Nonlinear Prediction and Analysis of Solar-Geophysical Disturbances within the Ionosphere

By

Nicholas Francis

Thesis submitted for the degree of Doctor of Philosophy
of the University of Leicester

April 2002

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Man is a rope stretched between the animal and the Overman - a rope over an abyss.

A dangerous crossing, a dangerous wayfaring, a dangerous looking-back, a dangerous trembling and halting.

What is great in man is that he is a bridge and not a goal: what is loveable in man is that he is an over-going and a down-going.

I love those that know not how to live except as down-goers, for they are the over-goers.

I love the great despisers, because they are the great adorers, and arrows of longing for the other shore.

I love those who do not first seek a reason beyond the stars for going down and being sacrifices, but sacrifice themselves to the earth that the earth of the Overman may hereafter arrive.

I love him who liveth in order to know, and seeketh to know in order that the Overman may hereafter live. Thus seeketh he his own down-going.

I love him who laboureth and inventeth, that he may build the house for the Overman, and prepare for him earth, animal, and plant: for thus seeketh he his own down-going.

I love him who loveth his virtue: for virtue is the will to down-going, and an arrow of longing.

I love him who reserveth no share of spirit for himself, but wanteth to be wholly the spirit of his virtue: thus walketh he as spirit over the bridge.

I love him who maketh his virtue his inclination and destiny: thus, for the sake of his virtue, he is willing to live on, or live no more.

I love him who desireth not too many virtues. One virtue is more of a virtue than two, because it is more of a knot for one's destiny to cling to.

I love him whose soul is lavish, who wanteth no thanks and doth not give back: for he always bestoweth, and desireth not to keep for himself.

I love him who is ashamed when the dice fall in his favour, and who then asketh: "Am I a dishonest player?"- for he is willing to succumb.

I love him who scattereth golden words in advance of his deeds, and always doeth more than he promiseth: for he seeketh his own down-going.

I love him who justifieth the future ones, and redeemeth the past ones: for he is willing to succumb through the present ones.

I love him who chasteneth his God, because he loveth his God: for he must succumb through the wrath of his God.

I love him whose soul is deep even in the wounding, and may succumb through a small matter: thus goeth he willingly over the bridge.

I love him whose soul is so overfull that he forgetteth himself, and all things are in him: thus all things become his down-going.

I love him who is of a free spirit and a free heart: thus is his head only the bowels of his heart; his heart, however, causeth his down-going.

I love all who are like heavy drops falling one by one out of the dark cloud that lowereth over man: they herald the coming of the lightning, and succumb as heralds.

Lo, I am a herald of the lightning, and a heavy drop out of the cloud: the lightning, however, is the Overman.

- Friedrich Nietzsche, 'Thus Spoke Zarathustra'.

"Nonlinear Prediction and Analysis of Solar-Geophysical Disturbances within the Ionosphere" By Nicholas Francis

Abstract

This thesis presents novel and robust nonlinear techniques that can be used to analyse and predict geophysical time series. The techniques presented can cope with the problems of noise and non-contiguity that are typical of solar-terrestrial data sets. Existing nonlinear analysis techniques are prone to giving spurious indications of nonlinear behaviour in such circumstances. These new techniques can be applied as either an alternative to established techniques, to determine less fragile and more qualitative nonlinear properties that may be present, or as a precursor - to determine whether or not it is viable to proceed with the application of standard nonlinear analysis methods. The methods in this thesis have a general applicability that extends to any time series prediction problem, and are not necessarily limited to geophysical applications. Furthermore, these techniques have been adapted to form the basis of a real-time ionospheric forecasting demonstrator.

Acknowledgements

I would like to thank the following individuals and organisations for their assistance and advice during the writing of this thesis:

Professor P S Cannon is acknowledged for initiating the research topics that underlie this body of work and for his supervision during my post-graduate studies. I would also like to thank Dr A Akram for his dependable and incisive guidance over the last five years and most of all for his understanding when I experienced difficulties. My thanks also go to Dr A G Brown, who contributed his understanding of mathematics and a considerable amount of code during the same period. In particular, he was responsible for development of the minimisation algorithm employed in Chapter 6 for the interpolation of missing data points. In addition, I would like to thank him for proof reading the mathematics in this thesis, in conjunction with Professor D S Broomhead, to whom I would also like to express my gratitude. I would also like to thank the Corporate Research Programme managers who funded this project during the course of my studies.

The following sources of data are acknowledged: Mr R Stamper at the World Data Centre at the Rutherford-Appleton Laboratory, Dr T K Yeoman at the University of Leicester for the SABRE data and information concerning his past studies of the data set, including the spectrogram of the SABRE data in Chapter 4.

I would also like to thank Mr D Smith at the University of Leicester for many of the diagrams that have been used throughout this thesis.

Finally, I would like to dedicate this thesis to my loving mother and grandparents, whose selfless support has made this work possible, and to my brother for his constant companionship.

Table of contents

List of Figures and Tables

Chapter 1: Introduction

1.1 Problem Definition

A long-standing goal of solar-terrestrial physics has been the prediction of specific events or indices that are indicative of geophysical activity. Linear correlation studies, that relate solar wind energy input to various indices of geomagnetic activity, have played a crucial role in understanding the sun - solar wind - magnetosphere - ionosphere coupled system and the time scales of variation within the solar-terrestrial environment. As an extension of these studies, linear predictive techniques have been used to deduce relationships between input and output data [e.g. *Milsom et al.*, 1989, 94]. Although linear techniques have achieved some measure of success, it is unlikely that they will ever provide an adequate description of the nonlinear physical processes that are taking place within the solar-terrestrial environment.

The fundamental limitations of linear analysis have led to increased interest in the application of sophisticated nonlinear techniques to solar-geophysical data sets. A number of experimental studies have demonstrated the importance of nonlinear behaviour within the solar-terrestrial environment [*Baker et al.*, 1990]. Attempts have also been made to derive nonlinear theoretical models from the fundamental nonlinear MHD equations [*Klimas et al.*, 1992, 1996]. However, the MHD descriptions are often very complicated and analytical nonlinear solutions are available only in a small number of simplified scenarios. In the absence of a tractable nonlinear theory derived from the full MHD equations, it is an attractive idea to look for nonlinear features in experimental data, in order to develop nonlinear models using empirical methods.

These empirical methods, which are a development of work over the last decade to find evidence of finite dimensional dynamics in fluid turbulence, have received much recent attention. However, these techniques have been developed for well-

controlled experimental systems or computer simulations. Consequently, they work best when applied to large sets of low-noise, stationary data. In practice, solar-terrestrial data sets are very different from this ideal; they are noisy and highly quantised, with many data dropouts. They are rarely stationary, due to the large range of physically significant time-scales. It follows that the naive application of many of the nonlinear techniques whose purpose is to extract detailed information (e.g., attractor dimensions and Lyapunov exponents) from data are almost certain to fail or give spurious results without additional safeguards. Consequently, alternative methods are required that can be applied to such time series as an alternative to determine more robust nonlinear properties, or as a precursor to validate the subsequent application of traditional nonlinear methods.

1.2 Summary of Previous Nonlinear Empirical Studies

In a few cases, the use of empirical nonlinear time series methods to characterise the behaviour of complex solar-terrestrial systems has met with success, when applied to specific data sets. Examples include dimensional analysis of magnetospheric systems, as well as auroral indices of activity and nonlinear prediction of geomagnetic indices [*Vassiliadis et al.*, 1992; *Sharma,* 1995]. In some of these studies, subsequent use of surrogates showed that the results obtained could not be demonstrated as significant indicators of nonlinear activity [*Vassiliadis et al.*, 1995]. This highlights the importance of employing safeguards and precursive studies.

Most recently, attention has turned to knowledge-independent methods derived from studies into artificial intelligence (AI) and the application of techniques based upon neural networks to geophysical prediction problems [*Cander et al.*, 1997; *Lamming et al.*, 1997; *Williscroft et al.*, 1996; *Wu and Lundstedt*, 1997]. These techniques are robust and more suitable for use with solar-terrestrial data.

1.3 Space Climate and Space Weather

This thesis concerns itself primarily with prediction of short-term events in the solar-geophysical system (space weather), rather than issues relating to long-term climatological trends (space climate). However, the division between these two areas is often unclear. For example, the 11-year solar cycle is a significant climatological feature for a number of parameters such as sunspot number or geomagnetic activity. On the other hand, the 27-day periodicity associated with the solar rotation is more closely related to space weather. The 27-day recurrence cycle is one of the most predictable features of space weather, particularly during the descending phase of the solar cycle, and its characteristics change over the solar cycle as a whole. Thus, space climate and space weather are inextricably linked such that they cannot be easily decomposed into two separate issues.

In many ways, the prediction of short term and impulsive space weather is analogous to terrestrial weather forecasting, but there are important differences between atmospheric and solar-terrestrial weather systems. Many meteorological systems are localised and it is therefore possible to make good limited-area weather forecasts. Space weather is global on the planetary scale due to the large spatial scale sizes and long correlation times for the solar-terrestrial plasma systems. Thus, disturbances originating from the sun give rise to global changes within the near earth environment (the magnetosphere and ionosphere).

Space borne monitoring capabilities are also much more limited than the capabilities provided by the network of terrestrial sensors used to monitor meteorological conditions. Current space weather prediction schemes are only able to utilise isolated satellites in the solar wind and magnetosphere, in conjunction with a limited number of ground stations used to provide ionospheric and magnetometer measurements. As a direct consequence, space weather measurement activities are often restricted to global or long-distance scales.

1.4 Overview of Applications

Space weather effects give rise to changes within the near earth environment on time scales varying from minutes to weeks. These effects may damage space-borne or ground-based technological systems, or endanger human health or life. The socio-economic aspects of space weather and forecasting are aimed at avoiding or reducing the consequences of disturbed conditions within the near earth environment. This can be achieved through system design and prediction systems that allow timely preventative measures to be taken.

The effects of space weather can manifest themselves in many different forms. Within the disturbed magnetosphere, dynamic magnetospheric processes may enhance existing energetic particle populations to levels that are hazardous to electronic space-based systems. Undesirable effects include particle-induced anomalies, lattice structure damage and charging effects. Cosmic rays produced by particle acceleration also pose a significant hazard to satellites and to high-altitude civilian aircraft and their passengers. Thermospheric drag on spacecraft can increase during periods of elevated geomagnetic activity, reducing space system life expectancy and functionality.

The impact of space weather is not necessarily limited to space based systems. Enhanced geophysical activity can give rise to powerful current systems in the auroral regions at high latitudes. These auroral systems can produce Geomagnetically Induced Currents (GIC) along long distance power cables and pipelines, leading to corrosion or transformer overload. GIC effects can be mitigated through appropriate network design but a measure of forewarning may help to reduce infrastructure damage. Enhanced activity in the ionosphere that is attributable to space weather, the main topic of this thesis, can also disturb high frequency (HF) / very high frequency (VHF) radar and communications, GPS and ultra high frequency (UHF) satellite communications. Magnetic storm effects can

also influence mining operations, which are dependent upon accurate geomagnetic information for guidance purposes.

All of these effects tend to be correlated (with the exception of cosmic rays) with solar activity and disturbances are more frequent at the peak of the solar cycle. Despite advances in radiation hardening techniques, it is also expected that effects on made-made systems will become more severe over time, due to the increasing sensitivity of electronic circuits utilising ever-smaller components. Consequently, solar-terrestrial environment predictions will have an increasingly important bearing upon the specification and operation of a number of services [*Koskinen et al.*, 1998].

At this point, it is worth drawing attention to the fact that the current state of the art for operational geophysical forecasting services falls far short of the predictive precision that is required for an acceptable model of future solar-terrestrial activity [*Joselyn*, 1995]. In practice, existing forecasting schemes are capable of little more than providing 'now-casts' using real-time data monitoring experiments. For temporal, fixed-point predictions, current in-service models often even fail to offer an increase in performance over reference persistence and recurrence models. In particular, predictions of storm events and disturbances are very poor. Disturbed conditions occur infrequently and the existing basic techniques are unable to predict such impulsive events with any degree of accuracy. This performance shortfall can be attributed to several causes, but includes an inadequate understanding of, and inability to model, the coupled solar - magnetospheric - ionospheric system. The widespread adoption of purely linear modelling techniques is also a contributory factor. Consequently, the adoption of knowledge independent (time-series) nonlinear modelling techniques, that can utilise typical geophysical data sets and incorporate physical understanding, is highly desirable and is a primary motivator of this thesis.

The historical record of space weather stretches back well over a century. The first ground-based effects were observed with the dawning of the electrical communications age in the middle of the 19th century. Around 1850, it was noticed that the first electric telegraph communications were disturbed or stopped during periods of intense auroral activity. Later, long-distance telephone connections became the most susceptible technological systems. During a space weather event on March 24, 1940, 80% of the long-distance telephone connections out of Minneapolis, Minnesota were put out of operation.

One of the most reported space weather events took place on March 13, 1989, in Quebec, Canada. A severe geomagnetic storm caused the complete failure of the local electricity distribution system, such that several million people in the area were without electric power for up to nine hours. The peak power lost during this period exceeded 20GWh. The effect cascaded through the system very rapidly, spreading from the initial point of failure to system-wide collapse in less than 90 seconds. Simultaneously, EM propagation was also severely affected. HF communications were degraded worldwide and VHF signals were subject to unusual propagation conditions. The range of VHF signals was greatly increased, giving rise to severe interference problems. Satellite operations were also disrupted - a Japanese communications satellite lost half of its redundant systems and a NASA spacecraft lost 5km in altitude due to increased atmospheric drag.

Most recently, several satellites have experienced operational anomalies due to space weather effects. Anik E1 and E2 both experienced circuit failures in 1994 and 1996 that led to significant operational difficulties. Telstar 401 also experienced anomalies that were attributed to a solar coronal mass ejection.

1.5 Aim of this Thesis

This thesis presents novel and robust techniques, based upon recent developments in AI and nonlinear theory, for the analysis and prediction of solar-geophysical

parameters. The techniques described in this thesis can cope with the problems of noise and non-contiguity that are typical of solar-terrestrial data sets. Existing nonlinear analysis techniques are prone to giving spurious indications of nonlinear behaviour in such circumstances. These new techniques are an alternative to established nonlinear techniques and can be used to determine less fragile and qualitative nonlinear properties that may be present in the data. Alternatively, they can be used a precursor, simply to determine the validity of standard nonlinear analysis methods. The methodologies presented can also be used for the purposes of time series prediction. Such methods have a general applicability that extends to any time series analysis or prediction problem.

These techniques have also been adapted to form the basis of a real-time ionospheric forecasting demonstrator, discussed at greater length in the final chapter of this thesis. This tool provides real-time predictions of the ionosphere that are of operational significance. Currently, the tool provides predictions for a single parameter at a single station, but could be expanded to cover a network of stations or adapted to create models of other solar-geophysical parameters.

Chapter 2: Solar-Terrestrial Physics

2.1 Introduction

This chapter provides a simple description of some of the relevant physical processes that underlie space weather and which are driven by solar activity. It deals with the solar origin of geophysical disturbances and the propagation of effects to the near Earth environment via the solar wind. The chapter then deals with the behaviour of the magnetosphere and ionosphere and the coupling of these systems to the solar wind. Finally, it covers solar-geophysical variations during normal and disturbed conditions and details previous attempts to model these variations.

2.2 The Sun and the Solar Wind

The sun is an unremarkable star that emits electromagnetic radiation over the entire frequency spectrum. The visible surface is known as the photosphere and approximates a black body at 6400 K, with a radius of approximately $7x10^5$ km. The chromosphere and the corona lie above the visible surface of the sun. The chromosphere extends a few thousand kilometres above the surface of the sun and reaches temperatures of up to 50,000 K. The corona can be observed at distances of up to 10^6 km and can reach temperatures of up to 1.5×10^6 K. After this point, the corona becomes more diffuse and merges with the solar wind.

The sun emits considerable energy in the form of electromagnetic radiation. The peak of the energy distribution occurs in the visible light range and is constant to within 1%. The UV and X-ray emission levels are much more variable, as the black body emission at these frequencies is supplemented by other processes in the suns atmosphere. These frequencies are of particular relevance to ionospheric issues as they directly affect ionisation within the upper atmosphere.

A flux of energetic particles also originates from the solar surface. This flux arises from imbalances between the force of gravity and internal pressure at the stellar surface. This leads to the ejection of protons, electrons, alpha particles and a small percentage of heavy ions from the corona, the outermost layer of the sun. These particles are then accelerated to escape velocity by magnetic processes in the atmosphere of the sun and travel outwards across the solar system. In conjunction with the diffuse corona, these accelerated plasma particles are collectively known as the solar wind. The steady solar wind thus pervades the whole solar system as far as the heliopause where it forms a shock boundary with the interstellar medium. This steady solar wind stream is punctuated by brief and localised periods of enhanced density and velocity that have their origins in certain features of the active sun.

The rotation of the sun about its own axis is very important in relation to geophysical phenomena because the active regions of the sun are not uniformly distributed across the visible surface. As the sun is not a solid body, it exhibits a differential rotation rate with latitude. The rotation period varies from 25 days at the equator to 33 days at 75° latitude. This differential rotation is also thought to drive the solar dipolar magnetic field by creating circulating currents beneath the surface of the sun. The solar magnetic field is subsequently carried by the solar wind as it travels outward across the solar system. The coupled particles and field are collectively referred to as magnetoplasma. The interplanetary magnetic field (IMF) flux is "frozen" into the plasma because the electrical conductivity of the plasma is very large. Thus, the motion of this magnetoplasma is dictated by the plasma component of the solar wind, because it has a much higher energy density than the magnetic component.

The sun itself also undergoes periodic variations in the form of the 11-year solar cycle that governs the level of overall activity. The size and number of sun spots, cooler patches (4000K) on the surface of the sun that tend to occur in clusters, are often taken as a measure of solar activity. The Wolf sunspot number is derived

from observation and used as an index of sunspot activity. Historical sunspot number records are available from 1700, but it is possible to infer indirect values from geological records covering the last two thousand years. The 11-year cycle is present for most of the historical record, excepting the Sporer and Maunder minima of the 15^{th} and 17^{th} centuries respectively. During these latter periods, solar activity was exceptionally low for some decades and coincided with harsh winters in northern Europe. In the light of these records, it is interesting to note that recent studies have provided evidence to suggest a much stronger link between climate change and solar activity than can be accounted for by the variation of the total energy output of the sun over time [*Robinson,* 1998].

Longer-term processes of uncertain origin modulate the 11-year cycle and account for the variation in duration and magnitude of each solar cycle. Within the historical record, the direct sunspot data shows a variation of a factor of four between the highest and the lowest peaks of 11-year activity. The length of each solar cycle can vary between nine and fourteen years, shorter cycle lengths being associated with stronger solar cycle activity. While sunspot number is one particular aspect of solar activity, it has been found that other solar features are also strongly correlated with the solar cycle. Many of these features are relevant to the prediction of solar-geophysical disturbances and will be dealt with later in this chapter.

2.3 The Magnetosphere

The magnetosphere is defined as being the region of space to which the magnetic field of the earth is confined by the immiscible solar wind magnetoplasma flowing outward from the sun [*Cowley*, 1996]. This region extends more than 60,000 km from the earth and takes the form of a cavity within the solar wind. The magnetic field of the earth is generated by currents flowing in the iron core. In the far field and in the absence of external effects, it is dipolar in form and is approximately aligned with the spin axis of the planet, which is also approximately aligned with

the spin axis of the sun. However, this dipolar field is distorted by the solar wind and IMF.

In the same way as the solar wind and the IMF are coupled together, terrestrial plasma is frozen into the magnetic field of the earth. Consequently, these two plasmas are immiscible and form distinct regions separated by a thin boundary layer, known as the magnetopause. Thus, the solar wind confines the magnetic field of the earth to a cavity surrounding the planet, referred to as the magnetosphere, as shown in Figure 2.1.

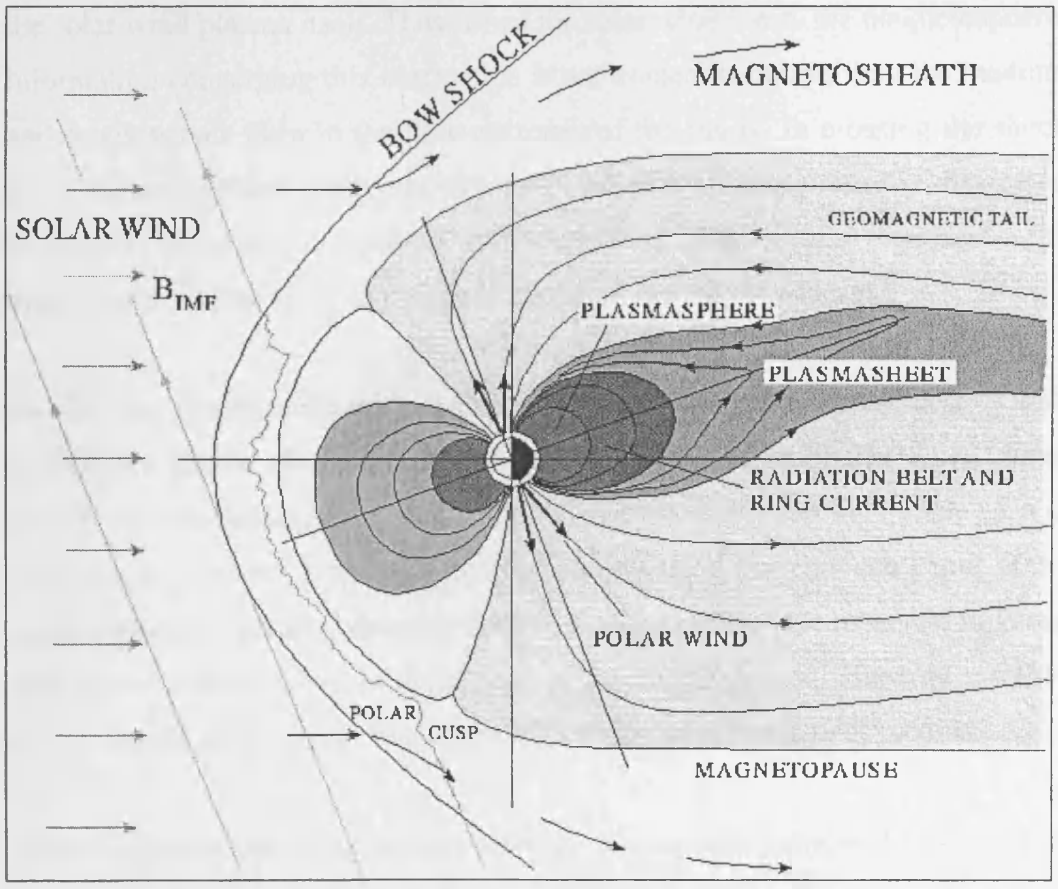


Figure 2.1 The structure of the magnetosphere

The volume of this cavity is determined by the pressure balance between the solar wind and the geomagnetic field. The equilibrium point (magnetopause) occurs at about ten earth radii on the upstream (day) side of the earth. On the downstream (night) side of the earth, the cavity is drawn out into an extended feature known as the magnetotail. This feature can extend out to greater than a thousand earth radii. At the magnetopause, the magnetic field usually undergoes a pronounced change in both magnitude and orientation. Consequently, a neutral current sheet forms at this interface.

The speed of the solar wind is greater than the speed of wave propagation within the solar wind plasma itself. Thus, when the solar wind meets the magnetosphere, information concerning this obstruction is not transmitted ahead into the medium and shock waves form in the flow upstream of the cavity. In crossing the shock zone, the solar wind plasma is slowed down and kinetic energy is dissipated thermally, forming a layer of hot, turbulent plasma that surrounds the magnetopause. This layer of plasma is known as the magnetosheath.

So far, the magnetosphere and the solar wind have been presented as two immiscible fluids. However, the frozen-in approximation breaks down under certain circumstances. One such circumstance occurs when very high current densities are present in the plasma. Sufficiently large currents can occur at the magnetopause boundary, violating the frozen-in condition. The magnetic field can then move independently of the plasma in the magnetopause, allowing the IMF and terrestrial field lines to interact through a process referred to as reconnection.

Reconnection refers to the process whereby the oppositely directed field lines on either side of the magnetopause diffuse into the boundary and become connected. The magnetic tension of the reconnected field lines causes them to contract along the current sheet away from the site of reconnection, accelerating the plasma as they do so. The plasma then moves towards the poles and joins the general circulation within the magnetosphere.

An additional reconnection mechanism occurs in the magnetotail, where open field lines from opposing poles reconnect and contract towards the earth. Consequently, the plasma on the night-side is accelerated along the magnetic field lines from the reconnection point towards the poles, precipitating heated plasma into the auroral regions of the high latitude ionosphere. The plasma on the other side of the reconnection point is accelerated towards the end of the magnetotail, back into the solar wind. This process of double reconnection forms a cyclical flow system within the magnetosphere and the high latitude ionosphere, as shown by Figure 2.2.

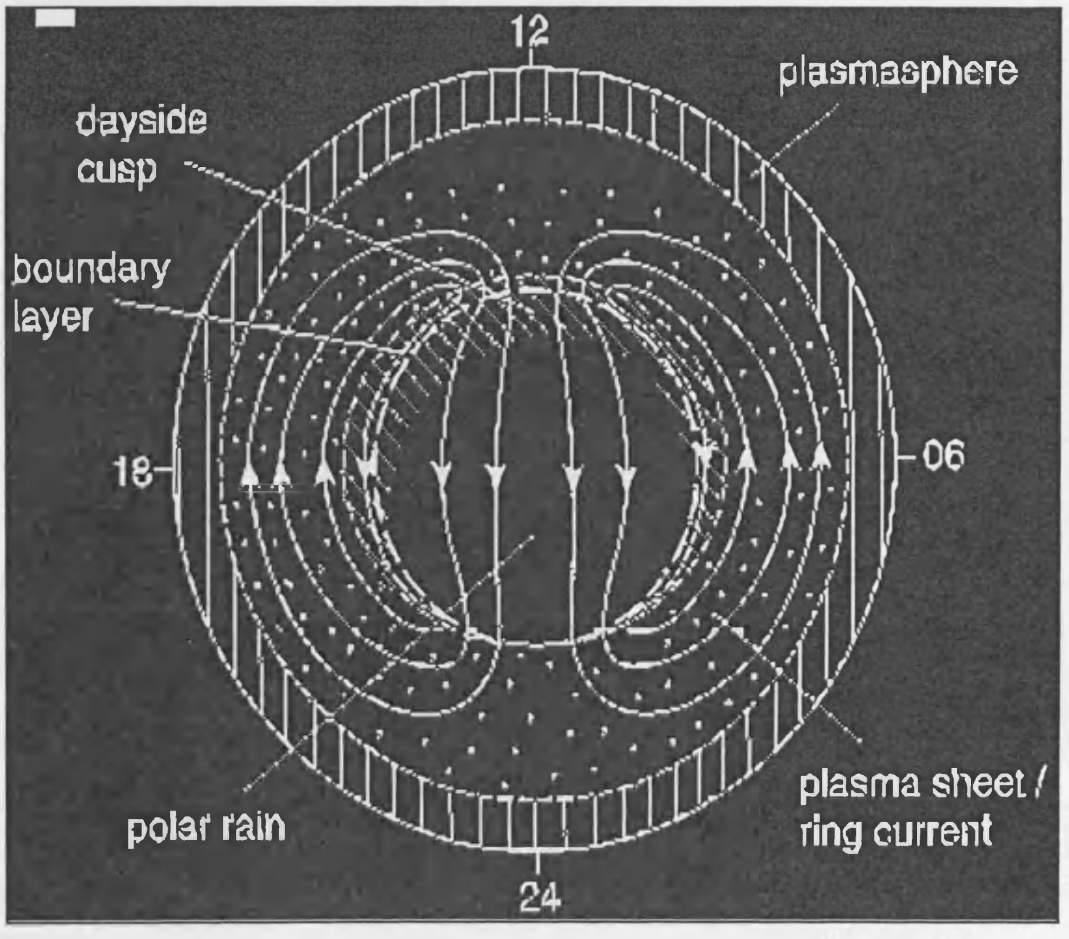


Figure 2.2 Ionospheric image of the magnetospheric flow

The strength of magnetospheric - ionospheric convection is modulated by the orientation and magnitude of the IMF. The dayside reconnection rate is strong when the IMF points south, opposing the alignment of the geomagnetic field. To a first approximation, reconnection does not occur when the IMF points north and therefore magnetospheric - ionospheric convection ceases. In the absence of reconnection, the magnetospheric field lines are frozen into the plasma originating from the ionosphere. The terrestrial plasma will therefore co-rotate with the earth and this process dominates the inner reaches of the magnetosphere in the equatorial plane.

The variability of magnetospheric activity on time scales of minutes and hours is associated with changes in the behaviour of the IMF, which affects reconnection on the sunward side of the magnetopause. The direction of the IMF is an important feature of this variability, but the continuity of the magnetospheric current systems is dependent upon reconnection in the magnetotail. Consequently, the magnetosphere does not evolve smoothly towards a new stable state of enhanced convection when reconnection occurs. This entire process of reconnection, at the magnetopause and within the magnetotail, evolves on a time scale of 1-2 hours in a process referred to as a magnetospheric substorm.

Three phases have been identified for the substorm process, as shown in Figure 2.3. In order of occurrence, these phases are referred to as growth, expansion and recovery. The growth phase (Figure 2.3a) corresponds to elevated reconnection levels on the sunward face of the magnetopause, attributed to periods when the IMF turns southward. The magnetospheric system adjusts to the altered magnetic field configuration, giving rise to large-scale excitation of the magnetospheric flow. During the expansion phase, the field lines become distended and rapidly fill with plasma. This process continues steadily for a few tens of minutes, before catastrophic reconnection occurs in the magnetotail and the distended field lines collapse inward towards the earth (Figure 2.3b), accelerating a great deal of plasma into the auroral regions. This plasma is strongly heated and compressed.

This produces an intense electron flux within the atmosphere at the feet of the associated magnetic field lines. This process is responsible for the auroral displays observed at high latitudes. Strong enhancement of the ionisation in these regions is observed with associated magnetic effects can be measured on the ground. On the other side of the reconnection point, the plasma within the recently closed field lines is ejected as the recovery phase begins. The ejected magnetoplasma is known as the plasmoid and can be detected as a burst of high-energy particles moving away from the earth. Continued reconnection then closes the open field lines, before the reconnection rate slackens and the magnetosphere assumes a pre-substorm configuration (Figure 2.3c).

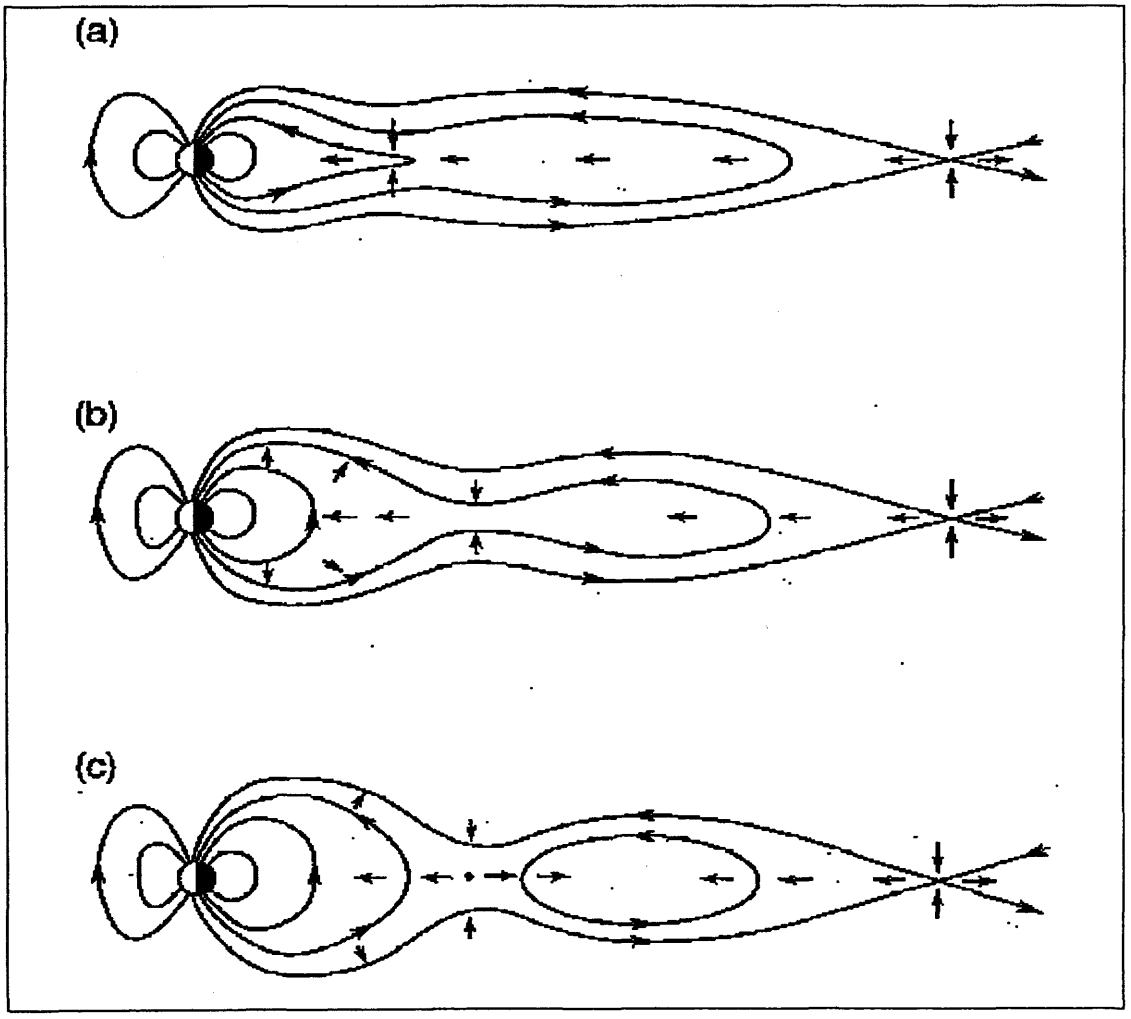


Figure 2.3 The magnetospheric substorm process

This process may run through several cycles while the IMF points south, each lasting about an hour.

2.4 *Solar-Geophysical Variation*

Geomagnetic activity displays a range of observed medium and long term periodicities, from the ~27 day solar rotation period (T_s) [*Chapman & Bartels*, 1940] to the 11 year solar sunspot cycle [*Delouis & Mayaud*, 1975], described above. Investigations of periodicities within geomagnetic time series have concentrated upon those parameters for which long time series are available, such as the magnetic indices, a_a and K_p. Direct spacecraft observations of T_s and $T_{s/2}$ (the second harmonic of T_s) periodicities have also been undertaken. *Gosling et al.* [1976], found that the ~27 day periodic solar wind variations occur most strongly during the declining phase of the solar cycle and are much weaker during the ascending phase of the solar cycle. This observation may be related to the theory that there is less activity during the latter phase with respect to coronal holes [*Jocelyn,* 1995], which are discussed in the following section.

Shorter time scale periodicities within the solar cycle are associated with the sector structure of the solar interplanetary magnetic field (IMF), which permeates the solar system. The solar magnetic field can be considered approximately dipolar at distances greater than a few solar radii. At these distances, the complex magnetic fields of the photosphere, the opaque visible light surface of the sun, can be neglected in models of the solar magnetic field. The high electrical conductivity of the solar wind plasma couples the solar dipolar magnetic field to the motion of the solar wind particles. As a result, the radial velocity of the solar wind, combined with the ~27 day rotation of the sun, distorts the solar magnetic field into a garden sprinkler configuration, often referred to as the Parker spiral.

The component of the IMF in the ecliptic plane can be directed inwards or outwards with respect to the sun, as the axis of rotation is inclined to the ecliptic

plane. The neutral current sheet exists in or near the equatorial plane and divides the inward IMF from the outward IMF. As the solar magnetic dipole is tilted from the axis of rotation of the sun, then a two-sector structure is observed from the earth as the sun rotates. However, perturbations can develop within the neutral current sheet, creating a four (or higher) cell sector structure at certain periods during the solar cycle.

2.5 *Solar-Geophysical Disturbances*

Although they do not directly affect the far-field dipolar approximation of the magnetic field of the sun (and hence the solar wind structure), photospheric magnetic fields are associated with features such as coronal holes, which are able to influence the solar sector structure. Coronal holes are a source of localised solar wind enhancements and persist for many solar rotations, producing a modulating effect upon the solar sector structure. These holes are cool regions of low coronal density above a quiet chromosphere - the transparent layer of gas above the photosphere. Magnetic field lines are weak and diverging (open) above the coronal hole and give rise to High Speed Solar-Wind Streams (HSSWS). These streams propagate away from the sun at speeds several times greater than the ambient solar wind speed. Signatures of these features persist for as many as ten solar rotations and create patterns of enhanced geomagnetic activity with a 27-day period.

The most short-term and energetic geophysical disturbances have been associated with solar features known as coronal mass ejections (CMEs). These events originate from regions of the sun where enhanced convection forces magnetic field lines of opposite polarity into close proximity. Sudden reconnection can then take place, which generates a considerable shock wave. This shock wave can accelerate the large mass of ionised plasma, frozen into the reconnected magnetic field lines, free of the gravity of the sun to form a CME. These magnetic clouds expand to widths of millions of kilometres as they travel away from the sun and

take two to three days to reach the earth where they may have a considerable impact on terrestrial and space based systems.

The frequency of occurrence of coronal holes and CMEs is correlated with position in the solar cycle. These events occur most frequently when solar sunspot activity is at a maximum, in the same fashion as solar flares. A solar flare is defined as a localised brightening of the photosphere that can take place over time scales ranging from a few minutes to several hours. These events can be attributed to the release of energy that has been gradually accumulated in a solar magnetic flux loop over the course of a day or so. When the loop collapses, the energy is released in a violent burst that produces high-energy particles and X-ray radiation. Solar flares are the source of many of the sporadic particle and electromagnetic emissions that affect the near earth environment. They are also responsible for the increased radio noise observed during periods of high flare activity.

Solar-geophysical disturbances can give rise to magnetospheric storms when they reach the near-earth environment. Storms events are characterised by continuous and very energetic substorm activity over several days. Significant global deviations in the geomagnetic field are also observed. The prolonged duration of these phenomena gives rise to global magnetospheric and ionospheric effects, over and above those normally associated with substorms, which are normally localised in extent to the high latitudes. This occurs primarily because the dissipative mechanisms in the magnetosphere, that result in short relaxation times for substorm activity, are not sufficient to prevent storm events from enhancing the particle population in the inner magnetosphere, which then produces global geomagnetic effects.

The classic magnetic storm is characterised by three phases. The initial phase is associated with the compression of the magnetospheric cavity as it encounters the enhanced solar wind. This causes an increase of the geomagnetic field at the surface of the earth, lasting only a few hours. At the same time, enhanced

reconnection takes place at the magnetopause and heated plasma is circulated into the magnetotail. The second phase of the magnetic storm occurs when heated tail plasma is circulated into the inner magnetosphere by increased convection. The positively charged ions in this plasma drift westward around the earth and the negative electrons drift eastward. Consequently, a current system circles the earth at a distance of several earth radii; this system is referred to as the ring current. The magnetic field produced by the enhancement of the ring current produces a significant global decrease in the strength of the H component of the geomagnetic field. This deviation reaches a maximum about 24 hours after storm commencement, before slowly decaying over the course of several days in the final phase of the storm.

2.6 The Ionosphere

The ionosphere is the ionised layer of the atmosphere that extends from 85 km in altitude to over 1000 km. This region affects all electromagnetic waves, up to a few GHz, that pass through it and therein lies its operational significance for man-made systems. Some systems are wholly dependent on the ionosphere for their operation, while the performance of other systems can be degraded by its presence. This region is highly variable in time and space, in such a way that it is very difficult to anticipate the nature of electromagnetic propagation through this medium. The ionosphere is particularly disturbed during magnetic substorms and in the auroral regions, as well as during periods of enhanced solar-geophysical activity. The high latitude ionosphere is particularly prone to instabilities. Consequently, propagation and prediction models have been developed to characterise the behaviour of the ionosphere and the passage of radio waves through it.

HF signals (3-30 MHz) can be "bounced" off the ionosphere, facilitating over-the-horizon propagation. However, the ionosphere is not a smooth reflective surface. Thus, the propagating signal can be scattered and reflected by these irregularities,

reducing the received signal strength at the receiver. In addition, multiple signals can interfere with each other. These processes adversely affect the integrity of the propagating signals. Shorter wavelength radio signals above the HF band are able to pass through the ionosphere but are still degraded by effects such as scattering and scintillation.

The ionisation of the upper atmosphere of the earth, which gives rise to the ionosphere, is produced by several mechanisms [*Hargreaves*, 1992]. The most important of these, at non-auroral latitudes, are the ionising actions of solar ultra-violet (UV) and X-ray radiation upon the upper atmosphere. This ionisation produces a sufficiently large number of free electrons that the propagation of electromagnetic radiation through this region is affected. Absorption in the atmosphere means that less intense radiation reaches the lower levels of the atmosphere leading to a reduced level of ionisation at these altitudes. However, the reduced density of the outer atmosphere means that there is less ionisation at high altitudes. These two opposing factors produce a peak in the level of ionisation at a certain altitude. This is a simplified view of the ionisation process. In reality, there are a number of different molecular and atomic species, which require different amounts of ionisation energy and which have specific concentration height profiles. In addition, different frequencies of EM radiation are attenuated differently as they pass through the atmosphere and interact with atoms and molecules to form different ion species. The plasma density at any given altitude is determined by the rate at which electrons are formed and the rate at which they recombine with positively charged ions to form neutral atoms or molecules. Diffusion can also affect the plasma density height profile by transporting charged particles to different altitudes.

Ionisation also occurs when energetic cosmic ray particles and energetic magnetospheric particles enter the upper atmosphere. At mid-latitudes, this is generally a second order effect in comparison to the ionising effects of solar electromagnetic radiation. However, these mechanisms are important in the high

latitude night-side ionosphere, where there is no direct solar illumination, at low altitudes where electromagnetic radiation cannot penetrate and in the polar cap region. In addition, this mechanism becomes more important during solar-geophysical disturbances when energetic particle populations are greatly enhanced.

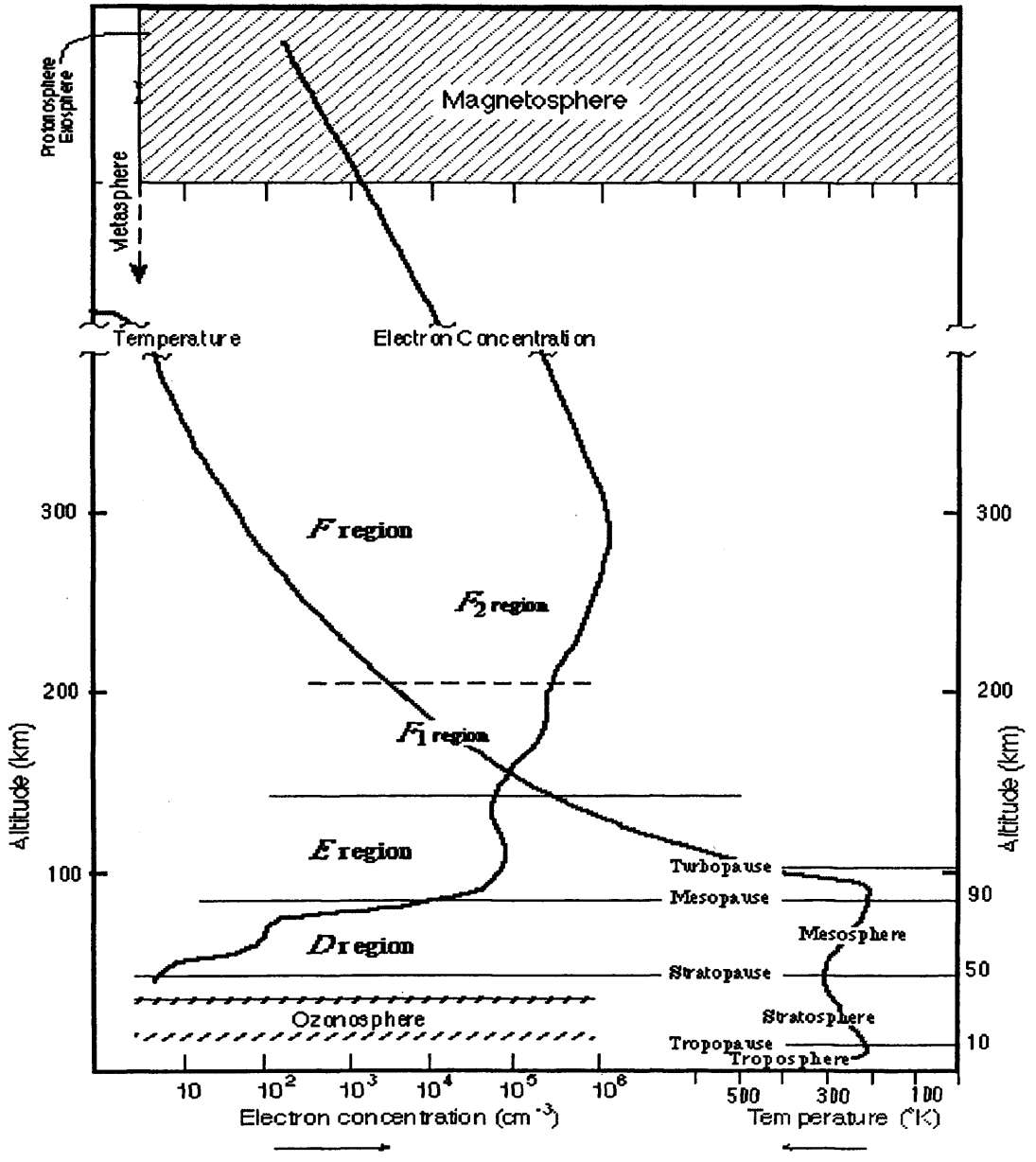


Figure 2.4 Ionospheric profile of electron density with altitude

The combination of these different factors produces a complex profile of electron density with altitude and gives rise to several distinct layers within the ionosphere, as shown in Figure 2.4. The principal layers are designated *D*, *E* and *F*, each being characterised by a different set of ionisation and loss processes. Fine structure can be found within each layer, giving rise to further sub-divisions within each region. The number of layers, their heights and their ionisation densities all vary in space and time. The high latitude ionosphere is subject to additional physical processes related to the magnetosphere, further complicating the behaviour of this region.

The *D*-region is the lowest layer of the ionosphere at 50-90 km in altitude, although there is little ionisation below 85 km. As the lowest region of the ionised atmosphere, it is produced by the most penetrating types of the ionising radiation, such as X-rays, extreme UV and solar cosmic rays. This layer is only very weakly ionised, the relative concentration of electrons is orders of magnitude lower than the neutral density. The electron density in this region exhibits strong diurnal behaviour, varying by two orders of magnitude. While the ionisation levels are low, the electron-neutral collision rate is high, leading to strong absorption at HF frequencies. At higher frequencies, losses due to this effect are negligible. In addition, the rapid recombination at this altitude, coupled with the direct radiation induced ionisation, means that *D*-region absorption is almost entirely a day-time phenomenon that disappears a few minutes after the cessation of direct solar illumination.

The *E*-region peaks at around 110 km, with a thickness of about 20 km, and can be attributed to soft X-rays and UV radiation. This layer is quite stable and predictable compared to the *F* layer above. However, the *E*-region also contains an unpredictable sub-layer, referred to as sporadic *E* (*Es*). This thin, optically dense layer acts as a smooth mirror and reflects electromagnetic signals accordingly. However, this layer is prone to transparent irregularities that can disrupt the seemingly ideal propagation properties of this region.

The *F*-region can be divided into two layers, the *F1* layer that lies between 150 and 200 km altitude and the *F2* layer that lies above. The *F2* layer is generally the denser of the two and is, therefore, of greater operational significance. The *F1* layer is dominated by solar ionisation and disappears soon after sunset. In contrast, the mid-latitude *F2* layer is maintained at night by virtue of the slow recombination rates and transport processes. At high latitudes, precipitation and convection processes maintain the *F2* region through the night, even in winter. The *F2* layer is highly variable in space and time on many different scales. The ionosphere also includes complex features such as the equatorial anomaly and the trough, which are beyond the scope of this discussion.

By processes already described, variation in the IMF and the solar wind in general are indirectly coupled to the ionosphere, particularly in the high latitude region. In the latter region, precipitation of energetic particles and plasma driven by the process of convection often dominates the electron density, particularly at night. Some of these high latitude enhancements also exhibit signatures at lower latitudes as these effects propagate towards the equator.

One important manifestation of the solar-terrestrial coupling is the geomagnetic storm, as discussed previously. This is closely related to the ionospheric storm. Ionospheric storms are characterised by an initial positive phase lasting several hours when the electron density and content are greater than normal. The main phase then follows and these quantities are reduced significantly below their normal values. The ionosphere then returns to its unperturbed configuration over a period of several days, referred to as the recovery phase. The effects are most pronounced at high and mid latitudes.

The mechanisms linking magnetic and ionospheric storms are not clearly understood, though of necessity such a link must exist given the high correlation between the two events. The ionospheric storm cannot be attributed to increased ionising flux, as solar EUV is approximately constant during these events. In

addition, the storm effects observed in the *F2* region do not extend to the *F1* region. Consequently, it is likely that the phenomena can be attributed to localised effects within the *F2* region. The major contenders include:

i) Heating from the magnetosphere that alters the chemical recombination rates in the *F2* region.
ii) Depletion of the protonosphere reducing the reservoir of available particles for the ionosphere.
iii) Changes in the neutral wind altering the transport processes in the *F2* region.
iv) Slow precipitation of particles from the ring current. This explanation is very attractive as it directly links magnetic and ionospheric storms.

2.7 *Solar-Terrestrial Forecasting*

The socio-economic effects of Space weather necessitate the adoption of techniques for predicting events that have an operational impact upon man-made systems. The origin of most geophysical variability can be traced back to the sun. Consequently, space weather forecasting of geoeffective events often benefits from the inclusion of precursive solar information and this requires monitoring of both the sun and the solar wind.

One such real-time monitoring instrument is the Solar Orbiting Heliospheric Observatory (SOHO), which has been placed at the L1 Lagrangian point between the earth and the sun. At this point, the gravitational forces of the earth and the sun exactly balance out. A satellite placed here will orbit the sun in such a way that it is stationary with respect the rotating rest frame of the earth. From the L1 point, direct observations of the sun can be used to provide a measure of forewarning for specific events that are deemed important for the task of predicting space weather.

For electromagnetic effects, SOHO can provide about eight minutes warning following an observed event. For particles, most features take 3-4 days to traverse the distance between the sun and the earth, although energetic protons may complete the journey in less than thirty minutes, if they have been accelerated to relativistic speeds. In addition to SOHO, satellites such as the Advanced Composition Explorer (ACE) also sit at the L1 point and measure the in-situ velocity, density and composition of the solar wind. This can provide information concerning the solar sector structure, which is essential for characterising reconnection at the dayside magnetopause.

In addition to satellite based instrumentation that monitors the sun and the solar wind, networks of ground stations specify the behaviour of the magnetosphere and ionosphere in real-time. It is essential that these networks are well distributed and capable of providing good quality data. Data gathered from these networks is then made available in real-time by regional World Data Centres (WDCs), in conjunction with a number of limited forecasts made at some of the regional centres. For example, ground based magnetometers are used to measure enhancements of the ring current and produce DST as an index of activity. A network of sensors also measures the total electron content (TEC) of the ionosphere across the globe. This parameter is useful for imaging the ionosphere using tomographic techniques. Ionosondes are used to measure the critical frequencies of regions within the ionosphere, at low to high latitudes.

However, characterisation of the dynamic regions of the auroral region and polar cap ionosphere requires additional instrumentation. The ionospheric convection patterns in these regions are determined in almost real-time by the SuperDARN [*Greenwald. et al, 1995*] system. This system comprises three pairs of coherent HF radars in the Northern Hemisphere and three additional radars in the Southern Hemisphere. An improved predictive capability for high latitudes using input from the SuperDARN network would be very useful. It is likely that self-determining methods can be evolved to utilise the high latitude data network to provide

accurate forecasts, given the success of this thesis in applying these techniques to the mid-latitude ionosphere.

The data gathered by these monitoring instruments can be used as input to physical or numerical models that describe the evolution of solar disturbances and their effects within the near-earth environment. However, our understanding of solar activity and associated disturbances is far from complete. Current in-service forecasting techniques are solely based upon direct observation of the sun or in situ measurement of the solar wind in real-time in conjunction with the inherent time delay between the origin of effects and their propagation to the near-earth environment. Such predictions, although qualitative or probabilistic in nature, are still sufficiently useful to meet some operational requirements.

The Space Environment Centre (SEC), based in the United States, offers a space weather warning service based upon such techniques (http://www.sec.noaa.gov). The SEC offers 24 hour forecasts of solar and geophysical activity and also acts as a repository of historical data that is relevant to space weather forecasting. It also offers an alert warning service to specific customers. The Helios Space Weather Centre in Lund, Sweden also offers a forecasting service (http://www.irfl.lu.se/HeliosHome/forecastservice.html). The service is based upon real-time predictions of space weather, gathered from various institutions, in conjunction with real-time data resources. A number of neural net models are also demonstrated.

The simple qualitative or probabilistic models that are currently employed by space weather services have reached the limits of their performance, as discussed in section 1.4. In the absence of a complete theoretical understanding of the physics of the coupled solar-terrestrial system, an alternative strategy is to use knowledge independent techniques. These methods can be used to determine the dynamics of the data and to model any deterministic behaviour that may be present. Such techniques can be applied to the ionosphere and magnetosphere and

the modelling of processes within these realms and the nonlinear interaction between them. This thesis has attempted to apply such nonlinear techniques to the characterisation and forecasting of the ionospheric aspects of space weather.

2.8 Neural Network Predictions of the Ionosphere

As described above, space weather forecasting requires the utilisation of environmental sensing data as input to physical or empirical models. Due to the incomplete understanding of the coupled solar-terrestrial system, a particularly attractive modelling solution is the application of neural networks (NNs) (see section 3.5). The application of nonlinear predictive techniques to the forecasting of the ionosphere has focused exclusively upon the use of such neural network based techniques.

The earliest NN prediction studies were not true forecasting papers, since they did not incorporate a temporal aspect, i.e. some or all of the input variables were coincident in time with the output variables to be predicted. Models of this type are more properly referred to as cross prediction or now-casting tools. For example, *Williscroft et al.* [1996] adopts such an approach for the prediction of daily and monthly *foF2* values at Grahamstown, South Africa using seasonal time information, solar activity and magnetic activity as input data. The models presented were assessed in terms of RMS error, but no indication of the standard deviation of the *foF2* time series under examination was included. It also lacked a comparison with simple models such as recurrence or persistence, which makes it very difficult to judge the relative success of the work.

Cander et al. [1997] predicted the state of the ionospheric parameter *foF2* at Slough station, during 1973, up to 5 hours ahead (identical to the test period employed for the hourly predictions in this thesis). The neural network used the previous history of the ionospheric parameter, along with magnetic and solar data, as input to provide the predictions. However, the input parameters included values

that extended in time past the point at which the prediction was being made (i.e. input information was included from one step ahead). Such information can never be available for real-time predictions and it is questionable whether or not the results have any validity at all for the one step ahead prediction.

Lamming et al. [1997] detailed the prediction of monthly median *foF2* values at Poitiers station, France using the historical monthly median time series, R12 and the time of day. R12 was coincident in time with the monthly median to be predicted, so this is another example of cross-prediction rather than forecasting. Consequently, it cannot be directly compared with the results contained within this thesis. This problem is compounded by the significant overlap between the training and testing time series, which is likely to undermine the veracity of the results.

Stamper [1996] investigated the application of NNs to predict the values of the monthly ionospheric indices of activity, I_{F2} and I_G, up to eighteen months ahead. These indices were derived from noon monthly median *foF2* values, averaged over twelve months to eliminate the annual and seasonal variations and to reveal the variation due to the solar cycle. The resultant NN model showed a clear advantage over the reference technique, which was based upon sunspot number. The author also highlighted the importance of precursive information for ionospheric predictions, a point that has also been made in this thesis.

McKinnell et al. [1999] presented a study concerned with the NN prediction of *foF2* using temporal, solar and magnetic inputs in conjunction with a previous history of the parameter to be predicted. This work was a direct extension of *Williscroft et al.* [1996] and the model from this previous paper was extended to provide hourly predictions into the future rather than just cross-prediction on timescales of days and months. A particularly novel aspect of this work was an attempt to predict the error on the prediction, in addition to providing the prediction itself. However, indication of the predictive performance of the model

was provided solely in terms of RMS error, with no reference to the standard deviation of the input data. Hence, it is difficult to judge the success of this predictive technique relative to the results contained in this thesis or reference persistence and recurrence models, although their model is shown to outperform the International Reference Ionosphere model, referred to in their text. However, no attempt has been made to assess the ability of the model to predict the errors on its own predictions. Notwithstanding these comments, the paper is clear and rigorous in its approach to the problem of ionospheric forecasting.

Wintoft et al. [1999] described a NN model that was used to predict *foF2* at Slough station from 1 to 24 hours in advance using temporal information, the AE index of auroral activity and a past history of the variable to be predicted. Comparison between the observed and predicted time series revealed that the model lagged behind the actual time series by at least one point. It is to be assumed that the NN is equivalent to a recurrence model, though this is not explicitly stated in the text. No reference models are provided to assess the results of this study, but the plots of observed versus expected output indicate that the model has difficulty modelling even mildly disturbed ionospheric conditions.

Whilst a number of useful contributions have been made, some of the recent NN based ionospheric forecast studies have lacked rigour with regard to the assessment of results and the nature of the inputs used to feed the predictive models. For example, in the absence of a suitable model to act as a control, it is often impossible to judge the effectiveness of a prediction scheme. Such a NN based prediction schemes might well offer no advantage over simple persistence or recurrence models, or could be considerably less effective in the worst cases. This is a common failing of studies that adopt a black box modelling approach using third party NN toolkits. It is a common misconception that such techniques shift the onus of understanding the physical system to be modelled from the user to the modelling tool. NNs are simply a means of constructing a nonlinear functional relationship between available input and the desired output; they do not

have the ability to capture abstract concepts in the same way as the human mind. Consequently, the context in which they are applied is of paramount importance. Successful model optimisation requires a great deal of effort and it must be approached in a studied manner. This thesis has tried to adopt such an approach.

Chapter 3: Mathematical Theory

3.1 *Introduction*

This chapter deals with the fundamental mathematical concepts that underlie the work reported in this thesis. The following section places nonlinear techniques in their historical context and outlines the stringent conditions that should be applied to these methods. Typical solar-terrestrial data sets fail to meet any or all of these conditions. The next section introduces the technique of singular value decomposition (SVD), which is used to provide the core of the neural network modelling framework and also the principal component analysis (PCA) filtering method. PCA is used both to create an orthogonal basis set from a time series for the surrogate analysis and also as a noise reduction technique for the neural network models. Section 3.4 describes a means of testing for the presence of nonlinear behaviour within a time series, based upon the application of a nonlinear predictive function [*Brown et al.*, 1999] to an orthogonal basis set created from the time series under study. This approach is utilised in chapter 4 to form the basis of the nonlinear analysis technique.

Section 3.5 introduces the basic concepts of the nonlinear neural network (NN) techniques that have been adapted to provide a novel, robust analysis and prediction scheme that is suitable for application to solar-geophysical time series. Section 3.6 details the construction of the Radial Basis Function (RBF) NN used to provide the nonlinear modelling capability used in this thesis. The same RBF techniques also form the basis of the time series prediction models described in chapters 5 and 6 [*Francis et al.*, 2000, 2001]. Section 3.7 details the construction of two hypotheses, based upon surrogate data techniques. These hypotheses are used to provide a level of confidence that the nonlinear behaviour found has physical meaning and is not just an artefact of the data or the analysis, before moving onto standard nonlinear analysis techniques. Finally, section 3.8 details

the precursive and supportive role of these techniques in relation to the application of standard nonlinear techniques.

Notation key for this chapter:

Matrices / Sets	A	italic upper case
Vectors	$\underline{v}$	italic underlined lower case
Variables / functions	x, f	italic lower case
Constants	N	italic upper or lower case
Manifolds / spaces / maps	$\mathbf{M}$	bold upper case
Set elements	$\{X_i\}$	italic upper case lower case element ordinand
Permutations	P_X	upper case upper case set ordinand

3.2 *Nonlinear Analysis*

Linear techniques have proved to be powerful tools in the characterisation and prediction of general time series. However, there are situations in which these techniques are unsuitable and nonlinear techniques need to be considered instead. Solar-terrestrial systems are one such situation, as the underlying mechanisms that govern these systems are thought to be nonlinear.

Nonlinear methods evolved from work done in the early 1970's to establish the mechanistic origins of fluid turbulence. The idea proposed [*Ruelle and Takens*, 1971] was that the apparent randomness of turbulent fluids could be related to the complex behaviour observed in low-dimensional dynamical systems that had a strange (chaotic) attractor. The original purpose of these methods was to demonstrate that the underlying attractor governing the observed dynamics of the system was of low dimension, using time series data. Several groups proposed

ways of doing this, including *Packard et al.* [1980], *Takens* [1981] and, more recently, *Sauer et al.* [1991].

The fundamental principle behind nonlinear prediction techniques is Takens theorem [*Takens*, 1981]. A brief statement of which is included:

"Let **M** be a compact manifold of dimension d. For triples (f, h, τ), where f is a smooth vector field on **M** with flow ϕ, h: $\mathbf{M} \rightarrow \mathbf{R}$ a smooth measurement function and the embedding delay $\tau > 0$, it is a generic property that the delay coordinate map $\mathbf{F}(h, \phi, \tau)$: $\mathbf{M} \rightarrow \mathbf{R}^{2d+1}$ is an embedding."

In essence, the theorem states that for typical real-valued measurements (the measurement function h) on typical low-dimensional dynamical systems (the compact manifold **M** of dimension d), it is possible to construct from the time series data alone a representation of the system (the coordinate map **F** with embedding delay τ). This construct will agree with the original, up to a nonlinear change of co-ordinates. In other words, it is possible, using this construction, to calculate from time series data any property of the system that is invariant to smooth changes of co-ordinates (the smooth vector field function f with flow ϕ).

In the first instance, this result was used to justify direct estimates of attractor dimensions using time series data. Calculations of this kind are, however, notoriously vulnerable to data limitations. Generally, they require large amounts of high quality statistically stationary data.

The implications of Takens' theorem, however, go beyond simply being able to estimate attractor dimension. A corollary of this theorem is applicable to any dynamical system, possibly of high or infinite dimension, which has a stable attracting set with dimension d. For any such system, there generally exists a

function, requiring at most $2d+1$ independent inputs, which can be used as a predictor for a time series arising from that system. Takens result guarantees the existence of nonlinear predictive functional models of the data. In addition, Takens result implies that where a given system is being monitored by making several simultaneous measurements, any single time series can be used to construct a nonlinear model of any of the others (a variant of this property is used in the next chapter). However, it provides no details about the dimension of the attractor or the derivation of the underlying function.

However, techniques are available that allow the determination of both local and global attractor characteristics. This information can be used to generate and validate prediction systems. For example, calculating the dimension of the attractor can help to determine the number of inputs required for modelling the dynamics of the system. Correlation dimension estimation [*Grassberger and Proccacia*, 1983] is perhaps the best known technique for calculating the dimension of the underlying attractor. The local scaling properties of the density of data points are investigated in this instance by calculating the distances between each pair of points in a set X of size N and counting (the Heaviside function θ) those that are less than a distance r apart to give the correlation function:

$$c(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \theta(r - |X_i - X_j|) \qquad (3.2.1)$$

The correlation dimension D_{corr} is then defined as:

$$D_{corr} = \lim_{N\to\infty} \frac{\log(c(r))}{\log(r)} \qquad (3.2.2)$$

Crudely, this technique involves determining how many pairs of phase space points are less than a distance r apart and then determining the power law dependence of this number upon r as it is increased. Each point in phase space is derived from n consecutive points within the time series and is referred to as a delay vector of embedding dimension n. A set of delay vectors is constructed by sliding a window of length n points across the time series, one point at a time. The power law dependence upon r will reach a limiting value, the correlation dimension estimate, as the embedding dimension is increased.

It is also desirable to characterise fundamental upper limits on prediction time scales. The calculation of Lyapunov exponents and Kolmogorov entropy provides an insight into this problem. Lyapunov exponents are a measure of the divergence along the phase space trajectory and, similarly, Kolmogorov entropy describes the 'information loss' along the path [*Ott*, 1993].

All the above techniques are based on certain assumptions regarding the quality of the data. The techniques have proven to be successful with time series that possess the following properties:

i. The data is well sampled in the time domain, such that the essential dynamics of the system have been captured.

ii. The underlying attractor is of low dimension, otherwise vast quantities of data are necessary for an accurate dimensional analysis. The amount of data required to characterise the properties of the attractor scales exponentially as the dimension of the attractor increases.

iii. Quantisation error is low enough to resolve the scaling properties of the attractor.

iv. The signal to noise ratio of the data is high. Any noise present in the system increases the amount of data required to obtain reliable estimates of the desired properties.
v. The underlying dynamics that produced the data come from a stationary process, i.e., a process in which the moments of the data are time invariant.
vi. The data forms a single contiguous sequence. This property is required for calculations of the attractor dimension, Lyapunov exponents and Kolmogorov entropy.

Solar-terrestrial data sets typically fail many or all of these stringent conditions. In addition, some of these conditions are mutually exclusive for geophysical data sets. For example, to get a stationary (or approximately stationary) data set, one must use only a short section of data relative to the length of the solar cycle. This may not provide sufficient coverage to reliably estimate the global properties of the attractor.

The properties outlined above have been the driving force behind the development of more robust nonlinear techniques derived from biological analogues that can be applied to time series for which these prerequisites do not hold true. These methods can be used either as an alternative or a precursor used to validate the subsequent use of established techniques in conjunction with nested surrogate hypotheses. The basic concept for the technique presented in the following chapter is determined by the physics believed to be behind the problem.

3.3 *Singular Value Decomposition and Principal Component Analysis*

Linear Singular Value Decomposition (SVD) is a powerful technique associated with matrix-based computations and analysis [*Golub and Van Loan*, 1989]. SVD is used principally to provide the optimal solution for a linear least squares problem. Applications of this technique include orthogonal filtering of an input time series (see section 3.4) and principal component analysis. The particular

algorithm adopted to calculate the SVD in this thesis is detailed in *Press et al.* [1992].

A complete proof and description of the SVD technique is lengthy and beyond the scope of this thesis and is contained in the above references, however the definition of SVD is as follows:

"Let A be a general real M-by-N matrix. The singular value decomposition (SVD) of A is the factorisation $A = U\Sigma V^T$, where U and V are orthogonal M-by-N and N-by-N unitary matrices respectively, r = min (M, N) and Σ is a matrix containing only non-zero real diagonal elements:

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \qquad \sigma_1 \geq \cdots \geq \sigma_r \geq 0$$

$$\sum_{i=1}^{M} U_{ik} U_{i\pi} = \delta_{k\pi} \qquad \begin{matrix} 1 \leq k \leq N \\ 1 \leq \pi \leq N \end{matrix}$$

$$\sum_{j=1}^{N} V_{jk} V_{j\pi} = \delta_{k\pi} \qquad \begin{matrix} 1 \leq k \leq N \\ 1 \leq \pi \leq N \end{matrix}$$

The σ_I are called the singular values, the first r columns of V the right singular vectors and the first r columns of U the left singular vectors."

SVD provides a numerically robust solution to the problem, $A\underline{x} = \underline{b}$. By noting that A^{-1} is equivalent to $(A^TA)^{-1}A^T$ then it is possible to substitute $A = U\Sigma V^T$ into the solution $\underline{x} = A^{-1}\underline{b} = (A^TA)^{-1}A^T\underline{b}$ to give $\underline{x} = V\Sigma^{-1}U^T\underline{b}$. SVD can then be used to derive V, Σ and U and hence complete the matrix inversion.

Solar-terrestrial data sets are typically very noisy. SVD can be applied to perform a process known as Principal Component Analysis (PCA) to pre-process the time series to allow selection of principal components to optimise separation of the

signal and noise subspaces [*Broomhead and King*, 1986]. Removing the noise subspace improves the performance of RBF NNs by reducing the effects of over fitting. This technique is explained more fully in the following section.

For an ideal error free measurement system, in which the data lies in a finite dimensional space, measurement data can be arranged in a matrix with sufficiently many columns, such that the matrix is known to be rank deficient. This means that some or all of the rows or columns of a matrix are linearly dependent upon each other. This linear dependency breaks down when any form of noise is present, as is the case in almost all real-life measurement systems. The resultant measurement matrix will then become full rank and independent. PCA can be used in these instances to determine a set of orthogonal basis vectors, which characterise the principal directions of variation within the vector space. A subset of the basis vectors can then be selected to produce the optimal separation of the signal and noise subspaces and to minimise redundancy in the input vector.

To achieve this optimal estimate, a sliding window, of length n, is passed along the data set, one point at a time, to construct the matrix of delay vectors, as mentioned above. Each window is a short time history of the data set. The SVD form of this matrix can be used to derive a set of n orthogonal filters. Each of these filters corresponds to the projection onto a principal vector, which in turn has an associated principal component (PC). Those filters that adversely affect the model accuracy are removed by setting the weight of the corresponding principal component to zero in Σ, incidentally reducing the dimensionality of the prediction problem and the associated processing overhead.

The magnitudes of the PCs can be plotted in order of decreasing size to assess the relative linear contribution of the output of each of these filters in characterising the variation of the time series. For a linear model, the efficiency of the predictive scheme can then be optimised by removing all filters that correspond to PCs

below a certain threshold magnitude, as determined by optimum model performance. PCA for a nonlinear model is more complicated as the magnitude of the relative linear contribution of a component is no longer necessarily correlated with the affect of that component upon the accuracy of the nonlinear model utilising the filtered input data. Thus, a simple sequential search algorithm must be applied to test each of the components in turn to determine whether adding or removing them has a positive effect upon the model output error. Each iteration adds or removes the PC that has the greatest positive effect upon the modelling accuracy until the process converges on the optimum model.

3.4 *Correlation and Dependence*

It is well known that nonlinear processes introduce dependence between data sets that is not captured by simple linear correlation statistics. A simple example of this is a time series generated by iteration of the sequence function (map):

$$b_{n+1} = 1 - 2b_n^2 \qquad (3.4.1).$$

The auto-correlation function of this time series is zero for every non-vanishing lag. There is, however, a nonlinear dependence between the data points that could be extracted from the time series by fitting a nonlinear model of the form $b_{n+1} = f(b_n)$, where f is the nonlinear mapping function described in equation (3.4.1).

The method employed in the following chapter, to look for nonlinear dependencies within a data set, is based on a generalisation of these simple facts. Assume that the time series of measurements is made on a dynamical system that evolves on a finite-dimensional state space **M** (i.e. it is not noise). Delay vectors are constructed from the data by treating consecutive data values as components of a $(d+1)$-dimensional vector. If T denotes a transpose matrix, the delay vector is:

$$\underline{z}_i = (b_i, b_{i+1}, \ldots, b_{i+d})^T \qquad (3.4.2).$$

A consequence of Takens' theorem is that for $d \geq 2m+1$ (m is the dimension of **M**) these vectors can be partitioned into two parts; a d-dimensional vector $\underline{x}_i$ and a scalar y_i, where

$$\underline{z}_i = (\underline{x}_i, y_i)^T \qquad (3.4.3).$$

such that there is a smooth (nonlinear) function f, which relates the two

$$y_i = f(\underline{x}_i) \qquad (3.4.4).$$

In general, y_i and the components of $\underline{x}_i$ are linear combinations of the components $b_i, b_{i+1}, \ldots, b_{i+d}$ of $\underline{z}_i$. As a special case, $\underline{x}_i$ can be the vector $(b_i, b_{i+1}, \ldots, b_{i+d-1})^T$ and the scalar y_i can be b_{i+d} (in which case $f(\underline{x})$ would be a predictor).

For the purposes of this thesis, the $\{\underline{x}_i\}$ need to be linearly uncorrelated with the $\{y_i\}$. This lack of correlation means that there is no non-trivial linear model that relates the two. Therefore, if a relationship of the form of equation (3.4.4) can be found which has modelling power, then there is evidence that there is some underlying low-dimensional, nonlinear mechanism at work.

There is a simple technique, based on the application of principal component analysis to all of the vectors $\{\underline{z}_i\}$, whereby this orthogonality can be arranged [*Broomhead and King*, 1986; *Horn and Johnson,* 1991]. The techniques of SVD and PCA are discussed in the previous section but are also presented here for completeness. To see this, consider the following matrix that gives the covariance of the components of the $\{\underline{z}_i\}$ averaged over a time interval, say $1 \leq i \leq N$:

$$\hat{I} = \sum_{i=1}^{N} \underline{z}_i \underline{z}_i^T \qquad (3.4.5).$$

Each term $\underline{z}_i \underline{z}_i^T$ of this sum is an outer product i.e. a (d+1) by (d+1) symmetric matrix whose (jk)th element is the product of the jth and the kth components of $\underline{z}_i$. Ξ is a real, (d+1) by (d+1) symmetric matrix and hence has an orthonormal set of eigenvectors, say $\{\underline{v}_k\}_{k=0}^{d}$. (In practice this set of eigenvectors is calculated using singular value decomposition of the matrix

$$Z = \begin{pmatrix} b_1 & \cdots & b_{d+1} \\ \vdots & \ddots & \vdots \\ b_{N-d} & \cdots & b_N \end{pmatrix}$$

by noting that, if the singular value decomposition of Z is $U\Sigma V^T$ (section 3.3), then:

$$\Xi = Z^T Z = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T,$$

i.e. the eigenvectors of Ξ, $\{\underline{v}_k\}$, are known as the right singular vectors of Z). It can be shown that time series generated by projecting the vectors $\{\underline{z}_i\}$ onto any two different principal components, corresponding to two principal vectors $\underline{v}_k$ and $\underline{v}_{k'}$, are uncorrelated. Hence, by writing:

$$(\underline{x}_i)_k = \underline{z}_i^T \cdot \underline{v}_k \qquad (3.4.6a)$$

$$y_i = \underline{z}_i^T \cdot \underline{v}_{k'} \qquad (3.4.6b)$$

it is possible to generate $\{\underline{x}_i\}$, a set of d-dimensional vectors with components $(\underline{x}_i)_k, k \neq k'$, which are uncorrelated with the $\{y_i\}$. A nonlinear modelling technique can then be applied to determine if there is evidence of a low-dimensional nonlinear mechanism in the data set in the form of equation (3.4.4).

3.5 *Neural Networks - Basic Concepts*

The past ten years have also seen a rapid growth in the evolution and application of Artificial Intelligence (AI) techniques to real-life problems. Advances in terms of available computing resources have made it possible to simulate and test self-adaptive models in ways that were not previously possible. The basic premise of "weak" AI states that it is possible to add features that are analogous to biological processes to computers to augment their capabilities. Over 600 million years of evolution, biological systems have managed to solve many of the problems that scientist's face when processing massive amounts of noisy, inconsistent and highly redundant data. Natural systems are robust, efficient, flexible and fault tolerant. Techniques such as expert systems, classification and recognition algorithms, fuzzy logic, genetic algorithms and self-organising maps are already in widespread use and have been developed from insights into biology, evolution and cognitive processes.

One particular branch of AI is relevant for the purposes of this thesis. This area comprises a group of models and associated techniques that are referred to as artificial Neural Networks (NNs) [*Hertz et al.,* 1991]. These methods have their roots in neurobiological studies of brain structure. In a fashion analogous to the neurons that compose the human brain, a generic NN is a collection of simple, highly interconnected processors ("units", "nodes" or "centres"), each of which has a small amount of local memory and processing power. The units are connected by channels of communication ("connections" or "links") that carry encoded data. In general, each unit is dependent only on local data and on the inputs received via direct connections. Such networks can be applied to many different tasks, depending on whether or not they are trained and what algorithms, functions and interconnections are used to encode data and functionality. Knowledge can be encoded in the network via a learning process or in the links used to connect the nodes. These networks have enormous potential for

parallelism, since the nodes are largely independent of each other, and their interconnectivity provides them with a high measure of nonlinear adaptability.

This thesis restricts itself to the use of supervised learning networks for the purpose of time series prediction, where a constrained set of basis functions is used to create a predictive model from a set of input-output training vectors. In this instance, a radial basis function neural network approach has been adopted and the exact implementation is described in the following section. The resultant model is validated upon test data from the same source as the training data to assess its performance in terms of its ability to generalise onto unseen data. If this process is successful, the NN model can be said to have characterised the underlying dynamic of the input data.

3.6 *Neural Networks - Radial Basis Function Implementation*

Functional approximation addresses the problem of fitting a set of vector–point pairs $\{\underline{X}_n, Y_n\}$, where $n = 1 \ldots N$, to a model of the form $Y = f(\underline{X})$, where f is a scalar function of a vector. Clearly, for a finite data set, there will be a class of functions in a function space, **F**, such that any $f \in \mathbf{F}$ satisfies these relationships.

As already described, matters are further complicated by the presence of a noise component that is associated with each measurement. By fitting the data exactly, the function inadvertently fits this noise as well. Therefore, the model is likely to perform sub-optimally on unseen data, as the unseen noise will not be predictable from that contained in the original data set. Any process that minimises the contribution of noise to the model, such as principal component analysis, will necessarily improve the predictive accuracy.

To address the problems of fitting and noise simultaneously, it is usual to select f so that the modelling error is minimised in an arbitrarily restricted class of

functions. Increasing the size of this class will result in a decrease in the error calculated over the original training data. However, on previously unseen test data drawn from the same distribution, the error will eventually begin to grow. This happens because the noise in the original data is modelled increasingly accurately, as the size of the function class is increased, until over-fitting occurs and the model is no longer general enough to predict unseen data. At the point where the test data error is minimised, the approximation is deemed optimal, since it models the general trends within the data without modelling the noise or characteristics that are specific to the training data.

There are numerous linear and nonlinear methods for determining such functional relationships between time series. The RBF neural network [*Broomhead and Lowe*, 1988] offers one approach to the solution of this problem and has been successfully applied to time series prediction problems in other areas [*Smith*, 1994]. In the most general case, given a set of vector–point pairs $\{\underline{X}_n, Y_n\}$, where $n = 1 \ldots N$, it is desirable to construct a function f that minimises the summation,

$$E = \sum_{n=1}^{N} (f(\underline{X}_n) - Y_n)^2 \qquad (3.6.1).$$

The RBF approach restricts f to a certain form, such that the solution of (3.6.7) is equivalent to that of a suitably posed linear least square problem, which can subsequently be solved using SVD (section 3.3). It is assumed that f is a linearly weighted sum of radially symmetric functions of the input data, wherein lies the nonlinear adaptability of the network, that is,

$$f(\underline{X}_n) = \sum_{i=1}^{p} w_i \varphi(|\underline{X}_n - \underline{c}_i|) \qquad (3.6.2).$$

The weights, $\{\omega_i\}$, are chosen to minimise equation (3.6.1), to complete the construction of the function, f. For the purposes of this thesis, φ is a cubic function of the form,

$$\varphi(r) = r^3 \qquad (3.6.3).$$

In this thesis, a number of individual basis functions, $\{\varphi_i\}$, were chosen with centres of radial symmetry, denoted by the set of vectors $\{\underline{c}_i\}$. These centres were placed at equally spaced intervals through the embedded time series. More complex methods of placing centres were investigated but provided marginal improvements in predictive accuracy at a high computational cost. Different basis functions were also employed but this was found to have very little influence on the predictive accuracy of the model, so cubic were selected for simplicity.

Given this formulation, equation (3.6.1) can be posed as a linear least squares problem of the form $M\underline{W} = \underline{Y}$, where $\underline{W} = [\omega_1, \ldots, \omega_p]$, $\underline{Y} = [Y_1, \ldots, Y_n]$ and

$$M = \begin{bmatrix} \varphi(\underline{x}_1 - \underline{c}_1) & \cdots & \varphi(\underline{x}_1 - \underline{c}_p) \\ \vdots & \ddots & \vdots \\ \varphi(\underline{x}_N - \underline{c}_1) & \cdots & \varphi(\underline{x}_N - \underline{c}_p) \end{bmatrix} \qquad (3.6.4).$$

Solving for $\underline{W}$ yields $\underline{W} = N\underline{Y}$, where $N = (M^T M)^{-1} M^T$ (if $p \leq n$) or $N = M^T (M M^T)^{-1}$ (if $p > n$), the Moore-Penrose inverse of a non-square matrix. We can use SVD (section 3.3) to solve this problem by noting that, if $M = U\Sigma V^T$, then $N = V\Sigma^{-1} U^T$.

As the problem reduces to a linear least squares form that can be solved using SVD, the RBF method has several advantages over the more commonly used Multi-Layer Perceptron (MLP) techniques. For a given set of centres (specified by quantity and individual locations), RBF NNs are assured of finding the global

optimum solution in a single pass, because the training process is reduced to a linear least squares problem. MLPs train using gradient descent algorithms, which determine optima on a nonlinear surface, given a set of initial conditions. There is no guarantee that a particular optimum is global and various initial conditions must be tested to find the best optimum solution. Gradient descent also relies on an iterative process, the back-propagation of training errors through the network, to adjust the network parameters to reach each optimum. Cross validation against a control data set determines when the network generalisation capabilities are optimum in each case, for any given set of initial conditions. These iterative processes are typically numerically intensive and arbitrary to a certain degree.

Consequently, RBF NNs are relatively cheap in computational terms and are more straightforward to use than MLPs. The disadvantage of the RBF approach is that its form restricts the search to a linear hyper-surface in the nonlinear function space (linear combination of nonlinear functions). Therefore, it can require more degrees of freedom (centres) to perform as well as the MLP, which searches a nonlinear hyper-surface in the nonlinear function space (nonlinear combination of nonlinear functions). This can sometimes be an issue when the amount of data is limited, as dictated by Takens theorem (section 3.2). In general, however, the ease of fitting the RBF more than compensates for any necessary increase in the complexity of the class of models. In addition, RBF NNs provide a straightforward functional relationship between input and output vectors, which facilitates understanding of the model constructed by the network. Rule extraction for MLPs is a much more complicated matter and the information contained in the knowledge matrix is not amenable to intuitive decomposition.

3.7 *Surrogate Data*

The final aspect of mathematical background needed in this thesis relates to the use of surrogate data and its application to determining the level of confidence of the nonlinear analysis results presented in Chapter 4. Essentially, given that it will

always be possible to find some relationship between any two finite data sets, the following question needs to be addressed: what is the level of confidence that the relationship that has been found has some physical meaning and is not just a result of some chance correlation within the data? In this section, the concept of surrogate data is introduced to provide an answer to this question.

Surrogate data methods [*Theiler*, 1992] are closely related to the 'bootstrap' methods of statistics. Bootstrap methods employ statistical tests to distinguish between a data set and an artificial attempt to replicate the data using a nested set of assumptions about the nature of the underlying distribution of the data. Surrogate techniques are designed to test hypotheses, such as the hypotheses outlined in the following paragraphs, by generating surrogate time series that share some statistical and / or nonlinear properties with the original time series. By comparing the original and surrogate data sets, it is possible to distinguish between processes that may have generated the data. This approach can be used in the present case to interpret the results of fitting a relationship such as equation (3.4.4).

The first null hypothesis (I) used in chapter 4 of this thesis states that the original time series was generated by a correlated Gaussian noise process. The first step is to ask how the original model compares with similar models of surrogate data sets that have the same correlation properties as the original data. Such surrogates can be generated by randomising the phases of the Fourier transform of the original data and applying the inverse Fourier transform to the result:

$$\{X_i\}_{i=1}^{N} \xrightarrow{F} F(X,\omega) \xrightarrow{\text{random } \phi} F(X,\omega)e^{i\phi(\omega)} \xrightarrow{F^{-1}} \{X'_i\}_{i=1}^{N} \qquad (3.7.1).$$

Here the notation $F(X)$ means the Fourier transform of the process X. $F(X,\omega)$ represents the component of the Fourier transform corresponding to the frequency ω. The time series $\{X'_i\}_{i=1}^{N}$ - though phase randomised - clearly has the same

power spectrum (and, therefore, the same auto-correlation function) as the original time series $\{X_i\}_{i=1}^{N}$. The generation of the surrogate data is equivalent to passing a white Gaussian noise process through a linear filter having a spectral response that is the same as the power spectrum for the original data. If the modelling errors (as quantified by the normalised mean square error) for such surrogate data sets are statistically indistinguishable from the fitting error on the original data, then it is not possible to reject the null hypothesis (I).

A more sophisticated null hypothesis (II) considers data generated by a correlated Gaussian noise process that was subsequently subjected to a fixed nonlinear transformation. This may be tested using surrogate data generated in a different way - essentially by shuffling the order of the original data values. The process is complicated by the fact that it is not sufficient to perform any random shuffle. A family of constrained shuffles is found which rearrange the original data so that it appears to have been generated by a process satisfying the null hypothesis (II). The process removes any detailed phase relationships that the original data may have had, whilst still approximating the original amplitude distribution.

Firstly, a randomly generated Gaussian series $\{Y_i\}_{i=1}^{N}$ is shuffled in such a way as to appear to be a monotonic nonlinear function of the original data $\{X_i\}_{i=1}^{N}$. To do this it is necessary to find mapping permutations Q_Γ and Q_X which respectively rearrange the sequences Γ and X into descending order (which are labelled $Q_\Gamma\Gamma$ and $Q_X X$). These mapping permutations have clearly defined inverses that transform the reordered sequences back to their original state. The desired monotone function, say ϑ, is defined implicitly by associating corresponding elements of the sets $Q_\Gamma\Gamma$ and $Q_X X$:

$$(Q_\Gamma\Gamma)_i = \vartheta\,(Q_X X)_i \qquad (3.7.2).$$

Since ϑ is defined point-wise on the set, it commutes with the permutation σ_X, i.e.

$$\vartheta\, \sigma_X X = \sigma_X \vartheta\, X \qquad (3.7.3).$$

Hence, it is possible to write,

$$\vartheta\, X_i = (\sigma_X^1 \sigma \Gamma)_i \qquad (3.7.4).$$

where i is the i-th element of the series. The set $\sigma_X^1 \sigma \Gamma$ - a permutation of Γ - is a sequence that is Gaussian distributed but, according to equation (3.7.1), appears to have been generated from X by the application of ϑ, a fixed monotonic function (i.e. the sequence increases when X increases and decreases when X decreases).

The second step in the procedure takes the resultant ordered random series and generates a number of new data sets using the phase randomisation technique summarised in equation (3.7.1). Finally, for each of these X'', a monotone relationship with the original data is constructed. In this case, however, X is shuffled to produce $\vartheta_{X''}^{-1} \vartheta_X X$ which is a time series that is distributed in a very similar fashion to X but appears to have been generated from X'' by the application of a fixed monotonic function. Whereas ϑ generates Gaussian distributed points from points distributed like X, ϑ^{-1} does the reverse. It is, therefore, the inverse of ϑ at least where the probability density of X'' is non-vanishing.

Formally, surrogate data sets can be thought of in similar terms to equation (3.7.2),

$$\{x_i\}_1^N \xrightarrow{\vartheta} \vartheta(X) \xrightarrow{F} F(\vartheta(X)-\mu) \xrightarrow{\text{random sie}} F(\vartheta(X)-\mu)e^{i\phi(\omega)} \xrightarrow{F^{-1}} X'' \xrightarrow{\vartheta^{-1}} \{x_i\}_1^N$$

although they are, in fact, permutations of the original data set. This representation demonstrates the main point of this approach: that the surrogate sets can be thought of as having been derived from correlated Gaussian processes by application of a fixed nonlinear function, ϑ^{-1}.

The particular correlation properties built into X'' are derived from those of X using ϑ in such a way that the correlation properties of $X = \vartheta^{-1}X'$ will resemble those of X. In essence, the original time series has been reordered as a phase randomised copy of a Gaussian time series with the same mean, variance and amplitude distribution as the original time series. All of the original phase information (and hence the nonlinear properties of the time series) is removed, but the surrogate bears a much closer resemblance to the original time series. This is because it is a constrained shuffle of the original points and approximately conserves the original correlation properties and amplitude distribution. In other words, while both surrogates have the same, or a similar, power spectrum as the original time series, only the type II surrogate approximately conserves the amplitude distribution of the original data.

In principle, it is possible to devise ever subtler surrogate data sets to test ever more subtle null hypotheses - possibly involving the use of established nonlinear techniques. However, if the original data cannot be distinguished from either of the surrogates described above, there would be little point in trying to do this.

3.8 *Broader Relation of Technique to Standard Nonlinear Methods*

The final point of the previous section bears further examination in the light of traditional nonlinear techniques (dimension, power spectrum, phase analysis, entropy, correlation, Lyapunov exponents etc). Established nonlinear methods are dependent upon the presence of meaningful phase information within the data. On their own, such methods are not sufficient to provide evidence of nonlinear dynamical behaviour within a time series, as they can give spurious results in the

presence of broad band noise signals [*Theiler*, 1992]. For example, it is possible to construct a coloured noise process with arbitrary power law spectrum and amplitude distribution that will give a positive result using dimensional or power law spectrum analysis. The type II surrogate is an example of just such a noise process and could be used to create spurious power law spectra. In addition, established nonlinear techniques require many good quality data points. Hence, there is a clear need for tests of nonlinearity that can be applied to analyse much smaller and less well behaved data sets.

Thus, there is an essential need for additional safeguards and significance tests when applying these standard methods. The checks take the form of a nested set of null hypotheses that can be used to identify and then classify any nonlinearity that may be present in the data. The surrogate type I test assesses whether the nonlinearity detected is significant enough to indicate that the data can be distinguished from a Gaussian noise process with the same statistical moments as the data. The type II surrogate test assesses whether the data contains the meaningful phase relations that are a necessary precondition for nonlinear dynamical behaviour. In essence, if the data cannot be distinguished from a type II surrogate then there is little to be gained performing any examination of the phase information. Likewise, any traditional nonlinear analysis method will give negative or spurious positive results. However, if both of these tests indicate the data contains significant nonlinear behaviour then it is reasonable to proceed to standard nonlinear techniques, given that there is sufficient data and that it is of good quality. In like manner, surrogate data hypothesis can be constructed around these subsequent analyses for the purpose of significance testing.

In summary, existing nonlinear analysis techniques are prone to giving spurious indications of nonlinear behaviour when given few or poor quality data points. Data processing techniques applied either before analysis, during measurement or collimation may also introduce artefacts. Even if conditions for the quantity and quality of the data are met then such techniques may well still provide spurious

results. The techniques presented in this thesis can be applied as either an alternative to established techniques, to determine less fragile and more qualitative nonlinear properties that may be present in the limited data, or as a precursor - to determine whether or not it is viable to proceed with the application of standard nonlinear analysis methods.

Chapter 4: Nonlinear Behaviour within the Solar Sector Structure

4.1 Introduction

This chapter reports on the analysis of data from the SABRE (Sweden And Britain Radar Experiment) VHF coherent bi-static radar, which was designed to probe the irregularities in the *E*-region to the north of the UK. These irregularities are indirectly related to the solar sector structure and other facets of the IMF. Using data from the SABRE VHF coherent radar, *Yeoman et al.* [1990] found evidence for two and four sector structures in the IMF during the declining phase of solar cycle (SC) 21. No such obvious harmonic features were present during the ascending phase of solar cycle (SC) 22. It was suggested, in a private communication with the author, that the structure of the heliospheric current sheet might exhibit nonlinear dynamical behaviour during the latter period. Such behaviour could account for the observed complexity of the spectrum during the ascending phase of the solar cycle.

A direct test of this suggestion, using established nonlinear methods, would require the computation of the fractal dimension of the data, for example. However, the quality of typical solar-geophysical data precludes the use of such traditional techniques for the quantitative characterisation of nonlinear systems without first conducting preliminary tests and establishing safeguards. Therefore, this thesis tries to answer a simpler question: is there any evidence that the SABRE data was generated by a (low-dimensional) nonlinear process? If this were the case, it would be a powerful indicator of nonlinear behaviour in the solar current sheet. By looking for relatively unsophisticated information, the results

should be less sensitive to imperfections in the data. If positive results are obtained using these methods then it is sensible to proceed to traditional nonlinear techniques, again using surrogates to validate the results.

4.2 *Previous Related Research*

Yeoman et al. [1990] analysed daily averages of the signal-to-noise ratio (S/N) from SABRE and found periodicities that could be matched to the solar rotation period and harmonics derived from the IMP8 satellite data. Evidence was found in both data sets for two and four sector structures during the declining phase of solar cycle 21 (SC 21), but no such features were present during the ascending phase of solar cycle 22 (SC 22). The spectrogram plot in Figure 4.1 shows the results of this analysis. The reason for this absence is not clear but may be related to the paucity of persistent co-rotating coronal holes during the ascending phase [*Jocelyn,* 1995]. The results are, however, consistent with those of *Gosling* [1976] for solar wind measurements. *Yeoman et al.* suggested, in a private communication [1996], that another possible cause of the observed spectrum complexity during the ascending phase was the intrinsic nonlinearity of the system. In other words, the observed broad band spectrum could be attributed to low dimensional nonlinear dynamics. These nonlinearities may originate very early in the system, in the solar current sheet for example, or may be due to some transfer mechanism between the sun and the ionosphere.

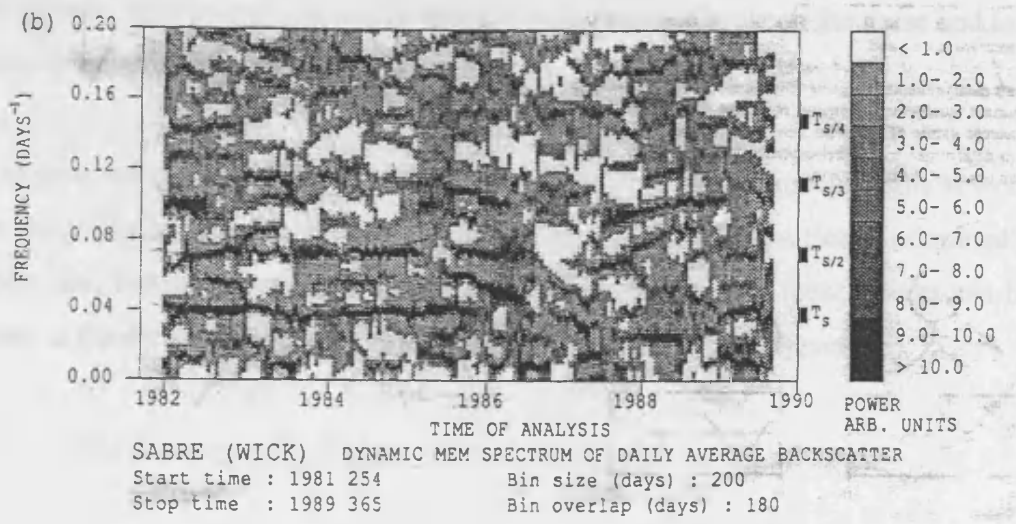


Figure 4.1 SABRE Spectrogram plot from *Yeoman et al.* [1990].

4.3 *Aim*

In this chapter, a technique is described that can determine the nonlinear properties of a data set. Firstly, principal component analysis is employed to create a set of orthogonal linear filters to separate the data into linearly uncorrelated time series. A nonlinear modelling study is then undertaken to look for nonlinear dynamical relationships between these time series, using radial basis function models (which can be thought of as a class of neural networks). The presence of such a relationship, indicated by the ability to model one filter output given another, would equate to the presence of nonlinear properties within the data [*Brown et al.*, 1999].

The data for the descending phase of SC 21 and the ascending phase of SC 22 are examined separately for evidence of nonlinear effects and their relative proportions during the two phases of the solar cycle. Statistical tests are then conducted to determine the significance of the observations, through the use of surrogate data techniques. By this means, it is possible to assess whether a similar result could have been obtained by using data generated by various noise

processes. This technique can be applied to distinguish between the noise and low dimensional nonlinear dynamics hypotheses.

The purpose is not to establish if there are nonlinear physical mechanisms at work in the underlying systems that determine the SABRE time series. Undoubtedly there are, but the purpose here is to establish whether or not these effects can be seen in the data using techniques based on finite dimensional dynamics.

4.4 *The Experimental Data*

This thesis makes use of experimental data from SABRE [*Nielson et al.*, 1983], which comprised two VHF, bi-static, phased array, coherent, backscatter radars. SABRE was operated over the period 1981 to 1994. The first station was located near Wick in Scotland, UK, and was operated by the University of Leicester. The second station was located at Uppsala, Sweden, and was operated by the Max Planck Institute, Germany. SABRE used a frequency of 150 MHz and measured both the received backscatter power and the Doppler velocity associated with irregularities within the ionospheric *E* region. The spatial resolution was approximately 20 km by 20 km, over a field of view of some 200,000km^2 covering the plasmapause boundary at high latitudes and the auroral oval during periods of high geomagnetic activity. Backscatter was recorded when ionospheric flows exceeded the threshold velocity required to generate irregularities (~300m/s). Measurements were produced every 20 seconds for each of the beams that comprised the field of view of the radar.

For this thesis, the Wick SABRE database was processed to produce a sequence of daily backscatter power values, averaged over the day and the field of view of the radar. If an average could not be constructed for a given day, due to an absence of data, an appropriate marker was inserted to indicate a "bad" data point in the time series. The Wick station was chosen as it operated almost continuously from 1981 to 1990. The Uppsala radar was operational for a much shorter period.

For the Wick station, approximately one third of all of the daily data points in the time series were classified as "bad". These missing points often appeared in contiguous blocks up to several months in length. This problem became worse near the end of the operational period of the radar. These discontinuities pose a significant obstacle for nonlinear prediction schemes, which generally require continuous time series.

A simple, but not particularly effective, solution to the problem of data gaps is to set all missing data points to a value of zero, after normalising the input time series (de-meaned and reduced to unit variance). This minimises the average mean square interpolation error, for a constant value interpolation, over the de-meaned time series as a whole.

For time series that contain a significant proportion of missing points, a more sophisticated linear interpolation scheme could also be used. However, due to the fragility of the nonlinear properties that have to be characterised, use of a linear interpolation scheme could seriously alter the apparent underlying dynamics of the system. This is particularly true for time series that contain a relatively high proportion of missing data points. Linear interpolation techniques are, therefore, highly undesirable in these circumstances.

A third simple approach is to discard those training vectors that contain missing data. In the absence of a suitably robust nonlinear method of dealing with missing data points, this was the adopted technique.

4.5 *Analysis Techniques*

The analysis technique described in this section is depicted by Flowchart 1 in Annex A. The SABRE radar data set considered covers the period from day 252, 1981 to day 84 of 1994, a total of 4579 days. Daily average signal-to-noise values were determined in like manner to *Yeoman et al.* [1990]. These values were

reduced to zero mean over the whole time series, to remove the need for a DC component in the modelling process. All missing data points were then set to zero to minimise their effects upon the subsequent principal component analysis.

The data set was divided into three portions. The first 2048 points covered the descending phase of SC 21, starting from day 252, 1981 to day 109, 1987 (Figure 4.2). The second set of 2048 points covered the ascending phase of SC 22 (Figure 4.3) from day 269, 1986 to day 127, 1991. These two sets overlapped by ~10%, to allow for the uncertainty in the position of the end of one phase and the beginning of the next. This division was made so that it was possible to assess the relative contribution of nonlinear effects within each of the two phases of the solar cycle. The remaining portion of the data was discarded due to the poor continuity of the time series near the end of the period of operation of the SABRE experiment.

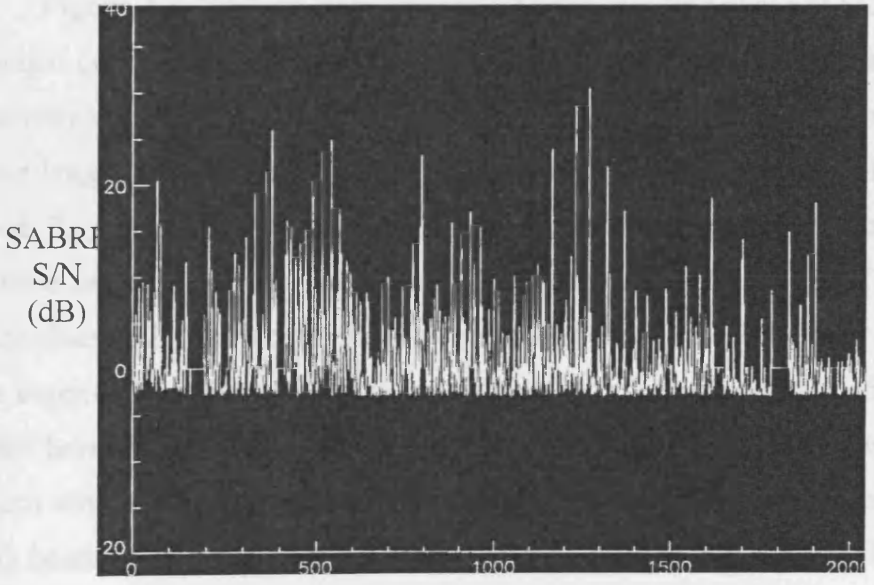


Figure 4.2 SABRE signal/noise (dB) for descending phase of SC 21.

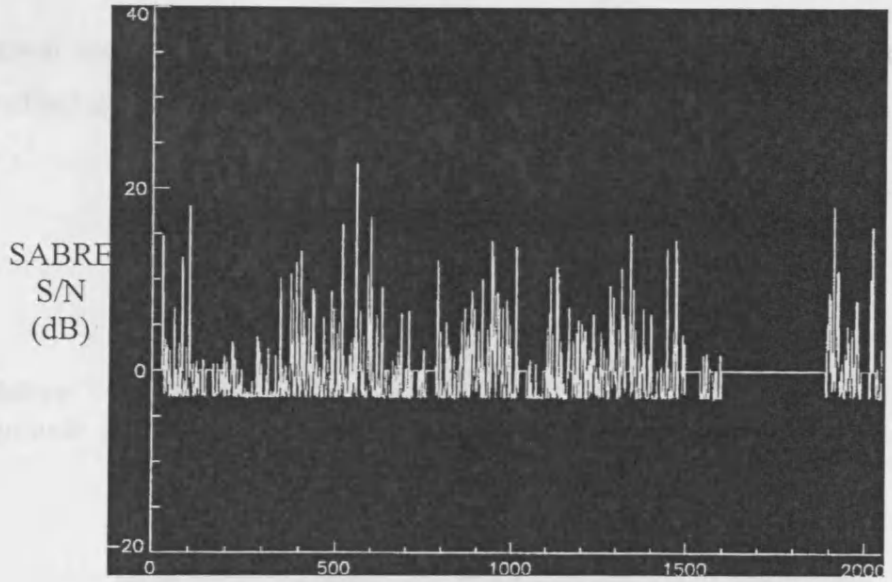


Figure 4.3 SABRE signal/noise (dB) for ascending phase of SC 22.

Principal component analysis (PCA) was performed, as described in section 3.3, separately on both of the retained portions of the data. For this analysis, a delay vector length of 29 was selected, a value that is very close to the solar rotation period, T_s. This choice is not central to the approach, however it leads to a more intuitive interpretation of the results. With this approach, some of the larger eigenvalues of Ξ, the covariance matrix, are found to come in pairs associated with eigenvectors that are sine/cosine filters at this fundamental period and its higher harmonics. This is to be expected due to the strong harmonic content present within the time series, and because the choice of window length has a direct bearing upon the features extracted by PCA. This effect is well known and is referred to by *Gibson et al.* [1992] as well as *Broomhead and King* [1986]. Figures 4.3 and 4.4 depict the magnitude of the contribution of each principal component filter to the variation of time series data from the descending phase of SC21 and ascending phase of SC22 respectively. The pairing of the eigenvalues can be clearly seen in Figure 4.4 and to a lesser extent in Figure 4.5. This property indicates that the data has a strong harmonic content. Given the results of the

original spectral analysis of *Yeoman et al.* [1990], it is not surprising to find that the effect is more pronounced in the descending phase of SC 21.

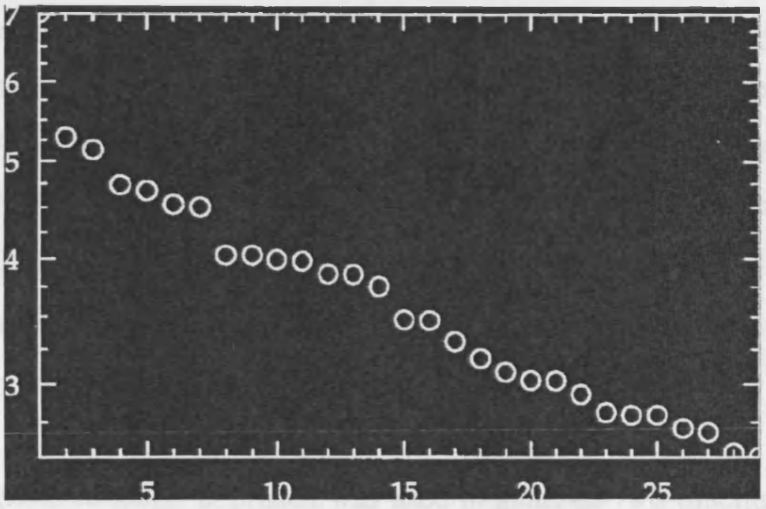


Figure 4.4 PCA spectrum for descending phase of SC 21.

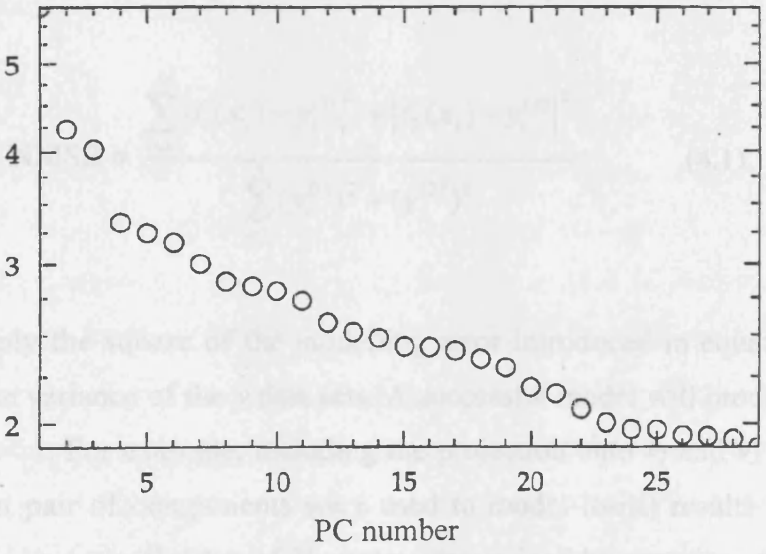


Figure 4.5 PCA spectrum for ascending phase of SC 22.

By construction, the output of any of these filters is linearly uncorrelated with the output of any of the others. Following the discussion of section 3.4, it is now necessary to look for evidence of nonlinear dependence between the filter outputs.

For the ascending and descending phases, two uncorrelated data vectors y and x (see equation (3.4.6a) and equation (3.4.6b)), were generated from the principal components. Two co-ordinates in y were adopted (in the notation of section 3.4): $y^{(1)}$, the projection of the data onto the eigenvector v_1, and $y^{(2)}$, the projection onto v_2 (v_1 and v_2 correspond to the ~T_s sine/cosine filters). The vector $\mathbf{x}$ was chosen to be the projection onto v_0 (which is a low pass filter that essentially generates the mean of the delay vector) and $\{v_k\}_{k=3}^{6}$ (the second and third pairs of eigenvectors that effectively pass higher harmonics of the fundamental ~T_s component). Including the remaining spectral information ($\{v_k\}_{k>6}$) with x does not appreciably alter the outcome of this analysis.

The radial basis function approach was used to generate two functions f_1 and f_2 that attempt to model $y^{(1)}$ and $y^{(2)}$ respectively (from x). The accuracy of the attempt was assessed using the normalised mean squared error (NMSE) of the model:

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{N} \left| f_1(x_i) - y_i^{(1)} \right|^2 + \left| f_2(x_i) - y_i^{(2)} \right|^2}{\sum_{i=1}^{N} (y_i^{(1)})^2 + (y_i^{(2)})^2} \qquad (4.1).$$

This is simply the square of the modelling error introduced in equation (3.6.1) scaled by the variance of the y data sets. A successful model will produce a value of NMSE << 1. For example, including the projection onto v_1 and v_2 into x (so that the first pair of components were used to model itself) results in an RBF model, using an optimal value of 30 centres determined by experiment, for which the NMSE is ~ 10^{-4}. If, however, the value of NMSE is close to unity, then the model has only been able to represent the mean of the data set. In this case, the modelling error on the data set is equivalent to the variation of the data set about its own mean.

Each of the filtered data sets, consisting of 2048 pairs of x_l, y_l, was split into two portions. The first 1600 pairs in each instance were used to fit the RBF model using the least squares procedure described in section 3.6. The final two hundred contiguous pairs from each time series were then used as test data to determine the optimum number of centres, and calculate their corresponding weights, for the model. This left a two hundred point buffer between the training and test set portions of the data, to avoid spurious correlations. The testing and optimisation of the model was achieved by plotting the number of centres against the model NMSEs calculated using the testing data sets. The fitting set NMSE will always decrease as the number of centres increases (it can be shown to reach zero when the number of centres is equal to the number of data points). The NMSE on the test set does not decrease indefinitely – for the reasons given in section 3.6. The optimum number of centres is the value that gives the smallest error before the test and training set errors begin to diverge significantly from each other; beyond this point, the model loses its ability to represent unseen data.

For each of the data sets, a number of surrogates, satisfying hypotheses (I) (Figures 4.5 and 4.6) and (II) (Figures 4.7 and 4.8), described in section 3.7, were created and the same filtering and modelling schemes applied. The surrogates were used to construct a distribution of possible values of NMSE. Statistical comparison of these distributions with the NMSEs obtained for the actual data set were then carried out to determine whether it was possible to reject either or both of hypotheses (I) or (II). Surrogates were assumed to have the same optimum number of centres for the RBF model as the original data set. The type I surrogate time series (Figures 4.5 and 4.6) possess only the same power spectrum as the original data. In particular, note that they do not bear close resemblance to the original data from which they were derived. The type II surrogate time series (Figures 4.7 and 4.8), however, bear a much closer resemblance to the original terms series. This occurs by virtue of being derived from the original data by re-ordering it in a way that essentially conserves the original amplitude distribution.

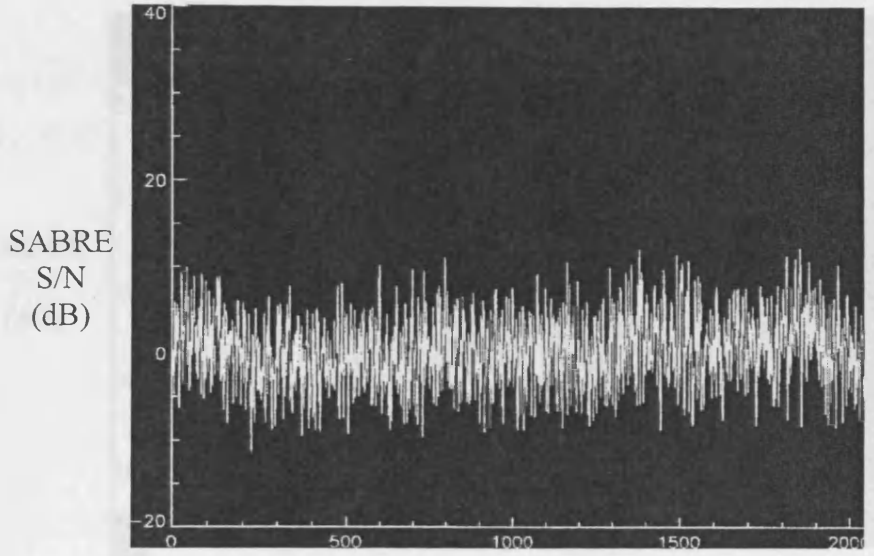


Figure 4.6 SABRE surrogate 'I' for descending phase of SC 21.

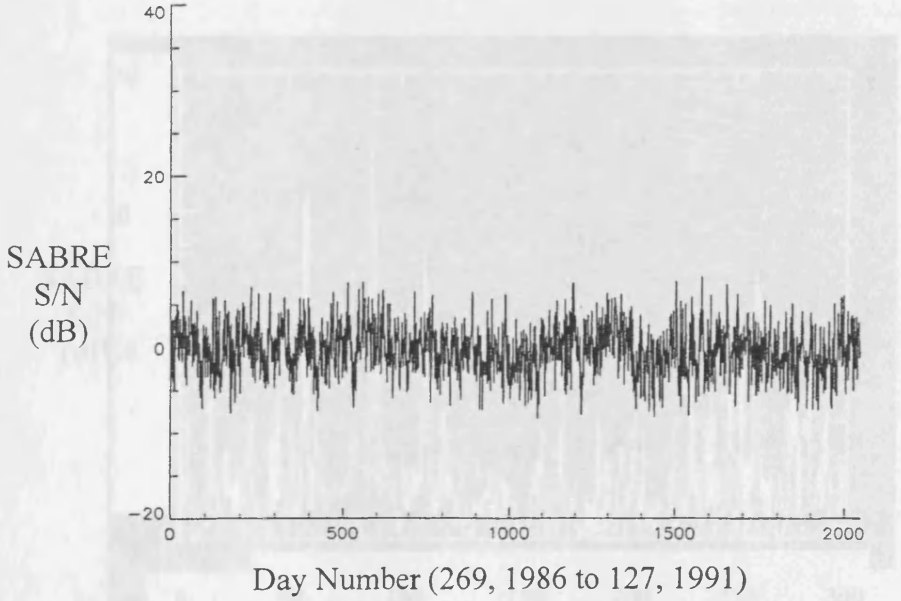


Figure 4.7 SABRE surrogate 'I' for ascending phase of SC 22.

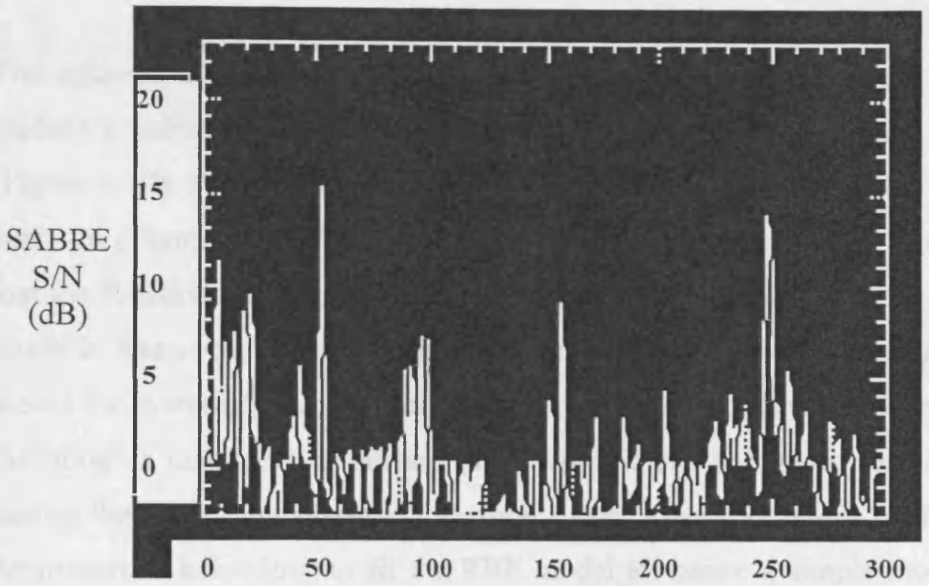


Figure 4.8 SABRE surrogate 'II' for descending phase of SC 21.

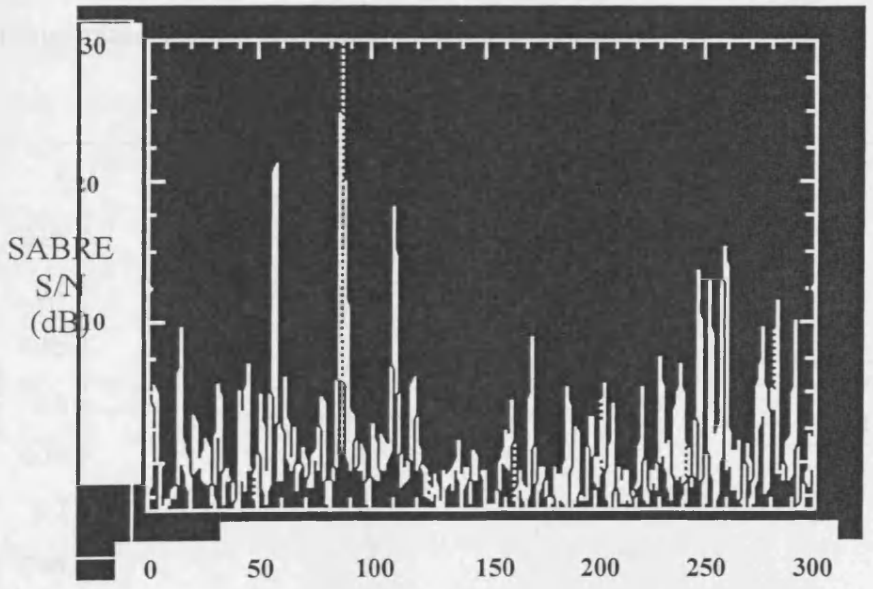


Figure 4.9 SABRE surrogate 'II' for ascending phase of SC 22.

4.6 *Results of Analysis*

The optimal number of centres (assuming the centre positions are randomly placed) required to model the descending phase of SC 21 was found to be ~68 (Figure 4.10). For the ascending phase of SC 22 the optimum number of centres was ~26 (Figure 4.11). The origins of this difference are obscure. It is possible that the functional relationship between the x and y components of the ascending phase is less complicated in the sense that it requires fewer basis functions to model the system dynamics during this period. On the other hand, it may be that the noise is just more important (and deterministic behaviour less important) during the ascending phase of the solar cycle. Therefore, there could be little deterministic behaviour to fit the RBF model to, hence a simple model would provide the optimal fit. The NMSEs for both the training and test sets are both close to unity for the ascending phase data, also indicating that there is no good RBF model for this period. Combined with the lack of observed linear periodic structure during the ascending phase, this fact tends to support the idea that the ascending phase exhibits primarily stochastic behaviour.

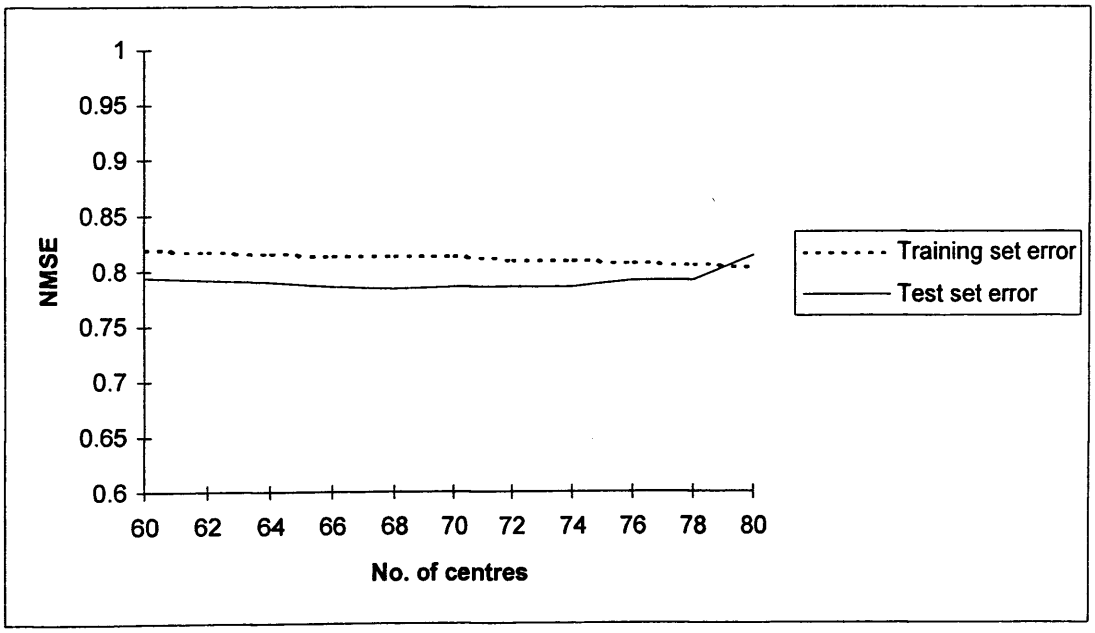


Figure 4.10 Optimal number of centres for descending phase of SC 21.

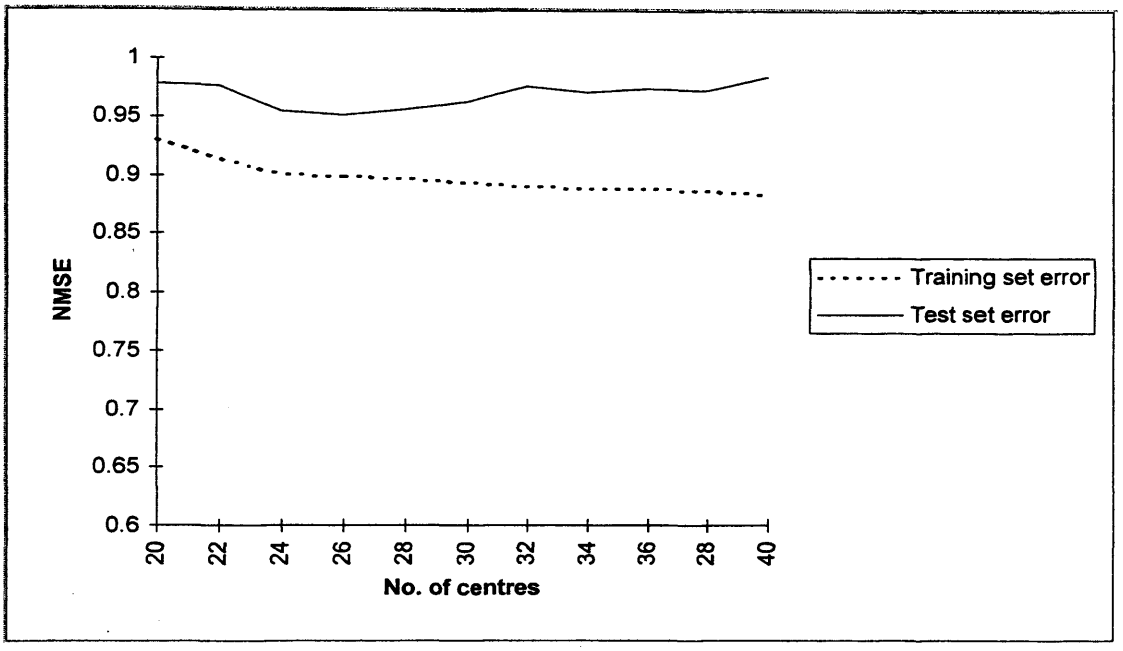


Figure 4.11 Optimal number of centres for model of ascending phase of SC 22

It should also be noted that the descending phase results show an anomaly in that the fitting NMSEs are greater than the testing NMSEs (Figure 4.10). Experimentation with different test sets revealed that this is a property of the time series. If the descending phase of the solar cycle is divided into two equal lengths, the anomaly follows the latter portion of the time series, regardless of whether or not it is used for training or testing the model. This anomaly can probably be attributed to non-stationarity of the statistical moments of the SABRE time series, taken over the solar cycle as a whole.

For surrogate data of type (I) – a correlated Gaussian process – any set of orthogonal, linearly filtered components of the data will be independent. A Student's test was employed to look for any significant difference between the fitting errors for the real data and the type (I) surrogates generated using the data. The t-statistic for a two-sided test concerning the population mean is given by:

$$t = \frac{\bar{x} - \mu_0}{s\sqrt{n}} \qquad (4.2).$$

where $\bar{x}$ denotes the mean NMSE of a number of surrogate data sets, μ_0 is the value of NMSE observed for either the ascending or descending phase of the solar cycle, s is the standard deviation of the NMSE of the surrogate data sets and $n = 5$ is the number of surrogate data sets employed. This quantity has a t-distribution with $(n\text{-}1)$ degrees of freedom. The means and standard deviations for both the descending and ascending phases are described in Table 4.1.

Using this value of t in conjunction with a table of probability values for the t-distribution provides a quantitative measure of confidence in the null hypothesis that the errors for surrogate data and real data have been drawn from the same sample. For SC 21, the value of the t-statistic for the descending phase is 8.47. A value of 10.72 is likewise obtained for the ascending phase of SC 22. Both results are significant at the 99.5% level, so there is a high level of confidence that the deviation from the mean of the surrogate time series is significant. Thus, the first null hypothesis must be rejected in both cases. In conclusion, therefore, there is some evidence for a nonlinear dependence within the data. However, the magnitudes of the test NMSE values suggest that, particularly during the ascending phase of SC 22 (see Table 4.1), there is no good low-dimensional nonlinear model of the dependence between uncorrelated components.

Data set	Training error	Test Error
Descending phase of SC 21	0.814457	0.783399
Mean of surrogates (method I)	0.968563	1.071137
σ of surrogates (method I)	0.001719	0.075972
Mean of surrogates (method II)	0.878556	0.981377
σ of surrogates (method II)	0.042104	0.246892

Ascending phase of SC22	0.898442	0.951368
Mean of surrogates (method I)	0.984017	1.011248
σ of surrogates (method I)	0.002897	0.012487
Mean of surrogates (method II)	0.813867	0.975038
σ of surrogates (method II)	0.056054	0.223164

Table 4.1 NMSE results for optimised RBF NNs for original and surrogate data

However, a nonlinear dependence does not necessarily imply the presence of nonlinear dynamics. Thus, it may not be necessary to invoke nonlinear dynamics to explain these results. An alternative to a nonlinear dynamical process would be that the data is the result of applying a fixed nonlinear transform to a Gaussian noise process (possibly a radar systems effect). It is possible to distinguish between these two types of nonlinear behaviour by estimating the significance of the NMSE results obtained for the type (II) surrogates. The normal statistic for a two-sided test concerning the population mean is given by:

$$p = \frac{\bar{x} - \mu_0}{s} \quad (4.3)$$

As the number of surrogate time series used in this instance was 1000, the normal distribution can be used in preference to the t-distribution. Using this value, in conjunction with a table of probability values for the normal distribution, provides a quantitative measure of confidence in the null hypothesis that the observations are consistent with a fixed nonlinear transformation of a correlated Gaussian process. For SC 21, the value of the normal statistic for the descending phase is 0.802. A value of 0.106 is likewise obtained for the ascending phase of SC 22. Neither of these deviations from the mean is significant, even at the 90% confidence level, so the null hypothesis cannot be rejected – the data provides

insufficient evidence to distinguish nonlinear dynamics from a nonlinearly transformed coloured Gaussian noise process.

4.7 Conclusions

Studies of the solar wind – ionosphere – magnetosphere coupling reveal periodicities of the order of T_s, the solar rotation period. From the SABRE data, T_s and higher harmonics can be clearly observed within the high latitude ionosphere from spectra extending over 200 days. This clearly indicates that a definite two or four region solar sector structure is present during the descending phase of SC 21. No such obvious structure was present during the ascending phase of SC 22, when the behaviour of the heliospheric current sheet is expected to become more complex and dynamic.

The possibility that the structure of the heliospheric current sheet may exhibit nonlinear behaviour during the latter period has been studied. A novel technique has been used to investigate this proposition. Many of the existing nonlinear data analysis techniques were developed for application to well controlled experimental systems and even numerical data. Such data sets are generally well sampled, stationary and do not suffer from missing points, in stark contrast to most solar-terrestrial data sets. Consequently, it is often inappropriate to apply these techniques to geophysical time series without precursive study and adequate safeguards. The technique employed in this instance, however, is sufficiently robust to be of use in instances where high quality data is not available and it has wider applicability to a range of real-world classification problems. Its robustness is, however, its limitation. In contrast to the other nonlinear analysis techniques, it provides little detailed information. Given that these techniques give positive results, it is then possible to apply more sophisticated nonlinear methods to obtain more precise information concerning the nonlinear properties of the data. Again, using surrogate hypotheses must be used to validate the results.

Although there is evidence of nonlinear behaviour in the SABRE data (rejection of null hypothesis I), in both the descending period of SC21 and the ascending period of SC22, it is not distinguishable from a nonlinearly transformed coloured Gaussian noise process in either case (acceptance of null hypothesis II). The absence of meaningful phase information then precludes the application of further nested hypothesis that might be constructed based upon the application of traditional nonlinear techniques. This topic is discussed at greater length in Chapter 3.

The RBF models of the dependency between filter outputs produce very large errors, particularly in the case of the ascending phase of SC22. This suggests that the former phase is dominated by the purely linear structure noted by *Yeoman et al.* [1990]. It also suggests an unresolved high dimensional (>10) or noise process is dominant during the latter phase of the time series, due to the absence of linear behaviour noted during that phase. Thus, it is not possible to support the hypothesis that increased nonlinear dynamical behaviour dominates, or even plays a significant part in, the behaviour of the ascending phase of SC 22. Contrary to expectations, the NMSE evidence seems to suggest that the descending phase of SC 21 exhibits more obvious nonlinear properties than the ascending phase of SC 22. The lower number of centres required to model the ascending phase of SC 22 also suggests that the ascending phase has less structure and is more noise-like than the descending phase.

These conclusions carry the proviso that this thesis has restricted itself to an investigation on the time scales dealt with by *Yeoman et al.* [1990]. The SABRE data used has been averaged to give a set of daily values. Thus, this analysis takes no account of nonlinearities that may be present on time scales of less than a day.

An absence of appreciable levels of nonlinear behaviour implies that the apparent lack of structure with the data set is attributable to stochastic processes, or high dimensional mechanisms that cannot be resolved given the data and the available

techniques. The very low-level nonlinear behaviour that was observed in this study could be attributable to the characteristics of the measurement system itself, or to some nonlinear transfer mechanism within the solar-terrestrial environment.

If stochastic processes do play a significant role in the observed geophysical systems, this will have significant implications for solar-terrestrial prediction. Stochastic phenomena are inherently unpredictable and are only tractable using probabilistic models. The source of this noise could lie in a number of areas. Instrument noise, the processing of the data, or the effects of non-contiguity of the time series could be masking or disrupting any structure that might be present. It is also conceivable that some mechanism in the solar wind, magnetosphere or ionosphere may be imposing a dominant noise-like process over the modulation of the solar sector structure upon the solar wind. Non-stationarity of the time series due to the long-term variation over the solar cycle may also give rise to misleading results. Finally, it may be that the solar sector structure itself exhibits stochastic behaviour during the ascending phase of SC 22. The near solar environment could be dominated by features that have a duration much shorter than the period of solar rotation, due to the paucity of coronal holes. This could account for the observed lack of recurrent structure based upon T_s, observed in the original study.

In conclusion, therefore, there is insufficient evidence within the SABRE data set to support the suggestion of increased nonlinear dynamical behaviour during the ascending phase of solar cycle 22. In fact, nonlinear dynamics would seem to exert very little influence within the measurement time series at all, given the observed data. Therefore, it is likely that stochastic or unresolved high dimensional nonlinear mechanisms are responsible for the observed spectrum complexity during the ascending phase of SC 22.

Chapter 5: Nonlinear Prediction of the Ionospheric Parameter *foF2*

5.1 *Introduction*

In this chapter, the application of nonlinear radial basis function (RBF) neural networks (NNs) to improve the accuracy of the prediction of ionospheric parameters is presented [*Francis et al.*, 2000]. Principal component analysis (PCA) is adopted for the purposes of noise and dimension reduction. Hourly, daily and monthly predictive models have been created for the Slough, UK, *foF2* time series. The quality of the model predictions is evaluated by comparison with corresponding predictions from reference persistence or recurrence models.

Each RBF NN offers a significant improvement over the performance of the corresponding reference model. The noonday model gives a performance improvement of approximately 60% over the baseline persistence model, for a one-day ahead prediction. For a one-hour ahead prediction, the hourly model offers an improvement of approximately 45% over the baseline 24-hour recurrence model. Finally, the monthly median model gives a performance improvement of approximately 40% over the baseline persistence model, for a one-month ahead prediction.

5.2 *The Experimental Data*

This chapter utilises the *foF2* time series corresponding to the ionosonde station at Slough, UK. The critical frequency of the *F2* layer of the ionosphere, *foF2*, is one of the most significant parameters of the ionised upper atmosphere. As this area often contains the highest electron concentration within the ionosphere, it generally determines the maximum usable frequency for HF communications. It also characterises that portion of the bottom side ionosphere that is farthest from

the ground, a fact that has direct bearing on the maximum usable range of a communication link. For a fixed elevation angle, transmission range increases as the maximum height of the point of reflection increases.

The first RBF model was trained and tested using noonday *foF2* values from Slough station covering the period from 01 January 1957 to 31 December 1990. This represents the entire available noonday record available on CDROM and provides a time series of 12,418 points. The first 10,000 were used to train the RBF model while the remainder of the points were used to test the predictive performance of the model on unseen data. The standard deviation of the training set was 2.65 MHz.

The second RBF model was trained and tested using hourly *foF2* values from Slough station. Much more data was available than could actually be used (1957-1990), given existing computer resources. An arbitrary decision was made to select data covering the period from 01 January 1971 to 31 December 1973. This enabled the model to generalise from year to year, whilst keeping processing overheads to an acceptable level. Consequently, the generality of this model may not extend over the whole of the solar cycle. The total length of the time series was 26,304 points. The first 23,000 were used to train the RBF model, while the remainder of the points were used to test the predictive performance of the model on unseen data. The standard deviation of the training set was 1.92 MHz.

The final RBF model was trained and tested using monthly median *foF2* values from Slough station covering the period from 01 January 1951 to 31 December 1995 (the entire available monthly median record). The total length of the time series was 12,960 points. The first 10,000 were used to train the RBF model, while the remainder of the points were used to test the predictive performance of the model on unseen data. The standard deviation of the training set was 2.31 MHz.

Each time series, as a whole, was normalised to zero mean and unit variance. Normalisation of each input time series removes the need for an offset component parameter for the radial basis functions.

5.3 *Missing Data Points*

The complete noonday time series contained approximately 8.9% missing data points. The hourly and monthly median time series contained approximately 6.6% and 1.4% missing data points, respectively. In addition, these missing points are not randomly distributed through each time series, introducing a potential source of sampling bias into each model. A missing data point may correspond to an equipment failure, or ionospheric conditions may prevent a clear measurement of the properties of the *F2* layer. Another class of data dropouts includes those values that fall outside the accepted range of variation for the *foF2* parameter. These discontinuities pose a significant obstacle for prospective nonlinear prediction schemes, which generally require continuous time series.

Potential strategies to deal with missing data points are similar to those elucidated in Chapter 4 and each carries with it the same advantages and disadvantages. In the absence of a suitably robust nonlinear method of dealing with missing data points, it was decided to reject incomplete input vectors.

5.4 *Creation of Data Vectors*

A one dimensional sliding window was passed along each time series, one point at a time, to construct the matrix of input vectors that are required to train the RBF model. Input vectors that contain one or more missing data points were then removed from the training set. The optimum window length was selected by assessing the relative performance of the one step ahead prediction model using a range of input window lengths. The optimal model parameters derived for the one

step ahead prediction were then also adopted for the models used to predict from 2 to 30 steps into the future, due to processing constraints.

The prediction step ahead increment was a single point for the hourly and daily predictions. However, the monthly median time series contains 24 concomitant hourly median values for each month on record. Therefore, it was necessary to set the prediction step ahead increment to 24, to ensure that the model predicts from 1 to 30 months ahead, rather than 1 to 30 points ahead.

5.5 *The Modelling Process*

In each instance, training and testing data were processed entirely separately; i.e. the sliding window used to create the input vectors was not allowed to overlap both of the data segments at the same time. The analysis technique described in this section is depicted by Flowchart 2 in Annex A.

Missing data points gave rise to approximately 90% rejection of possible input vectors for the noonday value predictions. The corresponding rejection rates for the hourly and monthly median predictions were approximately 60% and 15%, respectively. Thus, the problem was most severe for the noonday value test set. In this instance, the bias due to the effects of non-randomly scattered missing data points and the statistical uncertainty of the error measurements is likely to be more significant.

The training vector set was projected onto its principal axes using SVD. If PCA noise filtering was to be performed, principal components that adversely affected the prediction were removed, otherwise all of the principal components were used for the subsequent modelling process.

The remaining filter outputs were then used to train the RBF model. The optimum number of functional centres was determined by assessing the relative

performance of the one step ahead prediction model using a range of values for the number of centres.

A value of 140 centres and a window length of 90 days were found to give the optimal error results for the noonday model. The hourly model required 153 centres and a window length of 61 hours, whilst the monthly median model required 398 centres and a window length of 73 points.

Models were constructed, using the same parameters, to predict from 1 to 30 steps ahead, for each of the three models. Due to limitations on processing resources, it was not possible to optimise the predictions for each step value. The results were compared with the corresponding reference model predictions, using the root mean square (RMS) error.

5.6 *Additional Comments on the Median Model*

The monthly median time series used in the median model used all 24 hourly values for each month on record. Thus, the input window of length 73 points covers a period slightly longer than three months, in which the input values are not successive in the strictest sense. The additional complexity of the median model, reflected by the larger optimal number of centres, may be required so that the RBF model contains sufficient adaptability to process the median input vector correctly, given its more complex format.

The ideal solution to the monthly median prediction problem is to create 24 separate monthly median models, one for each hourly median. However, the median time series contains insufficient data points (approximately four hundred per hourly median) to successfully train and test a monthly median neural network model for each hour of the day in isolation. This problem was compounded by the low ratio between the number of available data points and the length of the optimal input window for such a model. These two factors led to an extremely low

number of input vectors, further reduced by the subsequent rejection of incomplete vectors. A composite model, which combines the training vectors for each separate hour together, in conjunction with an appropriate hourly time stamp for each vector, also proved non-viable. Insufficient training data was available to produce an RBF model with the required performance.

The technique finally adopted used an input window that included all the monthly median values over a period of about three months. It was the only candidate that produced results that were an improvement upon the reference model, for a one step ahead prediction. This median RBF model has enough input data to create a satisfactory model and utilises cross-prediction between hourly monthly median values to achieve prediction accuracy results that are comparable to the hourly and noonday RBF models.

It is interesting to note that longer-term information, covering one or more solar cycles, seems to be of little consequence to the one step ahead median predictions, for which the model was optimised. Under this scheme, increasing the input vector length led to degradation in the predictive accuracy of the model. This observation can probably be attributed to increased redundancy in the input vector or to under-training of the model due to increased input vector rejection. However, it is likely that multiple step ahead predictions will require additional longer-term inputs, covering one or more solar cycles, to produce optimal results.

5.7 *Reference Models*

A successful predictive model must offer a significant increase in performance over the reference technique to prove the value of the method employed. The persistence reference model predicts that the value of an observable at some specified point in the future will be identical to the current measurement of that observable. This model performs well for quiet time conditions, when the terrestrial environment is relatively undisturbed. However, it cannot predict the

onset of the short-lived impulsive disturbances that characterise periods of unusually high geomagnetic activity. Similarly, the recurrence reference model predicts that the value of an observable at some specified point in the future will be identical to the measurement of the same observable a set number of steps before the point to be predicted.

5.8 *Noonday foF2 Value Predictions*

5.8.1 *Results*

The performance of the noonday RBF NN model has been quantified in terms of the RMS error, and has been compared with the reference persistence model for prediction time scales of one to thirty days ahead (Figure 5.1).

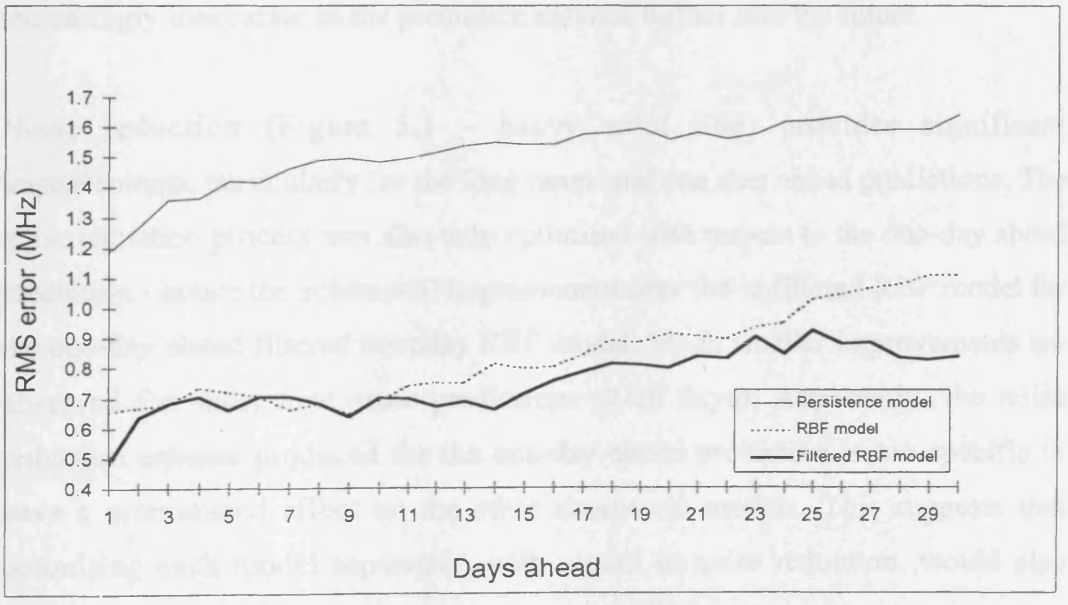


Figure 5.1 RMS prediction error for *foF2* noonday test values.

For a one-day ahead prediction, the noonday RBF model (no noise filtering) gave an RMS error of 0.655 MHz, in comparison with an RMS error of 1.186 MHz for the persistence model. This represents an increase in performance of

approximately 45% over the persistence model. Optimising the noise filtering process led to the removal of 36 of the PCs that adversely affected the accuracy of the model. For a one day ahead prediction, the noonday RBF model (with PCA noise filtering) gave an RMS error of 0.484 MHz; a relative increase in performance of 59% over the standard persistence model.

The standard unfiltered noonday RBF model (Figure 5.1 - dashed line) represents a significant improvement over the standard persistence model. This is particularly true up to 10 days ahead during which time the RBF predictive accuracy stays roughly constant. The performance increase margin narrows as the prediction time scale increases to thirty days and the model error rises more quickly. The likely cause of this effect is that the optimised model parameters were only derived for the one-day ahead prediction, due to limitations on available processing resources. The optimal one-day ahead model parameters become increasingly inaccurate as the prediction extends further into the future.

Noise reduction (Figure 5.1 – heavy solid line) provides significant improvements, particularly for the long range and one step ahead predictions. The noise reduction process was also only optimised with respect to the one-day ahead prediction - hence the substantial improvement over the unfiltered RBF model for the one-day ahead filtered noonday RBF model. Much smaller improvements are observed for short time scale predictions (2-10 days). Apparently, the noise reduction scheme produced for the one-day ahead prediction is too specific to have a pronounced effect on the other short-term models. This suggests that optimising each model separately, with regard to noise reduction, would also provide substantial improvements in terms of predictive accuracy.

Using noise reduction, the short-term prediction accuracy plateau now extends out to 15 days - five days longer than for the corresponding unfiltered noonday RBF model. Noise reduction provides tangible improvements for medium time scale predictions (10-25 days), giving a constant improvement in accuracy across the

whole period. The effects of the noise reduction scheme upon the longer-term predictions (25-30 days) are even more pronounced - probably due to the effects of the solar rotation, which gives rise to recurrent structure with a 27-day period. This periodic structure could account for the fact that the one step ahead noise reduction scheme is also applicable for the predictions one solar rotation ahead.

Figure 5.2 presents the RMS difference between the noonday RBF model and the corresponding persistence model for each prediction step. This information allows assessment of the similarities between the model produced by the RBF and the reference persistence model to be made, for varying time scales. The effects of noise reduction on these features can also be determined.

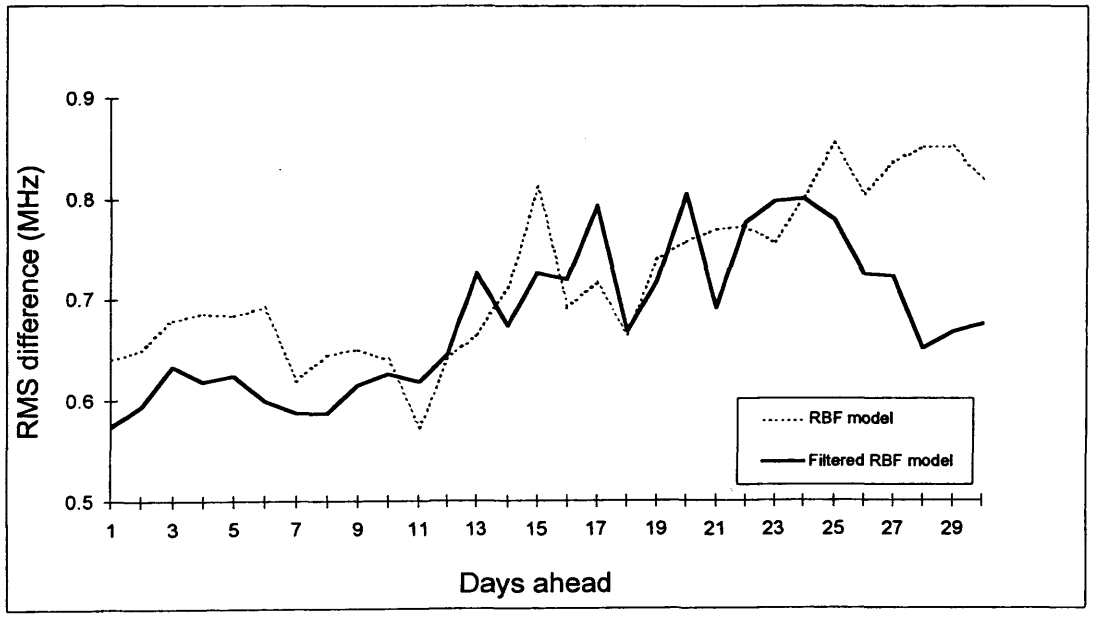


Figure 5.2 RMS difference between RBF and persistence models.

The divergence of the unfiltered noonday RBF model from the corresponding persistence model increases markedly after twelve days. This point coincides with the longer term narrowing of the performance increase margin noted earlier. This implies that, while the RBF and persistence model errors are converging, the actual structure of the RBF model is diverging from that of the persistence model.

It is, therefore, possible that including persistence (or first differences of the original time series) information into the RBF model could provide a worthwhile accuracy improvement over the standard RBF models, for medium term predictions. This divergence can also be observed for the filtered noonday RBF model. However, in the latter instance, the relative difference undergoes a dramatic decrease after 25 days. This effect is coincident with the improvement in predictive accuracy over the unfiltered noonday RBF model noted in Figure 5.1. In that graph, the filtered noonday RBF model is markedly superior to the unfiltered noonday RBF model for prediction time scales greater than 25 days. These two observations lead to the deduction that the noise reduction process improves the ability of the RBF to successfully model structure that can be attributed to periodicities based upon the solar rotation (25-30 days).

5.8.2 *Strategies for Improving the Noon Daily Predictions*

A detailed comparison between the testing data set and the corresponding prediction output time series (Figure 5.3) reveals those areas where the noonday RBF prediction is least effective. The predicted *foF2* values at the onset of periods of unusually high geomagnetic activity lag behind the actual time series. In addition, the model underestimates the peak magnitude of these events. These events account for a significant proportion of the overall RMS error. However, in this respect the performance of the noonday RBF models is still superior to the reference persistence model.

Three factors could account for the relatively poor behaviour of the model during periods of increased geophysical activity. Firstly, periods of elevated ionospheric activity are rare in comparison with quiet time conditions. Therefore, the training process is biased against the prediction of these infrequently occurring events, and in favour of accurate prediction of quiet time conditions. Secondly, the time scales associated with the evolution of such disturbances are very much smaller than the resolution of the input time series. Therefore, the input time series does not

contain sufficient information about the dynamics of these events to accurately model their rapid development. Use of additional higher resolution inputs would be required to address this issue. Thirdly, the current RBF model relies solely upon a single input time series to produce a prediction. Certain data sets are capable of providing precursive information that has a direct bearing upon the prediction of *foF2*. This might include direct measures of solar variability, solar wind activity or magnetospheric time series. Including these time series as additional inputs to the model would provide a measure of forewarning that could improve the prediction accuracy during periods of increased ionospheric activity. However, incorporating additional inputs into the model will increase the amount of training data required by the model. If insufficient training data is available, the performance of the model will be severely degraded.

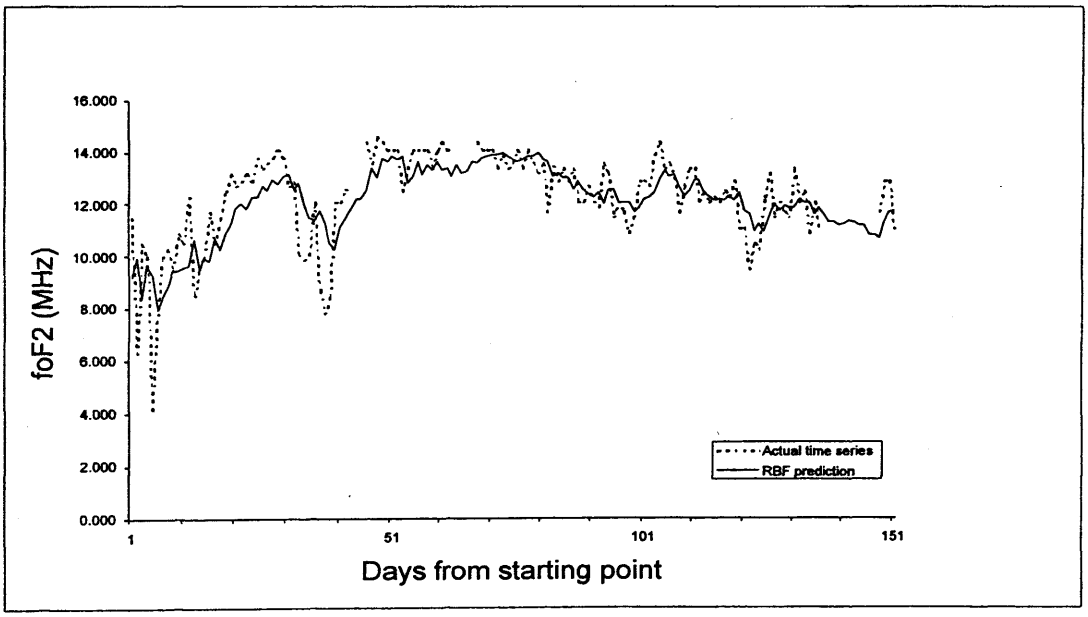


Figure 5.3 Comparison of predicted RBF and actual *foF2* daily values.

Generally, as the number of model inputs increases, the number of degrees of freedom required by the model also increases, reflecting the increased complexity of the modelling problem. As a rule, the model must be trained using at least twice

as many training vectors as there are degrees of freedom (hidden nodes or centres) in the model. Ideally, the number of training vectors should exceed the number of degrees of freedom by a factor of ten or more. Use of additional time series inputs will also increase the amount of processing time required in a polynomial fashion.

It is also possible that the tracking of storm onsets could be improved using a positive feedback mechanism to improve the responsiveness of the RBF model. One possible method of achieving this would be to auto-correlate errors on previous predictions into the current prediction

Finally, geophysical theory suggests that linear functions may be less prone than nonlinear functions to over-fitting for quiet time conditions. Cursory studies have suggested that a hybridised technique, containing both linear and nonlinear aspects, provides further enhancements of the models capabilities. In such a hybrid model, the linear functions are used to model the linear characteristics of the time series, while the nonlinear functions are used to model the residue, using the complete time series as input. This suggestion seems to be supported by observation of the behaviour of the relative difference between the RBF and persistence models, as outlined in the previous section.

Given the quality of the noonday time series, it may be preferable to construct an alternative daily time series such that input vector rejection is less of an issue. The number of missing points could also be reduced by using persistence or recurrence to fill in the each gap. However, the interpolation process should only be allowed to search up to a set number of steps into the past to find a suitable replacement point, to limit the deleterious effects of this scheme.

5.9 *Hourly foF2 Value Predictions*

5.9.1 *Results*

Figures 5.4 and 5.5 describe the relative performance of the hourly *foF2* value RBF NN model, in terms of the RMS error, for prediction time scales of one to thirty hours ahead. This time, the results are compared with both the recurrence and the persistence models. The recurrence model utilises the 24-hour cyclic nature of the hourly time series such that the value of the predicted measurement is the same as the measurement 24 hours before the time for which the prediction is required.

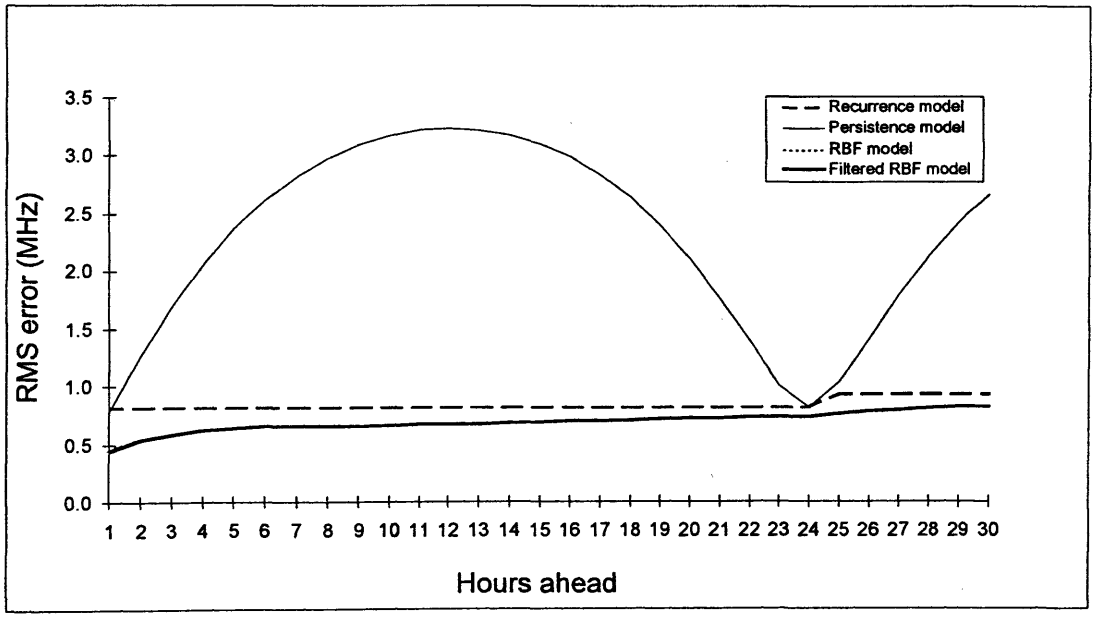


Figure 5.4 RMS prediction error for *foF2* hourly test values (large scale).

For a one-hour ahead prediction, the hourly RBF model (no noise filtering) shown in Figure 5.4 gave a RMS error of 0.452 MHz, in comparison with a RMS error of 0.781 MHz for the persistence model - an increase in performance of approximately 42.1%.

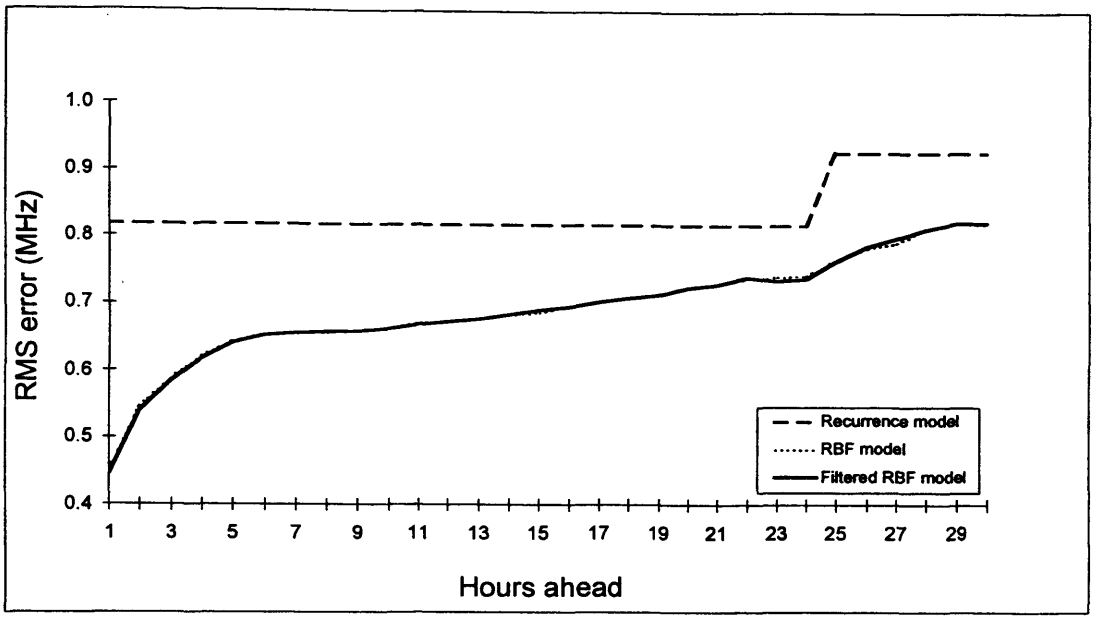


Figure 5.5 RMS prediction error for *foF2* hourly test values.

Optimising the noise filtering process led to the removal of 15 of the PCs that adversely affected the accuracy of the model. For a one-hour ahead prediction, the RBF model (with PCA noise filtering) gave an RMS error of 0.447 MHz, an increase in performance of 42.7% over the standard persistence model. In comparison, the RMS error for the 24-hour recurrence model was 0.818 MHz. The corresponding unfiltered and PCA filtered RBF model values relative to the 24-hour recurrence model were 44.7% and 45.4%, respectively.

For a one-hour ahead prediction, the persistence model gave slightly better performance than the 24-hour recurrence model. However, at prediction time scales greater than one hour, the strong diurnal variation of *foF2* generally led to a poor correspondence between the actual value of the point to be predicted and the persistence prediction. To provide a more effective reference model it was necessary to default to the 24-hour recurrence model. This model gave a constant RMS error for all prediction time scales up to 24 hours ahead, as shown in Figure 5.4. For predictions more than 24 steps into the future, the recurrence model utilises the value of the measurement from two cycles (48 steps) in the past.

Figure 5.5 shows the variation of model output error as the number of hourly steps ahead for the prediction increases, for both the PCA filtered and unfiltered RBF models, but on an expanded scale. The error rises steeply from one to six hours ahead, but the model still provides a significant improvement over the baseline recurrence model. For longer-term predictions, the model performance degrades more slowly and converges towards a constant margin of improvement over the reference model. The rapid degradation of the model between one and six hours ahead can probably be attributed to short-term features that are being modelled, but which are not predictable on medium to long time scales. It may also be a function of the optimisation process, as outlined previously.

For the hourly RBF model, noise reduction produced only marginal benefits. Indeed the improvement in performance is so small that it can only be discerned for one or two points on Figure 5.5. The principal component analysis technique is intended to remove those features of the time series that are present in the training data, but which are absent from the testing data. The absence of any significant enhancement of the performance of the model, using this noise reduction process, indicates that such features are not present in the training set. The input time series covers a three-year period at the bottom of a solar cycle. This time interval is likely to be approximately stationary with respect to the long-term variation of the solar cycle. Hence, the statistical moments of the time series may not alter to any significant degree over the time series as a whole, unlike the noonday *foF2* value time series. The dominant diurnal variation could also give rise to a lack of redundancy in the input vector, such that our noise reduction scheme can only offer very limited benefit. However, while the technique did not offer a significant performance advantage in terms of accuracy, it still reduced the dimensionality of the modelling problem, thereby reducing the processing overhead.

Figure 5.6 displays the RMS difference between the hourly *foF2* model and the persistence model, to assess the similarity of the RBF model to a simple climate based model. Currently, the modelling software only produces a direct

comparison with the persistence model and makes no direct comparison with the recurrence model. Little can be deduced, therefore, besides the dominant diurnal influence. However, the persistence model defaults to the 24-hour recurrence model in the case of the 24 step ahead prediction. In this instance, the RMS difference between the RBF and recurrence/persistence models was very much lower than the RMS prediction error of either model in isolation (Figure 5.4 and 5.5). This observation indicates that the 24 step ahead RBF prediction is little different to the 24-hour recurrence model or one day persistence model prediction.

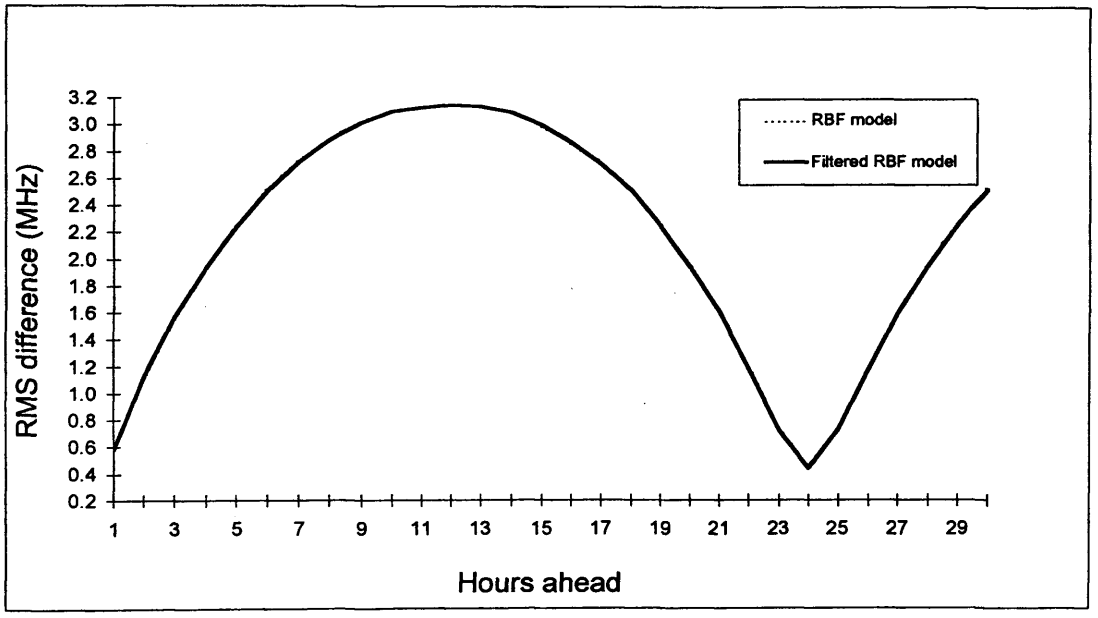


Figure 5.6 RMS difference between RBF and persistence models.

The apparent similarity between the 24 step ahead RBF prediction and the 24-hour recurrence/persistence model is an interesting result when taken in conjunction with the fact that the noonday RBF model performs significantly better (~60%) than the one day persistence model. Furthermore, the RMS prediction error to test set standard deviation ratio is approximately constant for the time series for all hours in the day. Therefore, there is no reason to believe that noon is inherently more predictable than any other time of the day. We conclude, therefore, that the hourly *foF2* model could be improved such that the 24-step

ahead RBF prediction error is comparable to the one-day ahead noonday *foF2* RBF prediction error.

5.9.2 *Strategies for Improving the Hourly Predictions*

Figure 5.7 compares the testing data set and the corresponding prediction output time series and shows that the RBF model provides high quality predictions one hour ahead. It has also successfully modelled the elevated activity on the final day of the plot, something a simple model such as recurrence would not be able to match. However, there are a number of potential improvements that could be made to the hourly RBF model. These improvements illustrate the complexity of predicting geophysical variables and particularly the need to keep in mind the background physical processes. To elucidate the potential improvements to the hourly *foF2* model, we first examine the possible causes of the discrepancy between the hourly and daily model errors for 24 hour / one day ahead predictions.

One obvious possible cause of the discrepancy is simply that the model predictions are optimised with respect to the one step ahead case. The dynamics of short-term hourly predictions may differ markedly from the longer-term hourly predictions, necessitating the adoption of optimisation for each individual hourly prediction time scale. However, attempts to re-optimise the hourly models for 12 and 24 step ahead predictions have only produced marginal improvements in predictive accuracy. Therefore, it seems likely that the cause of the anomaly lies elsewhere.

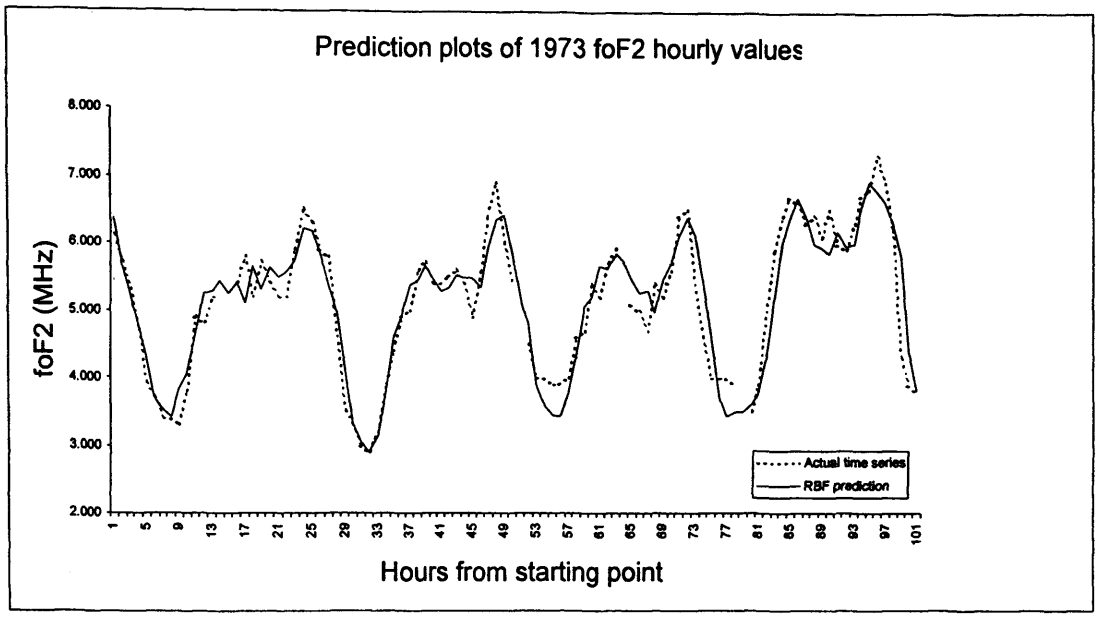


Figure 5.7 Comparison of predicted RBF and actual *foF2* hourly values.

Another possibility relates to the nature of the inputs used to produce the hourly time scale predictions, compared with those used to produce the daily time scale predictions. The latter utilises longer term information that is not characterised by the short input window length (<3 days) adopted by the hourly time scale predictions. Additional inputs, in the form of a longer term daily sampled time series, could, therefore, provide improvements. Inclusion of longer-term information is also likely to improve the ability of the model to track the non-stationarity of the time series over the whole solar cycle, should sufficient resources become available to process the entire available historical hourly record.

The dominant, presumably largely linear, diurnal influence present within the hourly *foF2* time series is likely to mask more subtle influences within the data. This occurs because the adaptability of the RBF NN is primarily focused on the dominant short-term variation. As such, little may be gained by optimising the model with respect to hourly prediction time scales, because the model is dominated by a linear periodic component. Nonlinear models are susceptible to this kind of interference and compound the problem by over-fitting a linear signal,

in comparison with a purely linear model. Removing the dominant diurnal variation in the pre-processing phase is likely to mitigate this problem.

Using first and second differences of the data, in conjunction with the complete hourly time series, would be one way of mitigating this effect. Another would be to use a hybrid linear / nonlinear model in such a way that the linear aspect of the modelling process is used to characterise the diurnal variation. The nonlinear aspect can then be used to predict the deviation from the purely diurnal variation. Any or all of these techniques should improve the predictive accuracy of the model.

The hourly model has been trained and tested using data from a portion of one solar cycle. Processing constraints prevented the use of a time series containing data from multiple solar cycles. Therefore, the generality of the model is likely to be compromised with respect to other periods within the solar cycle. In addition, the nature of the diurnal variation is such that the RBF model is modelling very different characteristics for a given hour of the day. Currently, the optimisation process finds a non-optimum solution, in the sense that it minimises the error over the day as a whole and not for any given hour. A separate model for each hour would enable the entire available time series for each hour, covering several solar cycles, to be used to train and test the RBF model. These models would then be able to generalise over the entire solar cycle. These individual models would also decompose the original composite hourly model such that the dominant diurnal variation no longer masks more subtle variations within the hourly time series. In essence, the existing prediction problem would be rendered down into a number of specific sub-problems that can be modelled more accurately, at the expense of an increased processing overhead.

5.10 *Monthly foF2 Value Predictions*

5.10.1 *Results*

Figure 5.8 describes the variation of RMS error for the monthly median RBF NN model, for one to thirty months ahead, in relation to the reference persistence model. In this instance, rather than use a prediction of the sunspot number to predict the median *foF2,* we have chosen to attempt this prediction based upon the time series of *foF2* in isolation.

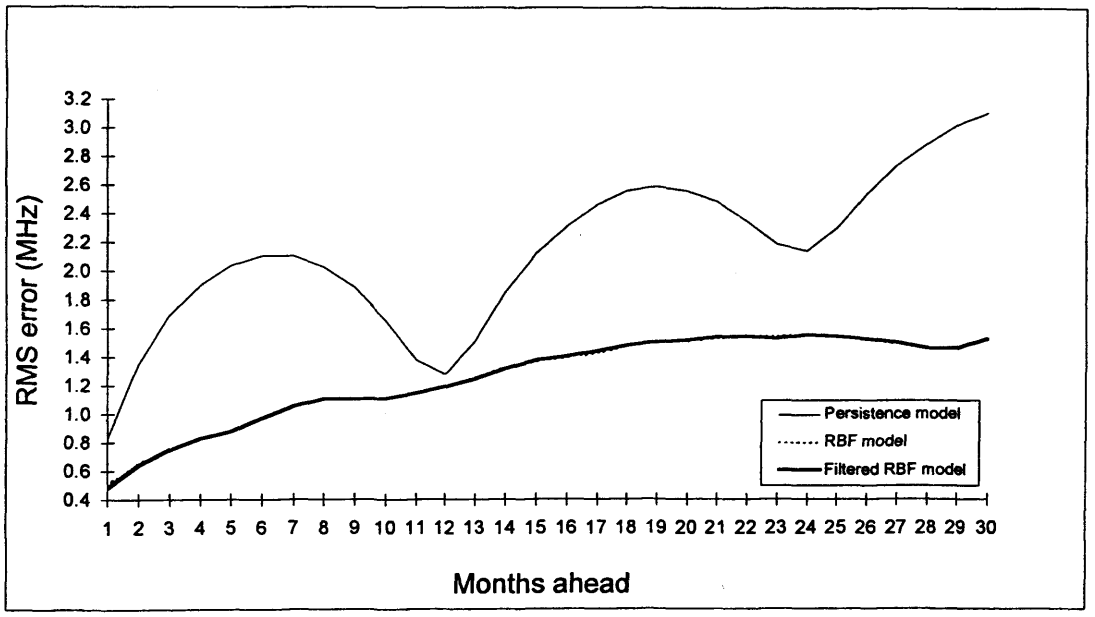


Figure 5.8 RMS prediction error for *foF2* monthly test values.

For a one-month ahead prediction, the monthly median RBF model (no noise filtering) RMS error was 0.499 MHz, in comparison with an RMS error of 0.829 MHz for the standard persistence model. This was an increase in performance of approximately 40% over the persistence model. Optimising the noise filtering process led to the removal of 22 of the PCs that adversely affected the accuracy of the model. For a one-month ahead prediction, the monthly median RBF model

(with PCA noise filtering) produced an RMS error of 0.488 MHz, a relative increase in performance of 41% over the standard persistence model.

As expected, the persistence model error curve displays clear recurrent variation on a time scale of twelve months. The monthly median RBF model is a significant improvement over the persistence model for almost all time scales. In addition, the RBF model does not exhibit any cyclic structure over the prediction time scales shown in Figure 5.8, indicating that it has characterised the annual variation successfully. For predictions up to eleven months ahead, the RBF model captured variations that were not characterised by simple median climatological models and consequently the RBF model errors were lower than the persistence model. However, the cyclic (one year) nature of the monthly median time series, coupled with the approximate stationarity with respect to solar cycle variation over the course of a single year, means that the persistence model and the RBF model achieve very similar results for the twelve month ahead predictions. This does not hold true for the 24-month ahead prediction, as non-stationary features of the time series become significant at these time scales. The RBF NN has better characterised this long-term non-stationarity, as evidenced by the much larger performance increase margin at 24 months, relative to the reference persistence model.

The RBF model performance degrades relatively smoothly from one to thirty months ahead and contains no prominent features. The rate of increase of prediction error decreases with time and this is most likely due to the convergence of the model dynamics towards the underlying climatological variation of the median values. For predictions greater than 18 months into the future, the RMS error is almost constant, indicating that linear features are likely to be dominant at these time scales.

The monthly median RBF model exhibited the greatest disparity between the predictive accuracy of the one step ahead and 30 step ahead predictions. The ratio

of the 30 step ahead prediction error to that of the one step ahead prediction was approximately 3 to 1, at least 50% higher than either of the other predictive models. A limited specific optimisation of the 30-month ahead prediction yielded some gain, but the improvement was of a lower order than that observed with the noonday value RBF model. Therefore, it is unlikely that this effect arises directly from the limitations of the optimisation method, as outlined earlier in chapter for other time scales. It is probable that the cause of this effect can be attributed to the format of the monthly median input vector. The input vector covers a period of approximately three months, which is very short in comparison with the longer-term prediction time scales

Figure 5.9 presents the RMS difference between the monthly median RBF model and the corresponding persistence model for each of the prediction time scales. The plot shows a strong annual variation in the RMS difference between the RBF model and the persistence model. This is due to the inability of the persistence model to successfully characterise the annual variation of the *foF2* parameter. Although the 12 month ahead RMS prediction errors are very similar in terms of RMS error (<0.1 MHz apart), the RMS difference between the two is very much higher (>0.5 MHz), indicating that they are achieving similar results in different ways. This suggests that further improvement of the 12-month ahead RBF prediction is possible, such that it incorporates those features of the reference model that are not currently characterised by the RBF model.

Noise reduction produces only marginal benefits for the monthly median *foF2* model. The improvement in performance is so small that it cannot be discerned in Figure 5.8. The absence of any significant enhancement of the performance of the model, using noise reduction, can probably be attributed to the smoothing inherent in any median time series, which removes the noise-like features of the data. Again, the technique does not offer a significant performance advantage in terms of accuracy, but still reduces the dimensionality of the modelling problem and thereby reduces the processing overhead.

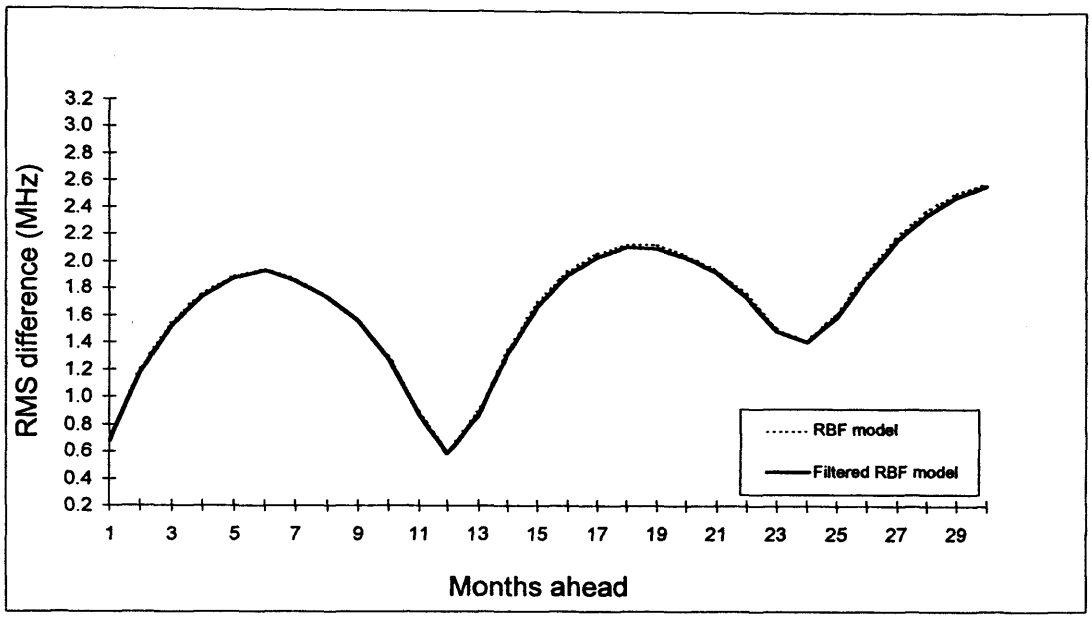


Figure 5.9 RMS difference between RBF and persistence models.

5.10.2 *Strategies for Improving the Monthly Median Predictions*

As already discussed, the ideal solution to the monthly median prediction problem is to create 24 separate monthly median models, one for each hourly median. However, the median time series contains insufficient points to create isolated hourly models. A composite model, which combines the training vectors for each separate hour together, in conjunction with an appropriate hourly time stamp for each vector, also proved non-viable. Insufficient training data was available to produce an RBF model with the required performance.

The actual input window used in this thesis, constructed from the median time series, covers a period slightly in excess of three months. This window length was determined to be optimal for the one-month ahead prediction. It utilises any or all of the values for a given month, as the data window slides across the entire time series. Consequently, the length of the input window is very short in comparison with the long-term prediction time scales. The longer-term variations in the monthly median parameter may not be characterised successfully by such a short

input vector. This is the most likely cause of the high prediction error ratios between the short and long-term monthly median predictions. It may therefore be necessary to re-optimise the input window length for each prediction time scale. However, this approach may incorporate redundant medium term information into the model. A better solution may be to supplement the existing predictions with selected longer-term monthly median inputs, corresponding to the hour to be predicted. This would provide the model with information concerning the long-term variation of the monthly median time series for a specific hour, improving the longer-term performance of the RBF model. Keeping the number of selected longer-term additional inputs small would prevent a marked decrease in the number of available training vectors.

It is also likely that this approach would benefit the twelve-month ahead prediction, such that it incorporates more of the features of the persistence model. It should result in a more definite improvement in performance over the reference model for that particular prediction time scale. Whilst producing similar results in terms of predictive accuracy, the RBF and persistence models are achieving this in different ways. A hybrid linear / nonlinear approach should combine the best parts of each model to improve the ability of the RBF to model linear recurrent structure in the *foF2* time series. It should also improve the long-term predictive capabilities of the model, as the features that are dominant at these time scales seem to possess a linear structure.

In comparison with the daily time scale predictions, the one month ahead prediction is significantly more accurate than the thirty day ahead model. However, this difference may be due to the smoothing inherent to any median time series. It is not possible, solely based on this evidence, to assume that the noonday RBF model would benefit from longer term inputs, though this is likely to be the case.

5.11 *General Remarks*

This chapter has presented a nested hierarchy of predictive models that operate upon different time scales. Prediction accuracy decreases as the sampling time scale increases from hourly to monthly values. Thus, the hourly one step ahead prediction is the most accurate, while the monthly median one step ahead prediction is the least accurate. However, the one step ahead prediction errors, for each of the models, all lie in the range 0.44 to 0.50 MHz. The range of variation of the error results over the different timescales is much less (~10%) than the variation between the standard deviations for each input time series (~40%). This observation runs contrary to the following preconceptions concerning the expected performance of the models. Due to the averaged nature of the monthly median time series, it was expected that the corresponding median model would exhibit the lowest error. In the same fashion, we expected the hourly model to exhibit the highest prediction error, due to the short-term dominant diurnal variation swamping other, more subtle, influences within the time series. In addition, the hourly time series should be more impulsive than the other time series.

An irreducible error in *foF2* predictions may be responsible for this discrepancy. The *foF2* parameter, which is used to derive all of the time series employed in this thesis, varies considerably between hourly measurements. This can be due to the effects of gravity waves (travelling ionospheric disturbances), sudden ionospheric disturbances and to uncertainties in the measurement techniques. The magnitude of this variation is thought to vary between 5% and 30% [*Professor T. R. Robinson, private communication*]. Therefore, aliasing effects introduce an ultimate prediction accuracy threshold. The RBF models will only be able to meet, but not exceed, this threshold. Assuming an average within hour variation of 15%, we can directly equate this figure to a normalised mean error (NME) of 0.15. It is then possible to compare this value to NME values for the RBF models. The hourly, noonday and monthly median models produced NME values of 0.202

MHz, 0.170 MHz and 0.209 MHz respectively, for one step ahead predictions. The values are very close to the proposed irreducible error figure and could explain why the model prediction accuracy statistics do not occur in the hierarchical order that the underlying physical preconceptions suggested. The fundamental hourly sampling information limit has been reached in each instance. The noonday RBF model NAE may be lower because the ionosphere is more stable at noon than at other times of the day. The noonday within hour variation would, therefore, be lower than the average value of the within hour variation, taken over the course of an entire day.

In comparison with the reference model, the noonday one-step RBF model provides a relative improvement of 60%, the largest performance increase margin of the three models. The monthly median value RBF model provides the smallest relative improvement (41%). The smoothed nature of the median time series means that a linear persistence model already provides a reasonable baseline model of the median time series, for the one step ahead prediction. The hourly value RBF model provides a relative improvement of 45%. Again, the reference recurrence model can provide a good approximation to the dominant linear diurnal variation. The larger relative improvement observed for the noonday model could be attributable to the fact that the noonday time series contains no smoothing or dominant short-term variation that can be characterised by a trivial linear reference model. In such an instance, however, the RBF model is still able to produce a model with a predictive accuracy in line with the other RBF models.

It is also interesting to note that the noonday model training errors were found to be much higher (~40%) than the corresponding test errors. It is more usual for the testing error to be higher than the training error, due to the fact the model is being applied to unseen data. This effect was also observed with the hourly model errors, but the magnitude of the effect was much smaller (~5%). No such effect was observed with the monthly median model errors.

The origins of this effect may lie in one of several areas. Firstly, statistical processes may be responsible for this effect. The large number of testing vectors that have been rejected during the modelling process may result in a lower level of statistical significance for the RMS prediction error results, particularly for the noonday predictions. The non-random distribution of missing points could also introduce a potential source of bias. This is supported by the fact that the magnitude of the effect seems to correlate with the number of input vectors that were rejected during the creation of each of the models. The development of an optimal nonlinear interpolation scheme will allow a more detailed examination of this issue. Secondly, physical processes could be responsible for this effect. Improvements in measurement and equipment accuracy over time may result in a decreased noise contribution in the later portions of the time series. Alternatively, long-term non-stationarity based upon the solar cycle may be responsible for more predictable behaviour during the portions of each time series used to test the hourly and noonday models.

Essentially, the statistical nature of a non-stationary time series alters on time scales that are much longer than the prediction time scales, but that are short enough to be present in the input time series as a whole. The hourly value time series covers the period from 1971 to 1973, at the solar activity minimum. This time series is quite short with respect to the solar cycle period and is approximately stationary, as the rate of change of solar activity is minimal at this point. Hence, the effects of non-stationarity upon the hourly value error results are quite small. The noonday value time series covers the period 1957 to 1990, encompassing three complete solar cycles. Thus, this data set is non-stationary, which could account for the large discrepancy between the observed training and test set errors. The monthly median value time series covers the period 1951 to 1995. In this instance, the parameter to be predicted is directly representative of the solar cycle variation. Therefore, the issue of non-stationarity does not arise in this instance, barring the presence of an undetermined non-stationary process that varies from solar cycle to solar cycle.

Preliminary studies, that partition the entire available noonday historical record into two equal parts and attempt to predict each set from the other, suggest that the observed discrepancy in the testing error is a property of the time series itself. This supports the hypothesis that non-stationary physical processes are responsible for this anomaly. Further investigation may well highlight the need to develop an adaptive modelling technique or hybrid linear / nonlinear model that deals more effectively with the problem of non-stationarity.

Further general enhancements of predictive accuracy may also be possible. Use of additional input time series, higher resolution data and individual optimisation for each number of prediction steps ahead should all yield improvements. Creating specialised predictive models for specific sub-problems should also increase prediction accuracy. An ensemble of such specialised models can then be used to cover the original broader problem.

Predictive accuracy may also respond to carefully selected presentation of an input time series. For each model, the most accurate result was recorded for the one step prediction. This may be due, in part, to the restricted optimisation noted earlier. However, it also possible that the sampling rate of the input data is coupled to the optimum prediction step number. Therefore, a prediction for a certain number of steps into the future might be best served by an input vector set sampled from the original time series at the same rate as the prediction step size.

5.12 *Conclusions*

This chapter has presented a simple technique that can cope with the problems of noise and non-contiguity that are typical of solar-terrestrial time series. The method of using nonlinear radial basis function (RBF) neural networks (NNs) to model the noonday variation of the critical frequency of the *F2* layer of the ionosphere, *foF2*, was implemented. The benefits of PCA noise reduction were assessed and found to offer a substantial improvement for the noonday RBF

model, in addition to providing worthwhile reductions in processing requirements for all three RBF models. This thesis has also highlighted the need for a novel nonlinear technique that interpolates missing data points in any specific time series. Optimising the models for each individual prediction time scale will bring further improvements in terms of accuracy.

The performance of the noonday *foF2* RBF model was compared with the reference persistence predictions. For a one-day ahead prediction, the unfiltered noonday RBF model produced an RMS error of 0.655 MHz, which is approximately 45% more accurate than the corresponding persistence prediction, which had an RMS error of 1.186 MHz. For a one-day ahead prediction, the noonday RBF model (with noise filtering) produced an RMS error of 0.484 MHz, a relative increase in performance of 59% over the corresponding persistence model. This accuracy was maintained for a ten to fifteen day period before the error started to rise significantly.

The performance of the hourly *foF2* RBF model was compared with the reference recurrence and persistence models. For a one-hour ahead prediction, the hourly RBF model (no noise filtering) produced an RMS error of 0.452 MHz, in comparison with an RMS error of 0.781 MHz for the standard persistence model. This represented a relative increase in performance of approximately 42.1% over the persistence model. For a one-hour ahead prediction, the hourly RBF model (with noise filtering) produced an RMS error of 0.447 MHz, a relative increase in performance of 42.7% over the standard persistence model. In comparison, the RMS error for the 24-hour recurrence model was 0.818 MHz. The corresponding unfiltered and noise filtered RBF model improvements over the recurrence model were 44.7% and 45.4%, respectively. The accuracy of the RBF model degraded rapidly up to 6 hours ahead, but then reached a stable level of improvement over the recurrence model.

The performance of the monthly median *foF2* RBF model was compared with the reference persistence model, for prediction time scales of one to thirty months ahead. For a one-month ahead prediction, the monthly median RBF model (no noise filtering) produced an RMS error of 0.499 MHz, in comparison with an RMS error of 0.829 MHz for the standard persistence model. This represents a relative increase in performance of approximately 40% over the persistence model. For a one month ahead prediction, the monthly median RBF model (with noise filtering) produced an RMS error of 0.488 MHz; a relative increase in performance of 41% over the standard persistence model. In this case, the model accuracy degrades smoothly from 1 to 30 months ahead.

In conclusion, the RBF NN models presented in this chapter are a significant improvement over the standard persistence and recurrence models, which act as references for the assessment of any proposed prediction scheme. Unlike many existing studies, that utilise input parameters from supplementary time series that are concomitant with the future point to be predicted, the techniques presented in this chapter could be used to form the basis of a real time ionospheric forecasting service. This chapter has also highlighted the need for a novel nonlinear technique that interpolates missing data points in any specific time series. This should be done in such a way that the effects of the interpolation are minimised for any specific model that uses that time series as an input. This issue is addressed in the next chapter.

Chapter 6: Development of a Nonlinear Interpolation Technique

6.1 *Introduction*

This chapter proposes a novel technique for the prediction of solar-terrestrial data sets that contain a significant proportion of missing data points [*Francis et al.*, 2001]. A nonlinear interpolation technique is employed to assign values to gaps in a time series. Each missing point is interpolated such that the error introduced into any specific predictive function is minimised. This interpolative technique has general application in any instance where the effects of interpolation upon a given analytical or predictive process need to be minimised or a complete time series needs to be constructed from incomplete data.

6.2 *Incomplete Data Sets*

As already discussed, incomplete data sets are a recurrent problem associated with geophysical parameters. Instrument failure or prevailing conditions often give rise to missing points in the time series of measurements. Discontinuities pose a significant obstacle for prospective nonlinear prediction schemes, which generally require continuous time series data. Relaxing this requirement necessarily leads to adapting the predictive techniques to cope with this particular problem. Three simple approaches are commonly used to overcome the problems associated with data gaps. These have already been discussed but it is useful to reiterate the issues involved.

The simplest solution is to set all missing data points to a value of zero, after the time series has been normalised to zero mean and unit variance. This minimises the root mean square error for a constant-value interpolation method, averaged over the entire set of missing values contained in the de-meaned time series. However, this approach is likely to be unsatisfactory for more than a small proportion of missing data points, depending on the data characteristics.

For time series that contain a larger proportion of missing points, a more sophisticated linear interpolation scheme could be used. However, geophysical systems such as the ionosphere are thought to exhibit significant levels of nonlinear behaviour, as previously discussed. Use of a linear interpolation scheme on such data could alter the apparent dynamics of the underlying nonlinear dynamical system, thereby degrading the performance of the modelling process. The magnitude of this distortion will increase as the proportion of missing data points increases. Linear interpolation techniques are, therefore, highly undesirable when using nonlinear models or time series.

A third approach is to discard those training vectors that contain missing data. This avoids the issue of interpolation altogether and prevents any possibility of providing the NN with misleading interpolated data. As the proportion of missing points in the time series increases, the proportion of rejected vectors also increases, in some fashion dependent upon the distribution of the missing data points and the method of constructing the data vectors. This would be the preferred technique in the absence of a suitably robust nonlinear method of dealing with missing data points. However, rejecting a proportion of the training vectors may compromise the model training process. This is more likely to occur with a lower ratio between the length of the available time series and the length of each training vector. Additionally, a significant bias will be introduced if the missing points are not randomly distributed through the time series.

In the absence of a standard and mathematically rigorous alternative, the rejection of input/output pairs containing missing data points was adopted in the previous chapter. The proportion of missing points observed over the hourly *foF2* time series in the previous chapter (6.6%) gave rise to a very high vector rejection rate (60%). It would be highly desirable, therefore, to make use of the information contained within these discarded vectors to improve the training of the RBF NN. This must be done in such a fashion that it does not alter the apparent dynamics captured by the modelling techniques or introduce excessive artificial bias. The

satisfaction of these requirements underlies the novel nonlinear interpolation technique outlined in this chapter.

Consequently, it has been necessary to devise a more complex nonlinear interpolation scheme to overcome the considerable drawbacks of the above techniques. This chapter presents a technique that minimises the effects of interpolation upon any given modelling process. The optimal value is interpolated for each missing training point, such that the RBF gives a minimal error over the training set as a whole. Incomplete training vectors can then be interpolated and included into subsequent modelling attempts, improving the predictive performance of the RBF model. This approach allows the extraction of the maximum amount of information from initially incomplete training vectors, whilst introducing a minimum amount of artificial bias.

6.3 *The Problem of Missing Data Points*

Given a time series $\{x_n\}_{n=1}^{N}$ consisting of measurements of a physical system in which some of the data points $\{x_{n_i}\}_{i=1}$ correspond to failed measurements, it is desirable to construct a k step ahead predictive model of optimal order d. This model aims to utilise the general input vector $(x_{n-d+1},\ldots, x_n)$ to predict x_{n+k}, where $d \leq n \leq (N-k)$.

RBF NN techniques provide one such means of creating a predictive model. These techniques require a matrix of input vectors derived from the input time series. This matrix of input vectors is constructed by sliding a one-dimensional d-length window along the input training time series, one point at a time. Each vector is then matched with the corresponding k step ahead point to be predicted (desired output) to form a set of augmented input/output pairs {X_n input, Y_n output} of the form $\{(x_{n-d+1},\ldots, x_n), x_{n+k}\}$, for $n = d$ .. $(N-k)$. The RBF modelling technique uses a constrained set of nonlinear functions f (equation 3.6.2) to

provide an approximation to the mapping between the input/output training elements. This mapping is the required predictive function. The model parameters (including d) are optimised to give the best functional solution by minimising the error

$$\mathbf{E} = \sum_{n=d}^{N-k} (f(\mathrm{X}_n) - \mathrm{Y}_n)^2 \qquad (6.1).$$

The presence of missing data points necessitates an adaptation of this scheme to cope with input/output pairs that contain one or more missing elements. Trivially, any input/output pair that contains a missing element has the corresponding entry deleted from the input vector set to form a subset, **M**, of input/output pairs that contain no missing elements and can be used in equation (6.1). This is the approached used in Chapter 5 and described in Flowchart 2 of Annex A.

This process of creating the final uninterpolated vector input set is represented pictorially in Figure 6.1. This figure depicts the construction of the input/output pair matrix (augmented array) from an input time series using a moving window. Missing points are indicated by '*' characters in the input time series and augmented array. Input/output vectors that contain a missing point are crossed out in the figure to indicate the fact that they are removed from subsequent stages of the modelling process.

Equation (6.1) is then summed over the input/output test pairs that contain no missing elements. It is worth noting that each missing element will be involved in the deletion of d+1 rows of the array, as the moving input window moves along the time series one point at a time. The proportion of rejected vectors will therefore initially increase geometrically as the proportion of missing points increases. As the proportion of missing points increases still further, the rate of increase of vector rejection will decrease, as the probability of multiple missing points occurring in the same input/output vector becomes significant. Finally, for

a given proportion of missing points, increasing the length of the sliding window will also increase the vector rejection rate. The distribution of missing points will also have an important effect upon the proportion of rejected vectors. In addition, non-randomly distributed missing points will introduce bias into the model. This will occur because the rejected vectors will be correlated with certain features present in the data and these features will be poorly represented in the training set.

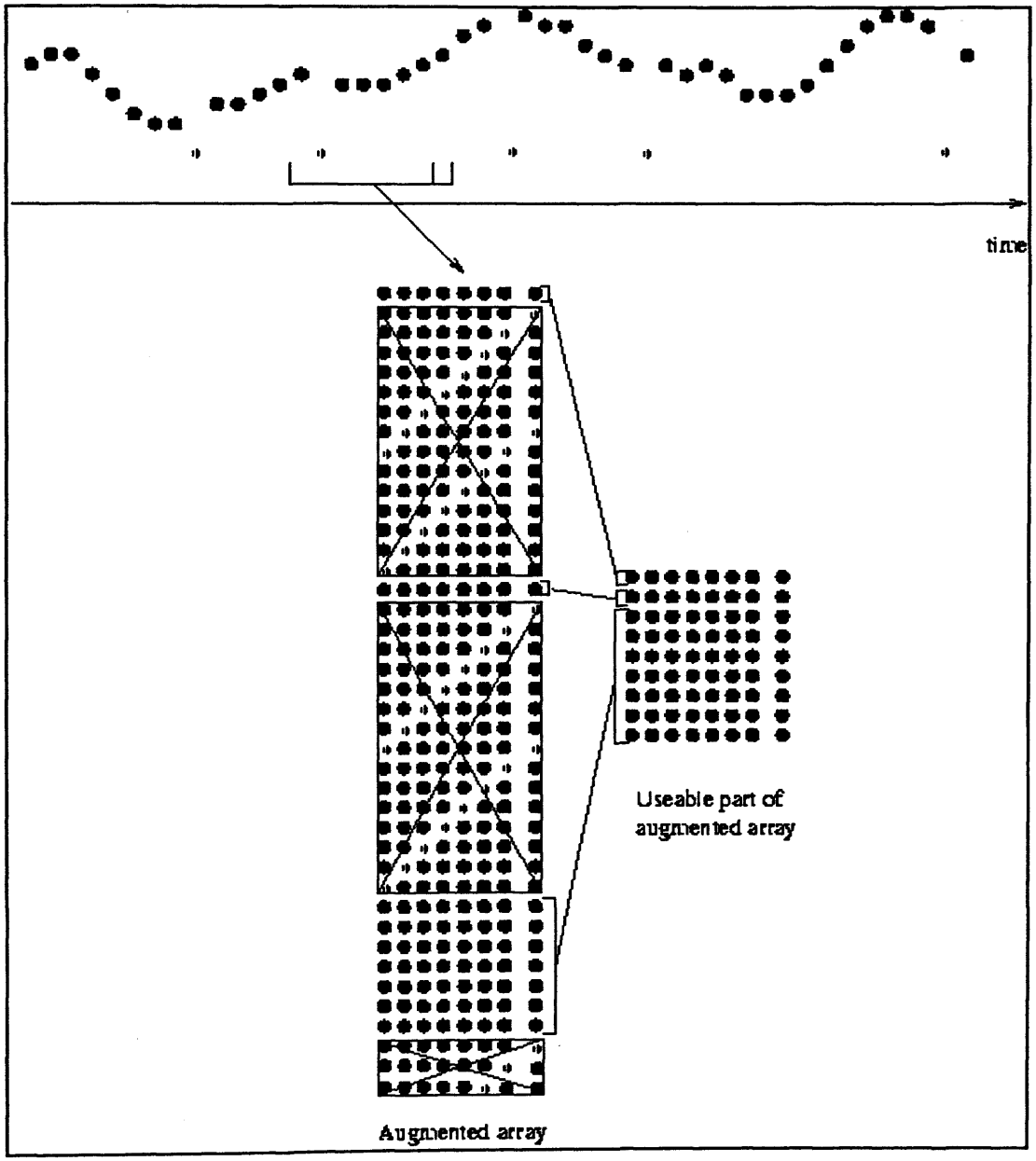


Figure 6.1 Pictorial representation of data vector construction

6.4 *Interpolation Theory*

Using the basic prediction framework outlined in the previous section as a starting point, several possibilities emerge for a prospective nonlinear interpolation scheme. In each instance, the interpolation scheme is only used to fill the gaps in the input time series. Gaps in the output time series are inviolate. Interpolating these points would compromise the veracity of the statistics used to assess the performance of the interpolation scheme. Fortunately, given an interpolated input time series, each missing output point only leads to the rejection of a single input/output pair. Thus, the residual number of rejected vectors after interpolation of the input time series is now only directly, rather than geometrically, related to the proportion of missing points.

The most obvious possibility for a nonlinear interpolation technique is based upon the direct application of the predictive function derived from the complete input/output pairs to the prediction of missing data points. Essentially, this method uses successive application of the one step ahead RBF NN model, as the input window moves from the start to the end of the time series, to provide a prediction for the missing values. It utilises complete vectors, or vectors that contain missing points that have been previously interpolated by the technique, to predict one step ahead to interpolate subsequent missing points. Subsequent models can then be derived using the interpolated input data set. The one step ahead model is the logical choice because it gives the best results in terms of predictive accuracy and should therefore provide the most accurate interpolation. However, once the RBF model moves far outside of its normal operational regime (i.e. complete input vectors only), this scheme can become highly unstable and either converge towards zero or infinity. Even if such asymptotic behaviour does not manifest itself, the interpolation process can result in distortion of the apparent dynamic of the system that is being modelled. This is a particular problem if the interpolation process is iterated in the hope of achieving better predictive results through successive interpolations of the input data.

It is possible, however, to control the stability of the nonlinear interpolation process by further constraining the gap filling technique. The previous method only considered the d points of the input window that determine each point to be interpolated, via the predictive model, f (equation 3.6.2). However, the interpolated point will also form a part of each of the d consecutive input vectors that determine the d subsequent points in the output time series.

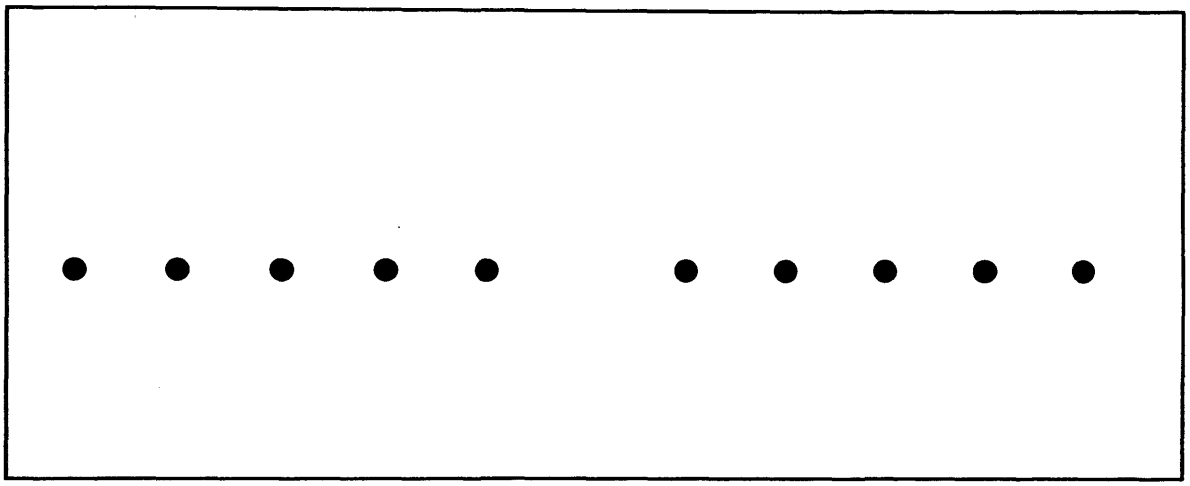


Figure 6.2 Pictorial representation of blanket of influence

The second possibility for a nonlinear interpolation technique, which was the one finally developed, considers this factor by widening the area of influence of each missing point. Thus, each missing point can be said to possess a blanket of influence of $2d+1$ points, as shown pictorially in Figure 6.2. Essentially, this region includes all those points that determine or are determined by the missing point. For a single blanket of influence, the effects of any interpolation do not need to be considered outside of this set of points.

Adjacent blankets of influence may overlap good data points without introducing interdependence, as these points are fixed values with respect to both zones of influence. In such a case, there is no interpolative interdependence between the missing points at the centres of each of the overlapping blankets of influence. However, multiple missing points within a time series may give rise to

overlapping and interdependent blankets of influence within the time series, such that two or more missing points are within d points of each other.

An example of overlapping blankets of influence is shown in Figure 6.3. In the case depicted, two overlapping blankets of influence encompass the missing point denoted by '#'. The interpolation problem then becomes multi-dimensional as the impact of this missing point will effect both blankets of influence and hence the interpolation of the two adjacent missing points. Each blanket may be inter-linked with several others, forming long chains of overlapping blankets of influence, along which the influence of a single missing point can propagate through mutual dependence. These chains can span a significant proportion of the time series if the proportion of missing points is large enough and the associated interpolation problem can become very high dimensional.

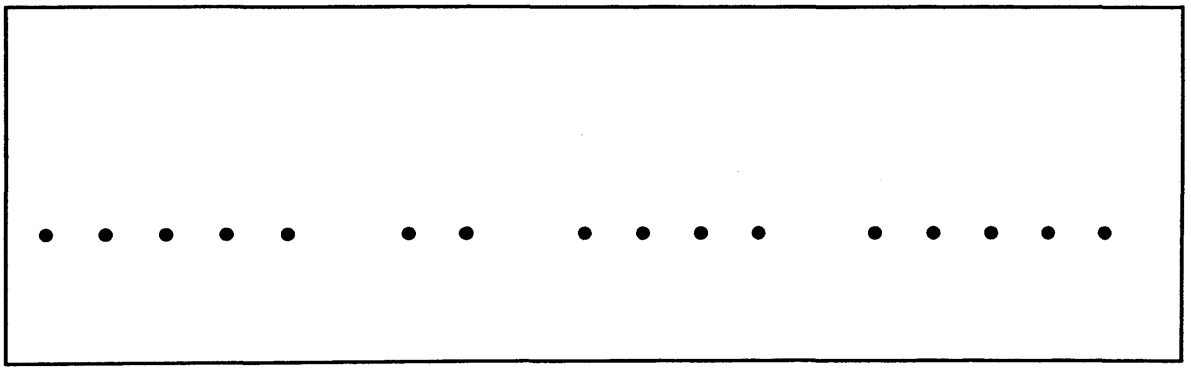


Figure 6.3 Pictorial representation of overlapping blankets of influence

Initially, consider the case where a solitary central missing point is surrounded by n good points on either side (Figure 6.2). The influence of the missing point is confined to this portion of the time series. It is desirable to interpolate the point in such a way that the error on the predictive function f (equation 3.6.2) derived previously from the complete vector only input/output pairs is minimised over the set of n input vectors that are derived from this set of $2n+1$ points. In a sense, the stability and accuracy of the interpolation process is preserved in both forward and reverse directions along the time series. Given a sequence of input points $\{x_s$-

$_{n},\ldots, x_{s+n}\}$, where x_s ($d < s < N\text{-}k$) is the only missing point and {X,Y} are the set of corresponding input / output pairs, this is achieved by minimising, with respect to x_s, the error

$$\mathbf{E} = \sum_{i=s}^{s+n} (f(\mathrm{X}_i) - \mathrm{Y}_i)^2 \qquad (6.2).$$

Local minima occur when $\frac{\partial \mathbf{E}}{\partial x_s} = 0$ and the second derivative is positive.

This minimisation technique can be extended to encompass overlapping blankets of influence, where two or more missing points are within n points of each other. The simple minimisation scheme outlined above is extended to encompass a region of the data, containing t missing points, which is bounded by n good points at either end of the set. Given a sequence of points $\{x_{s1-n},\ldots, x_{st+n}\}$ in which the subset of points $\{x_{s1},\ldots, x_{st}\}$, where ($d < s_1 \ldots s_t < N\text{-}k$), are the only missing points, it is necessary to minimise, with respect to the entire subset of missing points, the error

$$\mathbf{E} = \sum_{i=s_1}^{s_t+n} (f(\mathrm{X}_i) - \mathrm{Y}_i)^2 \qquad (6.3).$$

Without loss of generality, it is possible to assume that $s_1 + n \le s_2,\ldots, s_{t-1} + n \le s_t$, since if this were not the case then it would be possible to sub-divide the set into independent partitions. For a multi-dimensional minimisation problem, local minima occur when $\frac{\partial \mathbf{E}}{\partial x_{s1}} = \frac{\partial \mathbf{E}}{\partial x_{s2}} = \ldots = \frac{\partial \mathbf{E}}{\partial x_{st}} = 0$ and the matrix of second derivatives, the Hessian, is positive definite (positive determinant and positive valued at the turning point).

Many techniques exist for calculating local minima, usually based upon a variant of a gradient descent algorithm [*Press et al.*, 1992]. This thesis has employed conjugate gradient descent, which often converges more quickly than other gradient descent methods. It is guaranteed to find a minimum of a quadratic function within p steps (for a p-dimensional minimisation problem). Techniques such as steepest gradient descent require many multiples of p steps to approach this minimum, especially if the minimisation surface is strongly elliptical. The conjugate gradient descent technique is applied in this instance to solve the multi-dimensional minimisation problem outlined above, by combining the baseline RBF function derived in the previous chapter (equation 3.6.2) with equation (6.3).

6.5 *High Level Interpolation Methodology*

An interpolated model has been implemented for the prediction of hourly *foF2* values from 1971 to 1973 for Slough station, UK. The time series as a whole contains 6.6% missing data points, which gives rise to a vector rejection rate of approximately 60%, if incomplete input vectors are discarded. A missing data point might correspond to an equipment failure, or ionospheric conditions may prevent a clear measurement of the properties of the *F2* layer. Another class of data dropouts includes those values that fall outside the likely range of variation for the *foF2* parameter (0.5 to 17.5 MHz) and are probably due to clerical errors or poor measurement conditions. In general, missing points are not randomly distributed through each time series. This introduces a potential source of sampling bias into the uninterpolated RBF model.

In this study, the first 23,000 points of the hourly time series have been used to train and optimise the model, while the remaining 3304 points were used to test the ability of the model to generalise when applied to unseen data. A one-dimensional sliding window was passed along the data, one point at a time, to construct the matrices of input vectors required to train and test the RBF model. In each instance, training and testing data were processed entirely separately; i.e. the

sliding window used to create the input vectors was not allowed to overlap both of the data segments at the same time. The effectiveness of the interpolation process has been assessed relative to reference persistence and recurrence interpolated models.

Flowchart 2 in Annex A details the basic, non-interpolated RBF NN predictive modelling process adopted in the previous chapter and in *Francis et al.* [2000]. This procedure produces an uninterpolated RBF predictive function that acts as a bootstrap for the subsequent interpolation of missing data points, which is presented in this chapter. The optimal parameters derived in the previous chapter were also adopted for the purposes of this thesis, due to processing limitations.

By combining the baseline RBF function derived in the previous chapter (equation 3.6.2) with equation (6.3), the conjugate gradient descent method can be applied to find the optimal interpolated values for the missing points in the input time series.

Flowchart 3 in Annex A details the high-level process of applying the error minimisation algorithm to interpolate the gaps in the input time series. The basic interpolative scheme, outlined above, is initially used to fill in the gaps in the input time series using only the complete input vectors (~40%). Equation (6.1) is then summed over a set of vectors $\mathbf{M}^*$, which contains the interpolated input test vectors that do not correspond to missing output points, in addition to the set $\mathbf{M}$ of complete input vectors, to determine the new predictive model that includes the interpolated input data.

In this instance, the vector rejection rate is only dependent upon the output time series, so the model can utilise a much greater proportion of the total number of possible input vectors (~93%). In turn, this means that the second RBF model can subsequently be used to create a much better interpolation of the original non-contiguous input time series than the first RBF model - which only utilised

complete input vectors. This process can be iterated until the model error on the test set is minimised and the best interpolation (and the best predictive model) of the input time series is obtained. In this application, two iterations of the interpolation process were sufficient to achieve the optimal results. The optimally interpolated model was then used to predict one to thirty hourly steps ahead and the results were compared with the reference models described in the next section, using the RMS error.

6.6 *Reference Interpolation Models*

Three simple reference interpolation schemes were adopted for the purpose of comparison with the nonlinear gap-filling method. The first technique, the zero interpolation scheme, set all missing points in the de-meaned and normalised time series to a value of zero. The second used persistence to fill in missing data points. The third and final technique used the 24-hour recurrent structure of the hourly *foF2* time series to replace missing points. In addition, the complete vector only RBF model was included for comparison. However, the complete vector test set is much smaller than the interpolated vector test set, due to incomplete vector rejection. This means that the results are not directly comparable with the interpolated techniques, which use a significantly larger superset of the complete vector test set.

6.7 *Results and Discussion*

For each of the prediction schemes, Figure 6.4 shows a graph of root mean square (RMS) error, on the interpolated test vector set, plotted against number of hours. Using any of the proposed interpolation schemes, the vector rejection rate falls from 60% to 6.6%. Residual rejection occurs because training vectors are still rejected if there is a gap at the corresponding point in the required output time series.

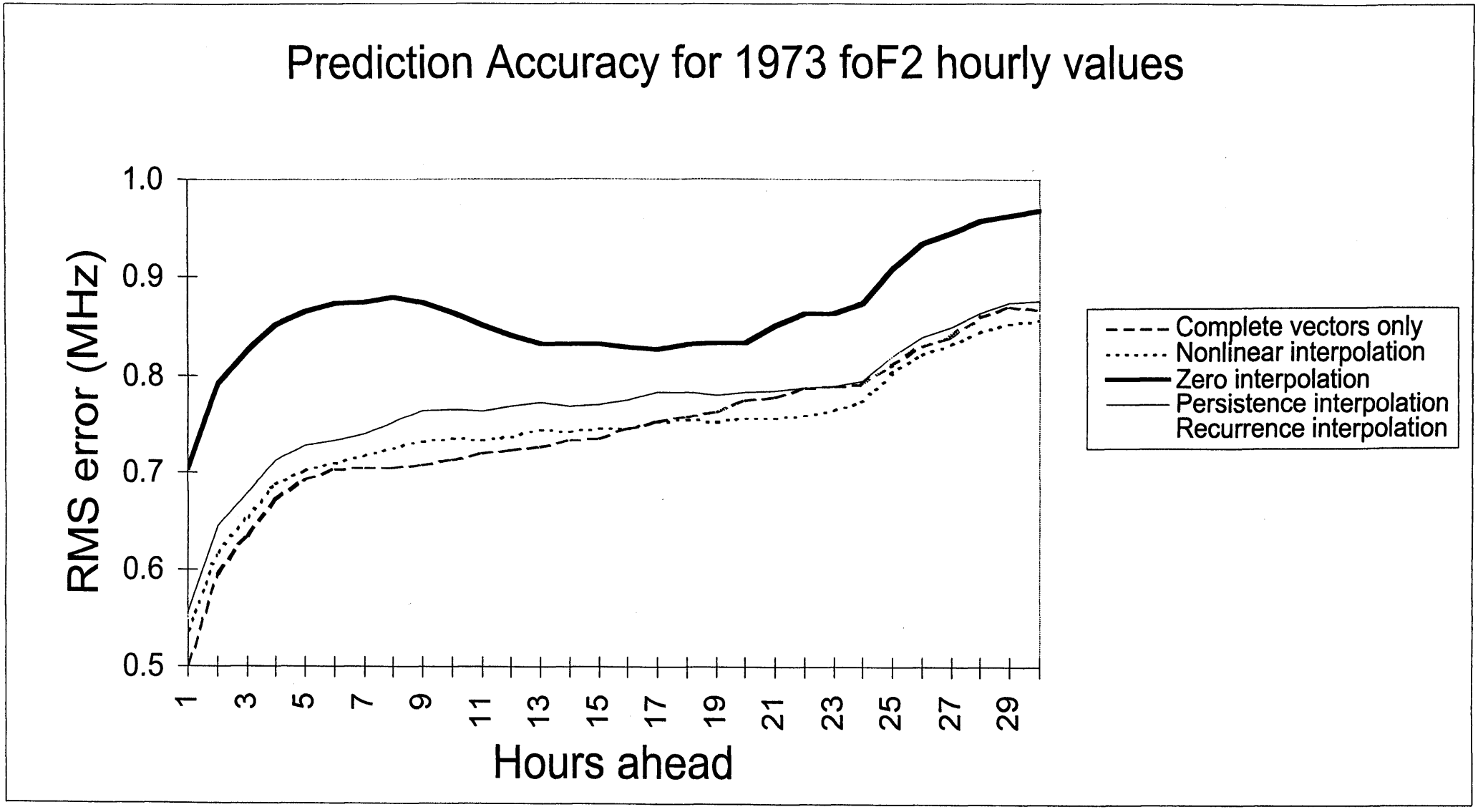


Figure 6.4 Relative performance of interpolation techniques

The graph clearly shows that the nonlinear interpolation technique consistently produces the best results in terms of relative predictive accuracy over the test set. The complete vector only (no interpolation) model is also included for completeness and outperforms the interpolated model up to 15 hours ahead. However, direct comparisons cannot be made with the interpolated models for reasons stated in the previous section. The nonlinear interpolation technique shows a 2.3% improvement in test set error over the recurrence interpolated model (the level of improvement rises to 3.8% if the incomplete test vector set is considered in isolation). In turn, the recurrence interpolated model shows a similar improvement relative to the persistence interpolated model. The zero interpolation model falls significantly behind these reference models, indicating the necessity of adopting at least some approach to deal with the problem of missing data points in geophysical time series.

The benefits of this technique are more apparent when input vectors containing missing data points are also used to test the model. In this case, the uninterpolated predictive model produces an RMS error of 0.653 MHz, 44.5% higher than the corresponding error for the model upon the complete vector only test set. Utilising the interpolation algorithm lowers the error by 26% to 0.486 MHz, almost equivalent to the performance of the uninterpolated model upon input vectors that contains no gaps (RMS error ~ 0.452 MHz). In fact, the performance of the interpolated model on incomplete vectors lagged just 7.1% behind the performance of the same model upon complete vectors. This clearly shows the benefits that can be achieved using this technique.

The complete vector only model curve (Figure 6.4) cannot be directly compared with the interpolated model curve, as both models have been applied to different test sets. Though it is not shown in the figure, it is interesting to note that the interpolated model produced lower RMS errors, relative to the complete vector only model, when both were applied to the complete vector only test set (M) instead of the interpolated vector test set (M^*). This shows the positive benefits

that can be accrued, in terms of improved network training, through interpolation of the input time series. This point is further strengthened by the fact, on the same test set, the zero interpolation model performed less accurately than the complete vector only model. From this evidence, it would appear that rejecting incomplete data vectors may be preferable to including them in a model without further processing, as zero interpolated vectors seem to provide the network with deleterious training material.

The improvement in overall predictive accuracy observed in this thesis might be considered relatively modest. This is most likely a side effect of the increased dimensionality of the RBF model optimisation problem, given the iterative nature of the interpolation methodology. The size of the optimisation space to be searched, in conjunction with associated limitations in terms of available processing resources, necessarily involved compromises in terms of the level of optimisation that could practically be achieved. The overall effect of these compromises is likely to have had a very significant effect on the performance of the optimisation process. This issue will be discussed at greater length in the following section.

In addition to the improvement in predictive accuracy provided by the interpolation process, the nonlinear predictive model is also enhanced in terms of its generality. The latter effect is arguably the more important of the two. Missing data points can give rise to correlated data vector rejection, as there may be features present in the time series as a whole that are not characterised in the complete vector only training set. While an improvement in predictive accuracy is a worthy goal, the model is of little use if it can only be applied to a limited set of circumstances with the indicated level of accuracy. Furthermore, it may not be straightforward to deduce in advance whether or not the complete vector model is going to be applicable for any given prediction. Such a limitation would preclude the use of an uninterpolated model for real-time predictions, where no a-priori knowledge concerning applicability is available.

Increasing the amount of training data available to an RBF NN can be vital if training material is a limiting factor on the predictive performance of the NN. If the number of available training vectors is adequate, then little may be gained in terms of predictive accuracy. In such instances, however, the interpolation technique outlined in this thesis can still be used to significantly reduce the effects of sampling bias and correlation of missing points with features present in the input time series.

For time series where simple periodic climatological models cannot be used to interpolate missing data, this technique will find the best interpolative solution for any given problem. In such a case, the resultant improvement in performance may well be greater than for the prediction of the hourly *foF2* time series. The dominant diurnal variation of this ionospheric parameter is well modelled by simple recurrence and this simple model can be used to provide a reasonably effective interpolation scheme.

6.8 *Restrictions on Use of Methodology*

This method is currently subject to several restrictions. First, the interpolation can only be carried out for a self-prediction problem. This means that only the time series history of the data to be predicted can be used as an input to the predictive model and the input time series itself must be one-dimensional. Secondly, the RBF function used for the interpolation must be for a single step ahead prediction.

In theory, any of these limitations could be circumvented through further development of the underlying interpolation technique. However, it may not be advisable to lift any or all of these restrictions. With regard to the first limitation, any attempt to fill gaps in the input time series using a model that predicts a different output time series is likely to be ill advised. Supplementing the self-prediction input time series with additional time series to improve the predictive

accuracy of the model, and hence the accuracy of the interpolation, is a more considered approach. For the prediction of the variation of hourly *foF2*, measures of solar activity or the interplanetary magnetic field could be appropriate. Any such supplementary time series will need to contiguous or contain very few missing points. Otherwise, the benefits of interpolating the original input time series are likely to be severely compromised. This is due to the fact that gaps in the supplementary time series will need to be considered during the process of rejecting incomplete input/output pairs and will contribute to increase the overall input vector rejection rate.

As a simple alternative, supplementary inputs could be interpolated separately before being used to enhance the original self-prediction problem. It might also be possible to evolve a co-minimisation technique, where each additional input time series added to the model necessarily requires a corresponding output time series. This would give rise to an output prediction for each input time series, some of which could be discarded if necessary. This multiple input / output RBF model could be used in conjunction with a multi-dimensional and interdependent analogue of equation (6.3). This would simultaneously find the optimal values for missing data points in all of the input time series in a fashion very similar to the interpolation of each time series in isolation. However, the minimisation would be further constrained by cross prediction between the time series and should achieve better results. This would partially be due to the inclusion of spatial information that can be utilised in the interpolation process. This could reduce the effects of under-sampling, as some constraints would not have a temporal aspect. It may also make the process less susceptible to choosing local optima that are based upon simple temporal properties of the data, such as recurrence.

The second limitation of the interpolation technique could be removed with only minor modifications to the underlying algorithm. The one step ahead predictive RBF model is usually the most accurate. As a direct consequence, it will provide the best average interpolation for any given missing point in the time series,

given a complete input vector. For consecutive missing points, however, it may be preferable to extend the existing scheme. Currently, the first gap of the consecutive series of missing points is filled by the one step ahead model using a complete input vector. The second gap then also has to be filled using the one step ahead model. This is achieved by moving the input window along one point to encompass the previously interpolated point, so that the second gap to be filled is the required one point ahead. It may arguably make more sense to continue to fill the first gap with the one step ahead model, but to fill the second gap with an optimal two step ahead model. This will allow the use of a complete input window rather than an interpolated input window. This would avoid using an interpolated input vector to interpolate the second missing point, improving the stability and accuracy of the interpolation process still further. By looking at the distribution of lengths of missing sections of the input time series, it should be possible to produce a scheme that trades processing resources against the number of different step ahead models available to be used by the interpolation algorithm. These models can be placed in a way that gives efficient coverage of the distribution of lengths of contiguous series of missing data points. This would necessarily involve an additional optimisation overhead. All of the models, not just the one step ahead model, would need to be fully optimised in terms of the model parameters to justify this approach. Otherwise, much of the performance advantage of using separate models would be lost.

Another possible improvement to the gap filling technique employs the idea of partial confidences for data points. Currently, valid and missing data points are assigned a confidence value of one or zero, respectively. Additional information is sometimes available for measurements of geophysical parameters. For *foF2*, each measurement has an accompanying URSI code that provides an indication of the reliability of the value as well as possible upper or lower bounds. These codes could be used to derive fractional confidence values for those points that are deemed less reliable. An input vector containing points with associated confidences of less than one could be assigned a lesser weighting in the training

phase. This would be achieved by weighting each input vector with the product of all of the individual confidence values for each input vector element. The resultant weighting for each complete input vector would be in the range (0,1].

The above idea could be extended to the utilisation of fractional confidences in the interpolation process. The filling procedure could be modified to take into account the confidences of surrounding points when interpolating a missing point, to improve the reliability of the gap filling process. Partial confidences could also be assigned to interpolated points in the *foF2* input time series. Interpolated points are currently assigned a default confidence value of one, in the absence of a rigorous alternative. The existing technique could be adapted to produce an appropriate partial confidence value for the interpolated point that minimises the error on the model output. The partial confidence of a filled point would be dependent on the accuracy of the predictive model employed, as well as the number of other missing points that are interdependent with it. This gradation of data quality, and attendant weighting of interpolated input vectors, would improve the general stability and subsequent iteration of the interpolation process, as described above.

6.9 *Optimisation Issues*

Model optimisation is a very important aspect of the work presented in this thesis. It influences the performance of the model in two respects. The first and most straightforward aspect of this influence is the optimisation of the interpolation of missing points detailed in section 6.4. Currently, the technique merely finds local optima for equation (6.3) from an initial starting condition. This chapter employed both a persistence initial condition and a zero vector initial condition, either of which led to a 2.3% improvement on persistence. There is no guarantee that such a solution is globally optimum. It may well be possible to find a lower cost solution, in terms of test set RMS error. The removal of persistence information, using linear feature extraction or by taking first

differences of the data would be desirable. This would remove the dominant persistence based minimum, and enable the search technique to concentrate on more globally optimum areas within the solution space.

Heuristic techniques, that contain a measure of backtracking, could also be used to conduct a more thorough and efficient global search of the solution space, whilst keeping the number of initial conditions used to an acceptable level. Such a modification would require significant extension of the existing interpolation framework, but should not dramatically increase the processing resources needed to find a solution.

The second aspect of the optimisation issue relates to the free parameters used in the RBF modelling process. In the previous chapter, the only significant compromise made with respect to this optimisation was that the optimal model parameters were only derived for the one step ahead model, due to processing constraints. This chapter then used the optimal one step ahead model to interpolate missing points. However, the iterative interpolation of the input time series necessitates a re-evaluation of the idea of the optimality of the one step ahead model. The optimal parameters for the one step ahead RBF model in the previous thesis were determined using the complete vector training and test data. These parameters are not necessarily valid for the interpolated RBF model, which uses the set of interpolated input vectors.

The input window length is one of the most important free parameters of the RBF model. For a time series that contains no gaps, a balance must be struck between not including sufficient information (input window too short) and including redundant information (input window too long) in each vector. Essentially, the correct input window length will just resolve the topology of the underlying dynamical system that belies the parameter to be modelled. This ideal situation is complicated by the presence of missing points and the limited number of points that are available for geophysical time series in the historical record. The number

of training vectors obtained from a specific number of consecutive points decreases as the length of the sliding input vector window increases. If the number of available time series points is limited then the optimisation of the input window length may be compromised, as the window length increases past a certain point. This can be attributed to inadequate training of the RBF model due to a lack of data, rather than increasing redundancy in the input.

This problem is compounded by the issue of input vector rejection, which is also directly related to the input window length. The proportion of rejected vectors will increase as the length of the input window increases, possibly compromising the training process. Thus, vector rejection is very likely to be a limiting factor on the optimisation of the uninterpolated RBF model. For the interpolated models, this limiting factor is removed and the interpolated model is free to adopt a longer optimal window length if it will improve the model performance.

These limiting factors will also directly influence the testing process in a similar fashion, by altering the number of available test vectors and hence the assessment of the optimum parameters as determined by the RMS error on the test set. For the hourly *foF2* time series, the interpolated RBF model uses more than twice as many test and training vectors as the uninterpolated model, as vector rejection has dropped from 60% to 6.6%.

The arguments applied to the choice of input window length also hold true for all of the other free model parameters. These parameters include the principal component selection, the number of functions used to construct the model and the placement of functional centres. The optimisation issues are also compounded by the iterative nature of the interpolation process. At each stage, the model must be optimised anew in some fashion that is dependent upon interpolations derived during previous optimisations. This very large meta-optimisation problem is based upon individual model optimisations that are already very demanding in terms of required processing resources. These limitations precluded re-

optimisation of the RBF model at successive stages of the interpolation process. It is very likely, therefore, that the current interpolated RBF model may be improved given greater processing resources.

In summary, there are many degrees of freedom in this meta-optimisation, with a degree of linkage between stages. Model parameters are not linearly separable and must be treated as a coupled whole. For example, the optimal PCA derived at one stage during the iteration of the interpolative process might discard information that could be vital for the next iteration. Feedback is therefore required in both directions, between iterations of the interpolation process. Again, heuristic approaches may be the appropriate response to this problem and it is likely that very significant gains in predictive and interpolative accuracy can be made using such techniques.

6.10 *Conclusions*

This chapter describes a novel and general scheme to deal with the problem of non-contiguity in geophysical time series. The scheme can be used to improve the viability, accuracy and generality of the underlying non-interpolated predictive model, derived in the previous chapter. This method can then be used to provide real-time predictions even when the input time series contains missing data points. This baseline nonlinear interpolation scheme provides a significant improvement upon simple schemes that rely upon persistence or recurrence interpolation of an observed ionospheric parameter.

In instances where simple periodic climatological models cannot be used to provide realistic interpolations of missing points, this technique will find the best interpolation solution for that particular problem. In such a case, the resultant improvement may be greater than for the prediction of the hourly *foF2* time series, where the dominant diurnal variation of this ionospheric parameter is reasonably well modelled by simple recurrence.

7.0 Conclusions

7.1 *Introduction*

The central theme of this thesis has been the development of novel and robust nonlinear techniques that can be used to analyse and predict geophysical time series. Simple linear techniques have achieved a measure of success in these areas, but are not adequate for the task of characterising the nonlinear dynamics that underlie the behaviour of the solar-terrestrial system.

The MHD equations that describe the solar-terrestrial environment are currently not thought to have tractable analytical nonlinear solutions. Consequently, it is an attractive idea to look for nonlinear features in experimental data, in order to develop nonlinear models and analytical tools using empirical methods. However, existing techniques in the field of nonlinear mathematics have been developed for well-controlled experimental systems or simulations. Therefore, these methods work best upon large sets of clean, stationary data. In practice, solar-terrestrial data sets are very different from this ideal; generally they are noisy and highly quantised, with many data dropouts and are rarely stationary, due to the large range of physically significant time-scales. As a result, many of the existing techniques are not suitable for application to such data and will give spurious results without additional safeguards. Thus, alternative methods are required to either assess nonlinear behaviour within data sets or validate the use of these standard nonlinear techniques.

The techniques presented in this thesis have been developed to cope with the problems of noise and non-contiguity that are typical of solar-terrestrial data sets. Existing nonlinear analysis techniques are prone to giving spurious indications of nonlinear behaviour in such circumstances. These new techniques can be applied as either an alternative to established techniques, to determine less fragile and more qualitative nonlinear properties that may be present, or as a precursor - to

determine whether or not it is viable to proceed with the application of standard nonlinear analysis methods. The methods employed have been derived from studies into nonlinear dynamical systems theory and artificial intelligence, specifically neural networks, and have a general applicability that extends to any time series prediction problem. Such tools can be used to create knowledge independent and self-determining nonlinear models that do not require a complete understanding of solar-terrestrial physics.

7.2 *Gains and Advancement*

This thesis has attempted to take a rigorous approach to the application of neural networks to the particular problem of solar-terrestrial prediction. By avoiding the widespread 'black box' approach to developing these modelling techniques, it has been possible to evolve analytical and predictive techniques that are robust and capable of handling typical geophysical data sets.

The first application of these techniques pertains to the analysis of the SABRE data, which can be used to infer the behaviour of the solar sector structure. In chapter 4, this data set was examined for evidence of nonlinear dynamical behaviour, as originally postulated by *Yeoman et al.* [*private communication*]. A combination of linearly orthogonal filtering and neural net based nonlinear prediction was applied to the problem, in conjunction with a set of nested testable hypotheses concerning the underlying nature of the SABRE data. It was determined that although there is evidence of nonlinear behaviour in the SABRE data, it is not distinguishable from a nonlinearly transformed coloured Gaussian noise process in either the descending or the ascending phase of the solar cycle. In such a case, standard nonlinear analysis techniques would either provide negative or spurious results so no additional knowledge could be gained through their application. Consequently, it is not possible to support the hypothesis that increased nonlinear dynamical behaviour dominates, or even plays a significant part in, the behaviour of the ascending phase of SC 22. Contrary to expectations,

the evidence seems to suggest that the descending phase of SC 21 exhibited more obvious nonlinear properties than the ascending phase of SC 22, which had less structure and was more noise-like.

Chapter 5 of this thesis presented a technique that can deal with the prediction of typical solar-terrestrial time series, or real-world data in general. Nonlinear radial basis function (RBF) neural networks (NNs) were used to model the variation of the critical frequency of the *F2* layer of the ionosphere, *foF2*, on hourly, daily and monthly time scales. The benefits of principal component analysis based noise reduction were assessed and found to offer a substantial improvement for the daily RBF model. Noise reduction was additionally found to reduce the processing requirements in constructing all of the RBF models.

The RBF NN prediction models were found to be significantly more accurate (40-60%) than the standard persistence and recurrence models, which act as references for the assessment of any proposed prediction scheme. RMS errors were in the range 0.45 to 0.5 MHz for the hourly, daily and monthly models. Many existing applications of AI techniques utilise input parameters from supplementary time series that are concomitant with the future point to be predicted. The NN techniques presented in this thesis do not suffer from such a drawback and form a viable basis for a real time ionospheric forecasting service.

Chapter 5 also highlighted the need for a novel nonlinear technique that optimally interpolates missing data points in any specific input time series. This should be done in such a way that the effects of the interpolation are minimised for any specific model that uses that time series as an input.

It was stressed that optimising the models for each individual prediction time scale would bring further improvements to the accuracy of the predictions. It was also noted that normalised prediction errors for all three model time scales were very close to the irreducible error associated with the input time series. All three

models used data derived from an hourly time *foF2* time series, which has a within hour variation of between 5-30% at mid-latitudes [*Professor T. R. Robinson, private communication*]. This variation is comparable to the observed magnitudes of the normalised RMS errors for the predictive models (between 20% and 25%). This indicates that the accuracy of current *foF2* predictive models could be improved by higher time resolution data.

Chapter 6 of this thesis described a novel and general scheme to deal with the problem of non-contiguity in geophysical time series. Missing data points are a major difficulty associated with all geophysical data sets. It is a particular problem for data measured by spacecraft, where telemetry often fails and large sections of data measurements can be lost. Similarly, ground based measurements are also prone to dropout, due to equipment limitations and failure. Simple linear interpolation schemes cannot be used to overcome this problem because the time series involved are nonlinear in nature.

The interpolated RBF models presented in Chapter 6 can be used to improve the viability, accuracy and generality of the underlying non-interpolated predictive model, derived in chapter 5. These methods can be used to provide real-time predictions even when the input time series contains missing data points. Such a capability is essential to ensure that it is possible to provide continuous predictions in the presence of data that contains a high proportion of data dropouts. In this study, the input vector rejection rate was reduced from 60% to 7%, with corresponding benefits to the model training process. However, the main advantage of this technique is apparent when data containing missing data points is used as an input to the model. For the data set considered, the uninterpolated predictive model produces an RMS error of 0.653 MHz. Utilising the interpolation algorithm lowered the error by 26% to 0.486 MHz, almost equivalent to the performance of the uninterpolated model upon data that contains no gaps (RMS error ~ 0.452 MHz).

7.3 *Broader Scope*

The techniques developed during the analysis of the SABRE data have general applicability as time series classification tools that can be applied to any data set. By removing the obvious trends and linear structure within a given time series, it is possible to apply these techniques to classify the contribution and nature of any nonlinear behaviour within the data set. This allows us to quantify the relative contributions of low dimensional nonlinear effects in comparison to stochastic processes or high dimensional nonlinear dynamics that cannot be resolved given an input time series of limited length.

It is in fact possible using these general techniques, developed in Chapter 3 and used in Chapter 4, to create a model of any nonlinear behaviour and subtract it from the time series. This process just leaves the noise and high-dimensional nonlinear components that the technique has been unable to classify. This residue could then be subjected to further analysis to quantify these aspects of the data and to categorise the features that the model finds difficult to predict. This residue ultimately defines the absolute limit of predictability for a nonlinear model of a particular time series.

This approach can be further extended by examining the behaviour of the prediction error for a set of predictive models for a specific time series that predict the time series a number of consecutive steps into the future. The predictive accuracy curve, as the models extend further into the future, provides a description of the temporal correlation of the system over time. Linear periodic systems exhibit a periodic correlation of the model error as a function of prediction time scale. In comparison, a constant level of correlation is observed for a purely stochastic process. Correlation for a nonlinear time series, however, should fall off sharply with time as the prediction extends further into the future.

Having examined the time series in question using this methodology, the nonlinear models constructed can be compared with nested sets of testable hypotheses to determine the presence and character of the nonlinear properties that might be present in the data in more detail. Although the SABRE data contained low-level nonlinear behaviour (can be distinguished from type I surrogate), the presence of nonlinear dynamical behaviour could not be determined (cannot be distinguished from type II surrogate). Consequently, there was no point proceeding to more complex surrogate tests or applying standard nonlinear analysis techniques, which are dependent upon phase information that has been shown to be absent from this data set. For time series that can be distinguished from type II surrogate data, subsequent surrogate tests might encompass tests for distinct types of nonlinear dynamical behaviour. Alternatively, surrogate tests could then be used to validate results obtained from standard nonlinear tests for dimensionality and power law spectrum, given that the methods presented in the thesis had already been used to show that nonlinear dynamics were present in the data and were not an artefact of the analysis process or the noise structure of the data. Standard nonlinear techniques are susceptible to this problem and employing surrogate tests to establish the presence of nonlinear dynamics and then assess the significance of the results obtained using the standard nonlinear techniques is an essential safeguard.

The nonlinear RBF modelling techniques presented in Chapter 5 have a general applicability that extends to any time series prediction problem, and are not necessarily limited to geophysical applications. Further applications of this work to geophysics are discussed in section 7.5.

The gap-filling technique presented in chapter 6 will find the best interpolation solution for any general problem and is most useful in those instances where simple periodic climatological models cannot be used to provide realistic interpolations of missing points. In such a case, the interpolation technique may provide a greater advantage than is evident for the hourly *foF2* time series, where

the dominant diurnal variation of this ionospheric parameter can be reasonably modelled by simple recurrence. In addition, this interpolative technique has general application in any instance where the effects of interpolation upon a given analysis process need to be minimised or a complete time series needs to be constructed from incomplete data.

7.4 *Future Theory Development*

Further enhancements of the underlying theoretical framework and the existing modelling capability fall into one of three categories: improved model optimisation, extension of existing methods and development of new techniques. These topics are dealt with separately in the following sections.

7.4.1 *Improved Model Optimisation*

As discussed in chapter 6, the determination of the best (uninterpolated) RBF model for a given time series is not currently optimal in the strictest sense of the word, though the RBF model itself is optimal for a given set of free parameters (number of centres, window length etc). Instead, the model free parameters are optimal within a set of constraints imposed by limitations on available processing resources. For the uninterpolated modelling process, it would be possible to overcome these constraints with greater resources and locate the true optimal configuration for each time series to be modelled using exhaustive search methods.

The iterative nature of the interpolated modelling process means that it is not possible to perform an exhaustive search across the whole of the solution space for this type of modelling scheme. The individual iterations of the modelling process cannot be decomposed into separate optimisations. Therefore, the set of free parameters increases in direct proportion to the number of iterations. However, the size of the solution space increases as the exponent of the number

of free parameters, as dictated by permutation theory. In addition, the objective function to be minimised in equation (6.3) changes at each iteration because the interpolated input vectors are used to create a new predictive model at each stage of the process. For the purposes of this thesis, the same parameter set has been used at each iteration of the modelling process. This is far from ideal but it does allow the interpolation to be carried out within realistic time scales. However, techniques are available that can be used to perform efficient searches within very high dimensional optimisation problems. Heuristic techniques are one such approach and can be readily applied to solve combinatorial optimisation problems of this type. These techniques have evolved in response to the need for 'rule of thumb' methods for determining near optimal solutions in very large search spaces. These techniques include methods such as taboo search, simulated annealing and genetic algorithms. Such methods could fully automate the optimisation of the model free parameters and result in a more thorough search of the solution space. Heuristic techniques may also be applied to the actual interpolation process itself. The resulting solutions would be more globally optimum, in terms of minimising the error introduced into the predictive function by interpolating points.

The next area of model optimisation relates to the fitting of the radial basis function set to the data. Currently, singular value decomposition (SVD) is used to determine the optimal weights by minimising the linear least squares error between the output of the constructed function and the input time series used as training data. However, it is possible to improve upon this technique by adopting nonlinear minimisation methods that do not limit the functional fitting process with linear constraints. Such techniques will improve the quality of the fit, but the RBF will no longer be a single pass process. Again, heuristics could be used to provide a mechanism for this approach.

The effectiveness of the nonlinear models that have been derived in this thesis has been determined using variants of the root mean square error. It has been

used as a means of representing the accuracy of the models as well as a means of assessing the optimal choice of free parameters for each model. The use of this error measure in the choice of free parameters can be deemed an optimisation issue. For example, RMS error stresses the accurate modelling of peaks at the expense of other characteristics within the data. Alternative error measures would increase the emphasis of different features within the data, according to requirements. The stringency of the testing criteria can also be increased by using the model predicted output (MPO) at one step as training input to the model at the next, such that the model output feeds back into the model input. Actual input test data is only used to start the process until MPO points are available, though the effectiveness of the model is still assessed using the original desired output for the test data. This improves the ability of the model to capture the underlying dynamics of the data, as opposed to creating just a one step ahead predictor.

7.4.2 *Extension of Existing Methods*

It has been noted that linear predictive models can be more effective than nonlinear models for periods of low geomagnetic activity. This observation can be attributed to the fact that linear models are less prone to over fitting when applied to purely linear data. It has also been noted that the efficacy of either modelling approach can be degraded by the presence of non-stationarity within an input time series.

A modified modelling scheme has been proposed to tackle the problems of nonlinear over-fitting and non-stationarity, but has not yet been tested. The first stage of this process employs physical knowledge concerning the input time series to remove the non-stationarity of the data. This could be achieved by taking first or second differences of the input time series or by removing simple climate based trends, such as recurrence, from the data. More sophisticated techniques might involve the application of adaptive nonlinear techniques such as Kalman filtering or RBF models of the non-stationarity in the time series.

The removal of the obvious long-term trends from the data enables the subsequent modelling procedures to focus more clearly upon the remaining features within the resultant stationary time series. The original time series would also be used as an input to subsequent modelling stages. This is necessary to allow the model to relate features within the stationary data to the original time series, as features in the former may be correlated in some sense to the non-stationarity of the original data.

A hybridised nonlinear / linear modelling technique could be implemented to produce a model of the data that is optimal with respect to both the linear and nonlinear aspects of fitting the input time series. A single pass process fits the linear and nonlinear adaptability of the model simultaneously. This approach has the advantage of simplicity and ease of use over separate linear and nonlinear modelling stages but it is more difficult to assess the relative contributions of each of the models.

It is also planned to develop a weighting scheme that will accord more importance to periods of unusually high geomagnetic activity in the training phase of model development. This type of functionality is related to the partial confidence concept suggested in Chapter 6. The weighting could be applied as a multiplier to the confidences associated with all sets of points that represent periods of special interest during the training process. Alternatively, training weight could be designated as an additional parameter to accompany the confidence level already associated with each data point.

The concept of partial confidences could also be extended to encompass the data quality field values that are often associated with geophysical time series. These fields provide additional information about the associated data point, such as noting uncertainty or whether the value is an upper or lower limit. This information could be used to modify the data confidence value and adjust the importance of that point in the training process. By going one step further still, it

might be possible to use the interpolation scheme to adjust the value of such data points in light of subsequent data points and predictions, even though they represent legitimate measurements.

Finally, the functions used to provide the nonlinear adaptability of the RBF modelling techniques need to be selected to suit the characteristics of the time series being modelled. For example, for a time series with a very wide dynamic range, it may be necessary to transform the input data logarithmically before applying the RBF model. This would improve the ability of the model to track dynamic events and estimate their magnitudes but at the expense of quiet time performance. A similar effect could also be achieved by using the auto-correlated errors on previous predictions to provide feedback to the current prediction.

7.4.3 *Development of New Techniques*

This thesis has highlighted areas where new techniques, not necessarily based upon time series methods, could be employed to widen the scope of the existing work and to provide additional functionality.

The first of these areas shifts the emphasis of the modelling from the prediction of exact time series values to the classification, extraction or prediction of patterns and features within the data. Fuzzy logic, Bayesian statistics and self-organising mapping (SOM) techniques could all be employed to classify patterns within the data and predict the occurrence probabilities of pre-defined events. A specific application would be prediction of geomagnetic storms from solar wind and magnetospheric data [*Chen et al.*, 1996, 1997, 1999]. Exact predictions are not required in this case and it is more appropriate to employ methods that estimate the state of a set of observable systems in a probabilistic fashion according to user-defined categories. Expert knowledge could also be incorporated into such systems more easily than is the case with time series based methods.

For a complete space weather forecasting tool, it is likely that many such models of varying types would need to be employed to achieve the overall aims of such a service. A hierarchical structure composed of individual time series and pattern-based models could be integrated by using expert systems, neural networks or statistical methods. An example of such an integrated model would be a pattern-based system to classify behaviour of the solar wind into disturbed or quiet time conditions. An appropriate ionospheric predictor could then be selected from a collection of models, where each model has been optimally trained for a specific set of observed conditions. As a second example, the outputs from a complete set of solar-terrestrial models and classification tools could be used to feed a classification system that would issue appropriate warnings to users.

Data related issues have been a recurrent theme throughout this work, particularly with respect to availability and continuity. In addition, real-time data made available on the WWW often falls far short of the quality that can be achieved for historical time series. The latter have usually been checked by an operator and may be significantly different from the automatically processed real-time data. Days or weeks may elapse before the automatically processed data is checked manually and corrected.

A significant bias may be present in the automatically processed data if the corrections turn out to be systematic rather than random in nature. Consequently, misleading results may be produced using a historically trained model to predict automatically processed real time values. A large error between the prediction and the automatic value may not necessarily indicate an inaccurate prediction. The accuracy of the prediction may not become apparent until the automated value has been manual checked. However, further analysis needs is required to assess the importance of this effect upon real-time forecasting schemes. Possible schemes to mitigate this effect could involve using principle component analysis or classification techniques to pre-process the data or provide a correction factor.

The simple nonlinear functional relationship between input and output data is an important advantage of RBF techniques. It is likely that this simple functional form will prove more amenable to intuitive decomposition and rule extraction than conventional neural network techniques. It may be possible to evolve self-determining rule based systems that would utilise the physical knowledge encapsulated within the RBF network and that these techniques might lead to a better understanding of solar-terrestrial physics.

7.5 *Future Modelling Requirements*

Although it is possible to suggest the methodological enhancements and additional functionality stated above, it is also desirable to apply the baseline nonlinear models and techniques presented in this thesis to investigate additional problems or applications within geophysics of a similar nature to those already described.

Due to processing limitations, the models in this thesis have necessarily been restricted to the problem of self-prediction, where the future value of a parameter is predicted solely from the history of its time series. However, as disturbances propagate from the sun to the near-earth environment, it is to be expected that additional time series inputs in the form of precursive information supplied by up-stream and solar monitoring would lead to substantial improvements in predictive accuracy. Satellites such as SOHO, WIND and ACE can provide such observational and in-situ time series. It is to be hoped that these capabilities will eventually be supplemented by satellites that image the ionosphere, the magnetosphere and the solar wind over greater spatial scales at high time resolution.

Models derived from data provided by these monitors can be used to provide precursive inputs to existing magnetospheric and ionospheric models. For example, solar wind data from ACE can be used to help model the behaviour of

the ring current system. The level of activity in this region of the inner magnetosphere is closely related to satellite risk levels and effects within the auroral region of the ionosphere. Existing models of *foF2* may also be improved by incorporating parameters indicative of behaviour within the solar wind and magnetosphere.

It was noted that the RMS errors for the hourly *foF2* model were very close to the irreducible error attributable to the within hour variation for the input time series. This suggests that utilising higher resolution input time series would yield improvements in predictive performance. In general, care must be taken during the determination of a model to ensure that the inputs accurately represent the long term and short-term variations of a parameter. The model performance will be degraded by aliasing effects without sufficient short-term information. Similarly, insufficient long-term information makes it difficult to follow any non-stationarity that may be present in the system. The model will cease to be effective as soon as the data used to train the model is no longer representative of the training data. It is often necessary to trade off between processing requirements, generality and accuracy during the training process to achieve specific operational aims. Consequently, it is likely that the predictive models evolved during the developmental phase of this project will need to be reassessed when they are incorporated into applications.

Any spatial correlation of parameters would allow the possibility of using the data from one station to predict data at another station. Initially, these stations would be chosen from approximately the same latitudes, at varying longitudes. The behaviour of ionosphere varies primarily by latitude so keeping this variable fixed would reduce the complexity of the modelling process. Subsequent tests would then predict readings from stations that lie along the same line of longitude, a much more taxing task for an ionospheric model. Cross prediction between stations could be used to provide additional constraints for the interpolation process and to give advance warning of activities that may be taking

place at one station before the effect has had time to propagate to the next. A complete set of predictions for a network of stations could be used to provide a contour map for a parameter over a given area. The geographical covariance function, derived from the prediction accuracy plotted as a function of the distance between any two stations, could be used in conjunction with a contouring technique to produce an ionospheric area map.

7.6 *Forecasting Demonstrator*

The techniques presented in this thesis have been adapted to form the basis of a real-time ionospheric forecasting demonstrator. The Ionospheric Forecasting Demonstrator (IFD) is a schedule-based event driven modular system that can be run in either interactive or automatic modes. It controls data download and processing (including optical character recognition) together with the prediction algorithms and the graphical output of results via WWW or email. Custom schedules can be specified to suit specific users. The model framework allows the incorporation of additional models, data sources, algorithms and output formats.

As an example of our capability, *foF2* forecasts based upon Slough, UK measurements and real-time, hourly updated *foF2* forecasts from 1 to 24 hours ahead are released online (http://rasp.dera.gov.uk/ifs/ifs.html). This scheme will be extended to include additional stations and ionospheric parameters.

By incorporating the future modelling requirements outlined in the previous section, the IFD could be broadened to encompass space weather. It could specifically incorporate magnetospheric and solar parameters that are relevant to the problem of satellite anomaly prediction. Such predictions are of great importance to both military and civilian HF users, as well as academia, and could potentially pull through to civil mobile communications, ESA, NASA, GPS operators and utility companies.

Bibliography

Baker, D. N., A. J. Klimas, R. L. McPherron, J. Buchner, 'The Evolution from Weak to Strong Geomagnetic Activity: an Interpretation in terms of Deterministic Chaos', *Geo. Res. Letters, Vol. 17, No. 1,* 41-44, Jan 1990.

Broomhead, D. S., G. P. King, 'Extracting Qualitative Dynamics from Experimental Data', *Physica D, 20,* 217-236, 1986.

Broomhead, D. S., D. Lowe 'Multivariable functional interpolation and adaptive networks', *Complex Systems 2*, 321-355, 1988.

Brown, A. G., N. M. Francis, D. S. Broomhead, P. S. Cannon, A. Akram, 'Is there Evidence for the Existence of Nonlinear Behaviour within the Interplanetary Solar Sector Structure?', *J. Geophys. Res., 104, A6*, 12537-12547, June 1999.

Cander, L. R., S. Stankovic, M. Milosavljevic, 'Dynamic Ionospheric Prediction by Neural Networks', *AI Applications in Solar-Terrestrial Physics,* conference proceedings, 1997.

Chapman, S., J. Bartels, *Geomagnetism*, Vol. 1, Ch. 12, Clarendon, Oxford, 1940.

Chen, J., P. J. Cargill and P. Palmadesso, 'Real-time Identification and Prediction of Geoeffective Solar Wind Structures', *Geophys. Res. Lett., 23*, 625-628, 1996.

Chen, J., P. J. Cargill and P. Palmadesso, 'Predicting Geoeffective Solar Wind Structures', *J. Geophys. Res., 102,* 14701-14720, 1997.

Chen, J., P. J. Cargill and P. Palmadesso, 'Prediction of Geomagnetic Storms Associated with Halo CMEs of January and April 1997', *Proc. Workshop on Space Weather, ESA WPP-155,* 403-406, 1999.

Cowley, S. W. H., 'The Earth's Magnetosphere', *Earth in Space, Vol. 8, No. 7,* March 1996.

Delouis, H., P. N. Mayaud, 'Spectral Analysis of the Geomagnetic Activity Index aa over a 103-year Interval', *J. Geophys. Res., 80,* 4681-4688, 1975.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Nonlinear Prediction of Ionospheric indices (*foF2*)', *AI Applications in Solar-Terrestrial Physics,* conference proceedings, 1997a.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead,' Nonlinear prediction of ionospheric parameters important for forecasting the battle-space environment', *Agard SPP Symposium,* conference proceedings, 1997b.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Novel nonlinear techniques for the prediction of noisy geophysical time series that contain a significant proportion of data dropouts', *ESTEC Space Weather Workshop, conference proceedings WPP155, 419-422,* 1998.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Nonlinear prediction of the hourly *foF2* time series and the nonlinear interpolation of missing points', *IEE Colloquium Proceedings #99/107,* March 1999a.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Nonlinear prediction and interpolation of the hourly *foF2* time series', Accepted for publication in *Physics and Chemistry of the Earth – EGS Proceedings special issue* (solicited paper), 1999b.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Prediction of the Ionospheric Parameter, *foF2*, on hourly, daily and monthly time-scales', *J. Geophys. Res., 105, A6*, 12839-12849, June 2000.

Francis, N. M., A. G. Brown, P. S. Cannon, D. S. Broomhead, 'Prediction of the Hourly Ionospheric Parameter, *foF2*, Incorporating a Novel Nonlinear Interpolation Technique to Cope with Missing Data Points', *J. Geophys. Res., 106, A12,* 30,077 (2000JA002227), December 2001.

Gibson, J., Farmer, J. D., Casdagli, M., Eubank, S., 'An Analytic Approach To Practical State Space Reconstruction', *Physica D, 57,1*, 1-30, 1992.

Golub, G. H., Van Loan, C. F., *Matrix Computations*, The Johns-Hopkins University Press, 1989.

Gosling, J. T., J. R. Asbridge, S. J. Bame and W. C. Feldman, 'Solar Wind Speed Variations: 1962-1974', *J. Geophys. Res., 81*, 5061-5070, 1976.

Grassberger, P., I. Procaccia, 'Measuring the Strangeness of Strange Attractors', *Physica D, 9, 189*, 1983.

Greenwald, R. A., K. B. Baker, J. R. Dudeney, M. Pinnock, T. B. Jones, E. C. Thomas, J. P. Villain, J. Cerisier, C. Senior, C. Hanuise, R. D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A. D. M. Walker, N. Sato, and H. Yamagishi, 'Darn/SuperDARN: a global view of the dynamics of high-latitude convection', *Space Science Reviews, 71*, 761, 1995.

Hargreaves, J. K., *The Solar-Terrestrial Environment*, Cambridge Atmospheric and Space Science Series (5), Cambridge University Press, UK 1992.

Hertz, J., Krogh, A., Palmer, R. G., *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies, Addison Wesley, 1991.

Horn, R. A., C. R, Johnson, *Topics in Matrix Analysis,* Cambridge University Press, UK, 1991.

Joselyn, J. A., 'Geomagnetic Activity Forecasting: The State of the Art', *Reviews of Geophysics, 33, 3,* August 1995, P383-401, Paper number 95RG01304.

Klimas, A. J., D. N. Baker, D. A. Roberts, D. H. Fairfield, 'A Nonlinear Dynamical Analogue Model of Geomagnetic Activity', *J. Geophys. Res., 97, A8,* 12,253-12266, Aug. 1992.

Klimas, A. J., D. Vassiliadis, D. N. Baker, D. A. Roberts, 'The Organised Nonlinear Dynamics of the Magnetosphere', *J. Geophys. Res., 101, A6,* 13089-13133, 1996.

Koskinen, H., T. Pulkkinen, 'State of the Art of Space Weather Modelling and Proposed ESA Strategy', Finnish Meteorological Institute Geophysical Research, 1998.

Lamming, X., L. R. Cander, 'Monthly Median Ionospheric Frequencies Prediction with Neural Networks', *AI Applications in Solar-Terrestrial Physics,* University of Lund conference proceedings, 1997.

McKinnell, L. A., A. W. V. Poole, 'The development of a Neural Network Based Short-Term *foF2* forecast program', Accepted for publication in *Physics and Chemistry of the Earth – EGS Proceedings special issue,* 1999.

Milsom, J. D., S. J. Goldsmith, D. W. Green, C. R. Telfer, N. S. Wheadon, 'Research and Development of a short-term Ionospheric Forecasting Service', Marconi Research Centre, *MTR 89/21*, 1989.

Milsom, J. D., N. G. Riley, C. R. Telfer, 'Recommendations for the Advancement of the MRC Short-term Ionospheric Forecasting Service', Marconi Research Centre, *MTR 87/94*, 1994.

Nielson, E., W. Guttler, E. C. Thomas, C. P. Stewart, T. B. Jones, A. Hedburg, 'SABRE – a New Radar Auroral Back-scatter experiment', *Nature, 304,* 712-714, 1983.

Ott, E., *Chaos in Dynamical Systems,* Cambridge University Press, UK, 1993.

Packard, N., J. Crutchfield, D. Farmer, R. Shaw, 'Geometry From A Time Series', *Phys. Rev. Lett. 45,* 712-716, 1980.

Press, W. H., S. A. Teulosky, W. T. Vettering, B. P. Flannery, 'Numerical Recipes in C (2nd Edition)', Cambridge Press, 1992.

Robinson, T. R., 'The impact of solar variability on climate', *Astronomy and Geophysics, 39*, 5.4, 1998.

Ruelle, D., and F. Takens, 'On the Nature of Turbulence', *Communications of Mathematical Physics, 20*, 167-192, 1971.

Sauer, T., J. A. Yorke, M. Casdagli, 'Embedology', *J. Stat. Phys., 65*, 579-616, 1991.

Sharma, A. S., 'Assessing the Magnetosphere's nonlinear behaviour; Its dimension is low, its predictability, high', *Reviews of Geophysics, Supp*, 645, July 1995.

Smith, L. A., 'Does a Meeting in Santa Fe Imply Chaos?', *Time Series Prediction*, 323-343, Santa Fe Institute, Addison-Wesley, 1994.

Stamper, R., 'Improved Prediction of I_{F2} and I_G Indices using Neural Networks', *IEE Proceedings – Microwaves, Antennas and Propagation, 143, 4*, 341-246, August 1996.

Takens, F., 'Detecting Strange Attractors in Turbulence', *Lecture Notes in Mathematics, 898*, pp366-381. Springer, Berlin, 1981.

Theiler, J., B. Galdrikian, A. Longtin, S. Eubank, J. D. Farmer, 'Using Surrogate Data To Detect Nonlinearity In Time Series', *Nonlinear Modelling and Forecasting*, Addison Wesley, 1992.

Vassiliadis, D., A. S. Sharma, K. Papadopoulos, 'Time Series Analysis of Magnetospheric Activity using Nonlinear Dynamical Methods', *Chaotic dynamics: Theory and Practice*, Plenum Press, New York, 1992.

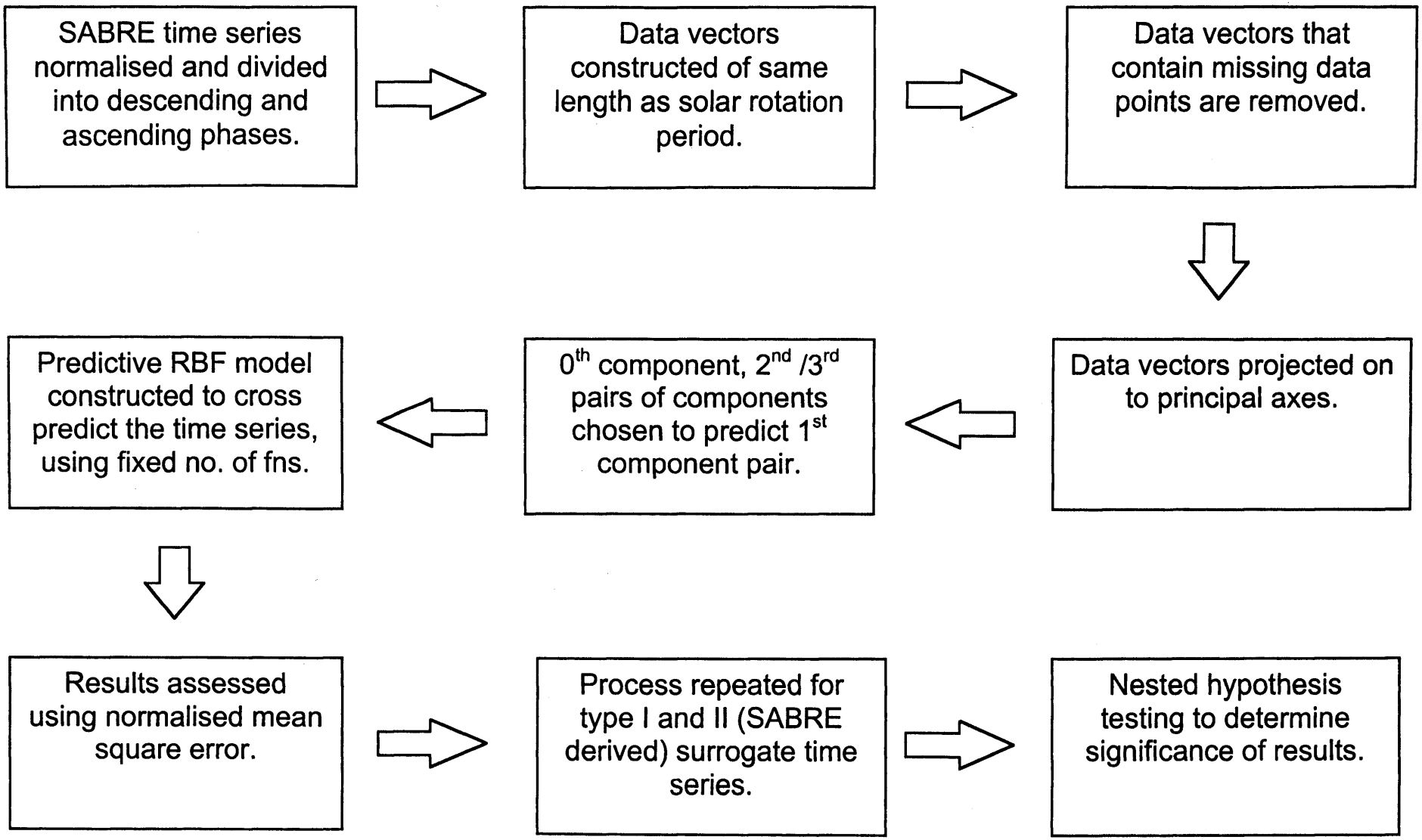
Vassiliadis D., A. J. Klimas, D. N. Baker, D. A. Roberts, 'A Description of the Solar-Wind Magnetosphere Coupling Based on Nonlinear Filters', *J. Geophys. Res., 100, A3*, 3495-3512, 1995.

Williscroft L. A., A. W. V. Poole, 'Neural Networks, *foF2*, Sunspot Number and Magnetic Activity', *Geophysical Research Letters 23, 24*, 1996.

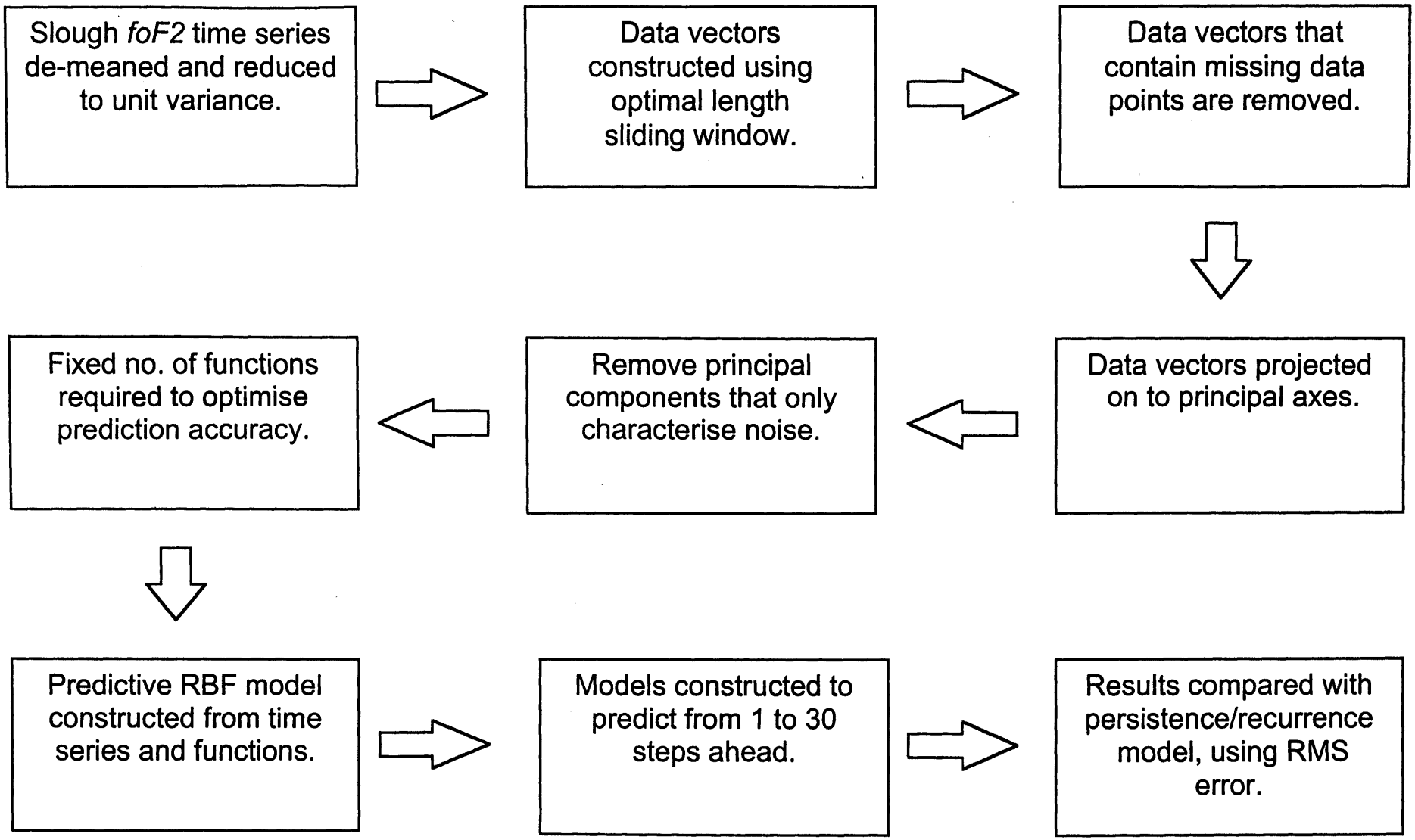
Wintoft, P., L. R. Cander, 'Ionospheric *foF2* Storm Forecasting using Neural Networks', Accepted for publication in *Physics and Chemistry of the Earth – EGS Proceedings special issue,* 1999.

Wu, J.G., H. Lundstedt, 'Geomagnetic Storm Predictions from Solar-Wind Data with the use of Dynamic neural Networks', *J. Geophys. Res., 102, A7,* 14255-14268, 1997.

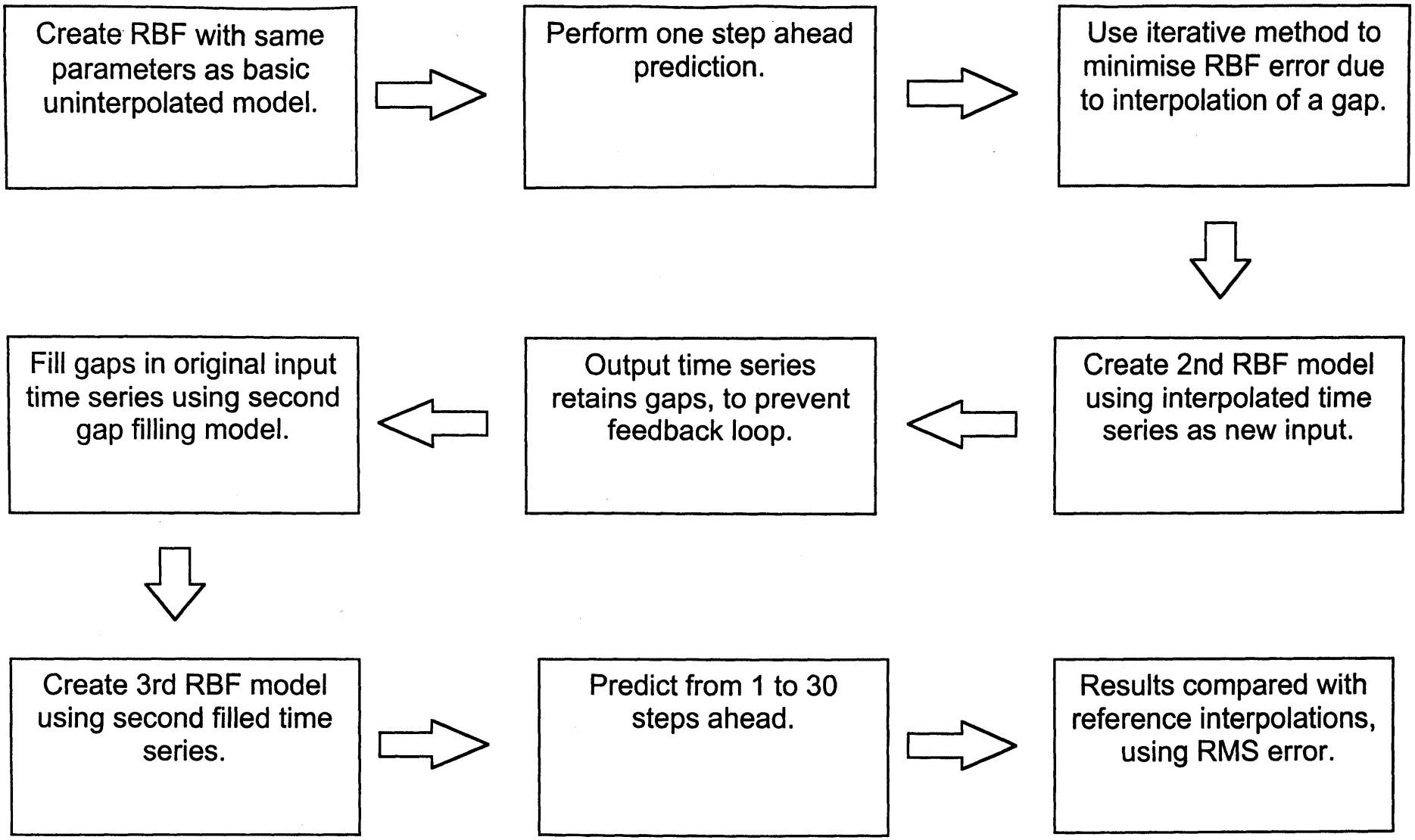
Yeoman, T. K., M. D. Burrage, M. Lester, T. R. Robinson and T. B. Jones, 'Long-term Variation of Radar-Auroral Back-scatter and the Interplanetary Sector Structure', *J. Geophys. Res., 95, A12,* 21123-21132, Dec., 1990.



Flowchart 1 Block diagram representing nonlinear analysis process



Flowchart 2 Block diagram representing non-interpolated modelling process



Flowchart 3 Block diagram representing interpolation modelling