Neural network-based ionospheric modelling over the South African region

Lee-Anne McKinnell[a,b*] **and Allon W.V. Poole**[a]

During the past decade, South African scientists have pioneered research in the field of ionospheric modelling using the technique of neural networks (NNs). Global ionospheric models have always been insufficient for the South African region owing to an historical paucity of available data. Within the past 10 years, however, three new ionospheric sounders have been installed locally and are operating continuously. These sounders are located at Grahamstown (33.3°S, 26.5°E), Louisvale (28.5°S, 21.2°E) and Madimbo (22.4°S, 30.9°E). The addition of a modern sounder at Grahamstown enlarged the ionospheric database for this station to 30 years, making this archive a considerable asset for ionospheric research. Quality control and online availability of the data has also added to its attraction. An important requirement for empirical modelling, but especially for employing NNs, is a large database describing the history of the relationship between the ionosphere and the geophysical parameters that define its behaviour. This review describes the path of South African ionospheric modelling over the past 10 years, the role of NNs in this development, the international collaborations that have arisen from this, and the future of ionospheric modelling in South Africa.

Introduction

The Hermann Ohlthaver Institute for Aeronomy (HOIA) at Rhodes University has focused much of its efforts over the past decade on ionospheric modelling. The area of interest has been the bottomside ionosphere, which forms that region of the upper atmosphere between about 90 km and 350 km. Measurements from the bottomside ionosphere are recorded by means of ground-based ionosondes, of which South Africa currently has three. These three devices are all digital pulse sounders, more commonly known as DPS systems or digisondes, manufactured by the University of Massachusetts Lowell Center for Atmospheric Research (UMLCAR). They are located at Grahamstown (Eastern Cape, 33.3°S, 26.5°E), Louisvale (Northern Cape, 28.5°S, 21.2°E) and Madimbo (Limpopo, 22.4°S, 30.9°E) (Fig. 1).

Currently, a commonly used global ionospheric model is the International Reference Ionosphere (IRI), which provides predictions for all possible upper-atmospheric parameters between 60 km and 1000 km.[1] The IRI model was established in the 1960s and is updated annually during special IRI workshops. At these workshops the IRI working group meets and discusses improvements to the model. The working group consists of scientists, including both authors of this paper, from various countries. The 2003 workshop was held at Rhodes University in Grahamstown.

Over the past 10 years, HOIA has contributed to the field of ionospheric modelling, in particular, by introducing the technique of employing neural networks (NNs). Owing to an historical paucity of available data for the southern African region, the IRI model is known to be inaccurate in this area. HOIA is providing new model solutions as well as an increased South African ionospheric database, both of which can be incorporated into future versions of the IRI.

Within South Africa, the main application for an accurate ionospheric model is in direction-finding systems.[2] These systems use a technique called single station location (SSL) for determining the position of an HF transmitter by means of ray tracing. SSL is dependent on radio waves being reflected by the ionosphere and, therefore, an accurate ionospheric model is essential. For more information on ray tracing and SSL, the reader is referred to McNamara.[3] The models developed by HOIA will be provided for SSL applications as well as to the ionospheric community for future research.

In this paper we review the path of South African ionospheric modelling over the past 10 years and the role of NNs in this development.

Neural networks

Neural networks have proved to be an ideal tool for the prediction of ionospheric behaviour, which is by nature highly non-linear. A major advantage of using NNs for the prediction of ionospheric behaviour over analytical methods is that no previous knowledge of the nature of the non-linear relationships is required. Briefly, an NN is a computer program that is trained by presenting to its input any number of multidimensional input vectors that correspond to a known measured output parameter. The NN learns to identify the relationship between the input vectors and the observed output.

The most essential requirement for training an NN is a large archived database describing the history of the relationship between the input and output parameters. This database usually takes the form of several input vectors, each with a corresponding output.

After determining the input space, the network architecture is designed to consist of at least three layers: an input, a hidden layer, and an output layer. Each layer contains a number of nodes; the input and output nodes correspond to the parameters within the known database, whereas a number of hidden

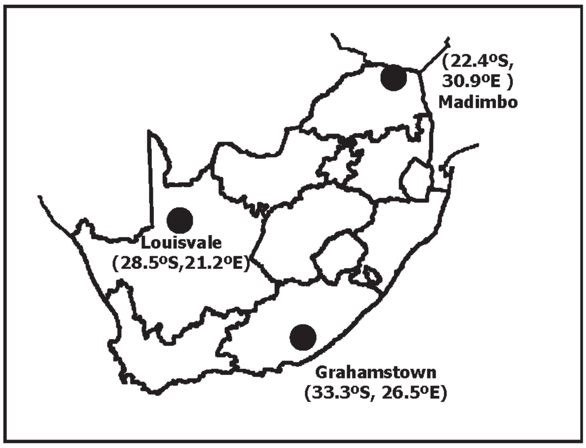

Fig. 1. Locations of the three ionospheric sounders in South Africa.

[a]Hermann Ohlthaver Institute for Aeronomy, Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa.
[b]Hermanus Magnetic Observatory, Hermanus, South Africa.
*Author for correspondence. E-mail: l.mckinnell@ru.ac.za

nodes are determined iteratively from the performance of the training procedure. There are connections between the nodes, which represent the feeding of the output from one node to the other, multiplied by a weight. The optimum architecture for the NN is determined by trial and error.

Before presenting the data to the network, the dataset is randomly split into training and testing subsets. This procedure is performed in order to prevent the training results from being biased towards a particular section of the database. In general, for the NNs trained for ionospheric modelling, 70% of the database is used for training and 30% for testing.

Training the NN is an iterative process that starts with randomly chosen weights in the NN model. The input vectors are ordered randomly, and then each is presented in turn to the network. In each case, an output is produced, which is compared to the measured output. An algorithm is then applied to update the weights in such a way as to minimize the difference. This process is repeated until the root mean square error between the given and predicted outputs on the testing set is stabilized. We have found a feedforward backpropagation algorithm to be optimum for ionospheric models.[4] Other groups who have employed NNs for ionospheric prediction purposes include Altinay *et al.*,[5] Kumluca *et al.*,[6] and Wintoft and Cander.[7] For more details on the workings of NNs, the reader is referred to Haykin.[8]

The data

As mentioned above, ionospheric data are currently being recorded at three South African field stations. To date, all of the ionospheric modelling has been accomplished using data from the Grahamstown station, because it holds the largest archived database of ionospheric data. Before 1996, a Barry Research Vertical Chirpsounder[9] was operated at this station, with the collected data being manually scaled. The data from the DPS system have been added to these manually scaled data, providing a database of 30 years' worth of ionospheric characteristics. The older data provided only virtual height (as opposed to real height) information and so the database of electron density profile information extends only from 1996. The quantity and quality of the Grahamstown archive makes it a valuable asset for ionospheric modelling.

For the usual research and practical purposes, the ionosonde is operated in a vertical incidence mode, where the transmitter and receiver are co-located. All of the ionospheric data mentioned above were collected in this vertical mode and provide information on the ionosphere above the Grahamstown area.

Each of the three South African DPS systems is set to perform a vertical incidence sounding on a continuous basis every half-hour; the resulting data are scaled using the UMLCAR automatic scaling software, Artist4.[10] All of the data collected from these sounders are archived at Rhodes University.

The DPS system can also be operated in drift mode, which is a scientific research mode that allows the calculation of precise angle of arrival and Doppler shift information from ionospheric echoes.[11] The Grahamstown DPS has operated in both vertical incidence and drift modes since 1999, and the data archived from the drift mode have allowed additional investigations into ionospheric tilts to be undertaken.

Ionospheric modelling

Predicting foF2

An important parameter in ionospheric modelling is the maximum electron density in the ionosphere, which is directly related to the measurable parameter foF2. This metric corresponds to the maximum frequency at which radio waves can be reflected at vertical incidence from the ionosphere and is measured in MHz. Within the Grahamstown database, 30 years of scaled foF2 values exist, from 1973 to 2003.

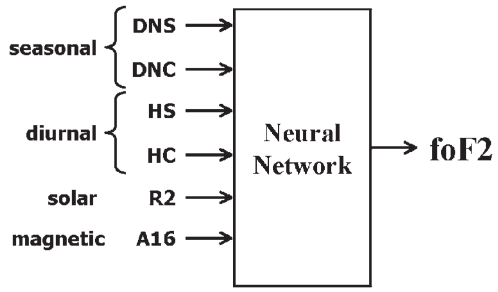

Fig. 2. A block diagram illustrating the inputs and output to a neural network trained to predict the Grahamstown foF2 value.

The value of foF2 is dependent on the season, time of day, solar activity and magnetic activity. Initially, NNs were employed to predict the 12:00 SAST (South African Standard Time) foF2 value for Grahamstown. This work was demonstrated, for the first time, in Williscroft and Poole,[4] and again, in more detail, in McKinnell.[12] These publications served to prove that NNs could be employed successfully for the prediction of ionospheric parameters. Comparisons with the IRI global model showed that, for Grahamstown, an NN-based model predicted the noon value of foF2 more realistically than the IRI.

This work was expanded to include all hours and the results were presented in Poole and McKinnell,[13] where a section on NNs for short-term forecasting of foF2 was also included. In Williscroft and Poole[4] and McKinnell,[12] it was shown how NNs could also be used to determine the optimum input parameters required for predicting a particular output. For the prediction of foF2, it was found that the optimum inputs representing the solar and magnetic variations were a 2-month running mean value of the daily sunspot number (R2), and a 2-day running mean value of the magnetic a_k index (A16). The magnetic index used was the local magnetic index from the Hermanus Magnetic Observatory. Although the latitude is also important for the prediction of foF2, it was not considered as an input at this stage as only Grahamstown data were used.

The seasonal and diurnal variations were represented by the quadrature components of the day number (DN) and the hour (HR). These components each form a separate input parameter and are defined as follows:

$$\mathrm{DNS} = \sin(2\pi \mathrm{DN}/365) \quad (1) \qquad \mathrm{HS} = \sin(2\pi \mathrm{HR}/24) \quad (3)$$

$$\mathrm{DNC} = \cos(2\pi \mathrm{DN}/365) \quad (2) \qquad \mathrm{HC} = \cos(2\pi \mathrm{HR}/24) \quad (4)$$

As an example, a block diagram showing the inputs and output to the NN trained to predict foF2 is shown in Fig. 2. To demonstrate the results from this Grahamstown foF2 model, Fig. 3 shows the measured and predicted 12:00 SAST foF2 values for a year of solar minimum (1996) and a year of solar maximum (2000).

The idea of providing an NN-based short-term foF2 forecast program was developed further; the details and results are presented in McKinnell and Poole.[14] In this paper, an NN was trained to predict the value of foF2 1, 2, 3, 4 and 25 hours ahead. This NN made use of past values of foF2 as part of the input space. If recent ionospheric data are available, therefore, it is possible to provide a prediction for the hours ahead. This paper also provided details of the development of an error network. The same input space is used to train an NN with the squared error, which is calculated by taking the square of the difference between the measured and predicted output parameters. Therefore, an estimate of the uncertainty over a prediction can be provided, where the uncertainty is dependent on the same inputs as the prediction. This is an extremely powerful feature of modelling with NNs, and was first reported in ref. 13.

In keeping with this line of work on the foF2 parameter, investigations were undertaken to determine the long-term trends in foF2 using NNs. Poole and Poole[15] report on the NN techniques used to isolate long-term variations from those due to geophysical dependencies. It was found that the long-term change in foF2 for Grahamstown is extremely linear, and negative for most hours and days. As an illustration of this result, Fig. 4 shows a graph of the difference between measured and predicted foF2 values (DfoF2) versus year. All 12:00 SAST values from 1973 to 2000 are shown. A trend line has been fitted to the data, the slope of which provides the average rate of change of DfoF2 with time.

Currently, a global foF2 model using NNs is being developed,[16] which makes use of a number of large ionospheric databases that exist for stations around the world. There is a need for a new foF2 global model that allows for latitudinal variations and is able to make accurate predictions for grid points between available stations. First results from this new model were presented at the 2003 IRI workshop, and the IRI community has expressed interest in including this model in future versions of the reference ionosphere.

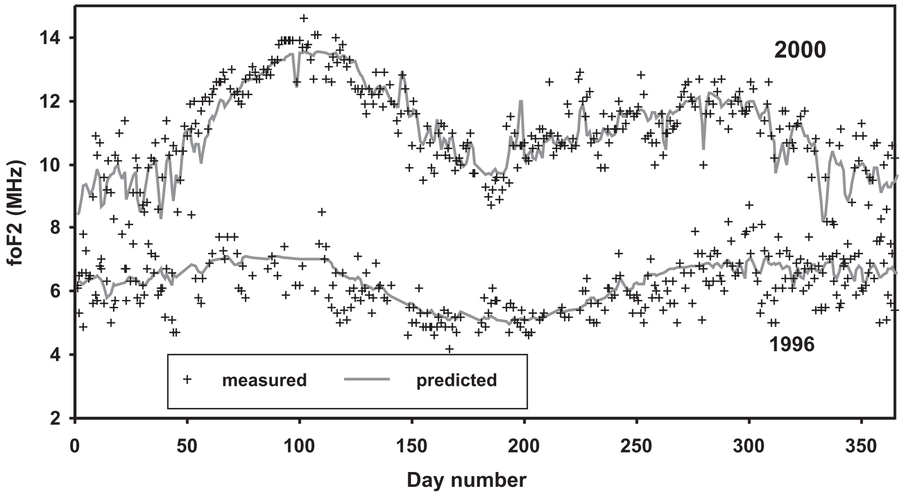

Fig. 3. Measured and predicted foF2 values for a year of solar minimum (1996) and a year of solar maximum (2000). The predictions resulted from the NN-based model.

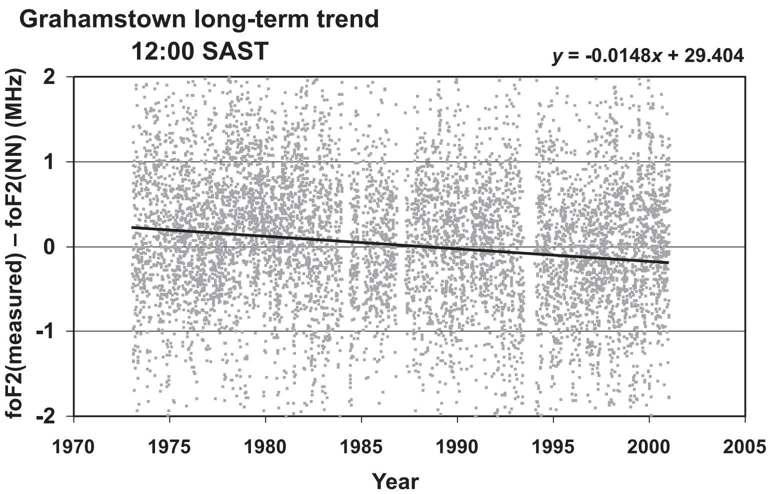

Fig. 4. Illustration of the average rate of change of foF2 with time. The difference in measured and predicted foF2 is plotted against year for all 12:00 SAST data for Grahamstown from 1973 to 2000.[15]

The LAM model

A new local ionospheric model for the complete bottomside electron density profile over South Africa was presented in 2002. This development, named the LAM model, involved using NNs to predict the parameters required for constructing the profile.[17] It is currently a single station model, since only Grahamstown data were included in its development.

To develop the LAM model, the electron density profile was split into two layers (E and F). Grahamstown DPS-recorded and Artist-scaled electron density profile data were used for training the NNs that formed the basis for the model. As part of the output from Artist, a description of the electron density profile is provided as sets of Chebyshev coefficients for each layer. These coefficients describe the shape of the profile for that layer. The peak parameters of each layer, and the E–F valley width are also provided. Five years of Artist-scaled data with electron density profiles, and 28 years of critical frequency data, were available for this development.

Details of the calculation of the E layer and F layer contributions to the LAM model are presented in McKinnell and Poole,[18] and McKinnell and Poole,[19] respectively. Although these particulars will not be described here, some important features will be noted. The construction of the LAM model-predicted profile makes use of a probability NN, which constituted a pioneering initiative in employing neural networks to determine the probability of existence of an F1 layer. For each layer, the predicted coefficients are used in conjunction with the predicted peak parameters in an analytical expression to determine the real height at any given frequency.[10] The valley region used by the UMLCAR model[10] was adopted to provide the E–F boundary and a smoothing technique for modifying the F1–F2 boundary was developed in order to provide a smooth and continuous profile. A block diagram illustrating the process that the LAM model follows in predicting the electron density profile is shown in Fig. 5.

The LAM model provides a realistic electron density profile that is representative of the average behaviour of the ionosphere over Grahamstown for a particular set of inputs, as well as an estimate of the uncertainty over the predicted profile. Predicted profiles from this model were compared with real DPS data and with profiles obtained from the IRI model. Six examples of these comparisons are shown in Fig. 6, and demonstrate that the NN-based LAM model is more successful at predicting the electron density profile for a particular set of inputs than the IRI.

Ionospheric tilts

As mentioned above, the Grahamstown DPS currently operates in both the vertical incidence and drift modes. The drift mode

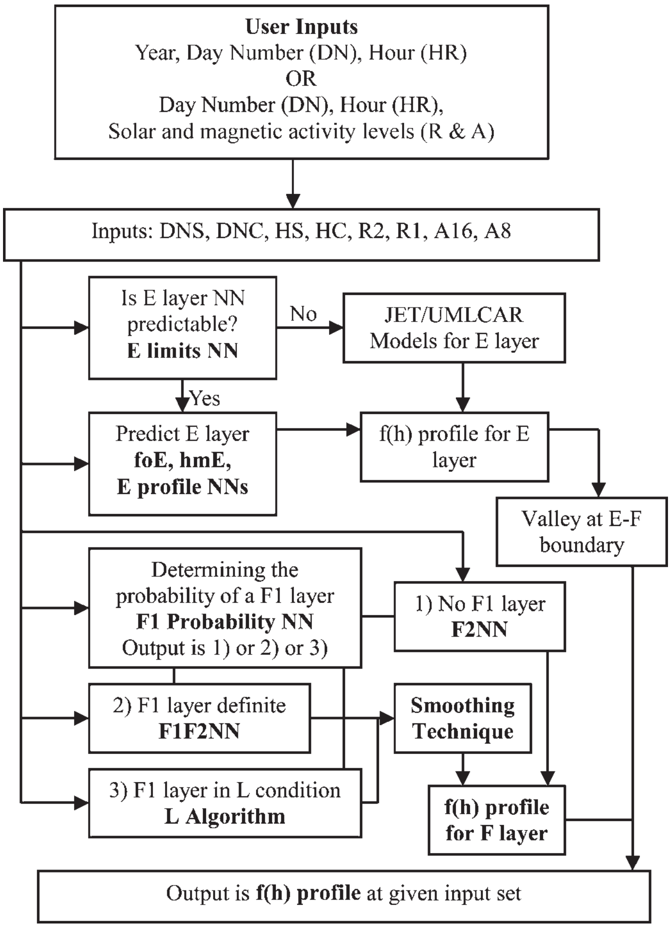

Fig. 5. A block diagram depicting the process that the LAM model follows when predicting an electron density profile for a particular set of inputs.[17]

provides information on angle of arrival, which is obtained through an interferometric receive antenna array. This information can be used to determine if there are any tilts in the ionosphere.[20] Tilt information is useful in direction-finding applications as bearing errors may be introduced if ionospheric tilts are not taken into account. Magnus[20] found a significant variation in the information on angle of arrival at sunrise, which was interpreted as a tilt variation. In addition, Magnus[20] showed that by employing NNs, future tilts could be predicted given enough training data. To achieve the larger database required, the Grahamstown DPS continues to collect drift mode data.

The F1 layer

The LAM model, which was based on Artist-scaled data for Grahamstown, highlighted a problem with the boundary between the F1 and F2 layers of an electron density profile. This problem took the form of a discontinuity at the boundary, which arose from the Chebyshev coefficient method of constructing the profile. The LAM model solved this problem by applying a smoothing technique to the electron density profile across the F1–F2 boundary.[17,19] However, this difficulty also occurs in the DPS Artist-scaled data and, since these data are being used in a real-time ray tracing application, it is necessary to find a solution. The F1 layer is particularly difficult to scale automatically because its existence is so variable. So, initially an analysis of all F1 layer Artist-scaled data was performed to provide statistics on how well Artist was scaling the F1 layer in the Grahamstown region. This analysis revealed that, taking into account all data from 1996 to 2003 that fell between 03:30 and 17:00 UT, 29% of it was incorrectly scaled by Artist in the F1 region.[21] Of the percentage that was incorrectly scaled, most fell into the area where a ledge would be expected on the ionogram, which is not taken into account by Artist. An algorithm is currently being developed that can be applied to real-time Artist data to correct for inaccuracies in the F1 region.

International collaborations

The ionospheric research team at HOIA are active participants in the IRI working group. The South African ionospheric data are archived and sent on a near real-time basis to the World Data Center in Boulder, Colorado. This makes the data more easily available to members of the ionospheric community to access for research purposes. Future versions of the IRI model will make use of these South African results. In the past, the data were not as accessible as they are today and were not quality checked. The quantity and availability of the South African observations has improved substantially over the last 10 years. There is now a node in the form of the Space Physics and Interactive Data Resource at Rhodes University, which serves to expedite the process of providing the World Data Center with ionospheric data. The NN-based models developed at Rhodes will become options for the IRI global model. This will make them accessible to the entire ionospheric community and serve as future testing stations.

Recently, collaboration between the group at Rhodes and the Department of Communications and Wave Propagation at Graz University of Technology in Austria resulted in an additional model, called IMAZ (for Ionospheric Model for the Auroral Zone). IMAZ is an NN-based ionospheric model for the lower ionosphere, which was built primarily using data from the European Incoherent Scatter Radar (EISCAT) facility, located in Tromsø, Norway. It was developed specifically for the D and E regions of the ionosphere (70–150 km). More details on this model can be found in refs 22 and 23.

Conclusion and future work

NNs have proved to be successful tools for ionospheric modelling, and have had a significant impact on the way in which models for the South African region are developed. The ease with which NNs can be re-trained should additional data become available is one of the many advantages of this method. The requirement for a good reliable ionospheric model for the South African region can be met and constantly improved over the years to come as more observations are recorded and added to the database.

Future work includes completing the global foF2 model and updating the LAM model. The global foF2 model will be incorporated into future versions of the IRI as a replacement for the current foF2 model, which, due to a paucity of data in the past, contains inaccuracies in the prediction of foF2 in the southern hemisphere. A plan for the LAM model is to include data collected at the Louisvale and Madimbo ionospheric stations and re-train all the NNs, thereby expanding the model to be a more representative South African one. Current work on the F1 region will also be included to improve local predictions. In addition, an algorithm is to be developed to correct scaled data for this region and to provide more realistic, real-time electron density profiles.

In conclusion, over the past decade the ionospheric research team at Rhodes University has made a significant contribution, in respect of modelling and data accessibility, to the ionospheric research community. As well as providing data for practical

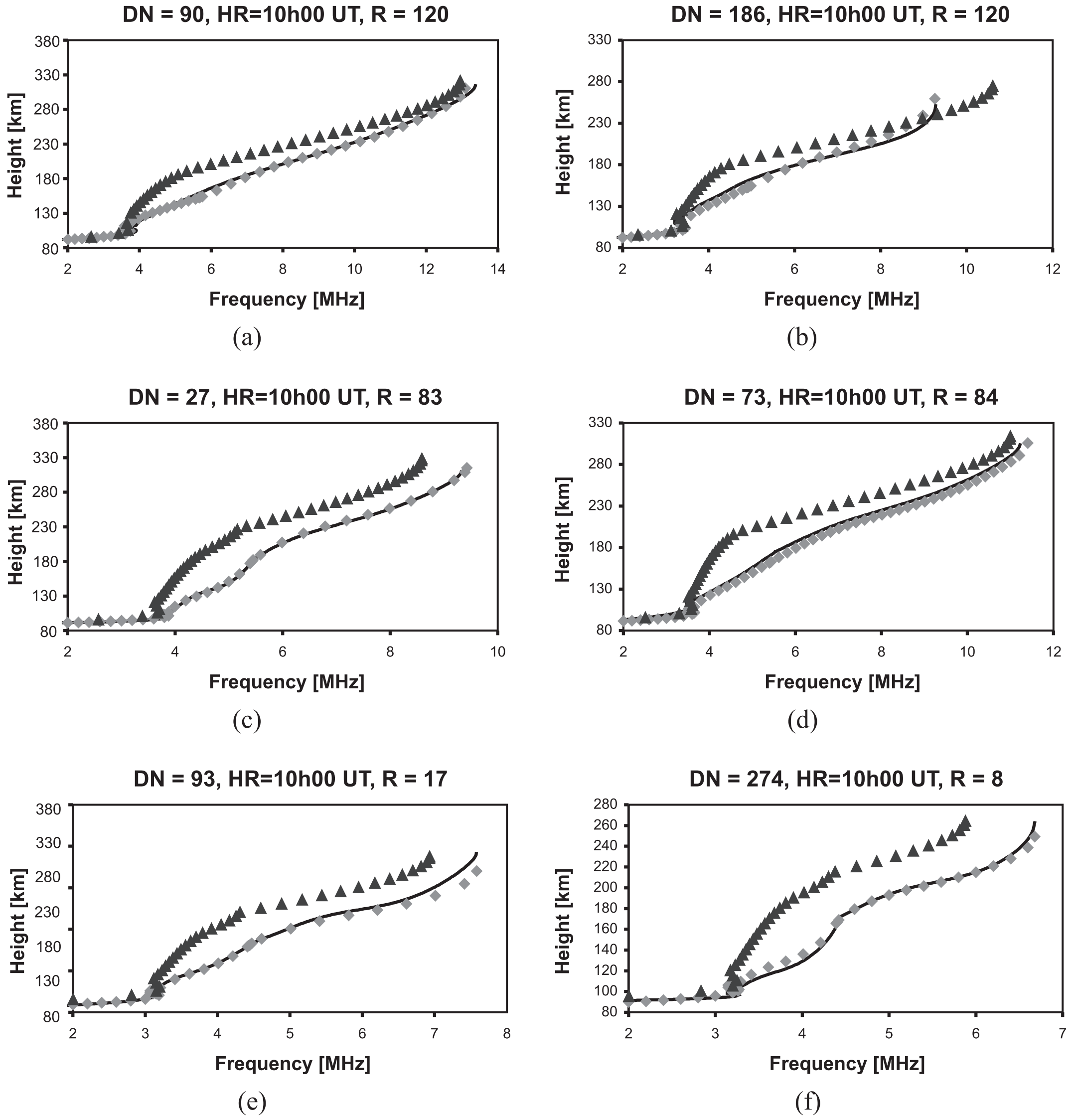

Fig. 6. Six examples of actual midday SAST DPS profiles for three levels of solar activity. The equivalent LAM model and IRI profiles for 2001 are also shown for comparison.[17] Legend: —, actual profile; ◆, LAM model; ▲, IRI profile.

purposes within the country, this work has led to international collaborations and to improving models for the South African region.

1. Bilitza D. (1990). *International Reference Ionosphere.* National Space Science Data Center, Boulder, Colorado.
2. Coetzee P.J. (2004). Applications of the IRI in southern Africa. *Adv. Space Res.* **34**(9), 2075–2079.
3. McNamara L.F. (1991). *The Ionosphere: Communications, Surveillance, and Direction Finding.* Krieger Publishing Company, Malabar, Florida.
4. Williscroft L.A. and Poole A.W.V. (1996). Neural networks, foF2, sunspot number and magnetic activity. *Geophys. Res. Lett.* **23**(24), 3659–3662.
5. Altinay O., Tulunay E. and Tulunay Y. (1997). Forecasting of ionospheric frequency using neural networks. *Geophys. Res. Lett.* **24**, 1467–1470.
6. Kumluca A., Tulunay E., Topali I. and Tulunay Y. (1999). Temporal and spatial forecasting of ionospheric critical frequency using neural networks. *Radio Sci.* **34**(6), 1497–1506.
7. Wintoft P. and Cander L.R. (1999). Short term prediction of foF2 using time delay neural networks. *Phys. Chem. Earth (C)* **24**, 343–347.
8. Haykin S. (1994). *Neural Networks, a Comprehensive Foundation.* Macmillan, New York.
9. Poole A.W.V. and Evans G.P. (1985). Advanced sounding (2): first results from an advanced chirp ionosonde. *Radio Sci.* **20**, 1617–1623.
10. Huang X. and Reinisch B.W. (1996). Vertical electron density profiles from the digisonde network. *Adv. Space Res.* **18**(6), 121–129.
11. Scali J.L., Reinisch B.W., Heinselman C.J. and Bullett T. (1995). Coordinated digisonde and incoherent scatter radar F region drift measurements at Sondre Stromfjord. *Radio Sci.* **30**(5), 1481–1498.
12. McKinnell L.A. (1996). *A new empirical model for the peak ionospheric electron density using neural networks.* M.Sc. thesis, Rhodes University, Grahamstown.
13. Poole A.W.V. and McKinnell L.A. (2000). On the predictability of foF2 using neural networks. *Radio Sci.* **35**, 225–234.
14. McKinnell L.A. and Poole A.W.V. (2000). The development of a neural network based short-term foF2 forecast program. *Phys. Chem. Earth (C)* **25**(4), 287–290.
15. Poole A.W.V. and Poole M. (2002). Long-term trends in foF2 over Grahamstown using neural networks. *Annali di Geofisica* **45**, 1.
16. Oyeyemi E. and Poole A.W.V. (2004). On the development of a global foF2 empirical model using neural networks. *Adv. Space Res.* **34**(9), 1966–1972.
17. McKinnell L.A. (2002). *A neural network based ionospheric model for the bottomside electron density profile over Grahamstown, South Africa.* Ph.D. thesis, Rhodes University, Grahamstown.
18. McKinnell L.A. and Poole A.W.V. (2003). A neural network based electron density model for the E layer. *Adv. Space Res.* **31**(3), 589–595.
19. McKinnell L.A. and Poole A.W.V. (2004). Predicting the ionospheric F layer using neural networks. *J. Geophys. Res.* **109**, A08308, doi:10.1029/2004JA010445.
20. Magnus L.G. (2001). *Expanding the capabilities of the DPS ionosonde system.* M.Sc. thesis, Rhodes University, Grahamstown.
21. Jacobs L.J., Poole A.W.V. and McKinnell L.A. (2004). An analysis of automatically scaled F1 layer data over Grahamstown, South Africa. *Adv. Space Res.* **34**(9), 1949–1952.
22. McKinnell L.A. and Friedrich M. (2003). Towards neural network based models for the ionospheric D-region. In *Proc. 16th ESA Symposium on European Rocket and Balloon Programmes and Related Research, St Gallen, Switzerland,* ESA SP-530, pp. 369–373. European Space Agency.
23. McKinnell L.A., Friedrich M. and Steiner R.J. (2004). A new approach to modelling the daytime lower ionosphere. *Adv. Space Res.* **34**(9), 1943–1948.