THE UNIVERSITY OF CALGARY

Wide Area Ionosphere Grid Modelling in the Auroral Region

by

Susan Skone

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

DECEMBER, 1998

ABSTRACT

Dual frequency GPS receivers enable the estimation of absolute ionospheric delay and total electron content (TEC) along the signal path. By using a number of reference stations, each equipped with a dual frequency receiver, it is possible to estimate values of the vertical ionospheric delay at a set of designated grid points (in latitude and longitude) on an ionosphere shell. This type of ionosphere delay modelling is employed in wide area differential GPS (WADGPS) networks, where grid accuracies generally depend on the temporal/spatial correlations of TEC. These models can suffer degraded performance in regions, such as the high latitude auroral zone, where spatial gradients and temporal variations of electron density may differ significantly from assumptions. Given that future extensions of WADGPS applications include implementation of safety-critical systems to support air navigation in auroral regions, such as the wide area augmentation system (WAAS - *FAA*[1994]), it is important to establish the impact of auroral effects on such GPS networks.

In the research presented here, a detailed study of auroral effects on GPS is conducted, and performance of the ionosphere grid model is investigated in the auroral region. The development of a Canadian wide area network in 1996, by Natural Resources Canada (NRCan) [*Caissy et al.*, 1996], allowed compilation of an extensive GPS data set from the auroral region. In this thesis, periods of enhanced auroral activity (over Canada) are identified from ground-based magnetometer signatures, using the Canadian Space Agency's CANOPUS MARIA array. Signatures of auroral TEC associated with these disturbances are established using corresponding GPS observations from the NRCan network. A wide area ionosphere grid model is developed, and degraded grid accuracies associated with the auroral features are estimated, under various modelling assumptions. It is observed that grid accuracies are degraded by a factor of 2-5, relative to typical

accuracies, during periods of enhanced auroral activity. In order to mitigate these effects, a modified grid algorithm is developed in which auroral disturbances are detected from time series of TEC observations. Model parameters are adjusted, consistent with local ionospheric processes, which allows improvements in the grid performance of 15-25 percent. Detailed analysis of a substorm event is conducted, where the largest degradations in grid accuracies are observed during the substorm expansive phase. The potential of monitoring auroral phenomena using GPS observations from a wide area network is explored briefly

.

ACKNOWLEDGEMENTS

I would like to first acknowledge the various groups providing data used in this thesis. In particular, I am especially grateful to Natural Resources Canada for providing me extensive amounts of GPS data, which formed the fundamental data set for my research. I also thank the Canadian Space Agency, for allowing me to access magnetic field data from the CANOPUS magnetometer array. I acknowledge groups at NASA, NOAA and the University of Washington, who also contributed data in support of my research.

I am grateful to my supervisor, Elizabeth Cannon, for her support and guidance throughout my research, and for her review of this manuscript. I also wish to acknowledge the Department of Geomatics Engineering and my fellow graduate students, for creating an enjoyable working environment. I also thank David Schwarz and my parents, Anne and John Skone, for their unconditional support and encouragement. My research was supported through funding received from a variety of sources: NSERC, the Amelia Earhart Fellowships, the Killam Fellowships, and the University of Calgary.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEI	auroral E-ionization
CONUS	Continental United States
CWAAS	Canadian wide area augmentation system
FFT	Fast Fourier Transform
GIVE	grid ionosphere vertical error
GPS	Global Positioning System
IGP	ionospheric grid point
IPP	ionospheric pierce point
LAAS	local area augmentation system
MLT	magnetic local time
NRCan	Natural Resources Canada
PRN	pseudorandom noise code
TEC	total electron content (el/m^2)
TECU	total electron content unit (10^{16} el/m^2)
UIVE	user ionosphere vertical error
VPL	vertical protection level
VAL	vertical alert limit
VTEC	vertical total electron content
WAAS	wide area augmentation system
WADGPS	wide area differential GPS

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The Global Positioning System (GPS) is a satellite navigation system in which a receiver's position is computed using range information contained in the satellites' RF signals. Twenty-four satellites currently provide worldwide coverage and positioning capability. GPS satellite signals must travel through the dispersive ionosphere prior to being received at the Earth's surface; this causes ranging/positioning errors dependent on the given signal frequency and ionosphere electron content. This effect is the single largest source of GPS positioning error (up to tens of metres), outside of Selective Availability (errors intentionally imposed by the U.S. military) [cf. *Fraile-Ordonez*, 1995; *Daniell et al.*, 1996].

Current GPS research is focused on developing differential networks capable of reducing these positioning errors, such that GPS can support precise positioning applications requiring metre-level accuracies. Such accuracies may be achieved through the use of differential GPS positioning techniques - the calculation of ranging errors by a monitor station, with corrections broadcast to a user's remote receiver. A sparse array of GPS monitor stations is referred to as a Wide Area Differential GPS (WADGPS) network. Estimates of ionospheric effects are computed using observations from each station in the wide area network and a grid of ionospheric corrections (or, equivalently, total electron content - TEC) is formed and transmitted to remote receivers for correction of local positioning results. An extension of the WADGPS network is the wide area

augmentation system (WAAS), a safety critical system designed to support precision approach air navigation [*Enge and Van Dierendonck*, 1996]. In order to meet WAAS positioning accuracies, the ionosphere model must provide reliable estimates of ionosphere corrections, within specific error bounds. The safety critical aspect of this system, coupled with an upcoming solar maximum (year 2000), has created an increasing focus on ionospheric research (in the GPS community) in the past few years.

To date many ionospheric correction algorithms have been proposed and tested to meet this specification [i.e. *Mannucci et al.*, 1993; *Komjathy and Langley*, 1996b; *Wu et al.*, 1996; *Draganov et al.*, 1996]. A common approach is to model the ionospheric delay at grid points in real-time, using pseudorange and carrier phase measurements from stations in the network and a Kalman filter technique. These grid accuracies generally depend on estimated temporal and spatial correlations of TEC. Ionospheric grid models can suffer degraded performance in regions, such as the auroral zone, where spatial gradients and temporal variations of electron density may differ significantly from assumptions [cf. *Skone and Cannon*, 1998]. This is an important consideration for WAAS implementation in high latitude regions such as Canada and Alaska.

The auroral zone is characterized by particle precipitation events (i.e. auroral E-ionization and magnetospheric substorms) which cause significant variations in both the vertical and horizontal electron density gradients. These effects are expected to increase in the next few years, approaching solar maximum. During this time, wide area networks will be implemented, or current wide area applications extended, to support enroute through precision approach flight operations in the auroral region (i.e. WAAS [*FAA*, 1997]). Given the variable nature of auroral region TEC, it is therefore important to establish signatures associated with enhanced auroral activity and substorm phenomena. Additionally, and more importantly, it is necessary to establish reliable estimates of ionosphere grid accuracies during such periods.

There have been few previous studies of variations in auroral region TEC, using GPS measurements, to date. *Hunsucker et al.* [1995] conducted an analysis of GPS-derived TEC variations during periods of auroral E-ionization (AEI), using observations from a single reference station in Fairbanks, Alaska. *Coker et al.* [1995] extended this work to investigate the potential of using TEC signatures to detect periods of AEI. Several Stanford researchers are currently studying the tracking performance of GPS receivers in the recently-established National Satellite TestBed (NSTB) Alaska stations, during periods of ionospheric scintillation.

In 1992, the MITRE Corporation and the United States Phillips Laboratory conducted investigations of large-scale spatial gradients in GPS-derived TEC observations at mid-latitudes. Comparisons with 1981 Faraday rotation data were used to establish estimates of grid accuracies at solar maximum [*Klobuchar et al.*, 1995]. These results do not reflect auroral region grid accuracies, however, given the regional nature of ionospheric phenomena. Similarly, investigations of ionospheric grid accuracies during periods of ionospheric and magnetic storms have been conducted primarily for the mid-latitude Continental United States [*Chavin*, 1996; *Chavin*, 1997; *Conker et al.*, 1997]. For future wide area network implementation and applications, it is necessary to also conduct a detailed analysis of ionospheric grid accuracies in the auroral region. Research of this nature is the focus of this thesis.

Additionally, several such performance evaluations of wide area grid algorithms have relied on planetary indices to identify disturbed ionospheric conditions. The use of such indices (i.e. Kp, AE) to infer levels of ionospheric activity is ambiguous, however, since these indices are based on magnetometer measurements made at various stations distributed around the globe. A large planetary index may not necessarily indicate enhanced ionospheric activity in the wide area network region, and results derived from studies of indices have been inconclusive [*Mannucci et al.*, 1997]. In identifying periods

of enhanced auroral activity in this thesis, measures are taken to ensure that a reliable, consistent data set is identified.

In 1996, the Canadian Active Control System developed a real-time GPS corrections service, based on WADGPS [*Caissy et al.*, 1996]. As part of this program, Natural Resources Canada has established a wide area network of ten GPS reference stations across Canada. Continuous data collection at these stations enabled the archiving of an extensive set of GPS observations from the auroral region. This provided a preliminary opportunity to study auroral effects on GPS and WADGPS applications, as conducted in this thesis.

1.2 OBJECTIVES

Given the future implementation of WADGPS applications in the auroral region, and the lack of comprehensive data sets describing auroral effects on GPS, the objectives of this thesis are threefold:

1) Derive a consistent data set of GPS observations from the auroral region, representative of enhanced auroral activity, and establish both spatial and temporal characteristics of ionospheric irregularities.

2) Develop a wide area ionospheric grid model and assess the performance of this model during the periods of enhanced auroral activity, under various modelling assumptions. Assess the impact of degradations in accuracy with respect to existing WADGPS error bounds.

3) Modify existing ionosphere grid algorithms to improve performance in the auroral region, and establish the feasibility of integrity monitoring via "real-time" detection of auroral disturbances from the GPS data.

1.3 OUTLINE

The first three chapters of this thesis introduce relevant background material and discuss the ionosphere grid model formulation. An introductory discussion of basic GPS theory is provided in Chapter 2, in addition to an overview of large-scale TEC variations and basic ionosphere morphology. Concepts of wide area differential GPS (WADGPS) networks are also introduced. Mathematical formulation of the various ionosphere model components is discussed in Chapter 3, and practical implementation issues are addressed. An overview of auroral zone phenomena is provided in Chapter 4, with definitions of various components of auroral activity. Ground-based signatures of auroral activity are also illustrated.

The research components of this thesis are organized into four chapters. In Chapter 5, a reliable data set is derived for periods of enhanced auroral activity, and spatial/temporal correlation properties estimated. This data set is used throughout the analyses of Chapters 6 and 7.

Chapter 6 provides a detailed analysis of ionosphere grid accuracies under various grid model assumptions. The grid accuracies are interpreted with respect to existing WAAS error bounds, and the reference station coverage necessary for achieving optimal accuracies is investigated. In Chapter 7, an adaptive technique is introduced for improving grid model performance in auroral regions. The mathematical formulation of this technique is discussed in detail, and improvements in grid accuracies are assesssed.

The detailed analysis of a magnetospheric substorm event is presented in Chapter 8. This analysis includes a description of auroral and magnetospheric morphology during the event, as observed from several space-borne instruments. The potential of detecting auroral disturbances in "real-time", using GPS observations, is briefly discussed.

CHAPTER 2

INTRODUCTION TO GPS THEORY AND IONOSPHERIC EFFECTS

2.1 THE GLOBAL POSITIONING SYSTEM (GPS)

In 1973, the United States Department of Defense (DoD) approved a program to develop a satellite positioning system based on *radio ranging* techniques [*Parkinson et al.*, 1995]. It was intended that this be primarily a military system, in which ground-based receivers could compute precise positions based on ranging information derived from artificial satellites' RF signals. Additionally, this system would provide worldwide, all-weather, and continuous positioning capabilities, with an end-state orbital configuration of 24 satellites (21 + 3 spares). This positioning system was to be called the *Global Positioning System* (GPS), in reference to the global coverage, and the artificial satellites referred to as NAVSTARs.

The GPS radio ranging technique was to be based on continuous wave signals, in contrast to earlier satellite laser ranging (i.e. STARLETTE [*CNES*, 1975]) and radar-based (i.e. U.S. DoD C-band, S-band radar) techniques, which required relatively large, immobile and expensive antenna/telescope configurations at the ground receiver. Recent technological advances from other satellite positioning projects were to be incorporated into the GPS program. These included satellite prediction algorithms developed for the TRANSIT Doppler positioning system [*Guier*, 1962], and precision timing standards from the TIMATION project [*Easton*, 1980].

The GPS ranging system was also designed as a *one-way* system [cf. *Wells et al.*, 1987], where signals are transmitted only from the satellite to a passive user. Such properties are a requirement of military systems, where full control is retained by system operators. In such one-way systems, however, there is an offset between the independent satellite and receiver clocks, which translates into a position error (the receiver position being calculated from the estimated delay between times of signal transmission and reception). This timing offset must be included as an unknown in the position solution, in addition to the three position coordinates, for 3-dimensional stand alone point positioning.

While the initial objective of the GPS program was to "drop 5 bombs in the same hole" [cf. *Parkinson et al.*, 1995], numerous civilian applications have been developed from GPS technology; these range from precision farming to search and rescue operations, in addition to the more conventional land, marine and air navigation applications. Additionally, the global nature of GPS allows geodetic ties to be established between continents and islands, bridging large distances and replacing conventional levelling methods. The GPS system consists of three segments: space, control, and user segments. These are described in the next three sub-sections.

2.1.1 The Space Segment

This segment consists of the GPS satellites, which transmit on two modulated signals. The signal structure is described in Section 2.2. The orbital configuration includes 6 orbital planes, inclined at 55 degrees with respect to the Earth's equatorial plane, with 4 satellites distributed approximately equally in each plane. Orbits are nearly circular, with slight perturbations due to effects such as nonsphericity of the Earth (perturbations of the gravity field) and solar radiation pressure [cf. *Hoffmann-Wellenhof*, 1994]. This orbital configuration is illustrated in Figure 2.1.

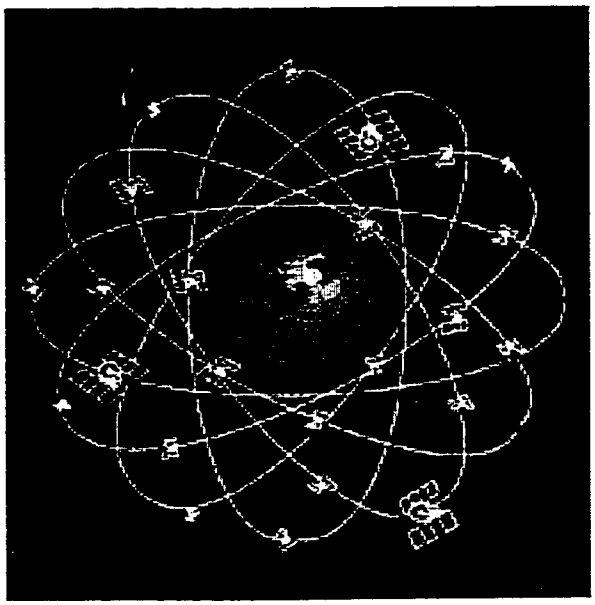


Figure 2.1 GPS orbital configuration.

Orbital altitudes are ≈20200 km, with periods of ≈12 hours, such that two repeated ground traces are produced per sidereal day. Each satellite is equipped with highly accurate timing standards derived from rubidium and cesium clocks (stabilities of one part in 10^{13} [*Bowen et al.*, 1986]). The development of such timing accuracies allows precise synchronization of the satellite clocks, and autonomy of satellite operations (where satellite timing need not be controlled continuously by the ground control segment).

2.1.2 The Control Segment

This segment consists of several monitoring stations distributed globally, with a master control station near Colorado Springs [cf. *Leick*, 1995]. Kepler parameters describing the orbital motion in an inertial reference frame are computed at the ground control stations, which track the satellites. These orbital parameters are uploaded to the

satellites and transmitted to users (in the *broadcast ephemerides*) via modulation of the GPS carrier signal (the *navigation message*). The control segment is also responsible for calibrating the satellite clocks and computing satellite clock corrections. These corrections are also provided to users via the navigation message, which is updated and uploaded to the satellites periodically (approximately once per day [*Zumberge and Bertiger*, 1996]).

The control segment is also responsible for establishing *GPS Time*, which is defined as the number of seconds elapsed from Saturday midnight of the present week [cf. *Kaplan*, 1996]. GPS Time is synchronized with UTC (Universal Time Coordinated) at the microsecond level, within an integer number of seconds (modulo 1 s). The integer shift between seconds GPS and seconds UTC arises from the continuity of GPS Time, which has not been adjusted for leap seconds since implementation in January, 1980.

2.1.3 The User Segment

The user is responsible for computing local position/navigation solutions. By measuring ranges to four satellites simultaneously, a unique position solution (in 3-dimensions) can be computed, in addition to the receiver clock offset. The GPS system was designed to provide users with redundant measurements, however, where 6-11 satellites are typically in view [*Spilker*, 1996a]. In most cases, an overdetermined position solution is computed, which minimizes the effect of random range errors. Ranges to individual satellites are calculated in the receiver using a correlation technique, where a replica of the satellite signal is generated and matched to that received from the satellite. Combined with knowledge of satellite positions, derived from the broadcast ephemerides, a position solution can be computed using the method of least squares. Satellite positions are computed in cartesian coordinates in the Earth-centred Earth-fixed WGS-84 reference

frame [*Defense Mapping Agency*, 1984], such that a user's positions can be referenced to the WGS-84 reference ellipsoid in latitude, longitude and height.

2.2 GPS SIGNAL STRUCTURE

The GPS signal is broadcast on two frequencies: L1 = 1575.42 MHz and L2 = 1227.6 MHz. The L2 signal is encrypted and intended primarily for military users. By using squaring (or similar) signal processing techniques in the receiver, however, the L2 signal can be recovered for civilian use [cf. *Hoffmann-Wellenhof*, 1994]. Each signal consists of a continuous wave carrier signal modulated with coded sequences of RF energy. In order to differentiate between the various satellite signals, a unique *pseudo-random noise* (PRN) code is associated with each satellite. These codes consist of a digital sequence of random bits (zeroes and ones), generated by a shift register, which repeats periodically (hence the *pseudo*-randomness) [*Spilker*, 1996b]. These codes were first developed for use in the Air Force 621B project (1972) and are an essential element of GPS operations.

This *code division multiple access* (CDMA) technique allows all satellites to broadcast on the same frequency, and receivers are able to uniquely identify individual satellites, by correlating received signals with replicas of the code sequence (the cross-correlation between different satellite codes being extremely poor). Modulation of the carrier signals with these codes results in a spread spectrum signal, which is resistant to jamming. Two types of PRN codes are used: the coarse acquisition code (C/A) intended primarily for civilian use, and the precise (P) code for military users [cf. *Parkinson et al.*, 1995]. Elements of the two GPS signals (on L1 and L2) are given in the following equations (where carrier amplitudes have been neglected for simplicity).

$$L1(t) = P(t)D(t)\cos(f_1 t) + CA(t)D(t)\sin(f_1 t)$$
$$L2(t) = P(t)D(t)\cos(f_2 t) \qquad (2.1)$$

where
$\cos(f_1 t)$, $\cos(f_2 t)$ = L1, L2 carrier waves
CA(t), P(t) = C/A (coarse acquisition), P (precise) codes
D(t) = navigation message

From these signals, two types of positioning services are available: the standard positioning service (SPS) using the C/A code, and the precise positioning service (PPS) using the P code [cf. *Leick*, 1995]. Differences in these services arise from the higher frequency used for the P code (10.23 MHz) versus the C/A code (1.023 MHz). The higher modulation bandwidth allows a reduction of noise in the received P code signal.

2.3 GPS OBSERVABLES

As stated in Section 2.2, GPS receivers compute satellite ranges using correlation techniques, where received signals are compared with replicas generated in the receiver. Numerous steps are involved in the signal processing, including various filtering and intermediate frequency algorithms. For a detailed discussion of GPS receiver design and signal processing, see *Van Dierendonck* [1996].

Initially, correlations of the PRN code are used to generate estimates of the time delay between signal transmission at the satellite to reception at the receiver. Multiplication by the speed of light results in derivation of a *pseudorange*. The received signal is then demodulated, resulting in recovery of the underlying carrier signal. The received phase is measured against the receiver-generated phase, and a phase offset is computed in fraction of cycles (which can be converted to a *carrier phase range* through

multiplication by the carrier signal wavelength). This observable is more precise than the pseudorange but is ambiguous by an integer number of cycles. A Doppler measurement (range-rate) is also derived from estimates of the received signal's frequency.

2.4 GPS ERROR SOURCES

Inherent in GPS range measurements are several errors, resulting from a variety of sources (Figure 2.2). Errors can arise from inaccuracies in estimates of satellite position and satellite clock corrections (broadcast ephemerides), tropospheric/ionospheric effects along the signal propagation path, and receiver noise generated though signal processing errors. These effects are included in the following equations describing the pseudorange (PR) and carrier phase range (Φ) observables (after *Lachapelle* [1995]), where all quantities are in units of distance:

$$PR = \rho + d\rho + cdt - cdT + io + trop + \varepsilon_{PR} \tag{2.2}$$

$$\Phi = \rho + d\rho + cdt - cdT - io + trop + \varepsilon_{\Phi} + \lambda N \tag{2.3}$$

where

ρ = geometric range

$d\rho$ = orbital errors

dt = satellite clock error

dT = receiver clock error

io = ionosphere error

trop = troposphere error

λ = carrier wavelength

N = carrier phase integer ambiguity (in number of cycles)

ε_{PR}, ε_{Φ} = pseudorange and carrier phase range noise and multipath, respectively

Estimates of satellite positions and clock corrections are controlled by the GPS ground control segment and are dependent on the length of prediction interval. These errors are on the order of 2-4 m [*Zumberge and Bertiger*, 1996]. The tropospheric effects can be reduced through modelling [i.e. *Hopfield*, 1969], while ionospheric errors can be estimated from a dual frequency receiver. Derivation of the ionospheric effect is discussed in detail in Section 2.6.1.

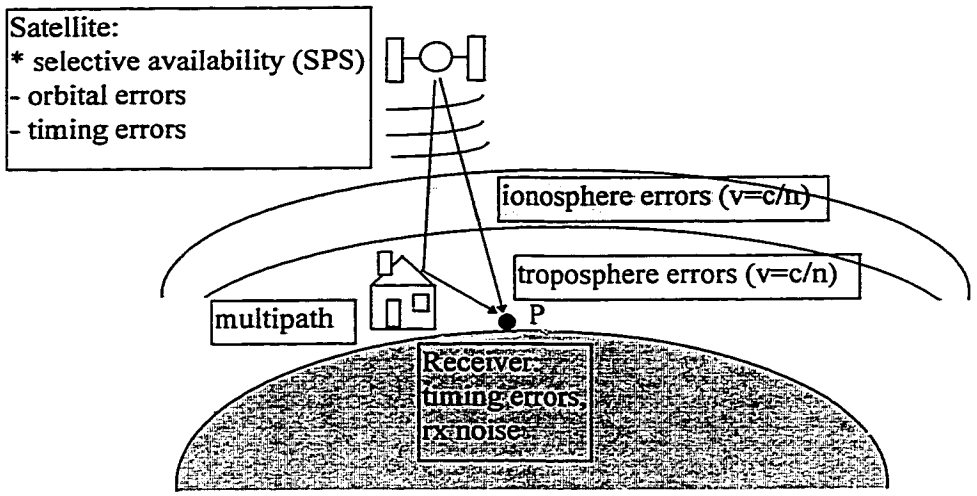


Figure 2.2 Illustration of contributing error sources in GPS range observations.

A further error source is multipath, which results from the reflection of signals from surfaces in the antenna environment. Multipath can be both specular (with predominant frequencies dependent on the distance from reflector to antenna) and diffuse (random scattering from irregular surfaces). These errors depend on the given signal frequencies, with a maximum phase multipath of approximately one-quarter of the carrier phase wavelength (≈5-6 cm for L1 and L2) [cf. *Hoffmann-Wellenhof*, 1994]. Maximum code multipath is on the order of 1.0-1.5 times the code length (≈30 m P code, ≈300 m C/A code) [cf. *Braasch*, 1996]. Typical C/A code multipath errors are observed in the range 2-4 m [i.e. *Parkinson*, 1996]. These errors can be limited by the use of an RF-

absorbing ground plane or choke-ring at the antenna; receiver processing technologies have also been developed to mitigate multipath effects [*Fenton et al.*, 1991].

The largest source of ranging error results from an intentional degradation of ephemerides and manipulation of the satellite clock (dithering), referred to as *Selective Availability* (SA) [cf. *Van Graas and Braasch*, 1996]. These errors are imposed by the U.S. miltary for security reasons, in order to deny users of the SPS full precise positioning capabilities. It was intended initially, in designing the GPS system, that position results using the C/A code would have accuracies be on the order of 400 m. In practical implementation, however, accuracies of 15-40 m were obtained [cf. *Hoffmann-Wellenhof*, 1994], necessitating the addition of SA. The U.S. DoD specifies stand alone point position accuracies of 100 m (horizontal) and 156 m (vertical) at 95% probability. Recent intiatives have been taken by the United States to phase out SA by approximately the year 2006 [cf. *Kovach and Van Dyke*, 1997].

Table 2.1 Typical Range Errors (SPS)

Type of error	**Magnitude 1σ (m)**
SA	32.3
ionosphere	5.0
troposphere	1.5
SV clock and orbit	5.0
receiver noise	1.5
multipath	2.5
total error	**≈33.0**

A summary of typical uncorrelated range errors (C/A code) is given in Table 2.1 (after *Kaplan* [1996]). These errors propagate into the position domain by a *position dilution of precision* (PDOP) factor dependent on satellite geometry [cf. *Spilker*, 1996a]. This factor essentially describes the volume of a tetrahedron defined by the receiver and satellite positions. A smaller volume (i.e. all satellites grouped closely together) results in

a larger DOP factor and a larger position error. Estimates of position accuracies can be made under the assumption that the satellite range errors are uncorrelated, such that the position accuracy is equal to the total range error multiplied by the dilution of precision factor. Typical PDOPs range from 2 to 4. For the range error in Table 2.1, this implies a position accuracy of approximately 100 m.

2.5 WIDE AREA DIFFERENTIAL GPS

From the discussion of error sources in Section 2.4, it is observed that several error sources, particularly SA, are satellite dependent. If the coordinates of a receiver are known (designated the reference or monitor station), range corrections can be computed from the observed pseudoranges to each satellite. These can be sent to a remote receiver, via a telemetry link, where the range corrections are applied to the remote range observations. This technique forms the basis of *differential GPS* (DGPS), as pictured in Figure 2.3.

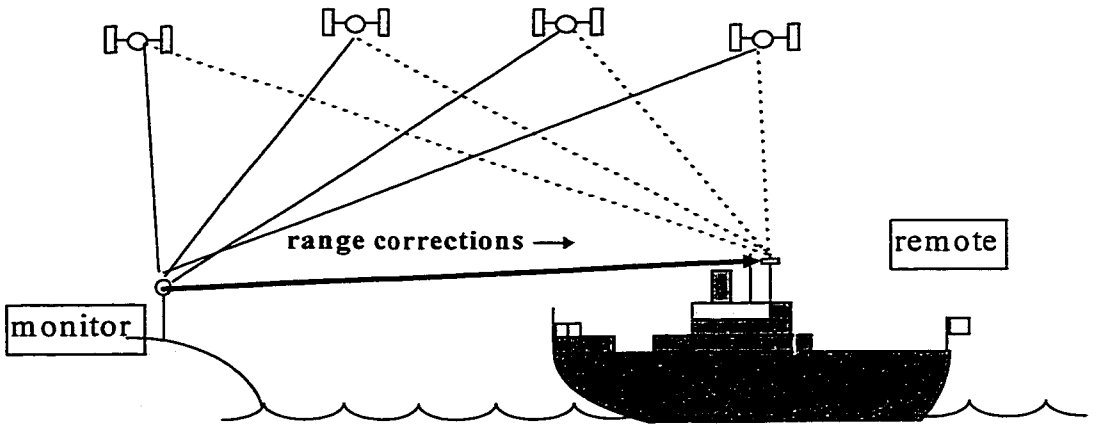


Figure 2.3 Illustration of DGPS. Range corrections are calculated for each satellite in view from the monitor receiver and transmitted to the remote.

In DGPS, both atmospheric and satellite-dependent errors are reduced significantly, allowing for position accuracies on the order of 1-10 m [cf. *Hoffmann-Wellenhof*, 1994], which are a factor of 10 times better than maximum stand alone point position accuracies (Section 2.4). These accuracies depend on the separation distance between the two receivers, with a degradation in positioning accuracies resulting from spatial decorrelation of atmospheric effects and differences in the projection angle of satellite orbital errors along the two lines-of-sight. The satellite clock error does not decorrelate over distance, but varies rapidly in time (the range error acceleration being as large as 0.019 m/s^2 [*DoD*, 1995]).

The DGPS range corrections are scalar quantities, incorporating all error sources in a single correction value. In *wide area differential GPS* (WADGPS), observations from a network of reference stations are used to model the individual components of range error over a region. By incorporating simultaneous measurents from the various receivers, satellite clock and orbital errors (including SA) are modelled over time, and transmitted to users in a vector consisting of the satellite clock error and a 3-dimensional orbital error. By equipping the reference stations with dual frequency receivers, measurements of the ionosphere range delay (or equivalently total electron content) along all satellite-receiver lines-of-sight can be used to model the ionospheric effect (see Section 2.6). Parameters describing the spatial distribution of ionospheric TEC are also supplied to the user. The wide area corrections are sent to users via geostationary communications satellites [i.e. *Loddo et al.*, 1996].

The three separate wide area corrections (clock, orbital, ionosphere) can be applied anywhere in the network area, using a local transformation, allowing users to achieve position accuracies on the order of 1-4 m [*Skone et al.*, 1997; *Kee*, 1996]. Due to the vector nature of the error components, there is no degradation of position accuracies with increasing distance from reference stations. Several wide area networks have been

developed worldwide: the Natural Resources Canada GPS•C corrections service [*Caissy et al.*, 1996], the European Geostationary Overlay Service (EGNOS) [*Nieto et al.*, 1997], and the SATLOC system in the United States [*Bertiger et al.*, 1997]. In the past few years, NASA has taken initiatives to develop a worldwide system, incorporating observations from up to 160 stations distributed globally [*Yunck et al.*, 1996].

An extension of WADGPS is the wide area augmentation system (WAAS). The WAAS architecture and operational concept were designed under FAA direction, to support enroute air navigation and precision approach [cf. *Enge and Van Dierendonck*, 1996]. Initially intended for operation in the Continental United States and Alaska, elements of the WAAS architecture are being adopted worldwide. Efforts are currently underway to assess the feasibility of extending WAAS applications in Canada by 2010 (CWAAS - *Nav Canada newsletter* [1998]). The WAAS is a safety critical system, with requirements that the vertical position error be bounded by the *vertical protection level* (VPL), with probability better than 0.9999999 [*FAA*, 1997]. For system availability, the VPL should be smaller than or equal to a *vertical alert limit* (VAL) of 15 m 99.9 percent of the time. The VPL is a 5.33σ statistical value. The vertical position standard deviation should therefore be less than or equal to 2.81 m (=15/5.33). Implicit in this error bound are accuracy requirements for the various range correction components (clock, orbital, ionosphere). Relevant to the research in this thesis are the ionosphere error bounds, which are provided in Chapter 3. While these error bounds are currently applicable for WAAS operation in CONUS and Alaska, it is likely that similar error bounds will be adopted by extended WAAS applications in Canada and Europe. For this reason, these error bounds are referred to throughout the analyses of ionosphere modelling accuracies in several chapters of this thesis.

2.6 IONOSPHERIC EFFECTS ON GPS SIGNALS

2.6.1 Derivation of Ionospheric Range Error

As discussed in Section 2.3, a GPS receiver measures two types of observables: carrier phase in number of cycles, and code group delay in seconds. These measurements are converted to carrier phase range and pseudorange measurements, respectively, through multiplication by the wavelength and speed of light in a vacuum. Implicit in GPS range measurements, therefore, is an assumption that the GPS signal travels at the speed of light (and the wavelength of the signal is equal to the wavelength in a vacuum) along its entire path length. This is equivalent to assuming that the phase and group indices of refraction, n_p and n_g, are equal to one throughout the propagation medium, where the carrier phase range Φ and pseudorange PR are calculated as follows:

$$\Phi = \int_{path} \lambda dN = \int_{path} \frac{\lambda_0}{n_p} dN = \int_{path} \frac{\lambda_0}{n_p} d\left(\frac{ct}{\lambda_0}\right) = \int_{path} \frac{c}{n_p} dt \qquad \text{(metres)} \qquad (2.4)$$

$$PR = \int_{path} v_g dt = \int_{path} \frac{c}{n_g} dt \qquad \text{(metres)} \qquad (2.5)$$

In this equation, v_g is the group velocity, c is the speed of light in a vacuum, λ is the true wavelength, λ_0 is the wavelength in a vacuum, dN denotes the differential number of cycles, and dt denotes the differential element of time. Φ and PR are both measured in units of length (nominally metres).

The assumption $n_g = n_p = 1$ is incorrect in regions such as the troposphere and ionosphere, where the index of refraction may differ significantly from one. In the ionosphere, the index of refraction depends on several factors. Specifically, the phase

index of refraction is given by the Appleton-Hartree formula [*Rishbeth and Garriott*, 1969]:

$$n_p^2 = 1 - \frac{X(1-X)}{(1-X) - \frac{1}{2}(Y\sin\theta)^2 \pm \left[\frac{1}{4}(Y\sin\theta)^4 + (1-X)^2(Y\cos\theta)^2\right]^{\frac{1}{2}}} \tag{2.6}$$

where θ is the angle between the magnetic field and the direction of signal propagation, and

$$X = \frac{\omega_p^2}{\omega^2}, \qquad Y = \frac{\omega_c}{\omega}, \qquad \omega_p^2 = \frac{Ne^2}{m\varepsilon_0}, \quad \text{and} \quad \omega_c = \frac{Be}{m} \tag{2.7}$$

where

N =electron density (el/m^3)

e = electron charge (C)

m = electron mass (kg)

ε_0 = permittivity of free space = 8.85×10^{-12} (farads/m)

B = ambient magnetic field (T)

The values ω_p and ω_c are referred to as the electron plasma and cyclotron frequencies, respectively. These two frequencies describe electron oscillations in the presence of a perturbation electric field, and the ambient magnetic field, respectively. Approximate estimates of these values are made under the assumptions

$N = 10^{12}$ el/m^3	(upper bound ionospheric electron density)
B = 32000 nT	(dipolar magnetic field at ionospheric altitudes)

such that $\omega_p \approx 9$ MHz and $\omega_c \approx 1$ MHz, and $X \approx 4 \times 10^{-5}$, $Y \approx 7 \times 10^{-4}$, for GPS signals ($\approx$1500 MHz). Note that the index of refraction is complex for signals where $X > 1$. In this case the ionosphere is "opaque", such that the signal is damped and unable to propagate. For the estimated values of X and Y, Equation 2.6 can be expanded and higher order terms ignored, such that (after *Brunner and Gu* [1991])

$$n_p = 1 - \frac{X}{2} - \frac{X^2}{8} \pm \frac{XY}{2}\cos\theta \tag{2.8}$$

The $\pm$ in Equations 2.6 and 2.8 derives from *ordinary* and *extraordinary* propagation modes of the electromagnetic wave, as it travels in an ambient magnetic field. For maximum solar conditions, *Brunner and Gu* [1991] have determined that the higher order terms in Equation 2.8 only contribute 1-2 mm of range error for GPS users. The phase index of refraction can therefore be approximated by the formula:

$$n_p \cong 1 - \frac{X}{2} = 1 - \frac{1}{2}\frac{\omega_p^2}{\omega^2} = 1 - \frac{1}{2}\frac{(2\pi)^2\left(\frac{Ne^2}{m\varepsilon_0}\right)}{(2\pi)^2 f^2} \tag{2.9}$$

which can be further simpified by substituting values for the constants e, m, and ε_0

$$n_p = 1 - 40.3\frac{N}{f^2} \tag{2.10}$$

such that the phase index of refraction depends only on the electron density N. A corresponding expression for the group index of refraction can be derived from the relation

$$n_g = n_p + f\frac{dn_p}{df} \tag{2.11}$$

such that

$$n_g = 1 - 40.3\frac{N}{f^2} + 80.6f\frac{N}{f^3} = 1 + 40.3\frac{N}{f^2} \tag{2.12}$$

and the phase and group velocities are given by

$$v_p = \frac{c}{1 - 40.3\frac{N}{f^2}} \approx c\left(1 + 40.3\frac{N}{f^2}\right), \qquad v_g = \frac{c}{1 + 40.3\frac{N}{f^2}} \approx c\left(1 - 40.3\frac{N}{f^2}\right) \tag{2.13}$$

The phase and group velocities differ from c by the same magnitude. The phase velocity exceeds c, while the group velocity (which carries the wave energy) is always less than c. The GPS range errors associated with ionospheric effects can therefore be derived by expanding Equations 2.4 and 2.5, and using the expressions in Equation 2.13, as follows

$$\Phi = \int_{path} v_p dt = \int_{path} \frac{c}{n_p} dt = \int_{path} c dt + \frac{40.3c}{f^2}\int_{path} N dt = \int_{path} ds - \frac{40.3c}{f^2}\int_{path} N ds \tag{2.14}$$

$$PR = \int_{path} v_g dt = \int_{path} \frac{c}{n_g} dt = \int_{path} c dt - \frac{40.3c}{f^2}\int_{path} N dt = \int_{path} ds + \frac{40.3c}{f^2}\int_{path} N ds \tag{2.15}$$

where integration over time has been transformed into an integration along the path length ds, using $dt = \frac{1}{v} ds$.

The second integral in Equations 2.14 and 2.15 corresponds to the range error associated with GPS signal propagation through the ionosphere. The pseudorange and carrier phase measurements therefore include range errors of equal magnitude and opposite sign. These correspond to a *group delay* and a *phase advance*. The second integral in Equations 2.14 and 2.15 can be evaluated by integrating the electron density along a 1 m^2 column along the signal path, such that the ionosphere range delay is

$$\frac{40.3}{f^2}\int_{path} Nds = \frac{40.3}{f^2}TEC \tag{2.16}$$

where TEC represents the total electron content along the signal path. The dispersive nature of the ionosphere ($dn/df \neq 0$) therefore allows calculation of the absolute TEC, if range measurements are available on two separate frequencies:

$$TEC = \frac{1}{40.3}\left(\frac{1}{f_1^2}-\frac{1}{f_2^2}\right)^{-1}\left(PR_1 - PR_2\right) \quad (el/m^2) \tag{2.17a}$$

The relative TEC can be derived from the phase measurements:

$$TEC = \frac{-1}{40.3}\left(\frac{1}{f_1^2}-\frac{1}{f_2^2}\right)^{-1}\left(\Phi_1 - \Phi_2\right) \quad (el/m^2) \tag{2.17b}$$

Equation 2.17 is the basic calculation used to derive TEC observations from dual frequency GPS receivers. TEC values are often expressed in terms of *total electron content units* (TECU), where one TECU equals 10^{16} el/m^2. TEC is also easily converted to ionosphere range delay for the L1 and L2 frequencies, using Equation 2.16, where

1) For L1, 1 TECU = 0.16 metres range delay

2) For L2, 1 TECU = 0.27 metres range delay

2.6.2 Practical Limitations in Deriving Ionosphere TEC

Precise absolute estimates in TEC, from formula 2.17, are limited by errors in the pseudorange and carrier phase range observables (cf. Section 2.4). While many common errors are eliminated in forming the linear combinations of PR_1-PR_2 and Φ_1-Φ_2 (dρ, cdt, cdT, trop), uncorrelated "noise" errors remain. Such errors include receiver noise and multipath effects. Calculation of the ionospheric effect therefore includes these uncorrelated error sources, such that phase-derived ionosphere observables (Equation 2.17b) have accuracies a factor of 10 or more better than the pseudorange-derived values. Clearly, the phase-derived estimate of TEC is preferable. As stated in Section 2.3, however, the carrier phase measurements are ambiguous, such that Equation 2.17b can be used to derive only relative variations in TEC. An optimal estimate of TEC is derived by exploiting both the absolute nature of the pseudorange-derived TEC and the "smooth" phase-derived TEC (see Chapter 3).

Additionally, errors in forming the linear combinations of PR_1-PR_2 and Φ_1-Φ_2 result from the presence of interchannel biases (timing offsets) between the L1 and L2 channels, in both the satellite and receiver. These biases are relatively constant [cf. *Wilson and Mannucci*, 1993] and result in inaccuracies in determining the absolute TEC; such biases must be solved for in wide area ionosphere modelling algorithms.

2.6.3 The Earth's Ionosphere

The ionosphere is defined as the part of the Earth's atmosphere where ions and electrons are present in quantities sufficient to affect propagation of radio waves [cf. *Rishbeth and Garriott*, 1969]. This region extends from approximately 60 km to more than 1000 km altitude. Electrons and ions are produced, primarily through absorption of solar ultraviolet radiation (UV) and x-ray radiation, which photoionizes and photodissociates atmospheric neutrals (Ar, He, N_2, O, O_2). Predominant production mechanisms depend on height, with the more intense x-ray radiation penetrating to lower altitudes. In high latitude regions, ions and electrons are also produced through collisions between energetic precipitating electrons and the neutral molecules.

From a simplified model of the ionization rate (constant absorption coefficient, single species atmosphere), *Chapman* [1940] derived a simple expression for the variation of ionized densities with height. Densities were found to increase exponentially with height, with peak heights being dependent on density profiles for the respective ionizable species.

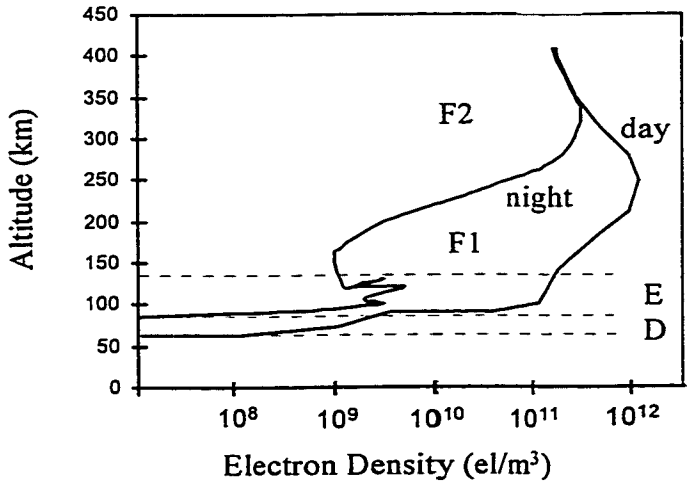


Figure 2.4 Ionospheric electron density profile, as a function of altitude.

The most important ionospheric constituent for GPS users is the free electrons, the effects of which were described in Section 2.6.1. The electron density profile is shown in Figure 2.4, where different regions of the ionosphere (D, E, F1 and F2) are identified. The nature of ionospheric processes/effects varies between these regions. These effects are described in *Rishbeth and Garriott* [1969]. The D region (60-90 km) is characterized by absorption of radio waves (low VHF), and is of little concern for GPS users. The E region (90-120 km) is characterized by lower electron densities than the F region, with the presence of irregularities at high latitudes. These irregularities, caused by energetic electron precipitation, are a feature of the auroral zone, and are described in more detail in Chapter 4. Higher electron densities are found in the F region (120-1000 km), with the largest values observed at 250 (350) km altitude in the dayside (nightside) sectors. These heights correspond to the F2 layer, which is the predominant source of absolute ranging errors for GPS signals.

Since photoionization and photodissociation are controlled by solar radiation, the pattern of free electrons is quasi-fixed with respect to the Sun, such that the ionospheric total electron content (electron density integrated along a 1 m^2 vertical column) will vary diurnally in an Earth-fixed reference frame. This is illustrated in Figure 2.5, where the peak total electron content (TEC) is observed at 1400 magnetic local time (MLT). This corresponds to the maximum number of electrons produced versus those lost in recombination. Minimum values of TEC are observed nightside (1800 to 0600 MLT).

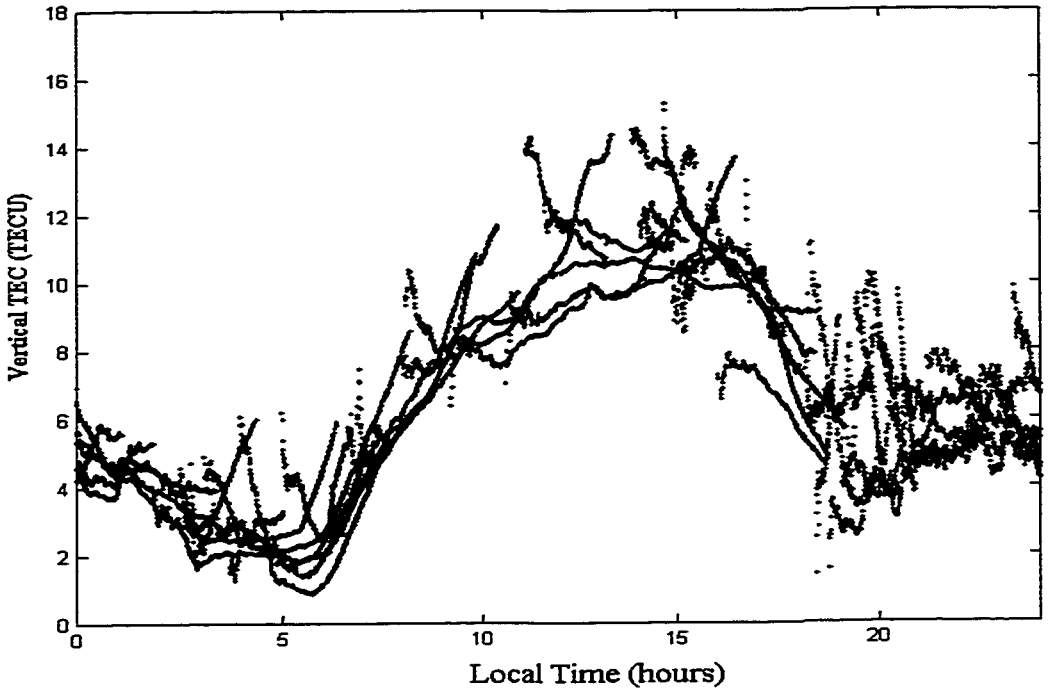


Figure 2.5 Diurnal variation of vertical TEC, as compiled from observations at reference station Ottawa on October 18, 1996.
The unit TECU is defined as 10^{16} el/m^2.

2.6.4 Solar Cycle

Solar sunspot activity undergoes periodic variations over several time scales, the principal cycle having an 11 year period, on average. A *solar maximum* was last observed in 1989, with the next solar maximum being predicted for the year 2000-2001 [*Kunches*, 1997]. Sunspot activity is directly correlated with the emission of solar electromagnetic radiation and the creation of solar flares, two phenomena which control the level of ionization in the Earth's ionosphere. Ionospheric electron density profiles therefore exhibit a dependence on the solar cycle, and absolute values of total electron content (TEC) have been observed to increase by a factor of 3 for solar maximum versus solar

minimum conditions [*Klobuchar et al.*, 1995]. For example, the values of vertical TEC in Figure 2.5, which are in the range 1 to 15 TECU, were derived for a solar minimum year (1996). At solar maximum, these values will increase to 3-45 TECU, with corresponding increases in the dawn and dusk large-scale gradients. This is equivalent to approximately 0.5 - 9.0 m vertical ionospheric delay, at the GPS L1 frequency.

In addition to the large-scale global increase in TEC during solar maximum, an increase in the frequency and magnitude of magnetic storms accompanies the enhanced solar flare activity. The number of "major" magnetic storms increases by a factor of 6 (18 days compared with 3) for solar maximum, versus solar minimum, conditions [*Kunches*, 1997]. A major storm is defined by NOAA as that during which Ap > 40 [cf. *NOAA/NGDC web site*, http://www.ngdc.noaa.gov/stp/GEOMAG/apstar.html], where Ap is the planetary A index (derived from a daily average of the Kp indices - see Section 5.2).

CHAPTER 3

THE WIDE AREA IONOSPHERIC GRID MODEL

For the wide area network (Section 2.5), multiple reference stations allow the separation and estimation of various GPS error sources. In this thesis, the focus is on estimation of the ionospheric total electron content, using an ionospheric grid algorithm. This chapter introduces the grid modelling concept, with details of the model specifications, observed accuracies, mathematical formulation, and practical implementation. It is intended that these discussions provide an overall description of the grid model. Several of these topics, particularly the model formulation, are discussed in more detail in later chapters, where specific aspects of the model are investigated.

3.1 WIDE AREA GRID CONCEPT

As outlined in Section 2.6, estimates of TEC (which are directly related to GPS range errors) can be derived from dual frequency GPS observations, using Equation 2.17. If all observations from a wide area network of reference stations are combined, the TEC can be modelled over the network region. An overview of the modelling of TEC at designated grid points is outlined as follows.

Slant TEC values are calculated for each satellite-receiver line-of-sight in the network of reference stations. These TEC observations are depicted in Figure 3.1, where each slant TEC observation depends on the satellite elevation angle E.

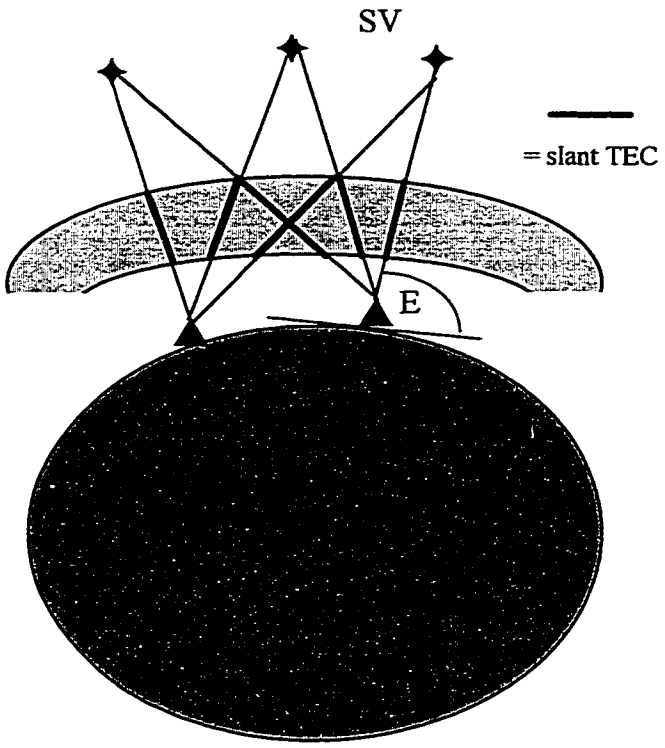


Figure 3.1 Slant TEC observations, for two reference stations and three satellites in the wide area network.

In order to derive a consistent set of TEC observations, it is necessary to eliminate the elevation dependence, mapping the slant TEC observations to vertical. In doing so, it is assumed that the majority of the TEC is concentrated in an infinitessimally thin shell at a fixed altitude. This assumption is based on the electron density profiles in Figure 2.4, where peak electron densities are concentrated at altitudes near 300 km. The mapping of slant TEC values to vertical at discrete pierce points on this shell is depicted in Figure 3.2. An observation *ionosphere pierce point* (IPP) is defined as the point at which a given satellite-receiver line-of-sight pierces the ionosphere shell. Once all observations have been mapped to vertical, estimation of the vertical TEC (VTEC) is reduced to a two-dimensional adjustment in latitude and longitude, where VTEC is estimated at designated grid points in the ionosphere shell.

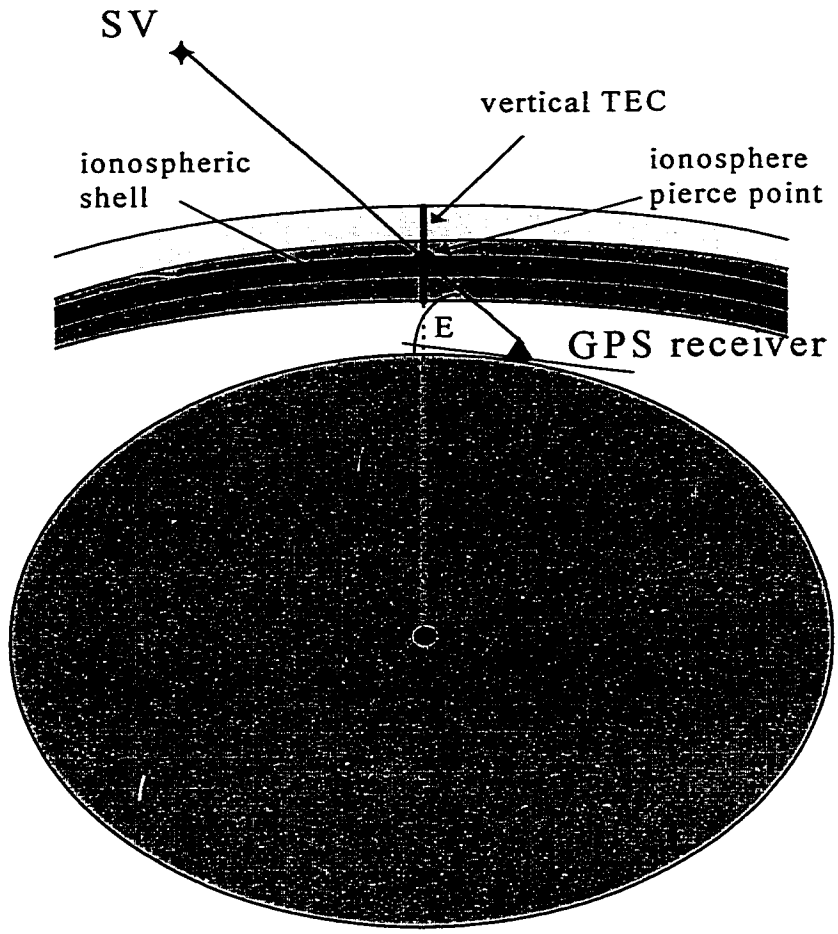


Figure 3.2 Mapping of slant TEC to vertical in the ionosphere shell.

The grid VTEC values are updated periodically and transmitted to remote receivers, where an interpolation algorithm is employed to estimate VTEC values for local GPS range observations. Remote users are then able to correct observed ranges for ionospheric effects, allowing an improvement in positioning accuracies.

3. 2 MODEL SPECIFICATIONS AND ACCURACIES

For the proposed Wide Area Augmentation Sytem (WAAS) in the United States and Alaska, original FAA specifications included a grid ionosphere vertical error (GIVE) ≤ 2 m 99.9% of the time [*FAA*, 1994]. This GIVE requirement has recently been dropped from FAA WAAS specifications, however. An implicit ionosphere accuracy requirement is now included in the vertical protection level (VPL) [*FAA*, 1997], which is calculated as $5.33\sigma_v$, where σ_v is the vertical position error standard deviation. The VPL is bounded by a vertical alert limit (VAL) of 15 m 99.9% of the time, which translates into σ_v = 15.0/5.33 $\approx$ 2.81 metres. The value σ_v accounts for all error sources, including the ionosphere. In the range domain, a goal of GIVE $\leq$ 1.5 m 99.9% of the time could be useful in satisfying overall accuracy requirements [*M.B. El-Arini*, private communications]. WAAS specifications also include an ionosphere message update interval of 5 minutes or less [*FAA*, 1997]. Proposed WAAS grid spacings are 5 degrees (square grid) at latitudes less than 55 degrees. For higher latitudes, grid spacings of 10 degrees are proposed [*FAA*, 1997].

To date, many ionospheric correction algorithms have been proposed and tested to meet this specification. A common approach is to model ionospheric TEC values at the fixed grid points in real-time, using a Kalman filter technique. Estimates are made of satellite and receiver L1-L2 interchannel biases, which must be included in ionospheric delay calculations (Section 2.6.2), in addition to parameters characterizing the vertical TEC [*Mannucci et al.*, 1993; *Mannucci et al.*, 1995]. Model parameters are generally approximated as Gauss-Markov or random walk stochastic processes, such that updated parameter values are correlated with estimates from earlier epochs. Typical slant delay accuracies of 25-50 cm rms (L1) [*Mannucci et al.*, 1995; *Kee*, 1996; *Skone et al.*, 1997], and vertical delay accuracies of 15-20 cm rms [*Bertiger et al.*, 1997; *Yunck et al.*, 1996; *Wilson et al.*, 1997], have been achieved under solar minimum conditions.

Recent research has focused on evaluating the performance of WADGPS ionospheric grid algorithms during periods of geomagnetic disturbances, in anticipation of the upcoming solar maximum (Section 2.6.4). Such studies have included investigations of the spatial and temporal correlation of ionospheric delays, under various levels of geomagnetic activity [*Klobuchar et al.*, 1995; *Chavin*, 1997], as well as the accuracy of ionospheric corrections derived using grid TEC values [cf. *Draganov et al.*, 1996; *Conker et al.*, 1997; *Mannucci et al.*, 1997]. This research has been conducted primarily using observations from the Continental United States (CONUS) region. Reported slant delay accuracies range from .40 m to 1.0 m rms. It is anticipated that ionospheric grid accuracies may be degraded by a factor of 2-3 [cf. *Bertiger et al.*, 1997] for large magnetic storms.

3.3 MODEL FORMULATION

Since only a limited number of simultaneous observations are available at a given epoch, the model must incorporate observations from varying positions in time and space. While several methods exist for deriving estimates of unknowns from such an observation set, a technique commonly used in existing wide area grid models is the Kalman filter. In this approach, optimal estimates of unknown parameters are derived from *a priori* assumptions about system processes and error statistics. Such an approach is employed in this thesis, and a brief derivation of the Kalman filter equations (with implicit assumptions) is included in Appendix A.

The Kalman filter employs two models: a measurement model describing the functional relationship between observations and unknowns; and a system model describing temporal variations of the unknowns. Prior to filter updates, the GPS observations are pre-processed to produce smoothed observations of slant TEC.

3.3.1 Smoothing TEC Observations

At a given epoch, estimates of the slant TEC are calculated for all reference station-satellite observations using Equation 2.17:

$$\text{absolute slant TEC} = f_1^2 f_2^2 (PR_1 - PR_2)/[(f_2^2 - f_1^2) \times 40.30] = TEC_{PR} \quad (3.1)$$

$$\text{relative slant TEC} = -f_1^2 f_2^2 (\Phi_1 - \Phi_2)/[(f_2^2 - f_1^2) \times 40.30] = TEC_{\Phi} \quad (3.2)$$

where

f_1, f_2 = 1575.42, 1227.60 MHz

PR_1, PR_2 = L1 and L2 pseudoranges (in metres)

Φ_1, Φ_2 = L1 and L2 carrier phase ranges (in metres)

The absolute slant TEC values are then smoothed using a technique similar to the method described in *Mannucci et al.* [1993], where pseudorange and carrier phase estimates of TEC are combined in a "levelling" process. *Eueler et al.* [1991] and *Goad* [1990] also describe the implementation of a Kalman filter approach to obtaining smoothed ionosphere delay estimates. As mentioned in Section 2.6.2, the absolute TEC estimates derived from pseudorange observations have much larger errors than the (ambiguous) phase-derived estimates. Since the ionosphere group and phase delays are equal in magnitude and opposite in sign, the precise carrier phase-derived estimates may be used to "smooth" the noisier pseudorange-derived values. In this method, it is assumed that the difference TEC_{PR} - TEC_{Φ} is a constant over time (in the absence of cycle slips), and equal to a linear combination of the carrier phase ambiguities. The constant offset between TEC_{PR} and TEC_{Φ} is determined from the following equation:

$$\begin{aligned} \text{offset}(t_j) &= \frac{1}{N}\sum_{i=1}^{j}\left(TEC_{PR}(t_i) - TEC_{\Phi}(t_i)\right) \\ &= \frac{1}{N}\left[(N-1)\,\text{offset}(t_{j-1}) + \left(TEC_{PR}(t_j) - TEC_{\Phi}(t_j)\right)\right] \end{aligned} \quad (3.3)$$

where the second expression represents a recursive relation, and the number of observations N increases over time. The smoothed value of TEC is then calculated as

$$TEC(t_j) = TEC_{\Phi}(t_j) + offset(t_j) \tag{3.4}$$

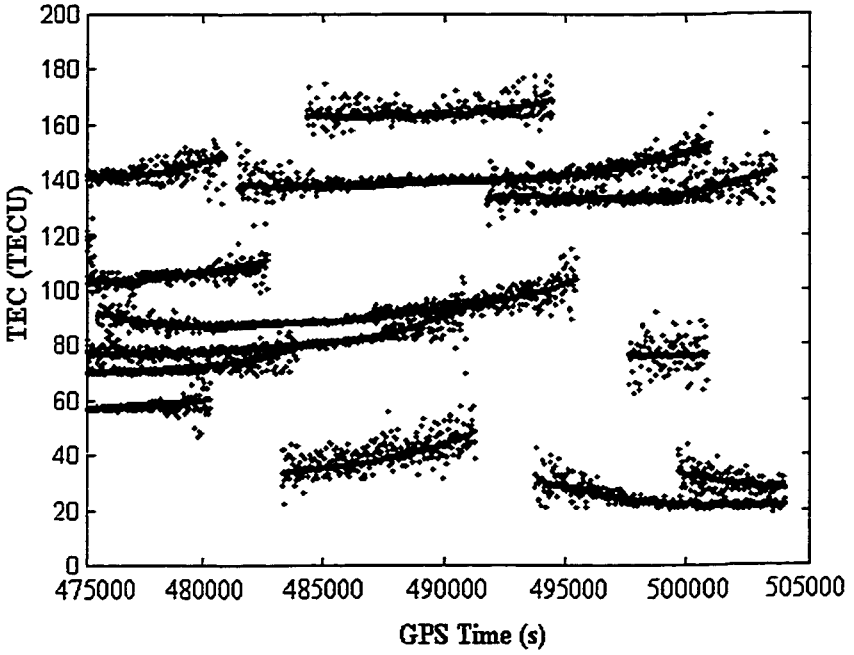


Figure 3.3 Smoothed TEC series, superposed on pseudorange-derived TEC values. Series are offset for comparison purposes.

Figure 3.3 illustrates the TEC smoothing algorithm, for various satellites, over an interval of approximately 8 hours. The smoothing algorithm generally converges after several minutes. For cases where cycle slips are detected, the algorithm is reinitiated.

3.3.2 Measurement Model

Each vertical TEC observation at a given epoch is expressed as a linear combination of vertical TEC values at designated grid points. The discrete-time measurement model is given as

$$TEC_{rs}(t_j) = M(E)\,\Sigma_{i=1}^{4}\, W_i(t_j) \times VTEC_i(t_j) + b_r + b_s + v \qquad (3.5)$$

where $TEC_{rs}(t_j)$ is the TEC measurement (for receiver r and satellite s) at time t_j, and $VTEC_i(t_j)$ are values of vertical TEC at the four grid points surrounding the observation pierce point (Figure 3.5). For an optimal filter, v represents random zero-mean observation errors, with variance $E[vv^T] = R$. If, however, the measurement model does not describe the relationship between observations and unknowns fully, additional errors (arising from the model deficiencies) may exist. In such circumstances, the filter is operating suboptimally. The constants b_r and b_s are the receiver and satellite interchannel biases, respectively. These biases can be as large as several metres [cf. *Wilson and Mannucci*, 1993] and are estimated in the grid adjustment.

In order to reduce the modelling problem to two dimensions, the mapping factor M(E) is included (Figure 3.4). This factor essentially maps vertical TEC from grid points on an ionosphere shell to a slant TEC value along the observation line-of-sight.

$$M(E) = \{1-[\cos(E)/(1+h/R_E)]^2\}^{-1/2} \qquad (3.6)$$

where VTEC is the vertical TEC in el/m^2, E is the satellite elevation angle, R_E is the Earth radius, and h is height of the ionosphere shell (both R_E and h in kilometres).

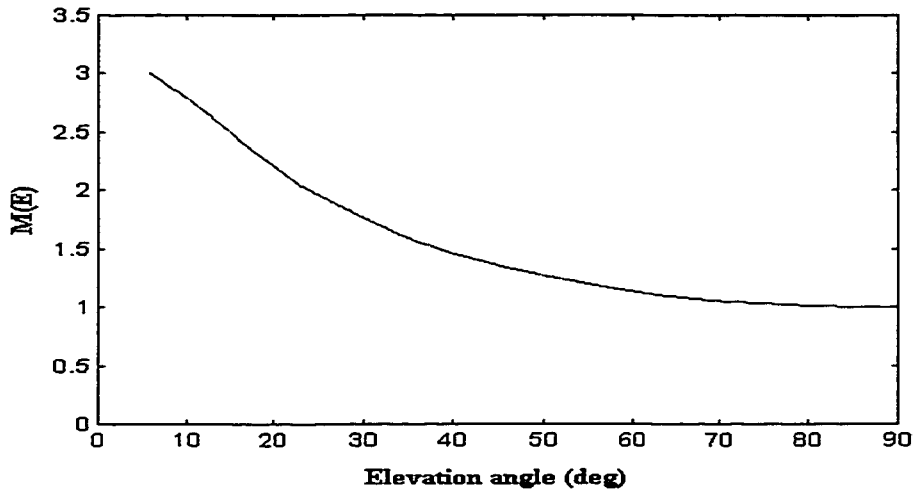


Figure 3.4 Mapping factor, for an ionosphere shell height of 350 km.

Derivation of this mapping factor and calculation of the ionosphere pierce point are provided in Appendix B. A standard shell height of 350 km is generally assumed [cf. *Mannucci et al.*, 1993; *Draganov et al.*, 1996; *RTCA*, 1998; *FAA*, 1997], corresponding to a mean centroid height for the electron density. Several authors, however, have determined that variations in the ionospheric centroid height can be significant, leading to inaccuracies in the estimated grid values of VTEC [*Komjathy and Langley*, 1996a; *Hansen et al.*, 1997]. *Conker et al.* [1998] demonstrated the following points, in considering the magnitude of errors introduced in mapping slant TEC observations to vertical, and in assuming that the ionosphere is a thin shell:

1) The errors are relatively stable, producing absolute average vertical ionosphere errors less than 15 cm (L1), and

2) Maximum error values exceed 2 m only in the region of very small elevation angles (E < 15 deg) and large average spatial gradients (> 0.3 m/100 km).

This issue is further addressed, in the context of the analysis conducted in this thesis, in Chapter 6.

The weighting factors $W_i(t_j)$ relate surrounding grid TEC values to the given measurement (Figure 3.5). The choice of spatial weighting function W is somewhat arbitrary. Algorithms may be based on inverse distance weighted averaging schemes [*Conker et al.*, 1997], bilinear interpolation [*Mannucci et al.*, 1993], or multiquadric weighting functions [cf. *Junkins et al.*, 1973]. It has been demonstrated, however, that similar results are obtained for the various spatial weighting functions [*El-Arini et al.*, 1994]. In tests of various measurement models for the purposes of this thesis, results (residuals) were found to agree at the centimetre level, under both quiet and active ionospheric conditions. For the purposes of this thesis, a multiquadric interpolation function is assumed:

$$
\begin{aligned}
W_1 &= W(1-t,1-u) \\
W_2 &= W(1-t,u) \\
W_3 &= W(t,u) \\
W_4 &= W(t,1-u) \\
W(t,u) &= t^2u^2(9-6t-6u+4tu)
\end{aligned}
\tag{3.7}
$$

This function was included in specifications for users of WAAS [*FAA*, 1994]. Recent specifications, however, have been changed to include a bilinear interpolation weighting function [*FAA*, 1997; *RTCA*, 1998]. Geometry of the observation pierce point, relative to the surrounding grid points, is shown in Figure 3.5. Calculation of the factors t and u is also illustrated.

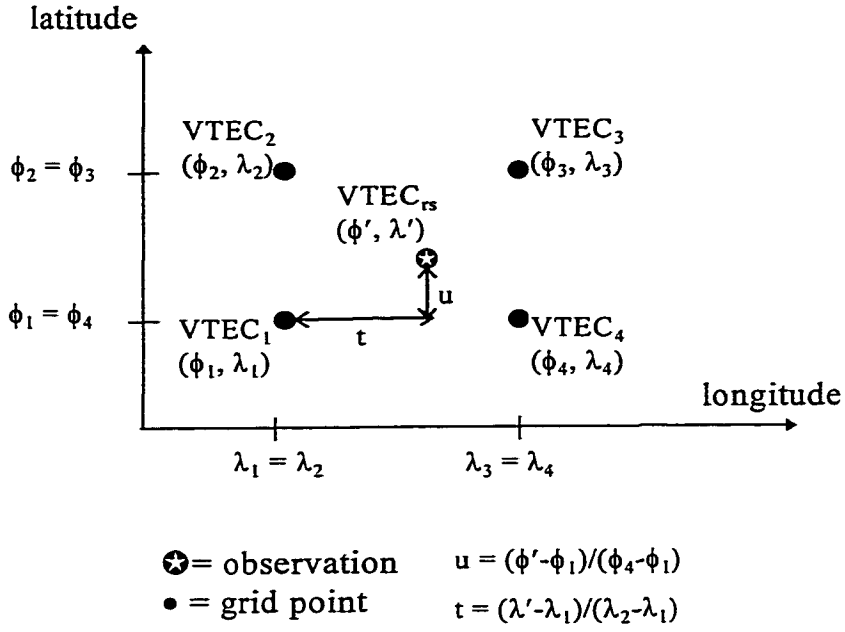


Figure 3.5 Observation pierce point, relative to surrounding grid points.

In addition to the grid-based ionosphere models, several authors have proposed the use of spherical harmonics and polynomial expansions in latitude and longitude [*Komjathy and Langley*, 1996b; *Wu et al.*, 1996; *Draganov et al.*, 1996; *El-Arini et al.*, 1994]. These models have performed with slant range delay accuracies on the order of 50 cm rms and less. Unlike the grid-based model, however, harmonic model coefficients depend on TEC observations from all regions of the ionosphere shell. Such models have a tendency to underestimate or overestimate TEC values in regions neighboring steep localised gradients [cf. *Mannucci et al.*, 1997]. In contrast, the grid model can be represented as a tesselation over a spherical surface, where individual rigid tiles are allowed to tilt and follow local gradients in TEC. An example of such a tiled surface is illustrated in Figure 3.6.

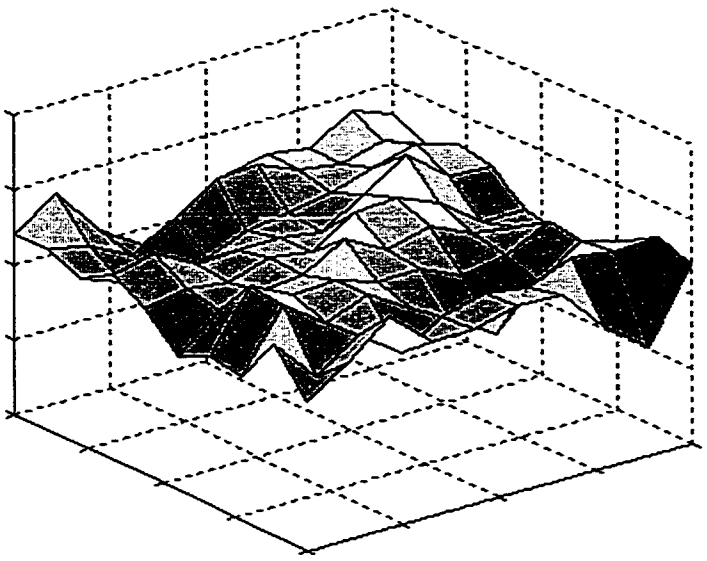


Figure 3.6 Example of surface tiling in two dimensions.

The multiquadric weighting function has been developed as a means of modelling large irregular surfaces, while allowing for resolution of fine-structure topographic features [*Junkins et al.*, 1973]. Such techniques are readily transferable to the 2-dimensional modelling of ionospheric features, as formulated in the ionosphere grid model.

3.3.3 System Model

The model unknowns ($VTEC_i$) are approximated as stochastic processes in time. Many grid algorithms use a first order Gauss-Markov system model. As demonstrated in a later chapter (Section 5.6.2.1), a first order Gauss-Markov process is consistent with observed temporal correlations in TEC, and the following system model is employed:

$$VTEC_i(t_{j+1})=e^{-\beta(\Delta t)}\, VTEC_i(t_j) + w(t_j) \qquad (3.8)$$

where $1/\beta$ is the *correlation time* and $\Delta t = t_{j+1} - t_j$ (prediction interval).

Equation 3.8 provides a statistical description of how TEC varies over time at each grid point. Vertical TEC values at a given grid point are only partially correlated with those at later epochs, the normalized autocorrelation function being $e^{-\beta(\Delta t)}$. The uncorrelated part of the predicted measurement $VTEC_i(t_{j+1})$ is described by a white noise sequence $w(t_j)$ with variance $q(t_j)$:

$$q(t_j) = \sigma^2[1-e^{-2\beta(\Delta t)}] \tag{3.9}$$

where $q(t_j)$ is referred to as *process noise*, and σ^2 represents the process spectral density ($2\beta\sigma^2$) divided by 2β. Values of β and σ can be derived from empirical autocorrelation functions for the random process VTEC(t). The value of σ^2 represents the combined variance of all contributing harmonic components of VTEC(t).

Receiver and satellite interchannel biases (b_r and b_s, Section 3.3.2) are also modelled as first order Gauss-Markov processes over time. The constant nature of these biases, however, results in a specification of correlation times much longer than those for ionospheric processes. As such, it is not necessary to solve for these biases at every update epoch, allowing greater redundancy. By monitoring observation residuals, drifts in biases can be detected over time and the relevant biases updated as necessary. For the purposes of this thesis, where the focus is on ionosphere modelling (i.e. estimates of grid point VTEC), bias estimates were provided by the control segment of the NRCan wide area network.

3.3.4 Prediction and Update Equations

The standard discrete Kalman filter equations are given as follows (after *Gelb* [1974]), where the superscripts - and + denote prediction and update, respectively.

1) Prediction (from time t_j to t_{j+1})

$$x^-(t_{j+1}) = \Phi(t_j, t_{j+1})\, x^+(t_j) + w(t_j) \tag{3.10}$$

$$P^-(t_{j+1}) = \Phi(t_j, t_{j+1})\, P^+(t_j)\, \Phi(t_j, t_{j+1})^T + Q(t_j) \tag{3.11}$$

2) Update (at time t_{j+1})

$$x^+(t_{j+1}) = x^-(t_{j+1}) + K[z(t_{j+1}) - H(t_{j+1})\, x^-(t_{j+1})] \tag{3.12}$$

$$P^+(t_{j+1}) = [I - KH(t_{j+1})]P^-(t_{j+1}) \tag{3.13}$$

where K is the gain matrix:

$$K = P^-(t_{j+1})\, H^T(t_{j+1})\, [H(t_{j+1})\, P^-(t_{j+1})\, H^T(t_{j+1}) + R(t_{j+1})]^{-1} \tag{3.14}$$

and the measurement model is given by

$$z(t_j) = H(t_j)\, x(t_j) + v(t_j) \tag{3.15}$$

The matrix elements are represented as follows:

$$x=\begin{pmatrix} VTEC_1 \\ \vdots \\ VTEC_n \end{pmatrix}, \quad z=\begin{pmatrix} VTEC_{rs,1} \\ \vdots \\ VTEC_{rs,m} \end{pmatrix}, \quad H=\begin{pmatrix} W_{11} & W_{12} & \cdots \\ W_{21} & \ddots & \\ \vdots & & W_{mn} \end{pmatrix} \tag{3.16}$$

$$R = = \begin{pmatrix} E[v_1v_1] & E[v_1v_2] & \cdots & E[v_1v_m] \\ E[v_2v_1] & E[v_2v_2] & & \\ \vdots & & \ddots & \\ E[v_mv_1] & & & E[v_mv_m] \end{pmatrix} \tag{3.17}$$

$$\Phi=\begin{pmatrix} e^{-\beta_1\Delta t} & 0 & \cdots & 0 \\ 0 & e^{-\beta_2\Delta t} & & \\ \vdots & & \ddots & \\ 0 & & & e^{-\beta_n\Delta t} \end{pmatrix} \tag{3.18}$$

$$Q=\begin{pmatrix} \sigma_1^2(1-e^{-2\beta_1\Delta t}) & 0 & \cdots & 0 \\ 0 & \sigma_2^2(1-e^{-2\beta_2\Delta t}) & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_n^2(1-e^{-2\beta_n\Delta t}) \end{pmatrix} \tag{3.19}$$

where n is the number of grid points, m is the number of observations, and P, R are the covariance matrices for estimates of the unknowns x (state vector) and the observations z, respectively. The process noise matrix is denoted by Q, while the H matrix describes the relationship between observations and unknowns and is referred to as the *design matrix*. In the notation of this chapter, I refers to the identity matrix.

Note that the off-diagonal terms in Equation 3.18 are assumed to be zero. These terms describe the relationship between temporal VTEC processes at different grid points.

While these terms are set to zero here, additional information could be supplied to the filter if these relationships were well-modelled. For the purposes of this thesis, however, in which periods of ionospheric disturbances are considered, such relationships can be random, time-varying and difficult to model. It is also likely that the correlation between temporal processes at adjacent grid points is relatively small, in regions of auroral disturbances, and the independence assumed in Equation 3.18 is adequate. Similarly, the off-diagonal terms in Equation 3.19 are assumed to be zero. These terms represent the cross-correlation between prediction (system model) errors at different grid points. This is an adequate assumption, for the random VTEC processes modelled here (provided the system model parameters σ and $1/\beta$ are representative values).

3.4 PRACTICAL MODEL IMPLEMENTATION

In order to initiate the filtering algorithm, four matrices must be specified: the design matrix (H), the transition matrix (Φ), the measurement noise covariance (R), and the process noise (Q). If the measurement covariance matrix (R) is fully populated, calculation of the gain matrix (K) requires inversion of an $m \times m$ matrix. By neglecting the off-diagonal terms, however, derivation of the gain matrix can be computationally simplified. In this section, measurement covariance values are estimated, and simplifying assumptions are investigated.

3.4.1 Derivation of Observation Noise Covariance Matrix R - Diagonal Terms

In order to derive typical measurement covariance values, sample observation errors were estimated from the smoothed values of TEC (i.e. Figure 3.3) for all satellites at a given reference station over a one day period. Observation errors $v(t_j)$ were derived as

$$v(t_j) = TEC_{PR}(t_j) - TEC(t_j) \qquad (3.20)$$

where $TEC(t_j)$ is the smoothed TEC value at time t_j (Equation 3.4) and $TEC_{PR}(t_j)$ is the noisy pseudorange-derived absolute TEC. In this case, the observation errors essentially represent errors (noise) in the $TEC_{PR}(t_j)$.

Figure 3.7 shows a sample time series of TEC observations and estimated errors for PRN 7 over a period of several hours. Note that the larger errors in Figure 3.7 correspond to smaller elevation angles (i.e. Figure 3.8), where multipath effects are largest. Such effects are propagated into the ionospheric smoothed TEC estimates.

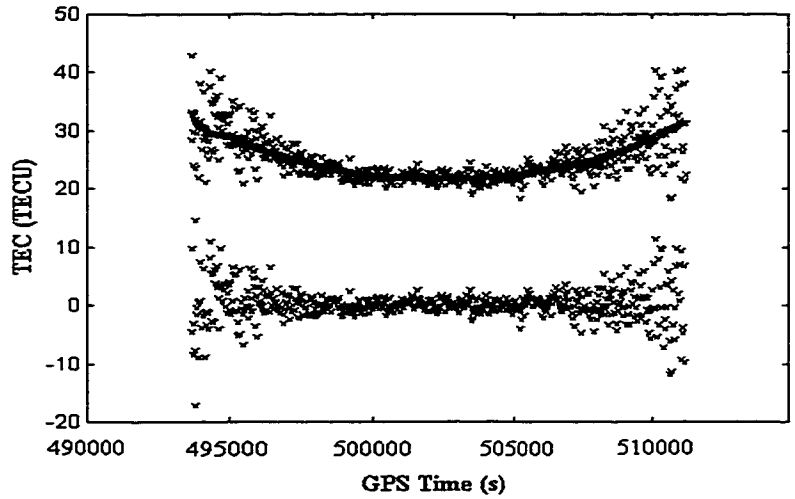


Figure 3.7 Smoothed TEC observations (upper plot) from a TurboRogue receiver, and estimated errors (lower plot) for PRN 7. The smoothed TEC values have been shifted for comparison purposes.

At a given time, TEC measurements from one reference station will have measurement errors of varying magnitudes, depending on the satellite elevation angles. Figure 3.9 shows estimated measurement errors over a one day period, as derived from observations at a single reference station (Albert Head, B.C.). At any given time, the probability distribution of errors in Figure 3.9a approximates a zero-mean Gaussian distribution. This is illustrated in Figure 3.9b.

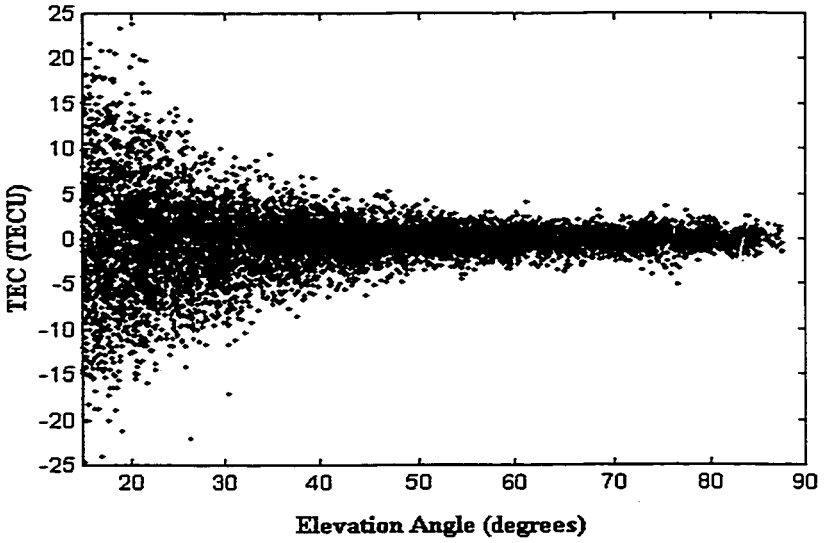


Figure 3.8 Variation of observation errors with elevation angle.

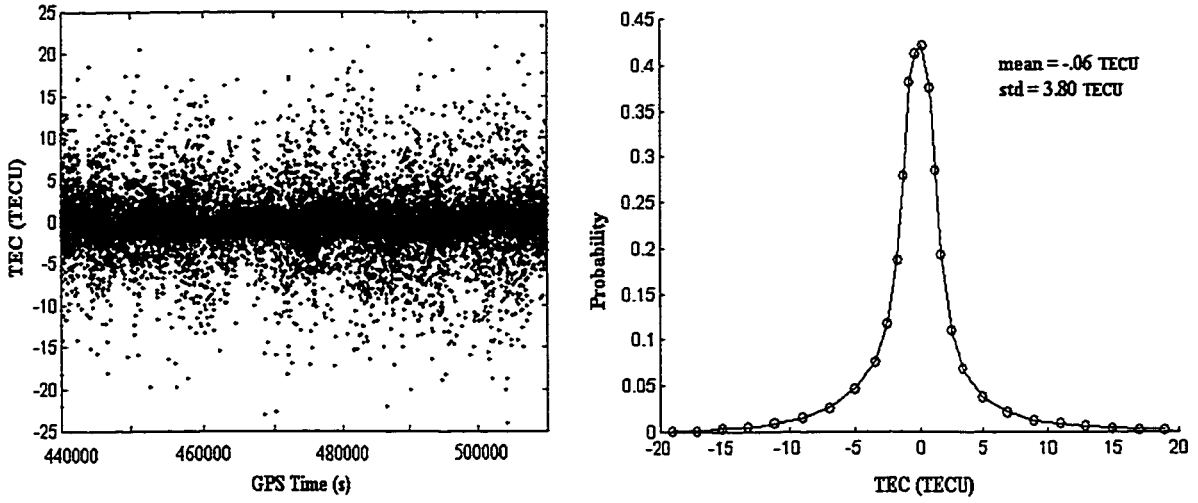


Figure 3.9 **a) Errors in TEC observations (TurboRogue receiver) from all observed series of TEC at Albert Head, B.C., over a one day period.**
b) Corresponding probability distribution of TEC measurement errors.

Correspondingly, the variance (second moment of the error distribution) associated with a given observation also depends on the satellite elevation angle. In implementing the ionosphere grid model in this thesis, diagonal terms in the R matrix are derived from comparisons between the pseudorange-derived and smoothed estimates of TEC (i.e. Figure 3.7), which reflect the greater uncertainties for larger elevation angles. Essentially, the variance of a given observation is computed as the variance of measurement errors (Equation 3.20) over the given satellite-receiver time series of TEC observations. These variance values primarily reflect noise in the TEC_{PR} estimates, and the variance of smoothed TEC values is actually significantly less (i.e. smoothed arc in Figure 3.7). The larger variance values are used, however, to reflect a range of uncertainties in the absolute value of the smoothed TEC, which depends on the estimated "offset" (Equations 3.3 and 3.4). For low frequency code multipath, the offset may be biased over the averaging time. An example of the variance estimation is shown in Figure 3.10.

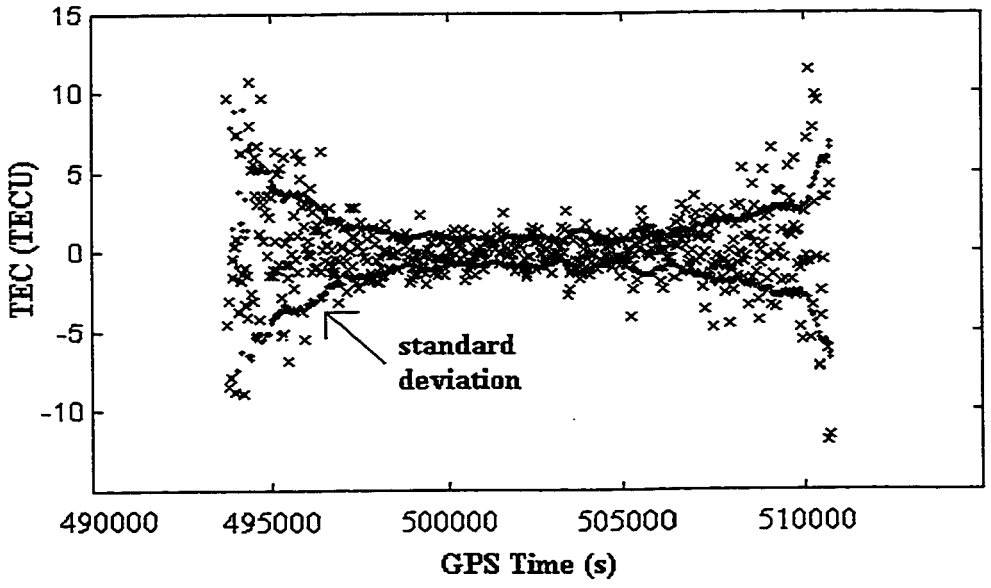


Figure 3.10 Derivation of measurement variance from estimates of observation uncertainties (20 degree elevation cutoff angle).

3.4.2 Derivation of Observation Noise Covariance Matrix R - Off-Diagonal Terms

In deriving elements of the R matrix, off-diagonal terms are considered negligible if observation errors are poorly correlated between observations: i.e.

$E[v_i v_k]=0 \qquad i \neq k$

This may occur when observation errors are derived from independent error sources. Errors in the observed values of TEC are primarily due to multipath effects and receiver noise (Section 2.4). Receiver noise is uncorrelated between receivers (i.e. different reference stations), and is generally uncorrelated between channels for a single receiver (i.e. between different satellites at the same reference station). Multipath effects are dependent on the antenna environment at a given station. These effects may be correlated for two satellite observations, observed from the same station, where the satellite-receiver lines-of-sight are separated by a small angle.

In order to investigate the significance of off-diagonal terms in the observation covariance matrix R, cross correlations of observation errors were derived for all available satellite pairs in the Albert Head (one day) data set. The cross-correlation coefficients are derived as

$$\rho_{XY} = \frac{E[XY] - E[X]E[Y]}{\sigma_{XX}\sigma_{YY}} \tag{3.21}$$

where E[XY]-E[X]E[Y] is the covariance, and σ_{XX} and σ_{YY} are the variances of the individual observation errors for two different satellites (denoted X and Y). The cross-correlation coefficients give a measure of relative magnitudes of off-diagonal, versus diagonal, terms in the measurement covariance matrix R. Figure 3.11 shows the cross-correlation coefficients derived for the Alberta Head data set, and the corresponding probability distribution.

From the distribution in Figure 3.11b, it can be observed that the correlation coefficients are consistently less than .40, and generally less than .20, indicating that off-diagonal terms in the measurement covariance matrix are approximately 80% smaller than the diagonal terms. For the purposes of this thesis, the off-diagonal elements of R are set to zero when implementating the ionosphere grid model. This allows simplication of the processing algorithm, as described in the next section. It must be noted, however, that this may result in suboptimal filter performance when simultaneous observations from a given station are separated by short distances (i.e. errors highly correlated). Modelling the error correlation under such circumstances is difficult, however, requiring knowledge/modelling of the antenna environment in order to estimate the specular (correlated) multipath component. Grid accuracies under the simplifying assumptions (off-diagonal terms set equal to zero) are shown to be consistent with results for existing ionosphere grid models [i.e. *Bertiger et al.*, 1997] in Chapter 6.

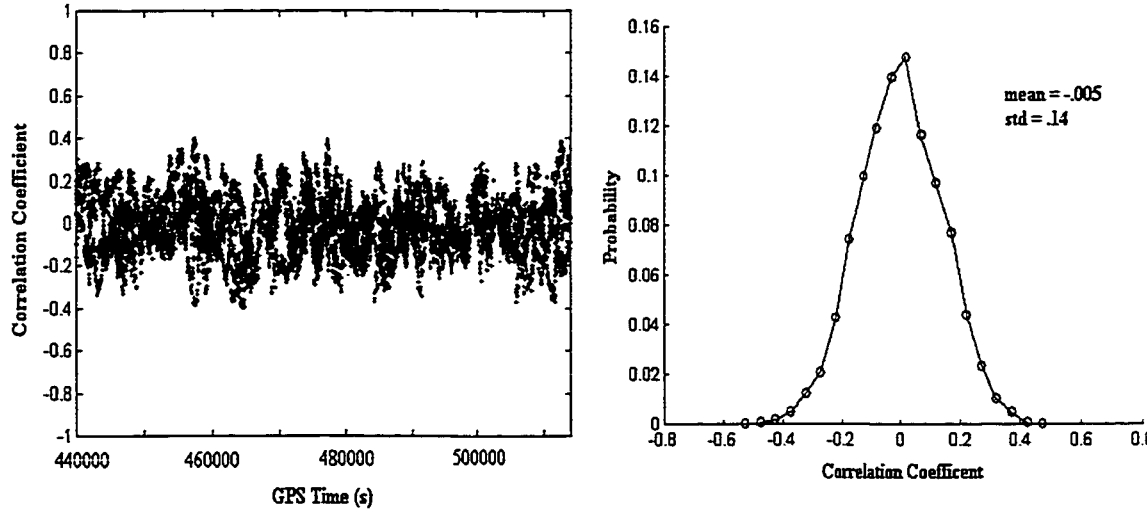


Figure 3.11 **a) Cross-correlation coefficients for all satellite pairs over a one day period, as observed from reference station Albert Head.**
b) Corresponding probability distribution.

3.4.3 Sequential Adjustment

For non-correlated measurement errors, the measurement noise covariance matrix R is a diagonal matrix. In this case, the Kalman filter update (Equation 3.14) can be computed in a series of steps, by folding each new measurement (at the given epoch) sequentially into the updated filter estimates. In this processing algorithm, the filter prediction is identical to Equations 3.10 and 3.11. The filter update, however, is computed as follows (for m measurements at time t_{j+1}).

For the first new measurement (at time t_{j+1}):

$$x^{+}_{1}(t_{j+1}) = x^{-}(t_{j+1}) + K[z(t_{j+1}) - H(t_{j+1})\, x^{-}(t_{j+1})]$$
$$P^{+}_{1}(t_{j+1}) = [I - KH(t_{j+1})]P^{-}(t_{j+1})$$

where K is the gain matrix:

$$K = P^{-}(t_{j+1})\, H^{T}(t_{j+1})\, [H(t_{j+1})\, P^{-}(t_{j+1})\, H^{T}(t_{j+1}) + \sigma^{2}_{1}]^{-1} \qquad (3.22)$$

Note that the design matrix H has dimensions 1 × u, and the term $H(t_{j+1})P^{-}(t_{j+1})H^{T}(t_{j+1})$ is a scalar. The value σ^{2}_{1} denotes the variance associated with the first new measurement at time t_{j+1}. The m × m inversion in Equation 3.14 has been reduced to a scalar inverse in Equation 3.22.

Updated estimates, $x^{+}_{1}(t_{j+1})$ and $P^{+}_{1}(t_{j+1})$, from this first step are then used as "predictions" for a second update using the second available measurement at epoch t_{j+1}. This process is repeated until all observations from time t_{j+1} have been incorporated in the final estimates of $x^{+}(t_{j+1})$ and $P^{+}(t_{j+1})$. Figure 3.12 illustrates this processing algorithm. This simplification of the Kalman filter algorithm allows greater computational

efficiency. For typical grid models, where 6-8 satellites are simultaneously observed from each reference station in the network, this sequential adjustment eliminates the inversion of a matrix with rank ≈70.

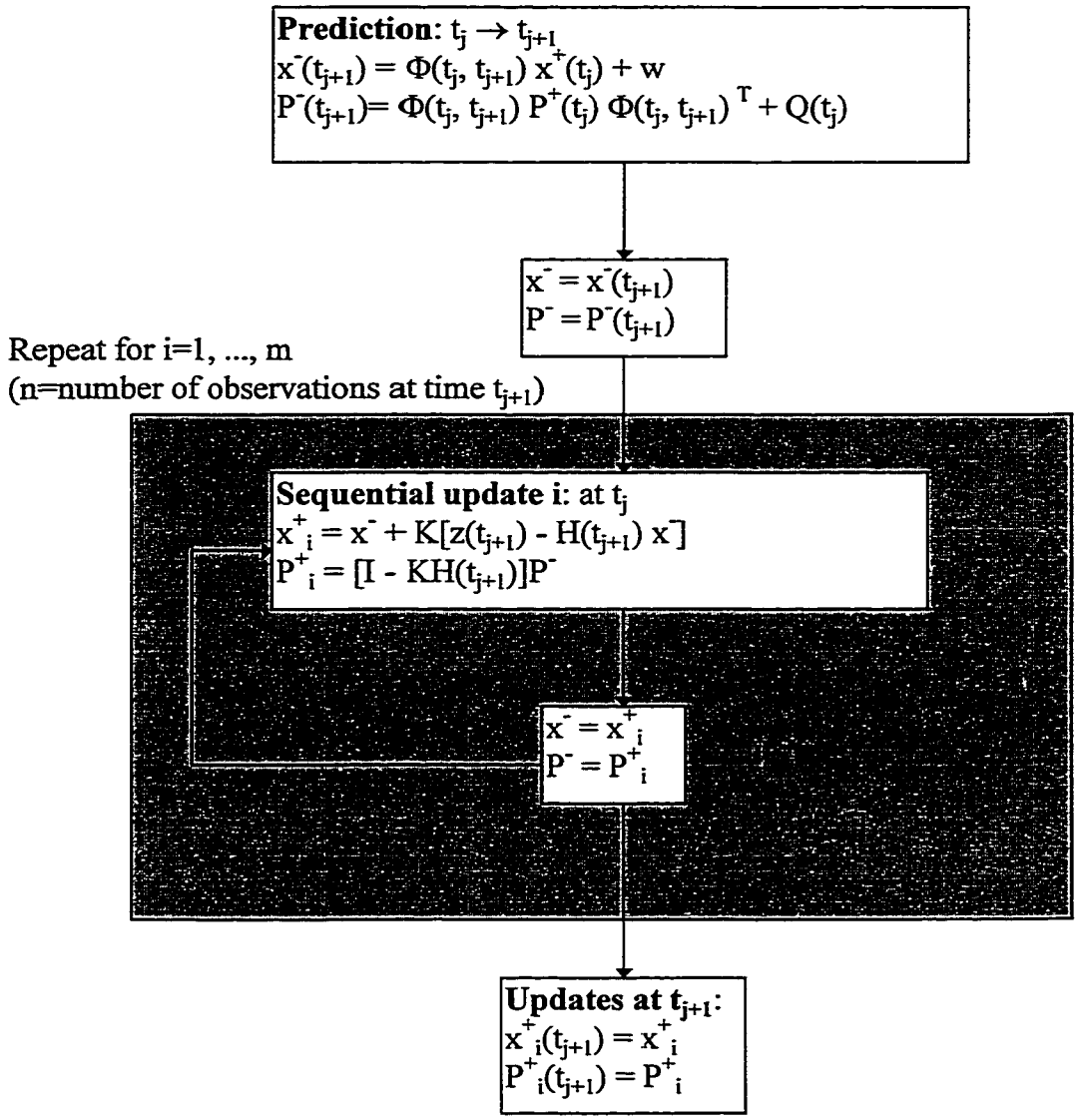


Figure 3.12 Sequential Kalman filter adjustment.

CHAPTER 4

AURORAL ZONE PHENOMENA

The auroral zone is a high latitude region of the Earth's ionosphere, where dynamic phenomena such as the aurora ("Northern Lights") take place. The unique nature of this region derives from complex interactions between the solar wind and the terrestrial magnetic field, some of which are still not entirely understood. In this thesis, the focus is directed towards studying manifestations of such processes in the auroral zone, and associated variations in the total electron content, rather than the underlying physical processes. It is important to understand the nature of auroral disturbances, and formation of the auroral zone ionosphere, however, in order to properly interpret TEC observations from this region. A brief description of the auroral oval and associated current systems, and an overview of enhanced auroral activity, are included in this chapter.

4.1 AURORAL OVAL

The auroral oval [*Feldstein and Starkov*, 1967] (Figure 4.1) is a high latitude region of the ionosphere, characterized by enhanced conductivity and energetic electron precipitation at E layer altitudes. This region is limited in latitudinal extent, located between approximately 65° and 75° geomagnetic latitude (nightside), with an average width of 5-7 degrees [*Rostoker and Skone*, 1993]. Under storm conditions, the nightside oval may expand to twice this width. Dayside oval boundaries are ≈ 75-80° geomagnetic latitude [*Akasofu*, 1968]. Auroral regions include much of Canada and Alaska, in addition to parts of Russia and Northern Europe. Under active conditions, large-scale electric currents (carried primarily by electrons) facilitate coupling of the auroral ionosphere and

solar wind. Intensifications of these currents result in the visible *aurora* or Northern Lights, and variable TEC.

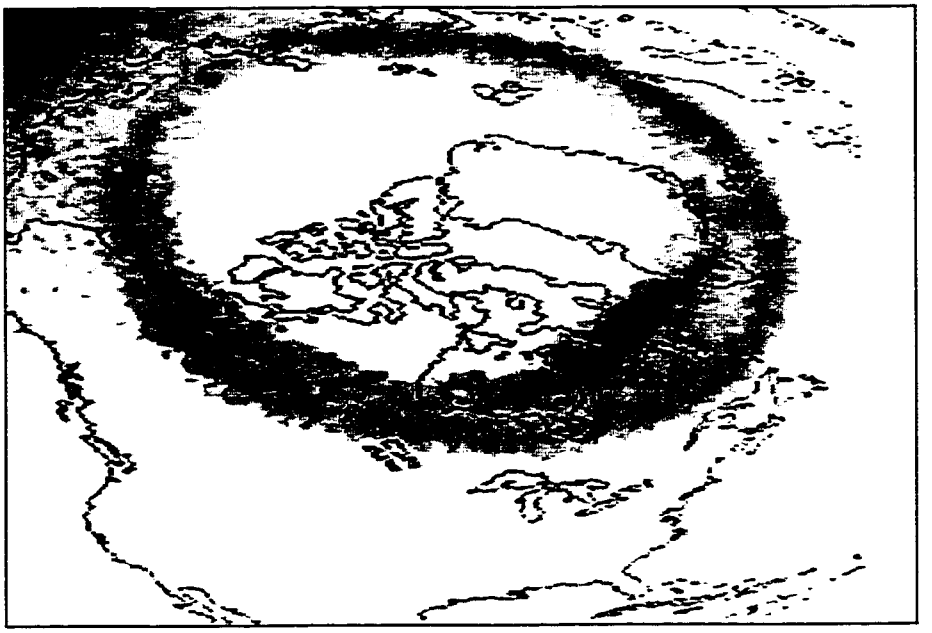


Figure 4.1 UV image of the nightside auroral oval, over North America (courtesy L. Frank, University of Iowa). Darker discrete structures within the oval correspond to the precipitation of energetic electrons.

4.2 AURORAL CURRENT SYSTEMS

Variations in auroral electron density are directly related to two current systems, which are components of an active auroral ionosphere: auroral electrojets and field-aligned currents.

4.2.1 Auroral Electrojets

In the active auroral zone ionosphere, anisotropic conductivities allow the formation of both horizontal (Hall) and vertical (Pedersen) electric currents, in

accordance with Ohm's Law. *Harang* [1946] established the location of large-scale east-west ionospheric currents at auroral zone latitudes. These currents, called electrojets, flow primarily in the E region (100-150 km altitude) and are fixed with respect to the Sun. The eastward electrojet flows across dusk, while a stronger westward electrojet flows across dawn, penetrating into the evening sector through midnight (as depicted in Figure 4.2). Variations in these currents, during periods of enhanced ionospheric activity, can be observed in ground-based magnetic field measurements.

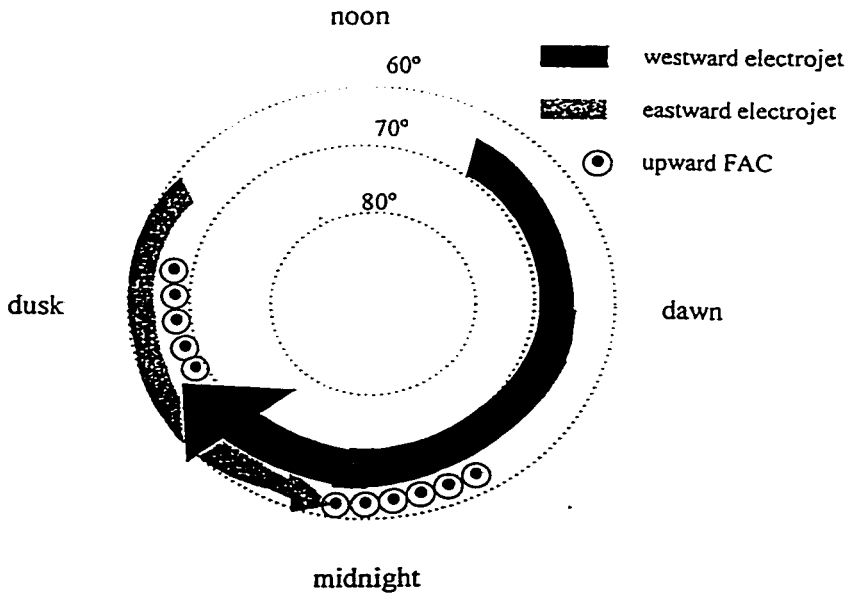


Figure 4.2 Auroral electrojet currents and field-aligned currents.

The electrojets are components of a three-dimensional current system which includes vertical *field-aligned currents* [*Birkeland*, 1908; *Iijima and Potemra*, 1978; *Zmuda and Armstrong*, 1974] - electrons flows aligned with terrestrial magnetic field lines. The field-aligned currents are colocated with the auroral oval, and serve to facilitate solar wind-ionosphere coupling. The strength of these currents is directly correlated with the level of auroral activity.

4.2.2 Field-aligned Currents

Figure 4.3 depicts the Earth's magnetosphere, where the geomagnetic field is modified and distorted through electrodynamic solar-terrestrial interactions, stretching into a nightside *magnetotail*. The magnetosphere boundary (magnetopause) serves to shield the Earth's magnetic field from the solar wind, such that solar particles do not enter the auroral ionosphere directly. Instead, energized magnetotail particles are precipitated into the auroral oval along magnetic field lines, via indirect solar-controlled processes.

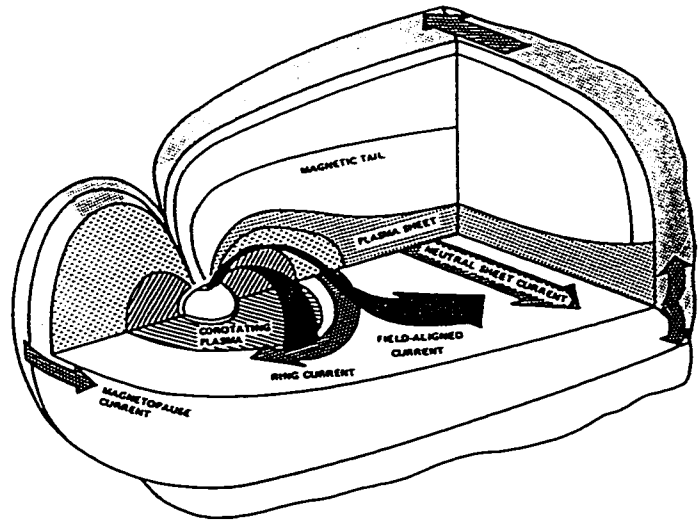


Figure 4.3 The Earth's magnetosphere (after [*Parks*, 1991]).

Figure 4.2 shows the locus of upward field-aligned currents in the auroral oval. These currents map (along terrestrial magnetic field lines) to the central plasma sheet (nightside magnetotail) [cf. *Lassen*, 1974] and boundary regions, which are populated by warm electrons (5 keV) [*Huang and Frank*, 1986] of both solar wind and terrestrial origin. These electrons may become energized through solar wind-magnetosphere

interactions and precipitate into the auroral ionosphere, via the field-aligned currents. Large-scale generation of field-aligned currents, corresponding to the Figure 4.2 pattern, results from oppositely directed plasma flows or, equivalently, diverging electric fields in the magnetotail and boundary regions [*Goertz and Boswell*, 1979; *Hasegawa and Sato*, 1979]. Small-scale instabilities map to localised current structures in the auroral ionosphere, resulting in localised enhancements/depletions of auroral TEC. Intense field-aligned currents are a feature of magnetospheric substorms, and auroral disturbances, where visible aurora are a manifestation of the precipitating electrons. The darker discrete structures in Figure 4.1 correspond to the locus of intense field-aligned currents.

4.3 COMPONENTS OF AURORAL ACTIVITY

While enhancements of the large-scale currents described in Sections 4.2.1 and 4.2.2 are correlated with the presence of aurora (and variable TEC), some distinctions between components of auroral activity must be made. The auroral signatures associated with enhanced auroral activity are caused by two distinct components: a directly driven process, and an explosive "loading-unloading" process [*Rostoker et al.*, 1987]. The directly driven system is characterized by the large-scale auroral electrojets and field-aligned currents (Figure 4.2). The dynamics of this system are such that input energy from the solar wind is dissipated in the ionosphere with no intermediate storage of energy in the magnetotail. Variations in the directly driven electrojet strength, in direct response to changes in the solar wind, take place over a time-scale of hours.

Short explosive periods of energy dissipation are associated with the "unloading" system, resulting from local formations of field-aligned current wedges. The wedge consists of two oppositely directed sheets of field-aligned current, separated in longitude, which are closed through a longitudinally localized westward current in the auroral ionosphere (Figure 4.4). This current system has a lifetime on the order of 10 minutes

[*Rostoker*, 1991]. Unlike the directly driven system, stored energy is released (or "unloaded") through internal instabilities in the magnetosphere. These intense smaller-scale current systems characterize magnetospheric substorms.

The visible aurora are created through collisions of precipitating electrons (as carried by the field-aligned currents) and neutral oxygen, nitrogen and hydrogen molecules at E-region altitudes (95-150 km). Diffuse aurora [cf. *Hargreaves*, 1992], which occur at lower auroral latitudes, are associated with enhancements of the directly driven system. The more structured forms of discrete aurora correspond to the presence of intense electron precipitation, resulting from "unloading" system activity.

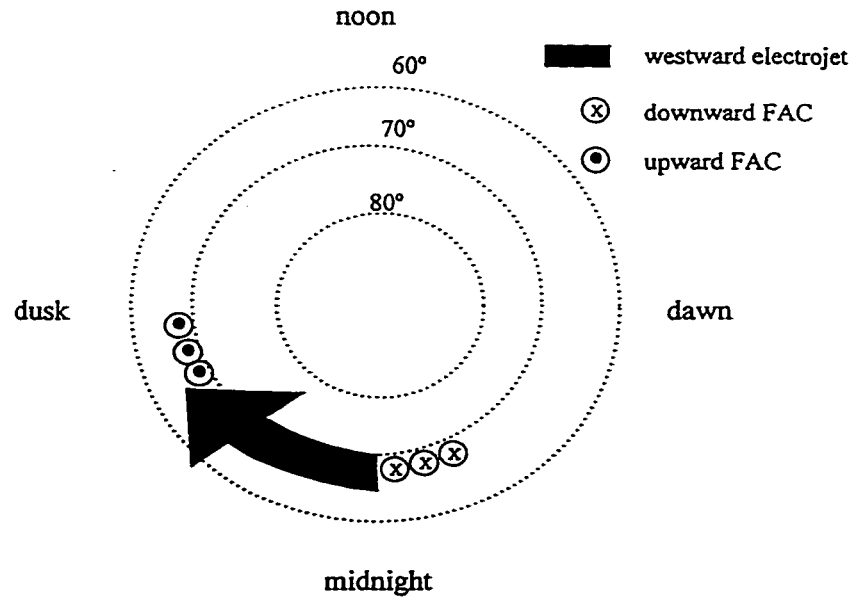


Figure 4.4. Localised current structure (current wedge) associated with intense electron precipitation during substorm events.

4.4 E-REGION IRREGULARITIES IN ELECTRON DENSITY

Ionization of the nighttime E layer, via precipitating energetic electrons in the aurora, is often referred to as Auroral E-ionization (AEI) [cf. *Hunsucker et al.*, 1995]. AEI facilitates the propagation of terrestrial RF signals and is often studied for communications applications, where it may be referred to as *radar aurora*. AEI is caused by enhancements of both the directly driven and unloading systems, and electron density profiles associated with AEI have been studied by several authors.

Structured irregularities can be extremely localised, with horizontal scale sizes ranging from 20 to 400 km [cf. *Coker et al.*, 1995]. These irregularities are also aligned along terrestrial magnetic field lines [*Hunsucker*, 1992], with electron densities concentrated at altitudes near 110 km. *Bates and Hunsucker* [1974] and *Baron* [1974] have observed peak AEI electron densities in the range $2\text{-}20\times10^{11}$ el/m^3. These values are comparable to maximum F region densities in Figure 2.4. Such increases in electron density can take place over a time scale of minutes.

It is important to note that precipitating electrons may also ionize neutrals along their entire path length, creating irregularities in the F region. These structures have relatively small scale sizes, ranging from several metres to 10 km [cf. *Hunsucker*, 1992]. These structures are associated with scintillation effects, with frequencies on the order of 0.5-1.0 Hz [*Hargreaves*, 1992]. Such effects are not studied in this thesis, where observations were available at only 30 s intervals.

4.5 ENHANCED AURORAL ACTIVITY - DEFINITIONS

In this thesis, TEC characteristics and implications for ionospheric grid models are studied during periods of "enhanced auroral activity". From the previous sections, it is

evident that there are several elements which may constitute such activity and a clear definition is required. Furthermore, in the GPS literature it is common to refer to enhanced ionospheric activity by a variety of terms: magnetic storms, geomagnetic disturbances, ionospheric storms, and substorms. These terms are used as general "catch-alls", intended to categorize, rather than define, the nature of ionospheric data being studied. Often, "storm" periods are identified using geomagnetic indices [cf. *Chavin*, 1996]. For the purposes of this thesis, such general categorization of enhanced auroral ionospheric activity is inadequate. In this section, distinct components of enhanced auroral activity are defined by nature of the associated physical processes, and the nomenclature clarified.

4.5.1 Magnetospheric Substorms

The term *magnetospheric substorm* is often applied to observed auroral disturbances. The concept of a substorm, however, requires the involvement of a significant portion of the near-Earth magnetotail, with well-defined field-aligned current signatures. In an attempt to clarify what classifies as a "substorm", a brief description follows.

There are three stages of magnetospheric substorm development: growth phase [*McPherron*, 1970], expansive, and recovery phases [*Akasofu*, 1968]. Each phase is defined by distinct auroral signatures. Initally, during the growth phase, energy is deposited in the magnetosphere and the auroral oval increases in size, the equatorward boundary moving southward. Part of the enhanced energy is stored in the magnetotail, while part is dissipated directly into the directly driven currents. As increasing energy is deposited in the auroral oval, conductivity increases, leading to enhanced electrojets and field-aligned currents of the directly driven system. This may be accompanied by

increased auroral disturbances in TEC, AEI and diffuse/discrete auroral intensifications [cf. *Murphree et al.*, 1991].

The expansive phase is characterized by substorm onset, in which multiple substorm current wedges (the "unloading" system), corresponding to intense discrete aurora, are observed near local midnight. These are triggered by unknown mechanisms in the magnetotail, such that a large amount of stored energy is released into the ionosphere over a short time. The dominant component of this substorm phase is the "unloading" system, and the largest temporal and spatial variations in electron density, and AEI, are observed during this phase. Multiple onsets may be observed during one substorm event. Auroral forms associated with this stage have been observed to develop and dissipate within 30-60 minutes [*Hargreaves*, 1992]. Distinct signatures associated with this phase of substorm development include the brightening of the most equatorward discrete arc [*Rostoker et al.*, 1980], followed by poleward and westward expansion of multiple current wedges. Near-Earth signatures include a dipolarization of the geostationary magnetic field [*Nagai*, 1982]. After this stage, a gradual "recovery" returns the oval to quiet time conditions.

Given this definition of a substorm, only those auroral events consisting of growth, expansive and recovery phases are herein classified as "substorms". Other enhanced auroral activity may consist of significant directly driven activity, with structured aurora, AEI signatures and auroral disturbances, but no clearly defined expansive phase. In such cases, TEC variations may still be significant.

4.5.2 Magnetic Storms, Ionospheric Storms, and Travelling Ionospheric Disturbances

The magnetic storm is a global disturbance, characterized by magnetic signatures in the mid-, and low-latitude regions. While these events may be associated with magnetospheric substorms and aurora, substorms are not a required element in the evolution of a magnetic storm.

The ionospheric storm is characterized by F region irregularities in electron density, at middle latitudes. This phenomenon is most likely controlled by the atmospheric neutral wind, although the physical processes are not as yet clearly understood [*Hargreaves*, 1992]. The ionospheric storm has demonstrated some correlation with auroral activity (and magnetospheric substorms), but the two phenomena are not necessarily connected.

During periods of intense auroral activity, travelling ionospheric disturbances may also be created. These phenomena resemble waveforms, propagating from auroral latitudes to the equator. These disturbances cause significant variations in F region TEC at middle latitudes [cf. *Chavin*, 1997].

4.5.3 Definition of "Enhanced Auroral Activity"

For the purposes of this thesis, enhanced auroral activity will be defined as periods during which there is evidence of auroral disturbances associated with the "unloading" system. During such times, diffuse and discrete aurora may be present, with significant AEI and associated irregularities in electron density [cf. *Hunsucker*, 1992]. This definition does not preclude periods of short-lived substorm expansive phases, but classification as a magnetospheric substorm event is not a criteria.

The terms AEI, auroral disturbances, and "enhanced auroral activity" are herein used interchangeably, although the terms are not strictly equivalent. The term AEI has been used by several authors, however, as a convention to define auroral disturbances with distinct TEC signatures. In order to better reference these previous studies, it is convenient to classify AEI signatures and irregularities associated with "enhanced auroral activity" in the same category. The enhanced auroral activity studied in this thesis may occur in conjunction with travelling ionospheric disturbances, and global magnetic or ionospheric storm effects at middle latitudes. No attempts have been made to verify the presence of such activity, however.

4.6 THE MAIN TROUGH

Field-aligned currents are not present in the region several degrees south of the auroral oval, called the *main trough*, which can result in extreme variations of ionospheric activity across the equatorward oval boundary. Peak electron densities in this region are on the order of 5×10^{10} el/m^3, at altitudes of 300-400 km [*Bates et al.*, 1973]. These densities are much less than those in the auroral oval, which results in sharp gradients across the trough/oval boundary. This has been referred to as the "Canadian border effect", in reference to the geographic location of the transition region. In general, the trough is located at latitudes 60-65° geomagnetic, and is present only in the nightside local time sector. The trough is observed in conjunction with enhanced auroral activity, particularly during substorm growth phase.

4.7 SIGNATURES

Since electric currents give rise to magnetic field perturbations, integrated effects of local current systems may be observed at ground-based magnetometer stations. The magnetic field signatures associated with the auroral electrojets and field-aligned current

wedges have been well documented in the literature [*Rostoker et al.*, 1980; *Clauer and McPherron*, 1974]. The westward auroral electrojet (Figure 4.2) causes a southward magnetic field perturbation at local ground stations. This is a large-scale signature, generally observed at auroral stations in the local time sector 2100 to dawn (where the westward electrojet is dominant). The "unloading" system current wedge (Figure 4.4) displays a similar signature (southward magnetic field perturbation), but is limited in local time and latitude, resulting in localised short-term geomagnetic disturbances.

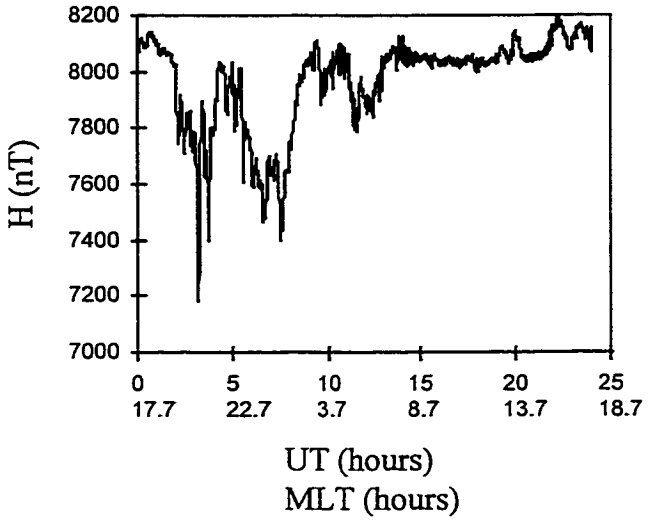


Figure 4.5 Magnetic field variations at Fort Churchill, versus hours UT and magnetic local time (MLT). The H component of the magnetic field is south-north, in the local horizontal plane.

Figure 4.5 illustrates the magnetic field signatures at ground station Fort Churchill (south-north component H), in a region of auroral disturbances. Note the large-scale negative perturbation from 0 to 10 UT, with higher frequency variations superposed. These are characteristic of an enhanced directly driven system (enhanced westward electrojet - Figure 4.2), with short-term enhancements of field-aligned current wedges (Figure 4.4). Irregularities in electron density (and TEC) are correlated with the variations in current strength througout the interval 0 to 10 UT.

CHAPTER 5

TEMPORAL AND SPATIAL CHARACTERISTICS OF TEC IN THE AURORAL REGION

Ionospheric phenomenology in the auroral region has been described in Chapter 4. Large variations in electron density, and TEC, are associated with enhanced auroral activity (AEI). The magnitude of such variations is an issue for wide area grid modelling, and few statistics on the temporal and spatial variability of TEC, in the auroral region, have been compiled from GPS data to date. The derivation of a consistent data set, representative of TEC variations in the auroral region, is important for WADGPS systems operating in Canada, Northern Europe and Alaska. Naturally, such statistics are useful for wide area ionosphere grid modelling. Additionally, however, extended applications exist in the development of local area augmentation systems (LAAS), which require double-difference accuracies (second order ionospheric gradients) on the order of centimetres.

In this chapter, an extensive set of GPS observations from the Natural Resources Canada (NRCan) wide area network is used to estimate spatial and temporal gradients of VTEC over Canada, during periods of auroral disturbances. A consistent data set is selected using local magnetic field data from the CANOPUS MARIA array, and enhanced auroral activity is identified from magnetic field signatures (as outlined in Section 4.7). Stochastic properties of TEC variations, in both time and space, are also verified using spectral analysis methods.

5.1 GENERAL CHARACTERISTICS OF AURORAL TEC

Figure 5.1 shows a specific example of localised variations in TEC during a period of enhanced auroral activity. The time series of TEC for PRN 3 and PRN 19 were calculated using dual frequency observations from several reference stations in the Natural Resources Canada wide area network (Figure 5.3). Figure 5.2 shows corresponding satellite pierce points, for each SV series plotted in Figure 5.1, and the region of auroral disturbance or AEI (see Section 4.5.3). The plots in Figure 5.1a (5.1b) correspond to SV pierce points inside (outside) the region of disturbance. Note the large variations in Figure 5.1a, which are caused by the local geomagnetic disturbances. In contrast, the "quiet" values of TEC vary smoothly in Figure 5.1b.

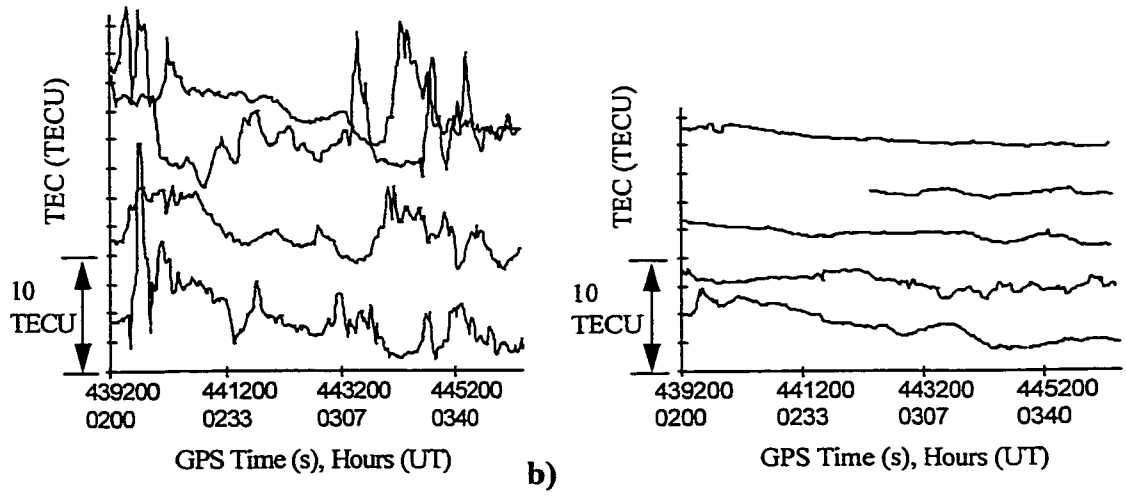


Figure 5.1 **a) Time series of TEC observations for satellite-receiver pairs with pierce points inside region of auroral geomagnetic disturbance: Oct. 18, 1996.**

b) Time series of TEC observations for satellite-receiver pairs with pierce points outside region of auroral geomagnetic disturbance: Oct. 18, 1996.

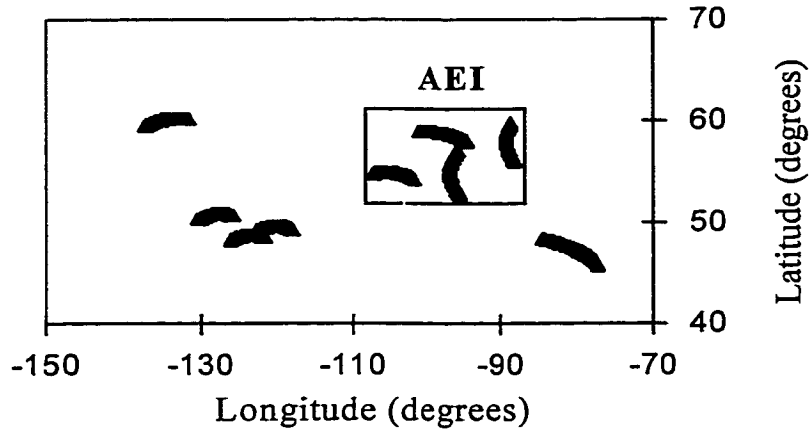


Figure 5.2 Ionospheric pierce points for PRNs 3 and 19, and selected NRCan reference stations, during the interval 0200-0400 UT: October 18, 1996.

The effects in Figure 5.1a are localised and it is apparent that, in order to derive representative statistics for regions of auroral disturbances, it is necessary to identify a consistent data set, based on local levels of geomagnetic activity. The selection of such a data set is described in the next section.

5.2 DATA SELECTION

Several studies [*Chavin*, 1996; *Draganov et al.*, 1996; *Mannucci et al.*, 1997] have attempted to determine the accuracy of WADGPS grid corrections during periods of "disturbed" geomagnetic activity. In such studies, it is common to test ionospheric grid performance using observations (or simulated observations, i.e. Bent model) from periods characterized by high values of geomagnetic activity indices, such as Kp, Ap, or AE. Derivation of these indices is described in *Mayaud* [1980]. A brief discussion follows here.

Essentially, these indices are based on variations in magnetic field measurements at various stations distributed around the globe. As described in Section 4.7, magnetic field signatures correspond directly to levels of ionospheric activity. The largest variations during a given period, from each contributing station, are combined and converted to a numerical index. Larger indices correspond to greater magnetic field perturbations. The Kp index is derived from sub-auroral zone stations, at three-hour intervals; Ap is a daily index, derived from the same data as Kp. The AE index is derived from auroral zone stations, at intervals of one hour.

The use of such indices to infer levels of local ionospheric activity is ambiguous, since these indices are based on global measurements and may not reflect disturbances in the local region of interest. Additionally, the transient nature of ionospheric processes allows significant variations in activity over several hours. Indices derived from several hours of data do not necessarily reflect ionospheric activity during the entire interval of observations. In order to derive a data set truly representative of ionospheric disturbances, the location/duration of enhanced ionospheric activity must be identified consistently and correctly. An alternative to indices must be employed. One such alternative involves making use of local magnetic field signatures, within the GPS network of interest. This approach is employed here.

For the purposes of this study, periods of enhanced auroral activity (as defined in Section 4.5.3) were identified from one-minute averaged magnetometer data, collected at twelve stations across Canada during the period April - October 1996. These stations form the CANOPUS MARIA array. This array, and other data sources, are described on the CANOPUS (Canadian Auroral Network for the Open Program Unified Study) web site: http://canott.dan.sp-agency/www/canopus_home.html. Magnetometer data from this array were provided courtesy of the Canadian Space Agency.

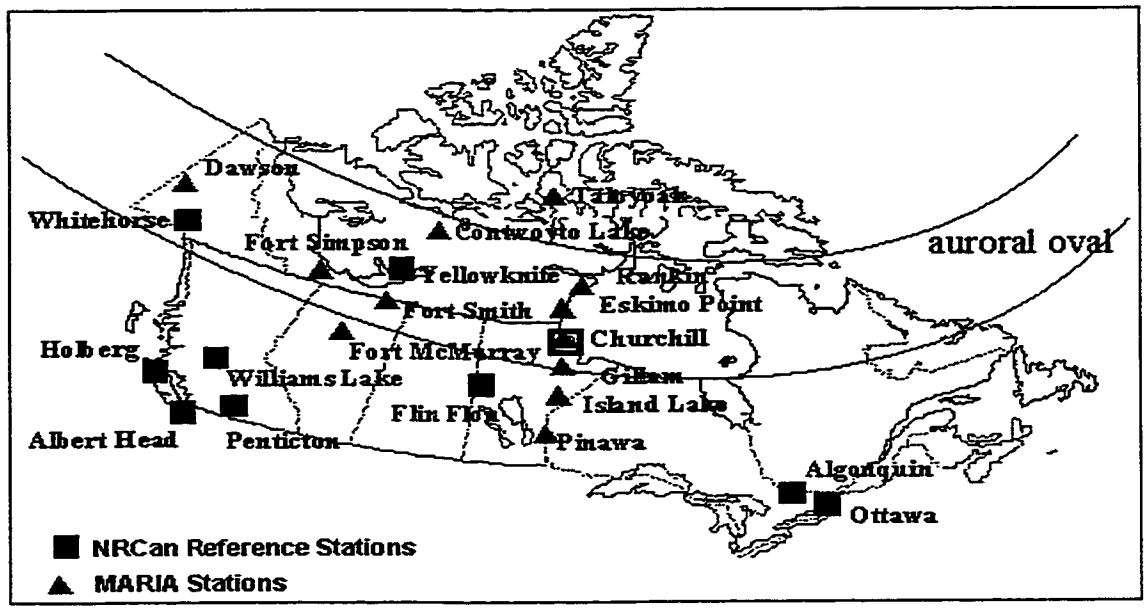


Figure 5.3 CANOPUS MARIA observatories and NRCan reference stations.

Geographic locations of the stations are illustrated in Figure 5.3. Note that six stations in central Canada are located along approximately the same meridian, while several stations are located in western Canada, providing good coverage of the auroral oval in both latitude and local time. Regions of geomagnetic disturbances were identified from the magnetic field data as follows.

The identification of geomagnetic disturbances is illustrated in Figure 5.4, using magnetic field signatures described in Section 4.7. In this figure, fluctuations in the X (South-North) component of the magnetic field are evident at several stations, with negative perturbations being evidence of auroral disturbances. Such negative perturbations begin at Rankin, Eskimo Point, and Churchill, during the period 0200-0230 UT. In contrast, the magnetic field perturbation is positive at Island Lake, indicating that Island Lake is at the equatorward boundary of electrojets characterizing the auroral oval. The correlation between negative magnetic field perturbations (AEI) and TEC variations

can be observed in Figures 5.1 and 5.4, for the time period 0200-0400 UT. The region of AEI in central Canada (as identified from Figure 5.4) is plotted in Figure 5.2, and corresponding TEC signatures are plotted in Figure 5.1a.

For the purposes of this thesis, geomagnetic disturbances at a given station are identified as total magnetic field perturbations greater than ≈700 nT. Periods of enhanced auroral activity are selected as those intervals during which magnetic field perturbations reached 700 nT. This threshold is chosen somewhat arbitrarily, based on observed TEC variations and magnetic signatures associated with AEI. If geomagnetic disturbances are identified simultaneously in magnetometer data from both central and western Canada, the region of disturbance is assumed to extend across Canada. If geomagnetic disturbances are identified in only a limited local time sector, the disturbances are assumed to extend only 500 km (radially, within the auroral oval) from stations in the region of disturbance. Similarly, the latitudinal extent of disturbances is inferred from measurements at stations along the central meridian. No attempts have been made to identify the nature of the disturbances. Some magnetic field perturbations correspond to localised AEI during periods of enhanced directly driven activity, while others may correspond to magnetospheric substorms and global disturbances. Ten periods of enhanced auroral activity, similar to those plotted in Figure 5.4, were identified in the auroral zone during April - October 1996. These periods are listed in Table 5.1.

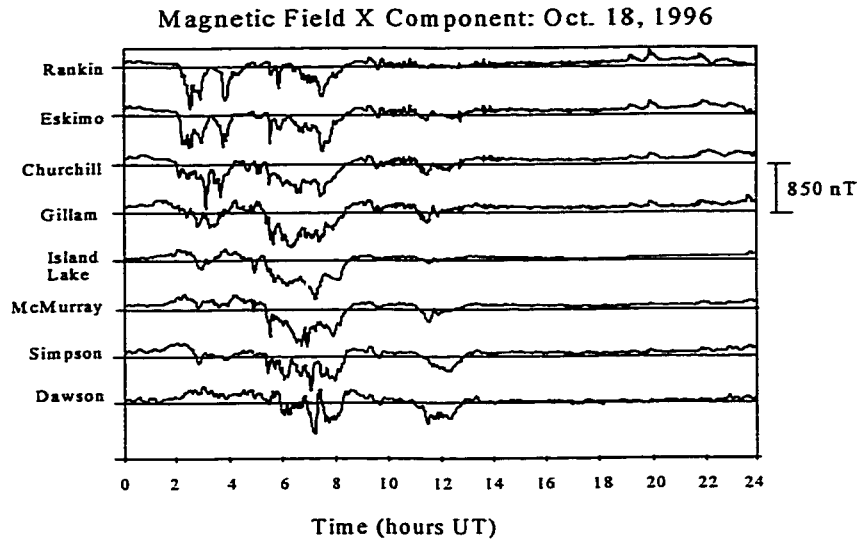


Figure 5.4 Magnetic field variations at MARIA stations: October 18, 1996. Quiet time values are plotted (straight lines) for reference purposes.

Table 5.1 Selected Days of Enhanced Auroral Activity

Month/Day (1996)	UT (hours, min)	Kp
4/17	0230-0500	4+
7/8	0700-0930	3-
8/23	0600-0700	3+
8/29	0400-0500, 0730-0900	6-, 5+
9/20	0530-0700,0830-0930	4, 4+
9/27	0530-0700	4
10/9	0430-0530	5-
10/18	0200-0400,0600-0800	5+, 5+
10/22	0830-1030	4
10/23	0130-0400	7+

Corresponding TEC values, and values of ionospheric delay, were calculated for the periods listed in Table 5.1, using GPS data from the Natural Resources Canada (NRCan) wide area network [*Caissy et al.*, 1996]. Locations of NRCan reference stations are shown in Figure 5.3. For comparison purposes, two sets of ionospheric delay data

were compiled simultaneously. These two data types will be referred to throughout several data analysis sections (Chapters 5, 6 and 7) of this thesis:

1) Observation pierce point located inside region of AEI/auroral disturbance. This type of data is herein refered to as "active" (i.e. observations plotted in Figure 5.1a).

2) Observation pierce point located outside region of AEI/auroral disturbance. This type of data is herein refered to as "quiet" (i.e. observations plotted in Figure 5.1b).

Note that the two types of data were collected *simultaneously*, during the intervals listed in Table 5.1, such that statistics for active and quiet data are valid for different regions of the ionosphere at the same times. While the data set consists of only 22 hours of observations, a full 7 months of observations were considered in selecting the final data set. The observations chosen here are representative of the largest auroral geomagnetic disturbances during those 7 months in 1996.

5.3 DATA PROCESSING

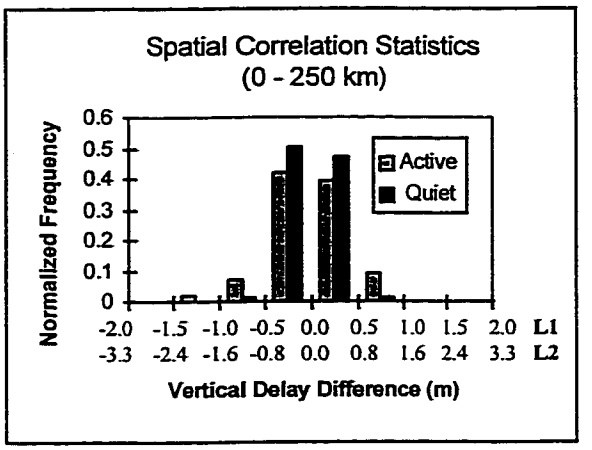
The GPS observations provided by NRCan were collected at 30 s intervals using TurboRogue receivers. Slant TEC values were calculated from a linear combination of the L1 and L2 range GPS observations (in codeless receiver mode) using Equations 3.1 and 3.2, and smoothed using the method of Section 3.3.1. An elevation cutoff angle of 20 degrees was used, in order to minimize multipath effects and obtain pierce point observations over a wide range of latitudes/longitudes.

As outlined in Chapter 3, satellite and receiver interchannel biases exist which can cause errors in TEC calculations as large as several metres. For purposes of this data set, all observed TEC values are corrected using satellite and receiver bias values provided by NRCan, which were computed using the entire 24 hours of data on a given day [cf. *Gao et al.*, 1994]. These biases have uncertainties on the order of 30 cm rms [*P. Héroux*, private communications]. Absolute values of ionospheric delay could therefore be calculated for individual observations, without necessarily employing a least squares grid algorithm to estimate biases simultaneously.

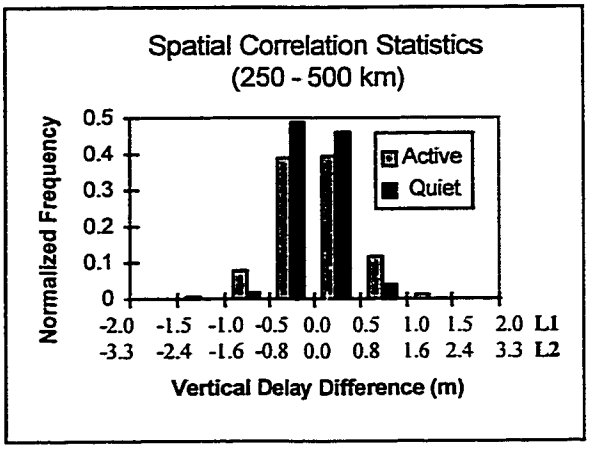
5.4 SPATIAL CORRELATION STATISTICS

In order to determine the spatial correlation of ionospheric delay values, the smoothed slant TEC values were mapped to vertical using the mapping factor (Equation 3.6). Corresponding pierce points were calculated for each observation on an ionospheric shell at 350 km altitude. At a given epoch, differences between the simultaneous vertical TEC values were calculated using all network observations classified as active. Four separate sets of statistics were compiled for the active data set, depending on the distance between pierce points: 0-250 km, 250-500 km, 500-750 km, 750-1000 km. Statistics were similarly compiled for the quiet data set. The spatial correlation statistics are shown in Figure 5.5, where vertical TEC differences have been converted to range delays for L1 and L2. Delay differences have been sorted into separate 0.50 metre bins (i.e. 0-0.5 m, 0.5-1.0 m,...) , for clarity.

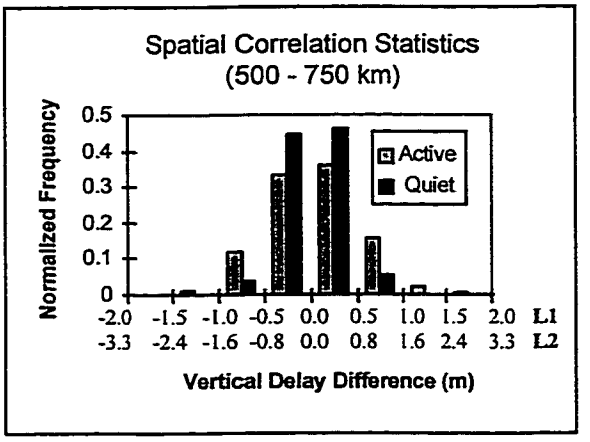
Figures 5.5a-b illustrate that delay differences are consistently larger for the active data, and increase with separation distance for both quiet and active data sets. The consistency of these results (active versus quiet) reflects consistency in selection of the active data set. Figure 5.6 shows the rms of vertical delay differences (L1), for both active and quiet regions, versus distance between pierce points.



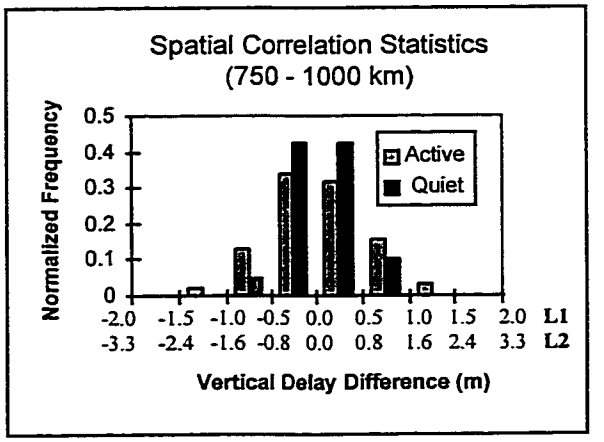
a) 0-250 km (sample size=5830)



b) 250-500 km (sample size=21342)



c) 500-750 km (sample size=15504)



d) 750 - 1000 km (sample size=12033)

Figure 5.5 Delay differences for varying pierce point separations.

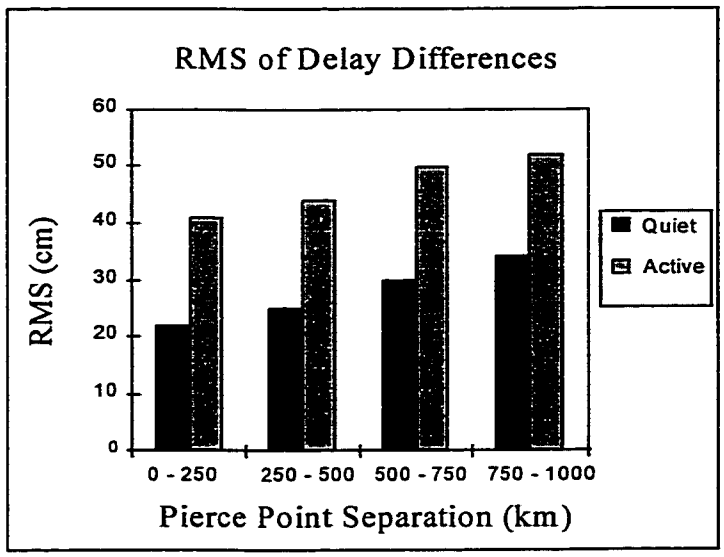


Figure 5. 6 Rms of vertical delay differences.

The active and quiet statistics differ considerably in magnitude. Delay differences are ≈20 cm larger in the regions of disturbance, as compared with regions outside of the disturbance. The difference (active versus quiet) is independent of pierce point separation (Figure 5.6), and is most likely due to both enhanced large-scale gradients and the existence of smaller-scale gradients associated with localised auroral disturbances. Figures 5.5a and 5.6 suggest that AEI scale sizes can be on the order of several hundred kilometres or less. This is consistent with the auroral GPS/TEC observations in *Hunsucker et al.* [1995]. Scale sizes associated with AEI are an important consideration when determining optimal ionosphere model grid spacing requirements in the auroral zone. Various grid spacings are evaluated in Chapter 6.

It is not possible to compile reliable statistics for various pierce point separations in the range 0-250 km, due to the smaller amount of available data. Figures 5.5 and 5.6 suggest, however, that such statistics are of interest for wide area grid modelling. Scale sizes in this range are further investigated using spectral analysis, in Section 5.6. A maximum spatial gradient of 0.058 TECU/km (≈ 1.0m/100km) was computed for the

active data set. This value corresponds to maximum values determined in *Chavin* [1997] for the dayside CONUS region, during "storm" conditions.

One further consideration in interpreting Figures 5.5 and 5.6 is the degree to which individual samples are statistically independent. Each set of statistics in Figure 5.5 has been compiled using data from the separate intervals listed in Table 5.1. For several of these intervals, consecutive observations from a given satellite-receiver pair are included in one set of statistics. With a 30 s sampling interval, these observations (samples) are correlated and are not entirely independent. In order to ensure that statistics in Figures 5.5 and 5.6 are reliable, a separate test was conducted using a subset of the observations with a 5 minute sampling interval. Statistics compiled from these observations are almost identical to those in Figures 5.5 and 5.6, and show the same trends in the data.

Finally, it must be noted that the data here represent ionospheric disturbances during a period of solar minimum (1996). As discussed in a later chapter (Section 6.8), the magnitude of such disturbances is expected to increase by a factor of ≈ 2 during intense storms at solar maximum.

5.5 TEMPORAL CORRELATION STATISTICS

In order to determine temporal variations in TEC during geomagnetically disturbed periods, Earth-fixed grid points were designated, and vertical TEC values were determined at these points using independent two minute data sets. At discrete two minute intervals, weighted averages were computed for each grid point, using only the previous two minutes of observations (mapped to the vertical at discrete points on the ionospheric shell). The grid points were spaced 3 degrees apart (in latitude and longitude), and only observations at pierce points within the eight surrounding grid points

were included in the weighted average for any given grid point. Each observation was weighted by the inverse of its distance from the grid point and 8-10 observations were typically included in each calculation.

For each grid point, the temporal changes in vertical TEC values were calculated for temporal delays of two, four, six, ... , sixty minutes. Rms values were then calculated for each set of temporal variations in TEC, as compiled for each temporal delay interval. Figure 5.7 shows the variance values plotted versus time. The slope in Figure 5.7 effectively represents process noise.

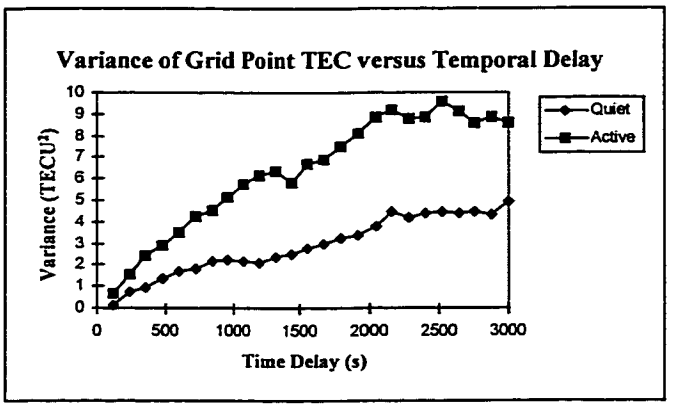


Figure 5.7 Vertical TEC variance, as a function of temporal delay.

From Figure 5.7, it is evident that variance values increase more rapidly for active, as opposed to quiet, ionospheric conditions. This follows from higher order temporal variations associated with particle precipitation events during geomagnetically disturbed periods (Figure 5.1a). Trends in the variance values (for time differences of less than 500 s) indicate process noise values of approximately 0.08 TECU/$\sqrt{s}$ during geomagnetic disturbances. The statistics for quiet ionospheric conditions in Figure 5.7 imply process noise values of approximately 0.04 TECU/$\sqrt{s}$, which corresponds to TEC

variations of 2.5 TECU/hour. This TEC rate of change is typical of TEC variations due to larger-scale gradients [cf. *Mannucci et al.*, 1993].

The process noise values derived from Figure 5.7 represent averages for both the quiet and active data sets. In reality, temporal variations for the active data set are distributed over a range of values. Figure 5.8 shows the distribution of temporal variations in grid vertical delay values (and vertical TEC) for the active data set, for temporal delays of two and five minutes. While the five minute differences are distributed over a larger range, neither the two minute, nor the five minute, variations exceed WAAS specifications (less than 2 m variation over two minutes).

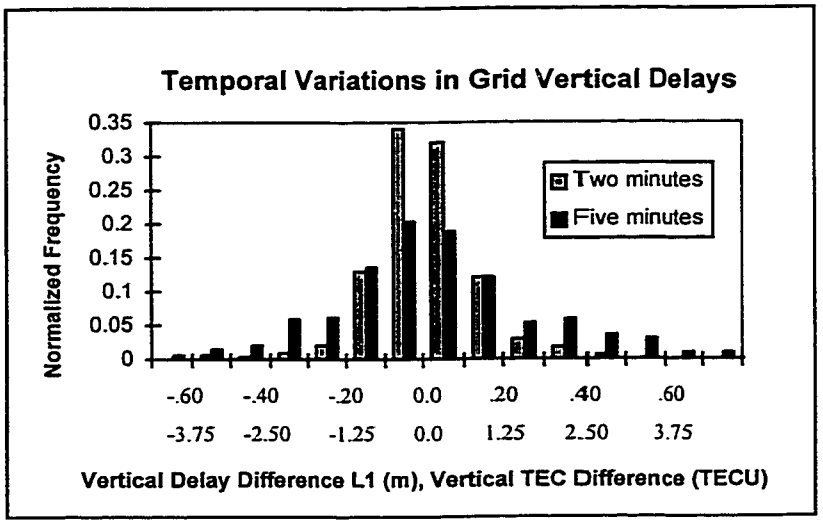


Figure 5.8 Temporal variations in grid vertical delays (VTEC), for 2 and 5 minute intervals.

It must also be noted that large short-term variations were often observed in time series of TEC, for a given satellite-receiver pair (i.e. plots in Figure 5.1a). While these time series actually represent combined temporal and spatial effects, an approximation of short-term temporal variations can be derived from a given series. Maximum variations

of up to 5 TECU over the 30 s sampling interval were observed for vertical TEC series. A more detailed study of TEC time series is included in the following section.

5.6 SPECTRAL ANALYSIS

In addition to the analyses presented in Sections 5.4 and 5.5, information concerning the spatial and temporal characteristics of TEC may be derived through spectral analysis of TEC time series. Such time series were shown in Figure 5.1. These TEC variations actually represent both spatial gradients and temporal variations, as the satellite-receiver line-of-sight moves through the ionosphere. Mathematically, these variations are represented by the convective derivative:

$$d/dt = \partial/\partial t + (\mathbf{v} \bullet \nabla)$$

where $\partial/\partial t$ denotes the partial derivative with respect to time, $\mathbf{v}$ is the satellite-receiver line-of-sight velocity in the ionosphere shell, and ∇ is the spatial gradient.

It is difficult to separate the effects of temporal variations and spatial gradients in a given satellite-receiver time series. It is useful, however, to determine *approximate* estimates of the time scales and scale sizes associated with auroral disturbances from such discrete time series through spectral analysis. This is done by deriving power spectral density profiles in the frequency domain, and wave number domain, for individual time series of VTEC. This approach allows resolution of smaller-scale variations than the analyses in Sections 5.4 and 5.5. The spectral analysis techniques are described in Appendix C.

5.6.1 Method

5.6.1.1 Time Domain to Frequency Domain

In order to derive statistics describing the spectral properties of TEC in the auroral region, the discrete Fast Fourier Transform (herein referred to as FFT) was computed for all satellite-receiver time series of smoothed VTEC in "active" or "quiet" regions, as identified in Section 5.2.

Separate statistics were computed for both the quiet and active data sets, for comparison purposes. For a given satellite-receiver time series, successive 128-point FFTs (i.e. N=128) were calculated. The 128 discrete samples represent a total time of 64 minutes (30 second sampling interval), with a Nyquist frequency of 17 mHz. The one-sided power spectral density (as defined in Appendix C) was calculated for each 128-point FFT, and the results at each discrete frequency averaged, to obtain representative statistics. In order to better approximate relative energy contributions from various parts of the power spectrum, mean squared amplitudes were also computed for harmonic components from several discrete frequency ranges. These calculations were carried out using a variation of Parseval's Theorem (Appendix C).

Corresponding zero-mean autocorrelation functions were also calculated for each 128-sample time series, using Equation C.17; these functions are defined for time lags in the range 0-1920 s (corresponding to one half of the overall sample length). Representative autocorrelation functions, for the active and quiet data sets, were computed by averaging the autocorrelation functions derived for each series in the given data set.

5.6.1.2 Spatial Domain to Wave Number Domain

In addition to analysis of TEC series in the time domain, spectral properties were also computed for TEC series defined in the spatial domain. Figure 5.11a shows corresponding distances, for each VTEC data point sampled at 30 second intervals, in the ionospheric shell at 350 km altitude. Note that the sampling interval, as defined in the spatial domain, is uneven. This results from the varying satellite-receiver line-of-sight velocities in the plane of the ionospheric shell, which depend on satellite elevation angle. These velocities range from 150 m/s at 20 degrees elevation, to $\approx$ 20 m/s for elevation angles greater than 80 degrees. For such unevenly sampled data, alternative techniques to the FFT must be used to map data into the frequency domain. An algorithm for spectral analysis of unevenly sampled data, herein referred to as the Lomb method, is outlined in Appendix C. This algorithm essentially computes least squares "best fits" to a linear combination of sines and cosines at each wave number resolved.

Due to the uneven sampling intervals, it is possible to resolve spectral characteristics in a range of wave numbers exceeding the average Nyquist wave number, and interpolate additional samples between data points. For the purposes of this analysis, an over-sampling factor of 4 is used. As for the FFT, discrete series of 128 sampled data points are analysed successively. In order to calculate statistics for entire data sets, the power spectral densities are binned in wave number increments and averaged.

In both the temporal and spatial analyses, no power spectral densities are plotted for a frequency of zero, and this harmonic component is excluded from amplitude estimates. The zeroth order component simply represents the mean value of a given TEC series, and the analyses of this section focus on relative variations only.

5.6.2 Results

5.6.2.1 Temporal analyses

Figure 5.9 shows the average power spectral density, as derived from VTEC series in the time domain, for both the active and quiet data sets. Figure 5.10 shows corresponding differences between the active and quiet power spectral densities. The power is concentrated in lower frequencies for both data sets, with the majority of power in frequencies less than 0.001 Hz. These frequencies correspond to larger scale trends in the data series (i.e. Figure 5.11b).

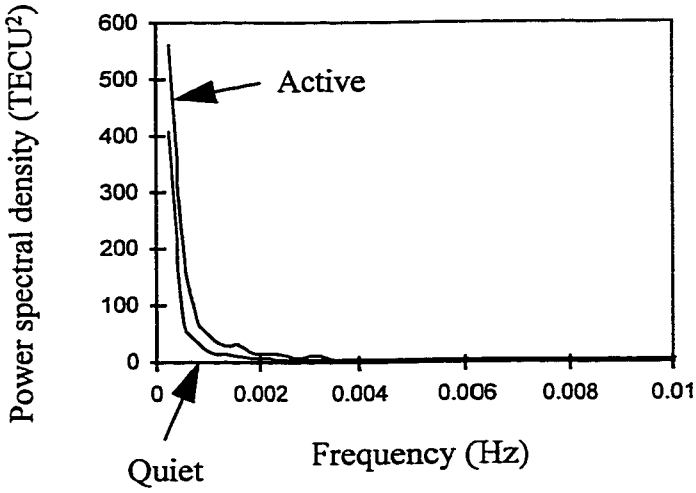


Figure 5.9 Power spectral density for the active and quiet data sets, in the frequency domain.

Of particular interest is the small peak in Figure 5.10, near 0.0015 Hz, with a period of ≈10 minutes. The increased power in this range, for the active data set, corresponds to auroral disturbances associated with AEI. Typical time-scales associated with auroral disturbances on the order of 10 minutes [*Rostoker*, 1991].

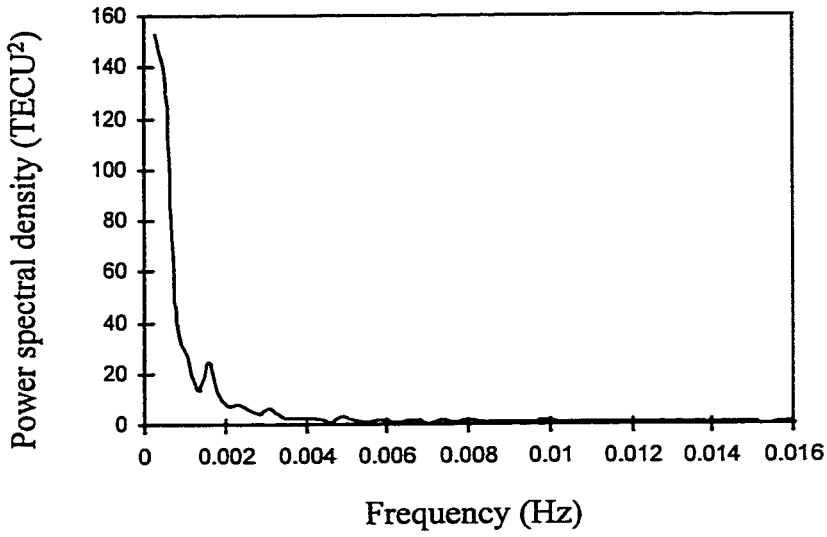


Figure 5.10 Differences in active and quiet power spectral densities, in the frequency domain.

As an example, Figure 5.11 illustrates the relative contributions of various harmonic components of VTEC, in a region of AEI, for a single series. Figure 5.11a shows the entire time series of VTEC (converted to L1 range delay), which includes all harmonic components, while Figure 5.11b shows relative contributions of harmonic components for two frequency ranges: < 0.001 Hz and 0.001-0.004 Hz. The latter range corresponds approximately to AEI time scales.

It is important to note that while the amplitudes associated with lower frequency components are relatively large, the lower frequency contributions are more slowly-varying and more easily modelled/resolved. The short-term variations in Figure 5.11b may have greater impact on wide area grid models, where nominal update intervals of 5 minutes are typical.

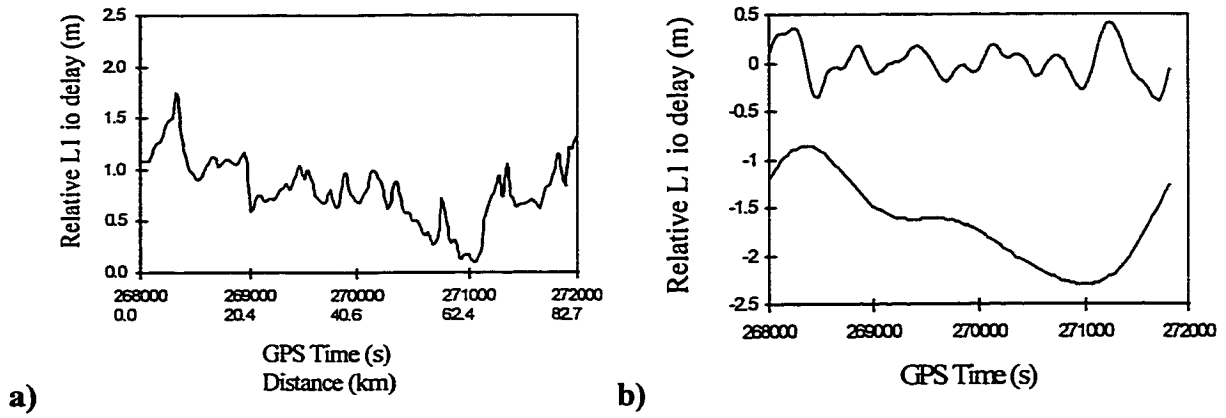


Figure 5.11 **a) Ionosphere range delay for SV 19, all harmonic components.**
b) Superposition of harmonic components of TEC, frequencies in ranges .001-.004 Hz (upper plot) and <.001 Hz (lower plot). The lower plot is offset, for comparison purposes.

It is useful to calculate average power spectral densities, in a given frequency range, in order to estimate the mean squared amplitudes of associated harmonic components, such as those plotted in Figure 5.11b. This may be done by making use of Parseval's Theorem (Appendix C). Figure 5.12 illustrates the averaged power spectral densities, for both the active and quiet data sets, in three frequency ranges. These correspond to mean squared amplitudes of the respective time series consisting of only those harmonic components in the given frequency range.

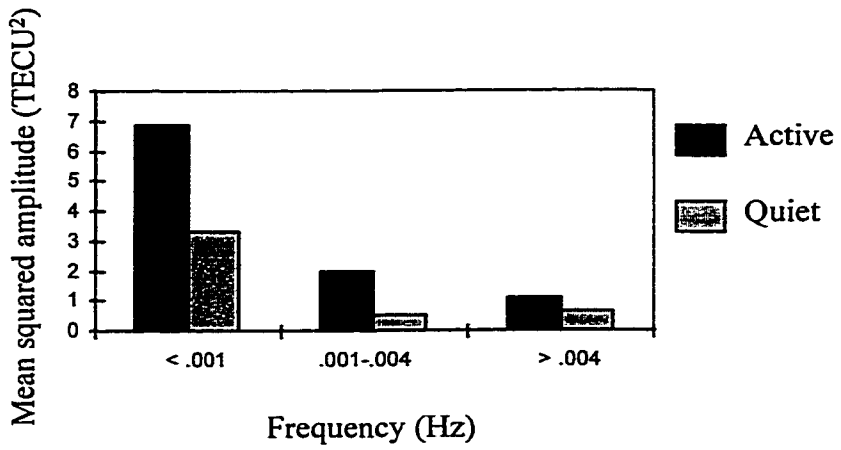


Figure 5.12 Mean squared amplitudes of harmonic TEC components, in three frequency ranges, for the active and quiet data sets.

Mean-squared amplitudes are consistently larger for the active, versus quiet, data set. In the range .001-.004 Hz, typical of auroral disturbances, the active data have a mean squared amplitude of 2.0 $TECU^2$. This translates to an ionospheric delay of approximately 23 cm at L1. Amplitudes at higher frequencies (> .004 Hz) are approximately 0.5-1.0 $TECU^2$, on the order of noise effects. Figure 5.12 suggests that short-term variations on the order of 20 cm (L1), over several minutes, are associated with auroral disturbances. This is consistent with the process noise values derived in Section 5.5.

Zero-mean autocorrelation functions, as derived for the active and quiet data sets, are shown in Figure 5.13. Both functions approximate first order Gauss-Markov processes. Theoretically, such processes fall off exponentially, approaching zero asymptotically. The negative autocorrelation values in Figure 5.13 result from periodicity associated with finite-length discrete data series. The maximum values of each function (at time lag 0) are equal to the process variance, or the mean power spectral density, as computed for the entire range of discrete frequencies. Correlation times of approximately 500 s are derived for both the quiet and active data sets, respectively. It is important to

note that autocorrelation values in Figure 5.13 are predominantly influenced by ionospheric (noise) effects for the active (quiet) data sets. Therefore, the correlation time of 500 s is representative of ionospheric variations only for the active data set.

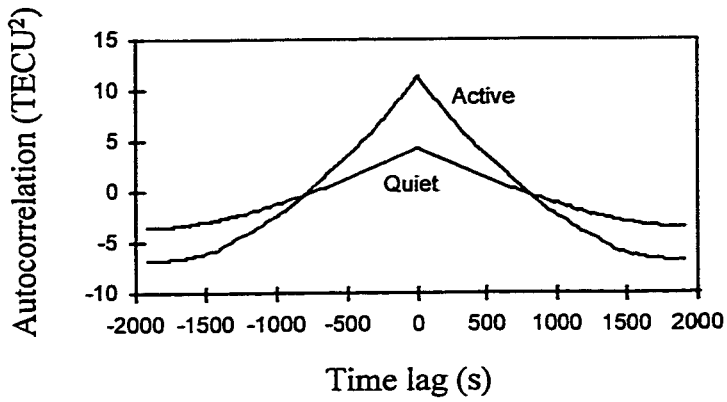


Figure 5.13 Discrete empirical autocorrelation function, as derived for all VTEC series in the active and quiet data sets.

5.6.2.2 Spatial analyses

Figure 5.14 shows the power spectral density derived from all satellite-receiver time series, for both the active and quiet data sets, where sampling intervals have been defined in the spatial domain.

The abscissa in Figure 5.14 is in units of wave number (km^{-1}), which are analagous to the frequency values in Figure 5.9. Variations in the power spectrum derive from oversampling inherent in the Lomb Transform method. The minimum wave number resolved is .0015 km^{-1} (wavelength 667 km). As in Figure 5.9, power spectral densities are larger for the active data set, particularly at lower wave numbers, (< 0.03 km^{-1}). Unlike Figure 5.9, however, the largest power is not found at the minimum resolved wave

number. This is in part due to the nature of the transform method used, and the binning (averaging) of values in discrete wave number ranges.

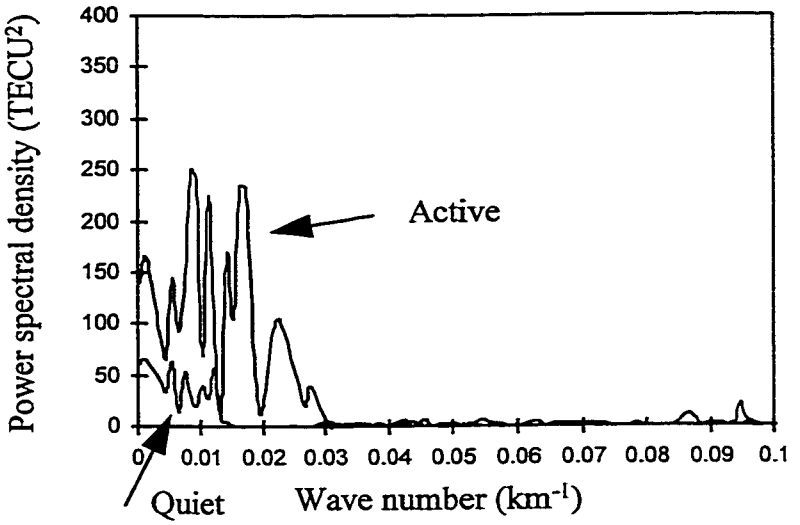


Figure 5.14 Power spectral density for the active and quiet data sets, in the wave number domain.

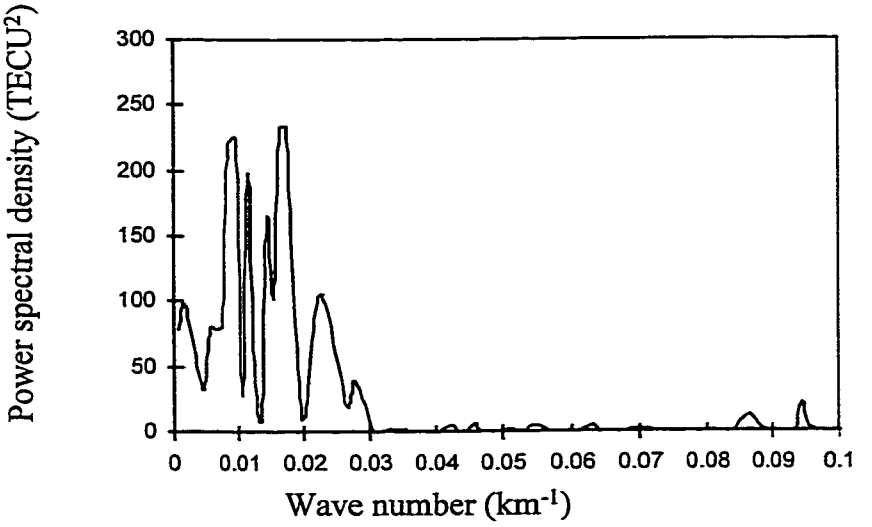


Figure 5.15 Differences in power spectral density for the active and quiet data sets.

Figure 5.15 shows differences in power spectral density between the active and quiet data sets. Larger differences are observed in the range .0015 to .03 km^{-1}, corresponding to wavelengths of 33 to 667 km; these values are consistent with typical scale-sizes observed for AEI [*Hunsucker et al.*, 1995].

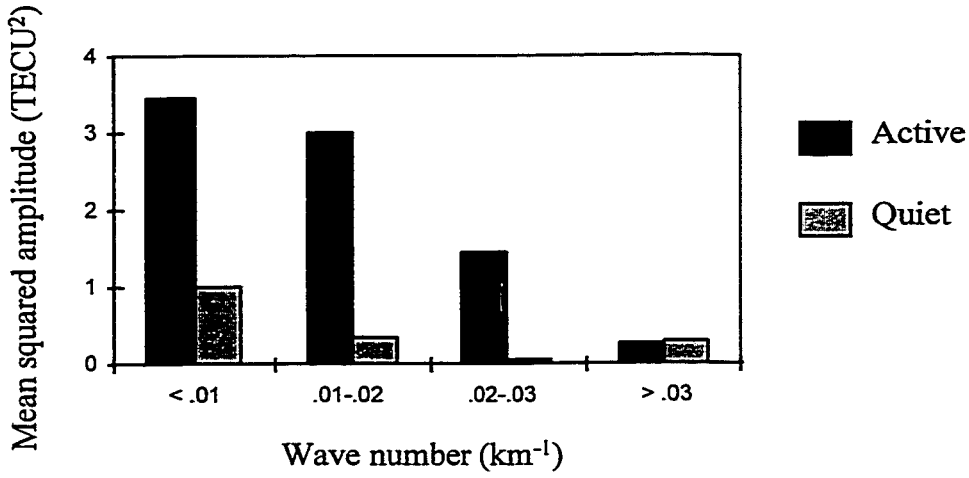


Figure 5.16 Mean squared amplitudes of harmonic TEC components, in four wave number ranges, for the active and quiet data sets.

Analogous to Figure 5.12, Figure 5.16 shows the average power spectral density, as calculated in various wave number ranges. The mean-squared amplitude of those harmonic components in the range .01-.02 km^{-1} is on the order of 3 $TECU^2$. This corresponds to an ionospheric delay of 28 cm (L1), with scale sizes in the range 50-100 km. These results suggest that small-scale ionospheric features with amplitudes on the order of 30 cm may exist in regions of auroral disturbances. This agrees with the statistics compiled in Section 5.4, where differences in ionospheric delay were approximately 40 cm (L1) for pierce point separations of less than 250 km. It must be emphasized, however, that the results of this section actually represent the combined effects of spatial and temporal variations as the satellite-receiver lines-of-sight traverse the ionosphere shell.

CHAPTER 6

INVESTIGATION OF GRID MODEL ACCURACIES

In Chapter 5, it was determined that small-scale temporal and spatial variations in TEC are associated with enhanced auroral activity. These irregularities are difficult to model, and the presence of such features has an impact on the accuracies of TEC estimates in the wide area ionosphere grid adjustment. In this chapter, grid accuracies are assessed during periods of auroral disturbances, using the grid model introduced in Chapter 3 and the active data set from Chapter 5. The tests conducted in this chapter were designed with regard to existing wide area ionosphere grid algorithms, where varying values of grid spacing and process noise are specified [i.e. *Mannucci et al.*, 1993; *Komjathy and Langley*, 1996; *Mueller and Davis*, 1998]. It is intended that test results do not necessarily provide definitive model parameters for all grid models, but that guidelines are established for implementing ionosphere grid models in the auroral region.

The grid model is tested using various grid spacings and system model parameters (process noise values). Accuracies are also assessed during a non-disturbed period, for comparison purposes. A detailed analysis of selected test results is illustrated for a selected period of enhanced auroral activity, and grid accuracies are assessed during a non-disturbed period, for comparison purposes. Compiled statistics (for the entire active data set) are then presented, and the effects of shell height assumptions and latency are discussed. Interpretation of the results, with respect to existing error bounds, are included for WAAS, and requirements for reference station coverage are investigated. Results are also used to predict the impact of auroral disturbances on users at solar maximum.

6.1 MODEL IMPLEMENTATION AND TESTING METHOD

For the tests conducted in this chapter, grid TEC maps are generated, and model accuracies tested, for the ten periods of disturbed auroral activity identified in Chapter 5 (Section 5.2, Table 5.1). The grid adjustment incorporates observations from all NRCan reference stations (Figure 5.3), using the Kalman filter formulation described in Chapter 3 and Appendix A.

6.1.1 Reference Frame

Although wide area user requirements often specify Earth-fixed grid points (i.e. WAAS [*FAA*, 1994; *FAA*, 1997]), many grid algorithms compute vertical TEC values in a quasi Sun-fixed coordinate system, where the grid points are defined in local time (hour angle relative to the Sun) and geomagnetic latitude [i.e. *Wilson and Mannucci*, 1993; *Gao et al.*, 1994; *Mueller and Davis*, 1998]. Estimates of TEC at overlapping Earth-fixed grid points can then be interpolated from updates of the "Sun-fixed" grid. This method generally allows better estimation of TEC in the dayside regions (the ionosphere being controlled by both the geomagnetic field and solar radiation) where large-scale spatial gradients vary more slowly in local time coordinates. For the purposes of this thesis, a solar-magnetic reference frame is chosen, where TEC is modelled in geomagnetic latitude and magnetic local time. The magnetic local time (MLT) at a given grid point (or observation pierce point) is defined as the flare angle between two dipole meridional planes, one containing the given grid point, the other containing the subsolar point. Calculation of these coordinates requires conversion from geographic to geomagnetic coordinates. Relevant transformation equations are given in Appendix D. While this coordinate system is not strictly "fixed" with respect to the Sun, the dipole axis being offset from the Earth's rotation axis, variations in the reference frame are slow with respect to filter correlation times in the grid adjustment.

This coordinate system is preferable for auroral phenomena, where the auroral oval is centred approximately on the geomagnetic pole. This is demonstrated in Figure 6.1, where Earth-fixed grid points are compared with solar-magnetic grid points (projected into geographic coordinates), for a 5 degree grid spacing. The grid point coverage in Figure 6.1a is approximately aligned with the auroral oval boundaries in Figure 4.1.

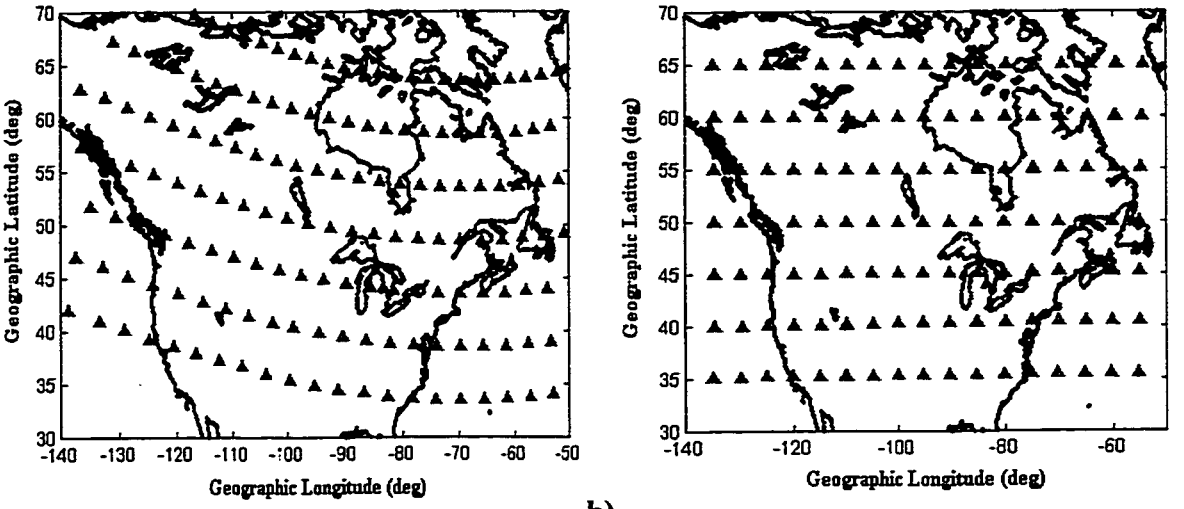


Figure 6.1 Distribution of ionosphere grid points (5°×5° grid) for a) solar-magnetic coordinates, and b) geographic coordinates.

An additional advantage of using the solar-magnetic reference frame, versus Earth-fixed coordinates, is the greater coverage of observation pierce points in the ionosphere shell. While the number of observations is equal in both coordinate systems, observation pierce points are derived over a larger range of solar-magnetic longitude (local time) than (Earth-fixed) geographic longitude, as the satellite-receiver line-of-sight traverses the ionosphere shell. Figure 6.2 shows observation pierce points for all line-of-sight observations in the NRCan network during a three hour period. For observations

made in a solar-magnetic reference frame, there is greater east-west coverage; this allows a more extensive distribution of observation pierce points.

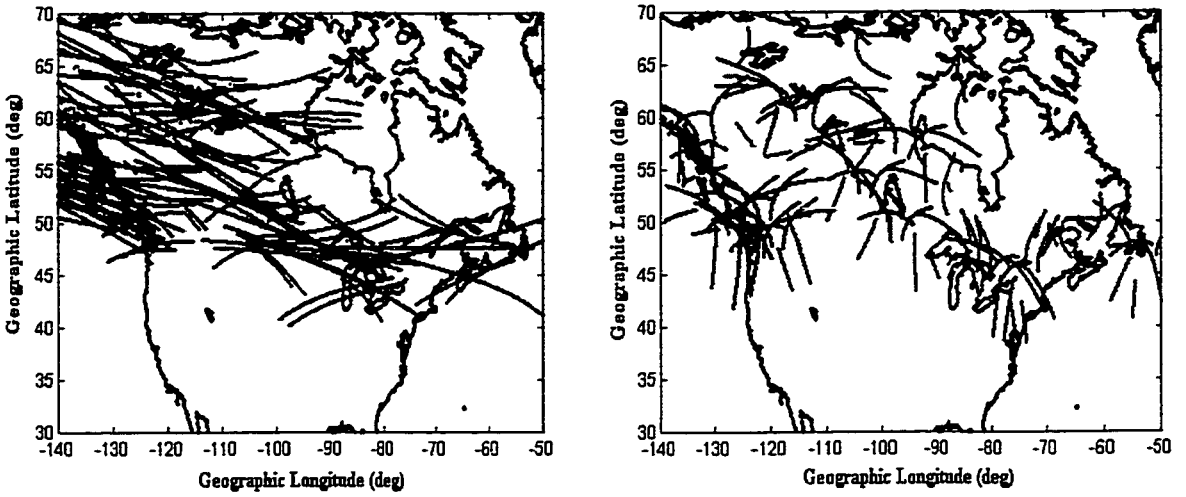


Figure 6.2 Distribution of observation pierce points over a 3-hour period for a) solar-magnetic coordinates, and b) geographic coordinates.

In this chapter, only accuracies of the underlying solar-magnetic grid model are investigated. It is assumed that these accuracies are similar to those for an Earth-fixed grid derived/interpolated from the solar-magnetic estimates of TEC.

6.1.2 Testing Method

Grid VTEC maps are generated using, alternately, spacings of 3°×3°, 5°×5°, and 10°×10°. These spacings are chosen to represent possible grid spacings within a range of typical values. Grid spacings of less than 3 degrees are also used in some models [i.e. *Mueller and Davis*, 1998] but, in order to derive reliable statistics here (with a relatively large sample size), 3 degrees is the minimum grid spacing investigated. Additionally, for each grid spacing, separate tests are conducted using two sets of system model parameters

(Equation 3.8 and 3.9), one set being typical of average ionospheric conditions, the other characterizing ionospheric temporal processes during enhanced auroral activity. The "average" or "quiet" process parameters are chosen as $1/\beta = 15$ minutes and $\sigma = 0.75$ TECU. These values are consistent with the statistics of Section 5.5, and correlation times in *Mannucci et al.* [1997].

During the test periods, updated grid TEC maps are generated at five minute intervals, using the previous five minutes of observations from the NRCan network. These GPS observations are available at 30 s intervals, with 6-8 satellites typically observed from a given reference station. An elevation cutoff angle of 20 degrees is assumed. Accuracies of these maps are assessed two different ways:

1) The grid adjustment is carried out excluding observations from one reference station (designated the "test" station). Vertical ionospheric delay corrections are then extracted from the grid VTEC estimates, at the test observation pierce points, using the multiquadric weighting functions (Equation 3.7). Differences between grid corrections and the observed VTEC values are computed and converted to vertical range delays (L1). The entire data set is tested in this manner using each test station alternately. Only test observations with elevation angles greater than 40 degrees are included in accuracy statistics.

2) The grid adjustment is carried out using observations from all stations in the network. The grid VTEC estimates are then used to calculate measurement residuals (corrections to the observations), which are converted to L1 range delays. Calculation of the measurement residuals, which are a by-product of the Kalman filter adjustment, is described in Appendix A. This test is similar to the Type 1 test, except that error estimates are determined using "test" observations which are also included in the grid adjustment. There is no independent "truth" data. This type of test reflects random observation errors and deficiencies in the filter formulation.

The two types of testing are herein referred to as "Type 1" and "Type 2" testing, respectively. In the Type 1 testing, an independent data set is used to assess the accuracy of the model. It is assumed that these accuracies are on the order of those observed by an independent single frequency user of the wide area network. Note, however, that grid accuracies are degraded by excluding a reference station from the grid adjustment. This condition is mitigated by using only Churchill, Flin Flon, Yellowknife, Holberg, Albert Head, Penticton, and Williams Lake as test stations. The remaining reference stations are not included in this type of testing, since the density of observation pierce points is significantly compromised in localised regions when these stations are excluded from the adjustment. Additionally, a given test observation is included in the accuracy statistics only when the surrounding grid points have been updated with no fewer than ten new measurements (i.e. no fewer than ten total measurements available from the four surrounding grid cells, at the given grid point, over the last five minutes). The threshold value of ten is derived from statistics describing the growth of range delay residuals - these statistics are presented in Section 6.6.

The test observations are collected over a five minute interval. Testing is then conducted using updated grid values from the beginning of this five minute interval. In real implementation of the grid model (i.e. WAAS), users will receive grid updates at approximately five minute intervals. These grid values are used throughout the following five minute interval, until the next grid updates are received. In such cases, errors may arise from latency effects. The Type 1 testing, as described above, allows for inclusion of these effects.

If the filter has been designed optimally, the Type 2 test (residuals) provides an indication of how much random error exists in a given observation. Ideally, magnitudes of the residuals should be on the order of measurement and system model uncertainties. Residuals can also, however, reflect how "poor" the model is (i.e. the residual absorbing an unmodelled portion of the measurement). In the analysis presented here, residuals with

magnitudes exceeding typical measurement accuracies are assumed to reflect shortcomings in the grid model. Unlike the first type of testing, however, estimates of grid accuracies are derived from observations included in the grid adjustment; this type of analysis can underestimate the true errors at an independent user site.

6.2 DETAILED ANALYSIS - OCTOBER 23, 1996

In order to illustrate some of the grid accuracy analyses conducted in this chapter, detailed results from October 23, 1996 are described in this section. Overall results, as derived for the entire 22-hour active data set are compiled and presented in later sections.

6.2.1 TEC Maps

Prior to assessing the grid accuracies, it is useful to consider some typical VTEC maps derived from the grid adjustment, for the Canadian sector. In Chapter 5, a detailed analysis of nightside auroral TEC irregularities was presented. Such features were characterized by transient small-scale spatial gradients. Large-scale gradients also exist near the dawn and dusk terminators. These are repeatable features which are observed in day-to-day TEC maps.

In order to generate sample TEC maps, NRCan observations from October 23, 1996 were used in the ionosphere grid adjustment. Magnetometer data for this day is shown in Figure 6.3, where significant auroral activity takes place 0130-0400 UT. Locations of the magnetometer stations are shown in Figure 5.3. A plot of vertical TEC observations from Fort Churchill is shown in Figure 6.4, where the VTEC variations are highly correlated with nightside magnetic field perturbations.

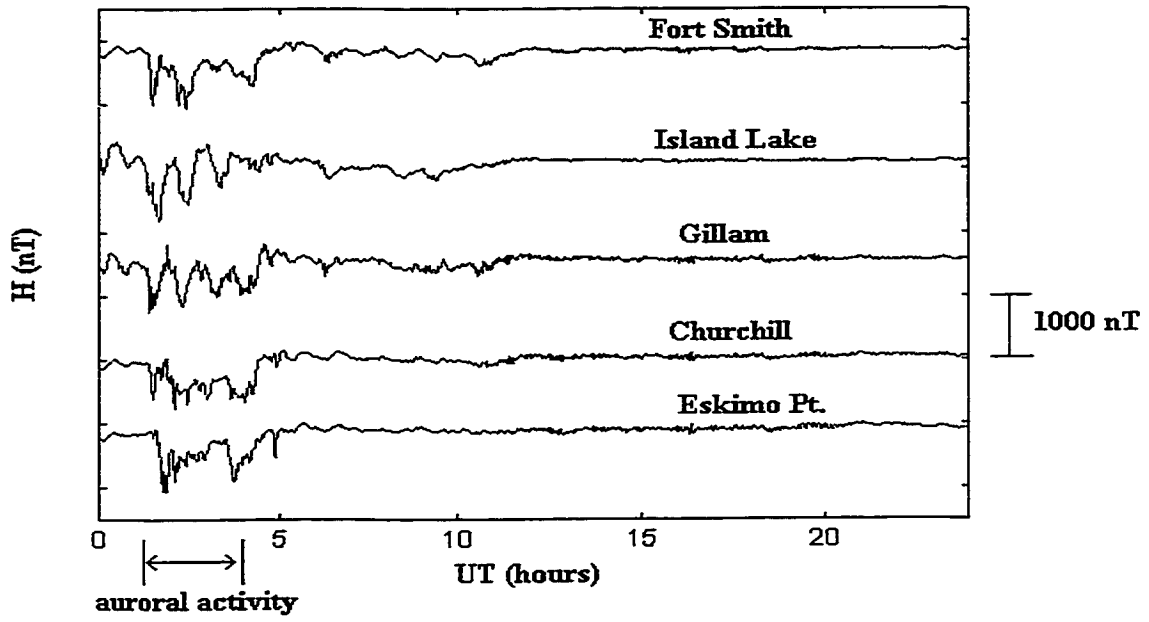


Figure 6.3 North-south magnetic field component (positive northwards), at five stations in the CANOPUS MARIA array October 23, 1996 (data courtesy of the Canadian Space Agency).

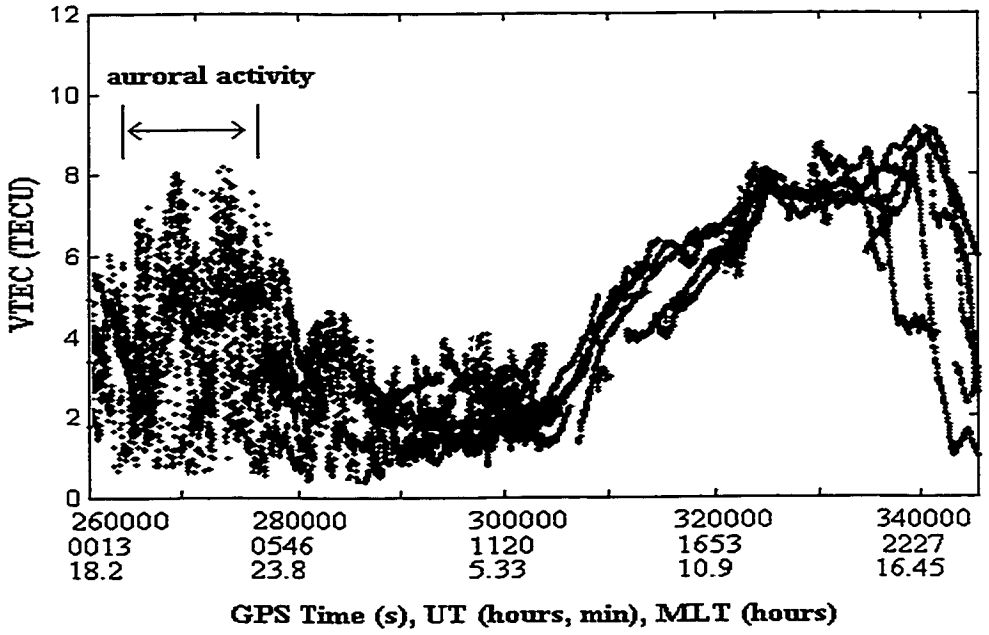
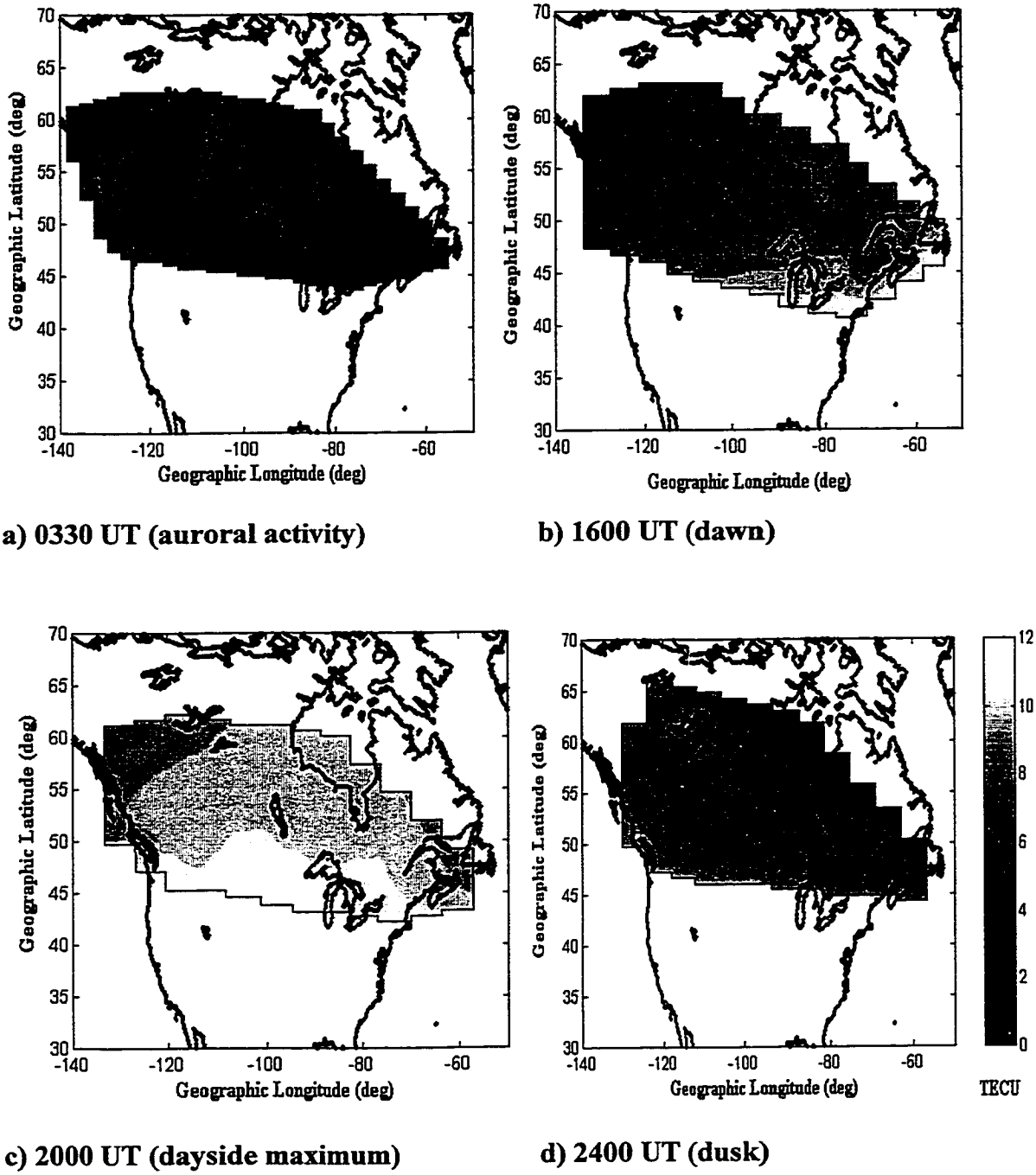


Figure 6.4 VTEC series observed from Fort Churchill on October 23, 1996.

From Figures 6.3 and 6.4, it is observed that auroral activity decreases by ≈0500 UT. It is therefore assumed that ionospheric TEC maps for periods later than this represent typical TEC configurations. The grid model used to generate these maps includes the average process parameters of $1/\beta = 15$ minutes and $\sigma = 0.75$ TECU in the system model. These model parameters are not necessarily optimal, but are sufficient for the purposes of this section, where the TEC maps are intended to illustrate only general features of ionospheric TEC. A grid spacing of 5°×5° is also used.

The series of maps in Figures 6.5b, c, and d show vertical TEC distributions at 1600, 2000, and 2400 UT. At these times, central Canada is located at approximately 1000, 1400, and 1800 local time. These maps therefore illustrate the spatial distribution of TEC near the dawn terminator, the dayside maximum, and the dusk terminator, respectively. Large-scale east-west gradients are observed in Figures 6.5b and d. TEC increases eastward near dawn, approaching the dayside maximum (Figure 6.5c), while TEC decreases eastward near dusk, into the nightside. North-south gradients are also observed in all three figures, particularly near the dayside maximum. This is consistent with global observations of north-south TEC gradients [*Klobuchar*, 1996]. Such large-scale gradients are generally well-modelled using GPS observations from the NRCan network, where the density of reference stations allows resolution of these features.

In contrast to Figures 6.5b, c, and d, Figure 6.5a shows a map of VTEC irregularities at 0330 UT. From Figure 6.3, this time coincides with a period of enhanced auroral activity. The overall magnitude of TEC in Figure 6.5a is much lower than the dayside magnitudes in Figure 6.5c, and gradients are localised over relatively small regions. These features are associated with the auroral disturbances observed near Fort Churchill in central Canada. It must be noted that these features reflect model estimates of grid TEC values at regularly spaced intervals (5°×5° grid) and most likely smooth over higher order features which are not resolved well by the model. The actual TEC distribution is more irregular.



a) 0330 UT (auroral activity)

b) 1600 UT (dawn)

c) 2000 UT (dayside maximum)

d) 2400 UT (dusk)

Figure 6.5 VTEC maps for October 23, 1996, converted to geographic coordinates.

6.2.2 Grid Accuracies - Various Grid Spacings

For a comparison of grid accuracies achieved for the various grid spacings, average system model parameters ($1/\beta$ = 15 minutes, σ = 0.75 TECU) are assumed. Figure 6.6 shows the test results for a 3°×3° grid. The "+" symbols represent accuracies derived from the active data set, during the period of auroral disturbance (0130 - 0400 UT) in a high latitude region. The "o" symbols represent all other test results; it is assumed that these results, for the period 0500 UT onwards, are more typical of average (quiet-time) accuracies. Rms values for L1 range delay accuracies are derived from all estimated accuracies over a given five minute interval, for a given test station (Type 1 test). The results for all seven test stations are plotted in Figure 6.6a. Rms values for the residuals are derived similarly (Type 2 test), incorporating all residual values computed for a given 5 minute interval in one test statistic (Figure 6.6b).

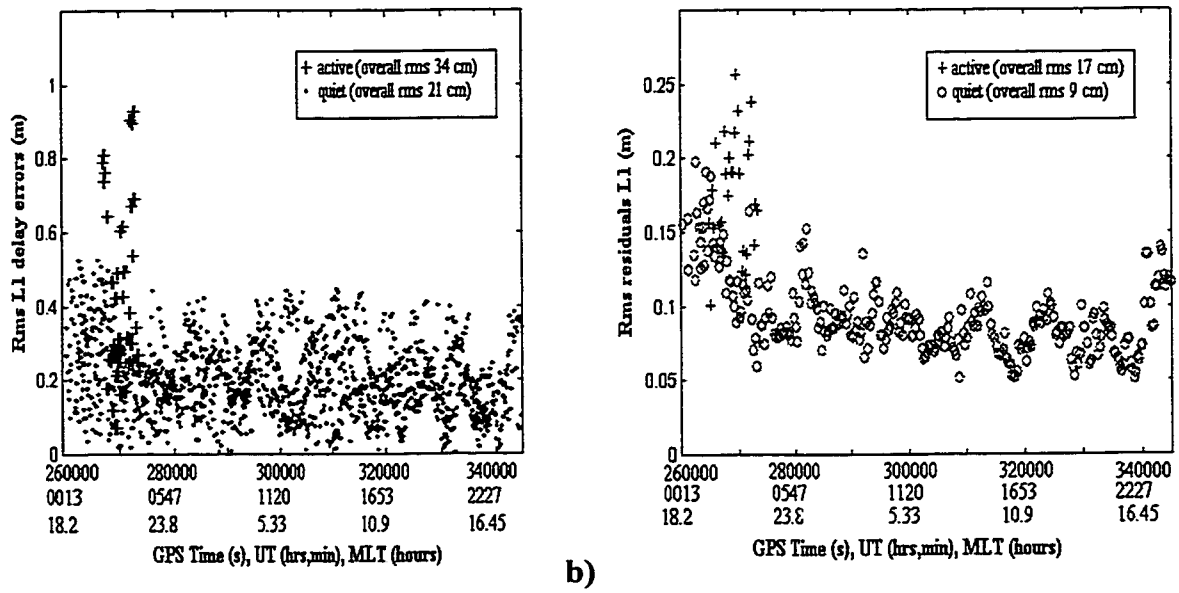


Figure 6.6 Rms values of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test) for October 23, 1996. The grid spacing is three degrees. MLT has been calculated approximately for central Canada.

In Figure 6.6a, the active data set has an overall accuracy of 34 cm rms, with errors as large as 90 cm, while the quiet data has an overall accuracy of 21 cm rms. In Figure 6.6b, testing of the residuals indicates accuracies of 17 cm rms for the active data, and 9 cm rms for the quiet data set. While magnitudes of the accuracies in Figure 6.6 are different for the two testing methods, both indicate degraded grid accuracies for the active data set. Larger rms values on the nightside result from the presence of auroral disturbances and associated irregularities in electron density. In some cases, quiet rms values exceed active rms values. This likely results from steep gradients at the equatorward boundary of the oval, near the main trough. These effects are investigated briefly in Section 6.2.3. Note that the large-scale gradients in Figure 6.5 are well-modelled, with dayside residuals of 5-12 cm. These values are on the order of random measurement errors.

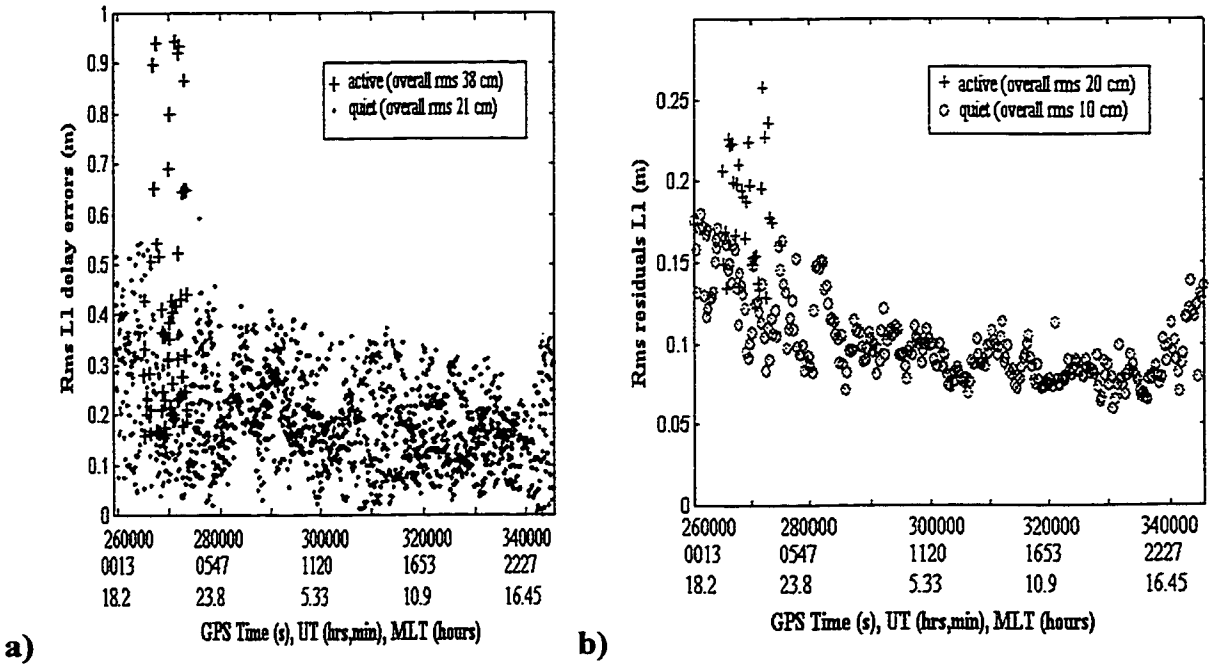


Figure 6.7 Rms values of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test) for October 23, 1996. The grid spacing is five degrees. MLT has been calculated approximately for central Canada.

Figure 6.7 shows results analagous to those in Figure 6.6, for a 5°×5° grid. In this case, accuracies are degraded for both the quiet and active data sets, versus the three degree grid spacing. In particular, in Figure 6.7a, the active data set has rms values 4 cm larger for the larger grid spacing. Residuals are also increased by several cm for the larger grid spacing.

For a 10°×10° grid (Figure 6.8), accuracies are further degraded. For the Type 1 testing method, accuracies are 48 cm for the active data set, which is 14 cm larger than for the three degree spacing; residuals are 6 cm larger. The maximum errors are also increased, for both the Type 1 and Type 2 testing, with errors in the L1 range delay being greater than one metre for a localised region.

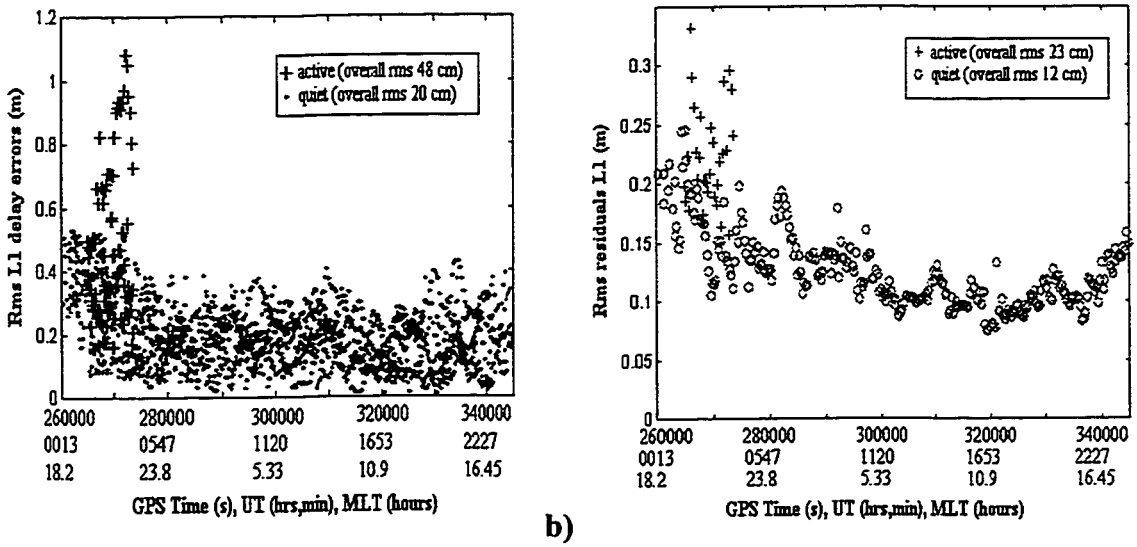


Figure 6.8 Rms values of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test) for October 23, 1996. The grid spacing is ten degrees. MLT has been calculated approximately for central Canada.

For all three grid spacings tested here, quiet rms values (Type 1 test) are approximately 20 cm; these accuracies are consistent with typical vertical ionosphere

delay accuracies observed by *Bertiger et al.* [1997]. Note that these quiet accuracies are not strongly dependent on grid spacing. This is consistent with Figures 6.5b, c and d, where scale sizes of TEC gradients are large compared with all three grid spacings. Errors in range delay accuracies, for the active data set (all three grid spacings) can reach values on the order of 1.0 m, which are five times larger than typical error bounds.

Larger residuals, for the active data set, are indicative of model deficiencies and suboptimal filter performance. In particular, the simple measurement model is based on an interpolation scheme (Section 3.3.2) which is not necessarily appropriate for localised irregularities in TEC. Additionally, the system model parameters may not be representative of the ionospheric processes. This issue is addressed in Section 6.3.2 and Chapter 7.

6.2.3 Development of Main Trough

In all three Figures of Section 6.2.2, a subset of the "quiet" nightside accuracies exceed dayside rms values. This results from two effects: 1) enhanced auroral activity being included in the quiet data set, and 2) poor modelling of steep gradients near the main trough (Section 4.6). The second effect is most predominant, since the quiet data set was derived almost exclusively from observations near or below the equatorward boundary of the auroral oval, and degraded "quiet" accuracies are observed prior to the onset of auroral disturbances (≈0130 UT).

An example of the main trough is shown in Figure 6.9, where TEC values are observed to increase by a factor of two over several degrees latitude. Note the contrast between Figure 6.9 and Figure 6.5d, where each map has been derived for similar local times. The gradients in Figure 6.5d, during a less active period of ionospheric activity, are considerably less steep.

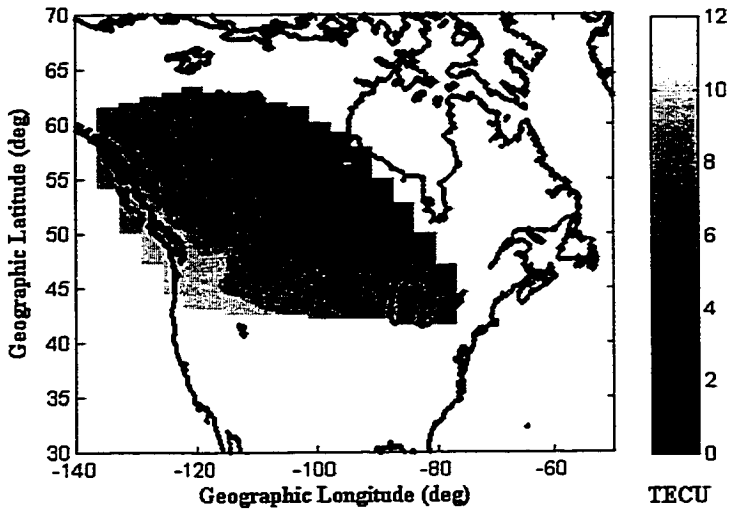


Figure 6.9 Map of VTEC for 0100 UT October 23, 1996. The steep gradients are associated with the main trough at latitudes below the auroral oval.

Larger residuals are observed in regions near the steeper gradients in Figure 6.9 (i.e. 15-20 cm residuals at 0100 UT - Figure 6.8b). While these values are larger than typical quiet time accuracies, they are not as poor as the larger errors (≈30 cm - Figure 6.8b) associated with high latitude auroral disturbances at later times.

6.3 ANALYSIS OF ACTIVE DATA SET

From the results of Section 6.2, it is observed that grid accuracies are degraded near regions of auroral disturbances. In order to derive more extensive and reliable statistics, grid accuracies are assessed for all periods included in the active data set (as identified in Section 5.2) and the results compiled. In order to analyse all the data from different days together, results have been binned in 30 minute intervals of local time and averaged. In this section, the grid model is tested using various grid spacings and system model assumptions.

6.3.1 Grid Accuracies - Various Grid Spacings

Figure 6.10 shows overall grid accuracies for the three different grid spacings. The grid model used to generate these plots includes average system model parameters ($1/\beta$ = 15 minutes and σ = 0.75 TECU). Consistent with the statistics from Section 6.2, better accuracies are observed for the smaller grid spacing, for both the Type 1 and Type 2 testing. In all cases, however, accuracies are a factor of 1.5-2 worse than those for quiet conditions. This is consistent with the scale size of localised geomagnetic disturbances (several hundred kilometres), as estimated in Chapter 5 (where 1° latitude ≈ 110 km, and 1° longitude ≈ 60 km in the auroral region). Results are particularly poor for the 10 degree grid spacing, where distances between grid points are ≈1000 km north-south and ≈600 km east-west. Accuracies are similar for all local times, with slightly larger values in the pre-midnight sector.

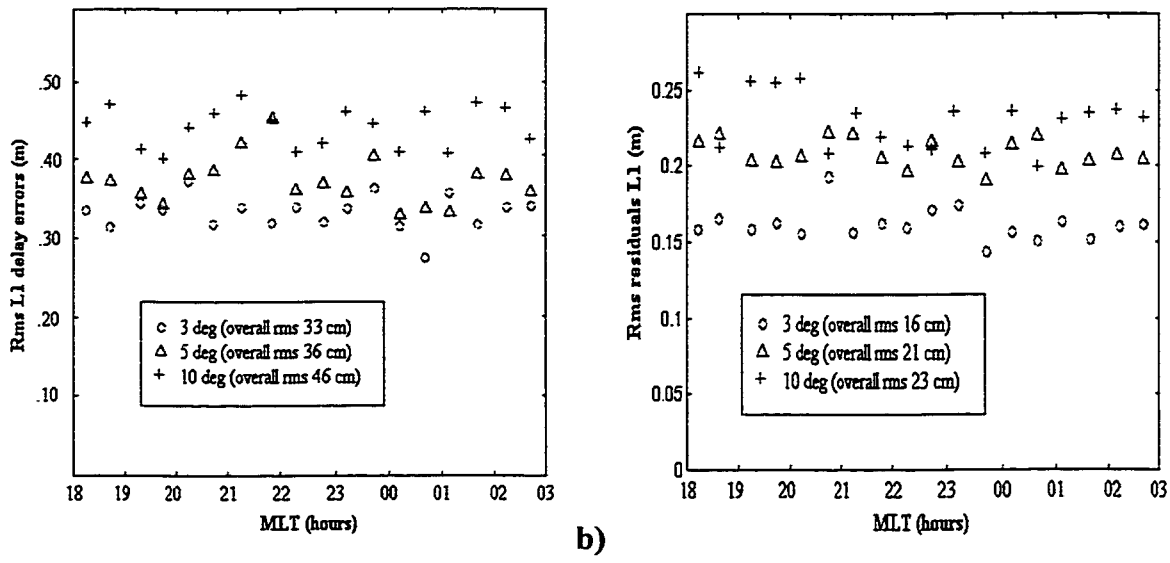
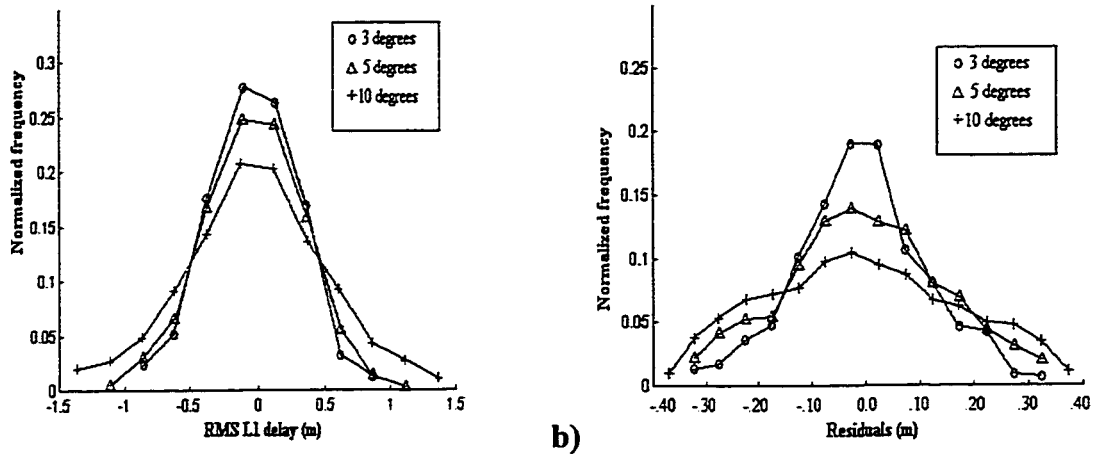


Figure 6.10 Rms values of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test) for the entire active data set.



a) b)

Figure 6.11 Probability distributions of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test).

Figure 6.11 shows probability distributions of results from Figure 6.10, where the errors for a 10 degree grid spacing are distributed over a larger range of values. Maximum overall errors of 1.0-1.50 m are consistent with those illustrated in Figures 6.6a, 6.7a, and 6.8a. These represent accuracies a factor of approximately 6 worse than those for non-disturbed periods.

For improved accuracies in the auroral zone (during periods of enhanced auroral activity), grid spacings of 5 degrees are preferred, and grid spacings of 3 degrees or less are best. This is in contrast to non-disturbed periods during which accuracies are similar for all three grid spacings. It is intended that the relative accuracies derived in this Section serve as guidelines for determining the density of reference stations (grid spacing) required to achieve desired WADGPS correction accuracies. An interpretation of these results, with respect to WAAS error bounds, is given in Section 6.7.

It must be noted, however, that the results presented here are derived under a condition that sufficient observations were available to update grid values. The

implementation of a smaller grid spacing does not necessarily guarantee better results under all circumstances. For a sparse network of reference stations over a large area, the density of observation pierce points may limit the minimum grid spacing. An optimal grid spacing depends on both the scale size of ionospheric TEC gradients, and the density of reference stations. The spacing of reference stations required to fulfill accuracy requirements will be discussed in Section 6.6.

6.3.2 Comparison of System Model Parameters

Results of the previous sections have been calculated using average values of system model parameters (Equations 3.8 and 3.9). In the absence of any information about local ionospheric processes, such typical values are generally used [i.e. *Mannucci et al.*, 1993]. These assumptions are not necessarily correct for regions of auroral disturbances, however. The total electron content can vary rapidly in such regions, and statistics describing these temporal variations were derived in Chapter 5. In this section, grid accuracies are computed for the active data set, using values of process noise (system model parameters) consistent with auroral disturbances. The system model parameters for active conditions are chosen as $1/\beta$ = 120 s and σ = 1.50 TECU, using the results of Sections 5.5 and 5.6. Figure 6.12 shows a comparison of rms values (Type 1 testing) for the 3 degree grid spacing, derived for both "quiet" ($1/\beta$ = 15 minutes, σ = 0.75 TECU) and "active" process parameters.

In Figure 6.12, it is observed that overall accuracies are improved by approximately 5 cm rms (for the 3 degree grid spacing) for the higher process noise values. Complete results, for all three grid spacings, are listed in Tables 6.1 and 6.2. These tables essentially summarize the results of this section and Section 6.3.1.

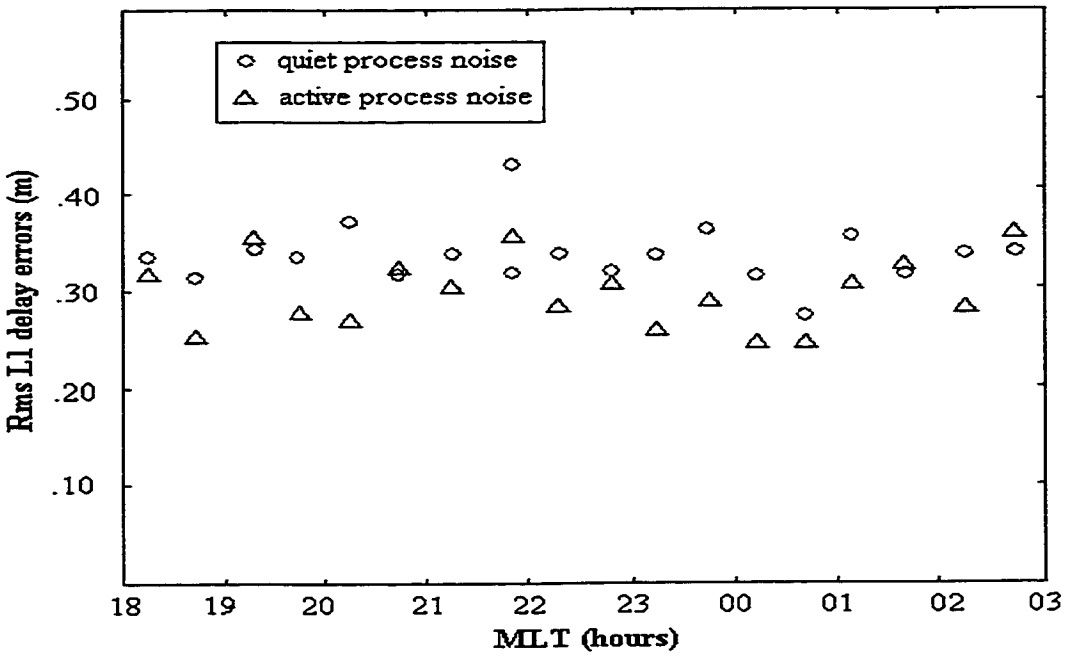


Figure 6.12 Type 1 test results for two sets of system model parameters (3 degree grid spacing).

Table 6.1 Rms Grid Accuracies (L1) for Type 1 Testing

RMS (m)	GRID SPACING		
System Model Parameters	3 degrees	5 degrees	10 degrees
quiet	.33	.36	.46
active	.28	.33	.47

Table 6.2 Rms Residuals (L1) for Type 2 Testing

RMS (m)	GRID SPACING		
System Model Parameters	3 degrees	5 degrees	10 degrees
quiet	.16	.21	.23
active	.14	.19	.22

Note that the results for larger process noise ("active" system model) are not significantly improved in Table 6.1 for the larger grid spacings, as compared with accuracies for the smaller process noise value. This may be due to the dominant effects of unmodelled spatial gradients (scale sizes of a few hundred km and less) near the test observation point. Overall, however, the results in Tables 6.1 and 6.2 imply that optimal grid performance may be achieved in the auroral zone by using a smaller grid spacing with larger process noise values, in regions of enhanced activity. It is anticipated that a form of adaptive filtering, in which process noise values are adapted to local levels of activity, will provide the best results. This type of modelling is explored in detail in Chapter 7.

6.4 INVESTIGATION OF IONOSPHERE SHELL HEIGHT ASSUMPTION

In order to reduce the grid model to a two-dimensional adjustment in latitude and longitude, it was necessary to include a mapping factor in the measurement model (Section 3.3.2):

$$M(E) = \{1-[\cos(E)/(1+h/R_E)]^2\}^{-1/2}$$

This factor essentially allows mapping of the slant TEC to vertical in the ionosphere shell, where it is assumed that the majority of TEC is concentrated in a thin shell at fixed altitude. This function depends on the assumed shell height (h), which is nominally taken to be a centroid height of 350 km, consistent with observed vertical distributions of electron density (Section 2.6.3). The ionosphere shell height is also used in calculations of the observation ionosphere pierce point (Appendix B). Errors in the assumed ionosphere shell height therefore corrupt both the mapping factor (magnitude of vertical TEC), and computed positions of the ionosphere pierce point. These effects are most significant for lower elevation angles.

This simplifying assumption misrepresents the real three-dimensional electron density distribution. For the purposes of real-time ionosphere modelling in the wide area GPS networks, this assumption has proven to be adequate under typical ionospheric conditions, where vertical range delay accuracies on the order of 20 cm (L1) have been observed [*Bertiger et al.*, 1997]. For regions of auroral disturbances, however, electron density profiles may differ significantly from Figure 2.4, where energetic electrons precipitate at altitudes of ≈110 km [*Baron*, 1974]. Several authors have observed "double-peaked" densities (F and E region peaks of similar magnitude) in regions of energetic electron precipitation [i.e. *Bates and Hunsucker*, 1974]. If the vertical electron density peaks near 110 km (i.e for intense auroral disturbances), the ionospheric shell height may be better approximated at these lower altitudes.

In order to assess the impact of shell height assumptions, testing was conducted in exactly the same manner as the tests in Section 6.3, using a 120 km shell height. Figure 6.13 shows differences between the absolute values of measurement residuals derived for grid adjustments using shell heights of 120 km and 350 km (5 degree grid spacing). It is observed that the differences are normally distributed, where approximately equal numbers of positive and negative values are observed, such that neither shell height assumption allows consistently better results.

Tests were also conducted using heights of 200, 250 and 300 km. In all cases, results were inconclusive. It appears that simply imposing a lower shell height in regions of auroral disturbances does not allow for improvements in the grid model accuracies. This likely results from the fact that both vertical and horizontal electron density profiles are highly irregular in regions of auroral disturbances. The auroral disturbances are characterized by east-west current wedges (Section 4.3), with scale sizes on the order of several hundred kilometres (Section 5.6.2.2), and vertical electron density profiles may differ significantly across a localised region. Imposing one shell height across a region is too simplistic an assumption.

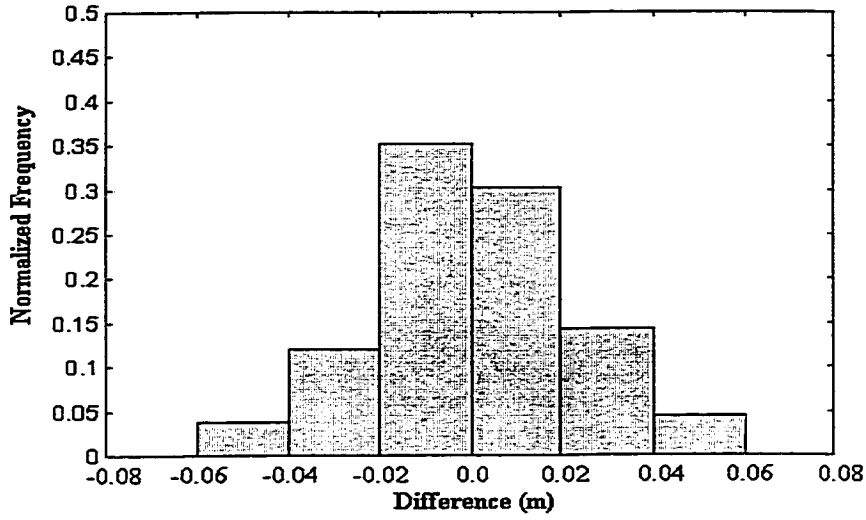


Figure 6.13 Probability distribution of differences between measurement residuals, as calculated for shell height assumptions of 120 km and 350 km. Differences are calculated as "120 km shell height" minus "350 km shell height".

In order to properly model TEC in auroral regions, it is suggested that a three-dimensional modelling technique, where electron densities are integrated along the satellite-receiver line-of-sight, be developed. For example, *Hansen et al.* [1997] has developed a method of ionosphere tomography, where vertical electron density profiles are estimated from a finite set of orthogonal basis functions. This technique requires a relatively long observation time to resolve the various coefficients, however, and it is possible that no combination of the (relatively simplistic) basis functions will allow for an adequate representation of complex auroral vertical electron density gradients. Developing a three-dimensional real-time ionosphere model for the auroral region would require significant additional research.

6.5 LATENCY EFFECTS

As outlined in the testing method (Section 6.1.2), the Type 1 test is intended to estimate typical grid accuracies observed at an independent user site. In conducting this type of testing, all test observations of VTEC during a given 5-minute interval were compared with values derived from grid point estimates of VTEC at the previous update. It is possible, therefore, that inaccuracies may grow during the 5-minute interval, as the "user" applies older updates of the grid model. In order to assess such latency effects, all results from the Type 1 testing have been binned according to the latency of grid point updates used to derive the grid accuracy statistics. Results for the entire active data set are presented in Figure 6.14. These statistics were generated using average values of system model parameters in the grid model.

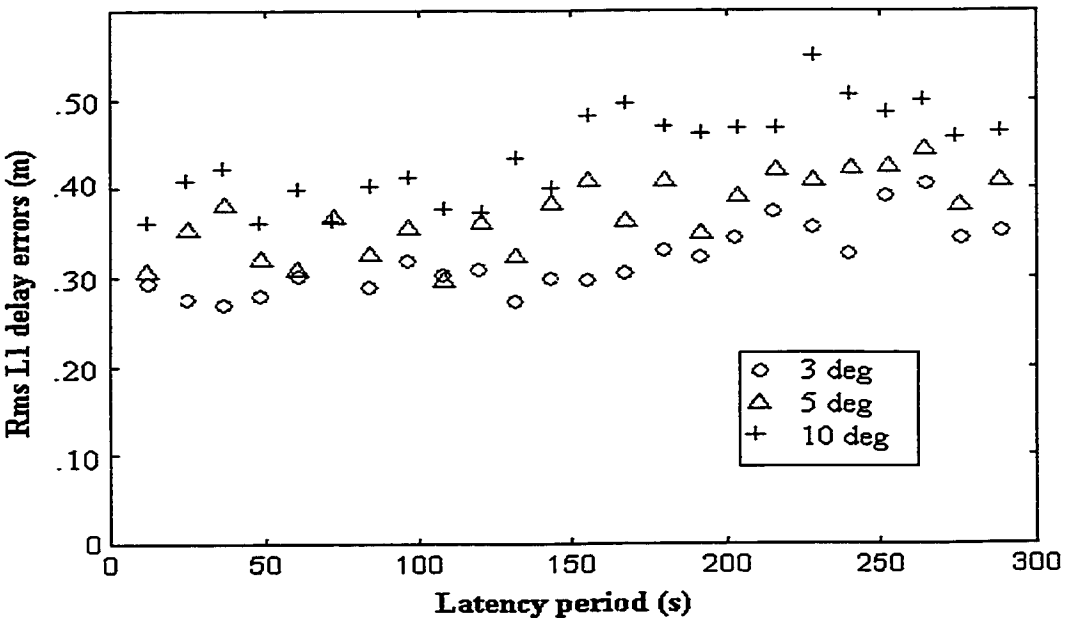


Figure 6.14 Grid accuracies (Type 1 testing) versus latency period, for three grid spacings.

For all three grid spacings, it is observed that inaccuracies in the grid point estimates grow for longer latency intervals. Accuracies are degraded by 5-10 cm for a latency period of 5 minutes, versus shorter delay intervals. These effects are due to the short-term temporal variations associated with auroral disturbances. Such degradation in grid accuracies, for latency intervals less than five minutes, are not observed in quiet-time statistics. These results should be considered when specifying network message update intervals for the wide area grid model. If ionospheric activity is monitored real-time, via growth of the observation residuals or other methods, it may be possible to shorten grid update intervals during periods of enhanced ionospheric activity.

6.6 OBSERVATION PIERCE POINT DENSITY REQUIREMENTS

In previous sections, improvements for smaller grid accuracies were achieved under restricted conditions. A requirement of the Type 1 testing (for deriving estimates of range delay errors) was the inclusion of at least ten measurements in updated grid estimates of TEC (see Section 6.1.2), for all grid points surrounding a given test observation. This condition is derived from the nature of the grid model, where only observations in the surrounding four grid cells contribute significantly to updates at a given grid point. Requirements for a minimum number of observations are equivalent to specifying the density of observation pierce points.

The threshold value of ten is based on the observed relationship between number of measurements and grid accuracies. Figure 6.15 illustrates rms of range delay errors (Type 1 testing) for the entire active data set versus the number of input measurements included in grid point updates. Results have been computed for grid spacings of 3, 5 and 10 degrees. Note that for grid spacings of 10 degrees the number of observations included in grid point updates is consistently greater than fifteen. Process parameters representative of active ionospheric conditions have been used.

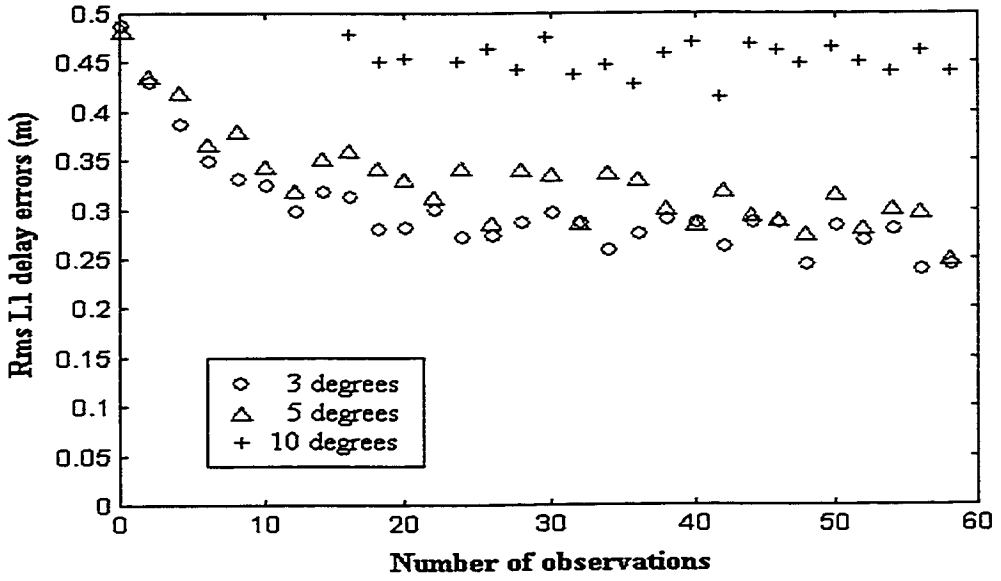
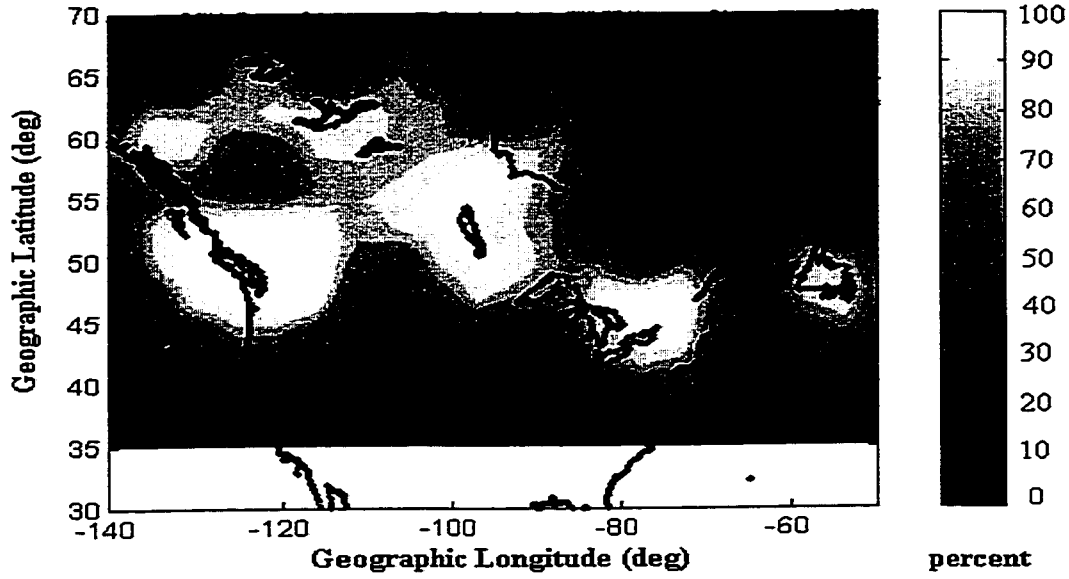


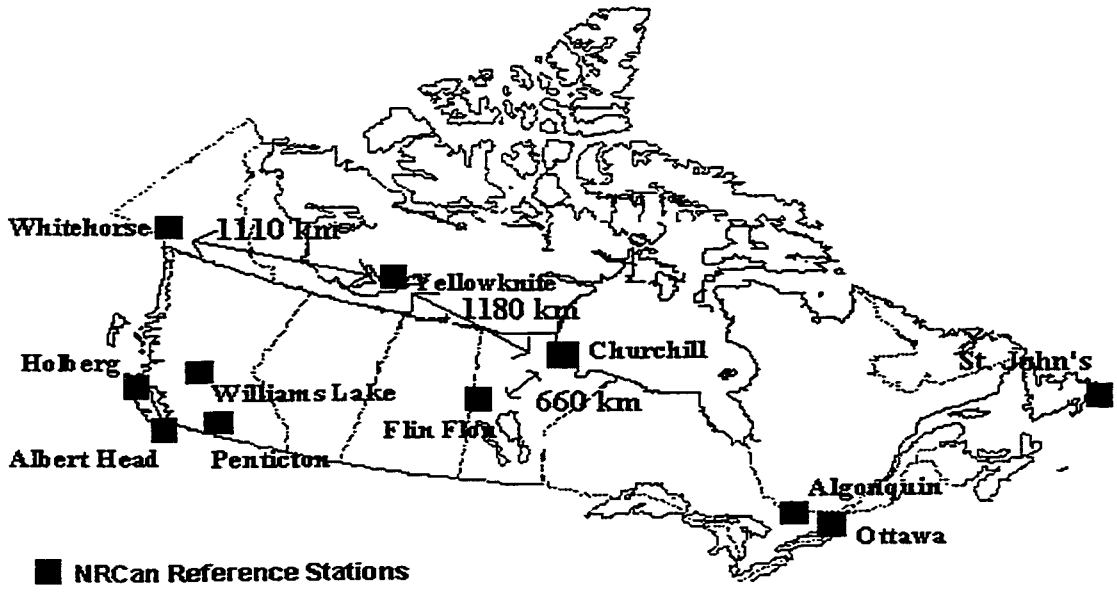
Figure 6.15 Grid accuracies (Type 1 testing) versus number of observations. Results have been computed using the entire active data set.

In Figure 6.17, it is observed that grid accuracies approach values of 30 cm rms (3 and 5 degree grid spacings) when ten or more measurements are included in updated estimates of VTEC at a given grid point. Accuracies are degraded for a smaller number of measurements, and the poorest accuracies are observed when no new measurements are available. In such cases, grid values are derived from system model predictions alone.

In practical designs of a wide area network, it is necessary to consider the spacing of reference stations required to meet ionosphere grid accuracy requirements. For optimal accuracies during periods of auroral disturbances, the results presented in Section 6.3.1 indicate that grid spacings of three degrees or less should be used. A requirement that ten or more new measurements be included in the update of a given grid point is also necessary. It should be noted, however, that no conditions have been placed on the distribution of these measurements. Both the number and spacing of the observations contributes to the overall grid accuracies. In the following analysis, reference station configurations required for improved grid accuracies are investigated.



a)

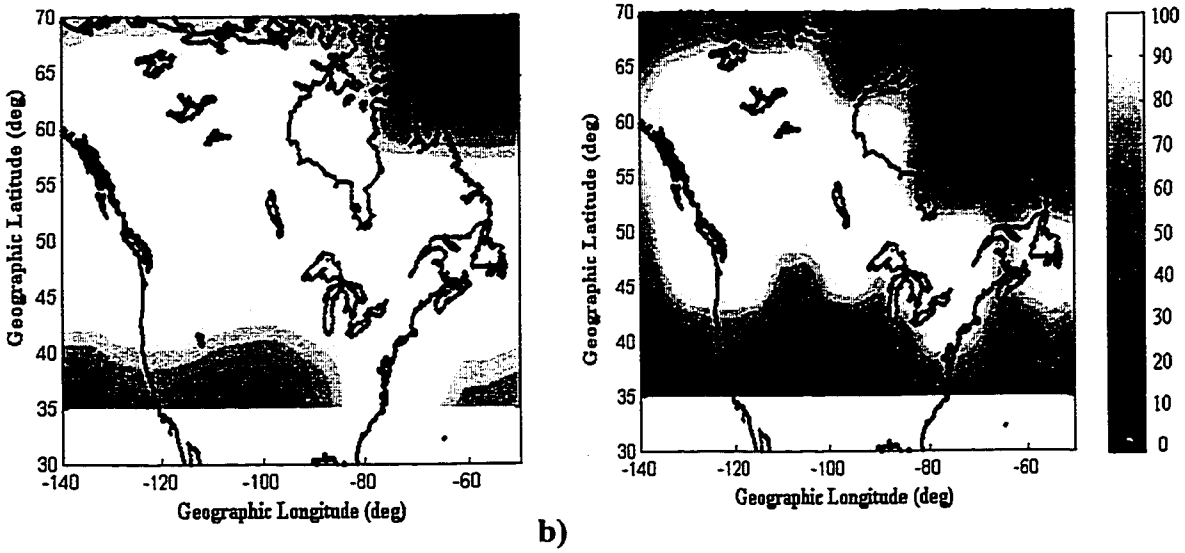


b)

Figure 6.16 a) Percentage of time ten or more measurements are included in a grid point update (5 minute update interval), for a 3 degree grid spacing. b) Distances between reference stations in the Natural Resources Canada wide area network.

Figure 6.16 shows the percentage of time ten or more measurements are included in grid updates for the NRCan network, for a three degree grid spacing. This criteria is met 100 percent of the time only in localised regions, where the density of reference stations is such that station separations are on the order of 600 km and less (the distance between Flin Flon and Fort Churchill being 660 km).

For station separations on the order of 1000 km, such as between Whitehorse and Yellowknife, coverage requirements are met only 60-80 percent of the time. In these regions degraded vertical range delay accuracies as large as 50 cm rms may be observed for 3 degree grid spacings during periods of enhanced auroral activity. This brief analysis suggests that in order to ensure full network coverage for such smaller grid spacings, and resulting improvements in grid accuracies, reference stations should be separated by distances of 600 km or less. For comparison purposes, Figure 6.17 shows results similar to those in Figure 6.16, for grid spacings of 5 and 10 degrees. In both cases, the majority of the wide area region meets accuracy requirements.



a) b)

Figure 6.17 Percentage of time ten or more measurements are included in a grid point update for a) a 10 degree grid spacing, and b) a 5 degree grid spacing.

6.7 INTERPRETATION OF IONOSPHERE GRID ACCURACIES

The analysis performed in this chapter is intended to provide general specifications, and typical accuracies, for all wide area grid models incorporating measurements from the auroral region. For models such as *Mannucci et al.* [1997], where VTEC is estimated at points in a triangular grid, each side 800 km in length, accuracies may be degraded by a factor of 2-5 (i.e. 10 degree grid spacing, Section 6.3.1), relative to typical accuracies of 20 cm rms, in the auroral region. Better accuracies are expected for the GPS•C network [*Skone et al.*, 1997], where a 3°×3° grid is implemented, provided coverage requirements are met (Section 6.6).

It is useful to also interpret these results specifically with respect to WAAS requirements. As mentioned in Section 2.5, the WAAS has been designed by the FAA as a safety critical system to support precision approach aircraft navigation. This system will be operational not only in the Continental United States, but also Alaska. Extension of the WAAS over Canada (i.e. CWAAS [*Nav Canada newsletter*, 1998]) may begin in the next few years. For high latitude users of this system, auroral disturbances (and their effects on ionosphere grid accuracies) are an issue. As stated in Section 3.2, WAAS error bounds specify a vertical alert limit of 15 m, which translates into ionosphere accuracy requirements. A value of grid ionosphere vertical error (GIVE) less than or equal to 1.5 m 99.9% of the time has been suggested by *El-Arini* [private communications], to satisfy overall error bounds. Grid spacings of 5 (10) degrees, at latitudes less than (greater than) 55 degrees geographic are also specified [*FAA*, 1997]. Implicit in WAAS requirements, therefore, is a grid spacing of 10 degrees in the auroral region.

The Type 1 testing method was designed to reflect, as closely as possible, typical model accuracies at an independent user site. Comparisons of these accuracies with WAAS error bounds, for a 10 degree spacing, indicate that overall grid accuracies are within tolerance levels during periods of enhanced auroral activity (Tables 6.1 and 6.2). It

is important to note, however, that the actual GIVE values (as transmitted to system users) are generally derived from very conservative bounds of real-time error estimates, where the estimated accuracies are multiplied by a confidence factor.

Several algorithms have been proposed to compute reliable GIVE estimates. *Conker et al.* [1997] have developed an estimation technique based on comparisons between observed values of TEC at numerous monitor stations and the values of TEC predicted (from grid model updates) at these "test" observation pierce points. A conservative tolerance bound is then constructed essentially through scaling various observed standard deviations (derived from the previous 5 minutes of observations) by a confidence factor, adding these values to the appropriate means, and finding the maximum value near a given grid point. For test stations colocated with network stations, derivation of the GIVE is similar to the Type 2 testing described in this chapter. For test stations located several hundred km from the network reference stations, derivation of the GIVE is similar to the Type 1 testing conducted here.

In order to provide approximate estimates of typical (conservative error bound) GIVE values derived for WAAS, the results in Tables 6.1 and 6.2 have been multiplied by a scaling factor of 5.43 [cf. *Conker et al.*, 1997; *Hahn and Meeker*, 1991] which is based on confidence level, the proportion of sample population to be bounded, and the sample size. In deriving this factor, *Conker et al.* have chosen a sample size consistent with the number of observations for a 5-minute satellite-receiver time series of TEC (in their case, 30 samples). It is assumed here that the values in Tables 6.1 and 6.2 are typical of rms values over such shorter intervals. This is a reasonable assumption, given that sample sizes of 30-40 were used to compute individual rms values plotted in Figures 6.6, 6.7, and 6.8 – and these values are consistent with those in Tables 6.1 and 6.2. Test samples are also assumed to be zero-mean (consistent with the derivation of an rms statistic). The GIVE estimates are summarized in Tables 6.3 and 6.4.

Table 6.3 GIVE Values for Type 1 Testing - Active Data Set

RMS (m)	GRID SPACING		
System Model Parameters	**3 degrees**	**5 degrees**	**10 degrees**
Quiet	1.79	1.95	2.50
Active	1.52	1.79	2.55

Table 6.4 GIVE Values for Type 2 Testing - Active Data Set

RMS (m)	GRID SPACING		
System Model Parameters	**3 degrees**	**5 degrees**	**10 degrees**
Quiet	.87	1.14	1.25
Active	.76	1.03	1.19

It is apparent that, if GIVE estimates are derived from independent monitoring stations (i.e. Table 6.3), a 10 degree grid spacing results in overall GIVE values on the order of 2.50 m during periods of enhanced auroral activity. GIVE values for the 3 and 5 degree grid spacings are much closer to satisfying approximate WAAS error bounds (i.e. GIVE $\leq$ 1.5 m). From Figure 6.11a, it is estimated that GIVE values greater than 5.43 m (i.e. 1.0 m rms $\times$ 5.43) were observed approximately 10 percent of the time, for the 10 degree grid spacing. While these estimates are rough, they nonetheless serve as a guideline for anticipated performance of the WAAS in Alaska and Canada (CWAAS).

A further observation can be made from Tables 6.3 and 6.4: the discrepancy between GIVE estimates derived for the Type 1 and Type 2 tests. Errors are underestimated for the Type 2 test, where grid accuracies are derived from observations included in the grid adjustment. If GIVE values are derived from observations at monitor sites colocated with network reference stations, the GIVE values will underestimate (by $\approx$1.0 m, 10 degree spacing) true grid point error bounds at independent test sites. Since it is intended that users derive *user ionosphere vertical errors* (UIVE) from transmitted GIVE values, this is an important issue for ensuring that users are provided with reliable estimates of the grid model errors.

One final observation is the option of improving grid accuracies by using a shorter update interval. (i.e. Figure 6.14). For an update interval of 2 minutes, grid accuracies (Type 1 test) can be improved by ≈5 cm, over those for update intervals of 5 minutes (10 degree grid spacing). This translates, under the assumptions made here, into a 30 cm improvement in GIVE error bounds. The WAAS message format is designed to allow for variable update intervals, with a maximum of 5 minutes [*FAA*, 1997]. If the level of ionospheric activity is monitored, update intervals could be shortened during intervals of enhanced auroral activity.

6.8 PREDICTIONS FOR SOLAR MAXIMUM

It is important to note that the active data set was compiled during a time of solar minimum (year 1996). As solar activity increases towards the solar maximum (year 2000), the magnitude and frequency of auroral disturbances and substorms is expected to increase. It is therefore useful to assess further degradations in auroral grid accuracies which may be oserved during the period of solar maximum.

Several authors have predicted that global grid accuracies will be degraded by a factor of 2-3 at solar maximum [i.e. *Klobuchar et al.*, 1995]. These predictions are based on observations of variations in the large-scale TEC gradients, however, which are not necessarily related to the unique auroral phenomena. It has been demonstrated, in Sections 5.1 and 5.2, that the magnitudes of auroral TEC variations are correlated with ground-based magnetic field perturbations. Typical amplitudes of auroral disturbances for the active data set were on the order of 1.40 TECU (Section 5.6). Corresponding magnetic field perurbations were in the range 700-1000 nT.

In 1989, during the previous solar maximum, magnetic field perturbations (similar to those in Figures 5.4 and 6.3) as large as several thousand nT were observed in Alaska

(i.e. March 1989). This indicates a further degradation in grid accuracies by a factor of ≈2 during intense substorm activity, as compared with the results in Tables 6.1 and 6.2 (and Tables 6.3 and 6.4). This is consistent with observed GIVE values on the order of 5.0 m during a severe storm at mid-latitudes in 1993 [*Conker et al.*, 1997]. It is important to note, however, that such intense storms are rare, even during solar maximum, occurring approximately one percent of the time [*Kunches*, 1997].

It is also likely that the greatest challenge will be actually tracking the GPS signals through scintillation effects during such severe ionospheric activity. The number of observations may be so limited that grid model updates cannot be computed. The type of results found in Section 6.7 will be more frequent during solar maximum, with "moderate-major" storms occuring approximately six times more often [cf. *Kunches*, 1997].

CHAPTER 7

ADAPTING THE IONOSPHERE GRID MODEL

The results of Chapter 6 show that grid accuracies are degraded in regions of auroral disturbances, and the magnitude of such errors may compromise system error bounds (i.e. WAAS). While it is difficult to model the somewhat random ionospheric features associated with enhanced auroral activity, simple adjustments of stochastic model parameters may allow for improvements in grid accuracies. For example, it was demonstrated in Section 6.3.2 that grid accuracies were improved, for a smaller grid spacing, when larger values of process noise are used in the system model. The use of shorter correlation times essentially allowed the model to follow more rapid variations in TEC, as appropriate during periods of enhanced auroral activity.

Signatures associated with the short-term auroral disturbances, such as amplitude and predominant frequencies/wavelengths, were identified in Chapter 5. If the presence of auroral disturbances can be detected, through knowledge of such signatures, grid model parameters may be adapted to better represent prevailing ionospheric conditions. In this chapter, an adaptive technique is explored, in which the grid system model parameters are adapted to reflect local ionospheric activity. This algorithm involves the detection of auroral disturbances, and estimation of model parameters for individual grid points, based on statistics derived from local time series of TEC. The potential of such an algorithm is demonstrated in testing of the active data set, where accuracies are improved in regions of auroral disturbances.

7.1 THE ISSUE OF SUBOPTIMAL FILTER PERFORMANCE

The ionospheric grid model developed in this thesis is based on a Kalman filter optimal estimation technique. As outlined in Appendix A, the Kalman filter equations represent an optimal linear estimate of unknowns (in this case VTEC at grid points), as derived from all available observations. The gain matrix K was derived under the condition of minimal variance, such that the *a posteriori* covariance of the estimated grid point VTEC is minimized (i.e. trace[P^+]=minimum). In this sense, the form of the gain matrix (Equation 3.14) reflects the optimal weighting of all observations, relative to each other, in time and space. Each element of the gain matrix is essentially the ratio between statistical measures of uncertainty in the predicted estimates, versus measurement noise. Given that elements of the gain matrix depend on *assumed* behaviors of system dynamics, measurement models, and error statistics, suboptimal filter performance may result under incorrect assumptions.

The simplistic weighting functions W (Section 3.3.2) are based on interpolation schemes, rather than knowledge of physical processes. In general, the system models are derived from stochastic properties, as opposed to deterministic models. In practical applications, the Gauss-Markov system model is derived from typical values of the first and second moments of VTEC random processes [i.e. *Mannucci et al.*, 1993], and no regional dependence is assumed (i.e. common values of σ and β in the Φ and Q matrices - Equations 3.18 and 3.19). In general, such assumptions are adequate. In the auroral region, however, the presence of localised disturbances can cause TEC processes to vary significantly between regions. In such circumstances, simplistic model assumptions may result in significant suboptimal filter performance. It is necessary to adapt individual terms of both the Φ and Q matrices, to best represent local conditions at the corresponding grid points. The development of such an adaptive grid algorithm forms the basis of this chapter.

7.2 REGIONAL DEPENDENCE OF TEC VARIATIONS

The regional dependence of auroral activity has been demonstrated in Chapter 5 (Figure 5.1). A further example of localised ionospheric activity is shown in Figure 7.1. This figure compares observations of vertical TEC, as derived from five satellite ionospheric pierce point (IPP) tracks in, or near, the auroral region on October 23, 1996. This is the same day studied in detail in Section 6.2. The TEC values were derived using simultaneous observations from PRN 19 at five stations in the Natural Resources Canada wide area network. Note that only relative values are plotted. Figure 7.2 shows the corresponding ionospheric pierce points for all satellite-receiver pairs.

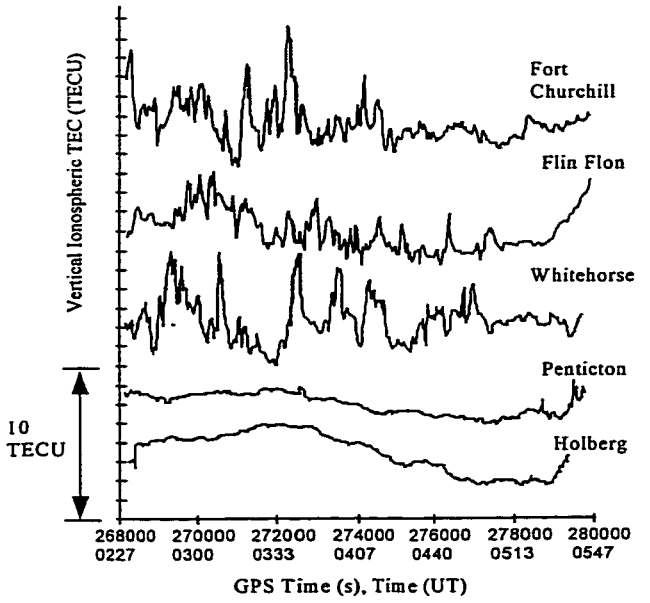


Figure 7.1 Vertical TEC for PRN 19 at five stations in the NRCan wide area network, October 23, 1996. Values are plotted on a relative scale.

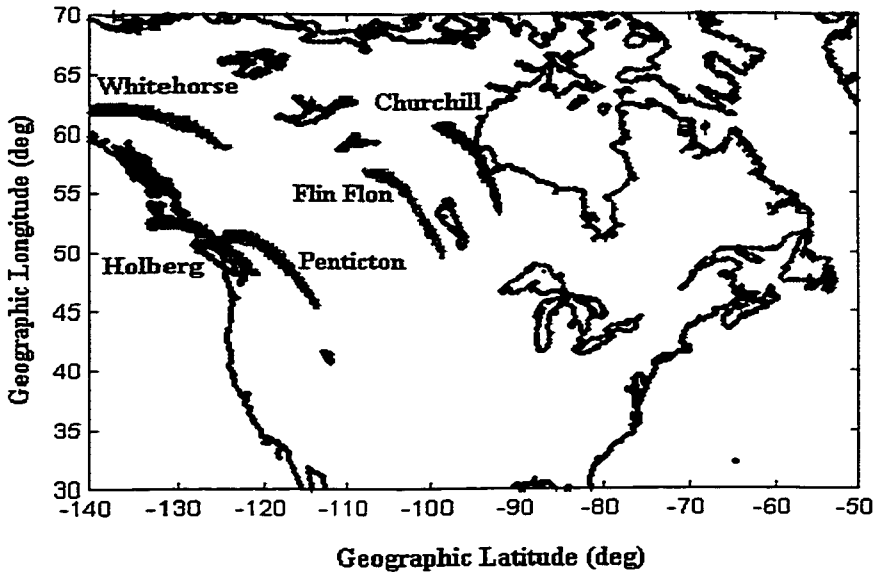


Figure 7.2 Ionospheric pierce point tracks (geographic coordinates) for PRN 19 and five stations in the NRCan wide area network.

Higher order variations in TEC (Flin Flon, Churchill, Whitehorse) are caused by auroral disturbances, while only large-scale TEC features are observed in lower latitude regions outside of the disturbances (Penticton, Holberg). It is clear that ionospheric activity may vary from region to region, and it follows that ionospheric temporal processes are regionally dependent. In order to allow for such regional dependence in the wide area grid model, the presence of enhanced auroral activity must be detected, and information describing the processes derived, from available real-time observations of TEC.

7.3 DESCRIBING IONOSPHERIC TEMPORAL PROCESSES

Equations describing the system model were provided in Section 3.3.3. For convenience, these equations are also included here, where the system model is given as

$$VTEC_i(t_{j+1}) = e^{-\beta(\Delta t)}\, VTEC_i(t_j) + w(t_j) \qquad (7.1)$$

where $1/\beta$ is the *correlation time*, and Δt is the prediction interval $t_{j+1} - t_j$. The process noise is given as

$$q(t_j) = \sigma^2[1 - e^{-2\beta(\Delta t)}] \qquad (7.2)$$

where σ^2 essentially represents the overall variance of the process (i.e. the variance of an infinitely long time series of TEC at the given grid point).

For stationary processes (i.e. no changes in statistical properties over time), σ^2 is derived as the variance of the entire random ensemble of processes $VTEC_i(t_j)$, evaluated at time t_j for the given grid point i. Generally, however, it is difficult to derive a representative ensemble of system processes for a given unknown, and the ergodic hypothesis is invoked. For an ergodic ensemble of random stationary processes, statistics describing the ensemble of functions can be derived from the values of one *representative* function, considered over all time. The process variance σ^2 at grid point i can then be approximated as $E[VTEC_i^2]-E[VTEC_i]^2$. For a stationary random process, the statistics are completely described by the discrete-time temporal autocorrelation function (Equation C.18, Appendix C).

In Chapter 5 (Section 5.6.2.1), statistics describing the correlation times ($1/\beta$) and mean squared amplitudes (σ^2) of auroral disturbances were estimated from a spectral analysis of 64-minute time series of vertical TEC variations. These time series were derived from individual series of satellite-receiver observations in the active data set. "Real-time" series of vertical TEC observations for a given satellite-receiver pair, such as those in Figure 7.1, can therefore be used to derive real-time estimates of system model

parameters. By deriving an empirical autocorrelation function for a given time series of TEC values in a local region, approximate values of process noise and correlation time may be derived for surrounding grid points.

7.3.1 Ergodicity and Stationarity Issues - Determining the Optimal Sample Length

A stationary random process is characterized by statistical properties (i.e. first, second moments) that are invariant in time. Stationary processes are ergodic if any statistic calculated by averaging over all members of the statistical ensemble at one point in time is equivalent to results obtained by averaging one representative process over all time.

As stated in the previous section, system model parameters of $1/\beta$ and σ can be derived from empirical autocorrelation functions. These autocorrelation functions are only representative of the statistical properties of a given process, however, if process statistics do not change appreciably during the time series observed to compute the autocorrelation function. For transient periods of auroral disturbances, the maximum length of sample time series is limited by this condition. Ideally, time series of several minutes are preferred. This is not possible, however, due to practical limitations in sampling interval (30 s for the data set used here), and the observation period required to ensure that the ergodic hypothesis is valid (i.e. the series has been observed long enough to reflect all possible behaviors). In order to derive statistics representative of enhanced auroral activity, the ionosphere process must be observed over an interval consistent with the period of auroral disturbances.

One further consideration is the inability to reliably derive the form of the autocorrelation function for large time shifts. This results from the finite nature of the data series, where such errors are proportional to the correlation time divided by time

interval of the sample data series [cf. *Gelb*, 1974]. For long data series (with lengths many times the correlation time) these errors are negligible. In order to assess the magnitude of these errors for auroral temporal processes, autocorrelation functions were derived from data series of various lengths, using the 22 hours of active data (Section 5.2). Results are shown in Figure 7.3.

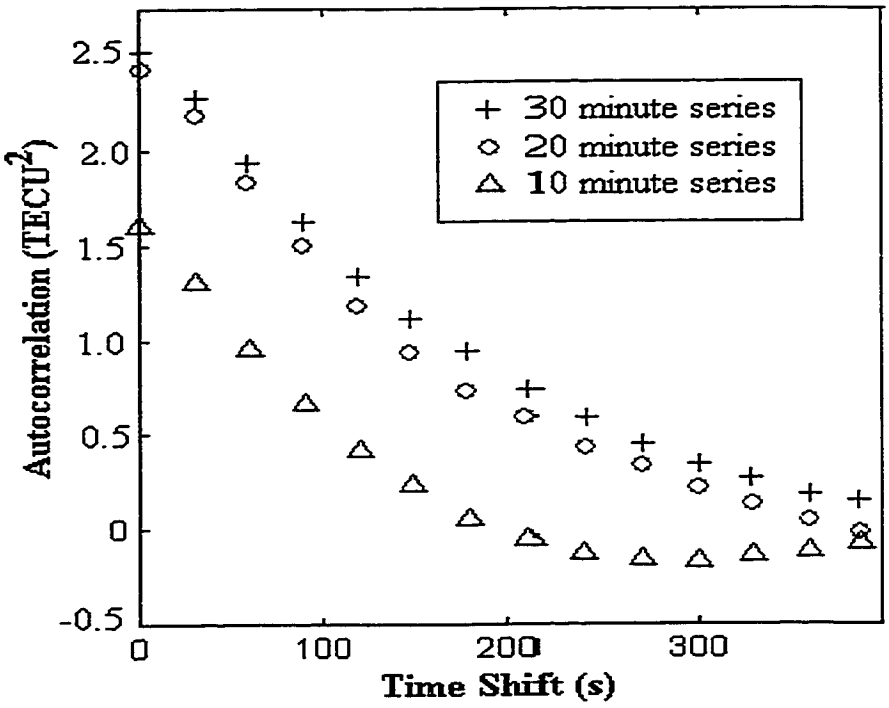


Figure 7.3 Autocorrelation functions for ionospheric processes in regions of enhanced auroral activity, as derived from time series of varying lengths.

In this figure, the autocorrelation function is significantly different for the 10 minute time series, as opposed to those series 20 and 30 minutes in length. Correlation times (where the autocorrelation coefficient is equal to σ^2/e) are underestimated for the shorter series, with a value of 100 s, compared with 180 s for the longer series. Additionally, the coefficient at time lag 0, which represents the overall variance of the

process, is smaller for the 10 minute series. Given that observed periods for auroral disturbances are on the order of 10 minutes [cf. *Rostoker*, 1991], it is likely that the 10 minute series is not long enough to observe all possible behaviors of the temporal process. In this case, ergodicity of the statistics is compromised. In contrast, few differences are observed between the autocorrelation functions for the 20 and 30 minute time series.

From these results, a sample interval length of 20 minutes has been chosen to satisfy the various criteria. The interval is reasonably short, such that computational efficiency is maximized, estimates are based on the most recent data, and the process can be assumed stationary. The interval is also long enough, however, such that small-scale ionospheric effects are represented in σ^2 estimates and an adequate number of data samples are available to derive the autocorrelation function. A maximum time shift of 10 minutes is used in deriving the autocorrelation function.

7.3.2 Typical Autocorrelation Functions

Figures 7.4 and 7.5 show empirical autocorrelation functions derived using two time series of ionospheric delay values from Figure 7.1. The functions are estimated from averages of correlation values derived at each epoch, using the most recent 20-minute interval of observations. The "best fit" theoretical functions are plotted for comparison, and it is evident that variations in TEC are well approximated as Gauss-Markov processes (exponential autocorrelation function) over time.

Correlation times are estimated to be 180s and 360s, respectively, for the observations at Whitehorse and Holberg. The correlation time is shorter in regions of auroral disturbances (Whitehorse), as expected. Note that the autocorrelation function for Holberg is not entirely representative of ionospheric processes, however. At Holberg, the

autocorrelation function is dominated by higher order effects associated with measurement noise. The correlation time would be significantly larger in the absence of such noise. The σ parameter better reflects differences in ionospheric activity near Whitehorse, versus Holberg. The standard deviation is a factor of 5 larger at Whitehorse, indicating the presence of enhanced local variations in TEC.

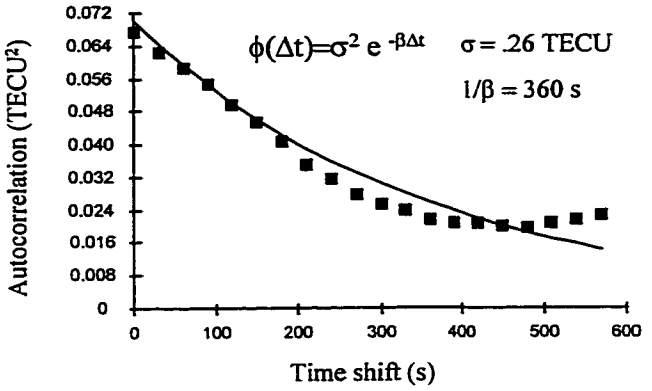


Figure 7.4. Zero-mean empirical autocorrelation function for PRN 19 observations of TEC at Holberg (squares). Theoretical "best fit" function is also plotted.

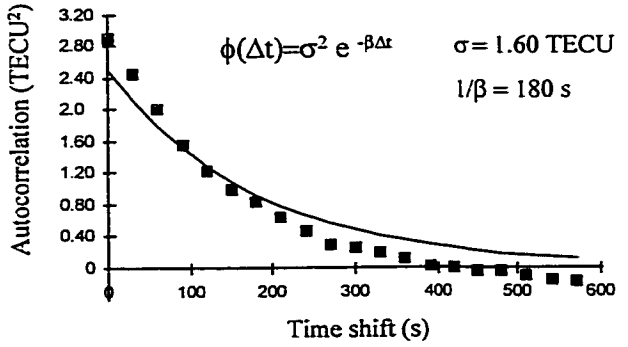


Figure 7.5 Zero-mean empirical autocorrelation function for PRN 19 observations of TEC at Whitehorse (squares). Theoretical "best fit" function is also plotted.

Note that the autocorrelation functions are derived under the assumption that TEC variations, as plotted in Figure 7.1, are representative of only temporal effects. The TEC variations are, however, attributed to combined effects of temporal and spatial gradients. In reality, it is difficult to separate the spatial and temporal gradients in time series observations of TEC for a given satellite-receiver pair. In deriving the autocorrelation function, it is intended that estimated parameters describing the temporal process may be derived. While these estimates include spatial correlation effects, they nevertheless serve to quantify *approximately* the enhanced temporal processes associated with geomagnetic disturbances. A process noise variance of 2.5 $TECU^2$ is derived from the temporal process parameters in Figure 7.5, for a five minute prediction interval (using Equation 7.2). This value is reasonable, as compared with process noise values in *Mannucci et al.* [1993], *Komjathy and Langley* [1996b] and *Mueller and Davis* [1998] (0.5-2.4 $TECU^2$) for larger-scale averaged TEC gradients.

7.4 ADAPTING THE SYSTEM MODEL

In adapting the grid model, it is intended that the correlation times $1/\beta_i$ and variance values σ_i^2 (Equations 7.1 and 7.2) be consistent with local ionospheric processes at each grid point i. This is achieved in two steps: (1) deriving estimates of $1/\beta$ and σ^2 for each observation, and (2) incorporating observed values of $1/\beta_i$ and σ_i^2 into estimates of process parameters at each grid point. Included in the first step is an algorithm designed to detect auroral disturbances.

7.4.1 Deriving Observed $1/\beta$ and σ

The values of variance and correlation time at each observation point (IPP) are computed using a method similar to that described in Section 7.3. At every epoch, an empirical autocorrelation function is derived using the most recent 20 minutes of vertical

TEC observations for a given satellite-receiver pair. The observations are filtered, prior to calculating the autocorrelation function, such that measurement noise and multipath effects are minimized (Section 2.4). A low pass filter with cutoff frequency 4.0 mHz (period = 4 minutes) is used. This filtering removes higher frequency measurement noise and multipath effects, while retaining spectral components with longer periods, such as those associated with AEI (periods of ≈10 minutes). The cutoff frequency is consistent with observed periods of specular multipath [cf. *Braasch*, 1996; *J.K. Ray*, private communications], for reflecting surfaces at distances of ≈10 m or more from the antenna [cf. *Leick*, 1995]. This low pass filtering does not remove all multipath components, however; for reflecting surfaces closer to the antenna, multipath periods may be longer, with frequencies less than 4 mHz. The estimated variance σ^2 is then tested, in order to detect the presence of small-scale ionospheric disturbances. The necessity of this detection test is outlined as follows.

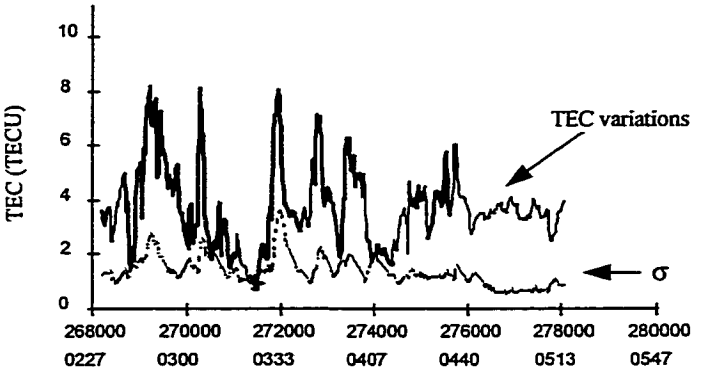
A typical autocorrelation function was shown for non-disturbed ionospheric conditions in Section 7.3 (Figure 7.4). A correlation time of 360 s was derived for this example, where the shorter correlation time reflects measurement noise rather than ionospheric variations. While low pass filtering of the observations limits these effects, some measurement noise (on the order of ionospheric variations) remains. In such cases, the statistics for non-disturbed periods do not truly represent local ionospheric conditions. Additionally, representative correlation times may not be derived reliably for quiet ionospheric conditions from the limited 20-minute window of observations. For such shortened time series, the autocorrelation function is computed for maximum temporal delays of 10 minutes (Section 7.3.1). Typical large-scale ionosphere correlation times are on the order of 15-30 minutes [cf. *Mannucci et al.*, 1997].

Given that the adaptive modelling technique is designed to improve grid accuracies in regions of auroral disturbances, and that limitations exist in deriving process

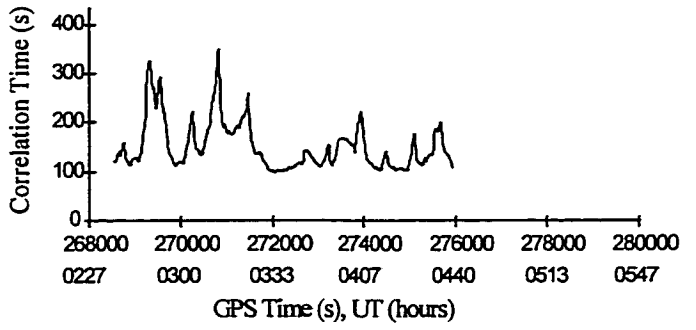
parameters for quiet ionospheric conditions, implementation of the adaptive model includes the detection of auroral disturbances. Process parameters are then allowed to vary only near such regions of enhanced auroral activity. A detection threshold of $\sigma^2 > 1.4$ TECU^2 is used. This minimum value has been chosen somewhat arbitrarily, based on observed amplitudes of slant TEC variations in the range 1-11 TECU [*Hunsucker et al.*, 1995] during periods of AEI. This is consistent with the statistics derived in Section 5.6, where a mean squared amplitude of 2.0 TECU^2 was estimated for auroral disturbances. For observations meeting the variance criteria, the computed values of $1/\beta$ and σ are "assigned" to the observation.

If an observation does not meet the detection criteria, it is assumed that no ionospheric disturbances are present at the observation pierce point, and default values of $1/\beta = 15$ minutes and $\sigma = 0.75$ TECU are assigned to the observation. These values are consistent with process noise values for quiet-average levels of ionospheric activity in the nightside auroral region (cf. Section 6.1.2), where TEC gradients are not typically steep. It has been found that the average values of parameters $1/\beta$ and σ best apply in this case.

Based on this estimation technique, the data in Figure 7.1 were tested for the presence of AEI and process parameters computed. Results are shown in Figure 7.6, where intervals of AEI are indicated in bold. Correlation times vary between $\approx$100s and 350s in Figure 7.6, while σ values are as high as 3.5 TECU (in regions of AEI). These values differ significantly from the average values of $1/\beta = 15$ minutes and $\sigma = 0.75$ TECU.



a)



b)

Figure 7.6 a) Vertical TEC variations (solid line) and corresponding σ (dotted line) for PRN 19 at Whitehorse. Intervals of detected geomagnetic disturbances are bold.

b) Corresponding values of estimated correlation time 1/β, for periods of auroral disturbances.

7.4.2 Deriving approximate values $1/\beta_i$ and σ_i at grid point

Once each observation has been assigned values of 1/β and σ, a weighted average of these values is computed at each grid point i, to estimate $1/\beta_i$ and σ_i (Equations 7.1 and

7.2). Only those observations located within the surrounding four grid squares are included in the calculation. An inverse distance weighting function is assumed. The system model parameters $1/\beta_i$ and σ_i then reflect local ionospheric conditions at each grid point i. If no observations are available within the surrounding four grid points, default average values $1/\beta = 15$ minutes and $\sigma = 0.75$ TECU are assumed. The matrices Φ and Q are then computed (Equations 3.18 and 3.19), using the estimated $1/\beta_i$ and σ_i values, and the grid adjustment carried out. Figure 7.7 illustrates the various components of the adaptive algorithm.

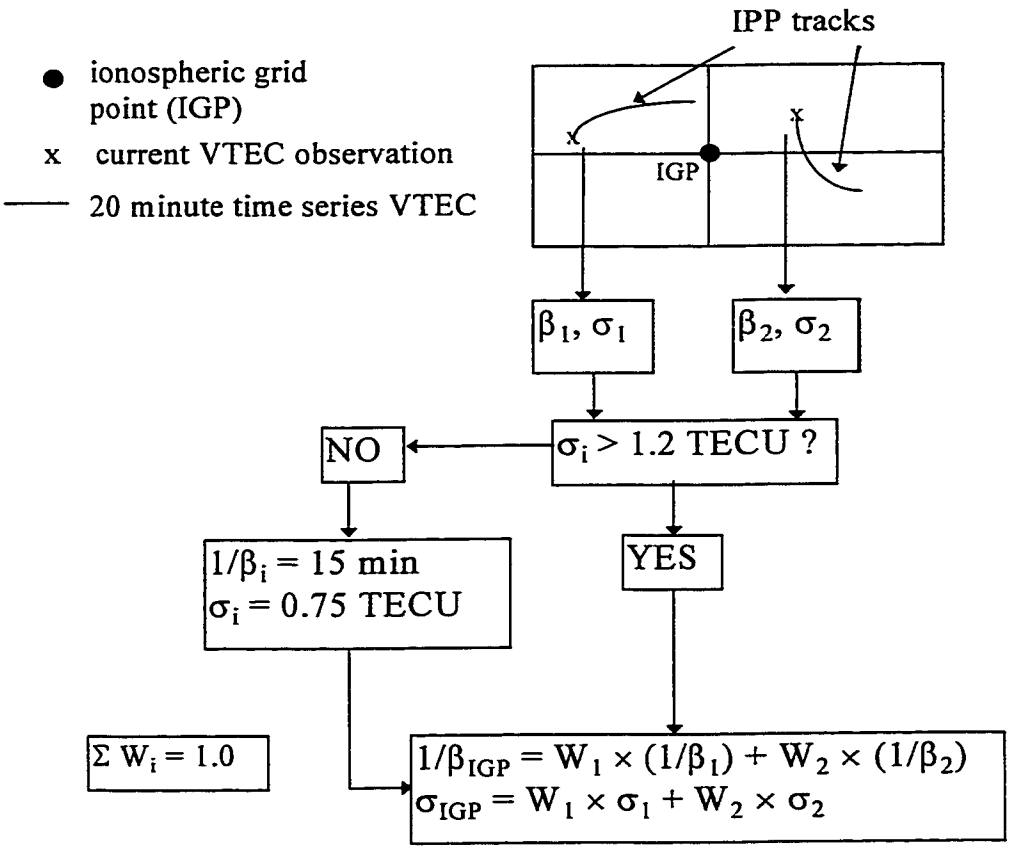


Figure 7.7 Algorithm for adapting the system model parameters. The notation "IGP" denotes the ionospheric grid points. Inverse distance weighting factors are denoted W.

As an example of the AEI detection technique, Figure 7.8a shows typical σ values. The larger σ values correspond to the periods of enhanced auroral activity in Figure 6.3. Only during the period 0 - 0400 UT do σ values consistently exceed the threshold for detection of auroral disturbances. Dayside values rarely exceed σ = 0.5 TECU. Figure 7.8b shows the spatial distribution of all detected auroral disturbances at 0230 UT. The region of enhanced auroral activity is generally consistent with typical auroral oval boundaries. Several disturbances have, however, been detected below the equatorward extent of the oval. These cases are indicative of the equatorward motion of the auroral oval during a period of enhanced directly driven activity. Real-time monitoring of auroral oval boundaries is addressed in Chapter 8.

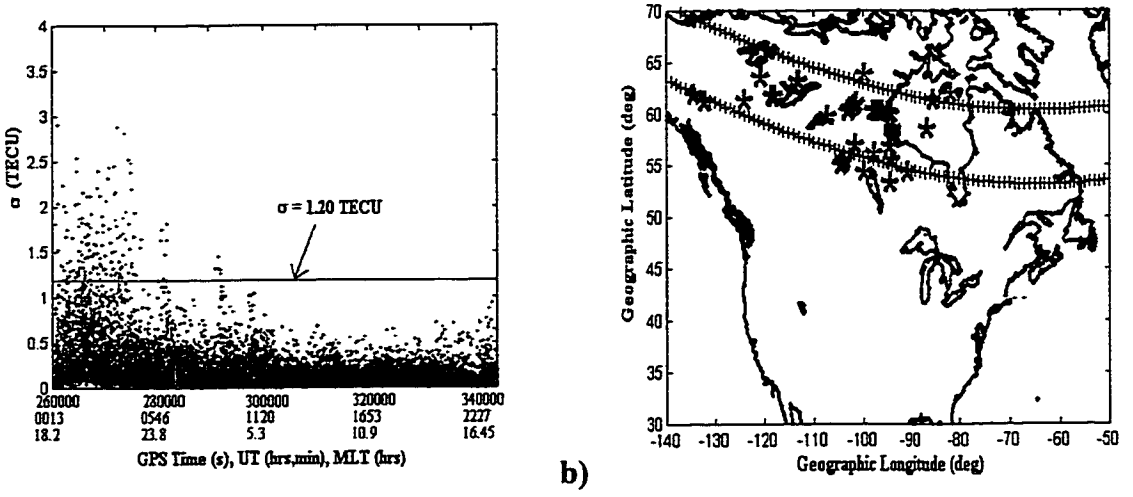


Figure 7.8 **a) Values of process σ for all VTEC time series observed from the NRCan reference stations, October 23, 1996.**

b) Spatial distribution of detected auroral disturbances at 0230 UT. Typical oval boundaries are plotted for comparison (++++). Boundaries are calculated from expressions in *Kamide and Fukushima* [1970], using parameters from *Rostoker and Skone* [1993].

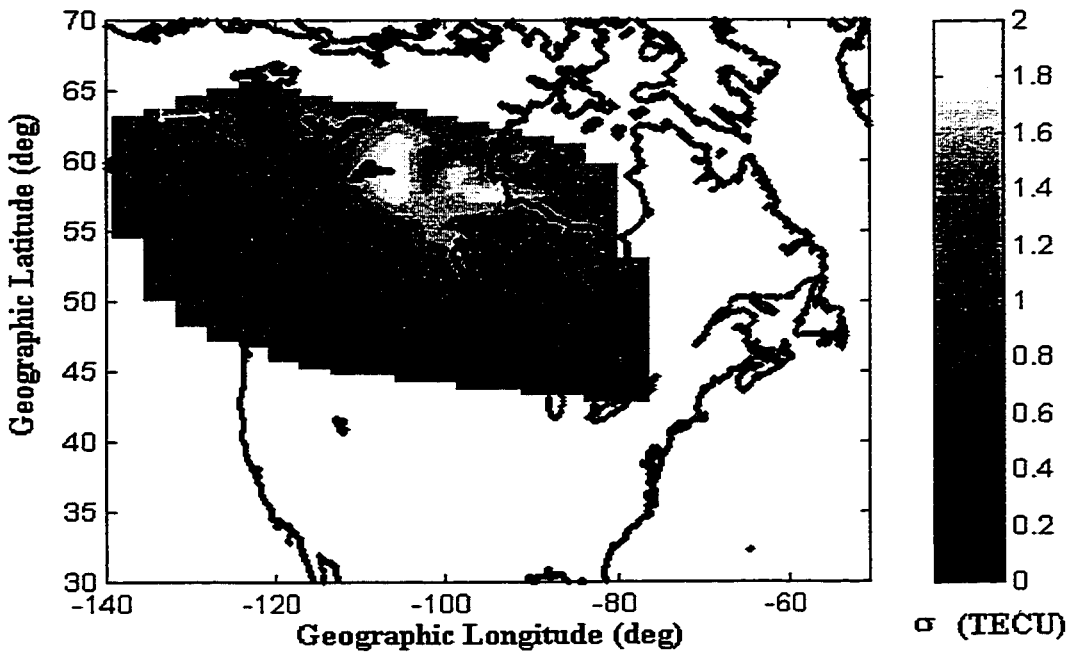


Figure 7.9 Map of process σ values, as interpolated from estimates of grid point process parameters, at 0230 UT October 23, 1996.

Figure 7.9 shows a corresponding map of the σ values at 0230 UT, as interpolated from estimates of process parameters at the designated grid points (5°×5° grid), where the largest variances are observed in a localised auroral region.

7.5 MODEL IMPLEMENTATION AND RESULTS

For purposes of testing the adaptive grid model, grid accuracies were computed for the same 22-hour active data set used in the analyses of Chapters 5 and 6. Results are presented for various grid spacings, and tests are conducted using both the Type 1 (rms of range delay accuracies) and Type 2 (residuals) tests described in Section 6.1.2.

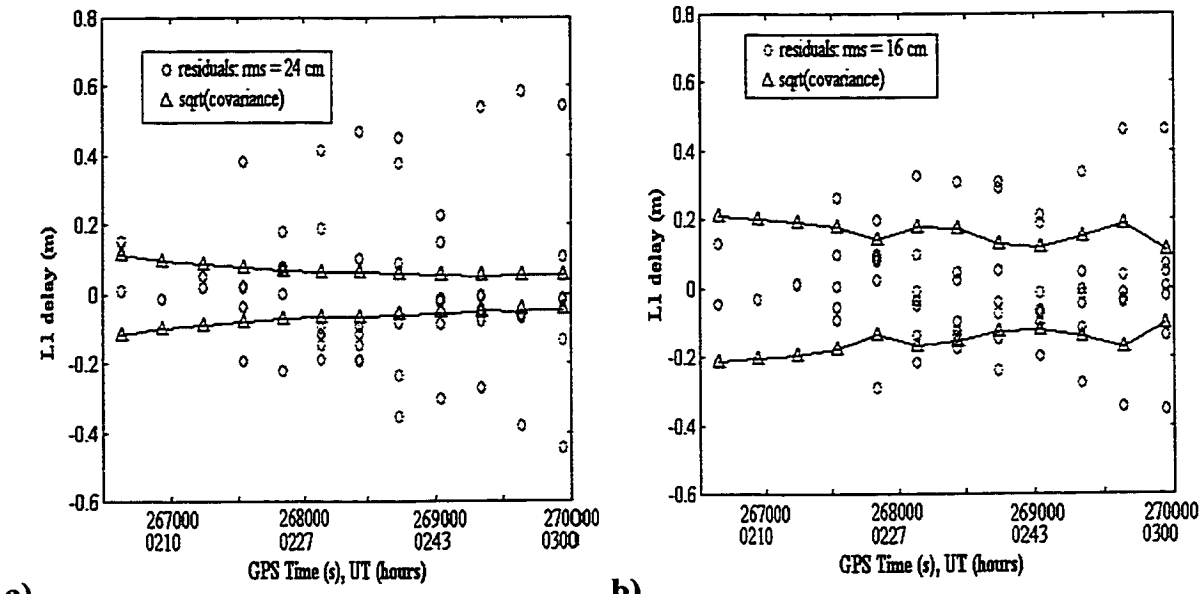
Two types of grid algorithms were tested, for comparison purposes. In the first test, a common system model was used, in which standard average values of $1/\beta_i = 15$

minutes and σ_i = 0.75 TECU are assumed for each term in Φ and Q (Equations 3.18 and 3.19). This test is conducted to generate results similar to those for existing ionosphere grid models [i.e. *Mannucci et al.*, 1993]. This model is herein referred to as the standard grid model. The adaptive system model was used in the second test, where individual terms in Φ and Q are estimated using the method outlined in Section 7.4. Prior to presenting the overall accuracies, detailed results from the October 23, 1996 data set are presented and optimality of the filter investigated.

7.5.1 Detailed Analysis - October 23, 1996

In adapting the grid system model, it is intended that the process parameters (for a given grid point) accurately reflect temporal variations in local regions of TEC. In this case, the Kalman gain matrix (Equation 3.14) better represents the optimal weighting of observations in both space and time, and the a posteriori covariance (P^+, Equation 3.13) of grid point VTEC better approximates true grid point VTEC accuracies.

As an example of filter performance, for the adaptive versus standard models, Figure 7.10 shows the covariance P^+ of VTEC (converted to L1 range delay) at a single grid point in the auroral region, during a one-hour period. In order to compare covariance values with model uncertainties, residuals are calculated and plotted for all TEC observations within the four grid cells surrounding the chosen grid point. It is assumed that these residuals reflect approximate model errors in the nearby grid point estimates. If the measurement and system models completely describe the observations, residuals should be on the order of measurement and system model uncertainties.



a) b)

Figure 7.10 Residuals calculated for a) the standard system model, and b) the adaptive system model. P^+ error bounds are also included.

Two important observations can be made from Figure 7.10:

1) The rms of residuals is decreased for the adaptive, versus standard, model by approximately 8 cm. This reflects improvements in the model performance. The maximum errors (largest residual values) are also reduced for the adaptive model, with the majority of values having magnitudes less than 40 cm and many having magnitudes less than 20 cm.

2) Overall, the adaptive covariance more accurately bounds the "true" errors (as reflected approximately by the residuals) in grid point ionosphere delay estimates. For the adaptive model, an error bound (P^+) on the order of 15 cm is observed, which is consistent with the calculated rms of residuals. In contrast, an error bound (P^+) of 5 cm is observed for the standard model, which is significantly different from the 25 cm rms residual error.

Additionally, one further observation is made from Figure 7.10. The covariance for the standard model converges to a value of 5 cm, independent of trends in the magnitudes of residuals. This is typical of filter performance for non-adaptive techniques, where covariance computations do not generally depend on input observations. In formulation of the Kalman filter, updates of the error covariance P^+ (Equation 3.13) are derived from knowledge of the transition matrix, the process noise, the measurement covariance, and the relationship between observations and unknowns (the design matrix). All of these matrices can be computed from a priori information and stochastic approximations, a fact which is often exploited in sensitivity analysis prior to filter implementation. There is no feedback between observations and filter models. For the adaptive filter, information from the observations is used to better estimate system model parameters. As observed in Figure 7.10b, this type of model does not converge to a steady state covariance; the model instead allows for an increase in error bound as ionospheric processes become more variable and the uncertainties are larger.

Note that the covariance derived for the standard model is much less than that for the adaptive model. In the Kalman filter formulation, where the model is optimal in a minimum variance sense (trace$[P^+]$ = minimum), it would appear that the standard model performs better. This is only the case, however, if the covariance values *truly* represent estimated accuracies in grid point estimates of TEC. For the adaptive case, where model residual errors are minimized *and* consistent with the estimated covariance, the grid model performance is closer to being optimal.

As an example of typical residuals for periods of enhanced auroral activity, model errors are estimated for the active data set during the period 0130-0400 UT on October 23, 1996. Figure 7.11 shows results for a grid spacing of 5 degrees. In this case, grid accuracies are consistently improved for the adaptive model, and begin to approach accuracies observed in Chapter 6 (Section 6.2.2) for non-disturbed periods.

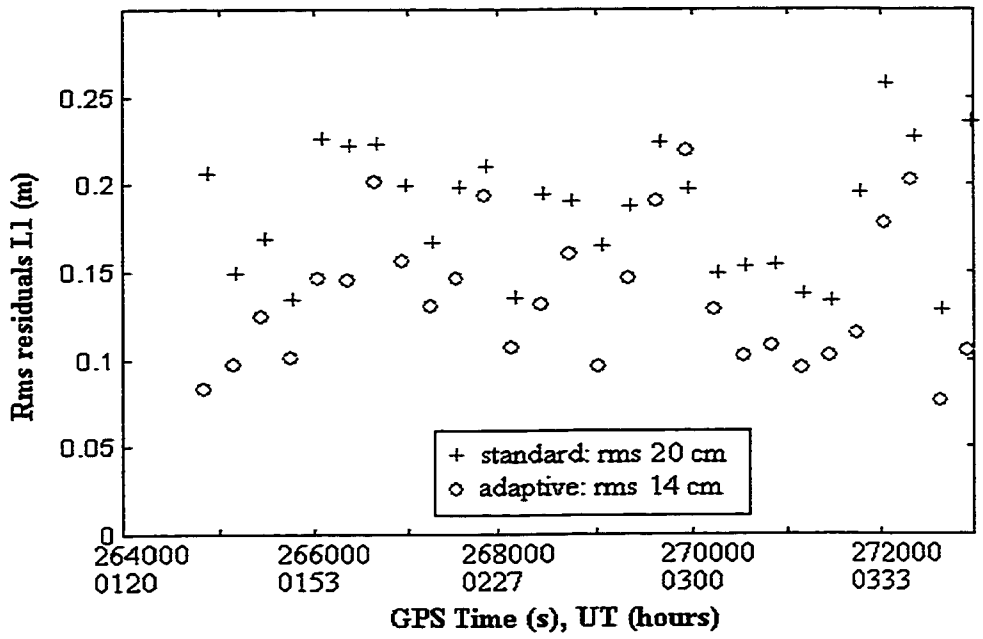


Figure 7.11 Rms of residuals calculated using the adaptive and system models, for a 5°×5° grid. Residuals are calculated from observations in the active data set on October 23, 1996.

7.5.2 Analysis of Active Data Set

Table 7.1 shows the compiled results for grid accuracies derived from the standard and adaptive grid models, for the three different grid spacings, for the entire active data set. These results are consistent with those derived for a subset of the data in Section 7.5.1. In particular, the greatest improvements are observed for the 3 degree grid spacing (≈25 percent improvement). Improvements are observed not only in model performance (as reflected in the Type 2 residual tests), but in estimated grid accuracies (Type 1 test) at independent user sites. Less significant improvements in accuracies for the 10 degree grid spacing are likely due to small-scale unresolved disturbances which dominate the error budget.

Table 7.1 Rms Grid Accuracies (L1) for Type 1 Testing

RMS (m)	GRID SPACING		
System model	**3 degrees**	**5 degrees**	**10 degrees**
standard	.33	.36	.46
adaptive	.24	.29	.40

Table 7.2 Rms Residuals (L1) for Type 2 Testing

RMS (m)	GRID SPACING		
System model	**3 degrees**	**5 degrees**	**10 degrees**
standard	.16	.21	.23
adaptive	.11	.16	.20

The probability distribution of the grid errors, for both the Type 1 and Type 2 tests, are shown in Figure 7.12. From comparisons with Figure 6.11, it is observed that the range of grid errors is minimized by using the adaptive method.

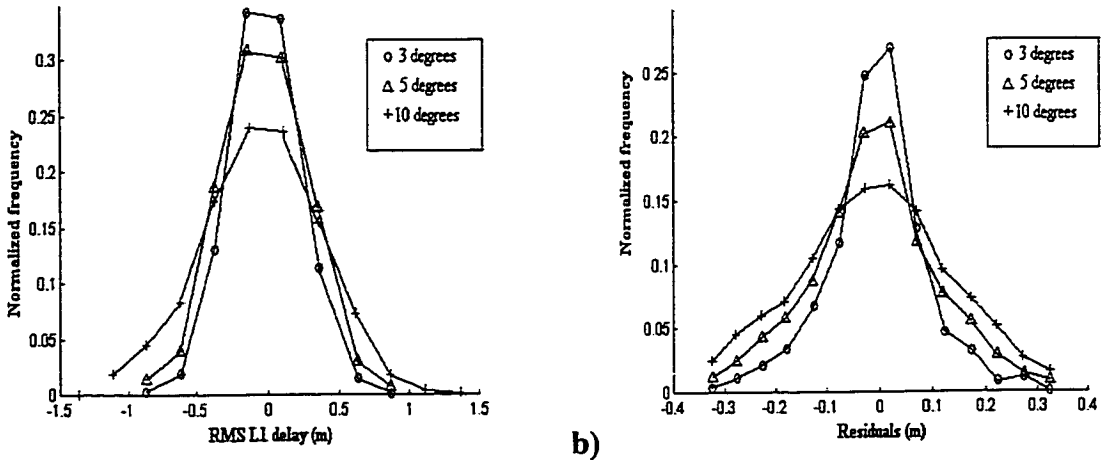


Figure 7.12 Probability distribution of a) range delay accuracies (Type 1 test), and b) residuals (Type 2 test), for the adaptive grid model.

While these accuracies indicate overall superior performance of the adaptive model, it is important to also ensure that this technique performs *consistently* better than

the standard model. An adaptive technique is not necessarily advantageous if it results in short periods of extremely poor grid accuracies. To assess consistency of the adaptive model performance, differences were computed between absolute values of observation residuals derived for the standard and adaptive models, for each individual observation. Differences are computed as standard minus adaptive, such that positive differences reflect larger residuals for the standard model. The results of this test are summarized in Figure 7.13.

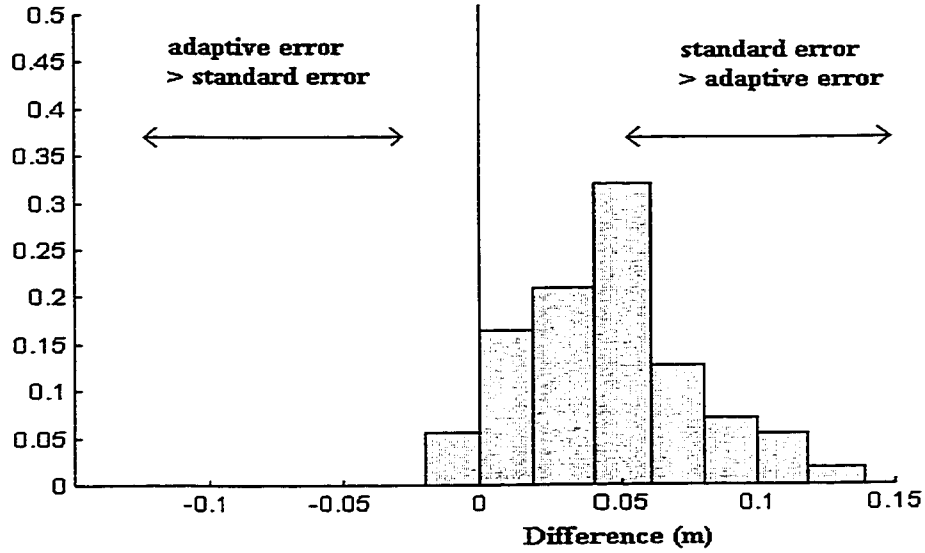


Figure 7.13 Distribution of residual differences between the standard and adaptive grid adjustments (standard minus adaptive).

The distribution of differences indicates larger (magnitudes of) residuals for the standard model approximately 95 percent of the time. These statistics therefore indicate consistently superior performance for the adaptive model. For the limited cases where adaptive model residuals are larger (negative values in Figure 7.13), the differences were less than 2 cm. These cases occur when adaptive model estimates of correlation time are less than 100 s. In such cases, despite the low pass filtering of observations (Section

7.4.1), noise effects dominate (generally for lower elevation angles - see Section 3.4.1). To minimize these effects, minimum correlation times of 100 s are specified. Additionally, model performance is monitored via the residuals.

7.5.3 Interpretation of Results

From the overall results in Tables 7.1 and 7.2, it is clear that adaptation of the system model allows for better accuracies (up to 25 percent improvement) during periods of enhanced auroral activity. Implementation of such a technique in existing wide area networks may be particularly important during the upcoming solar maximum, and it is useful to consider how the improved accuracies compare with WAAS error bounds.

As in Section 6.7, the results in Tables 7.1 and 7.2 can be multiplied by a factor of 5.43 to derive conservative GIVE error bounds. If it is also assumed that GIVE estimates are derived from independent monitor stations (i.e. Type 1 test - Table 7.1), typical GIVE values (for the adaptive model) range from 1.30 m (3 degree grid spacing) to 2.20 m (5 degree grid spacing). The GIVE values are improved by 30 - 50 cm over those for the standard model. If such an adaptive technique is combined with shorter message update intervals, further improvements in grid accuracies of $\approx$5 cm rms (Section 6.5), which translate to $\approx$25 cm in GIVE, may allow grid accuracies to satisfy system (i.e. WAAS) error bound requirements.

This adaptive technique is particularly beneficial in that it is transparent to the user. Unlike allowing for variations in the ionosphere shell height, or adjusting the grid spacing, no additional information must be sent to the user. Modifications to existing message formats are not required. Such an algorithm can therefore be implemented relatively easily in existing grid models.

It must be noted, however, that in actual WAAS implementation, grid updates are computed using 1 s data. In this case, the benefits of using such an adaptive technique have less impact on grid accuracies. The possibility does exist, however, of extending this adaptive model to estimates of spatial correlation parameters (i.e. implementation of a Gauss-Markov measurement model). For the data set studied here, refinements of the spatial model are not meaningful (due to the limited density of observations). For a larger WAAS network of more than 20 stations, the observation pierce point density may allow enhanced weighting of observations within a specified correlation distance.

7.6 THE POTENTIAL OF OTHER ADAPTIVE TECHNIQUES

It must be noted that alternative adaptive techniques exist. Such techniques often make use of the innovations sequence, which essentially consists of the "predicted residuals" (the $z-H\bar{x}$ term in Equation 3.12). Deficiencies in the system model are estimated from autocorrelation functions of this innovations sequence. System parameters may then be adjusted, such that the innovations sequence is "whitened" [cf. *Gelb*, ch. 9, 1974]. In the case of auroral ionospheric processes, it is difficult to implement such an adaptive technique, where filter error bounds are constantly changing. Additionally, it is difficult to separate the effects of measurement uncertainties, system model errors, and measurement model errors.

An adaptive filter which allows "pre-adjustment" detection of geomagnetic disturbances and isolation of the system model, as presented here, has performed better under testing. In modifying the grid model, it is intended only that real-time estimates of auroral processes be incorporated into the grid model updates. While this technique is somewhat simplistic, it has proven reliable under tests conducted here. In future, if data are available over shorter update intervals, and the density of observations is increased (i.e. with more reference stations), it may be possible to implement a more sophisticated

robust adaptive algorithm. In such cases, it may be possible to isolate measurement and system model deficiencies, using statistical testing of the innovations sequence and residuals.

7. 7 NETWORK INTEGRITY MONITORING

In wide area networks, users are generally provided a measure of estimated grid point accuracies (such as the grid ionosphere vertical error - GIVE - for WAAS). It is important, in deriving these estimates, to establish whether large error estimates are due to real variations in TEC, or reflect system abnormalities. In this section, a possible method of ionospheric integrity monitoring is briefly discussed. An example of substorm monitoring is included in Chapter 8.

In previous chapters, the correlation of TEC variations with ground-based local magnetic field perturbations was established (Sections 5.1 and 5.2). If such magnetic field data were available in real-time, these observations could be incorporated into wide area network integrity monitoring. It is preferable, however, to rely on network observations of TEC to derive measures of ionospheric activity levels. In this case, network integrity algorithms are self-contained, which is preferred for safety-critical commercial systems with liability issues.

A natural extension of the adaptive grid model developed here is the real-time detection and mapping of auroral disturbances. As demonstrated in Figure 7.8, the spatial distribution of auroral disturbances can be derived from variance estimates of TEC time series (the detection technique of Section 7.4.2). If the regional TEC variance levels, as derived from the adaptive algorithm, are monitored in network algorithms, the presence of enhanced auroral activity can be established. This technique may be used in conjunction with existing residual tests and GIVE estimates, to establish integrity of the

network. In such cases, message update intervals may be shortened, or grid spacings decreased, for periods of enhanced auroral activity (if the network design allows). Additionally, and perhaps more importantly, this mapping of auroral disturbances may be combined with knowledge of magnetospheric substorm signatures to allow wide area network operators a measure of storm tracking. This issue is addressed in Chapter 8, where various phases of a substorm event are studied in detail.

CHAPTER 8

ANALYSIS OF MAGNETOSPHERIC SUBSTORM EVENT

In Chapters 6 and 7, various aspects of the grid algorithm and accuracies were investigated during periods of enhanced auroral activity (for the active data set - Section 5.2). The identification of this active data set was based on local magnetic field signatures associated with the unloading current system (Section 4.3). As stated in Chapter 4, these signatures are often associated with the expansive phase of a magnetospheric substorm, but may also be observed during periods of enhanced directly driven activity (i.e. growth phase) [cf. *Murphree et al.*, 1991]. An expansive phase may not necessarily follow such precursor activity, and the classification of this activity as a magnetospheric substorm requires that several conditions be met. These conditions include a distinct evolution of electron precipitation patterns in the auroral ionosphere, in conjunction with near-Earth magnetic field and particle flux signatures. In the earlier chapters, no attempts were made to classify the nature of observed activity. In this chapter, however, a detailed analysis of a substorm event is conducted, using both ground-based and geostationary magnetic field signatures, in addition to observations of solar wind parameters and ultraviolet images of electron precipitation patterns.

This substorm event took place April 11, 1997, in the North American local time sector, and was not included in the 1996 active data set. This event was selected in order to illustrate various results of the previous chapters, under circumstances in which a variety of space-borne instruments were available to derive a clear picture of the substorm development. Grid accuracies are assessed during different phases of the substorm, using GPS observations from the Natural Resources Canada wide area network. This event was

relatively large (Kp = 7), and allows for independent verification of the statistics derived in earlier chapters. Additionally, the potential of monitoring substorm development using GPS, for both inclusion in integrity monitoring applications and space weather predictions, is investigated.

8.1 Description Of April 11, 1997 Substorm Event

For the substorm event studied here, several space-borne instruments were well-placed to make observations of various space weather parameters: near-Earth magnetic field variations, solar wind speed, and auroral emissions. A clear picture of this event may be developed from these observations.

The evolution of this substorm began with a solar flare on April 7, 1997. Subsequent enhancement of the solar wind speed is detected by the WIND satellite at approximately 1200 UT on April 10 (Figure 8.1).

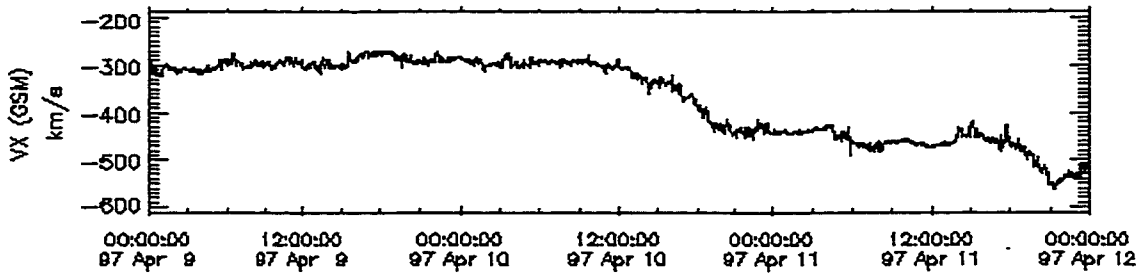


Figure 8.1. The solar wind velocity, X component (aligned parallel to the Sun-Earth line, positive sunwards), as detected by the WIND satellite (data courtesy of NASA).

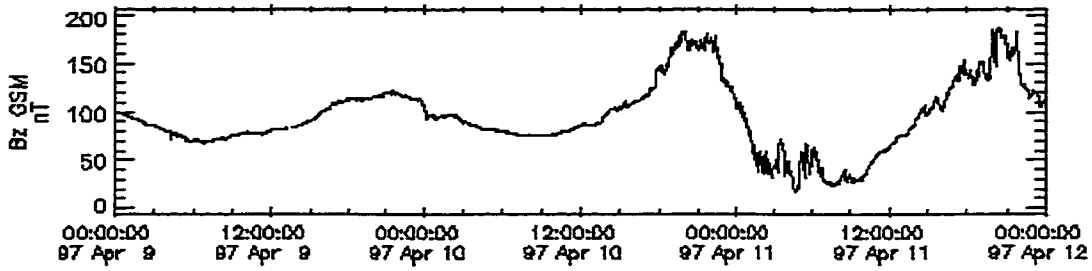


Figure 8.2 The geostationary magnetic field, Z component (aligned approximately N-S, positive northward). Typical diurnal variations are observed on April 9; substorm associated variations begin April 10 (data courtesy of NOAA).

This is accompanied by enhanced coupling of the solar wind and auroral ionosphere, leading to a period of enhanced directly driven activity. The growth of large-scale near-Earth electric current systems during this period (1200 UT April 10 onwards) is observed in geostationary magnetic field measurements. The geostationary satellite GOES 9 was located at ≈225 E geographic longitude in the North American sector. At approximately 1200 UT, April 10 1997, the near-Earth magnetic field began to vary, as compared with the "quieter" magnetic field on April 9 (Figure 8.2). These variations reflect a storing of energy in the magnetotail, some of which will be released into the auroral ionosphere at substorm onset, via a precipitation of energetic electrons. After 1200 UT, during the subsequent growth phase, stronger variations in auroral TEC are observed.

The substorm expansive phase (substorm onset) begins at approximately 0300 UT April 11, 1997 in the North American sector. Evolution of the substorm onset was observed in a series of ultraviolet images from the POLAR satellite, which was located at approximately 6 R_E altitude, over the North American sector. As energetic electrons (which have been accelerated along magnetic field lines) precipitate into the ionosphere,

they collide with neutral atmospheric constituents. Resulting emissions of visible and UV radiation can be detected optically. The POLAR imager was developed so as to be sensitive to vacuum UV wavelengths exclusively. The ultraviolet images essentially depict intensities of electron precipitation at altitudes 110-150 km. In Figure 8.3, the darker regions correspond to the most intense electron precipitation. The images are available at three minute intervals. Note, however, that only a subset of available images have been included in Figure 8.3.

The image at 0245 UT in Figure 8.3 shows a continuous auroral oval (the lighter region) at geographic latitudes 50 degrees, and possibly less (image resolution is limited by the field of view), over Canada. These latitudes are approximately 5 degrees south of typical oval boundaries, which are consistent with southward expansion of the oval during the substorm growth phase. At 0301 UT, a bright intensification is observed near the equatorward oval boundary. By 0306 UT, intensifications are observed at higher latitudes, progressively moving westward and poleward in the 0316 and 0322 UT images. These intensifications gradually dissipate until 0347 UT, when a second intensification occurs. During this expansive phase, aurora were reported as far south as New Hampshire and Boston. The movement of auroral disturbances, poleward and westward from an initial discrete equatorward arc, is a typical signature of the substorm expansive phase (cf. Section 4.5.1). At 0300 UT, the geostationary magnetic field (Figure 8.2) increases briefly. This is also a typical signature at substorm onset, resulting from a disruption of near-Earth electric currents as energetic electrons are diverted into the auroral region.

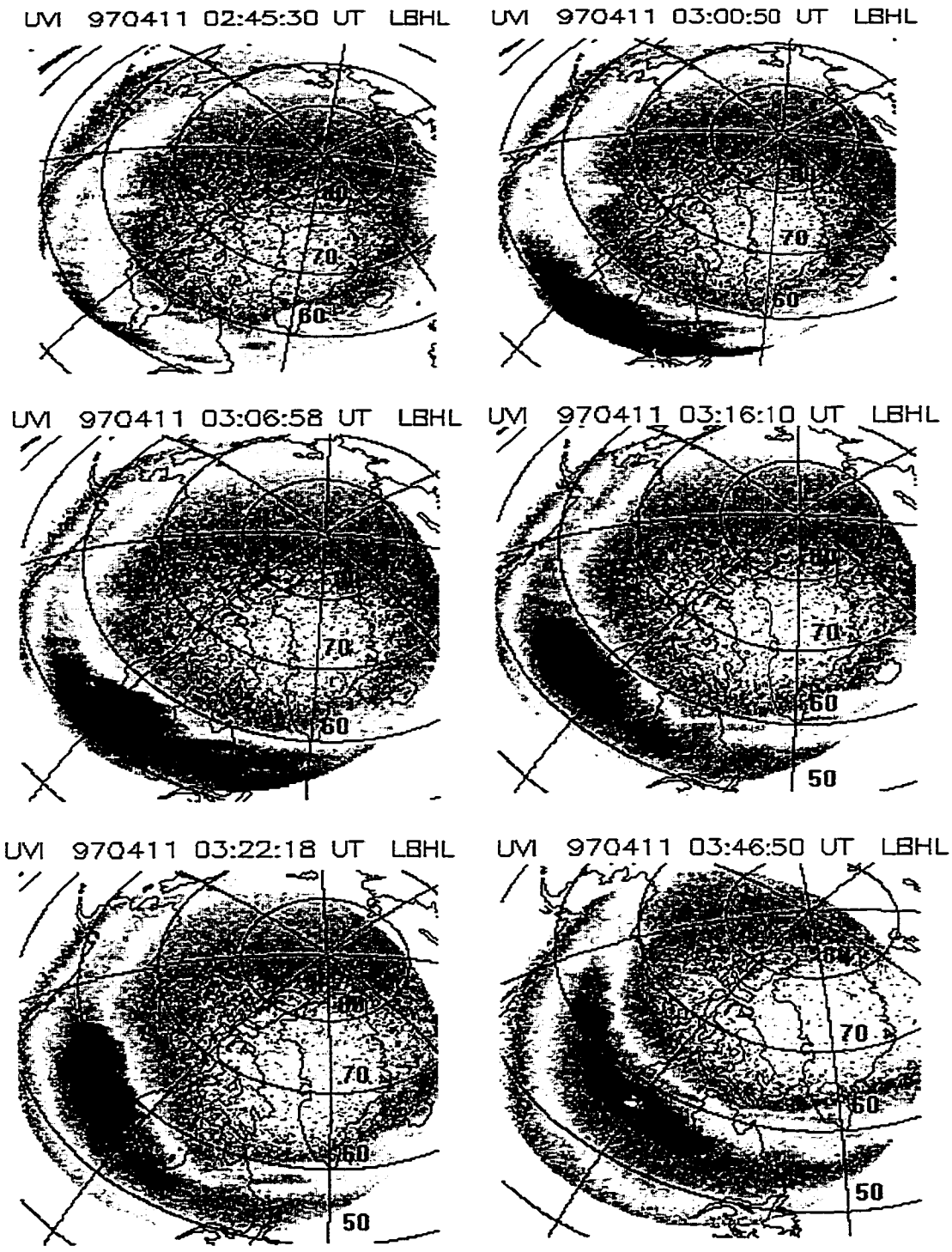


Figure 8.3 Series of ultraviolet images from the POLAR satellite, depicting auroral intensifications during the April 11, 1997 substorm (data courtesy of G. Parks, The University of Washington).

8.2 GROUND-BASED MAGNETOMETER AND TEC SIGNATURES

Evolution of the substorm event can also be observed at several stations in the CANOPUS MARIA array of magnetometer stations in Canada (these stations are pictured in Figure 5.3). Figure 8.4 illustrates the sequence of magnetic field perturbations across Canada, where initial negative perturbations begin at Gillam and Island Lake at 0300 UT, near the initial substorm onset intensification in Figure 8.3. Later magnetic field variations are observed at Fort Churchill and Fort Simpson, which are correlated with the auroral intensifications moving poleward and westward in Figure 8.3 (0316 and 0322 UT). The magnitudes of magnetic field perturbations are in excess of 1000 nT.

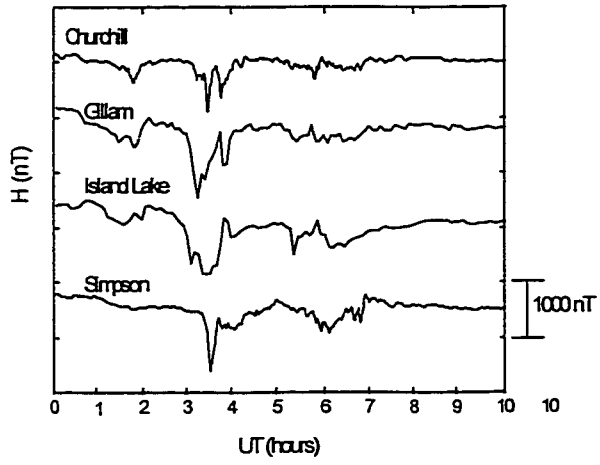


Figure 8.4 North-south magnetic field component (positive northward), at four stations in the CANOPUS MARIA array (data courtesy of the Canadian Space Agency).

Corresponding variations in TEC are computed from observations at several reference stations in the NRCan wide area network (these stations are shown in Figure 5.3). Figure 8.5 shows vertical TEC values derived from observations at four stations for

PRN 26. Corresponding tracks of ionospheric pierce points (IPPs) are shown in Figure 8.6, for each SV series plotted in Figure 8.5.

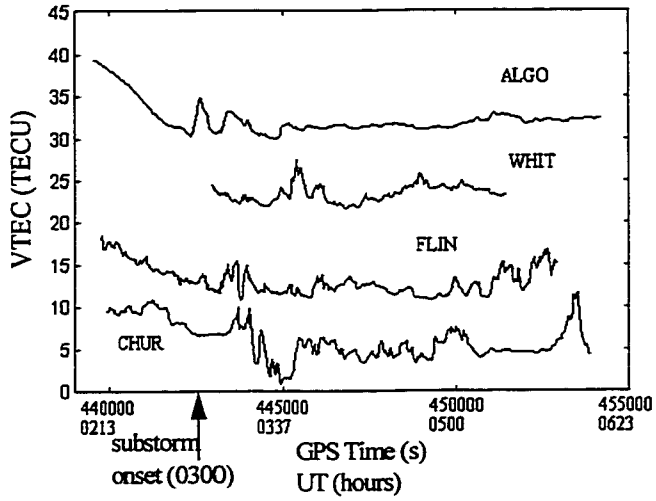


Figure 8.5 Vertical TEC variations for PRN 26, as observed from four reference stations in the NRCan network. Series have been offset for comparison purposes.

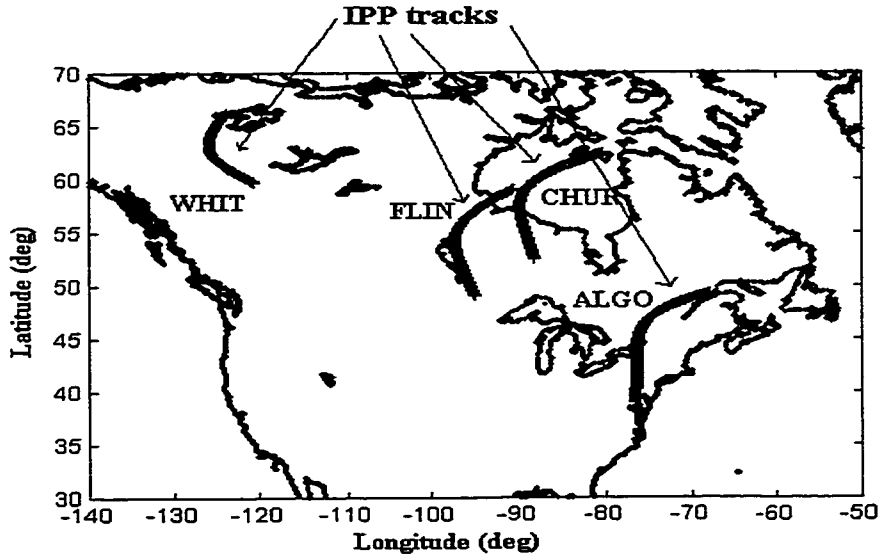


Figure 8.6 Line-of-sight ionospheric pierce points (IPPs), corresponding to Figure 8.5.

Prior to substorm onset (0300 UT), variations in TEC are observed at both Flin Flon and Fort Churchill. These variations correspond to enhanced variations in electron density, in the auroral oval, during the substorm growth phase. At substorm onset, large variations are observed at Algonquin, corresponding to the bright equatorward arc observed in Figure 8.3 (0301 UT). As the intensifications move poleward, increasingly large variations in TEC are observed at Flin Flon and Fort Churchill. At 0345 UT, effects of the second intensification (Figure 8.3) are observed further west, at Whitehorse.

8.3 SPECTRAL ANALYSIS OF TEC SERIES

In this section, approximate estimates of spatial and temporal parameters are derived by conducting spectral analysis of the TEC time series. This analysis is similar to that conducted in Section 5.6. In Figure 8.5, periods of 30 minutes on the x-axis correspond to distances in the range 40-300 km in the ionosphere shell, depending on the elevation angle of the satellite at a given station. For a given satellite-receiver series of TEC, 64-point Fourier and Lomb Transforms (Appendix C) are computed from successive windows of 64 observations and autocorrelation functions estimated. Only the TEC values between 0300 and 0400 UT (the expansive phase of the storm) are considered.

Figure 8.7 shows the normalized autocorrelation functions for both temporal and spatial variations in vertical TEC. For temporal variations, a correlation time of 120 s is derived, while the correlation distance is 50 km. These parameters are consistent with the statistics derived in Chapter 5 (Section 5.6), and both autocorrelation functions approximate a first order Gauss-Markov stochastic model.

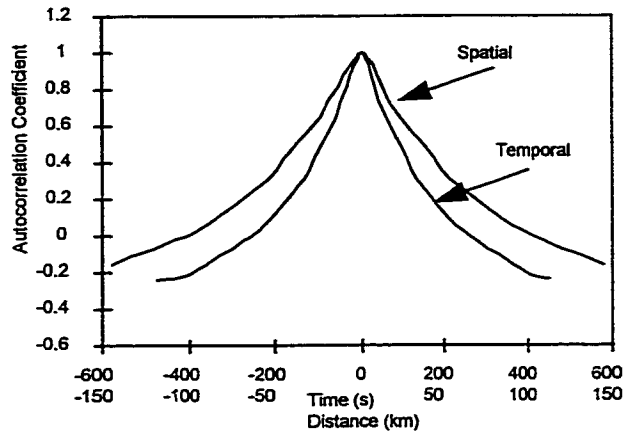


Figure 8.7 Normalized autocorrelation functions, in space and time, for TEC observations during the period 0300-0400 UT April 11, 1997.

Additionally, it is possible to isolate specific frequency/wavelength ranges, and estimate variances of the corresponding harmonic components (as in Section 5.6). For analysis in the frequency domain, mean squared amplitudes of $\sigma^2 = 3$ TECU^2 ($\sigma \approx 28$ cm L1 ionospheric delay) are derived for those temporal variations with periods in the range 5-20 minutes (i.e. approximate time scales of auroral disturbances). For analysis in the wave number domain, mean squared amplitudes of $\sigma^2 = 2.3$ TECU^2 ($\sigma \approx 24$ cm L1 ionospheric delay) were derived for those spatial variations with wavelengths less than 400 km (i.e. typical scale sizes of auroral disturbances [*Coker et al.*, 1995]). These values are also consistent with results derived in Section 5.6, for the active data set of 1996. Amplitudes as large as 8 TECU were observed for TEC series in the more intense regions of substorm onset.

8.4 IONOSPHERE GRID TESTING

8.4.1 Grid Accuracies

Magnitudes of grid inaccuracies, during different phases of the storm development, are evaluated in this section. The choice of grid model parameters includes a multiquadric weighting function (W) for the measurement model (i.e. Section 3.3.2), and the adaptive system model of Chapter 7. Updated grid solutions were estimated at five minute intervals, using the previous five minutes of observations. A grid spacing of 3 degrees is assumed for the underlying solar-magnetic grid (cf. Section 6.1.1), and estimates of TEC at Earth-fixed grid points are derived for a 5 degree spacing at latitudes below 55 degrees and 10 degrees at higher latitudes. This is consistent with WAAS specifications [*FAA*, 1997]. Rms of range delay accuracies are derived (i.e. Type 1 test - Section 6.1.2), in order to estimate accuracies of ionosphere range corrections at an independent user site.

Figure 8.8a shows grid accuracies for the period 0200-0300 UT, during the substorm growth phase. An overall accuracy of 36 cm rms was determined, for all available test observations. Larger rms values are observed in the region 45 to 60 degrees latitude, -90 to -120 degrees longitude. These degraded accuracies likely correspond to the TEC variations observed from Flin Flon and Fort Churchill (Figure 8.5), prior to substorm onset.

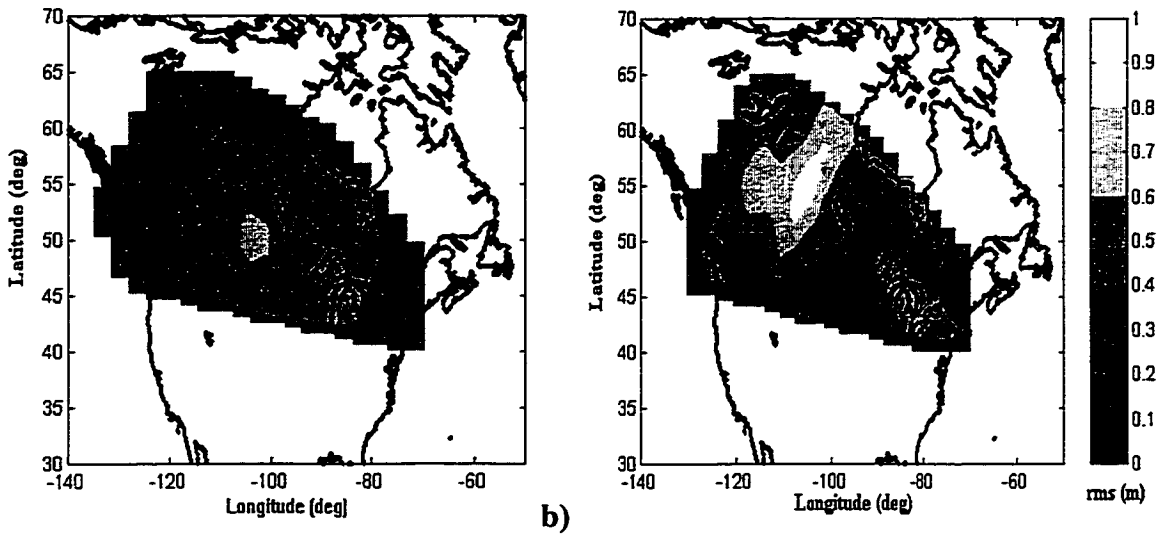


Figure 8.8 **a) Average grid accuracies during the growth phase period 0200-0300 UT, April 11 1997.**

b) Average grid accuracies during the expansive phase period 0300-0400 UT, April 11 1997.

Figure 8.8b shows grid accuracies for the substorm onset period, 0300-0400 UT. Larger rms values are observed in the high latitude region, with accuracies as poor as 90 cm rms in the region west of Hudson's Bay. This region corresponds to the intense electron precipitation observed in Figure 8.3. An overall accuracy of 48 cm was determined during this period, for all available test observations.

Note that the grid accuracies are better in Figure 8.8b, as compared with Figure 8.8a, for latitudes near 45 degrees. This may result from the equatorward auroral boundary beginning to recover and retract northward after the substorm onset. At the end of the growth phase, theaded grid accuracies are evident at latitudes as low as 45 degrees, consistent with observed equatorward expansion of the oval in Figure 8.3. It must also be noted that the adaptive technique allows for improved accuracies in this analysis. For

"typical" model parameters (i.e. the standard model of Section 7.5), accuracies are further degraded by 10-15 cm rms.

8.4.2 Interpretation of Grid Accuracies

It is observed that grid accuracies are degraded by a factor of ≈2 during the substorm growth phase, as compared with typical quiet time accuracies of 20 cm rms [cf. *Bertiger et al.*, 1997]. These larger errors are present over a time scale of hours, consistent with the period of growth phase development and enhanced directly driven activity. Due to the equatorward expansion of the oval during this period, such degraded accuracies affect low elevation observations for users as far south as 38 degrees geographic latitude.

Grid accuracies are further degraded during the substorm expansive phase, by a factor of 2-5 in local regions. These degraded accuracies, while larger than those observed during the growth phase, are transient and more localised. The magnitude of these accuracies, in the range .40-1.0 m, are consistent with those derived for the active data set in Chapter 6 (i.e. Section 6.2 - Type 1 test, Figure 6.8). If these accuracies are multiplied by a confidence factor of 5.43 in order to derive approximate GIVE error bounds (cf. Section 6.7), growth phase GIVE values are on the order of 2.0 m, and expansive phase GIVE values are in the range 2.0-5.0 m. These estimates are consistent with mid-latitude storm GIVE values derived in *Conker et al.* [1997] and exceed WAAS error bounds (if an error bound of GIVE ≤ 1.5 m is assumed - cf. Section 3.2). Note that the adaptive model, as used here, allows an improvement of ≈50 cm rms in the GIVE values. Overall, the results of this storm analysis are consistent with those determined in Chapter 6.

8.5 MONITORING OF IONOSPHERIC ACTIVITY

As briefly discussed in Section 7.7, the adaptive grid algorithm allows for real-time detection and mapping of auroral disturbances. Such algorithms can be incorporated into network integrity monitoring, to establish the nature of system abnormalities. Of particular interest is the potential of using TEC signatures to monitor the evolution of substorm development. As stated in Section 4.5.1, the various phases of substorm development are characterized by distinct signatures. Equatorward expansion of the auroral oval has been observed consistently during substorm growth phases. If oval expansion can be monitored during the growth phase, system operators may be able to assess the level of ionospheric activity and the probability of subsequent expansive phase activity. As an example, Figure 8.9 shows the distribution of auroral disturbances during the growth phase period, as derived from the adaptive grid model algorithm.

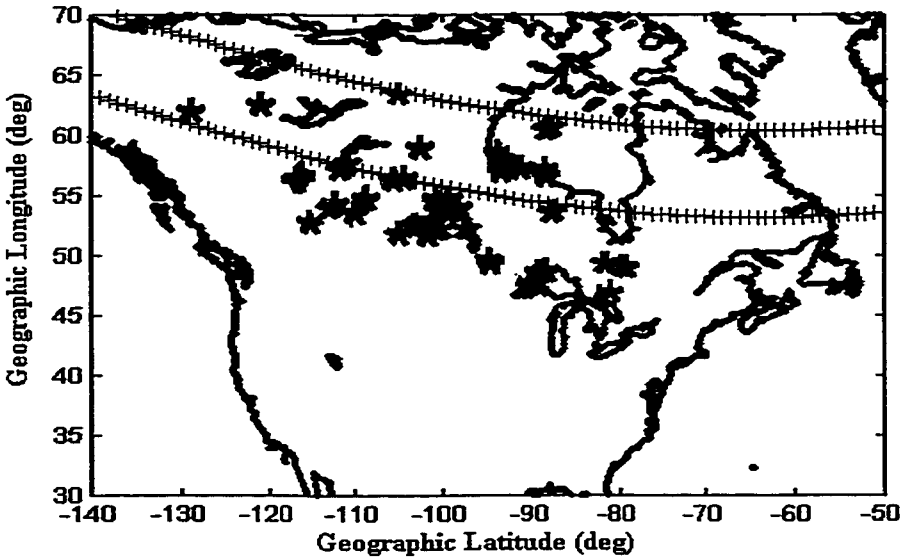


Figure 8.9 Distribution of auroral disturbances at 0230 UT April 11, 1997. Typical oval boundaries (++++) are plotted for comparison.

In this figure, many disturbances are detected below the "typical" equatorward oval boundary. This indicates an equatorward expansion of the oval of ≈5 degrees, which is consistent with observed oval boundaries in Figure 8.3 (0245 UT). While this oval expansion does not necessarily lead to the conclusion that an intense expansive phase will follow, it does reflect enhanced coupling of the solar wind-magnetosphere dynamo, with increased energy storage in the Earth's magnetotail. If loss of GPS signal lock and signal tracking errors are subsequently observed at several network reference stations, such system abnormalities may be attributed to intense scintillation effects at an expansive phase onset.

Development of the substorm expansive phase can also be monitored "real-time" from the GPS observations. This is demonstrated in Figure 8.10, where snapshots of the larger disturbances have been plotted at various times during the expansive phase. The poleward and westward motion of auroral features is observed in these maps (consistent with sequence of POLAR images - Figure 8.3), demonstrating the potential of monitoring auroral activity, and substorm development, using GPS observations. Additionally, as demonstrated in Section 6.2.3, the development of large-scale features, such as the main trough, can be observed from the GPS observations.

Note that the tracking of substorm development presented here is not necessarily representative of all substorm events. In some events, there are few precursors to the substorm expansive phase, such that tracking of growth phase auroral disturbances is not meaningful. Additionally, it is hard to establish a threshold limit of oval expansion at which an expansive phase will most likely occur (if it occurs at all). Monitoring ionospheric activity, using the GPS observations, simply allows a measure of oval expansion, the magnitude of auroral disturbances, and tracking of expansive phase development.

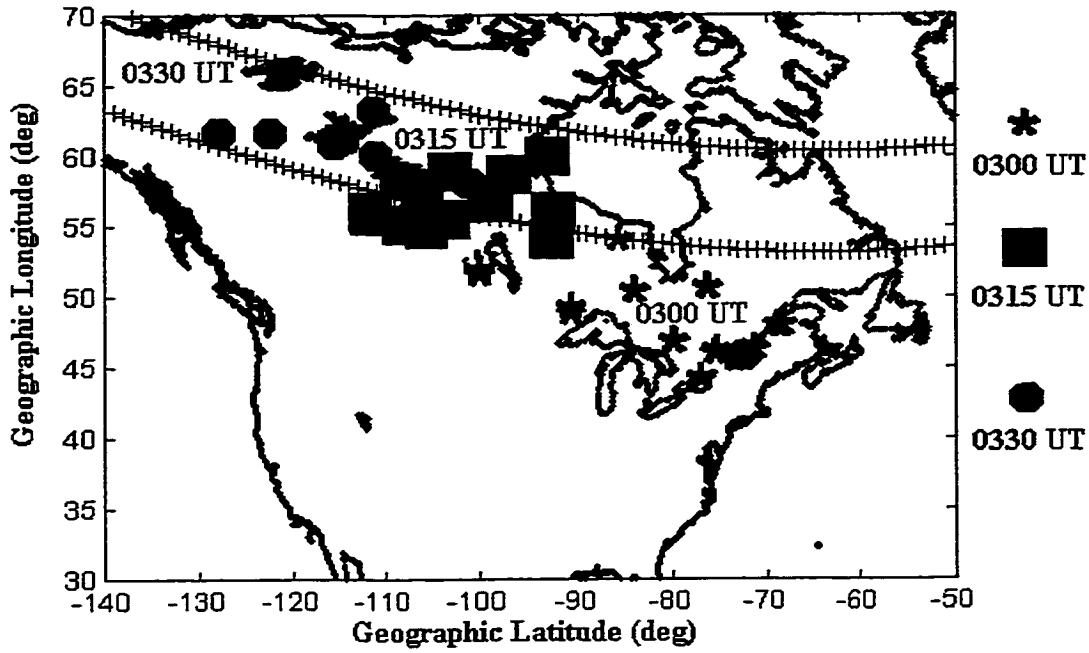


Figure 8.10 Development of the substorm expansive phase, at three different times during the period 0300-0400 UT. Only disturbances with $\sigma > 2$ TECU are plotted.

It is the intention, in presenting the analysis of this section, to also demonstrate the potential of including GPS observations in "nowcasting" of space weather. Ground-based magnetometer measurements have been used for many years to track substorm development and expansion of the oval boundaries. Magnetic field measurements from all latitudes are incorporated into various indices describing levels of ionospheric and magnetospheric activity (i.e. AE, Kp, Dst). In future, GPS measurements may allow for another form of ionospheric activity monitoring, with the derivation of alternative GPS-based local and global indices. With the establishment of wide area networks in many global regions, the density of GPS observations, coupled with the real-time nature of wide area network operations, may provide a useful tool in space weather predictions and space physics research.

CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS

As stated in Chapter 1, there has been an increasing focus on ionosphere grid modelling, as applied to wide area differential GPS networks, in the past few years. It is proposed that the wide area augmentation system (WAAS [*FAA*, 1997]), and similar systems (Canadian WAAS), be implemented in the auroral region. Given the safety critical aspect of these systems, and the volatile nature of auroral ionospheric phenomena, it is important to evaluate the performance of wide area ionospheric grid models in this region. In this thesis, the first comprehensive study of auroral effects on GPS, as applied to investigations of the grid model accuracies, has been conducted. An adaptive algorithm, designed to improve grid accuracies, has also been introduced. The results of this research are summarized here.

9.1 CONCLUSIONS

Global geomagnetic indices, which are often used in GPS-related studies of the ionosphere [i.e. *Chavin*, 1996], are not necessarily good indicators of local ionospheric activity. While extremely large indices may reflect the presence of global disturbances, other measures must be used to identify the duration and spatial extent of local disturbances. By using an extensive array of magnetometer stations within the wide area grid sector, levels of local activity were monitored, such that periods of auroral activity were identified consistently (the 1996 "active" data set). TEC observations compiled during these periods were found to have consistently larger spatial and temporal gradients, as compared with quiet periods. This is in contrast to earlier studies [*Mannucci*

et al., 1997], in which no consistent statistics were observed for disturbed levels of ionospheric activity. Correlation of local TEC variations with magnetic field perturbations was demonstrated.

It was established that auroral irregularities in TEC are transient processes, which can be approximated as first order Gauss-Markov processes. Correlation times of several minutes are associated with these disturbances, with dominant periods on the order of 10 minutes. Corresponding amplitudes are on the order of 1-2 TECU, which translates to 20-30 cm L1 range delay. Spatial correlation statistics indicate scale sizes on the order of several hundred kilometres or less, which are consistent with previous studies of TEC features associated with periods of AEI [cf. *Hunsucker et al.*, 1995].

The spatial correlation statistics suggest that grid spacings of 3 degrees or less are preferable, in order to resolve the small-scale ionospheric disturbances. For such grid spacings, accuracies on the order of 33 cm rms were observed, which are a factor of $\approx$1.5 larger than typical grid model accuracies. Degraded accuracies as large as 1.0 m were observed in local regions. For larger grid spacings of ten degrees, overall accuracies of 46 cm rms were observed (a factor of $\approx$2.5 worse than quiet time accuracies). Corresponding estimates of GIVE values were 2.50 m, for the larger grid spacing, exceeding approximate WAAS error bounds. By considering magnetic field perturbations from the previous solar maximum, it is predicted that these grid model errors (and GIVE values) will increase by a factor of 2 at solar maximum, during intense storms.

It has been established that the presence of auroral disturbances will challenge the capabilities of wide area networks operating in northern regions. In order to achieve accuracies on the order of 30 cm rms, and maintain a grid spacing of 3 degrees, reference station separations on the order of 600 km or less are recommended. Additional improvements of 5-10 cm rms in grid accuracies may be achieved by using a shorter

update interval for the ionosphere grid message. Investigations of inaccuracies associated with the ionospheric shell height assumption were inconclusive, with no improvements in grid accuracies observed for varying shell heights. It is anticipated that 3-dimensional modelling of the electron density, using a tomography technique, may allow for improvements in model accuracies.

In this thesis, a revised ionosphere grid model has been developed for the auroral region. This model allows variation of system model parameters, to best reflect local ionospheric conditions. This model differs from existing grid models in two ways: 1) the system model at each grid point is allowed to vary in time (i.e. *typical* constant values of process noise are not generally assumed), and 2) independent system models are used for each grid point, such that regional variations in ionospheric temporal processes are included in the grid adjustment.

Information about ionospheric temporal processes is obtained by deriving autocorrelation functions of observed TEC values. In this way, the model is able to follow large temporal variations in TEC associated with AEI and other geomagnetic disturbances in the auroral region. Grid accuracies are improved by 15-25 percent using such an adaptive grid model, as compared with accuracies for the non-adaptive case, and corresponding GIVE values are improved by 30-40 cm. It is anticipated that implementation of an adaptive ionospheric model may help to achieve grid accuracies within WADGPS network error bounds. The detection of auroral disturbances from GPS observations, a by-product of the adaptive technique, may be incorporated into network integrity monitoring algorithms, where message update intervals may be shortened to accommodate the rapid variations in TEC.

The evolution of a magnetospheric substorm (April 11, 1997) in the North American sector was monitored in detail, such that development of the distinct substorm

phases was established. During the substorm growth phase, equatorward motion of the auroral oval over the Northern United States was observed, with degraded grid accuracies of 40-60 cm rms (a factor of 2-3 worse than typical accuracies) at latitudes of 45 degrees. Such degraded accuracies affect users as far south as 38 degrees latitude, where lines-of-sight for low elevation satellites may penetrate the oval region.

During the substorm expansive phase, the most volatile period of the storm, significant variations in TEC were observed at various reference stations in Canada. Observed amplitudes of TEC variations reached 8 TECU, and the spatial distribution of auroral intensifications varied over a time scale of minutes (as observed in the POLAR UV images). Such rapid and significant variations in TEC are difficult to model, and corresponding wide area grid accuracies of .60-1.0 m rms were observed in the most intense region of the substorm onset. This region covered approximately 20 degrees of longitude, at latitudes several degrees above the Canada-U.S. border. These effects were consistent with those observed from the 1996 "active" data set. The possibility of monitoring levels of ionospheric activity, and the development of substorm phenomena, was explored for this event. In future, network GPS TEC observations may be used to establish local levels of ionospheric activity. Such a "self-contained" method of detecting ionospheric disturbances within the network is preferable to relying on independent data sources (such as geomagnetic indices).

9.2 RECOMMENDATIONS FOR FUTURE RESEARCH

While the research presented in this thesis has established guidelines for ionosphere grid model implementation in the auroral region (with associated estimates of grid model errors), there exists potential for future research. Clearly, a natural extension of this research is to continue investigating grid model accuracies, and the impact of network station spacing, as larger substorms are observed over the next few years.

Preliminary analysis of a major substorm from May 1998 indicates grid accuracies similar to those observed in the April 11, 1997 event. Larger events from August and September 1998 are also of interest. As this activity becomes more frequent, and increases in magnitude, the potential exists for establishing a definitive data base (the first of its kind) to describe the range of auroral effects on GPS applications.

A real-time reliable GIVE algorithm also needs to be developed, to bound the grid vertical ionospheric errors in the auroral region. Such an algorithm could be very useful for the Canadian WAAS precision approach phase of flight. The GIVE algorithm could be derived from existing algorithms [i.e. *Conker et al.*, 1997], with extensive testing to ensure reliability. It is also important to determine the success rate of bounding user errors with the user ionosphere vertical error (UIVE), as derived from the GIVE estimates.

A further issue associated with auroral activity is the presence of scintillation effects. In particular, there is presently a great deal of interest in anticipated GPS receiver performance during solar maximum. The presence of strong scintillation effects is expected to degrade receiver performance, such that tracking errors and loss of signal lock will be observed. Although the magnitude of auroral scintillation effects have been established in previous studies [i.e. *Aarons*, 1982], questions remain as to how extensive these effects will be (i.e. Will all satellites in view be affected simultaneously? How transient are these effects?). Studies of receiver performance during future substorm events will be valuable for all GPS applications.

In addition to wide area networks, GPS applications are currently being developed for *local area networks* where stations are spaced several hundred kilometres apart and accuracies at the centimetre-level are specified. In order to achieve such accuracies, higher order ionospheric effects must be modelled. Given the estimated spatial

decorrelation of TEC in the auroral region, challenges will exist in modelling the ionosphere in such networks during periods of enhanced auroral activity. A study similar to that conducted here is necessary to establish anticipated network performance.

REFERENCES

Aarons, J., Global morphology of ionospheric scintillations, *Proceedings of the IEEE*, **70**, no. 4, 360, 1982.

Akasofu, S.-I., *Polar and Magnetospheric Substorms*, Reidel, Dordrecht, 1968.

Baron, M.J., Electron densities within aurorae and other auroral E-region characteristics, Radio Science, **9**, no. 2, 341, 1974.

Bates, H.F., A.E. Belon and R.D. Hunsucker, Aurora and the poleward edge of the main trough, *Journal of Geophysical Research*, **78**, no. 4, 648, 1973.

Bates, H.F. and R.D. Hunsucker, Quiet and disturbed electron density profiles in the auroral zone ionosphere, *Radio Science*, **9**, 455, April, 1974.

Bertiger, W.I. et al., A prototype real-time wide area differential GPS system, *Proceedings of the ION GPS-97 National Technical Meeting,* Santa Monica, California, January, 1997.

Birkeland, K., "The Norwegian Aurora Polaris Expedition 1902-1903", Vol. I, 1, H. Aschehoug and Co., Christiana, 1908.

Bowen, R. et al., "GPS Control System Accuracies", in *Global Positioning System, Vol III*, Institute of Navigation, Washington, D.C., 241, 1986.

Braasch, M.S., "Multipath Effects", in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Brigham, E.O., *The Fast Fourier Transform*, Prentice-Hall, Inc., NJ, 1974.

Brunner, F.K. and M. Gu, An improved model for the dual frequency ionospheric correction of GPS observations, *Manuscripta Geodaetica*, **16**, 205, 1991.

Caissy, M., P. Héroux, F. Lahaye, K. MacLeod, J. Popelar, J. Blore, D. Decker and R. Fong, Real-time GPS corrections service of the Canadian Active Control System, *Proceedings of the ION GPS-96*, 1787, Kansas City, Missouri, September, 1996.

Chapman, S. and J. Bartels, *Geomagnetism*, Clarendon Press, Oxford, England, 1940.

Chavin, S., Ionospheric specification for the wide area augmentation system (WAAS) simulation studies, *Proceedings of the ION GPS-96*, 585, Kansas City, Missouri, September 1996.

Chavin, S., Magnetic storm modeling for the wide area augmentation system (WAAS) simulation studies, *Proceedings of the ION GPS-97 National Technical Meeting,* Santa Monica, California, January, 1997.

Clauer, C.R. and R.L. McPherron, Mapping the local time-universal time development of magnetospheric substorms using mid-latitude magnetic observations, *Journal of Geophysical Research*, **79**, 2811, 1974.

CNES: Centre National d'Etudes Spatiales, STARLETTE, Toulouse, 1975.

Coker, C., R. Hunsucker and G. Lott, Detection of auroral activity using GPS satellites, *Geophysical Research Letters*, **22**, 3259, December, 1995.

Conker, R., B. El-Arini, T. Albertson, J. Klobuchar and P. Doherty, Description and assessment of real-time algorithms to estimate the ionospheric error bounds for WAAS, *NAVIGATION: Journal of the Institute of Navigation*, **44**, 77, Spring, 1997.

Conker, R. and M.B. El-Arini, A novel approach for an ionospheric obliquity process responsive to azimuth variation, *Proceedings of the 54th Annual Meeting of the Institute of Navigation*, Denver, Colorado, June, 1998.

Cooley, J.W. and J.W. Tukey, An algorithm for machine calculation of complex Fourier series, *Math. Computation.*, **19**, April, 1965.

Daniell R.E., L.D. Brown and R.W. Simon, A new improved ionospheric correction algorithm for single frequency GPS receivers, *Proceedings of the ION GPS-96*, 635, Kansas City, Missouri, September, 1996.

Defense Mapping Agency, "World Geodetic System 1984 (WGS-84) - Its Definition and Relationship with local Geodetic Systems", DMA TR 8350.2 second edition, Fairfax, Virginia, 1984.

DoD, "The Global Positioning System - Charting the Future", National Academy of Public Administration, Washington, D.C., 1995.

Draganov, A., T. Cashin and J. Murray, An ionospheric correction algorithm for WAAS and initial test results, *Proceedings of the ION GPS-96*, 789, Kansas City, Missouri, September, 1996.

Easton, R.L., The Navigation Technology Program, in *Global Positioning System, Vol. I*, Institute of Navigation, Washington, DC, 20, 1980.

El-Arini, M.B., R. Conker, T. Albertson, J.K. Reegan, J.A. Klobuchar and P. Doherty, Comparison of real-time ionospheric algorithms for a GPS wide-area augmentation system (WAAS), *NAVIGATION: Journal of the Institite of Navigation*, **41**, no. 4, 393, winter 1994-95.

Enge, P.K. and A.J. Van Dierendonck, Wide Area Augmentation System, in *Global Positioning System: Theory and Applications, Vol. II*, ed. B.W. Parkinson and J.J. S Spilker, American Institute of Aeronautics and Astronautics, Washington, DC, 1996.

Eueler, H. and C.C. Goad, On optimal filtering of GPS dual-frequency observations without using orbit information, *Bulletin Geodesique*, **65**, no. 2, 130, 1991.

FAA, "Specification for the Wide Area Augmentation System", FAA-E-2892, 1994.

FAA, "Specification for the Wide Area Augmentation System", FAA-E-2892C (draft), 1997.

Fenton, P.C., Falkenberg, W.H., Ford, T.J., Ng, K.K. and A.J. Van Dierendonck, NovAtel's GPS receiver: the high performance OEM sensor of the future, *Proceedings of the ION GPS-91*, 49, Albuquerque, NM, September, 1991.

Feldstein, Y.I. and G.V. Starkov, Dynamics of auroral belt and polar geomagnetic disturbances, *Planetary Space Science*, **15**, 209, 1967.

Fraile-Ordonez, J.M., Ionospheric total electron content determination from dual-frequency GPS measurements for real-time applications in satellite navigation, *J.P.N.*, **3**, 88, 1995.

Gao, Y., P. Heroux and J. Kouba, Estimation of GPS receiver and satellite L1/L2 signal delay biases using data from CACS, *Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation*, 109, Banff, August 30-September 2, 1994.

Gelb, A., *Applied Optimal Estimation*, M. I. T. Press, 1974.

Goad, C.C., Optimal filtering of pseudoranges and phases from single-frequency GPS receivers, *NAVIGATION: Journal of the Institite of Navigation*, **37**, no. 4, Fall, 1990.

Goertz, C.K. and R.W. Boswell, Magnetosphere-ionosphere coupling, *Journal of Geophysical Research*, **84**, 7239, 1979.

Guier, W.H., Navigation using artificial satellites - the TRANSIT System, *Proceedings of 1st International Symposium for Use of Artificial Satellites in Geodesy*, Washington, D.C., 1962.

Hahn, G.J. and W.Q. Meeker, *Statistical Intervals, A Guide for Practitioners*, John Wiley and Sons, 1991.

Hansen, A., T. Walter and P. Enge, Ionospheric correction using tomography, *Proceedings of the ION GPS-97*, 249, Kansas City, Missouri, September, 1997.

Hapgood, M.A., Space physics coordinate transformations: a user's guide, *Planetary Space Science*, **40**, 711, 1992.

Harang, L., The mean field of disturbance of polar geomagnetic storms, *Terrestrial Magnetic and Atmospheric Effects*, **51**, 353, 1946.

Hargreaves, J. K., *The Solar-Terrestrial Environment*, Cambridge University Press, Cambridge, 1992.

Hasegawa, A. and T. Sato, Generation of field-aligned current during substorm, in *Dynamics of the Magnetosphere*, ed. S.-I. Akasofu, D. Reidel, Hingham, Massachusetts, 1979.

Hoffmann-Wellenhof, B. Lichtenegger, H. and J. Collins, *Global Positioning System: Theory and Practice*, Springer-Verlag, New York, third edition, 1994.

Hopfield, H.S., Two-quartic tropospheric refractivity profile for correcting satellite data, *Journal of Geophysical Research*, **74**, no. 18, 4487, 1969.

Huang, C.Y. and L.A. Frank, A statistical study of the central plasma sheet: Implications for substorm models, *Geophysical Research Letters*, **13**, 652, 1986.

Hunsucker, R.D., Mini-review: auroral and polar-cap ionospheric effects on radio propagation, *IEEE Transactions on antennas and propagation*, **40**, no. 7, 818, 1992.

Hunsucker, R.D., C. Coker, J. Cook and G. Lott, An investigation of the feasibility of utilizing GPS/TEC "signatures" for near-real time forecasting of auroral-E propagation at High-HF and Low-VHF frequencies, *IEEE Trans. on Antennas and Propagation*, **43**, no. 10, 1313, 1995.

Iijima, T. and T.A. Potemra, Large-scale characteristics of field-aligned currents associated with substorms, *Journal of Geophysical Research*, **83**, 599, 1978.

Junkins, J.L., G.W. Miller and J.R. Jancaitis, A weighting function approach to modeling of irregular surfaces, *Journal of Geophysical Research*, **78**, 11, 1794, 1973.

Kaplan, E.D., *Understanding GPS Principles and Applications*, Artech House, Norwood, Massachusetts, 1996.

Kee, C., Wide Area Differential GPS, in *Global Positioning System: Theory and Applications, Vol II*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Klobuchar, J.A., P.H. Doherty and B. El-Arini, Potential ionospheric limitations to GPS wide-area augmentation system, *NAVIGATION: Journal of the Institute of Navigation*, **42**, 353, 1995.

Klobuchar, J.A., Ionospheric Effects, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Komjathy, A. and R. Langley, The effect of shell height on high precision ionosphere modelling using GPS, *IGS Workshop*, Silver Springs, Maryland, 1996a.

Komjathy, A. and R. Langley, An assessment of predicted and measured ionospheric total electron content using a regional GPS network, *Proceedings of the ION GPS-96 National Technical Meeting*, Santa Monica, California, January, 1996b.

Kovach, K.L and K.L. Van Dyke, GPS in ten years, *Proceedings of the ION GPS-97*, 1251, Kansas City, Missouri, September, 1997.

Kunches, J.M., Now it gets interesting: GPS and the onset of solar cycle 23, *Proceedings of the ION GPS-97*, 225, Kansas City, Missouri, 1997.

Lachapelle, G., "ENSU 625: GPS Theory and Applications", Fall term lecture notes, The University of Calgary, 1995.

Lassen, K., Relation of the plasma sheet to the nighttime auroral oval, *Journal of Geophysical Research*, **79**, 3857, 1974.

Leick, A., *GPS Satellite Surveying*, second edition, John Wiley and Sons, U.S.A., 1995.

Loddo, S., Flament, D., Benedicto, J. and P. Michel, EGNOS, the European regional augmentation to GPS and GLONASS, *Proceedings of the ION GPS-96*, 1143, Kansas City, Missouri, September, 1996.

Lomb, N.R., *Astrophysics and Space Science*, 39, 447, 1976.

Mannucci, A.J., B. Wilson and C. Edwards, A new method for monitoring the Earth's ionospheric total electron content using the GPS global network, *Proceedings of the ION GPS-93*, 1323, Salt Lake City, Utah, 1323, September, 1993.

Mannucci, A.J., B. Wilson and D. Yuan, An improved ionospheric correction method for wide-area augmentation systems, *Proceedings of the ION GPS-95*, 1199, Palm Springs, California, September, 1995.

Mannucci, A.J., B.A. Iijima and B.D. Wilson, Wide area ionospheric delay corrections under ionospheric storm conditions, *Proceedings of the ION GPS-97 National Technical Meeting,* Santa Monica, California, January, 1997.

Matsushita, S. and W.H. Campbell, *Physics of Geomagnetic Phenomena*, Vol. II, Academic Press, New York, 1967.

Mayaud, P.N., *Derivation, Meaning, and Use of Geomagnetic Indices*, AGU, Washington, D.C., 1980.

McPherron, R.L., Growth phase of magnetospheric substorms, *Journal of Geophysical Research*, **75**, 5592, 1970.

Mueller, T. and P. Davis, Real-time WADGPS ionospheric correction filter, *Proceedings of the ION GPS-98 National Technical Meeting*, Santa Monica, California, January, 1998.

Murphree, J.S., R.D. Elphinstone, L.L. Cogger and D. Hearn, Viking Optical Substorm Signatures, in *Magnetospheric Substorms*, Geophysical Monograph 64, ed. J.R. Kan, T.A. Potemra, S. Kokubun and T. Iijima, AGU, Washington, DC, 1991.

Nagai, T., Observed magnetic substorm signatures at synchronous altitude, *Journal of Geophysical Research*, **187**, 4405, 1982.

Nav Canada, *SatNav newsletter*, issue 2, June, 1998.

Nieto, J., M.A. Molina, M.M. Romay, J. Cosmen, M.L. de Mateo, R. Roman and L. Andrada, Assessment of EGNOS system and performance: early test system, *Proceedings of ION GPS-97*, 1345, Kansas City, Missouri, September, 1997.

Numerical Recipes in C, ed. W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, Cambridge University Press, Cambridge, 1997.

Parkinson, B.W., Stansell, T., Beard, R. and K. Gromov, A history of satellite navigation, *NAVIGATION: Journal of the Institute of Navigation*, **42**, no. 1, Special Issue, 1995.

Parkinson, B.W., GPS Error Analysis, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Parks, G.K., *Physics of Space Plasmas, An Introduction*, Addison Wesley, Redwood City, CA, 1991.

Rishbeth, H. and O.K. Garriott, *Introduction to Ionospheric Physics*, Academic Press, New York, 1969.

Rostoker, G., S.-I. Akasofu, J. Foster, R.A. Greenwald, Y. Kamide, K. Kawasaki, A.T.Y. Lui, R.L. McPherron and C.T. Russell, Magnetospheric substorms - definition and signatures, *Journal of Geophysical Research*, **85**, 1663, 1980.

Rostoker, G., S.-I. Akasofu, W. Baumjohann, Y. Kamide and R.L. McPherron, The roles of direct input of energy from the solar wind and unloading of stored magnetotail energy in driving magnetospheric substorms, *Space Science Review*, **46**, 93, 1987.

Rostoker, G., Some observational constraints for substorm models, in *Magnetospheric Substorms*, Geophysical Monograph 64, ed. J.R. Kan, T.A. Potemra, S. Kokubun and T. Iijima, AGU, Washington, DC, 1991.

Rostoker, G. and S. Skone, Magnetic flux mapping considerations in the auroral oval and the Earth's magnetotail, *Journal of Geophysical Research*, **98**, 1377, 1993.

RTCA, "Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment", RTCA/DO-229A, SC-159, Washington, DC, June 8, 1998.

Skone, S., M.E. Cannon, K. Lochhead, P. Héroux and F. Lahaye, Performance evaluation of the NRCan wide area system, *Canadian Aeronautics and Space Journal*, **43**, no. 2, 92, 1997.

Skone, S. and M.E. Cannon, Auroral zone considerations for WADGPS, *NAVIGATION: Journal of the Institute of Navigation*, **45**, no.2, 117, 1998.

Spilker, J.J., Satellite Constellation and Geometric Dilution of Precision, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996a.

Spilker, J.J., GPS Signal Structure and Theoretical Performance, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996b.

Van Dierendonck, A.J., GPS Receivers, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Van Graas, F. and M.S. Braasch, Selective Availability, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

Wells, D. et al., *Guide to GPS Positioning*, Canadian GPS Associates, Fredericton, New Brunswick, 1987.

Wilson, B.D. and A.J. Mannucci, Instrumental biases in ionospheric measurements derived from GPS data, *Proceedings of the ION GPS-93*, 1343, Salt Lake City, Utah, 1343, September, 1993.

Wilson, B.D., A.J. Mannucci, D.N. Yuan and B. Iijima, Monitoring ionospheric electron content using a global GPS network, *Proceedings of the ION GPS-97 National Technical Meeting,* Santa Monica, California, January, 1997.

Wu, B, P. Nicolaides and T.N. Upadhyay, Ionospheric error compensation for GPS receivers using real-time ionospheric model, *Proceedings of the ION GPS-96*, 575, Kansas City, Missouri, September, 1996.

Yunck, T., Y. Bar-Sever, W. Bertiger, S. Lichten, U. Lindqwister, A. Mannucci, R. Muellerschoen, T. Munson, L. Romans and S. Wu, A prototype WADGPS system for real-time sub-meter positioning worldwide, *Proceedings of the ION GPS-96*, 1819, Kansas City, Missouri, September, 1996.

Zmuda, A.J. and J.C. Armstrong, The diurnal flow pattern of field-aligned currents, *Journal of Geophysical Research*, **79**, 4611, 1974.

Zumberge, J.F. and W.I. Bertiger, Ephemeris and Clock Navigation Message Accuracy, in *Global Positioning System: Theory and Applications, Vol. I*, ed. B.W. Parkinson and J.J. Spilker, American Institute of Aeronautics and Astronautics, Washington, D.C., 1996.

APPENDIX A

THE DISCRETE KALMAN FILTER

The Kalman filter is a recursive linear filter, in which only past estimates from the previous epoch are stored. This approach employs two models: a measurement model describing functional relationships between observations and unknowns, and a system model describing temporal variations of the unknowns. At a given epoch, estimates of the unknown parameters are derived by combining both *predicted* values of the unknowns (as derived from the system model) and information from observations at the present epoch (as derived from the measurement model). The optimal combination is determined by using an *optimal linear estimator*.

A.1 MEASUREMENT MODEL

The parametric measurement model at time t_j is expressed as follows:

$$z(t_j) = H(t_j)x(t_j) + v(t_j) \tag{A.1}$$

where

z = observation matrix (m×1)

H = "design matrix" (m×n)

x = matrix of unknown parameters (state vector) (n×1)

v = random measurement error (m×1)

where n = number of unknowns and m = number of observations.

It is assumed that the model H describes the functional relationship between z and x completely, such that the vector v represents only random zero-mean measurement errors:

$$E[v]=0$$
$$E[vv^T]=R \qquad (A.2)$$

where R is the measurement covariance matrix (m×m).

A.2 SYSTEM MODEL

The system model describes variations in the unknown parameters over time:

$$x^-(t_{j+1}) = \Phi(t_j, t_{j+1})\, x^+(t_j) + w \qquad (A.3)$$

where

$\Phi(t_j, t_{j+1})$ = "transition matrix" (n×n)

w = random errors in the prediction model (n×1)

and

"+" and "-" superscripts denote predicted and updated estimates, respectively

The transition matrix Φ relates the estimate of x at time t_j (i.e. $x^+(t_j)$) to that predicted for time t_{j+1} (i.e. $x^-(t_{j+1})$). It is assumed that Φ describes the functional relationship between x^- and x^+ completely, such that w is a zero-mean white sequence:

$$E[w]=0$$
$$E[ww^T]=Q \qquad (A.4)$$

where Q is the process covariance (n×n), or *process noise*.

A.3 DERIVATION OF OPTIMAL LINEAR ESTIMATOR

In order to calculate an optimal estimate of the parameters x at time t_{j+1}, the predicted values x^- (as derived from observations at previous epochs) must be combined with the new measurements at time t_{j+1}: i.e.

$$x^+(t_{j+1}) = K_1 x^-(t_{j+1}) + K_2 z(t_{j+1}) \tag{A.5}$$

where K_1, K_2 are essentially linear weighting factors. Equation A.5 can be rewritten in terms of one weighting matrix K at time t_{j+1} (where the time dependence is dropped for convenience):

$$x^+ = (I-KH)x^- + Kz \tag{A.6}$$

In the derivation of Equation A.6, it is assumed that any errors in the estimates x^- and x^+ are zero-mean: i.e.

$$E[\delta x^-] = E[\delta x^+] = 0 \tag{A.7}$$

where δx^- and δx^+ denote errors in the estimates x^- and x^+, respectively. The updated estmation errors can be expressed (from Equation A.6) as

$$\delta x^+ = (I-KH)\delta x^- + Kv \tag{A.8}$$

Under these conditions, the estimator is *unbiased.* Associated error covariances are derived as follows:

$$P^- = E[\delta x^- \delta x^{-T}] \tag{A.9}$$

and, using the covariance law,

$$P^{+} =$$

$$\left(\frac{\partial x^{+}}{\partial x^{-}}\right) E\left[\delta x^{-} \delta x^{-T}\right] \left(\frac{\partial x^{+}}{\partial x^{-}}\right)^{T} + \left(\frac{\partial x^{+}}{\partial v}\right) E\left[v v^{T}\right] \left(\frac{\partial x^{+}}{\partial v}\right)^{T} + \left(\frac{\partial x^{+}}{\partial x^{-}}\right) E\left[\delta x^{-} v^{T}\right] \left(\frac{\partial x^{+}}{\partial v}\right)^{T} + \left(\frac{\partial x^{+}}{\partial v}\right) E\left[v \delta x^{-T}\right] \left(\frac{\partial x^{+}}{\partial x^{-}}\right)^{T} \tag{A.10}$$

such that the updated covariance at time t_{j+1} is calculated as

$$P^{+} = (I-KH)P^{-}(I-KH)^{T} + KRK^{T} \tag{A.11}$$

where $E[\delta x^{-} v^{T}] = E[v \delta x^{-T}] = 0$

The Kalman filter is characterized by the form of the matrix K, which is referred to as the *Kalman gain matrix*. An optimal choice of K can be derived as that for which the weighted scalar sum of trace[P^{+}] is minimized. For this derivation, K is optimal in the *minimum variance* sense, where the sum of squared errors in estimates of x^{+} are minimized. Under this condition, a solution for K is derived as follows:

$$K = P^{-}(t_{j+1})\, H^{T}(t_{j+1})\, [H(t_{j+1})\, P^{-}(t_{j+1})\, H^{T}(t_{j+1}) + R(t_{j+1})]^{-1} \tag{A.12}$$

For such values of K, the expression for P^{+} is simplified to

$$P^{+}(t_{j+1}) = [I - KH(t_{j+1})]P^{-}(t_{j+1}) \tag{A.13}$$

A.4 KALMAN FILTER EQUATIONS

Given this overview of the Kalman filter derivations, the complete Kalman filtering recursive equations are given as follows, where the algorithm has been divided into two steps: prediction and update.

1) Prediction (from time t_j to t_{j+1})

$$x^-(t_{j+1}) = \Phi(t_j, t_{j+1})\, x^+(t_j) + w \tag{A.14}$$

$$P^-(t_{j+1}) = \Phi(t_j, t_{j+1})\, P^+(t_j)\, \Phi(t_j, t_{j+1})^T + Q(t_j) \tag{A.15}$$

2) Update (at time t_{j+1})

$$x^+(t_{j+1}) = x^-(t_{j+1}) + K[z(t_{j+1}) - H(t_{j+1})\, x^-(t_{j+1})] \tag{A.16}$$

$$P^+(t_{j+1}) = [I - KH(t_{j+1})]P^-(t_{j+1}) \tag{A.17}$$

Input matrices Φ, H, R and Q, are required for every prediction and update. Initial *a priori* estimates of x and P must also be supplied at t_0. As the filter approaches steady state, in the absence of any adaptive updates of filter covariances, P^+ will converge to a constant value. The magnitude of this value may be estimated through *a priori* sensitivity analysis, since P^+ does not depend on the observation input.

The *measurement residuals* are essentially the corrections to the observations required to make them consistent with filter estimates of unknowns. They can be calculated as

$$r(t_{j+1}) = H(t_{j+1})x^+(t_{j+1}) - z(t_{j+1}) \tag{A.18}$$

A.5 OPTIMALITY OF ESTIMATOR

As outlined in Section A.3, K is derived for the optimal linear estimator in a minimum variance sense, using statistical models for x and z. Other "optimal" estimators may be derived from different conditions, such as deterministic arguments (i.e. weighted least squares) or the derivation of a maximum likelihood estimator from statistical properties of z. The estimators derived from these various conditions are identical, however, if x (where x is treated as a quasi-observable) and z are assumed to be gaussian random variables. In such cases, the form of K in Equation A.12 is the unique optimal linear estimator.

Additionally, Equation A.6 assumes a linear form for the estimator K. Nonlinear estimators also exist, and the optimal estimator takes a linear form only under certain conditions. If the time-varying process modelled in Equation A.12 is a random gaussian process, such that w are normally distributed, the optimal estimator is a linear operator. Given these criteria, the Kalman gain matrix in Section A.3 is *the* optimal estimator only if z, x and w are random gaussian variables. In practical applications, w is generally derived from an empirical autocorrelation function, for which an equivalent gaussian process may be derived, and the measurement errors v are assumed to approximate a random gaussian distribution.

APPENDIX B

IONOSPHERE PIERCE POINT CALCULATION

Given the coordinates (latitude, longitude, height) of a reference station receiver, and the ephemeris for all visible satellites, the points at which satellite-receiver lines-of-sight intersect the ionosphere shell may be calculated. The calculation is performed as follows, using geometry arguments derived from Figure B.1.

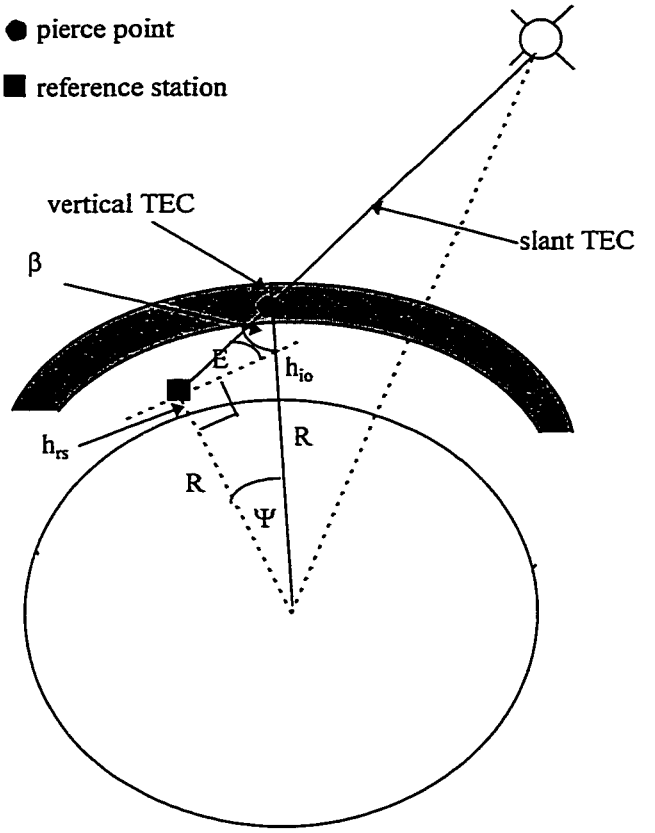


Figure B.1 Geometry of pierce point calculation.

The pierce point is offset from the reference station by an angle Ψ, as defined from the centre of a spherical Earth with mean radius 6371.0 km. This angle may be calculated from the relation

$$\beta + (E+\pi/2) + \Psi = \pi \tag{B.1}$$

such that

$$\Psi = \pi - (E+\pi/2) - \beta = \pi/2 - E - \beta \tag{B.2}$$

where E = satellite elevation angle at the reference station. Additionally, the angle β may be calculated from the sine rule:

$$\frac{\sin(E+\pi/2)}{R+h_{io}} = \frac{\sin\beta}{R+h_{rs}} \tag{B.3}$$

and

$$\sin\beta = \frac{R+h_{rs}}{R+h_{io}}\sin(E+\pi/2) \tag{B.4}$$

where R = mean Earth radius, h_{rs} = height of reference station, and h_{io} = altitude of ionosphere shell. All quantities on the right hand side of Equation B.4 are known, and the value of β can be computed. Substituting Equation B.4 in Equation B.2 gives

$$\Psi = \pi/2 - E - \sin^{-1}\left(\frac{R+h_{rs}}{R+h_{io}}\sin(E+\pi/2)\right) \tag{B.5}$$

where all angles are defined in radians. The latitude and longitude of the pierce point are then derived from known coordinates of the reference station and the offset angle Ψ:

$$\phi_{pp} = \phi_{rs} + (\Psi \cos(azim)) \tag{B.6}$$

$$\lambda_{pp} = \lambda_{rs} + (\Psi \sin(azim))/\cos(\phi_{pp}) \tag{B.7}$$

where azim is the azimuth angle of the satellite, and ϕ_{pp}, λ_{pp} and ϕ_{rs}, λ_{rs} are the latitude and longitude of the pierce point and reference station, respectively.

In order to map vertical TEC to slant TEC in the ionosphere shell, the geometry of Figure B.1 is considered, where

$$\cos\beta = \text{vertical/slant} \tag{B.8}$$

and

$$\text{slant} = \text{vertical}/\cos\beta = \text{vertical} \times (1-\sin^2\beta)^{-1/2} \tag{B.9}$$

where the expression for β is given in Equation B.4. The factor $(1-\sin^2\beta)^{-1/2}$ can be rewritten as mapping factor M(E):

$$M(E) = (1-\sin^2\beta)^{-1/2} = \{1-[\cos(E)/(1+h/R_E)]^2\}^{-1/2} \tag{B.10}$$

APPENDIX C

SPECTRAL ANALYSIS METHODS

C.1 FOURIER TRANSFORM - EVENLY SAMPLED DATA

C.1.1 Continuous Fourier Transform

For a continuous time series h(t), the Fourier Transform is calculated as

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{2\pi ift}dt \qquad (C.1)$$

The *total* power in signal h(t) can be calculated in either the time domain or the frequency domain. Both calculations give the same result. This is the basis of Parseval's theorem, which can be written as follows

$$\text{Total power} = \int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} h(t)h^*(t)dt = \int_{-\infty}^{\infty} H(f)H^*(f)df = \int_{-\infty}^{\infty} |H(f)|^2 df \qquad (C.2)$$

where * denotes the complex conjugate.

C.1.2 Discrete Fourier Transform

For discrete time series, sampled at evenly spaced intervals Δ in the time domain, it is possible to derive a representative power spectrum in the frequency domain via the discrete Fourier Transform:

$$H_n = \sum_{k=0}^{N-1} h_k e^{2\pi ikn/N} \qquad \text{(C.3)}$$

where H_n is periodic with period N, such that $H_n=H_{N-n}$.

The discrete Fourier Transform can then be considered as a mapping of N sampled values h_k, at times t_k (multiples of Δ), into the N+1 frequencies

$$f_n = \frac{n}{N\Delta} \qquad n=-N/2,...,N/2 \qquad \text{(C.4)}$$

The convention n=-N/2,...,N/2 is most commonly used, such that the Fourier Transform is defined in a frequency range symmetric about 0. In this case, the upper and lower frequencies are defined by the sampling interval Δ, which limits the highest resolvable frequency, the Nyquist critical frequency, $f_{N/2} = f_c = 1/2\Delta$. The smallest non-zero frequency which can be resolved is $f_1 = 1/N\Delta$. The Fourier Transforms corresponding to n=-N/2 and n=N/2 are identical, such that N independent values of H_n are derived from the N h_k samples of input.

Aliasing of power into the discrete frequency range will occur if the Nyquist frequency is too small. Such effects are considered in the choice of sampling interval and number of samples. The discrete Fourier Transform can be calculated using the Fast

Fourier Transform algorithm [*Cooley and Tukey*, 1965]. Such an algorithm is described in *Brigham* [1974]. This algorithm is implemented here by using a modified version of the "realft" computer program from *Numerical Recipes* [1997].

The power spectral density for a given harmonic component of frequency f_n is derived as

$$\text{power}(f_n) = \frac{1}{N}|H_n|^2 \qquad \text{(C.5)}$$

where the factor 1/N is a normalization constant (such that the power spectral density is defined per unit time, length, or other unit of h_k). Several different conventions are available to represent the power spectrum of a given function. For real-valued series h_k, the two-sided power spectral density is symmetric, and it is convenient to represent the power spectrum as a *one-sided* power spectral density distribution, which is defined at N/2+1 positive frequencies f_n as follows

$$P(f_0) = \frac{1}{N}|H_0|^2 \qquad n=0$$

$$P(f_n) = \frac{2}{N}|H_n|^2 \qquad n = 1,2,...,(N/2-1)$$

$$P(f_{N/2}) = \frac{1}{N}|H_{N/2}|^2 \qquad n=N/2 \qquad \text{(C.6)}$$

where

$$f_n = \frac{n}{N\Delta} \qquad n=0,1,...,N/2$$

Note that the values H_0 and $H_{N/2}$ are real. The factor 2 for n=1,2,...,(N/2-1) is necessary, in order to include contributions from corresponding negative frequencies (i.e. n=-1, -2,....,-(N/2-1)).

The discrete form of Parseval's Theorem is written as follows:

$$\sum_{k=0}^{N-1}\left|h_k\right|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\left|H_n\right|^2 = \frac{1}{N}\sum_{n=-(N/2-1)}^{N/2}\left|H_n\right|^2 \qquad \text{(C.7)}$$

If both sides of equation C.7 are multiplied by 1/N, the resulting equation can be used to calculate the mean squared value (or mean squared amplitude) of the time series h_k.

$$\frac{1}{N}\sum_{k=0}^{N-1}\left|h_k\right|^2 = \frac{1}{N^2}\sum_{n=-(N/2-1)}^{N/2}\left|H_n\right|^2 = \frac{1}{N^2}\left|H_o\right|^2 + \frac{2}{N^2}\sum_{n=1}^{N/2-1}\left|H_n\right|^2 + \frac{1}{N^2}\left|H_{N/2}\right|^2 \qquad \text{(C.8)}$$

A useful extension of equation C.8 involves calculating the total power in a selected subset of the discrete frequency range. In this case, the power spectral density summed from f_a to f_b, and multiplied by 1/N, gives the mean squared value of the time series consisting of only those harmonic components lying in the range f_a to f_b: i.e.

$$\frac{1}{N}\sum_{k=0}^{N-1}\left|c_k\right|^2 = \frac{2}{N^2}\sum_{n=a}^{b}\left|H_n\right|^2 \qquad \text{(C.9)}$$

where $f_0 < f_a$, $f_b < f_{N/2}$ and the c_k's correspond to harmonic components for frequencies ranging from f_a to f_b. This equation is useful for isolating and estimating relative amplitude contributions from various subsets of the overall frequency range.

C.2 LOMB METHOD - UNEVENLY SAMPLED DATA

The Fourier Transform, as developed in the previous sections, is a useful tool for mapping discrete, evenly sampled, data series into the frequency domain. In some cases, however, only unevenly sampled values are available. In these cases an expanded, evenly sampled, data set can be generated by interpolating between observed values. The expanded data set can then be transformed using the Fourier Transform. This technique is adequate for data series in which spurious data dropouts occur, such that the interpolated samples form a small fraction of the entire data set. If the interpolated data form a significant fraction of the expanded data, however, the computed power spectrum may not adequately reflect "observed" signal energy, since the Fourier Transform technique weights all data, real and interpolated, equally.

In this case, a superior method exists for calculating power spectral densities; this method is herein referred to as the Lomb method [*Lomb*, 1976]. In this approach, the power spectrum is derived in a least squares adjustment, where amplitudes corresponding to harmonic components of the series h_k are estimated from only observed values. In this method, no interpolation of data is required. The power spectral densities are derived as follows:

$$P(f_n) = \left\{ \frac{\left[\sum_{i=1}^{N} (h_i - E[h]) \cos 2\pi f(t_i - \tau)\right]^2}{\sum_{i=1}^{N} \cos^2(2\pi f(t_i - \tau))} + \frac{\left[\sum_{i=1}^{N} (h_i - E[h]) \sin 2\pi f(t_i - \tau)\right]^2}{\sum_{i=1}^{N} \sin^2(2\pi f(t_i - \tau))} \right\} \qquad (C.10)$$

where

$$E[h] = \frac{1}{N}\sum_{i=1}^{N} h_i \qquad \text{and} \qquad \tau = \frac{1}{4\pi f} a\tan\left(\frac{\sum_{i=1}^{N}\sin 2\pi f t_i}{\sum_{i=1}^{N}\cos 2\pi f t_i}\right)$$

Equation C.10 is equivalent to the power derived from the f_n harmonic component of h(t): i.e. the power in the series $h(t) = c_1\cos(2\pi ft) + c_2\sin(2\pi ft)$, where c_1 and c_2 are the "best fit" coefficients to the series h_k.

A Nyquist frequency of $1/2\Delta$ may be defined, where Δ is the average sampling interval over the N sampled data points. In this case, however, frequencies are resolved in a range exceeding the Nyquist frequency (since some sampling intervals will be less than Δ).

C.3 AUTOCORRELATION

Stochastic processes are often characterized by their autocorrelation functions. Such functions are essentially a measure of how well a given function is correlated with a second version of itself, superposed and shifted by a constant amount. These functions are not restricted to temporal processes, but can also be applied to any one-dimensional real sequence of values.

The autocorrelation of a continuous process h(t) can be derived as

$$\phi(\tau) = \int_{-\infty}^{\infty} h(t+\tau)h(t)dt \qquad \text{(C.11)}$$

where τ denotes the offset between series. By taking the Fourier Transform of Equation C.11 and applying the "Correlation Theorem" [*Brigham*, 1974], the integral in Equation C.11 can be conveniently rewritten as

$$\int_{-\infty}^{\infty} h(t+\tau)h(t)dt \Diamond H(f)H^*(f) = |H(f)|^2 \qquad (C.12)$$

where H(f) denotes the continuous Fourier Transform of h(t), H*(f) the complex conjugate, and $\Diamond$ symbolizes inversion.

In general, if it often desirable to derive empirical autocorrelation functions for a series of evenly sampled data h_k, in order to estimate stochastic properties of the associated process. In the discrete case, Equation C.11 becomes

$$\phi_j = \phi(\Delta j) = \frac{1}{N-j}\sum_{k=1}^{N-j} h_k h_{k+j} \qquad (C.13)$$

after *Gelb* [1974], where Δ is the sampling interval and Δj denotes the offset between series. Note that the larger the offset, the fewer terms in the summation. This is to be considered when estimating an adequate sample size. In practice, ϕ_j are generally computed for only the values j=0,...,N/2. For a shift of $\Delta j = 0$, Equation C.13 becomes

$$\phi_0 = \phi(0) = \frac{1}{N}\sum_{k=1}^{N} h_k h_k = E[h^2] = \sigma^2 + E[h]^2 \qquad (C.14)$$

where $E[\cdot]$ denotes the expectation value, and σ is the standard deviation. If Equation C.14 is normalized, i.e.

$$\phi(\Delta j) = \frac{\frac{1}{N-j}\sum_{k=1}^{N-j} h_k h_{k+j} - \frac{1}{(N-j)^2}\sum_{k=1}^{N-j} h_k \sum_{k=1}^{N-j} h_{k+j}}{\left(\frac{1}{(N-j)}\sum_{k=1}^{N-j} h_k^2 - \left(\frac{1}{(N-j)}\sum_{k=1}^{N-j} h_k\right)^2\right)\left(\frac{1}{(N-j)}\sum_{k=1}^{N-j} h_{k+j}^2 - \left(\frac{1}{(N-j)}\sum_{k=1}^{N-j} h_{k+j}\right)^2\right)} \quad \text{(C.15)}$$

an *autocorrelation coefficient* may be derived, ranging from -1 to 1.

It is often useful to approximate the continuous function $\phi(t)$ from the discrete values ϕ_j. Since the series h_k is not defined by a continuous function, a mathematical expression for $\phi(\tau)$ cannot be derived from Equation C.13 directly. In order to derive an estimate of the continuous function $\phi(\tau)$, the summation in Equation C.13 must be computed for each of the N/2+1 possible time shifts Δj and a functional form of $\phi(\tau)$ "best fit" to the computed ϕ_j. A more elegant way of deriving this function is to use a discrete form of the "Correlation Theorem" [*Brigham*, 1974], such that

$$\frac{1}{N-j}\sum_{k=1}^{N-j} h_k h_{k+j} \;\Diamond\; H_n H^*_n = |H_n|^2 \quad \text{(C.16)}$$

In this case, the discrete Fourier Transform H_n is computed for the series h_k, and multiplied by the complex conjugate of itself. The inverse Fourier Transform of the product $H_n H^*_n$ is then a vector, the components of which correspond to autocorrelation values ϕ_j for each time lag Δj.

$$\phi(\Delta j) = \frac{1}{N}\sum_{n=0}^{N-1} |H_n|^2 \, e^{-2\pi i j n/N} \quad \text{(C.17)}$$

While Equations C.13 and C.17 can both be used to calculate the autocorrelation functions of a given sequence h_k, these two equations, as written here, are not equivalent.

Discrepancies result from the periodicity of the Fourier Transform, where the series h_k is implicitly assumed to be periodic with period N. If Equation C.13 were rewritten with the summation from 1 to N-1, and the series h_k extended, with period N, the two equations would be equivalent. For practical purposes, the autocorrelation function is generally computed only for time lags in the range $-N/2 < j < N/2$, such that aliasing effects at higher time lags are minimized.

Empirically derived autocorrelation functions are useful statistical tools. Information concerning the periodicity of a given process (data series) may be derived, in addition to the first and second moments of the corresponding probability distribution functions. Several stochastic processes, such as white noise, coloured noise, Gauss-Markov processes and the random walk, are commonly defined by their autocorrelation functions and corresponding spectral densities.

APPENDIX D

CONVERSION OF GEOGRAPHIC TO GEOMAGNETIC COORDINATES

Given the geographic latitude and longitude of a point on the ionosphere shell, corresponding values of geomagnetic latitude (Φ) and longitude (Λ) may be derived from the following equations (after *Matsushita and Campbell* [1967]):

$$\sin\Phi = \sin\phi\cos\phi_{dip} + \cos\phi\sin\phi_{dip}\cos(\lambda-\lambda_{dip}) \quad \text{(D.1)}$$

$$\sin\Lambda = \cos\phi\sin(\lambda-\lambda_{dip})/\cos\Phi \quad \text{(D.2)}$$

where ϕ and λ denote the geographic latitude and longitude, respectively, of a given point on the ionosphere shell. The geographic latitude and longitude of the north geomagnetic pole are denoted by ϕ_{dip} and λ_{dip}. Expressions for these values are given by *Hapgood* [1992] as follows:

$$\phi_{dip} = 78.8 + 4.283\times10^{-2}[(MJD-46066)/365.25] \quad \text{(D.3)}$$

$$\lambda_{dip} = 289.1 - 1.413\times10^{-2}[(MJD-46066)/365.25] \quad \text{(D.4)}$$

where MJD is the modified Julian date, which is the time (in days) from 0000 UT on 17 November 1858.