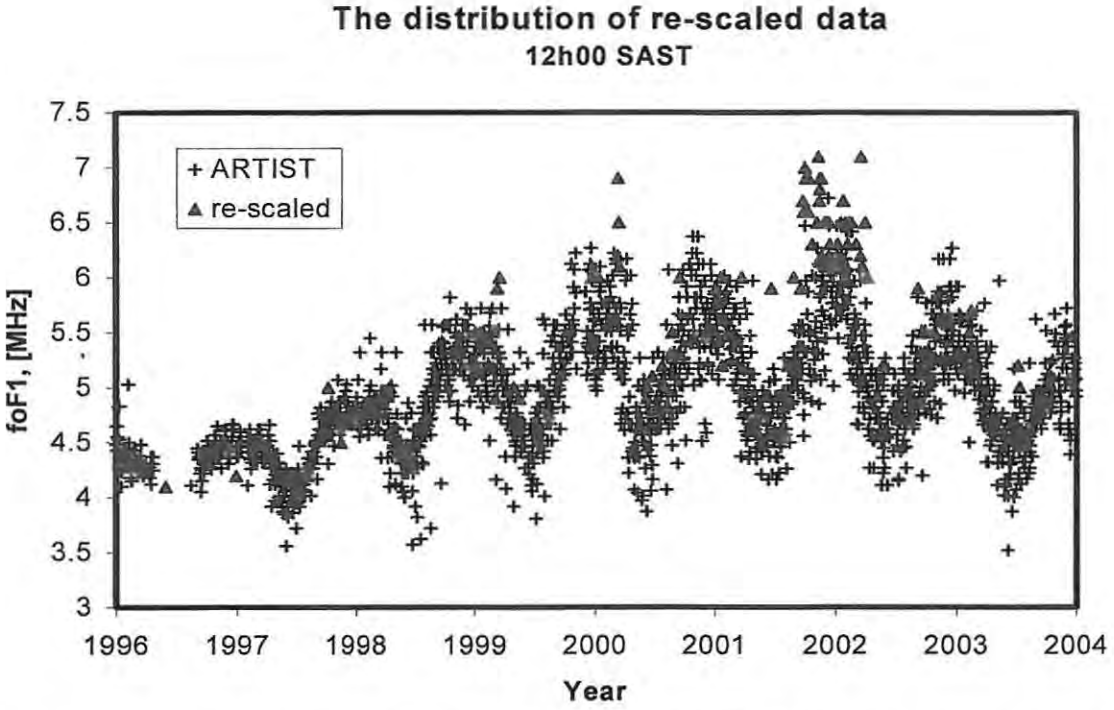


Figure 2-6: The distribution of the 12h00 SAST ARTIST foF1 values and the manually re-scaled values of type (5) error are shown.

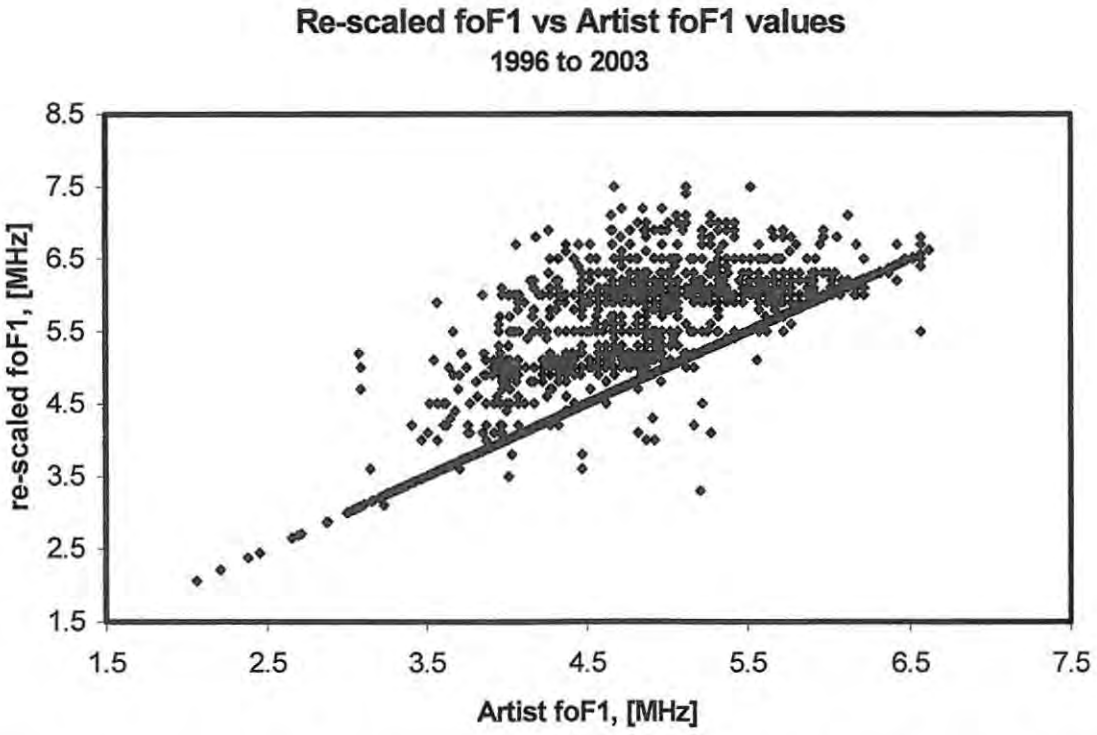


Figure 2-7: A graph of the manually re-scaled foF1 values of type (5) error versus ARTIST foF1 values for the period 1996 to 2003.

Figure 2-7 shows the comparison between ARTIST and the manually re-scaled foF1 values over the years 1996 to 2003. Only type (5) error is shown, where both ARTIST and the re-scaled data possess a value for foF1. From the graph it can be seen that ARTIST frequently underestimates the value of foF1.

As a check to determine if the re-scaling process was performed correctly, the percentage of L-conditions assigned to each hour was compared with previous manually scaled data from 1973 to 1996 obtained at the Grahamstown station. The L state was chosen for this comparison as, firstly, there is a particularly high incidence of L-condition occurrence at the latitude of Grahamstown, and secondly since the L-condition is sometimes difficult to identify. Since the state of the F1 layer is influenced by the season, solar activity and the time of day, the two data sets were compared over summer (November – January) and winter (May – July) months, and at both high (125 - 135) and low (10 - 20) values of R2, which is a two-month running mean of the daily sunspot number.

Figure 2-8 shows that the two data sets demonstrate very similar patterns of L-condition occurrence, although at low R2 the re-scaled ARTIST data detects a higher incidence of the L state. This is most probably due to there being fewer records at low R2 in the period 1996 to 2003 than at high R2. This disparity between high and low R2 continues to present problems in the F1 layer modelling attempts, and will be further discussed in Chapter 3.

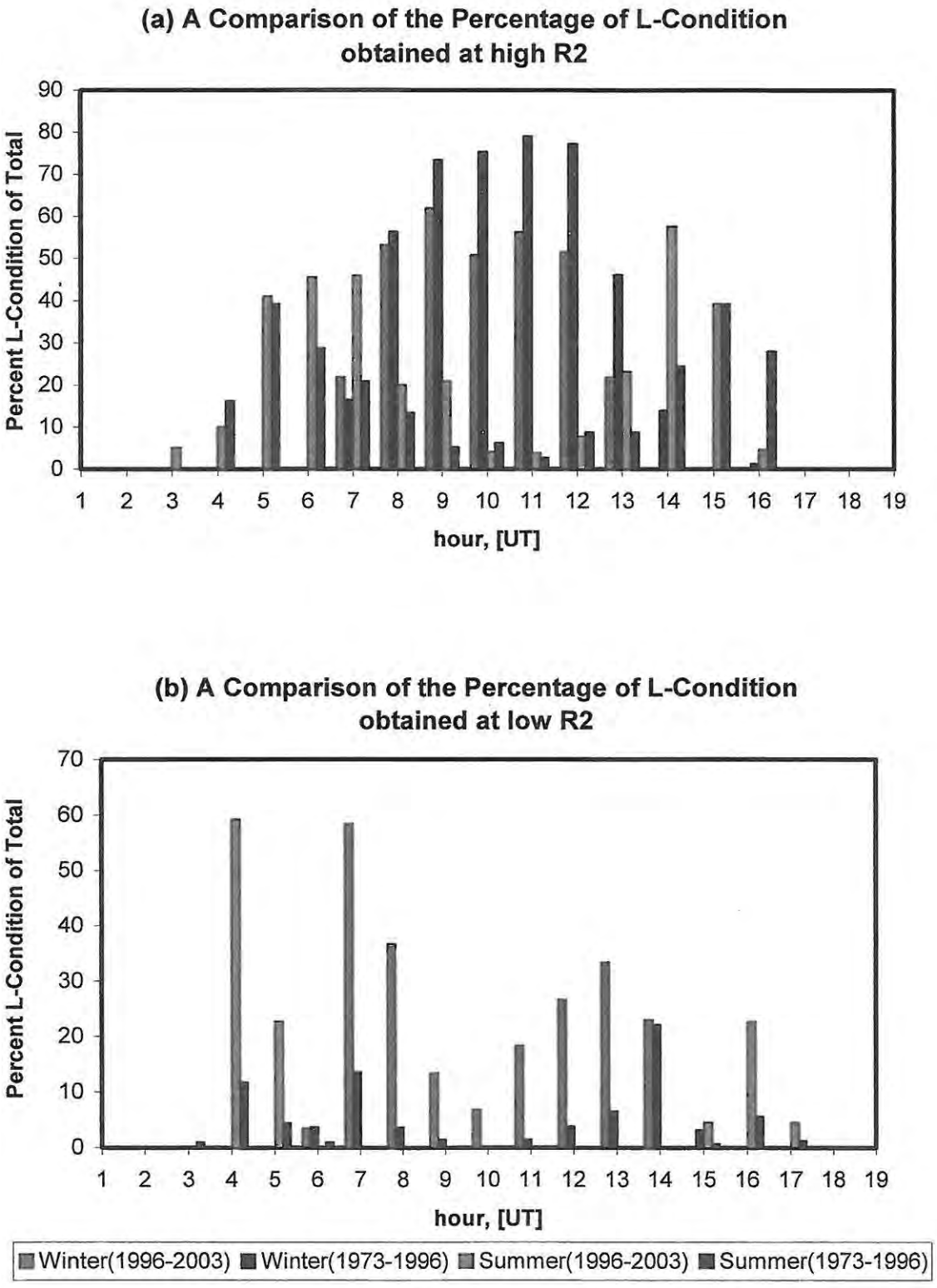


Figure 2-8: A comparison of the relative percentage of L-condition between manually re-scaled ARTIST ionograms (1996-2003) and previous manually scaled ionograms (1973-1996) with hour and season at (a) high R2 and (b) low R2 values.

2.4 Results

Having taken into account all of the above-mentioned errors it was found that, on average, ARTIST incorrectly scales foF1 29% of the time between the hours of 0330 and 1700 UT. The percentage of incorrectly scaled ionograms varies with hour and month and is illustrated in figure 2-9.

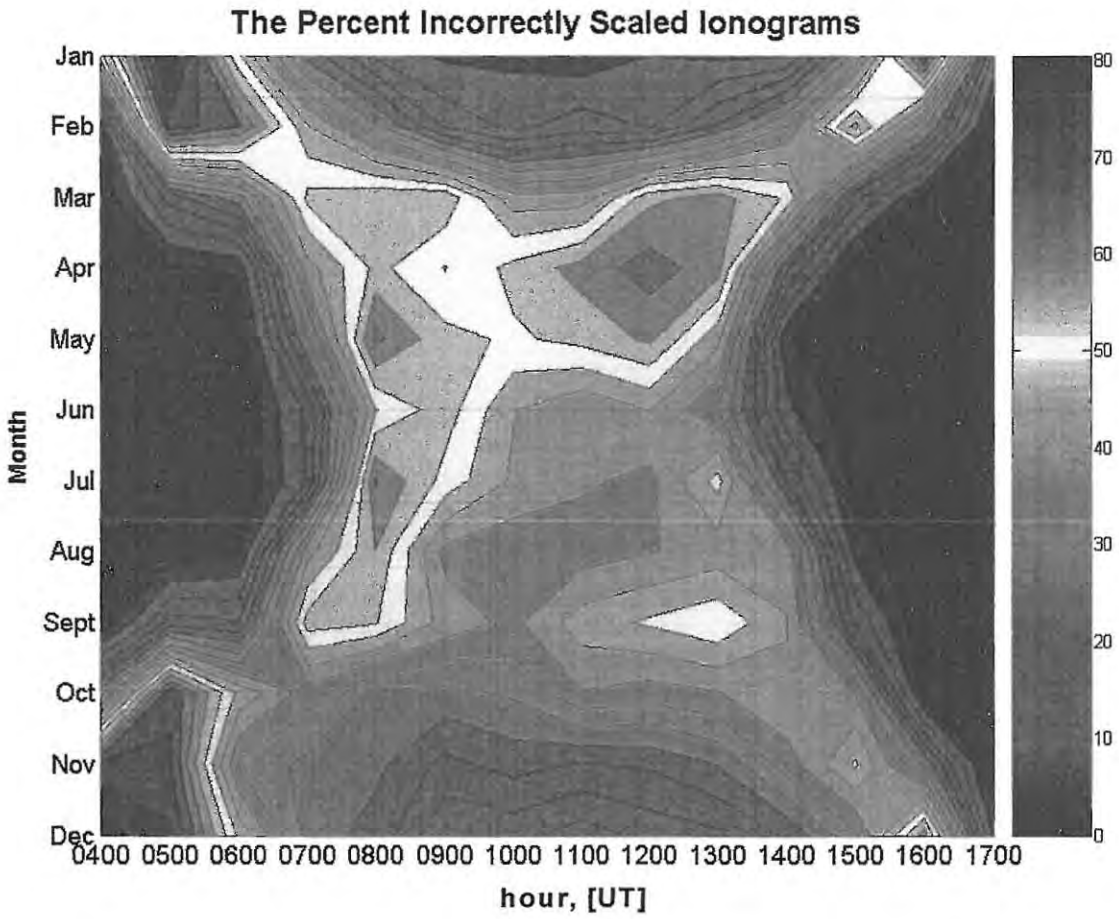


Figure 2-9: A contour map showing the number of incorrectly scaled ionograms as a percentage of the total ionograms examined, varying with hour and month.

From the contour map it can be seen that the highest incidences of incorrectly scaled ionograms correspond to those hours and months where L-condition status is expected to most likely occur, i.e. in summer during sunrise and sunset and in winter during daylight hours.

Figure 2-10 shows an analysis of the various types of inaccuracies encountered. The percentages shown are of the total number of re-scaled ionograms and not of the total number of ionograms examined. Although ARTIST was designed to disregard instances of L-condition, such instances have been included for the purpose of evaluating the performance of ARTIST over Grahamstown, where the incidence of the F1 layer being in an L-condition state is high.

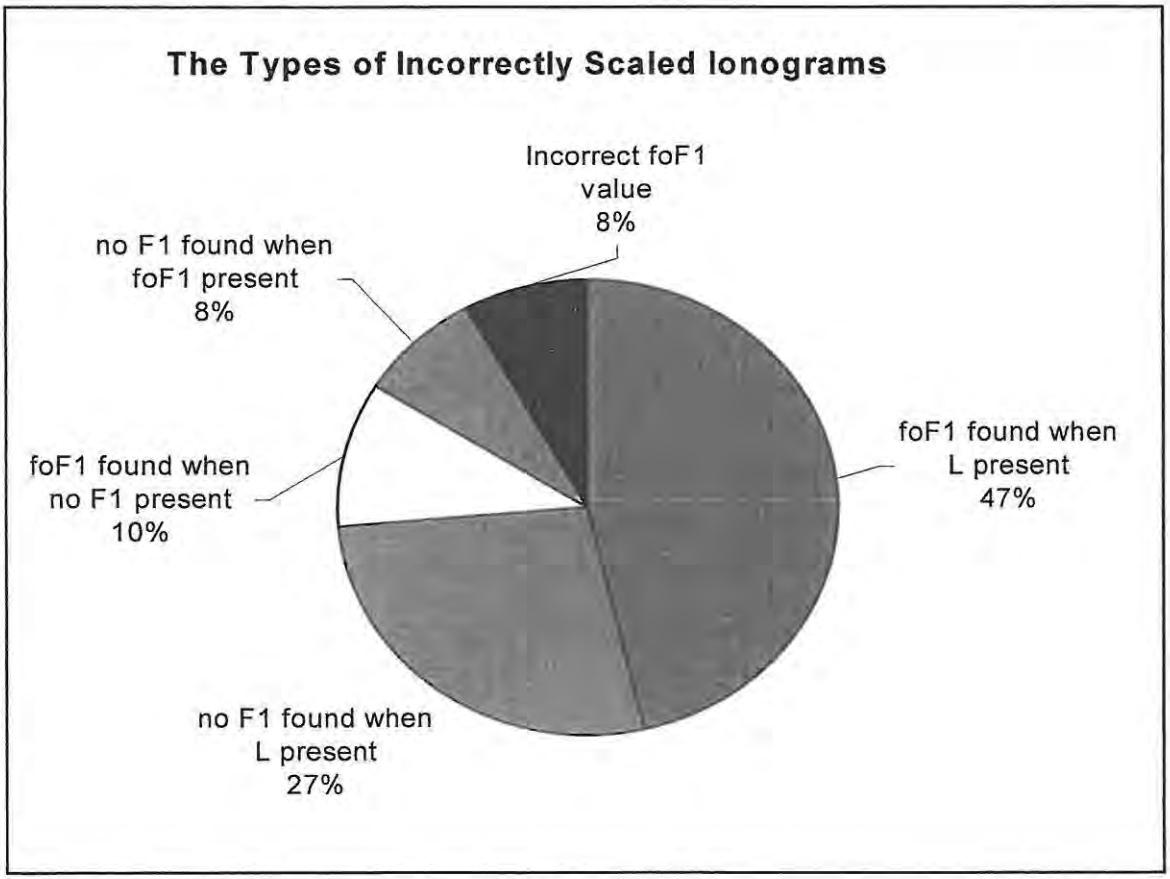


Figure 2-10: An analysis of the types of incorrectly scaled ionograms.

Figure 2-11 shows a time distribution of incorrectly scaled ionograms as a percentage of the total number of ionograms examined. It was noted that ARTIST fails at times to differentiate between true foF1 and spurious cusps of foF1 and therefore often underestimates the value. Sometimes ARTIST would identify foE as being foF1, particularly when either no F1 layer is present, or when an L-condition is present.

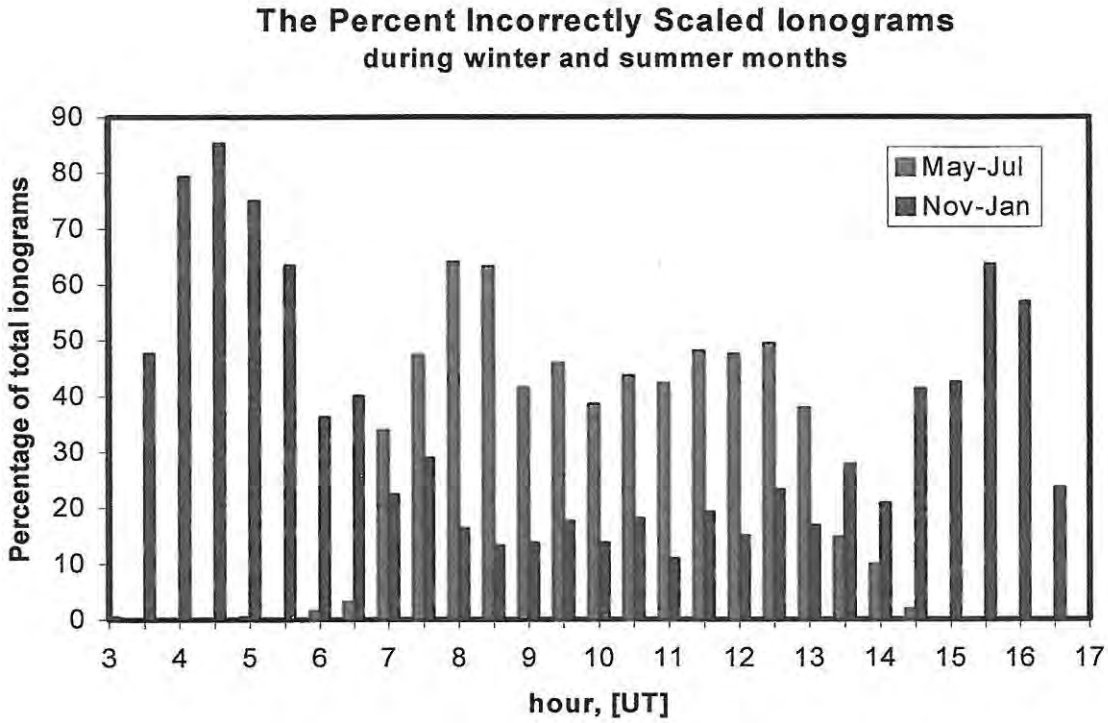


Figure 2-11: The time distribution of incorrectly scaled ionograms during summer and winter months.

Figure 2-12 shows the time distribution of three types of errors ARTIST encounters as a percentage of the total number of corrected ionograms. The L-condition errors have been excluded from this graph but are shown in figure 2-13, which depicts the time distribution of the L-condition errors as a percentage of the total number of re-scaled ionograms.

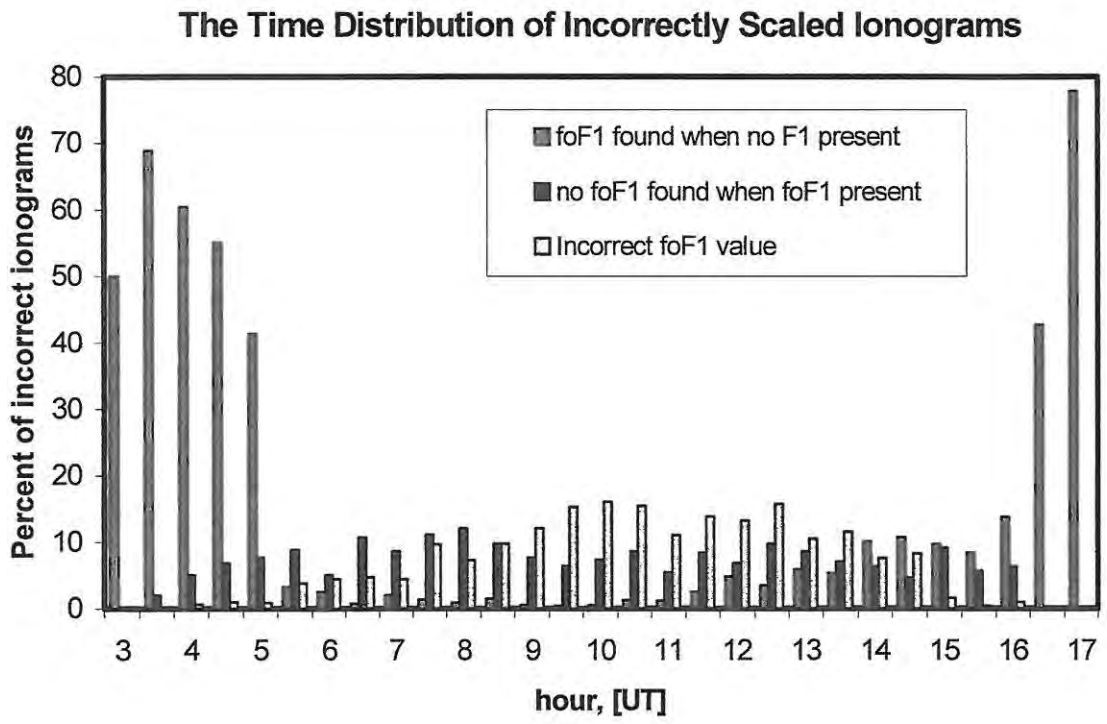


Figure 2-12: The time distribution of three types of errors encountered by ARTIST.

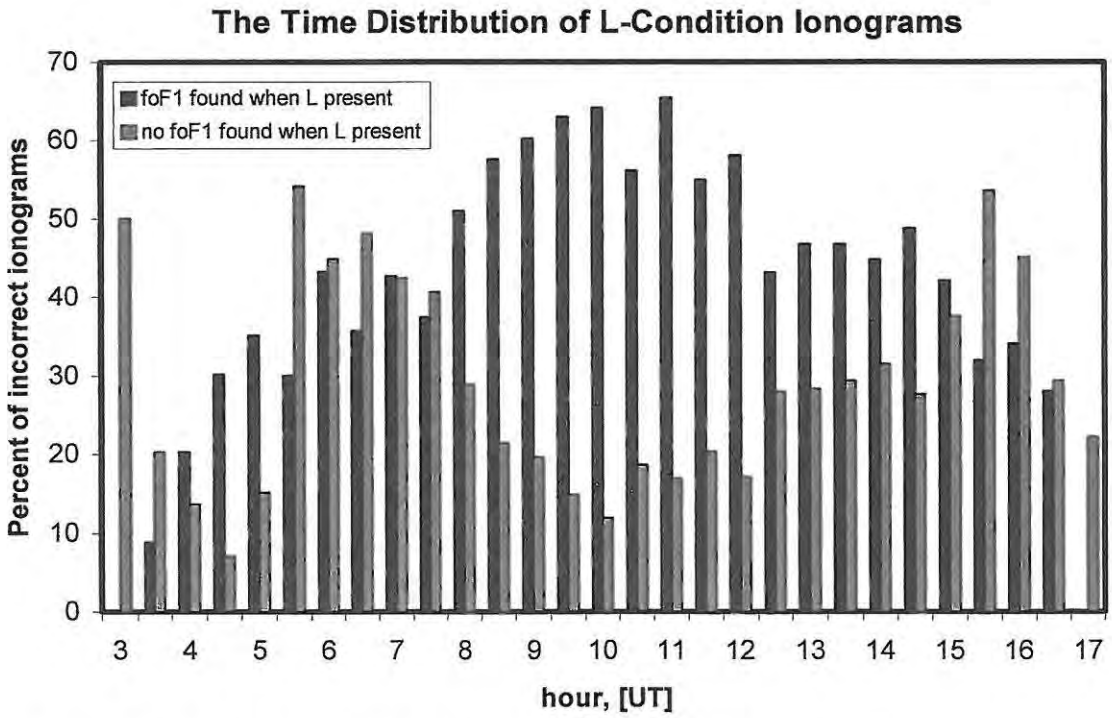


Figure 2-13: The time distribution of L-condition re-scaled ionograms.

2.5 Conclusion

An analysis of the automatically scaled F1 layer data for the Grahamstown, South Africa station has been completed. From the results of this analysis it is evident that a total reliance on the ARTIST software to successfully characterise the real height profile in the F1 region for ray tracing purposes would be impractical. However, for much of the time, the real-time profile is better than a model profile based on averaged historical data. It has therefore been proposed that the best strategy is a judicious mix of real-time ARTIST-derived profiles with back-up model profiles during times when ARTIST is likely to be incorrect. The above analysis will be very relevant to this strategy.

Chapter 3

NEURAL NETWORKS AND THE INPUTS

3.1 Training Neural Networks

Neural Networks (NNs) have for many years been used as a tool to model the behaviour of the ionosphere by predicting the different parameters that characterize the various regions of the ionosphere. A NN can be described as an information processing system *(Fausett, [1994])* that has certain characteristics in common with biological neural networks, and to some extent imitates the parallel processing paradigm of the human brain. The "nodes" which are described below fulfil the role of neurons in this software brain. The network architecture shown in figure 3-1 was developed using the cascade learning architecture *(Haykin, [1994])*, and consists of a set of units that constitute an input layer, one or more hidden layers of computational nodes, and an output layer. The connections between each node represent a feeding of output from one unit to the input of the other, multiplied by some weight, represented in the diagram by black dots. The nodes in the hidden layer have two functions:

(i) they perform the arithmetic sum of their respective weighted inputs,

(ii) they transpose this sum to a value in the range ±1 via a non-linear, continuous and differentiable sigmoid function.

The output y is then scaled back to match the input data.

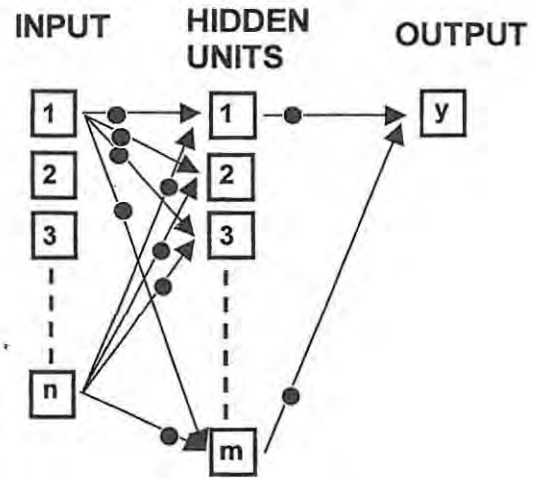


Figure 3-1: A diagram depicting the architecture of a NN. The black dots represent the weights that apply to each connection.

Training is an iterative process that starts with randomly assigned weights at each node. Before training begins, the dataset is firstly normalized so that the values of each of the input parameters and the output parameter lie in the range –1 to +1. The dataset is then split randomly into a training set (70%) and a testing set (30%). The process of training involves altering the weights to each node by an algorithm (backpropagation) in such a fashion as to minimize the difference, δ_i, between the desired output value and the output value calculated by the NN for each of the training vectors. The difference is calculated for each training vector, and after every epoch the NN program produces a Mean Square Error (MSE), as:

$$\mathbf{MSE} = \sum_{i=1}^{N} \delta_i^2 / N \qquad (3\text{-}1)$$

This method of training is termed feed-forward backpropagation. An epoch is one complete cycle through which every training input vector is presented to the NN. After a specified number of epochs the testing set is presented to the NN and the output for each vector determined and the MSE calculated. This is done in order to check that the NN is not over-training. Over-training

occurs when the NN either identifies a pattern based on the order of the input vectors (which is minimized if the training data is randomised) or else learns a relationship specific only to the training input set. Over-training is identified when the testing MSE values increase but the training MSE values continue to decrease.

The only way to determine the optimum architecture of the NN for a particular dataset is to add hidden nodes one at a time and alter the number of epochs until the MSE values are minimized. Version 4.2 of the Stuttgart Neural Network Simulator (SNNS) was the NN software package used for training the NNs required for this project. More details on the methods of training NNs are given in *SNNS [1995a]*. SNNS *(SNNS, [1995b])* was developed by the University of Stuttgart Institute for Parallel and Distributed High Performance Systems.

One of the most important requirements for training NNs is access to a large database. For ionospheric modelling the dataset should ideally span at least 22 years, being 2 solar cycles. For the purposes of this thesis, five sets of 25 NNs were trained and are described below:

1. An F1 Probability NN (Probability NN1) that determines the probability of the F1 layer being in any one of the three states, N, F or L at any time, including all the available Grahamstown data from 1973 to 2003.
2. An F1 Probability NN (Probability NN2) trained using only re-scaled DPS data. This dataset only includes the eight years of data from 1996 to 2003.
3. An foF1 Prediction NN (foF1 NN1) that predicts the value of the critical frequency of the F1 layer, using the full 30 years of data as in (1) above.

4. An foF1 Prediction NN (foF1 NN2) using the eight years (1996 to 2003) of manually re-scaled DPS data.
5. An hmF1 Prediction NN that predicts the value of the real height of reflection at foF1 using only DPS data as the scaled Verti data does not provide real height information.

25 NNs were trained for each of the five requirements described above since each NN has random starting weights and therefore the end result is unique. By training 25 NNs with the same architecture an average value for each weight can be used and thereby produce a mean output value for each input vector, thus minimizing statistical variability.

The reason behind training the probability NNs and the foF1 value prediction NN using both the data sets spanning the years 1973 to 2003 and 1996 to 2003 was to establish to what extent a NN trained with the abbreviated dataset could reproduce the results of one trained with the full dataset. This could be used to assess the validity of the NN described in (5) above, for which no data prior to 1996 is available.

3.2 The Inputs

The electron density in the ionosphere exhibits both diurnal and seasonal variation. It can also be influenced by changes in solar and magnetic activity *(McNamara and Reinisch, [1995])*. In order to train the NN, input parameters that represent these influencing factors on the electron density in the F1 region need to be determined. As a measure of diurnal variation, the hour of the day (HR), given in Universal Time [UT], is used. The value of the hour variable varies from 0 to 23.5 in steps of 0.5. The seasonal variation can be represented by the day number (DN), which varies from 1 to 365. As explained in *Poole and McKinnell [2000]* the DN and HR are split into their

sine and cosine components in order to more accurately denote the cyclic nature of these two input parameters. The resulting four inputs are calculated as follows:

$$DNS = \sin\left(\frac{2\pi \times DN}{365}\right) \quad (3\text{-}2)$$

$$DNC = \cos\left(\frac{2\pi \times DN}{365}\right) \quad (3\text{-}3)$$

$$HS = \sin\left(\frac{2\pi \times HR}{24}\right) \quad (3\text{-}4)$$

$$HC = \cos\left(\frac{2\pi \times HR}{24}\right) \quad (3\text{-}5)$$

As a measure of solar activity, a two month running mean value of the daily sunspot number, R2, was used. This R2 value has been proven *(Williscroft and Poole [1996], McKinnell [1996])* to be the optimum solar activity input for the purposes of predicting the peak electron density of the F1 layer. Figure 3-2 depicts the variation of the value of R2 over the two time periods that made up the different training sets. From figure 3-2 it can be seen that the period 1996 to 2003 constitutes a rise from solar minimum to solar maximum and the majority of the data for this period lies at solar maximum. Therefore the NNs that were trained using data from this period perform well at high values of R2, but at low values of R2 the performance is slightly impaired.

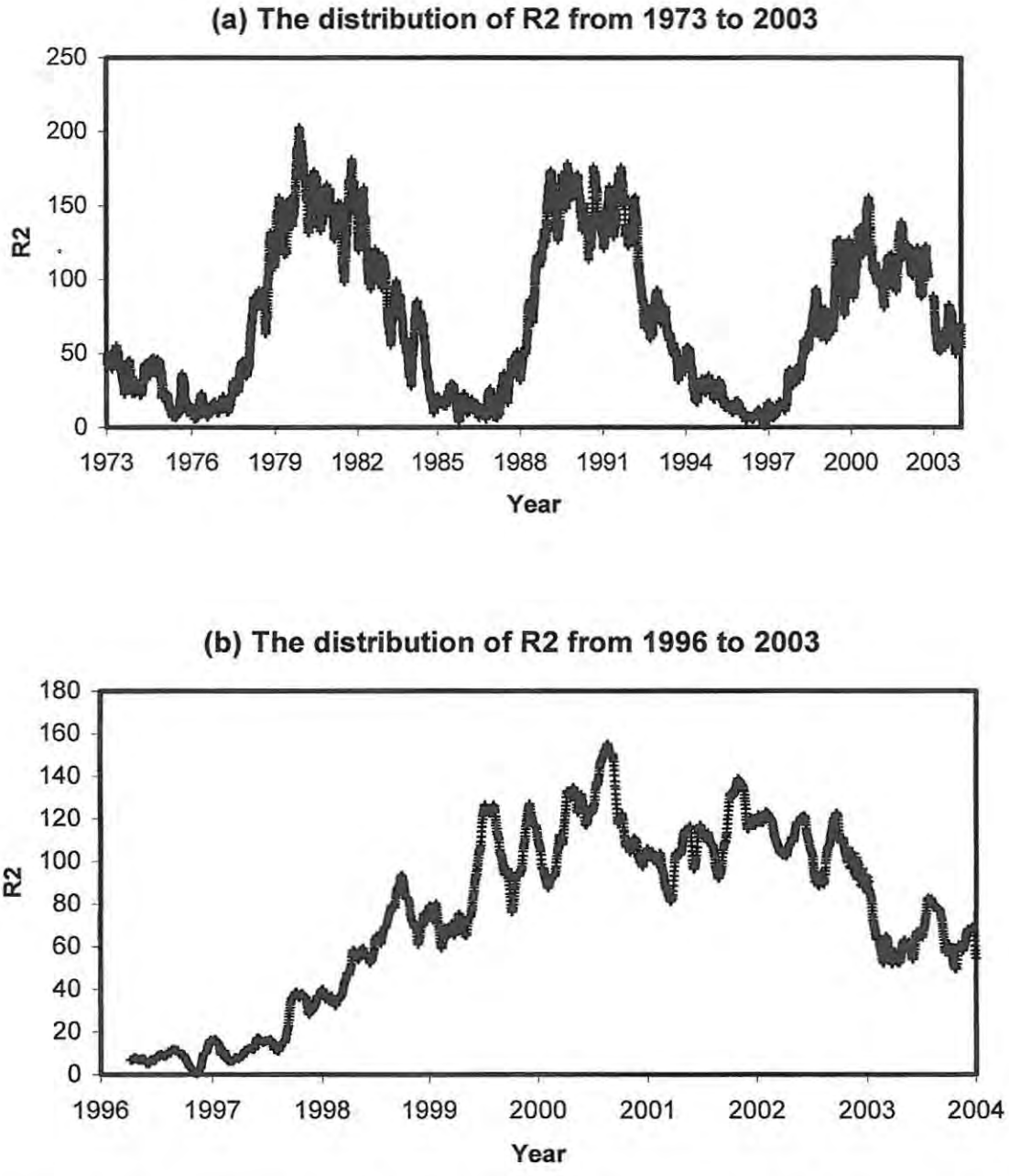


Figure 3-2: A graph showing the variation of the value of R2 over the period (a) 1973 to 2003 and (b) April 1996 to 2003.

As a measure of magnetic activity, A8, a 24-hour running mean of the 3-hourly magnetic index, a_k, has been shown *(McKinnell, [2002])* to be optimal for predicting the peak electron density in the F1 region. The magnetic data was obtained from the Hermanus Magnetic Observatory (HMO). Unlike the value of R2, A8 varies hourly and therefore during the period 1996 to 2003 there is an adequate distribution of both high and low values, as shown in figure 3-3.

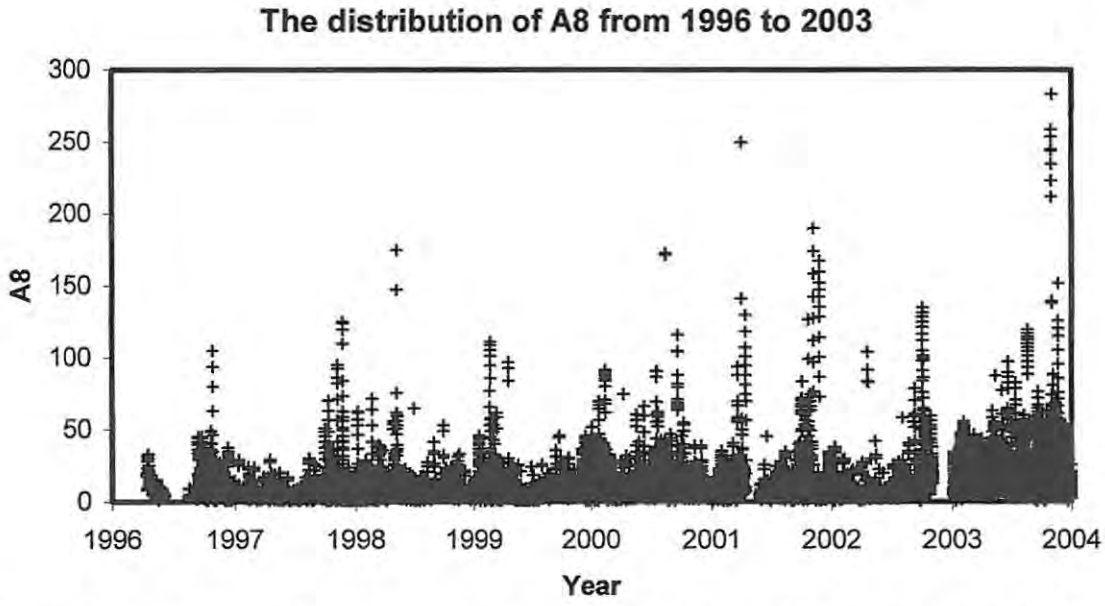


Figure 3-3: A graph showing the distribution of the value of A8 over the period 1996 to 2003.

3.3 Testing the NNs

In order to interrogate the performance of the NNs, artificial data is presented as the inputs, and the NN predicts output values for the given inputs. The NNs were interrogated over days corresponding to autumn equinox (DN = 81), winter solstice (DN = 173), spring equinox (DN = 265) and summer solstice (DN = 356), as well as with values representing both high and low values of R2 and A8.

The low and high values of R2 and A8 were chosen so that 70% of the data lies symmetrically between the values.

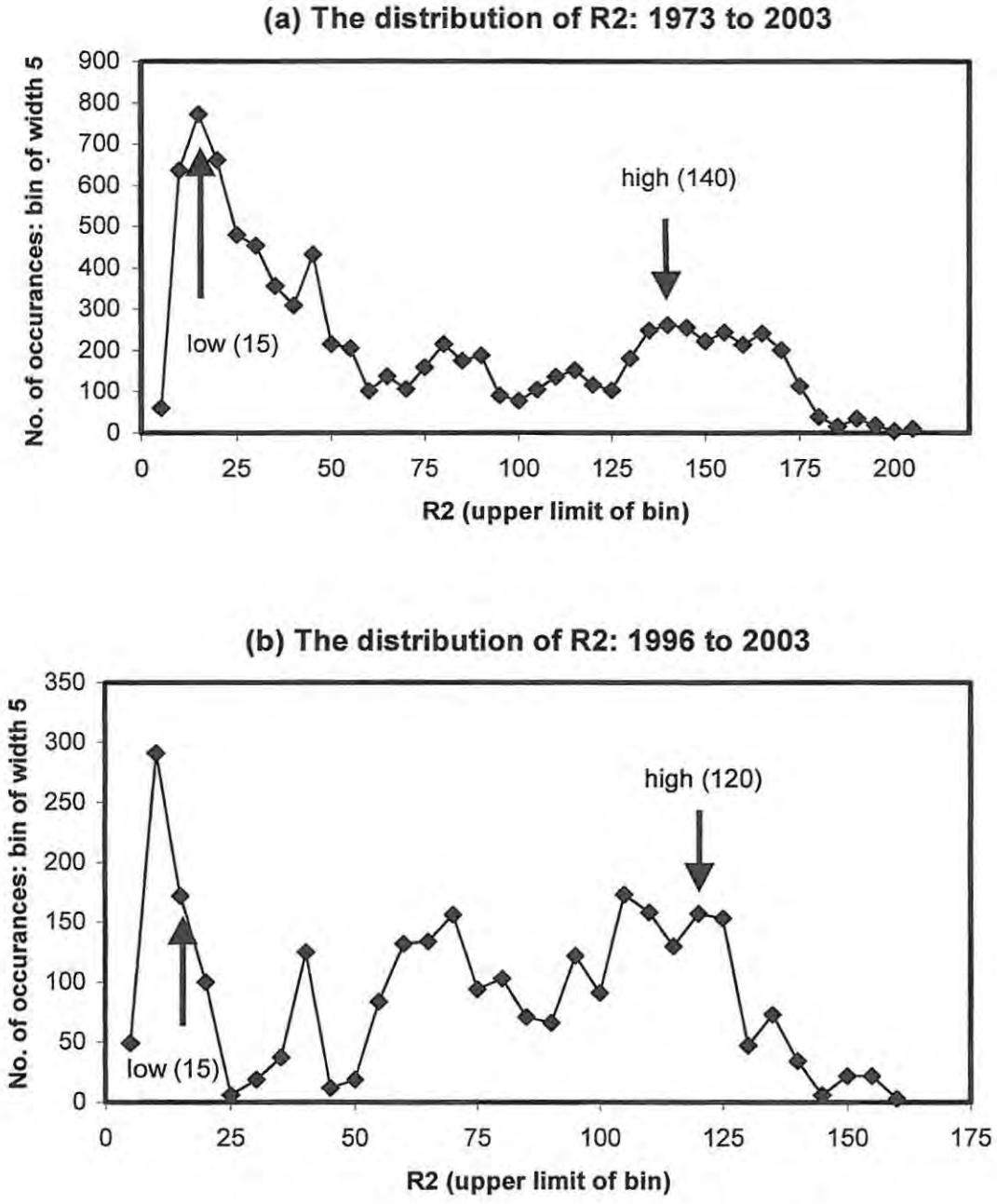


Figure 3-4: Graphs illustrating the values chosen to represent the low and high values of R2 for the period (a) 1973 to 2003 and (b) 1996 to 2003.

Figures 3-4 and 3-5 are the graphs produced (based on an histogram analysis) to determine the distribution of the data, illustrating the high and low values of R2 and A8 respectively that were used to test the various NNs.

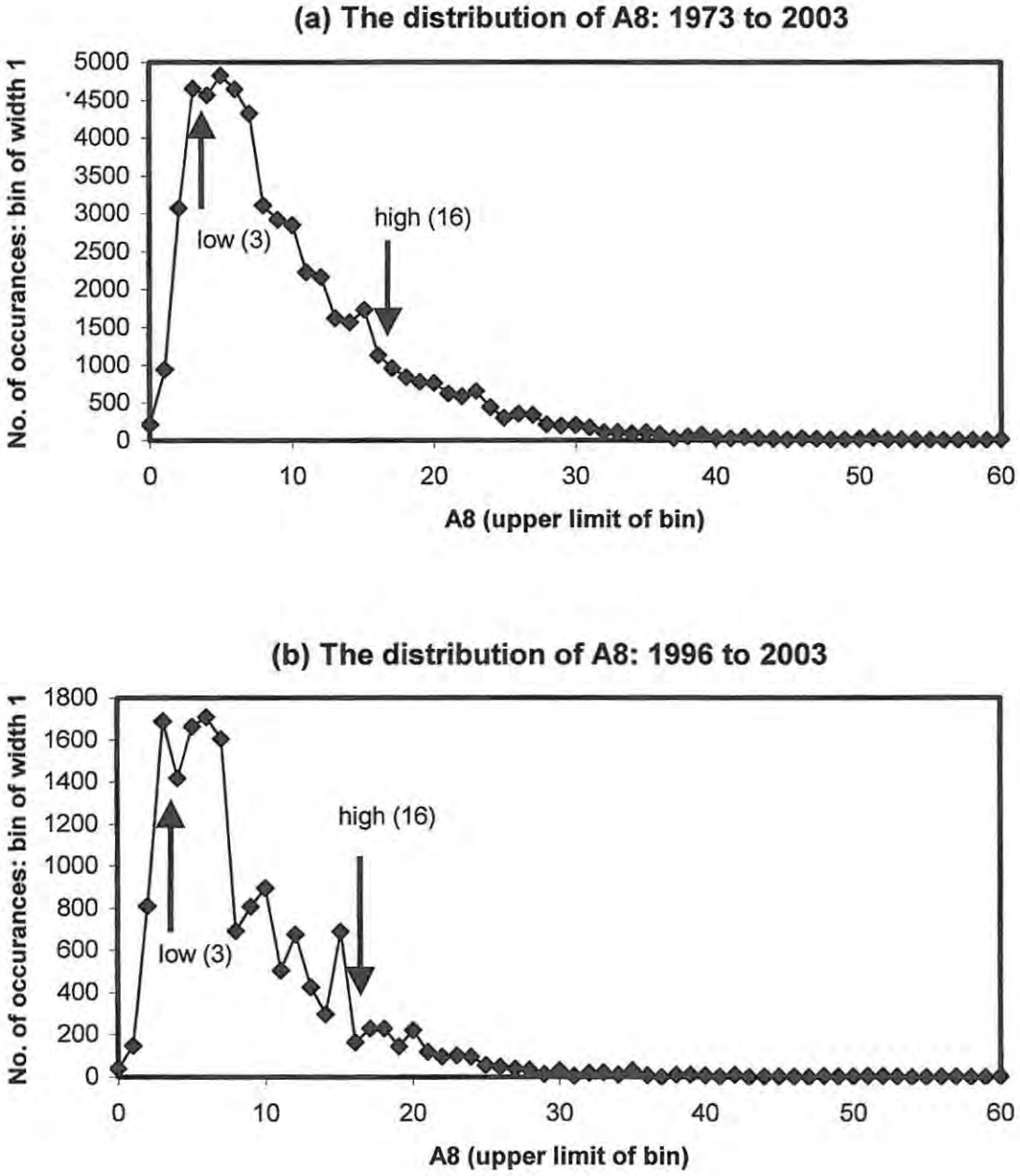


Figure 3-5: Graphs illustrating the values chosen to represent the low and high values of A8 for the period (a) 1973 to 2003 and (b) 1996 to 2003.

As shown in figures 3-4 and 3-5, different values were chosen depending on the period over which the NN was trained. However, for comparison purposes, the high and low values determined for the 1973 to 2003 dataset were used to interrogate the 1996 to 2003 Probability NN and foF1 Prediction NN. Only the hmF1 NN was interrogated with the high and low values obtained for the 1996 to 2003 data.

Another method to assess the performance of the NNs is to compare the actual measured values with those predictions given by the NN. The rms error between the measured and predicted values gives an indication of the performance of the NN and can be used as a comparison between NNs.

Although NNs are very good at interpolating values within the given dataset, they are poor at extrapolating. Therefore it should be noted that the NNs ought not to be interrogated with input values that lie outside of the training dataset. This is particularly relevant for the 1996 to 2003 dataset as there are still parts of the input space that are not sufficiently covered by the dataset. The green sections in figure 3-6 are examples of areas in the solar activity and seasonal input space for which the 1996 to 2003 NNs were not sufficiently trained and that should be avoided when testing these NNs.

3.4 Estimating the Uncertainty

A method for determining the uncertainty in the predicted output values from the NN was developed by *Poole and McKinnell [2000]*. This method involves finding a statistical gauge of the difference between the measured and predicted values. For each input vector in the original

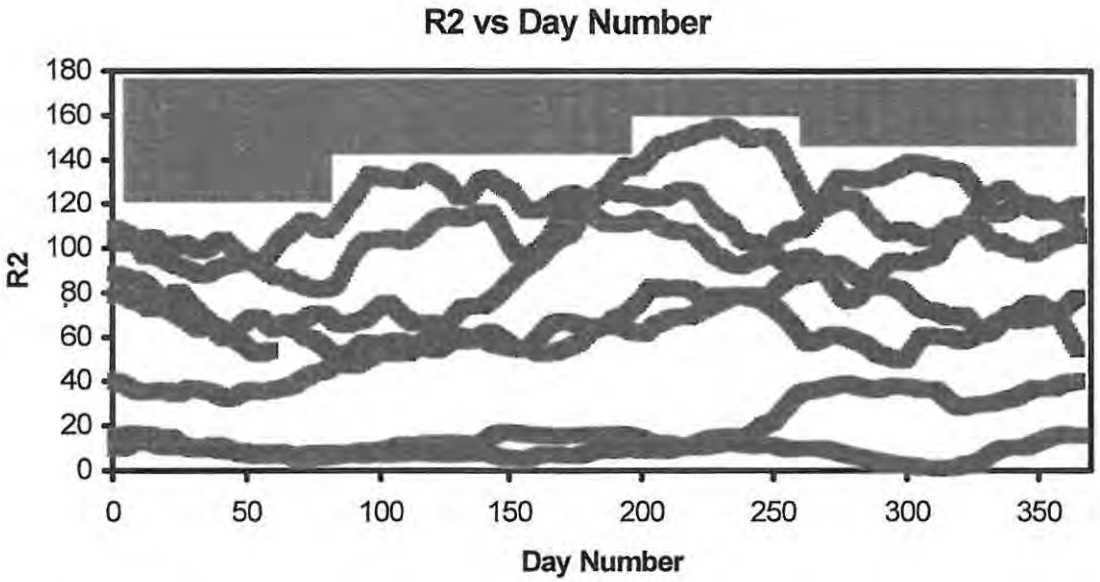


Figure 3-6: This graph illustrates the extent to which the input space (R2, DN) is covered by the available data from 1996 to 2003. The green boxes are examples of areas that should be avoided when interrogating the NNs.

dataset used to train the NN the difference between the actual and the predicted values is calculated and then squared. A second NN is then trained with the same input data but with the squared difference as the output. The output from this second NN provides a mean squared error, the square root of which represents an estimate of the uncertainty on the first NN's prediction.

For each of the 25 NNs that were trained to predict the output value for the three required NN models (foF1 1996 to 2003, foF1 1973 to 2003, and hmF1), a second set of 25 NNs were trained to estimate the uncertainty between the predicted output and the actual value. The error for the two probability NNs was not calculated as a probability inherently implies an uncertainty, and the need to apply some confidence level to the probability was not deemed sufficient to warrant the extra research.

In the next chapter the results from each of the five sets of NNs are given and are compared to results obtained from the LAM model.

Chapter 4

NEURAL NETWORK RESULTS

4.1 F1 Probability NNs

As discussed in the previous chapter, two separate NN sets were trained to predict the probability of occurrence of the three different F1 existence categories: P(N), P(F) and P(L). The inputs and outputs for the Probability NN are shown in figure 4-1. According to *McKinnell [2002]*, the state of the F1 layer is influenced by season, time of day and solar activity. Magnetic influences are imperceptible and therefore no measure of magnetic activity was included as an input.

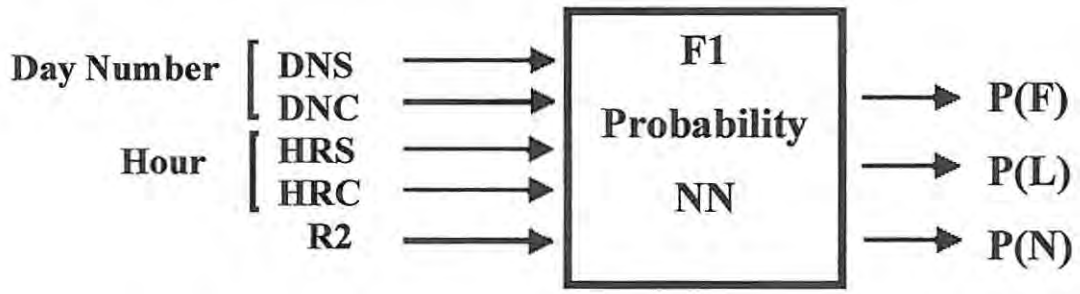


Figure 4-1: A block diagram depicting the inputs and outputs of the F1 Probability NN.

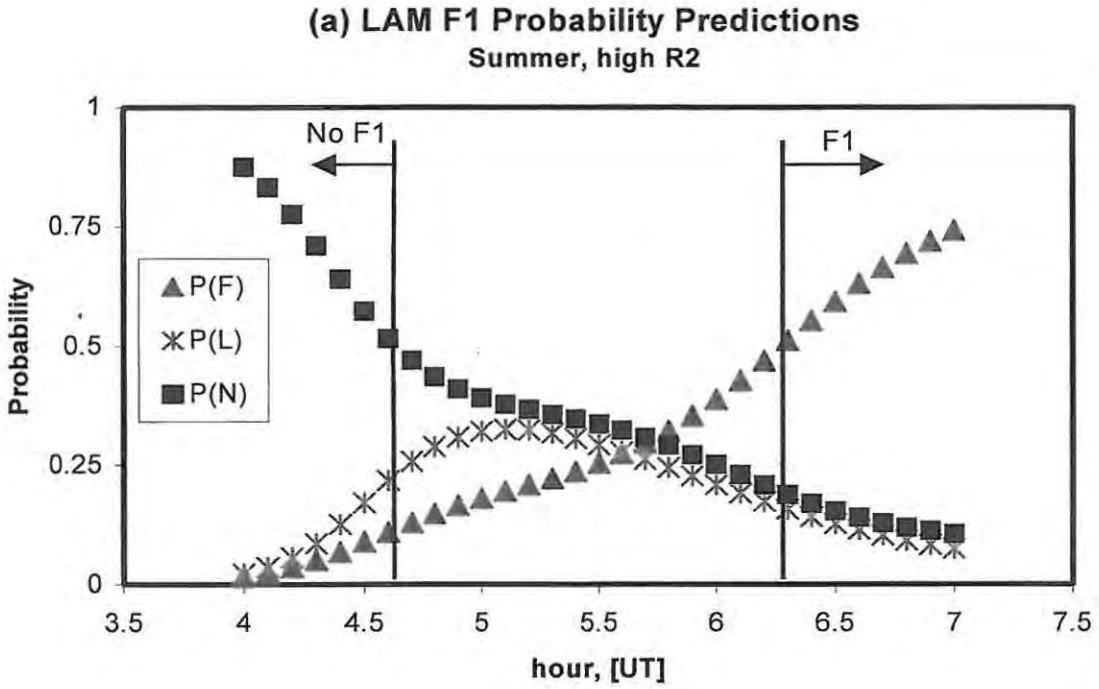
For each training input vector, the state of the F1 layer was determined and the output was set as shown in table 4-1. The predictions from the Probability NN then give a value between 0 and 1 for each state as an output. In order to determine the category that has the highest probability, certain conditions need to be applied to the predicted output from the NN. If P(N) is greater than 0.5 the category is N and it is most probable that no F1 layer exists. If P(F) is greater than 0.5 it is determined that a definite F1 layer is present. If both P(N) and P(F) are less than 0.5 then an L-condition is concluded to be most probable. For each input vector, the sum of P(N), P(F) and P(L) will be 1.

CATEGORY	P(N)	P(F)	P(L)
NO F1 LAYER PRESENT: N	**1**	**0**	**0**
DEFINTE F1 LAYER: F	**0**	**1**	**0**
L-CONDITION: L	**0**	**0**	**1**

Table 4-1: The values for each output to set the state of the F1 layer to be N, F or L are shown.

The optimum architecture for the 1973 to 2003 Probability NN (Probability NN1) was found to have 35 hidden nodes while for the 1996 to 2003 Probability NN (Probability NN2) only 25 hidden nodes were used. Both NNs were trained over 1000 epochs.

Two distinct patterns of F1 occurrence are apparent. Firstly, over the course of a day the F1 layer exhibits at least two of the states, and in certain instances all three states. This case in point is especially evident in summer at high solar activity. Figures 4-2 to 4-4 illustrate the probability of occurrence of each F1 state at the hours of transition from one state to the next, comparing the LAM model results to those predictions obtained from Probability NN1 and Probability NN2. The graphs show the average progression between no F1 layer (N) and a definite F1 layer (F), with L-condition marked between the black lines. Table 4-2 summarises these results in the form of the hours between which an L-condition is most probable. The variation between the different NNs is expected and is due to the NNs being trained with different datasets, as well as the fact that all NNs are inherently unique due to their random starting weights.



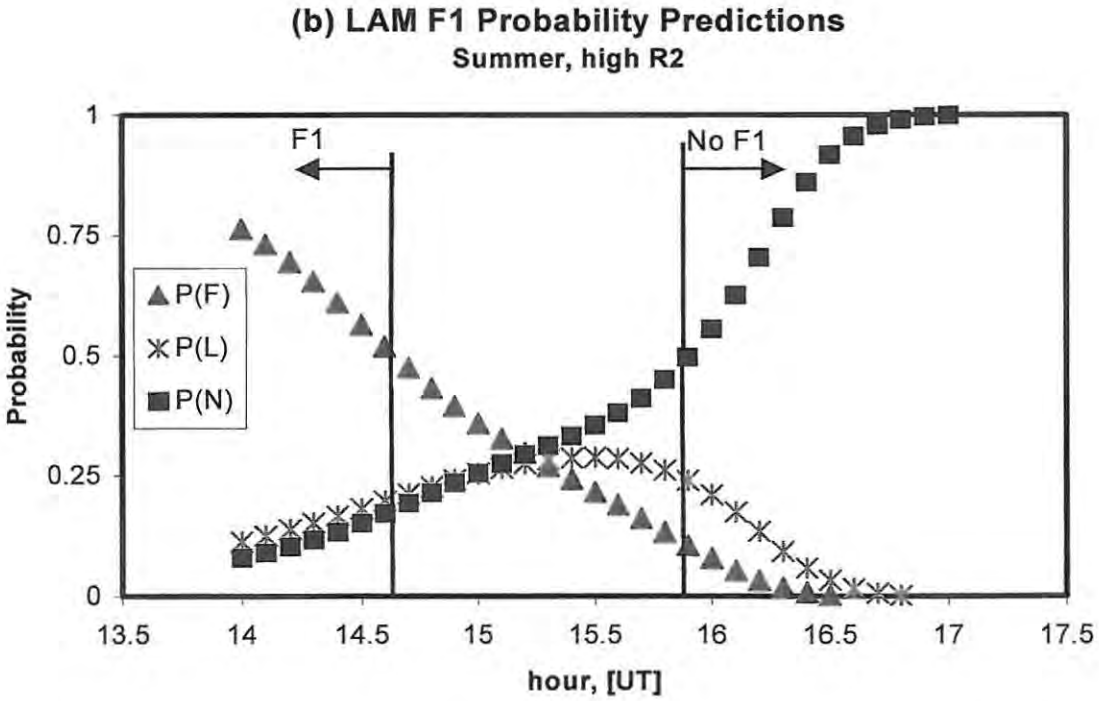
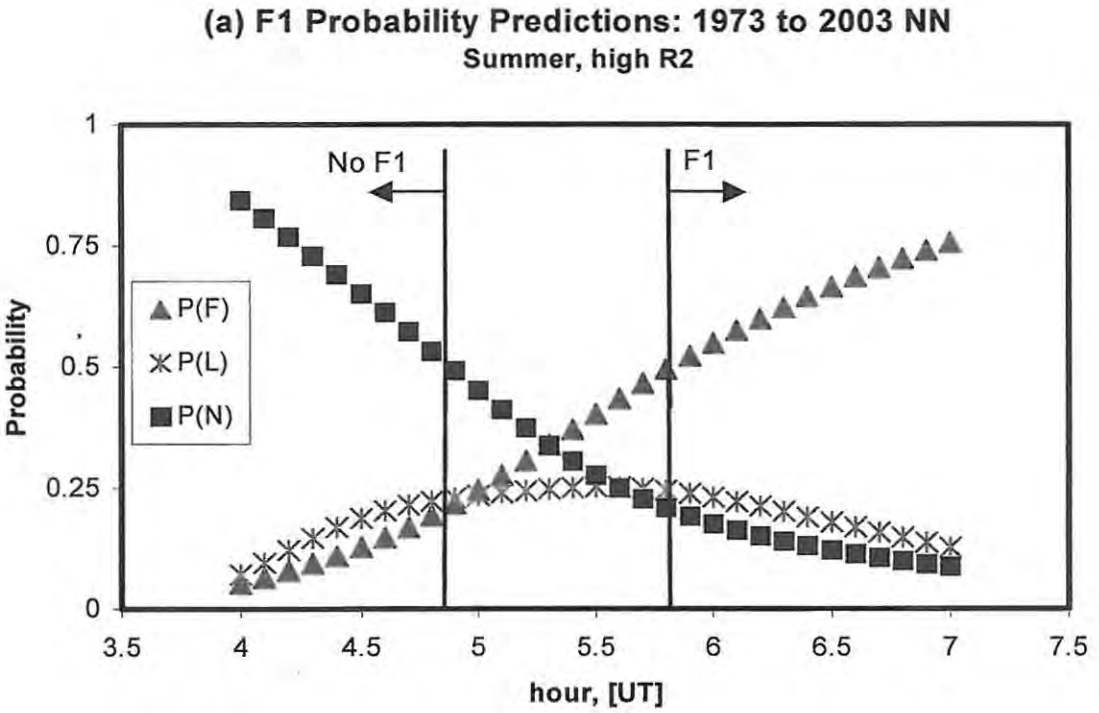


Figure 4-2: These graphs indicate graphically the output from the LAM F1 Probability NN for a summer day at high R2. The occurrence probability of an F1 layer is shown for the (a) start and (b) end of the day. On the graphs the vertical black lines indicate the hours between which an L-condition is most probable.



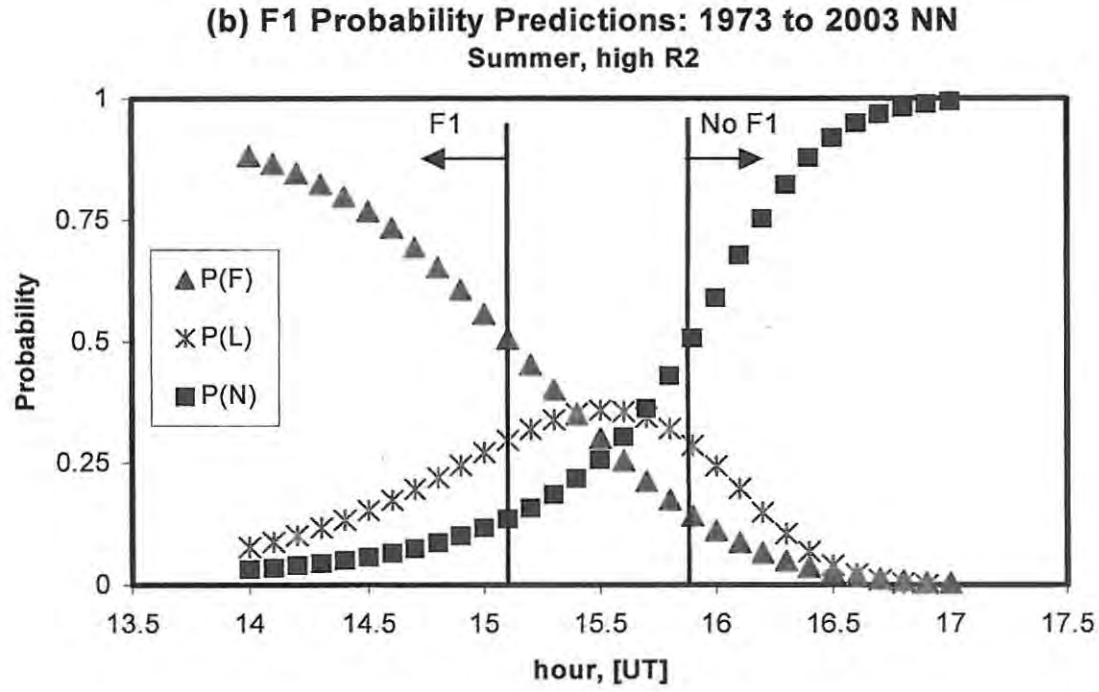
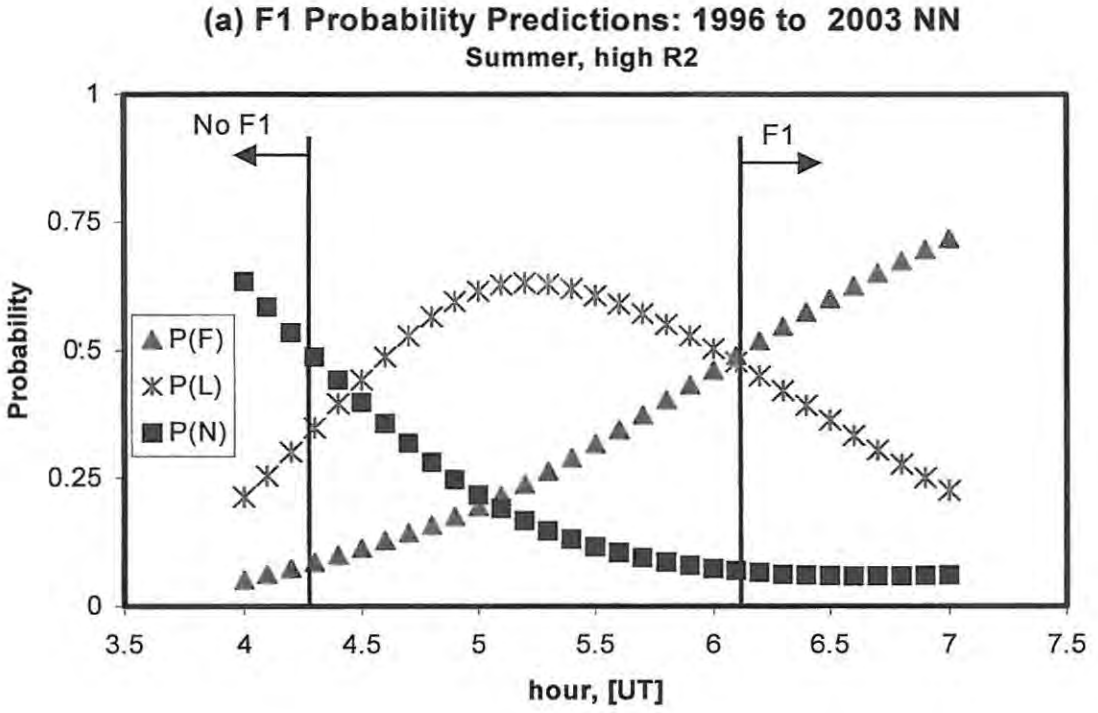


Figure 4-3: These graphs indicate graphically the output from Probability NN1 for a summer day at high R2. The occurrence probability of an F1 layer is shown for the (a) start and (b) end of the day. On the graphs the vertical black lines indicate the hours between which an L-condition is most probable.



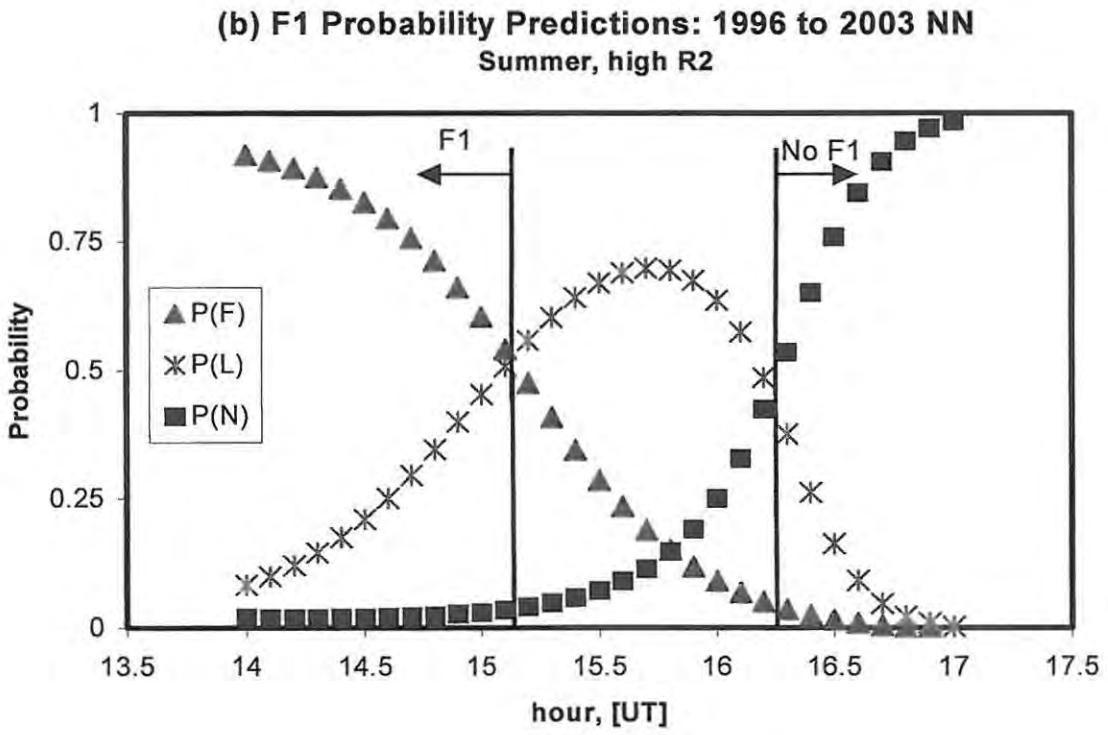


Figure 4-4: **These graphs indicate graphically the output from Probability NN2 for a summer day at high R2. The occurrence probability of an F1 layer is shown for the (a) start and (b) end of the day. On the graphs the vertical black lines indicate the hours between which an L-condition is most probable.**

	THE TIMES BETWEEN WHICH AN L-CONDITION IS MOST PROBABLE FOR A SUMMER DAY (DN=356) AT HIGH R2	
NEURAL NETWORK MODEL	**START OF DAY [UT]**	**END OF DAY [UT]**
LAM F1 PROBABILITY NN	0436 – 0618	1436 – 1554
1973 TO 2003 PROBABILITY NN	0454 – 0548	1506 – 1554
1996 TO 2003 PROBABILITY NN	0415 – 0609	1509 – 1615

Table 4-2: A summary of the hours [UT] between which an L-condition is expected for the three different F1 Probability NNs tested on a summer day (DN = 356) at high R2. The F1 layer is expected to be in the F state between these two periods of L-condition and in the N state during the night.

In the case of Probability NN2, the difference can also be attributed to the fact that the majority of the data used to train the net was at high solar activity which possibly biased the net to predict a greater instance of L-condition occurrence.

The second pattern of F1 occurrence omits the formation of a definite F1 layer and an L-condition exists throughout daylight hours. This situation tends to occur at high R2 in autumn and winter and is illustrated in figure 4-5, which shows the F1 occurrence probability for autumn equinox at high R2. It can be seen that the 1973 to 2003 and 1996 to 2003 Probability NNs display very similar results. The LAM model predicts an L-condition between the hours of 07h00 and 14h00 UT for an autumn day at high R2.

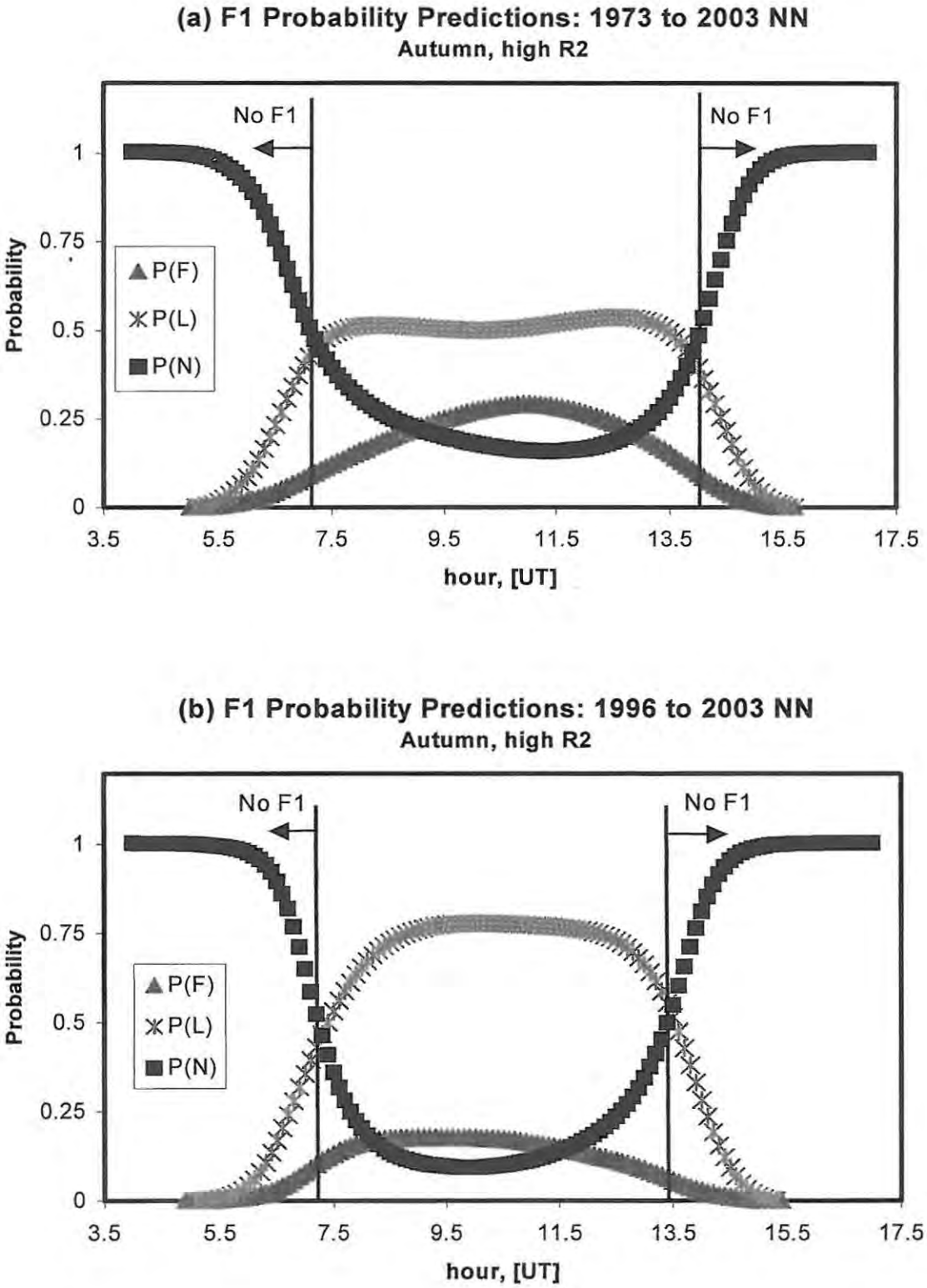


Figure 4-5: **These graphs illustrate the output from the (a) Probability NN1 and (b) Probability NN2 for an autumn day at high R2. The occurrence probability of an F1 layer is shown and highlights an example of when an L-condition is most probable during daylight hours.**

As a comparison to these predicted results, the percentage of L-condition occurrence is shown in figure 4-6. Only those records with an R2 value above 100 were extracted from the 1973 to 2003 dataset to plot the graphs. Autumn and summer periods were taken to extend 35 days on either side of autumn equinox and summer solstice.

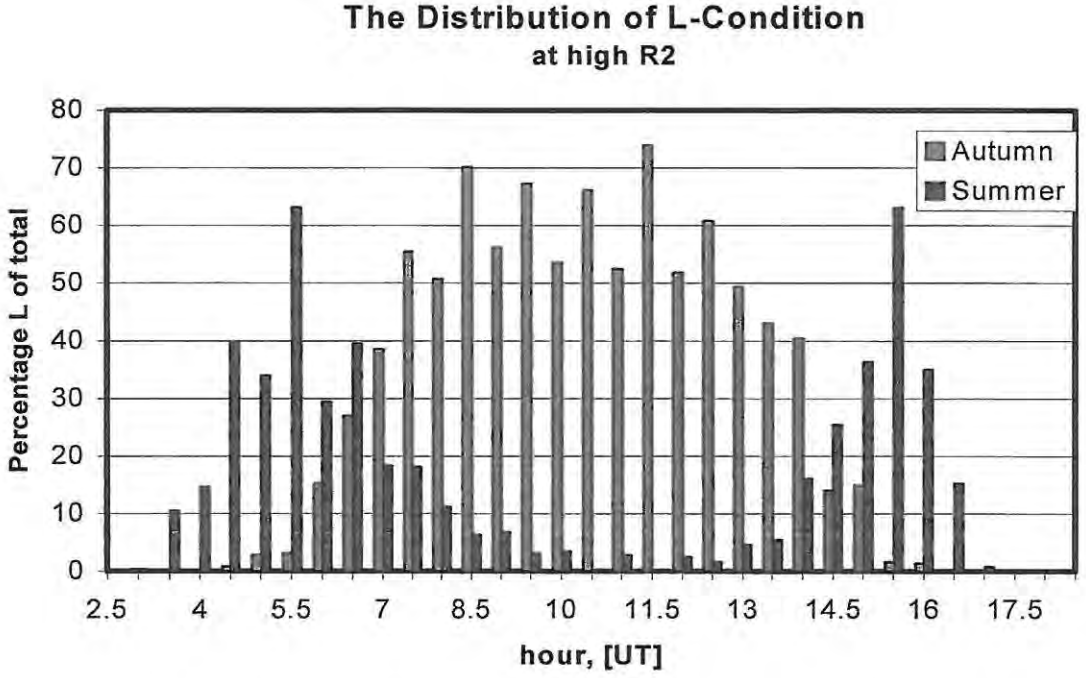


Figure 4-6: A graph showing the distribution of L-condition as a percentage of the total number of records for those periods in autumn and summer having an R2 value greater than 100.

From figure 4-6 it is evident that L-condition is prevalent at 05h30 [UT] and at 15h30 [UT] during summer days and between 07h30 and 13h00 [UT] during autumn days.

When an L-condition is predicted as being likely, the LAM model applies the L-condition algorithm *(McKinnell, [2002])*, explained in Chapter 1.

4.2 foF1 Prediction NNs

When the F state is predicted to be present, an foF1 Prediction NN can be used to predict the value of the critical frequency of the F1 layer. Two different NN sets were constructed and trained to predict foF1: one set using data from 1973 to 2003 (foF1 NN1) and the other including data from only 1996 to 2003 (foF1 NN2). According to *McKinnell [2002]*, season, time of day, solar activity and magnetic activity all influence the value of foF1. The inputs to and output from the foF1 Prediction NN are shown as a block diagram in figure 4-7.

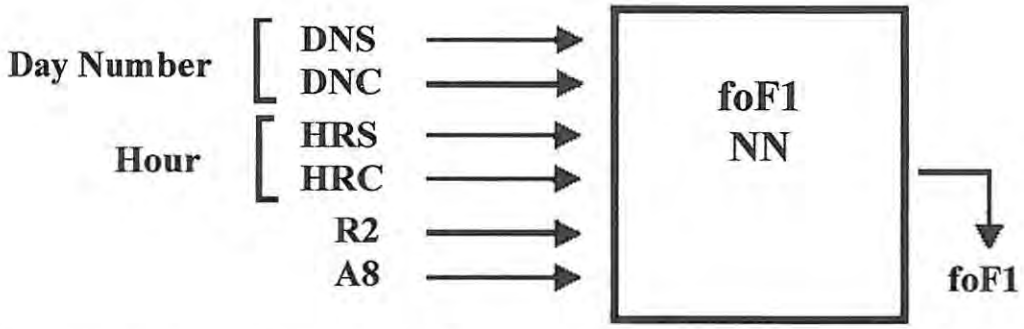


Figure 4-7: A block diagram indicating the inputs and output for the foF1 Prediction NN.

The foF1 NN1 was tested by predicting the 12h00 SAST foF1 values for a year of high R2 (1980) and a year of low R2 (1976) and comparing them with measured values. Figure 4-8 shows the comparison between the LAM model results and the predictions given by foF1 NN1 for these two years. The rms errors between the measured and predicted foF1 values are indicated on the graphs. Probability NN1 was interrogated with the same dataset to find instances of when no F1 layer was probable, and thus the gaps in the 1980 graphs correspond to those areas where the Probability NN predicts a value of more than 0.5 for P(N). At times when the Probability NN predicts an L-condition to be present, the foF1 values from the foF1 prediction NN have been included.

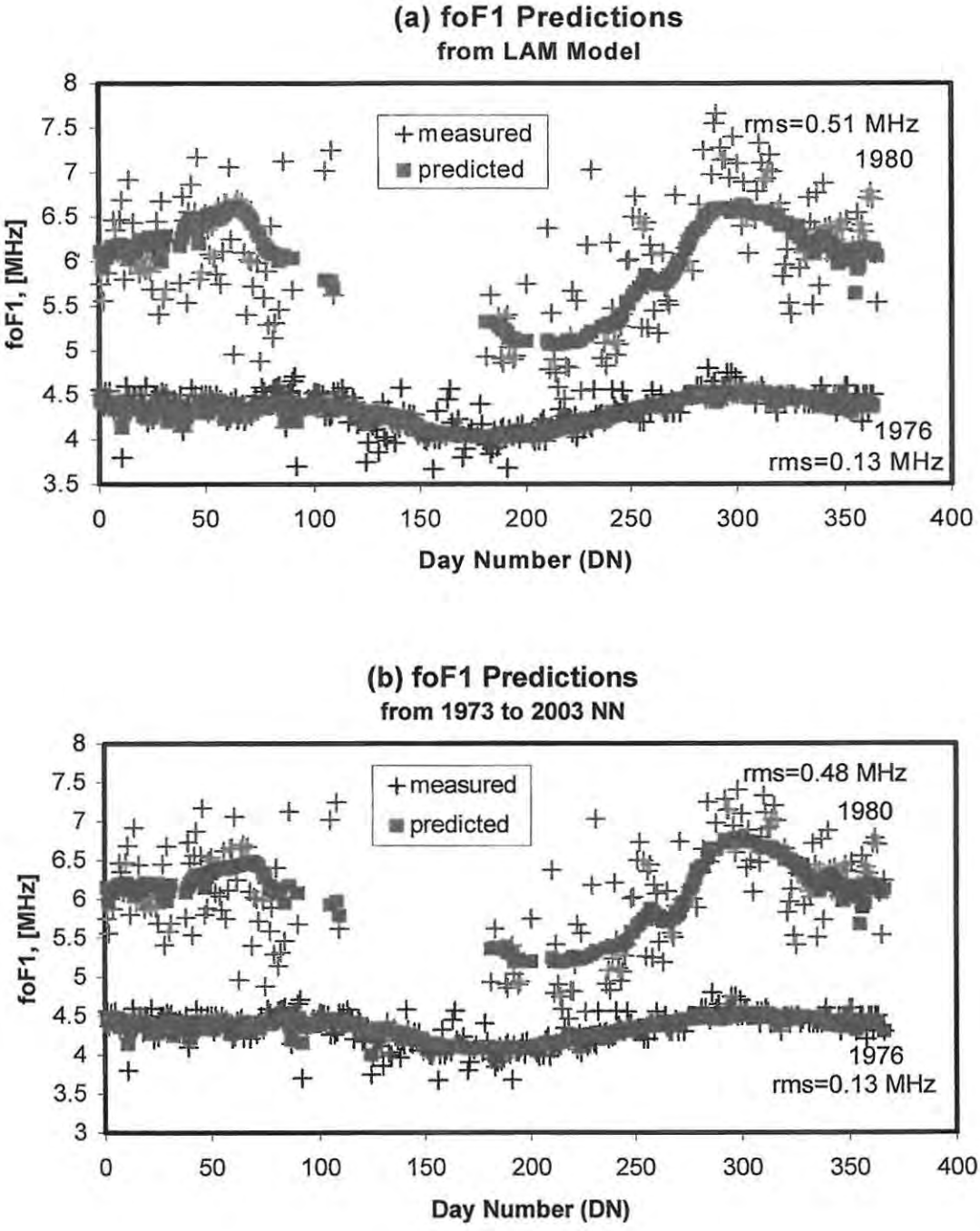


Figure 4-8: The measured and predicted 12h00 SAST foF1 values are shown for 1976 (low R2) and 1980 (high R2) for (a) the LAM model foF1 NN and (b) foF1 NN1. The rms errors are indicated on the graphs and show a slight improvement for the 1973 to 2003 foF1 NN at high R2.

From figure 4-8 it can be seen that the rms errors between the measured and predicted values are slightly better for the 1973 to 2003 NN than for the LAM model at high R2. During the low R2 year, the values are the same. Since foF1 NN1 was trained with only an additional 3 years of data as compared with the LAM model, no great improvement was expected. The additional 3 years (2001 to 2003) are over a high solar activity period and this could explain the slight improvement on the LAM model at high R2 but not at low R2.

The LAM model was tested at 12h00 SAST for 2001, a year for which no training data was included in the original LAM model dataset. Both foF1 NN1 and foF1 NN2 were tested with the same 2001 data and the measured and predicted values are shown in figure 4-9, along with the LAM model results. The LAM model prediction is compared with ARTIST determined values (figure 4-9 (a)) while the two foF1 prediction NNs are compared with the adjusted measured values determined by the re-scaling (figure 4-9 (b) and (c)). For this reason, there are visible differences between the two measured sets. These comparisons are illustrated this way since the individual NNs were trained with slightly different measured sets, though interrogated with the same test set. The rms errors between the measured and predicted values are indicated on the graphs and show an improvement for both foF1 prediction NNs on the LAM model. The similar results obtained from foF1 NN1 and foF1 NN2 show that even though less than one solar cycle was used to train the latter NN set, the results can still be seen to be valid, at least for high R2. The gaps in the graphs indicate those times that Probability NN1 predicts no F1 layer to be present. It should be noted, however, that the year 2001 was included in the foF1 NN1 and foF1 NN2 training sets, and therefore the NNs were interrogated with "seen" data.

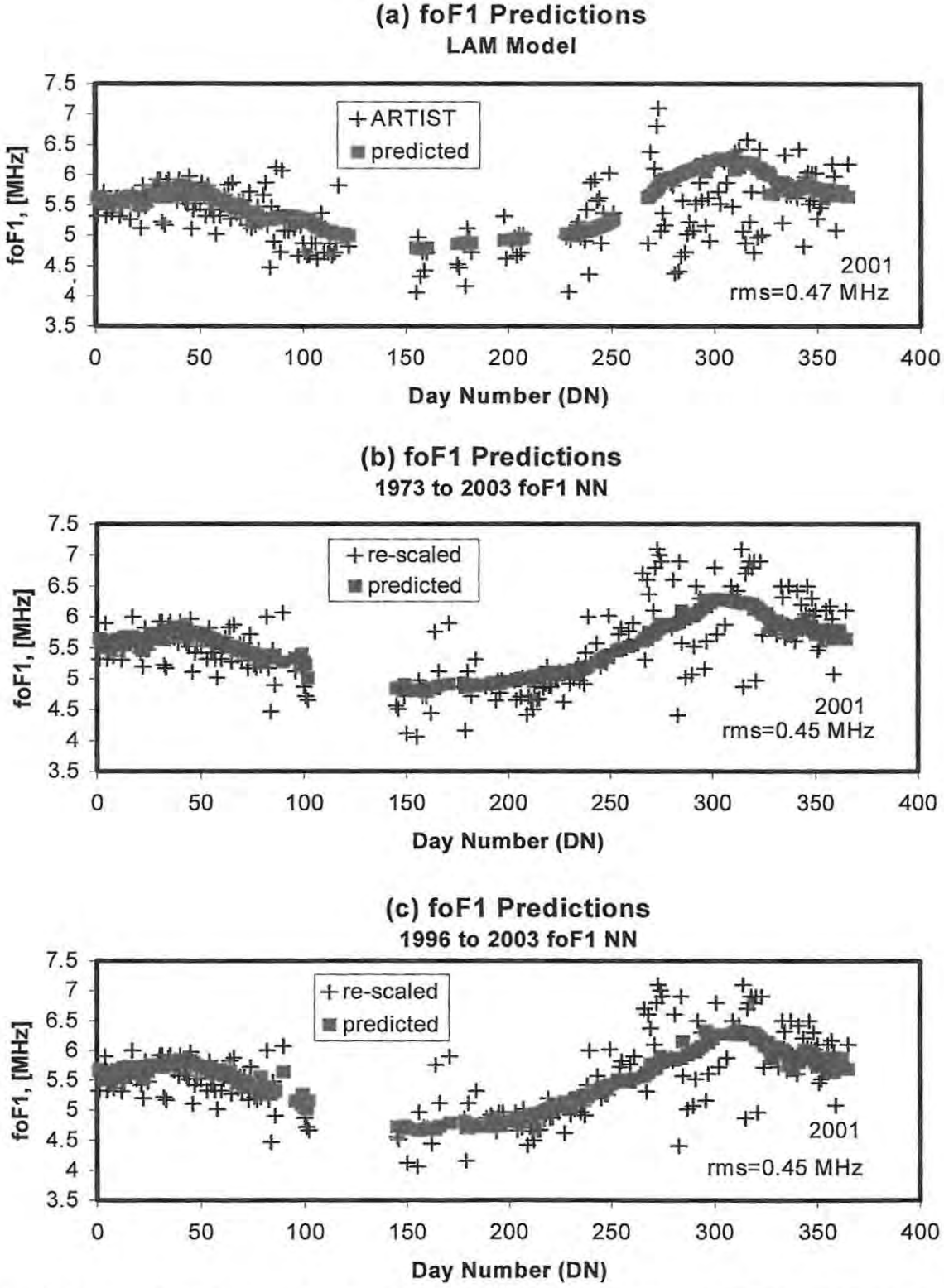


Figure 4-9: The measured and predicted 12h00 SAST foF1 values for the year 2001 are shown for the (a) LAM model foF1 NN, (b) foF1 NN1 and (c) foF1 NN2. The rms errors are indicated on the graphs. Differences between the LAM measured values and the two foF1 NN measured values are due to those ARTIST values having been re-scaled manually.

Figure 4-10 shows the 12h00 SAST measured and predicted values from 1996 to 2003 for both the (a) foF1 NN1 and (b) foF1 NN2. These graphs were generated as it was suspected that the re-scaled data might have unrealistically shifted the average prediction of the NNs towards higher values of foF1 since ARTIST often underestimated the value, shown in figure 2-6. Figure 4-10 shows that this was not the case. If it had been, the 1996 to 2003 NN would have shown much higher predicted values of foF1 than the 1973 to 2003 NN. The rms errors between the measured and predicted values are indicated on the graphs and this error is marginally better for foF1 NN2.

Figure 4-11 shows the measured and NN predicted foF1 values for the available months in 2004, a year that was not included in any of the training datasets. Results are shown from the (a) LAM model, (b) foF1 NN1 and (c) foF1 NN2. The rms errors are shown on the graphs and are similar for all three NNs, only differing in the third decimal place. However, from the graphs it can be seen that for the 1996 to 2003 foF1 NN the predicted foF1 values are slightly overestimated at the beginning of the year. This is probably due to 2004 being a year of relatively low solar activity, the maximum R2 value in this period being 60, but foF1 NN2 was trained with data over a period of mostly high solar activity. It should be noted that the measured values used are ARTIST values and not re-scaled values.

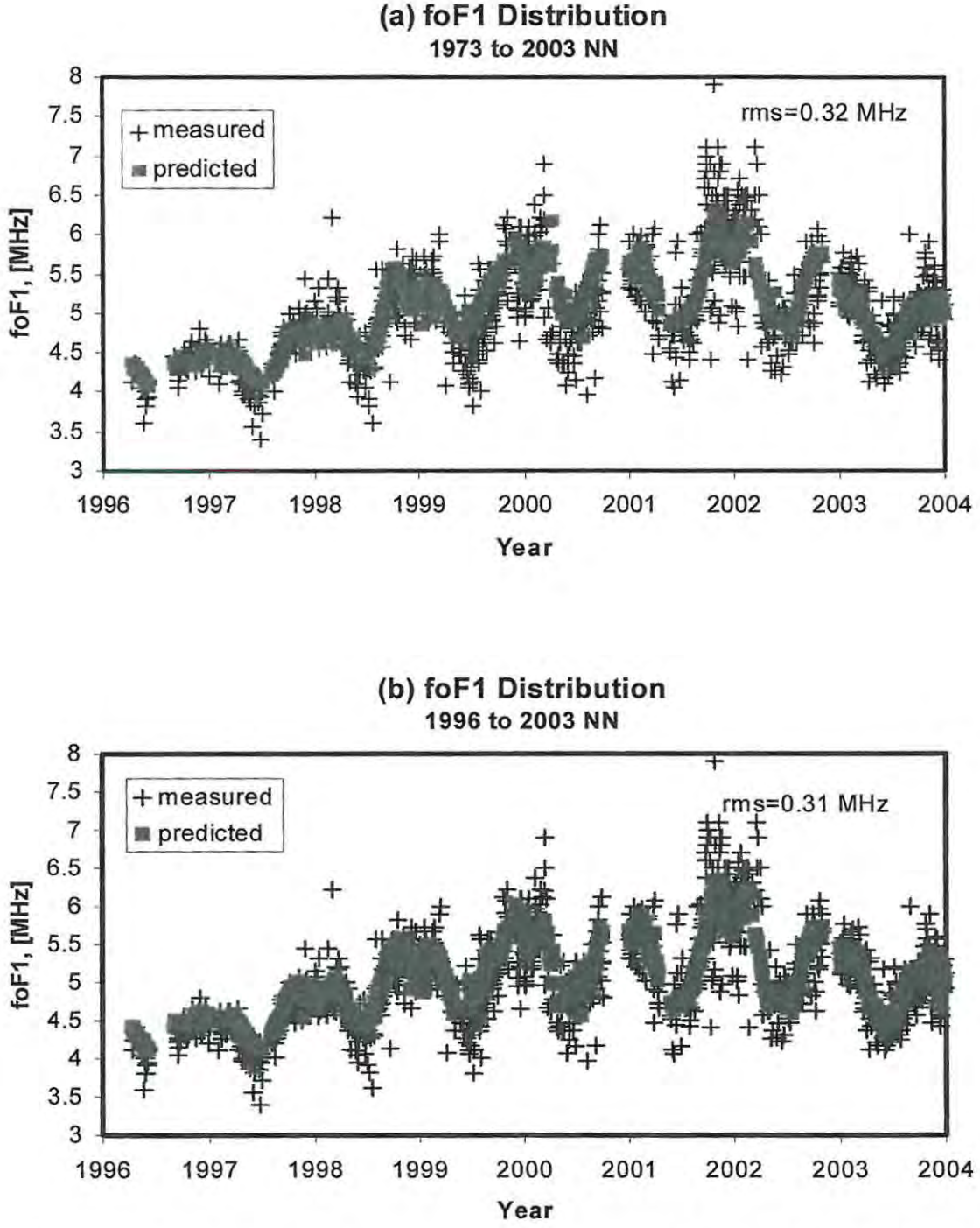


Figure 4-10: The measured and predicted 12h00 SAST foF1 values for the period 1996 to 2003 for both the (a) foF1 NN1 and (b) foF1 NN2. The rms errors are indicated on the graphs.

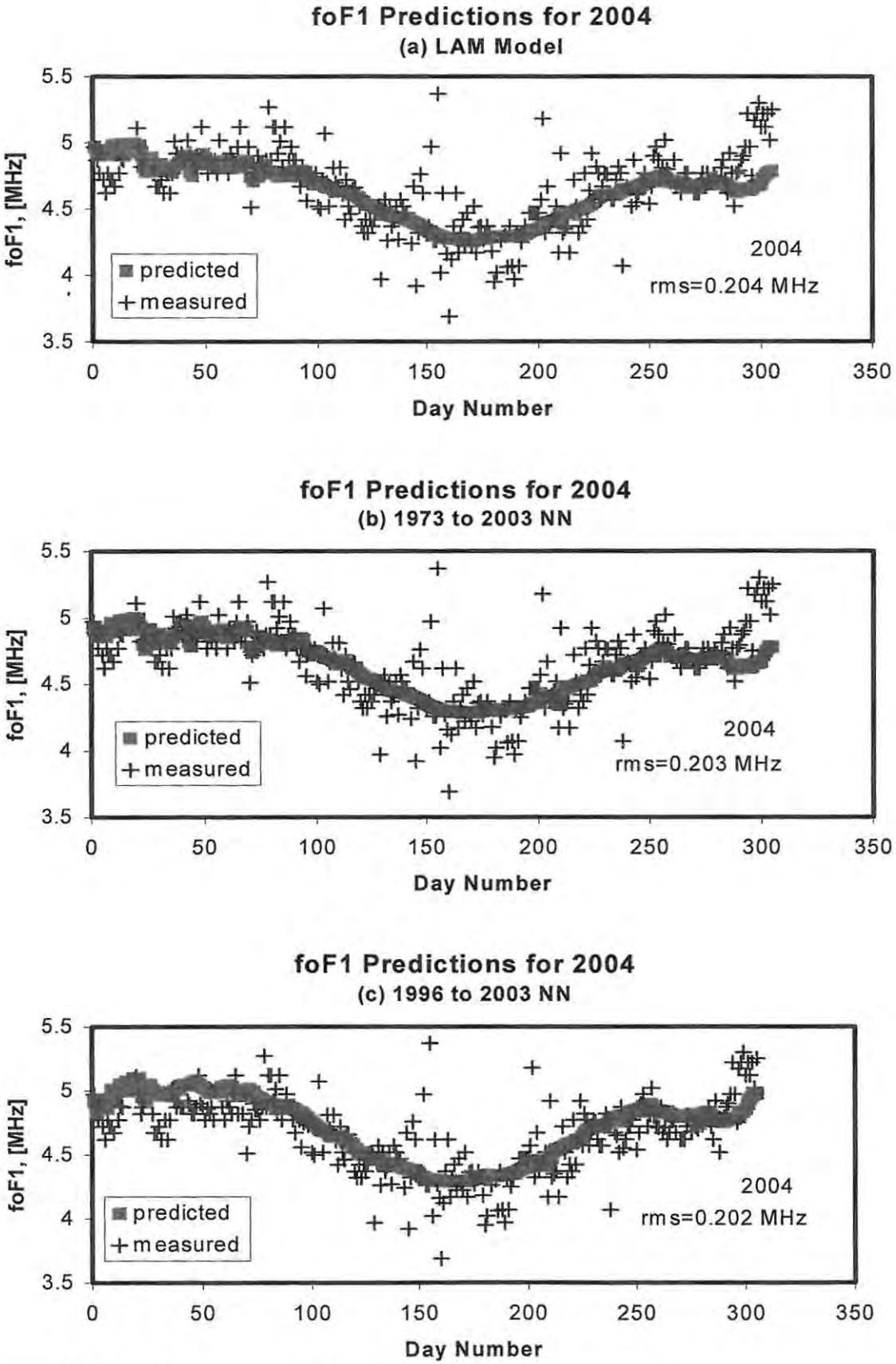


Figure 4-11: The measured and predicted 12h00 SAST foF1 values for 10 months of 2004 for (a) the LAM model, (b) foF1 NN1 and (c) foF1 NN2. The rms errors are indicated on the graphs.

To illustrate the error network prediction, figure 4-12 shows the measured and foF1 NN1 predicted foF1 values for 12h00 SAST for the year 1975. The error bars represent the predicted estimate of the uncertainty on the foF1 NN1 prediction. This uncertainty was predicted using the method of training 25 additional NNs with the squared difference between the measured values and the original predicted values as the output, described in Chapter 3.4.

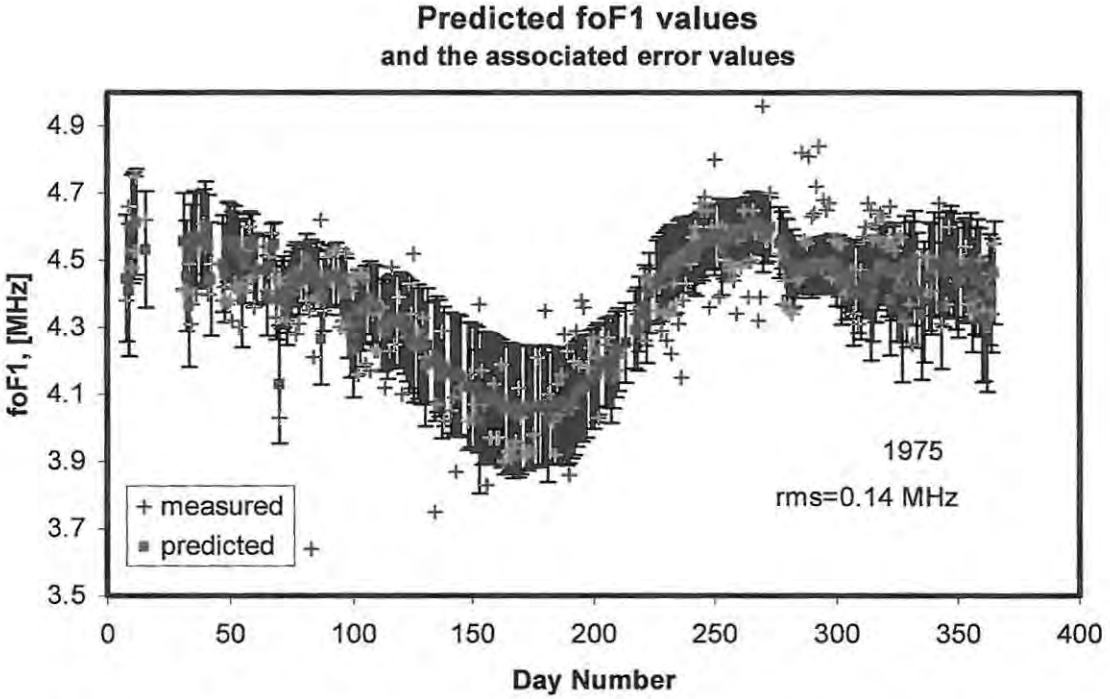
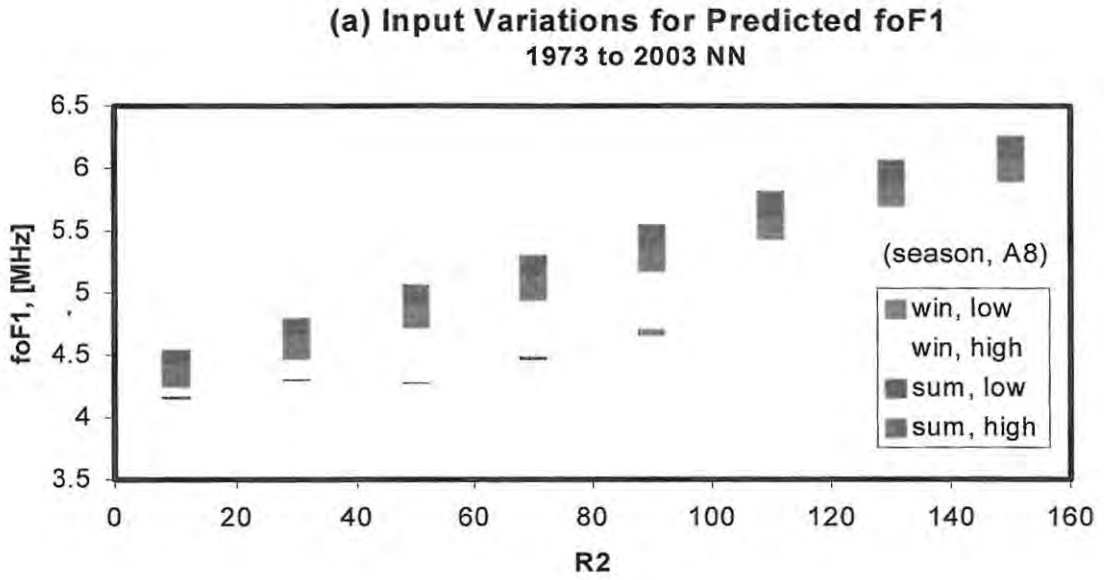


Figure 4-12: The measured and predicted 12h00 SAST values for 1975 from the 1973 to 2003 foF1 NN, with the predicted error associated with each input vector. The rms error between the measured and predicted foF1 values is indicted on the graph.

To investigate the response of foF1 to changing solar activity, the day number was set to 173 (winter solstice) and 356 (summer solstice) and the hour to 12h00 SAST while R2 was allowed to vary for both high and low values of A8. The resulting predictions from this input set are shown in figure 4-13 for the (a) foF1 NN1 and (b) foF1 NN2. Figure 4-13 shows that under all conditions the value of foF1 increases with increasing solar activity.



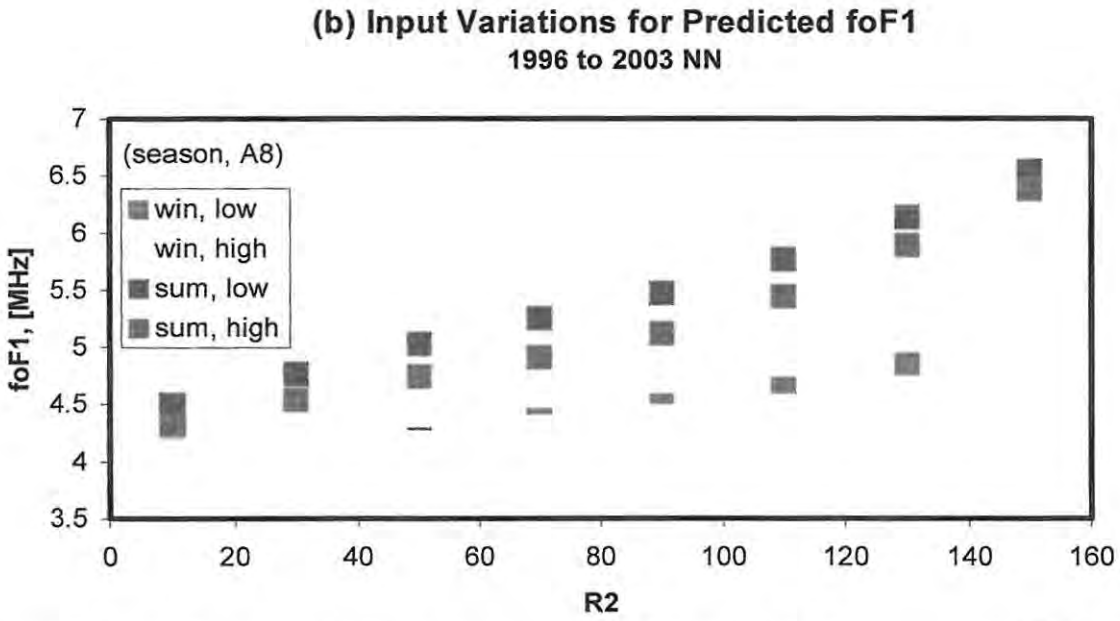


Figure 4-13: These graphs illustrate the seasonal variation in the response of foF1 to changing solar activity at low and high levels of magnetic activity for the (a) foF1 NN1 and (b) foF1 NN2. The break in the winter graphs at higher R2 is due to those conditions where the probability of no F1 layer being present is high.

Figure 4-13 is confirmed by the results shown in figure 4-8 whereby the values of foF1 for 1980, a year of high solar activity are greater than those values for 1976, a year of low solar activity.

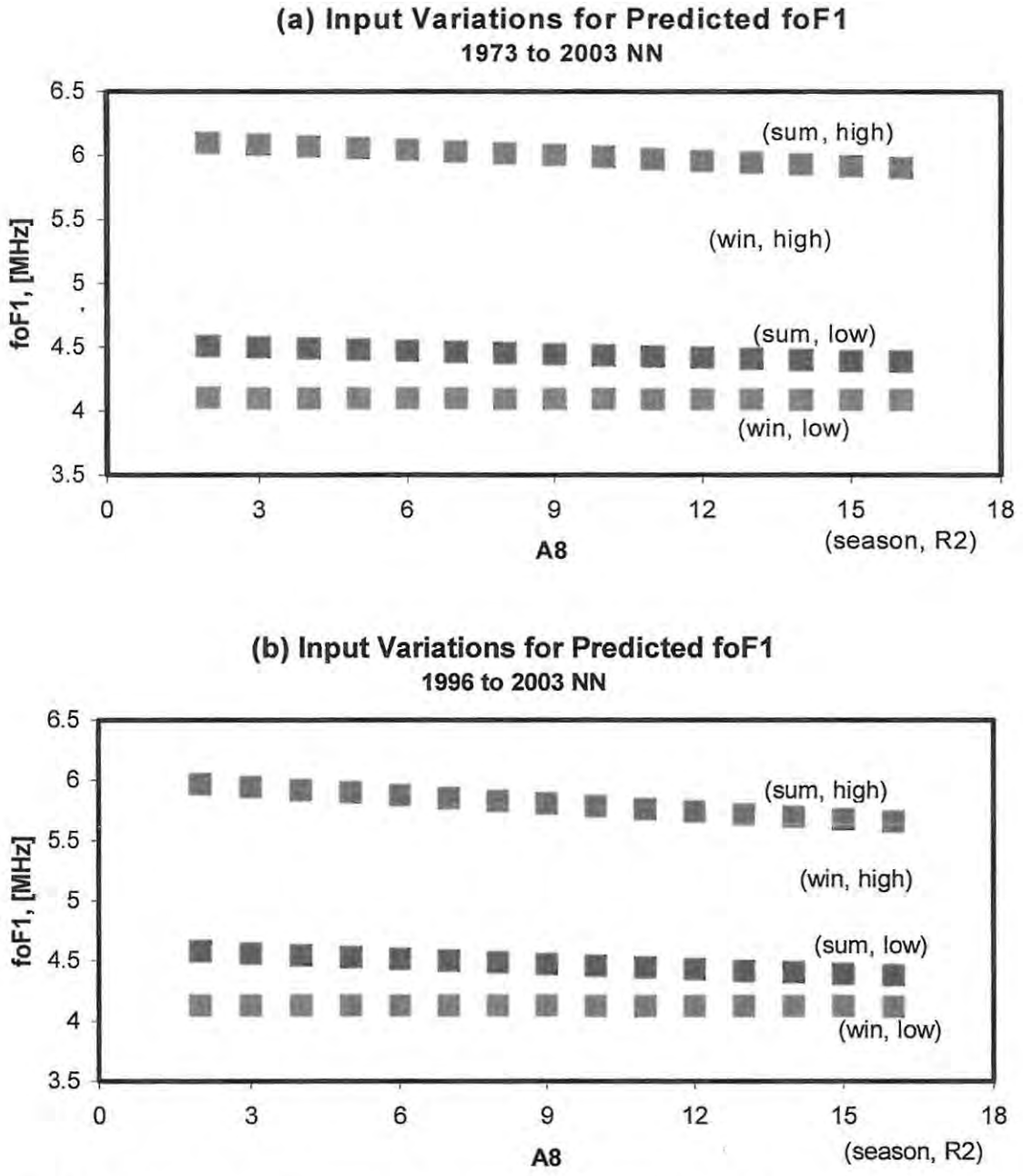


Figure 4-14: These graphs represent the seasonal response of foF1 to increasing magnetic activity at both high and low levels of solar activity for (a) foF1 NN1 and (b) foF1 NN2.

Figure 4-14 shows the response of foF1 to increasing magnetic activity for (a) foF1 NN1 and (b) foF1 NN2. As can be seen, the magnetic activity has a very small influence on the value of foF1. This is most noticeable at high solar activity during summer when the value of foF1 decreases with increasing magnetic activity and at high solar activity during winter when the value of foF1 increases with increasing magnetic activity. This result is

also evident from figure 4-13. The results shown in figures 4-13 and 4-14 were also obtained by *McKinnell [2002]*.

To further investigate the response of foF1 to changing input parameters, the diurnal variation of foF1 was predicted for the four combinations of high and low values of R2 and A8 for summer and autumn (figure 4-15) and winter and spring (figure 4-16). The comparison between the predicted outputs from the (i) LAM model, (ii) foF1 NN1 and (iii) foF1 NN2 are shown. The start and end hours for an F1 layer being present were predicted by the F1 Probability NNs and only the results for those hours for which the F and L states were predicted as likely are shown. Again, these graphs show that the difference between the predicted values at high and low A8 is minor compared with the difference between the values at high and low R2. The slight increase in the difference of the predicted values of foF1 between high and low magnetic activity, seen most noticeably in winter at high R2 and in summer at low R2, for the 1996 to 2003 NN as compared to the LAM model predictions can be attributed to the fact that during the period 1996 to 2003, a period of high solar activity, very few records of foF1 in winter months were included in the training data due to the L-condition being present at those times and therefore the NN is being interrogated in areas for which it was not trained. Even so, the trends are the same and the values comparable.

Throughout this Chapter, the predictions from the 1973 to 2003 foF1 NN and the 1996 to 2003 foF1 NN have been compared and show that although a much smaller dataset was used in training the latter NN, the results are on a par with the NN that was trained over a full solar cycle. These results are important for the next section, which deals with the prediction of hmF1, the real height of reflection at the critical frequency of the F1 layer, a case where only data from 1996 to 2003 was available for training.

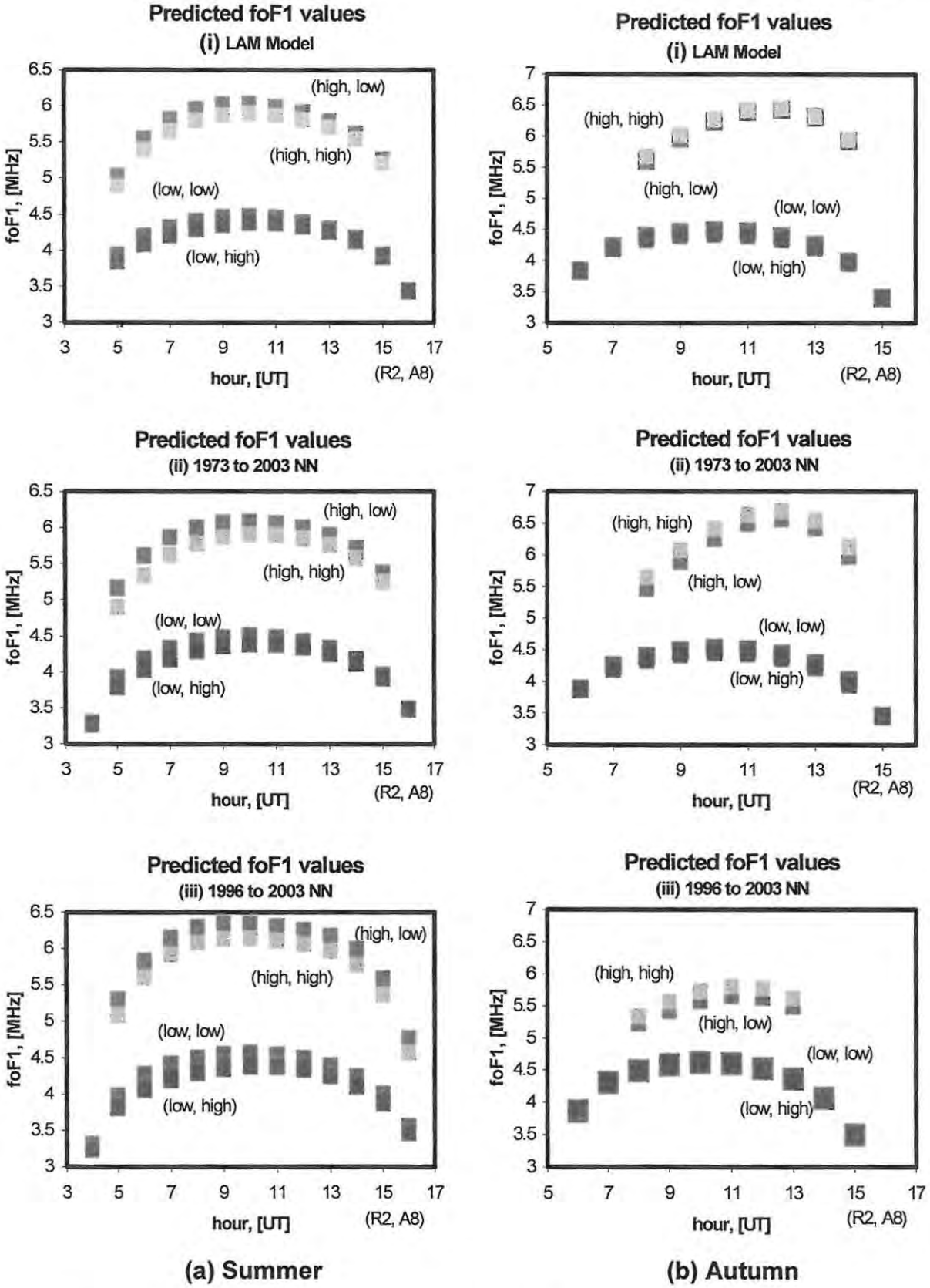
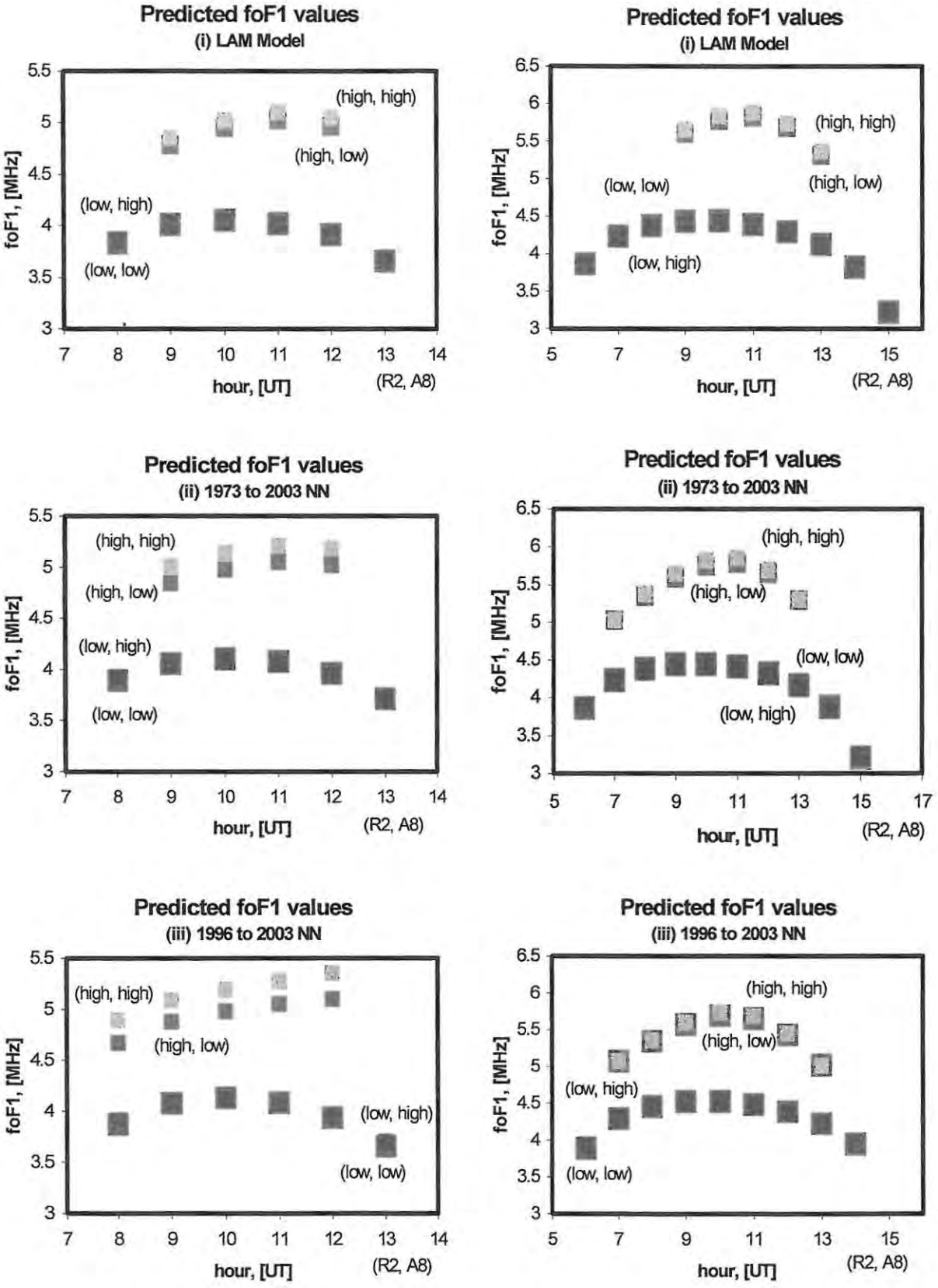


Figure 4-15: These graphs show the predicted foF1 values for artificial input data sets that were designed to investigate the relationship between the input parameters R2 and A8, and foF1. The graphs on the left show the response at (a) summer solstice, and the ones on the right at (b) autumn equinox for (i) the LAM model, (ii) foF1 NN1 and (iii) foF1 NN2. The low and high labels refer to the typical values of R2 and A8.



(a) Winter **(b) Spring**

Figure 4-16: Similarly to Figure 4-15, these graphs show the relationship between the input parameters R2 and A8, and foF1. The graphs on the left show the response at (a) winter solstice, and the ones on the right at (b) spring equinox for (i) the LAM model, (ii) foF1 NN1 and (iii) foF1 NN2. The low and high labels refer to the typical values of R2 and A8.

4.3 hmF1 Prediction NN

The existing real height data for the Grahamstown station was collected with the DPS system and so only data from 1996 to 2003 was available for training the hmF1 prediction NN (hmF1 NN). This presented some problems, as that period does not span a full solar cycle. However as shown in the previous chapter, this limited dataset made adequate predictions for foF1 as compared with the larger dataset. Furthermore, all those records for which foF1 had to be re-scaled were eliminated from the hmF1 training data because the information could not be considered accurate.

The inputs and output for the hmF1 NN are depicted in figure 4-17. According to *McKinnell [2002]*, season, hour, solar activity and magnetic activity all influence the value of hmF1. The optimum architecture was found to consist of 30 hidden nodes in the one hidden layer, and the NN was trained over 2000 epochs.

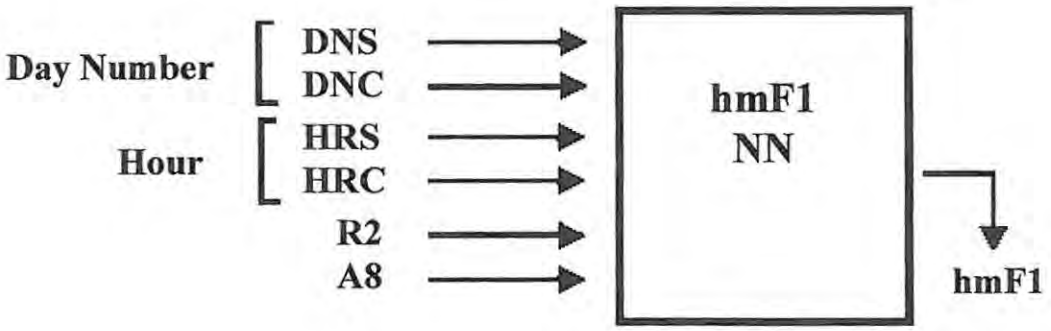


Figure 4-17: A block diagram illustrating the hmF1 NN inputs and output.

The available 12h00 SAST measured and predicted hmF1 values from 1996 to 2003 for (a) the LAM model and (b) the hmF1 NN are shown in figure 4-18. The rms error between the measured and predicted values was 12.21km for the LAM model and 11.66km for the hmF1 NN. The differences between the measured values in the two graphs are due to records having been eliminated from the 1996 to 2003 dataset as the foF1 value for such records was found to be incorrect. It is clear from this figure that the hmF1 NN is noticeably better at predicting the high values of hmF1

in summer than the LAM model. This is due to the larger dataset used, as well as the fact that during winter the L-condition predominates during daylight hours and no hmF1 value exists.

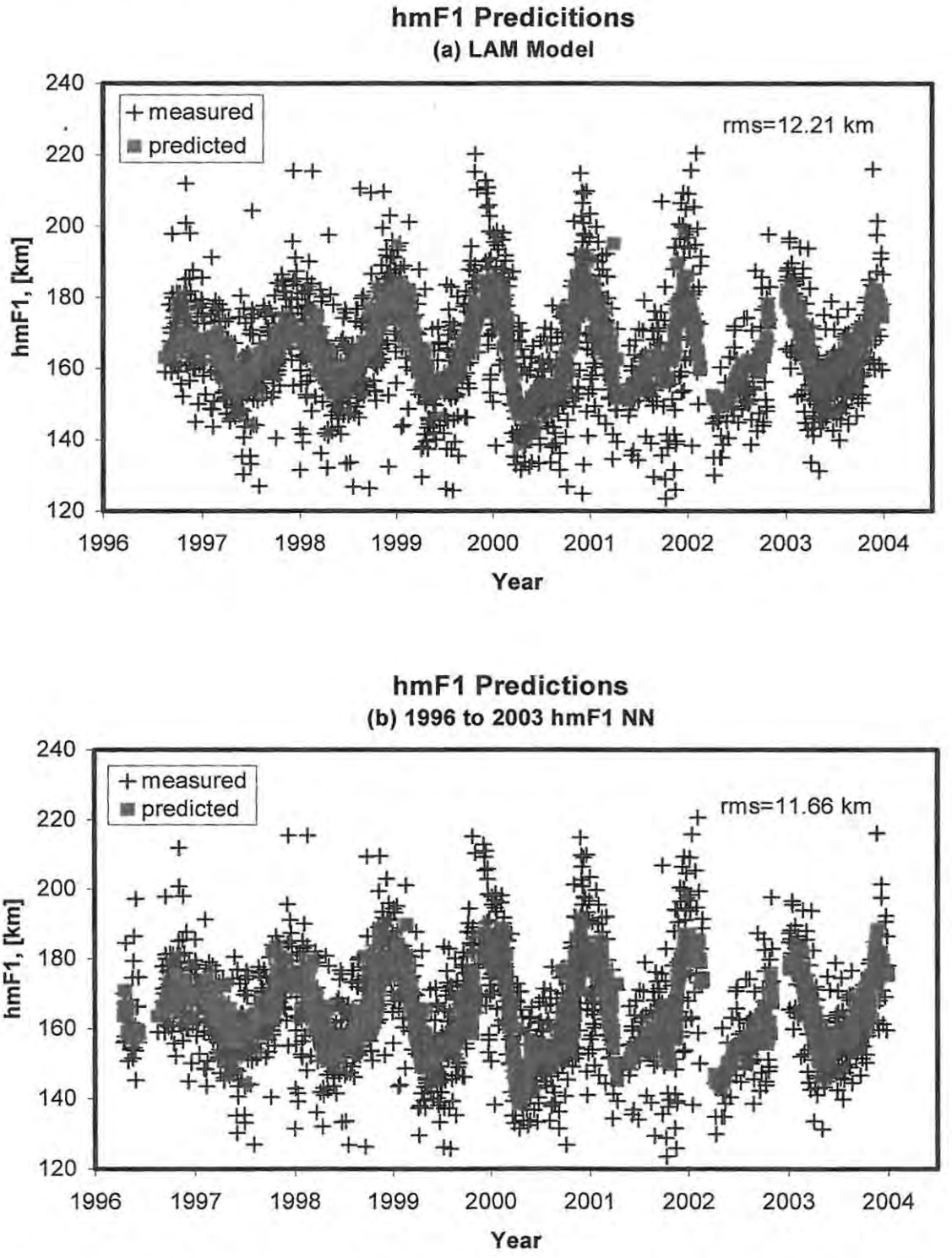


Figure 4-18: The measured and predicted 12h00 SAST hmF1 values are shown for (a) the LAM model and (b) the hmF1 NN. The LAM model hmF1 NN was trained using data only from 1996 to 2000. The rms errors for both NNs are indicated on the graphs.

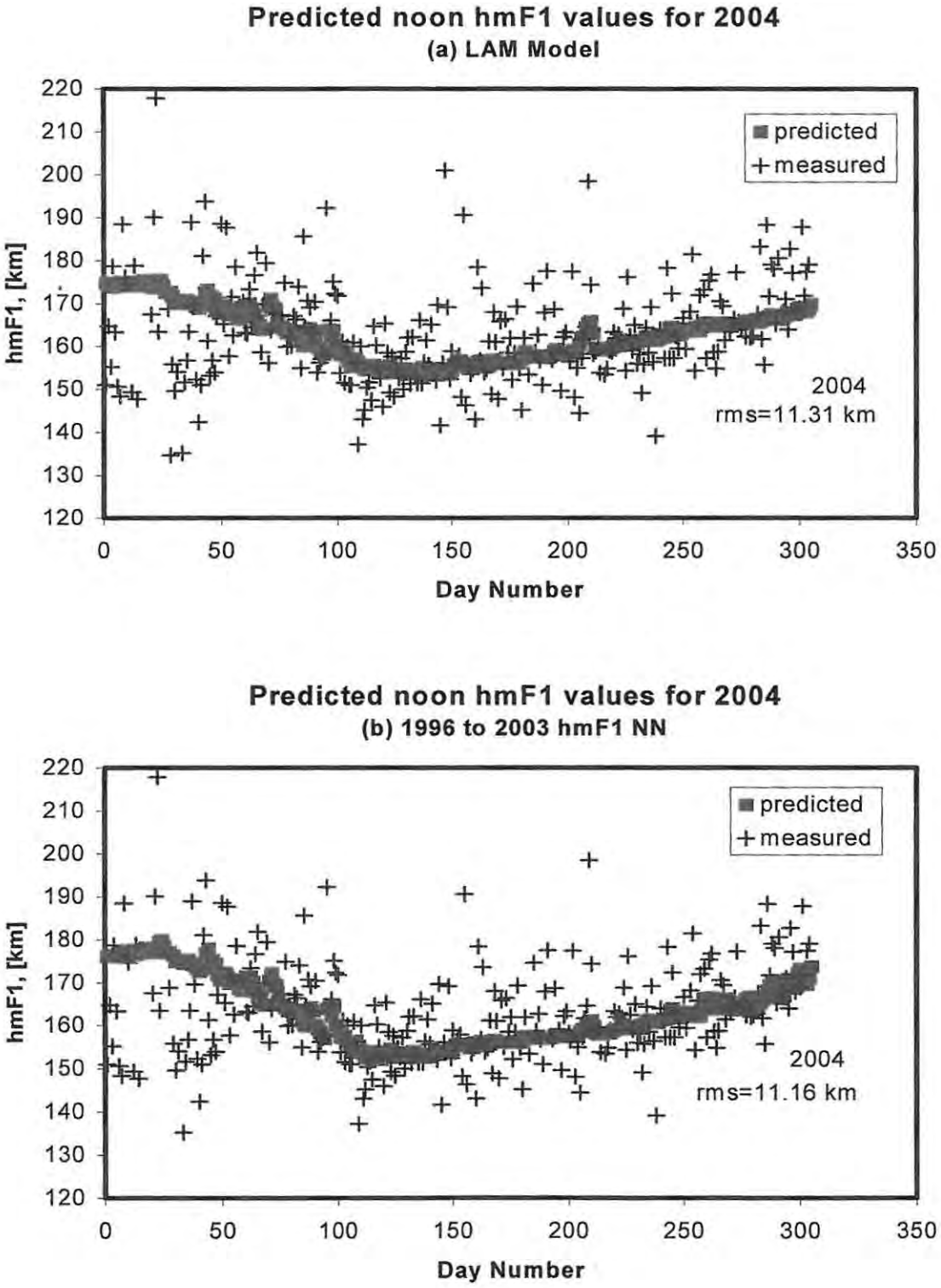


Figure 4-19: The available 2004 measured and predicted 12h00 SAST hmF1 values for (a) the LAM model and (b) the hmF1 NN. The rms errors are indicated on the graphs.

Figure 4-19 shows the measured and predicted 12h00 SAST hmF1 values for 10 months in 2004, a year for which no data was included in the hmF1 training set. A comparison is shown between (a) the LAM hmF1 predictions and (b) the hmF1 NN predictions. The rms errors indicate that there is a slight improvement in the predictions by the hmF1 NN on the LAM model.

In order to investigate the seasonal influence on the value of hmF1, predicted values for 12h00 SAST were generated over one year at the four combinations of high and low R2 and A8, shown in figure 4-20. From the graphs it can be seen that there is a much greater change in the value of hmF1 with season at high R2 than at low R2. The faded lines at high R2 are those areas for which the 1973 to 2003 Probability NN predicted the probability of the F state to be lower than 0.5 and no definite F1 layer exists.

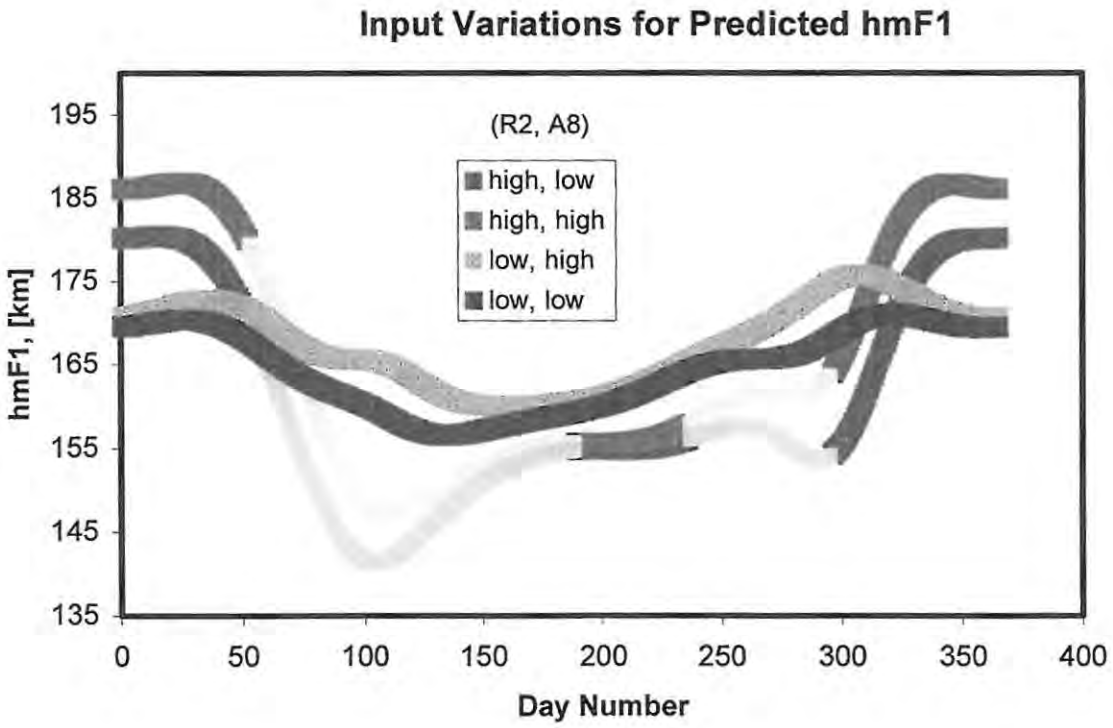


Figure 4-20: Predicted hmF1 values for 12h00 SAST over one year at the four combinations of high and low R2 and A8 are illustrated in this graph. The faded areas at high R2 are those days for which an L-condition is most probable.

As with the foF1 prediction NNs, the response of hmF1 to the changing input parameters of R2 and A8 was investigated. Figure 4-21 shows the response to increasing R2 during summer and winter and at both high and low values of A8. Unlike foF1, which increases under all conditions with increasing R2, hmF1 decreases with increasing R2 during winter months. The contribution of A8 is most noticeable in summer at higher values of R2. The gaps in the winter graphs are due to those areas for which the probability of the F1 layer being in the F state is less than 0.5, as predicted from the 1973 to 2003 Probability NN.

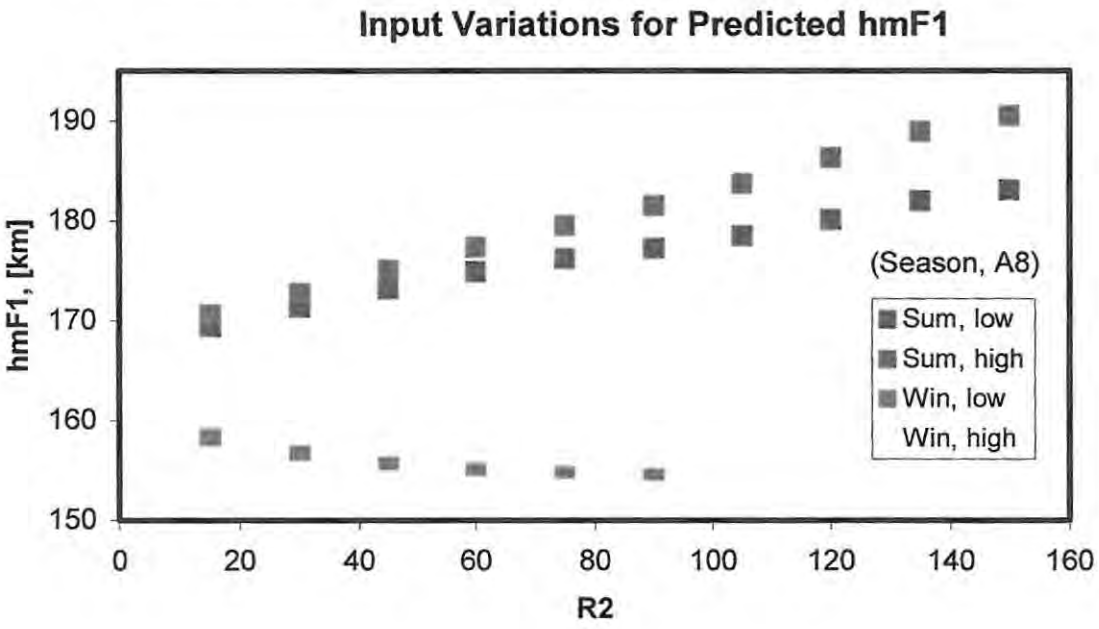


Figure 4-21: The seasonal variations in the response of hmF1 to changing solar activity at low and high levels of magnetic activity are shown.

Figure 4-22 shows the response of hmF1 at increasing values of A8 and at high and low values of R2 for summer and winter. The influence of A8 on hmF1 is most noticeable at high R2 during summer and otherwise is negligible.

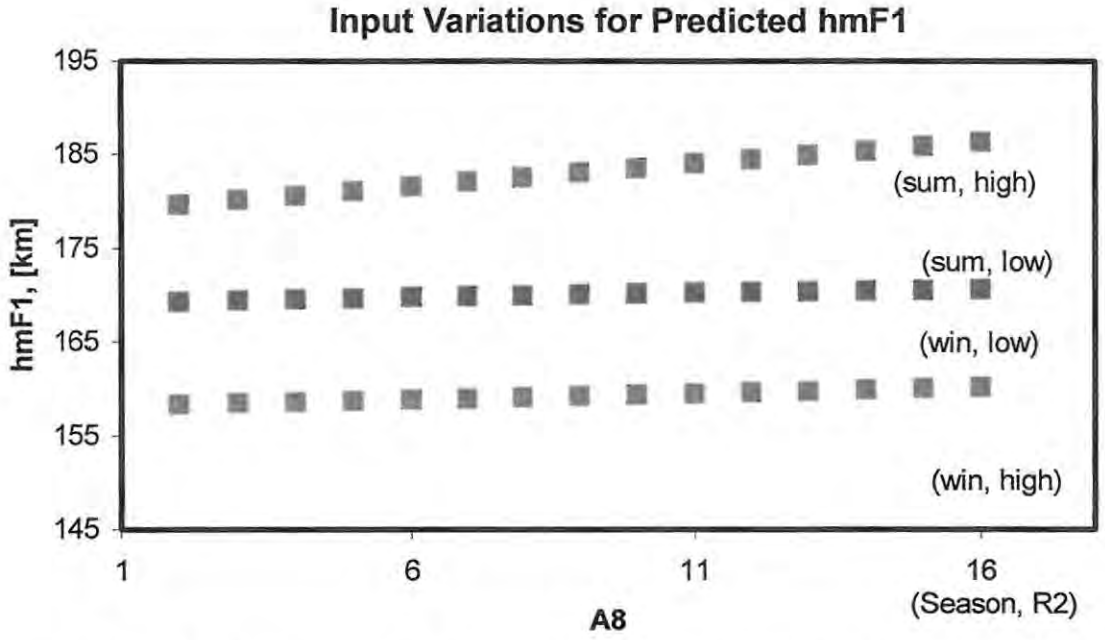


Figure 4-22: The seasonal variations in the response of hmF1 to changing magnetic activity at low and high levels of solar activity are shown.

In order to investigate the diurnal response of hmF1 to the changing input parameters, hmF1 was predicted at the four combinations of high and low R2 and A8 for summer and winter solstice. The results are shown in figures 4-23 (summer) and 4-24 (winter). The graphs show a comparison between the predicted values at high and low R2 for (a) low A8, and (b) high A8. Also shown are the comparisons between (i) the LAM model results and (ii) the 1996 to 2003 hmF1 NN results. The graphs only differ significantly during summer at high R2 and during hours less than 10h00 UT. This is most likely due to the 1996 to 2003 hmF1 NN being trained with a larger dataset, and the results from this NN are more realistic than the results obtained from the LAM model. The predictions shown in figures 4-23 and 4-24 confirm the findings shown in figures 4-21 and 4-22 that predict that during summer hmF1 is higher at high A8 than at low A8 and increases with increasing R2, and during winter there is minimal difference between the value of hmF1 at high and low A8, but the value decreases with increasing R2.

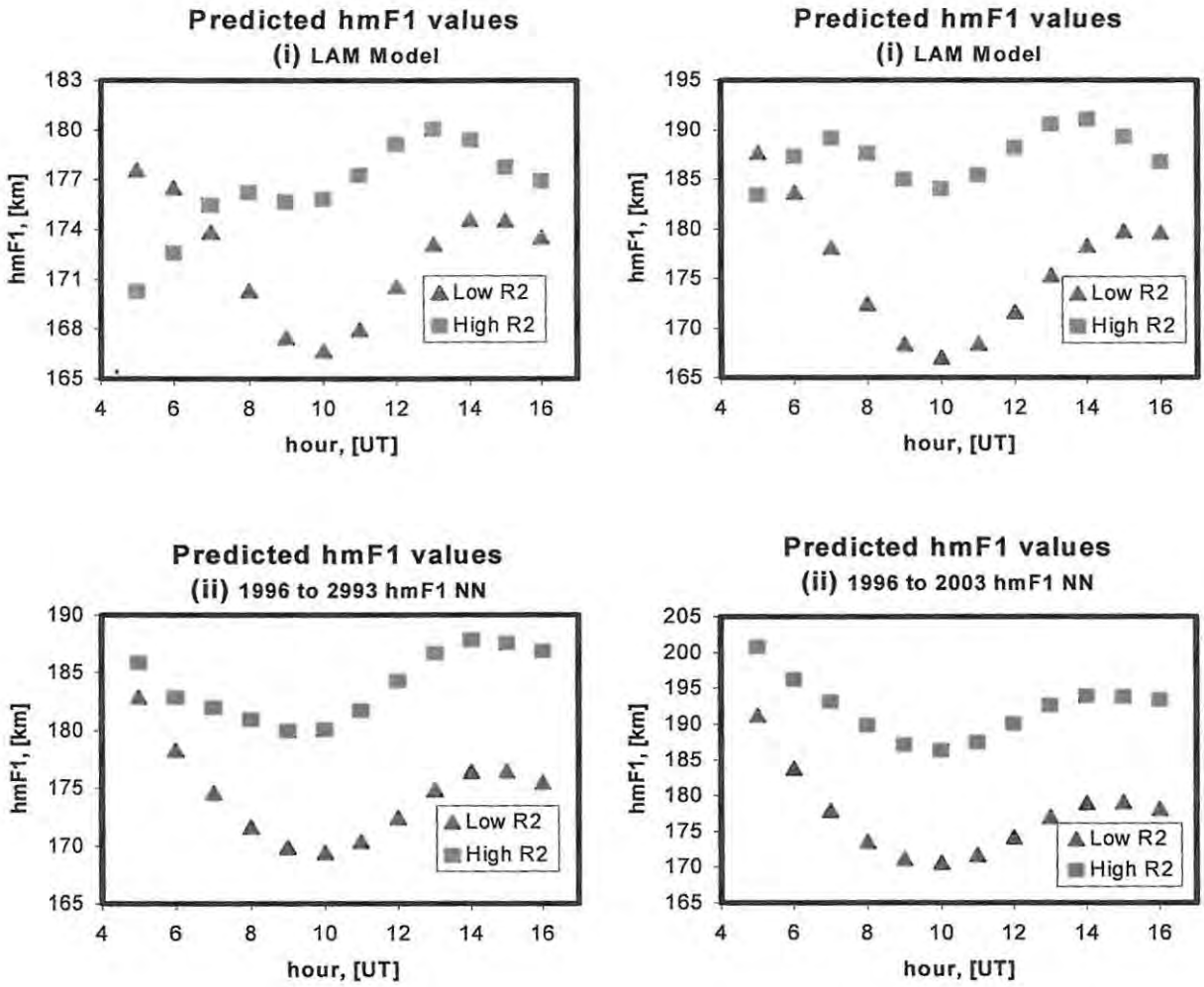


Figure 4-23: This figure shows the diurnal variation in hmF1 for a summer day at low and high values of R2 and at (a) low magnetic activity and (b) high magnetic activity. A comparison between the (i) LAM model predictions and (ii) the hmF1 NN predictions are shown.

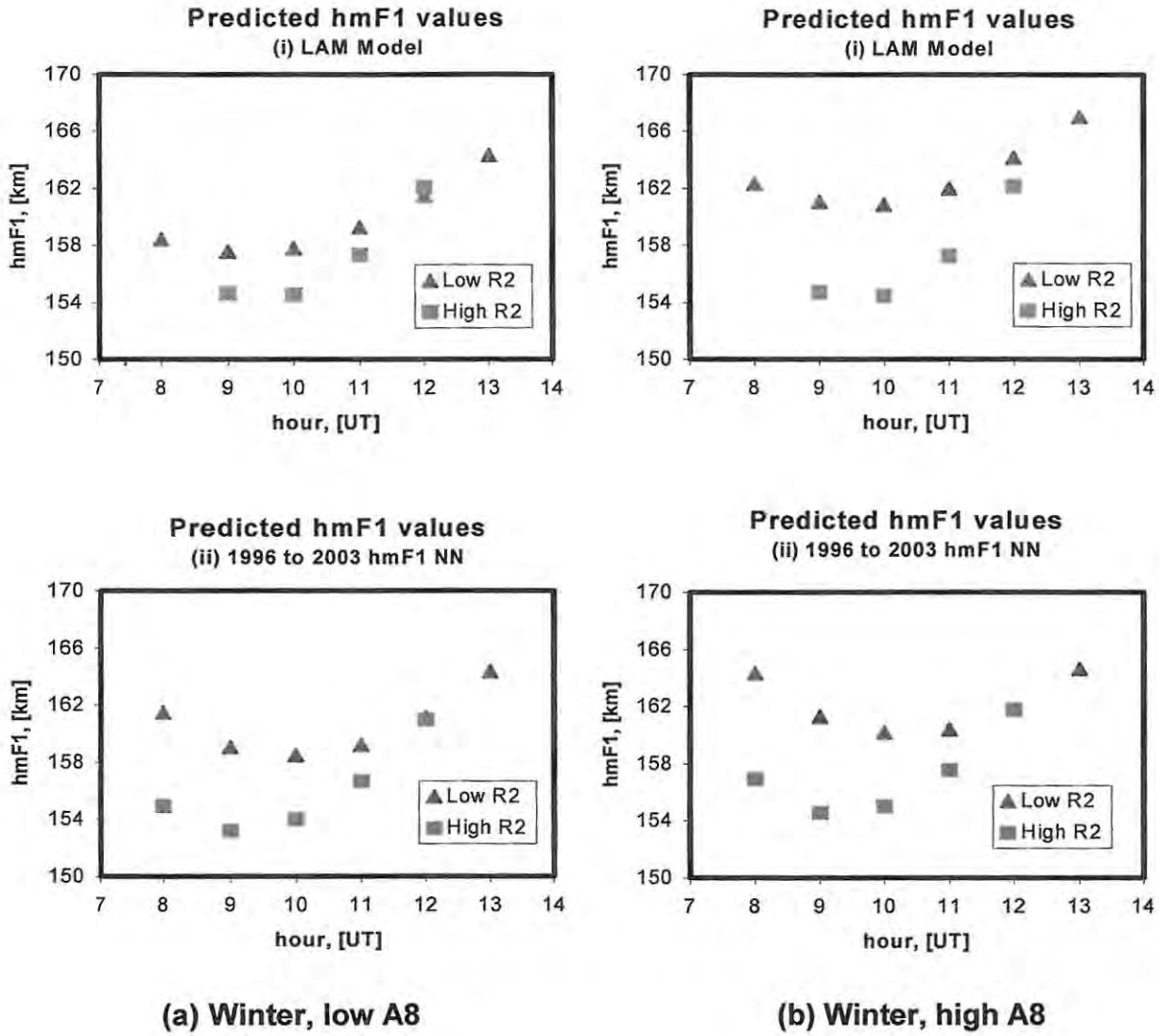


Figure 4-24: The diurnal variation in hmF1 for a winter day at low and high values of R2 and at (a) low magnetic activity and (b) high magnetic activity. A comparison between the (i) LAM model predictions and (ii) the hmF1 NN predictions are shown.

4.4 Conclusion

The results given in this chapter for the Probability NNs, the foF1 Prediction NNs and the hmF1 Prediction NN all confirm the predictions found by the LAM model. From the rms errors calculated between the measured and predicted values it was shown that by training the NNs with the increased dataset, the predictions have been improved.

Chapter 5
CONCLUSION

The LAM model is an empirical neural network (NN) based model for the entire bottomside ionosphere over Grahamstown. In developing the LAM model it was suspected *(McKinnell, [2002])* that the F1 layer was not adequately covered due to poorly scaled data obtained from the automatic scaling program ARTIST. The aim of this thesis was to update the LAM model in order to improve the NN ionospheric predictions in the F1 layer. In order to accomplish this, the automatically scaled ARTIST F1 data had to first be assessed for accuracy. Thus automatically scaled Grahamstown DPS data from 1996 to 2003 was examined and the value of foF1 was determined manually. Furthermore all instances of L-condition were manually recorded since ARTIST does not record L-condition status.

An extensive analysis of the performance of ARTIST in scaling the F1 layer over Grahamstown has been completed, and the times and conditions under which ARTIST can be considered accurate has been established. It was found that ARTIST incorrectly scales the F1 layer 29% of the time between the hours of 0330 and 1700 UT. 74% of these errors are due to L-condition being present. L-condition normally occurs in summer at the hours of sunrise and sunset and in winter during daylight hours. This relates to the fact that the majority of the errors occurred between the hours of 04h00 and 05h30 UT and 15h30 and 16h00 UT in summer and between 08h00 and 12h30 UT in winter. The results of this F1 layer analysis over Grahamstown showed that a total reliance on the ARTIST software to successfully characterise the real height profile in the F1 region would be impractical.

Therefore, for ray tracing purposes, it has been proposed that a mix of real time ARTIST-derived profiles with back-up model profiles be used.

The instances of L-condition identified during the re-scaling process were used to train an F1 probability NN using data from 1973 to 2003. This NN predicts the probability of the F1 layer being in each of the three states (F, N, L) at any given day, hour and solar activity period. The results from this NN closely resemble both the LAM model predictions and the trends found for L-condition occurrence established from the analysis.

The re-scaled foF1 values were used to train an foF1 prediction NN using data from 1973 to 2003. Figure 5-1 shows the predicted foF1 values from this 1973 to 2003 foF1 NN (foF1 NN1) versus the LAM model predicted values for 12h00 SAST for the years 1996 to 2003. The value of R^2 gives an indication as to how well the data approximates a straight line. The higher the R^2 value (with a maximum of 1), the better the fit. From this graph and the R^2 value it is apparent that the foF1 NN1 predictions closely resemble the LAM model predictions. However, it has been shown in Chapter 4 that predictions obtained from this NN produce an improvement in the rms errors between the measured and predicted foF1 values over the LAM model predictions.

Results obtained from training both a probability NN and an foF1 NN using data from only 1996 to 2003 gave some confidence when training the hmF1 prediction NN using data from only this period. Figure 5-2 shows the results from the 1996 to 2003 hmF1 prediction NN versus the predictions from the LAM model for 12h00 SAST for the years 1996 to 2003.

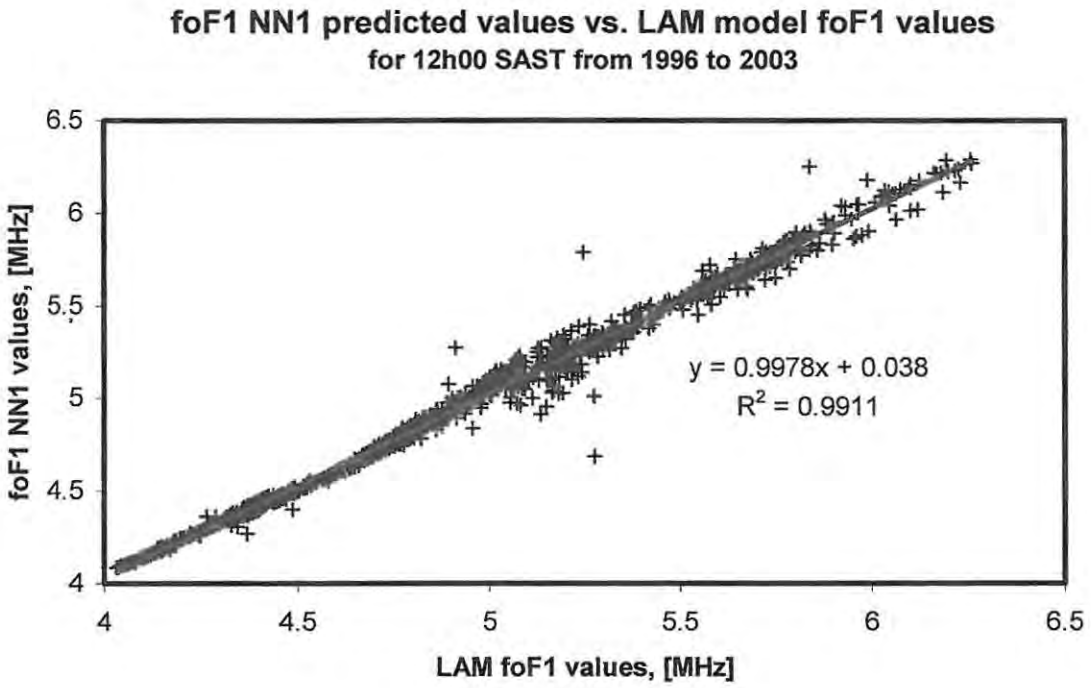


Figure 5-1: A graph of the predicted foF1 values from the 1973 to 2003 foF1 NN versus the predictions from the LAM model for 12h00 SAST from the years 1996 to 2003.

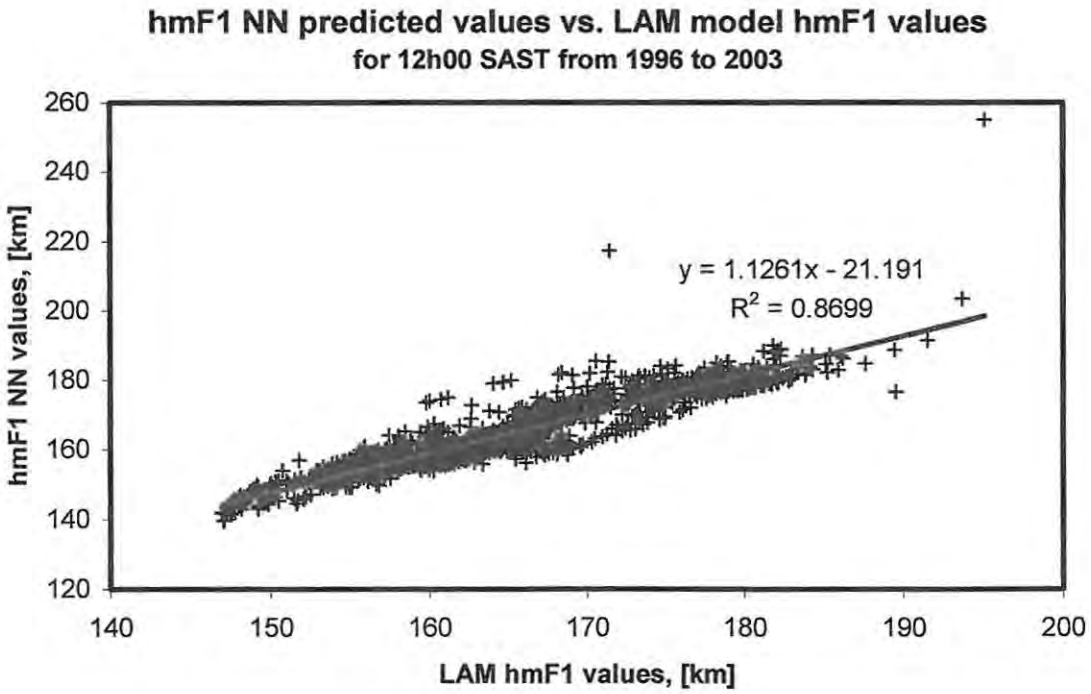


Figure 5-2: A graph of the predicted hmF1 values for 12h00 SAST for the years 1996 to 2003 obtained from the hmF1 NN versus the predicted hmF1 values from the LAM model.

From figure 5-2 it can be seen that there is less coherence between predictions from the LAM model and the hmF1 NN than from the foF1 NNs. However, results from Chapter 4 indicate that the hmF1 NN gives better predictions than the LAM model. This was expected, and indeed hoped for, as more data was used in this NN training than in the current version of the LAM model.

In conclusion, it has been shown in Chapter 4 that by incorporating the DPS manually re-scaled data from 1996 to 2003 into the various datasets used to train the different F1 layer NNs, the predictions obtained are an improvement on the LAM model predictions. These models could further improve the IRI predictions for foF1 and hmF1 in the Southern African region. Furthermore, the NNs can be used to improve ray tracing results for GrinTek Ewation.

Future work includes updating the LAM model as more data becomes available. It is also intended that data from the other two South African stations, Madimbo and Louisvale, be incorporated in order to produce a model that covers the whole of South Africa. This would require the DPS data from these stations to be manually re-scaled, as was done for the Grahamstown station, as currently no L-condition information has been collected from Louisvale and Madimbo. In addition, using the information recorded in this thesis, a remedial algorithm could be developed to modify the F1 layer produced from real-time DPS measurements to more accurately reflect true F1 layer status. This would ensure more realistic profiles being available for real-time applications.

REFERENCES

Baker D. C., Ionosonde stations in Southern Africa – A review of current status and future prospects, retrieved 2005 at URL http://www.ips.gov.au/IPSHosted/INAG/uag-104/text/baker, 1998.

Baker D. C., and Burden J. J., The multisegmented parabolic model for ionospheric electron density distribution: A convenient model for HF engineering applications, *Proceedings of the Fifth International Conference on HF Radio Systems and Techniques*, Edinburgh, 22-24 July, 1991.

Bilitza D., and Rawer K., New options for IRI electron density in the middle ionosphere, *Adv. Space Res.*, **10**, 11, 7-16, 1990.

Bilitza D., International Reference Ionosphere – Status 1995/1996, *Adv. Space Res.*, **20**, 9, 1751-1754, 1997.

Bilitza D., *International Reference Ionosphere*, National Space Science Data Center, 1990.

Bilitza D., IRI-2001, at URL http://umlcar.uml.edu/IRI-2001/, retrieved 2004.

Chen J., Bennett J. A., and Dyson P. L., Synthesis of oblique ionograms from vertical ionograms using quasi-parabolic segment models of the ionosphere, *J. Atmos. Terr. Phys.*, **52**, 277-288, 1990.

Coetzee P.J., Applications of the IRI in Southern Africa, *Adv. Space Res.*, **34**, 9, 2075-2079, 2004.

Davies K., *Ionospheric radio*, Peter Peregrinus Ltd, 1990.

DuCharme E.D., Petrie L.E., and Eyfrig R., A method for predicting the F1 layer critical frequency, *Radio Science*, **6**, 369-378, 1971.

DuCharme E.D., Petrie L.E., and Eyfrig R., A method for predicting the F1 layer critical frequency based on Zurich smoothed sunspot number, *Radio Science*, **8**, 837-839, 1973.

Fausett L., *Fundamentals of neural networks*, Prentice-Hall Inc., 1994.

Galkin I.A., *Standard Archiving Output (SAO) format*, at URL http://ulcar.uml.edu/~iag/SAO-4.htm, retrieved 2003.

Hargreaves J.K., *The upper atmosphere and solar-terrestrial relations*, Van Nostrand Reinhold Co. Ltd., 1979.

Haykin S., *Neural networks, a comprehensive foundation*, McMillan Publishing Company, Inc., 1994.

Huang X., and Reinisch B.W., Vertical electron density profiles from the digisonde network, *Adv. Space Res.*, **18**, 6, 121-129, 1996.

Jacobs L., Poole A.W.V., and McKinnell L.A., An analysis of automatically scaled F1 layer data over Grahamstown, South Africa, *Adv. Space Res.*, **34**, 9, 1949-1952, 2004.

McKinnell L.A., *A new empirical model for the peak ionospheric electron density using neural networks*, Thesis approved for the Master of Science degree of Rhodes University, Grahamstown, 1996.

McKinnell, L.A., *A neural network based ionospheric model for the bottomside electron density profile over Grahamstown, South Africa*, Thesis approved for the Doctor of Philosophy degree of Rhodes University, Grahamstown, 2002.

McKinnell L.A., and Poole A.W.V., Predicting the Ionospheric F layer using neural networks, *Journal of Geophysical Research*, **109**, A08308, doi: 10.1029/2004JA010445, 2004.

McNamara L.F., and Reinisch B.W., Observations of the mid-latitude F1 region, *Adv. Space Res.*, **15**, 2, 67-75, 1995.

McNamara L.F., *The ionosphere: communications, surveillance, and direction finding*, Krieger Publishing Company, 1991.

Poole A.W.V., and McKinnell L.A., On the predictability of foF2 using neural networks, *Radio Science*, **35**, 1, 225-234, 2000.

Ramakrishnan S., and Rawer K., Model electron density profiles obtained by empirical procedures, *Space Research* **XII**, 1253-1259, Akademie-Verlag, Berlin, German Democratic Republic, 1972.

Reinisch B.W., and Huang X., Automatic calculation of electron density profiles from digital ionograms 3. processing of bottomside ionograms, *Radio Science*, **18**, 3, 477-492, 1983.

Rishbeth H., and Garriott O.K., *Introduction to ionospheric physics*, Academic Press Inc., London, 1969.

SNNS, Stuttgart Neural Network Simulator user manual, Version 4.2, *University of Stuttgart, Institute for Parallel and Distributed High Performance Systems (IPVR)*, 1995a.

SNNS, Stuttgart Neural Network Simulator, retrieved 2003, at URL http://www-ra.informatik.uni-tuebingen.de/SNNS/, 1995b.

University of Massachusetts Lowell Center for Atmospheric Research (UMLCAR), *Digisonde Portable Sounder (DPS) system technical manual*, Version 4.0, 1996.

Wakai N., Ohyama H., and Koizumi T., *Manual of ionogram scaling*, Radio Research Laboratories, Ministry of Posts and Telecommunications, Japan, 1985.

Williscroft L.A., and Poole A.W.V., Neural networks, foF2, sunspot number and magnetic activity, *Geophys. Res. Let.*, **23**, 24, 3659-3662, 1996.