Annales
Geophysicae

Nonlinear forecasts of *fo*F2: variation of model predictive accuracy over time

A. H. Y. Chan and **P. S. Cannon**[*]

Centre for RF Propagation and Atmospheric Research, QinetiQ, Malvern, UK
[*]Also at the Department of Electronic and Electrical Engineering, University of Bath, UK

Abstract. A nonlinear technique employing radial basis function neural networks (RBF-NNs) has been applied to the short-term forecasting of the ionospheric F2-layer critical frequency, *fo*F2. The accuracy of the model forecasts at a northern mid-latitude location over long periods is assessed, and is found to degrade with time. The results highlight the need for the retraining and re-optimization of neural network models on a regular basis to cope with changes in the statistical properties of geophysical data sets. Periodic retraining and re-optimization of the models resulted in a reduction of the model predictive error by ~ 0.1 MHz per six months. A detailed examination of error metrics is also presented to illustrate the difficulties encountered in evaluating the performance of various prediction/forecasting techniques.

Key words. Ionosphere (ionospheric disturbances; modeling and forecasting) – Radio science (nonlinear phenomena)

1 Introduction

Variations in the solar, magnetospheric, and ionospheric characteristics can affect a variety of ground-based and space-borne technological systems (e.g. Hargreaves, 1995; Feynman and Gabriel, 2000). Disturbances in the ionosphere can degrade radio propagation and satellite communications; solar flares can cause positional errors of several kilometers in ground-based navigation systems, and the Global Positioning System (GPS) can be affected by variations in electron density in the ionosphere. Magnetic storms can induce currents in long-distance pipelines and cable networks. Magnetospheric particles and solar proton flares can affect spacecraft by causing radiation and structural damage. Consequently, predictions of geomagnetic storms have a significant bearing upon the operation of a number of services (Joselyn, 1995).

The importance of nonlinear behavior within the solar-terrestrial environment has been demonstrated by Baker et

al. (1990), and attempts have been made in modeling the nonlinear dynamics in geomagnetic activity (e.g. Klimas et al., 1992; Vassiliadis et al., 1995). However, due to the incomplete understanding of the physics of the Sun-magnetosphere-ionosphere coupled system, many theoretical and empirical models often fail to accurately predict ionospheric disturbances and geomagnetic storm events (Joselyn, 1995). An attractive alternative approach would be to adopt knowledge-independent modeling techniques that can cope with the problems of noise and non-contiguity typically found in geophysical data sets.

A number of studies have investigated the application of neural networks (NNs) to geophysical prediction problems (e.g. Lundstedt, 1992; Williscroft and Poole, 1996; Wu and Lundstedt, 1997; Francis et al., 1997, 2000; Cander et al., 1998; Wintoft and Cander, 2000). In this work, we shall refer to short-term (e.g. 1-hour ahead) predictions as forecasts, in order to distinguish these from the more commonly referred-to longer-term (e.g. monthly median) predictions. Many of the above studies have used multiple inputs. For example, Lundstedt (1992) used solar section boundary data, coronal mass ejection data, solar wind and coronal hole data to train neural networks for the prediction of a number of effects including geomagnetic induced currents, while Wu and Lundstedt (1997) used solar wind and D_{st} data as inputs to neural networks for the forecasting of geomagnetic storms. Williscroft and Poole (1996) predicted daily and monthly noon values of the ionospheric parameter *fo*F2 at Grahamstown, South Africa, using seasonal time information, solar and magnetic activities as input data. Cander et al. (1998) presented a 1-hour ahead forecasting technique for *fo*F2 and the total electron content (TEC) using input data which included *fo*F2/TEC, a daily sunspot number and D_{st} index. Wintoft and Cander (2000) used *fo*F2 data, together with magnetic activity index A_E, time-of-day and seasonal information to forecast *fo*F2 values for 1 to 24 h ahead. The above types of models are more commonly known as cross-prediction models.

In a rather different approach, Francis et al. (1997, 2000)

Correspondence to: A. H. Y. Chan (achan@QinetiQ.com)

have used only the ionospheric parameter *fo*F2 as input for their neural network models to predict *fo*F2. This is an example of a self-prediction model. Despite their simplicity, the models were shown to offer significant improvements over the performance of reference persistence and recurrence models on hourly, daily, and monthly time scales. For example, for 1-hour ahead forecasts at Slough, UK, the hourly model offers an improvement of $\sim 42\%$ and $\sim 45\%$, respectively, over the persistence and the 24-hour recurrence models. For the 1-day ahead predictions, the daily noonday model gives an enhancement in model prediction accuracies over the persistence model of $\sim 60\%$. The monthly median model shows an improvement of $\sim 40\%$ over the baseline persistence model for 1-month ahead predictions. Such a comparison with simple recurrence and persistence models is a minimum prerequisite of any rigorous assessment of new forecasting algorithms, and without it, performance claims with respect to new algorithms are rendered meaningless.

An assessment of the true value of space weather prediction schemes is, however, more complicated than a simple (albeit valuable) comparison with recurrence and persistence models. The space weather environment is nonstationary over a number of time scales, ranging from periods of days to 11 years or more. However, typical data sets are shorter than 11 years, and even when these long data sets are available, it may not be possible to undertake the necessary matrix operations on the data set as a whole (due to limited computing resources). Consequently, the neural network models and their associated error statistics usually presented are quite specific to a particular epoch. This was illustrated forcefully when a re-optimized version of the Francis et al. (2000) 1-hour ahead forecasting model was incorporated into our real-time forecasting system – the Ionospheric Forecasting Demonstrator, IFD (http://www.cpar.qinetiq.com). It soon became clear that the predictive capability of the IFD was degrading with time. In this paper, we describe a number of methods for assessing the long-term performance of NN models, and illustrate how the predictive accuracy of a NN model can be maintained in a non-stationary environment.

2   Analysis approach

Solar-terrestrial data sets are typically very noisy. The Time Series Analysis Routines (TSAR) described by Smith et al. (1998) employ novel and robust methods that can cope with the problems of noise and data dropouts. The detailed mathematical theory behind the TSAR software can be found in Smith et al. (1998). Here, we shall give a brief summary of the Radial Basis Function Neural Network (RBF-NN) model used in this study.

2.1   Principal component analysis

The use of principal component analysis (PCA) in the pre-processing of the time series allows for the separation of the signal and noise subspaces, and improves the performance

of the NNs by reducing the effects of over-fitting through the removal of the noise subspace. PCA is usually undertaken by employing linear singular value decomposition (SVD). The rows of the matrices used in the SVD calculation are obtained by sliding a window, of length n, across the time series one point at a time. A set of n orthogonal linear filters is generated, and a subset containing the principal components that adversely affect the model accuracy can then be deselected to act as a noise filter.

2.2   Radial basis function neural network

The modeling of time series data seeks to fit the input/output data points (X_n, Y_n), where $n = 1 \ldots N$, and N = total number of data points in the time series, to a model of the form $Y = f(X)$. The RBF-NN offers one approach to the solution of this problem, and has the advantage over the more commonly used Multi-Layer Perceptron (MLP) techniques of being able to find a globally optimum solution to a time series prediction problem in a single pass training process that determines the appropriate model weights (Broomhead and Lowe, 1988). (MLPs can only produce locally optimum solutions through an iterative training process.)

In the RBF approach, $f(X_n)$ is assumed to be a linearly weighted sum of radially symmetric functions of X, such that

$$f\left(X_n\right) = \sum_{i=1}^{N} \omega_i \varphi_i \left(\mid X_n - c_i \mid\right), \tag{1}$$

where ω_i are the weights of the functions φ_i, and c_i are the centers of radial symmetry.

The set of model parameters, including window length and number of centers, is optimized to give the best solution to the function $f(X_n)$ such that the error E is minimized

$$E = \sum_{n=1}^{N} \left[f\left(X_n\right) - Y_n\right]^2. \tag{2}$$

The effectiveness of the prediction model can be quantified in terms of the normalized root-mean-squared error (NRMSE), which is essentially the root-mean-squared error (RMSE) divided by the standard deviation, σ, of the input data

$$\text{NRMSE} = \sqrt{\frac{\sum\limits_{n=1}^{N} \left[f\left(X_n\right) - Y_n\right]^2}{\sum\limits_{n=1}^{N} \left[X_n - \overline{X}\right]^2}} = \frac{\text{RMSE}}{\sigma}, \tag{3}$$

where $\overline{X}$ is the mean value of $X_n, n = 1 \ldots N$. A NRMSE of zero indicates a perfect prediction, while a NRMSE of unity indicates that the model is no more effective than taking the mean of the data.

A subjective measure of the effectiveness of the technique can be seen in Fig. 1, which shows a 1-hour ahead RBF-NN model following the variations in *fo*F2 through a storm period in February 2000. For this particular RBF-NN model,
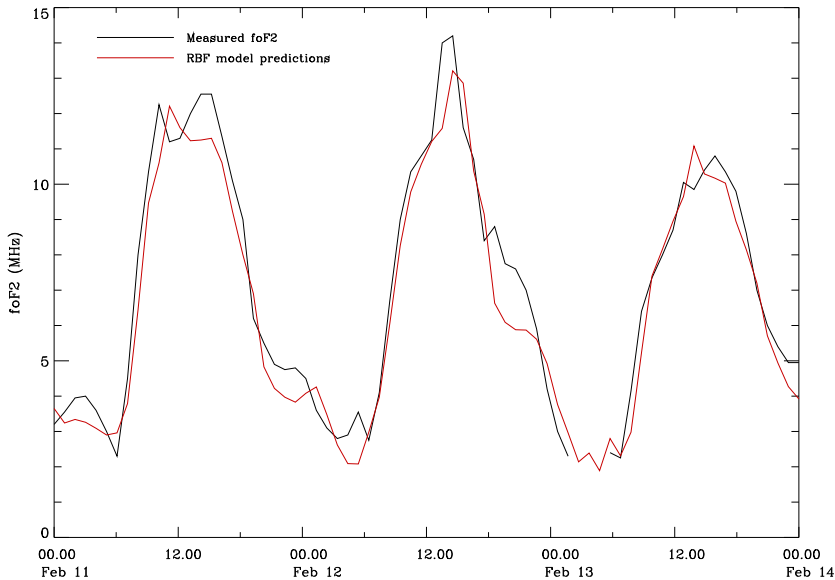
Fig. 1. Measured *fo*F2 and 1-hour ahead forecasts during a storm event in February 2000. The gap in the measured *fo*F2 time series represent missing data.

the NRMSE was found to be 0.26, representing an improvement of ∼ 35% and 42% over the reference persistence and 24-hour recurrence models, respectively.

3 Data analysis

3.1 Data description

For this study, hourly *fo*F2 measurements from the UK ionosonde station at Chilton (51.6° N, 358.7° E) from June 1995 to October 2000 are used. Measurements at the Chilton station started in the mid-1990's, and after an initial settling-in period, the percentage of missing data points per year fell to ∼ 4% in 1997, a value which has since been maintained.

As is typical in many geophysical data sets, the time series used in this study contains many data dropouts (e.g. associated with instrument failure, and data values which fall outside the accepted range of the *fo*F2 parameter). These discontinuities in the data time series pose a significant obstacle for prospective nonlinear prediction schemes, which generally require continuous data. A nonlinear interpolation technique, which minimizes the effects of interpolation upon any given modeling process, has been developed by Francis et al. (2001) to deal with data gaps in the time series. For 1-hour ahead forecasts, this has been shown to provide an overall improvement of 2.3% and 3.8% over the 24-hour recurrence and persistence interpolation schemes, respectively. This nonlinear interpolation technique, however, is computationally and time intensive, and as a result, has not been utilized in this study. Instead, missing data points are interpolated using the 24-hour recurrence values.

3.2 Model description and test error analysis

A number of optimized NN forecasting models were generated using the data time series of *fo*F2 from the Chilton ionosonde station. The optimization process involves adjusting the input vector window length and centers, as discussed in the previous section. Each model contains 1.5 years' worth of data, of which 75% of the data points available were used to train and optimize the model, and the remaining 25% were used to test the model's predictive accuracy on unseen data. The remainder of the period of available data (up to October 2000) was then used to evaluate the long-term performance of the model.

The data periods used for each model, the data characteristics, such as the mean and standard deviation, σ, of the input data, and the percentage of missing data, are detailed in Table 1. The data are characterized by improving quality, a higher mean value, and increased variability as time progressed. Also included in the table are the model test errors. As an indication of the effectiveness of the nonlinear modeling routines against the more common linear modeling techniques, the model test errors (based on the last 25% of the input data time series) are compared with that from the reference persistence and recurrence models, which assume that the data point Y_{n+1} will be the same as Y_n and Y_{n-23}, respectively (Table 2). The considerable improvements of the RBF-NN models over persistence predictions can be seen to range from approximately 25 to 50%, while improvements over 24-hour recurrence predictions are between ∼ 28 and 58%.

3.3 Long-term model error analysis

While it is common to present model errors such as those given in Table 2, in this work we are also interested in the long-term performance of the neural network models. The

Table 1. Description of prediction models

Model no.	Train & test period	Predict from	% of missing data	Mean (MHz)	σ (MHz)	Test NRMSE
1	1 Jun 95–31 Dec 96	Jan 97	15.7	4.24	1.12	0.346
2	1 Jan 96–30 Jun 97	Jul 97	12.0	4.08	1.13	0.310
3	1 Jul 96–31 Dec 97	Jan 98	9.45	4.17	1.11	0.268
4	1 Jan 97–30 Jun 98	Jul 98	4.38	4.33	1.35	0.281
5	1 Jul 97–31 Dec 98	Jan 99	4.05	5.00	1.56	0.222
6	1 Jan 98–30 Jun 99	Jul 99	4.35	5.47	1.96	0.228
7	1 Jul 98–31 Dec 99	Jan 00	4.08	6.11	2.01	0.180

Table 2. Comparison of RBF model normalized root-mean-squared errors (NRMSE) with persistence and 24-hour recurrence model NRMSE

Model no.	RBF model NRMSE	Persistence NRMSE	Improvement of RBF model over persistence (%)	Recurrence NRMSE	Improvement of RBF model over recurrence (%)
1	0.346	0.506	31.6	0.478	27.6
2	0.310	0.432	28.2	0.569	45.5
3	0.268	0.430	37.7	0.469	42.9
4	0.281	0.377	25.5	0.598	53.0
5	0.222	0.385	42.3	0.429	48.3
6	0.228	0.338	32.5	0.540	57.8
7	0.180	0.355	49.3	0.357	49.6

normalized root-mean-squared errors (NRMSE) for 1-hour ahead forecasts for the series of models from January 1997 to October 2000 are shown in Fig. 2. Each point represents a monthly averaged value (calculated by summing over all available data points in the month and then dividing by the total number of data points), and each curve starts after the previously described test period has ended (see Table 1). Immediately apparent is the degradation which occurs with time, and the benefit accrued from periodic retraining and re-optimizing of the neural network models. Model retraining every year, and most beneficially every 6 months, is necessary.

It can also be clearly seen that the NRMSEs are smaller in winter than in summer, indicating that the RBF-NN models perform better during winter. To examine the reason for this difference, we need to look at the variation in the *fo*F2 time series over the course of one year, an example of which is shown in Fig. 3. During the winter months, the *fo*F2 data exhibits a clear diurnal variation, whereas the summer variation is unclear, with the peak-to-peak variation being smaller and almost noise-like. As can be seen from Eq. (3), if all other factors remain unchanged, then the smaller variation in the summer data will result in a higher NRMSE. From

the perspective of the neural network developer, our model is, therefore, more successful in making the winter forecasts. However, while the NRMSE provides a useful measure of the model's success, the normalization obscures the absolute error associated with the forecast. For systems assessment, this might be more important.

The absolute errors can be best seen in Fig. 4, which shows the un-normalized root-mean-squared errors (RMSE) for each of the models over the same time scale as seen in Fig. 2. Again, the advantages of periodic retraining and re-optimizing of the models can be seen. The absolute errors can be reduced by up to ~0.1 MHz per six months as a result. Since the standard deviation of the input data is not taken into account when computing RMSE, we can see that the absolute errors are smaller in summer than in winter. This can be attributed to the fact that the variation from highest to lowest *fo*F2 values within the period of one day are greater in winter than in summer, thus producing higher absolute errors in the winter forecasts. The anomalously high RMSE for March 2000 in Fig. 4 corresponds to an unusually high mean *fo*F2 input value for the same month, which is a result of the high peak values (above 10 MHz) seen in the hourly *fo*F2 data throughout the month.
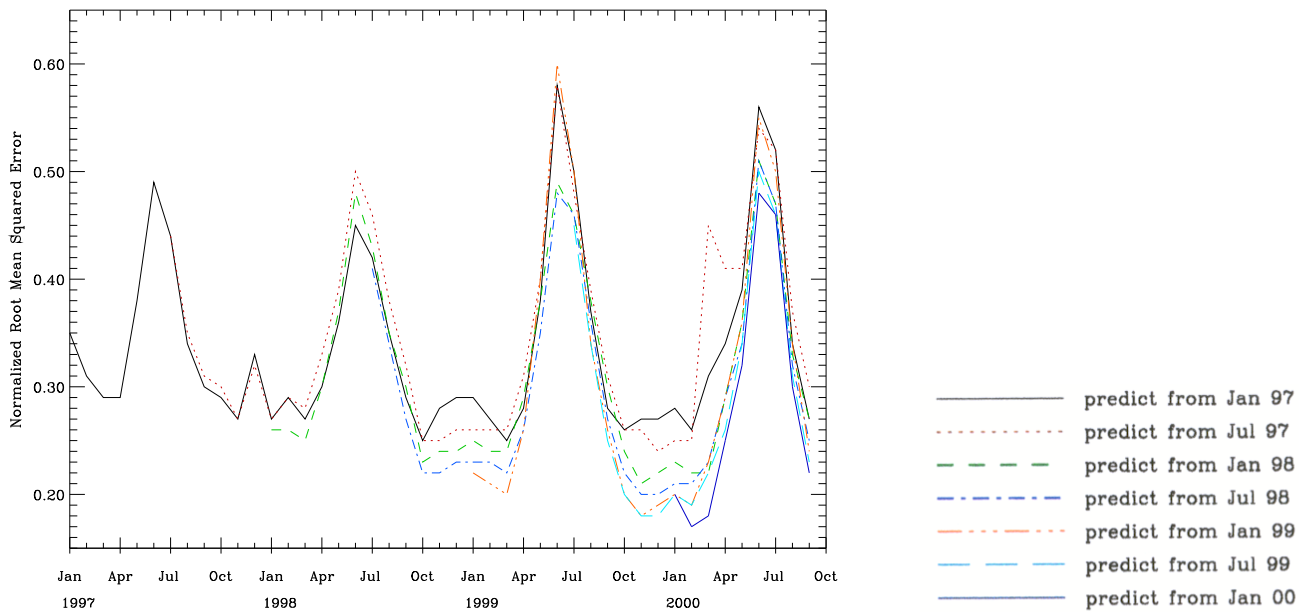
Fig. 2. Normalized root-mean-squared errors for 1-hour ahead forecasts of *fo*F2 from January 1997 to October 2000.
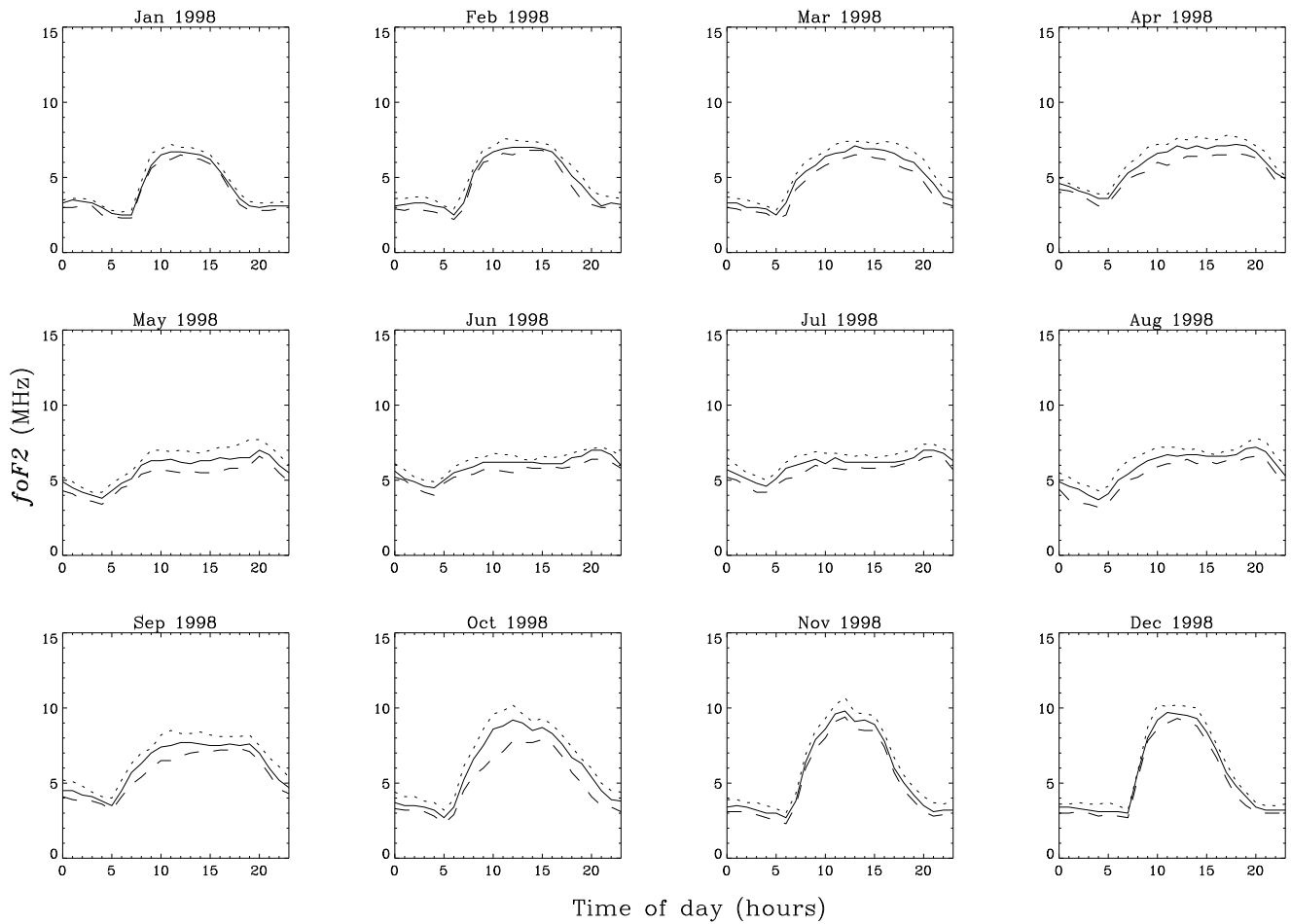

Fig. 3. Monthly median *fo*F2 values for the year 1998 (solid line). Dashed and dotted lines represent the upper and lower quartiles, respectively.
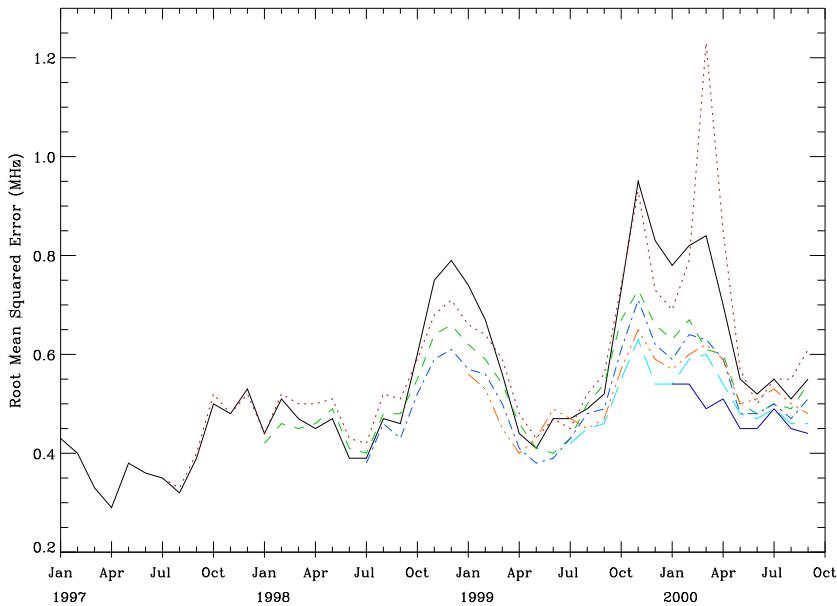
Fig. 4. Root-mean-squared errors for 1-hour ahead forecasts of *fo*F2 from January 1997 to October 2000 (Legend as in Fig. 2).
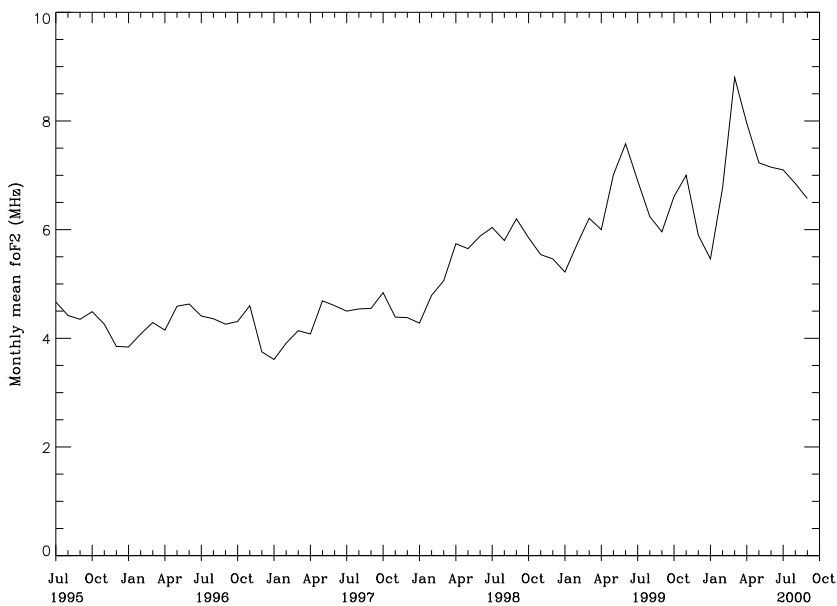

Fig. 5. Monthly mean *fo*F2 values from July 1995 to October 2000.

Also seen in Fig. 4 is the general increasing trend in the RMSE. The time period over which the forecasts are made is during an ascending phase of the solar cycle, as can be seen in Fig. 5, which shows the monthly mean *fo*F2 values from July 1995 to October 2000. To compensate for this increasing trend in the mean *fo*F2 data, we have demeaned the root-mean-squared errors by dividing by the monthly mean of the input data (Fig. 6). The annual variation in RMSE still remains.

3.4 Retraining versus re-optimizing

We have shown in the previous section the benefits of periodic re-optimization of neural network models to cope with nonstationary data sets. This is a consequence of developing models with input data sets that are shorter than the longest characteristic period of the data. Ideally, models should be trained using longer series of data, to take into account the large range of physically significant time scales seen in geophysical data sets. In particular, models trained using 11 or more years of data would be able to account for the variation in *fo*F2 data over the solar cycle. Unfortunately, processing constraints meant that the length of the input data time series in this study was limited to 1.5 years; this still contained ∼ 13 000 data points. Furthermore, the optimization process for the neural network model took roughly 1–2 weeks of computing time.
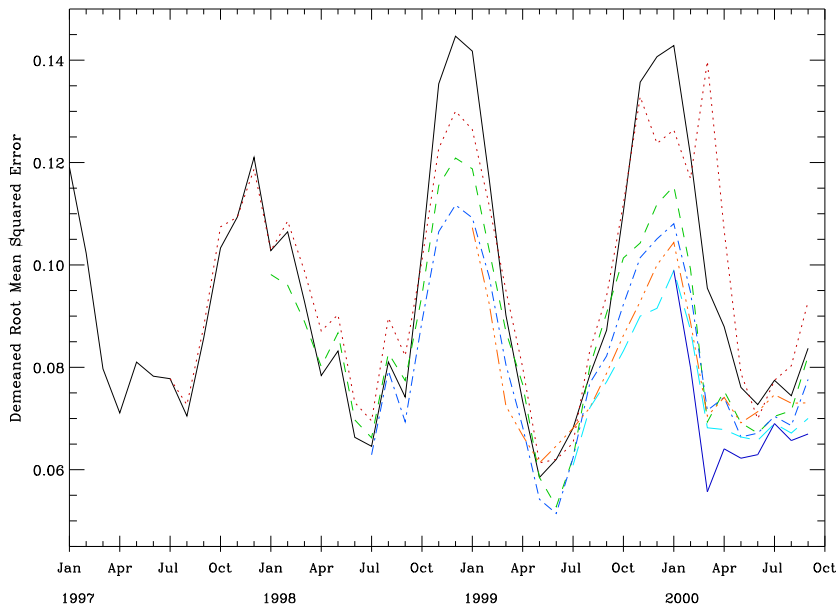
Fig. 6. Demeaned root-mean-squared errors for 1-hour ahead forecasts of *fo*F2 from January 1997 to October 2000 (Legend as in Fig. 2).
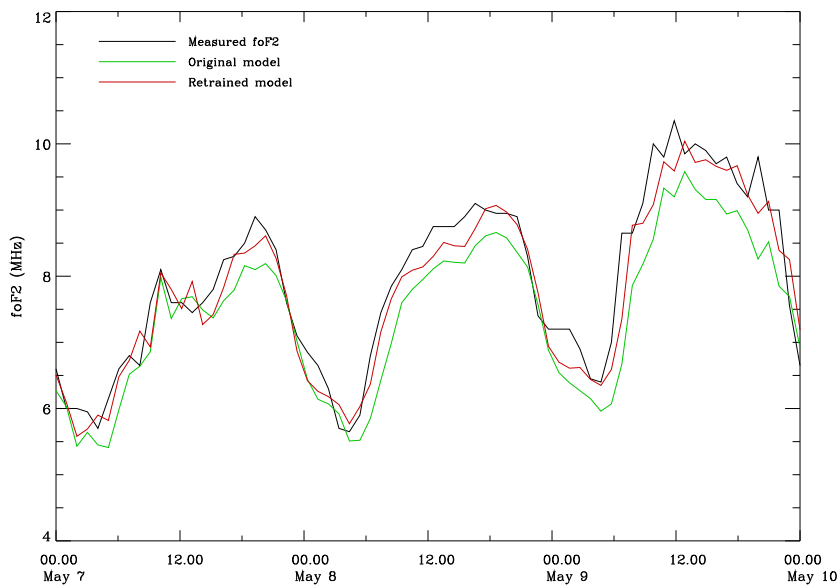

Fig. 7. Measured *fo*F2 and 1-hour ahead forecasts from original and retrained models for the time period 7–9 May 2000.

In the absence of adequate resources, we have investigated an alternative approach to improving the model predictive accuracy on a new data set. The approach simply retrains the neural network model to obtain new weights while still using previously optimized window length and centers. This approach saves considerable computer processing time, in that a "new" model can be generated in a matter of hours rather than weeks. Figure 7 shows the actual *fo*F2 time series, together with the comparison of nonlinear forecasts performed using an out-of-date model, and that using a retrained model. Even though not fully optimized, the retrained model shows clear improvements over the out-of-date model. Thus, in the absence of adequate processing time, simply retraining the model rather than re-optimizing can prove to be a useful

option to improve model forecasts.

4 Conclusions

The use of radial basis function neural network (RBF-NN) models for the 1-hour ahead forecasts of *fo*F2 at a northern mid-latitude location has been shown to provide improvements of up to 50% over the reference persistence prediction models, and up to 58% over the 24-hour recurrence models. Analyses have demonstrated the importance of retraining and re-optimizing ionospheric forecasting models at regular intervals to cope with the non-stationary data set. Specifically, model predictive errors were found to be reduced by

~ 0.1 MHz per six months as a result of periodic retraining and re-optimization. Our studies have also shown that, in the absence of adequate computing resources and time, retraining of the model without re-optimizing still proves to be beneficial.

This work further illustrates the problems of benchmarking one analysis technique against another. Neither the simple root-mean-squared (RMS) errors, nor the demeaned RMS errors, nor the normalized RMS errors (NRMSE) offer the definitive error analysis approach. While the normalized RMS error is an excellent metric for understanding and comparing predictive schemes, the RMS error is invaluable for quantifying the absolute errors. The former is needed to provide a meaningful measure of the model's success; the latter is a necessary requirement for understanding the utility of these prediction techniques in practical applications. Clearly, no one metric tells the complete story, and it is incumbent on all forecasters to provide a realistic range of error metrics.

References

Baker, D. N., Klimas, A. J., McPherron, R. L., and Buchner, J.: The evolution from weak to strong geomagnetic activity: An interpretation in terms of deterministic chaos, Geophys. Res. Lett., 17, 41–44, 1990.

Broomhead, D. S. and Lowe, D.: Multivariable functional interpolation and adaptive networks, Complex Systems, 2, 321–355, 1988.

Cander, Lj. R., Milosavljevic, M. M., Stankovic, S. S., and Tomasevic, S.: Ionospheric forecasting technique by artificial neural network, Electronic Letters, 34, 16, 1573–1574, 1998.

Feynman, J. and Gabriel, S. B.: On space weather consequences and predictions, J. Geophys. Res., 105, A5, 10 543–10 564, 2000.

Francis, N. M., Brown, A. G., Broomhead, D. S., Cannon, P. S., and Akram, A.: Nonlinear prediction of the ionospheric parameter foF2, Proc. 2nd Workshop on AI Applications in Solar-Terrestrial Physics, ESA WPP-148, 219–223, 1997.

Francis, N. M., Cannon, P. S., Brown, A. G., and Broomhead, D. S.: Nonlinear prediction of the ionospheric parameter foF2 on hourly, daily, and monthly timescales, J. Geophys. Res., 105, A6, 12 839–12 849, 2000.

Francis, N. M., Brown, A. G., Cannon, P. S., and Broomhead, D. S.: Prediction of the hourly ionospheric parameter foF2 using a novel nonlinear interpolation technique to cope with missing data points, J. Geophys. Res., 106, A12, 30 077–30 083, 2001.

Hargreaves, J. K.: The solar-terrestrial environment, Cambridge University Press, pp. 420, 1995.

Joselyn, J. A.: Geomagnetic activity forecasting: The state of the art, Rev. Geophys., 33, 3, 383–401, 1995.

Klimas, A. J., Baker, D. N., Roberts, D. A., Fairfield, D. H., and Buchner, J.: A nonlinear dynamic analogue model of geomagnetic activity, J. Geophys. Res., 97, A8, 12 253–12 266, 1992.

Lundstedt, H., Neural networks and predictions of solar terrestrial effects, Planet. Sp. Sci., 40, 4, 457–464, 1992.

Smith, R. J. K., Brown, A. G., and Francis, N. M.: Time Series Analysis Routines (TSAR) V4.0 User Manual, DERA, DERA/CIS(CIS2)/WP98048, 1998.

Vassiliadis, D., Klimas, A. J., Baker, D. N., and Roberts, D. A.: A description of the solar wind-magnetosphere coupling based on nonlinear filters, J. Geophys. Res., 100, A3, 3495–3512, 1995.

Williscroft, L. A. and Poole, A. W. V.: Neural Networks, foF2, sunspot number and magnetic activity, Geophys. Res. Lett., 23, 24, 3659–3662, 1996.

Wintoft, P. and Cander, L. R.: Ionospheric foF2 storm forecasting using neural networks, Phys. Chem. Earth C, 25, 4, 267–273, 2000.

Wu, J. G. and Lundstedt, H.: Geomagnetic storm predictions from solar wind data with the use of dynamical neural networks, J. Geophys. Res., 102, A7, 14 255–14 268, 1997.