Department of Geomatics Engineering

Evaluation of 2-Dimensional Ionosphere Models for National And Regional GPS Networks in Canada

(URL: http://www.geomatics.ucalgary.ca/links/GradTheses.html)

by

Yongjin Moon

April 2004

UNIVERSITY OF
CALGARY

UNIVERSITY OF CALGARY

Evaluation of 2-Dimensional Ionosphere Models for National And Regional

GPS Networks in Canada

by

Yongjin Moon

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

APRIL, 2004

ABSTRACT

The purpose of this research is to determine a 2-dimensional ionosphere model, applicable in real-time, that provides optimal accuracy of vertical delays at ionospheric grid points (IGPs). This model is based on GPS delays measured at ionospheric pierce points (IPPs), as observed from dual-frequency GPS tracking stations. Two algorithms are selected for possible implementation: spherical harmonic model and thin plate spline interpolation. These methods are based on two-dimensional estimation on an ionospheric shell at 350 km altitude. The input observations are computed as slant delays using dual frequency GPS observations.

In the spherical harmonics model, the coefficients and receiver differential code biases are estimated every 5 minutes in a real-time mode using a Kalman filter. Thin-plate spline is a 2-dimensional generalization of the cubic spline in 1 dimension. The basic idea of this method is to build a function that passes through the grid points and minimizes the roughness of the surface.

The performance of the two algorithms is evaluated in terms of accuracy of the residuals between observed vertical TECs (VTECs) and estimated VTECs at IGPs in two GPS networks: Canadian Active Control System (CACS) and Western Canada Deformation Array (WCDA). The evaluation results show that the thin plate spline outperforms the spherical harmonics model in both networks. The spatial and temporal variation of the geomagnetic storm is also investigated by plotting vertical TEC maps over Canada. In considering the VTEC maps generated for the storm period, it is observed that regions with large enhancements of vertical TEC coincide well with the locations of stations with larger rms values.

ACKNOWLEGEMENT

I would like to thank my supervisor, Dr. Susan Skone, for her valuable advice and support through my graduate studies. Dr Gao and Dr Maundy are also thanked for providing insightful comments and advice on my thesis as members of the examination committee.

My thanks also go to my colleagues: Sudhir Shresta, Victoria Hoyle, Natalya Nicholson, Ruben Yousuf.

I would like to express my gratitude to Dave Clark, my neighbour and good friend. My family's great friends Sharon and Stephen are also appreciated for their kindness and generosity.

Finally, I would like to express my deepest thanks to my family: my wife (Pil Sook), son (Junghoon) and parents in Korea for their unconditional support and understanding through my study in Canada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CACS	Canadian Active Control System
CDGPS	Canada-wide Differential Global Positioning Sytem
CONUS	Continental United States
DGPS	Differential Global Positioning System
EGNOS	European Geostationary Navigation Overlay System
GPS	Global Positioning System
IGP	Ionosphere Grid Point
IGS	International GPS Service
IPP	Ionosphere Pierce Point
MCS	Master Control Station
MET	Multipath Estimation Technology
MMT	Multipath Mitigation Technology
MSAS	Multifunction Transportation Satellite-based Augmentation System
MSV	Mobile Satellite Ventures
SA	Selective Availability
SCHA	Spherical Cap Harmonic Analysis
SED	Storm Enhanced Density
NRCan	Natural Resources Canada
TEC	Total Electron Content
TECU	Total Electron Content Unit
UT	Universal Time
UTC	Universal Time Coordinated
WADGPS	Wide Area DGPS
WAAS	Wide Area Augmentation System
WCDA	Western Canada Deformation Array
VTEC	Vertical Total Electron Content

LIST OF SYMBOLS

P	measured code pseudorange
r	true or geometric range
c	speed of light
δt_u	receiver clock bias
δt^s	satellite clock bias
I_P	code ionospheric delay
T_P	code tropospheric delay
ε_P	code measurement noise
$T_{GD,r}, T_{GD,s}$	receiver and satellite modulation offsets in units of second
A, ϕ	amplitude and phase of the direct signal
$\Delta\phi$	phase shift of the reflected signal
γ	amplitude damping coefficient due to the reflection
Φ	phase range
ϕ	phase measurement
N	integer ambiguity (cycle)
I_Φ	phase ionospheric delay
T_Φ	phase tropospheric delay
ε_Φ	phase measurement noise
λ_{wl}	wide-lane wave length
N_{wl}	wide-lane ambiguity
v_p	phase velocity
v_g	group velocity
n_p	phase refractive index
n_g	group refractive index
N_e	electron density
$\rho*$	non-dispersive delay, which includes geometric range, tropospheric delay and clock offset
$\varepsilon_{\phi_{wl}}$	wide-lane measurement noise in units of cycle
λ_{nl}	narrow-lane wavelength
N_{nl}	the narrow-lane ambiguity
$\varepsilon_{\phi_{nl}}$	narrow-lane measurement noise in units of cycle
$\langle\cdot\rangle$	mean value
F	threshold factor for cycle slip detection
R_E	earth radius
E	satellite elevation angle

H	altitude of the ionospheric shell
$M(E)$	vertical to slant mapping function
b_r, b_s	receiver and satellite code interchannel biases
NSAT	number of satellites observed in one epoch
φ	geomagnetic latitude of an IPP
ζ	geomagnetic longitude of an IPP
n, m	integer degree and order of Legendre function, respectively
C_{nm}, S_{nm}	unknown spherical harmonic coefficients
P_{nm}	normalized associated Legendre functions
$\dfrac{1}{\beta}$	correlation time
w	white noise

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND

The Global Positioning System (GPS) is a satellite navigation system that provides 3-dimensional positions, velocity and time information. GPS consists of 24 satellites allocated in 6 orbital planes to ensure that at least 4 satellites are visible anywhere on earth. Due to the fact that GPS is a space-based radio navigation system, GPS is subject to several error sources: satellite ephemeris and clock error, atmospheric (troposphere and ionosphere) propagation error, signal multipath error and receiver noise. The errors due to satellite position and clock bias, and atmospheric propagation, are correlated for GPS receivers observing the same satellites. Differential GPS (DGPS) takes advantage of this spatial correlation of those errors listed above to achieve improved accuracy.

In DGPS, a GPS receiver located at a known position (reference station) estimates the range errors and transmits the error estimates to the GPS users in the coverage area. The GPS users apply the error estimates generated at the reference station to their measurements and improve the accuracy of their position estimates. Since the correlated error sources decorrelate with the distance between the reference station and the users, many reference stations would be needed to cover a wide area. Wide Area DGPS (WADGPS) (Kee, 1996) can provide differential corrections to GPS users over a wide area with a minimum number of reference stations.

In WADGPS, a network of reference stations is deployed over a wide area and the raw data from the reference stations are processed at master stations to estimate the satellite ephemeris and clock error, and ionospheric delay separately in a vector mode. The separate error components are broadcast to the users through geostationary satellites or a network of FM stations (Misra and Enge, 2001). Several WADGPS services are

operational or being deployed in United States, Europe and Japan: Wide Area Augmentation (WAAS), European Geostationary Navigation Overlay System (EGNOS) and Multifunction Transportation Satellite-based Augmentation System (MSAS). Recently, Canada has also developed a WADGPS service, referred to as Canada-wide Differential GPS (CDGPS). CDGPS provides WADGPS corrections, referred to as GPS*C, through the MSAT-1 communication satellite operated by Mobile Satellite Ventures (MSV)(Kassam *et al.*, 2002). The GPS*C corrections are generated by Natural Resources Canada (NRCan) using the data from the Canadian Active Control System (CACS) network.

As mentioned above, one of the main tasks of WADGPS services is to model the spatially correlated errors. With the Selective Availability (SA) off, the ionosphere is the largest error source. Due to the fact that the ionosphere is a dispersive medium, the absolute ionospheric delay and total electron content (TEC) along the signal path from a GPS satellite to a user receiver can be derived from dual frequency GPS measurements. The data from a set of reference stations, each equipped with a dual frequency GPS receiver, are used to estimate vertical ionospheric delays at predefined IGPs on an ionosphere shell.

For single frequency GPS users, information describing vertical ionospheric delays at specified IGPs is very important for accurate positioning. The quality of vertical ionosphere delays at IGPs is dependent on the spatial and temporal distribution of the observed slant delays at ionospheric pierce points (IPPs). The spatial distribution of IPPs in Canada is limited – due to the sparse network of reference stations. Hence reliable interpolation algorithms for modeling the ionosphere are required. The purpose of this thesis is to evaluate several interpolation methods and find the optimal algorithms for different GPS networks in Canada.

The ionospheric models can be divided into two groups: grid-based interpolation methods and mathematical fitting methods. In a grid-based interpolation algorithm, grid points of an appropriate spacing are designated throughout the region of interest and vertical TEC

(VTEC) at the IGPs are interpolated from values at IPPs. For function-based approaches, coefficients describing the spatial distribution of VTEC are derived, and VTEC at designated IGPs are extracted from the underlying function. In this research, two algorithms are selected for evaluation and implementation: thin-plate spline and spherical harmonics model.

El-Arini *et al.* (1995) conducted comparisons of a number of these techniques, for evaluation of potential implementation in real-time wide area augmentation systems. Results indicated similar accuracies for all methods in the Continental United States (CONUS) region. Such methods were not evaluated specifically for regions of sparse observations, however, and differences exist in operational implementation of various models. Such methods may also provide different accuracies under various ionosphere conditions. It is necessary to evaluate different model approaches in Canada for optimal accuracies.

1.2 OBJECTIVES

The first objective of this thesis is to evaluate the two ionosphere model algorithms mentioned above in two different GPS networks: Canadian Active Control System (CACS) and Western Canada Deformation Array (WCDA). The CACS is a sparse network that covers the whole of Canada and the WCDA is a medium-size regional GPS network that is distributed in the western part of British Columbia. The purpose of selecting those two networks is to evaluate the performance of the two algorithms in networks of different dimensions-for the analysis of both wide area national and regional models. The algorithms are also evaluated for a range of ionospheric activity.

The second objective is to investigate the spatial and temporal variation of vertical TEC over Canada during a geomagnetic storm and asses its impact on the performance of the algorithm. The spatial variation of vertical TEC is represented with vertical TEC maps for the whole of Canada. For investigation of the temporal variation of vertical TEC, three epochs are selected from the geomagnetic storm period, and time snapshots of

vertical TEC maps at those epochs are plotted. The performance of the algorithm will be studied with the temporal and spatial variation of the vertical TEC over Canada.

1.3 THESIS OUTLINE

Chapter 2 describes the basic theory of GPS positioning. The principle of GPS measurement is given and the GPS observation error sources are detailed. The combination of GPS observables is also provided.

Chapter 3 describes the characteristics of GPS signal delay in the ionosphere. The principle of deriving ionospheric delays and related data processing techniques (carrier phase smoothing, cycle slip detection and correction, and multipath estimation) are presented in detail.

Chapter 4 presents 2-dimensional ionosphere models. The basic theory of spherical harmonics and thin plate spline is given and the implementation methods of the models are described in detail.

Chapter 5 shows the evaluation results of the two ionosphere models in two national and regional GPS networks in Canada. Accuracies of the different models are assessed. The spatial and temporal variation of a geomagnetic storm is also studied with the vertical TEC maps over Canada, and model accuracies between assessed for increased ionospheric conditions.

Chapter 6 gives conclusion and recommendations.

CHAPTER 2

GPS OVERVIEW AND ERROR MODELS

The GPS is a worldwide navigation system that consists of 24 satellites constellation in low earth orbit and ground tracking stations. GPS was originally designed for military uses to make sure that a user at any point on earth obtains accurate position, velocity and time. But now the number of non-military users of GPS surpasses those of military. The civil usage of GPS is very diverse. It includes surveying, geodesy, deformation monitoring of engineering structures, time transfer, aviation, space navigation and land navigation.

The basic observable of GPS is the measurement of the distance (range) between a satellite and a receiver. Each GPS satellite broadcasts radio signals containing information about their position and time. A receiver compares the time sent by GPS satellite with its own time and coverts the time difference into a range. The 3 dimensional position of the user is estimated using the range and satellite position information from at least 4 satellites. The accuracy of the position is dependent on the accurate time and position of satellites. Accurate time is achieved through the use of atomic clocks onboard every GPS satellite. Accuracy and integrity of GPS position are ensured by the ground control and tracking stations.

2.1 SYSTEM CONFIGURATION

GPS has three segments: the space segment, the control segment and the user segment. The space segment consists of 24 satellites placed in 6 orbital planes inclined at 55 degree relative to the equatorial plane. GPS satellites are in nearly circular orbit with an approximate altitude of 26,560km. The orbital period of GPS satellites is approximately 12 hours, with stationary ground tracks. The configuration of the satellites has been

designed to make sure that a minimum of 4 satellites can be seen in the sky anywhere on earth. Each satellite broadcasts ranging signals and navigation data, which is basic information for estimating the position of the user.

The control segment is composed of ground-based facilities, which includes monitor stations, a master control station (MCS) and uplink antennas. Monitor stations, located globally in longitude are equipped with GPS receivers, atomic clocks and data links to the MCS. GPS data collected by monitor stations are sent to the MCS, which processes the raw data and generates information of satellite orbit estimation, prediction and clock bias. The information calculated by the MCS is formatted into updated navigation messages and transmitted to the satellites via one of the uplink antennas. The navigation message is updated at least once per day. The control segment also determines GPS time, which is defined as the number of seconds elapsed from Saturday midnight of the present week. GPS time is based on the measurements from atomic clocks and differs from Universal Time Coordinated (UTC) by a nearly integer number of seconds because it is not incremented by leap seconds (Seeber, 2003). On January 1, 2003, the difference between GPS time and UTC was 13 seconds.

The user segment consists of the GPS receivers and the user community. Each GPS receiver is equipped with an antenna, clock, hardware and software. A GPS receiver detects the signal from a GPS satellite and converts into position, velocity and time estimates. Although GPS was developed for military applications, the civil uses of GPS are rapidly increasing with the decrease in the receiver price.

2.2 GPS SIGNALS

Each satellite transmits on two L-band radio frequencies: L1 (1575.42 MHz) and L2 (1227.60 MHz). These frequencies are coherent multiples of the fundamental frequency (10.23 MHz) of the atomic (Cesium) clock onboard each satellite. Each sinusoidal carrier

with the frequency mentioned above is modulated with ranging codes and navigation data.

The ranging code is a unique sequence of binary digits 0 and 1. This binary sequence apparently looks like random signals, but in reality follows a mathematical rule designed to ensure that all satellites transmit unique codes without interfering with each other. There are two kinds of code: a coarse/acquisition (C/A) code and an encrypted precision (P (Y)) code. C/A code consists of 1023 chips and repeats itself each millisecond. The chipping rate of C/A code is 1.023 Mcps (megachips/s), with each chip's duration time of 1 μs, which corresponds to a length of 300m (Spilker, 1996). P code is a very long sequence and repeats itself once a week. The chipping rate of P code is 10.23 Mcps, with the each chip's duration time of 0.1 μs, which corresponds to a length of 30m (Spilker, 1996). P code is allocated for military uses and encrypted with the Y code to limit access to these signals for non-authorized users.

The navigation message is a binary digit code containing information about satellite ephemeris, satellite health status, almanac and clock corrections. Each satellite transmits navigation messages at a rate of 50 bits per second. The whole message consists of 25 frames, each of which contains 1500 bits of information (Leick, 2004). A frame is divided into five 300-bit subframes. The first subframe contains information about clock corrections, including the error and drift rate of the clock. The second and third subframes contain satellite ephemeris parameters. The fourth subframe contains ionospheric parameters, UTC and almanac. The fifth subframe contains almanac and health status for all satellites. At the 50 bps rate it takes 6 seconds to transmit a subframe, 30 seconds for a frame and 12.5 minutes for the whole navigation message.

Each navigation message code is superposed with ranging codes using modulo-2 addition (Spilker, 1996). In the case that binary digits of the code and the navigation message have the same value (1 or 0), the added result is 0. If the values are different, the added result

is 1. The composite binary digit is imposed on two carriers L1 and L2 in a bi-phase modulation process (Spilker, 1996). If the bit is 0, the phase of the carrier remains unchanged. If the bit is 1, the phase of the carrier is shifted by 180 degrees. L1 is modulated with C/A and P (Y) code, while L2 is modulated with P (Y) code. Because L1 is modulated with two codes, two carriers are used on L1: the in-phase component and the quadrature component. The on-board clock generates the in-phase component and the quadrature component is obtained by shifting the in-phase component in phase by 90 degrees (Kaplan, 1996). The two GPS signals can be described mathematically by the following equations (Misra and Enge, 2001).

$$S_{L1}^{\ k}(t) = A_P P^k(t)D^k(t)\cos(2\pi f_{L1}t + \theta_{L1}) + A_C C^k(t)D^k(t)\sin(2\pi f_{L1}t + \theta_{L1})$$

$$S_{L2}^{\ k}(t) = B_P P^k(t)D^k(t)\cos(2\pi f_{L2}t + \theta_{L2}) \tag{2-1}$$

where

A_P and A_C are the amplitudes of P and C/A code components, respectively

B_P is the amplitude of L2 signal

$P^k(t)$ is P code sequence assigned to satellite k at time t

$C^k(t)$ is C/A code sequence assigned to satellite k at time t

$D^k(t)$ is navigation data for satellite k at time t

θ_{L1}, θ_{L2} are phase offsets

2.3 GPS OBSERVABLES

GPS measurements can be divided into two types: code measurements and phase measurements. Both provide a measure of the range between a satellite to a receiver.

Code observations are made by correlating the internal code generated within the receiver with those from the satellites. By shifting the internal replica of the code in time until a maximum correlation occurs, the transit time of the signal can be measured. The pseudo range, defined as the distance between the satellite and the receiver's antenna, is calculated by multiplying the signal transit time with the speed of the light. However, the pseudo range is biased by non-synchronization of the clock in the satellite and the receiver clock. This bias causes the pseudo range to differ from the geometric distance (or true range). In addition to clock bias, several error sources are included in the code pseudorange: ionosphere effect, troposphere effect, multipath effect and receiver noise.

The code pseudorange can be described mathematically by the following equation (Misra and Enge, 2001).

$$P = r + c(\delta t_u - \delta t^s) + I_P + T_P + \varepsilon_P + cT_{GD,r} - cT_{GD,s} \qquad (2\text{-}2)$$

where

P	is measured code pseudorange
r	is true or geometric range
c	is speed of light
δt_u	is the receiver clock bias
δt^s	is the satellite clock bias
I_P	is the code ionospheric delay
T_P	is the code tropospheric delay
ε_P	is the multipath error and receiver noise
$T_{GD,r}, T_{GD,s}$	are receiver and satellite modulation offsets in units of second

A phase observable is the difference between the received carrier phase from the satellite and the internal receiver carrier phase. The GPS receiver makes an initial measurement of the fractional phase and accumulates the changes to the phase. The initial integer phase cycle is ambiguous, but remains the same as long as the receiver does not lose the signal. The carrier pseudorange can be calculated by multiplying the measured phase by the wavelength of the carrier. This can be expressed in the following mathematical equation (Misra and Enge, 2001).

$$\Phi = \lambda\phi = r + c(\delta t_u - \delta t^s) + I_\Phi + T_\Phi + \lambda N + \varepsilon_\Phi \qquad\qquad (2\text{-}3)$$

where

Φ is phase range

ϕ is phase measurement

λ is the wavelength of the carrier

N is the integer ambiguity (cycle)

I_Φ is the phase ionospheric delay

T_Φ is the phase tropospheric delay

ε_Φ is the phase measurement noise

2.4 GPS OBSERVATION ERRORS

2.4.1 Satellite Clock and Ephemeris Error

Satellite ephemeris and clock information is determined by the Master Control Station based on the data from five monitoring stations of the ground segment. The Master Control Station estimates the current states of the satellite ephemeris and clock bias and predicts the future values of these parameters. This updated information is transmitted in

the navigation message. There are, however, residual errors in this estimation and prediction process, which grow with time from the last control station upload.

The broadcast error can be decomposed into 3 components: radial, along-track and cross-track. Along-track and cross-track component errors are several times larger than the radial component (Misra and Enge, 2001). But these components do not affect pseudorange measurements; only the projection of the ephemeris error along the line-of-sight direction creates a pseudorange error. Typical pseudorange error due to the radial component of the broadcast ephemeris error is 1~2m (Seeber, 2003).

The GPS clocks show some systematic and random errors due to the instability of the oscillator. The Master Control Station estimates and predicts clock errors in relation to the GPS time, which is the reference time scale used for all GPS clocks. GPS time is based on an atomic time scale and is manupulated to remain within one microsecond of UTC. The satellite clock error is available to GPS users via broadcast clock error coefficients in a polynomial form.

$$\Delta t = a_{f0} + a_{f1}(t - t_{0c}) + a_{f2}(t - t_{0c})^2 + \Delta t_r \qquad\qquad (2\text{-}4)$$

$$t_{\text{GPS}} = t - \Delta t$$

where

a_{f0} is clock bias

a_{f1} is clock drift

a_{f2} is clock drift rate

t_{0c} is reference time

Δt_r is small relativistic clock correction due to orbital eccentricity

Because the coefficients are predicted values, some errors exit even after time corrections. The typical error due to the clock error is 1~2 m in range (Seeber, 2003).

2.4.2 Ionospheric Error

The ionosphere is a part of the upper atmosphere that extends from 50km to 1000km above the earth surface. The main constituents of the ionosphere are ionized gases that are produced from the solar radiation. The ultraviolet radiation from the sun excites molecules in the atmosphere, producing free electrons and ions. The main effects of the ionosphere on the GPS signals are due to the free electrons. The free electrons change the propagation velocity of the GPS signals, generating ranging errors.

The density of electrons is dependent on the intensity of the solar radiation, which is a function of several factors including local time, season and solar cycle. In the daytime, the strong solar radiation breaks down molecules and produces many free electrons, with the peak electron density at 2 p.m local time. At night, without further ionization, the free electrons recombine with the ions, reducing electron density. The seasonal variation of the electron density is due to the revolution of the earth around the sun: the angle of solar radiation changes with season. The electron density also varies with the eleven-year solar cycle, which is closely related with the number of sunspots. The solar cycle is a measure of the periodic variation of the sun's activity. During the maximum of a solar cycle, an increased level of solar radiation produces more electrons in the ionosphere.

The ionosphere can be divided into several layers according to the electron density: D, E and F region (Rishbet and Garriot, 1969). Figure 2.1 (Davies, 1990) shows the vertical electron density profile of the ionosphere. The D region is a lowest part extending from 50km to 90km above the ground. The ionization in the D region is caused by X rays and Lyman α radiation. Due to the fact that the density of the air in this region is relatively

high, ions and free electrons quickly recombine at night, resulting in the disappearance of the D region.
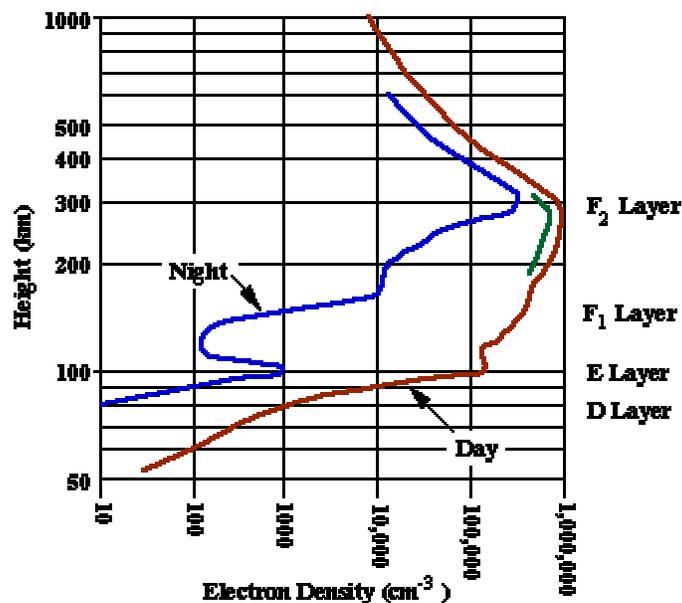
Figure 2.1 Vertical electron density profile of the ionosphere (Davies, 1990).

During daytime the D region absorbs radio signals at frequencies below approximately 7MHz and the electron density at these lower altitudes is relatively low, which means that the effect of the D region on GPS signals is negligible. The E region, extending from 90 to 140 km above earth surface, is produced by soft x rays. Because electrons and ions recombine quickly at night, the E region disappears after sunset - except during large geomagnetic storm events in which large increases in the E region electron density may be observed. The effect of the E region on GPS signals is generally negligible.

The F region lies between 140 and 1000 km and causes the most significant effect on GPS signals. The ionization in the F region is caused by the solar ultraviolet radiation. During daytime, the F region is split into two regions: F1 and F2 region. The F1 region extends from 140 to 210 km and accounts for 10 % of GPS signal delays. The F2 region extends from 210 to 1000 km and shows the highest electron density. The peak of the electron density in the F2 region occurs approximately between 250 and 400 km in

altitude. The altitude and electron density of the F region, however, depends on the time of day, the season and solar cycle.

The ionosphere is a dispersive medium, which means that the ionospheric effect on radio signals is a function of the frequency. Because of this dispersive nature of the ionosphere, GPS modulation (P code and C/A code) and carrier phase observables experience different effects in the ionosphere; the code is delayed and the phase is advanced (details are discussed in the next chapter). The amount of the code delay and phase advancement depends on the total electron content (TEC) along the signal path. The TEC is defined as the number of free electrons in a $1m^2$ column along the signal path from the satellite to the receiver (el/ m^2). The typical path delay at zenith direction is 1m at night and increases to 5~15 m during daytime (Grewal *et al.*, 2001).

2.4.3 Tropospheric Error

The troposphere is a lower part of the atmosphere and extends up to 50km from the earth's surface. The main constituent of the troposphere is dry gas (N_2, O_2); water vapor is a minor component but its effect on GPS signals are difficult to model. In contrast with the ionosphere, the troposphere is non-dispersive medium, which means that tropospheric effect on GPS signals is the same for both code and carrier phase observables and is independent of the frequency. The refractive index is larger than unity, and the GPS signal is delayed in the troposphere. The speed of the GPS signal is lower than that in vacuum and the resultant pseudorange becomes longer.

The signal delay in the troposphere is divided into two components: dry delay and wet delay. The dry delay is caused by the dry gases in the troposphere and can be estimated accurately by using ground temperature and pressure data with a theoretical model. The wet delay is due to water vapor and is difficult to estimate because water vapor density shows high temporal and spatial variations. Fortunately the predictable dry delay constitutes 90 % of the total tropospheric delay, while the wet delay make up only 10 %

(Janes *et.al.*, 1991). The tropospheric delay is often expressed as a product of the zenith delay (delay in the zenith direction) and a mapping factor that is a function of elevation angle. The mathematical expression of the tropospheric delay is

$$T(E) = T_{z,dry}m_{dry}(E) + T_{z,wet}m_{wet}(E)$$
(2-5)

where

$T(E)$ is total tropospheric delay at elevation angle E

$T_{z,dry}$ is dry zenith delay

$T_{z,wet}$ is wet zenith delay

$m_{dry}(E), m_{wet}(E)$ are dry and wet mapping functions at elevation angle E, respectively

There exist several troposphere models for the zenith wet and dry delay. These models are based on some approximation regarding variations of temperature and pressure with altitude. Two widely used troposphere models are the Saastamoinen and Hopfield models. In the Saastamoinen model (Saastamoinen, 1972), the zenith dry and wet delays are expressed as a function of temperature, pressure and partial pressure of water vapor at the receiver location. The mathematical description of the Saastamoinen model is given as (Misra and Enge, 2001)

$$T_{z,dry} = 0.002277(1 + 0.0026\cos 2\phi + 0.00028H)P_0 ,$$
(2-6)

$$T_{z,wet} = 0.002277(\frac{1255}{T_0} + 0.05)e_0$$
(2-7)

where

P_0 is total pressure at the receiver location

ϕ is the latitude of the receiver

H is the height of the receiver antenna

e_0 is the partial pressure of water vapor at the receiver location

T_0 is the Kelvin temperature at the receiver location

The Hopfield model (Hopfield, 1969) is based on a quartic zenith model of the refraction, which is derived empirically from extensive measurements. The zenith dry delay and wet delay are expressed as follows

$$T_{z,dry} = 77.64 \cdot \frac{10^{-6}}{5} \frac{P_0}{T_0}[40136 + 148.72(T_0 - 273.16)] \qquad (2\text{-}8)$$

$$T_{z,wet} = \frac{10^{-6}}{5}(-12.96 \cdot T_0 + 3.718 \cdot 10^5)\frac{e_0}{T_0^{\;2}} \cdot 11000 \qquad (2\text{-}9)$$

The typical zenith dry delay is 2.3m at sea level, but increases to about 20m near horizon (Seeber, 2003).

2.4.4 Multipath Error

Multipath occurs when a signal arrives at an antenna via multiple paths due to reflection (Figure 2.2). The antenna receives the direct line-of-sight signal and secondary reflected signals from the nearby objects. The reflected signals are delayed with respect to the direct signal due to the longer path caused by reflection; the reflected signals interfere with the direct signal with a different phase and amplitude. The resultant composite signal shows a distorted phase and amplitude from the original ones, which results in a range error. The range error depends on the relative amplitude, delay and phase of the reflected signal to the direct signal (Grewal *et al.*, 2001).
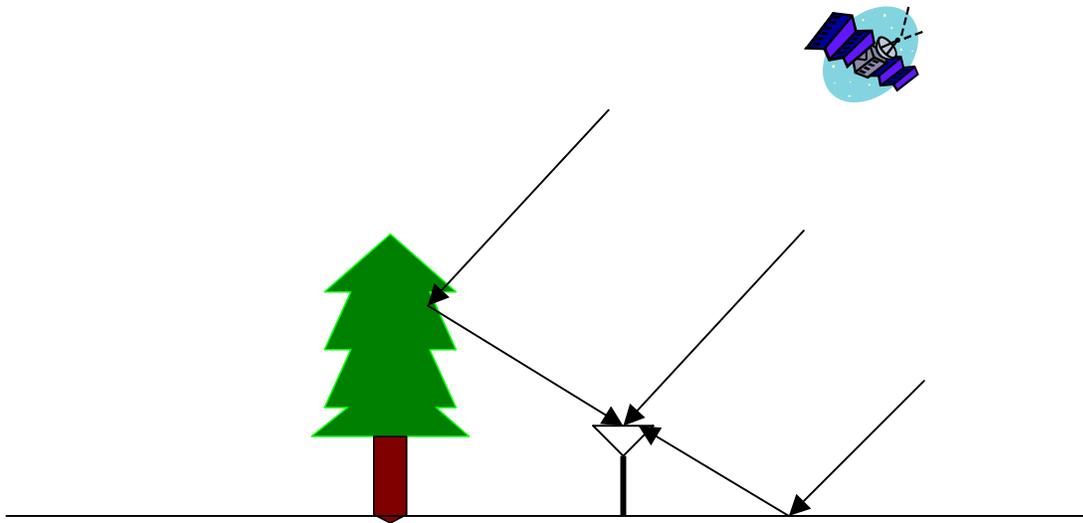
Figure 2.2 Multipath.

Both GPS code and phase observables are susceptible to multipath, but code ranges are more highly affected. The magnitude of code range error ranges from 1m to 5m, while phase range error is less than 5cm. The upper bound of the phase range error can be estimated assuming a simplified multipath situation. When an antenna receives a direct signal and a reflected signal, the composite signal can be represented as the sum of two signals

$$\underbrace{A\cos\phi}_{\text{direct signal}} + \underbrace{\gamma A\cos(\phi+\Delta\phi)}_{\text{reflected signal}} \qquad (2\text{-}10)$$

where

A, ϕ are the amplitude and phase of the direct signal

$\Delta\phi$ is the phase shift of the reflected signal

γ is an amplitude damping coefficient due to the reflection

Using trigonometric relations, the phase measurement error can be derived as

$$\delta\phi = \arctan\left(\frac{\gamma \sin \Delta\phi}{1 + \gamma \cos \Delta\phi}\right) \hspace{3cm} \text{(2-11)}$$

Because the damping coefficient γ varies from 0 to 1, the maximum value of the phase error $\delta\phi$ is 90 degrees, or 5cm in L1 range.

The most basic method of reducing multipath is to locate an antenna at a place with minimum reflecting objects. Another method is to use an antenna that has low response in the direction of the reflecting objects. Multipath can also be reduced by adopting receiver technologies like narrow-correlator (Van Dierendonck *et al.*, 1992), Multipath Estimation Technology (MET) (Townsend *et al.*, 1995), or Multipath Mitigation Method (MMT) (Weil *et al.*, 2000).

2.4.5 Receiver Noise

GPS code and phase measurements are affected by random noises. These include thermal noise, dynamic stress in code and phase tracking loops, oscillator noise, interference from other signals and signal quantization noise. The thermal noise and dynamic stress are the dominant error sources in both phase and code measurements. The thermal noise per chip is the same for C/A code and P code (Ward, 1996). But, because the chip width of C/A code is ten times that of P code, the thermal noise of C/A code is ten times higher that that of P code.

The receiver noise is dependent on the signal strength, which is a function of the satellite elevation angle (Van Dierendonck, 1996). Modern receiver can track the phase with the precision of 1% of a cycle (Seeber, 2003).

2.5 LINEAR COMBINATION OF GPS OBSERVABLES

New GPS observables can be generated by linear combination of dual-frequency phase and code observables. These are used for various methods, including derivation of TEC, ambiguity resolution and cycle slip detection. The linear combination of two phase (or code) observations Φ_1 and Φ_2 can be expressed like

$$\Phi = n_1\Phi_1 + n_2\Phi_2 \qquad \text{(Phase)} \qquad\qquad (2\text{-}12)$$

$$P = n_1 P_1 + n_2 P_2 \qquad \text{(Code)} \qquad\qquad (2\text{-}13)$$

where n_1 and n_2 are arbitrary numbers.

2.5.1 Geometry-free Linear Combination

The geometry-free linear combination is generated by subtracting the phase (or code) observable of L2 from that of L1 signal. This is also referred to as the L4 combination. In the case of phase observables, the geometry-free combination is derived using equation (2-3)

$$\Phi_{gf} = \Phi_{L1} - \Phi_{L2} = \lambda_{L1}N_{L1} - \lambda_{L2}N_{L2} - I_{\phi,L1} + I_{\phi,L2} + \varepsilon_{\Phi_{L1}-\Phi_{L2}} \qquad\qquad (2\text{-}14)$$

In this combination, the geometric terms including range, clock offsets and tropospheric delay are eliminated and only ionospheric delay and ambiguity terms remain. The code geometry-free combination is derived in a same way as that of phase observables using equation (2-2);

$$P_{gf} = P_{L1} - P_{L2} = I_{P,L1} - I_{P,L2} + \varepsilon_{P_{L1}-P_{L2}} \tag{2-15}$$

From equation (2-15), it can be deduced that ionospheric delay can be measured directly from the code geometric-free combination.

2.5.2 Wide-lane Linear Combination

The wide-lane linear combination is obtained by subtracting the L2 phase (or code) observable in unit of cycles from that of the L1 signal.

$$\phi_{wl} = \phi_{L1} - \phi_{L2} = \frac{\Phi_{L1}}{\lambda_{L1}} - \frac{\Phi_{L2}}{\lambda_{L2}} = \rho'(\frac{1}{\lambda_{L1}} - \frac{1}{\lambda_{L2}}) + \frac{1}{c}(f_{L1}I_{\phi,L1} - f_{L2}I_{\phi,L2}) + (N_{L1} - N_{L2}) + \varepsilon_{\phi_{L1}-\phi_{L2}}$$

$$= \rho*(\frac{f_{L1}-f_{L2}}{c}) + \frac{1}{c}(f_{L1}I_{\phi,L1} - f_{L2}I_{\phi,L2}) + (N_{L1} - N_{L2}) = \frac{\rho*}{\lambda_{wl}} + \frac{1}{c}(f_{L1}I_{\phi,L1} - f_{L2}I_{\phi,L2}) + N_{wl} + \varepsilon_{\phi_{wl}}$$

$$\tag{2-16}$$

where

$\lambda_{wl} = \dfrac{c}{f_{L1} - f_{L2}} = 0.862\text{m}$ is wide-lane wave length

N_{wl} is called the wide-lane ambiguity

$\rho*$ is non-dispersive delay, which includes geometric range, tropospheric delay and clock offset

$\varepsilon_{\phi_{wl}}$ is wide-lane measurement noise in units of cycle

Because the wavelength of the wide-lane combination is much longer than those of L1 and L2, this combination is used for cycle slip detection or ambiguity resolution. Equation (2-16) can be rewritten in metric units by multiplying it with λ_{wl}.

$$\Phi_{wl} = \lambda_{wl}\phi_{wl} = \rho* + \frac{1}{f_{L1}-f_{L2}}(f_{L1}I_{\phi,L1} - f_{L2}I_{\phi,L2}) + \lambda_{wl}N_{wl} + \lambda_{wl}\varepsilon_{\phi_{wl}} \tag{2-17}$$

Once the wide-lane ambiguities are resolved, the above equation can be written as

$$\Phi_{wl} = \frac{f_{L1}\Phi_{L1} - f_{L2}\Phi_{L2}}{f_{L1}-f_{L2}} = \frac{154}{34}\Phi_{L1} - \frac{120}{34}\Phi_{L2} \tag{2-18}$$

The wide-lane combination for code observables is derived similarly using equation (2-2).

$$P_{wl} = \rho* + \frac{1}{f_{L1}-f_{L2}}(f_{L1}I_{P,L1} - f_{L2}I_{P,L2}) + \varepsilon_{P_{wl}} = \frac{f_{L1}P_{L1} - f_{L2}P_{L2}}{f_{L1}-f_{L2}} = \frac{154}{34}P_{L1} - \frac{120}{34}P_{L2}$$

$$\tag{2-19}$$

One drawback of the wide-lane combination is the increase of the measurement noise due to large coefficients: it increases by a factor of six times of that of L1 or L2 alone.

2.5.3 Narrow-lane Linear Combination

The narrow-lane observable is generated by adding L1 phase (or code) in unit of cycles with that of L2.

$$\phi_{nl} = \phi_{L1} + \phi_{L2} = (\frac{f_{L1} + f_{L2}}{c})\rho* + \frac{1}{c}(f_{L1}I_{\Phi,L1} + f_{L2}I_{\Phi,L2}) + N_{L1} + N_{L2} + \varepsilon_{\phi_{L1}+\phi_{L2}} \qquad (2\text{-}20)$$

$$= \frac{\rho*}{\lambda_{nl}} + \frac{1}{c}(f_{L1}I_{\Phi,L1} + f_{L2}I_{\Phi,L2}) + N_{nl} + \varepsilon_{\phi_{nl}}$$

where

$\lambda_{nl} = \dfrac{c}{f_{L1} + f_{L2}} = 10.70 \text{ cm}$ refers to the narrow-lane wavelength

$N_{nl} = N_{L1} + N_{L2}$ refers to the narrow-lane ambiguity

$\varepsilon_{\phi_{nl}}$ is narrow-lane measurement noise

Equation (2-20) can be rewritten in metric units by multiplying with λ_{nl}.

$$\Phi_{nl} = \lambda_{nl}\phi_{nl} = \frac{f_{L1}}{f_{L1} + f_{L2}}\Phi_{L1} + \frac{f_{L2}}{f_{L1} + f_{L2}}\Phi_{L2} + \lambda_{nl}N_{nl} + \lambda_{nl}\varepsilon_{\phi_{nl}} \qquad (2\text{-}21)$$

$$= \frac{154}{274}\Phi_{L1} + \frac{120}{274}\Phi_{L2} + \lambda_{nl}N_{nl}$$

Due to the small wavelength of the narrow-lane observables, more precise measurements are possible. Similarly the narrow-lane combination can be formed for code observables.

$$P_{nl} = \rho* + \frac{1}{f_{L1} + f_{L2}}(f_{L1}I_{P,L1} + f_{L2}I_{P,L2}) + \varepsilon_{P_{nl}} = \frac{f_{L1}}{f_{L1} + f_{L2}}P_{L1} + \frac{f_{L2}}{f_{L1} + f_{L2}}P_{L2} \qquad (2\text{-}22)$$

$$= \frac{154}{274}P_{L1} + \frac{120}{274}P_{L2}$$

CHAPTER 3

IONOSPHERIC EFFECTS ON GPS SIGNALS

As described in Section 2.4.2, the velocity of GPS signals is altered in the ionosphere, resulting in variation of the signal transit time. The velocity of radio signals in a medium can be formulated in terms of the refractive index in the medium. If the refractive index is a function of the signal's frequency, the medium is called dispersive, which is the case of the ionosphere. By using this dispersive nature of the ionosphere, the GPS signal delay in the ionosphere can be estimated using dual-frequency measurements.

In this chapter, a mathematical description of the refractive index in the ionosphere is presented, and ionospheric effects on GPS signals are derived using the dispersive nature of the ionosphere.

3.1 REFRACTIVE INDEX OF IONOSPHERE

The refractive index of a medium is defined as the ratio of the signal speed in the medium to that in the vacuum and can be described by

$$n = \frac{c}{v} \tag{3-1}$$

where c is the speed of the radio wave in the vacuum, v is the speed of the wave in the medium.

In a dispersive medium like the ionosphere, the carrier phase velocity is different from the modulated signal velocity, which is called the group velocity. Consider two signals

with frequencies f_1, f_2 (or $\omega_1 = 2\pi f_1$, $\omega_2 = 2\pi f_2$) and phase velocities v_1, v_2 traveling in the x direction. The sum of two signals can be expressed as (Kaplan *et al.*, 1996),

$$S(x,t) = S_1(x,t) + S_2(x,t) = \sin\omega_1\left(t - \frac{x}{v_1}\right) + \sin\omega_2\left(t - \frac{x}{v_2}\right)$$

(3-2)

Using trigonometric relation, the following equation can be derived (Kaplan, 1996)

$$S(x,t) = 2\cos\frac{1}{2}(\omega_1 - \omega_2)\left[t - \frac{x}{\dfrac{\frac{1}{2}(\omega_1 - \omega_2)}{\frac{1}{2}\left(\dfrac{\omega_1}{v_1} - \dfrac{\omega_2}{v_2}\right)}}\right] \times \sin\left[\frac{1}{2}(\omega_1 + \omega_2)t - \frac{1}{2}\left(\frac{\omega_1}{v_1} + \frac{\omega_2}{v_2}\right)x\right]$$

(3-3)

The first part is the modulated signal and moves with the group velocity v_g

$$v_g = \frac{\frac{1}{2}(\omega_1 - \omega_2)}{\frac{1}{2}\left(\dfrac{\omega_1}{v_1} - \dfrac{\omega_2}{v_2}\right)} = \frac{\pi(f_1 - f_2)}{\pi\left(\dfrac{f_1}{v_1} - \dfrac{f_2}{v_2}\right)} = \frac{\left(\dfrac{v_1}{\lambda_1} - \dfrac{v_2}{\lambda_2}\right)}{\left(\dfrac{1}{\lambda_1} - \dfrac{1}{\lambda_2}\right)} = v_1 - \lambda_1\left(\frac{v_2 - v_1}{\lambda_2 - \lambda_1}\right)$$

(3-4)

where $\lambda_1 = \dfrac{v_1}{f_1}$, $\lambda_2 = \dfrac{v_2}{f_2}$ are the wavelength of each signal.

Assuming that $v_2 - v_1$ and $\lambda_2 - \lambda_1$ are very small, the above equation can be written as

$$v_g = v_p - \lambda_p \frac{dv_p}{d\lambda_p}$$ (3-5)

where $v_1 = v_p$, $v_2 = v_1 + dv_p$, $\lambda_1 = \lambda_p$ and $\lambda_2 = \lambda_1 + d\lambda_p$.

The above equation can be rewritten in terms of phase refractive index and group refractive index, which are defined respectively as $n_p = \frac{c}{v_p}$, $n_g = \frac{c}{v_g}$:

$$n_g = n_p + f \frac{dn_p}{df}$$ (3-6)

The phase refractive index in the ionosphere is described by the Appleton-Hartree equation (Klobuchar, 1996);

$$n_p{}^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \left[\frac{Y_T^2}{4(1 - X - iZ} + Y_L^2 \right]^{1/2}}$$ (3-7)

where

$X \qquad = N_e e^2 / \varepsilon_0 m w^2$

$Y_L \qquad = eB_L / mw = f_H \cos\theta / f$

$Y_T \qquad = eB_T / mw = f_H \sin\theta / f$

$Z \qquad = v / w$

$N_e \qquad$ is electron density (el/m^3)

$e \qquad = -1.602 \times 10^{-19}$ coulomb (electron charge)

ε_0 $= 8.854 \times 10^{-12}$ farad/m (permittivity of free space)

m $= 9.107 \times 10^{-31}$ kg (rest mass of an electron)

θ is the angle of the ray with respect to the Earth's magnetic filed

v is the electron-neutral collision frequency

f_H is the electron gyro frequency

B_T, B_L are the longitudinal and transverse components of the magnetic field with respect to the direction of propagation

If absorption and geomagnetic effects are negligible, the above equation can be simplified by letting $Y_L = Y_T = Z = 0$.

$$n_p^2 = 1 - X \Rightarrow n_p = \sqrt{1 - X} \tag{3-8}$$

Equation (3-8) can be expanded using a Taylor series and approximated by taking only up to the first-order term

$$n_p = 1 - \frac{X}{2} - \frac{X^2}{8} \dots \cong 1 - \frac{X}{2} = 1 - \frac{N_e e^2}{2\varepsilon_0 m (2\pi f)^2} \approx 1 - \frac{40.3 N_e}{f^2} \tag{3-9}$$

The approximation error arising from taking only up to the first-order term is about 1~2 mm in the GPS range during maximum solar activity (Brunner and Gu, 1991). The group refractive index can be derived from equation (3-6) and equation (3-9) as

$$n_g = n_p + f \frac{dn_p}{df} = 1 + \frac{40.3 N_e}{f^2} \tag{3-10}$$

Using equations (3-1), (3-9) and (3-10), the phase and group velocity can be derived as

$$v_p = \frac{c}{1 - \frac{40.3 N_e}{f^2}} \tag{3-11a}$$

$$v_g = \frac{c}{1 + \frac{40.3 N_e}{f^2}} \tag{3-11b}$$

From equations (3-11a) and (3-11b), it can be deduced that the phase velocity in the ionosphere exceeds the group velocity and free-space velocity.

3.2 IONOSPHERIC DELAY

As mentioned in Chapter 2, the basic GPS measurement is based on the signal transit time between a satellite and a receiver. But as a result of equation (3-11), the transit time of the carrier phase is decreased and that of the modulated code is increased. The change of the transit time can be converted to a phase advance and a code pseudorange delay through comparisons with the geometric range between the satellite and the receiver. The amount of the observed range compared with the geometric range is

$$I = \int_{SV}^{reciever} n \, dl - \int_{SV}^{reciever} dl = \int_{SV}^{receiver} (n-1) \, dl \tag{3-12}$$

In the case of the carrier phase, equation (3-12) can be rewritten using equation (3-9)

$$I_\Phi = \int\limits_{SV}^{receiver} (1 - \frac{40.3N_e}{f^2} - 1)dl = -\int\limits_{SV}^{receiver} \frac{40.3N_e}{f^2}dl = -\frac{40.3}{f^2}\int\limits_{SV}^{receiver} N_e dl \qquad (3\text{-}13)$$

Similarly, the code range delay can be derived using equation (3-10)

$$I_P = \int\limits_{SV}^{receiver} (1 + \frac{40.3N_e}{f^2} - 1)dl = \int\limits_{SV}^{receiver} \frac{40.3N_e}{f^2}dl = \frac{40.3}{f^2}\int\limits_{SV}^{receiver} N_e dl \qquad (3\text{-}14)$$

The total number of electrons along the signal path from the satellite to the receiver is defined as the total electron content (TEC) and expressed mathematically as

$$TEC = \int\limits_{SV}^{receiver} N_e dl \qquad (3\text{-}15)$$

The TEC is expressed in units of electrons/m^2. Using equations (3-15), (3-13) and (3-14), the ionospheric range errors are rewritten as

$$I_\Phi = -\frac{40.3TEC}{f^2} \qquad (3\text{-}16a)$$

$$I_P = \frac{40.3TEC}{f^2} \qquad (3\text{-}16b)$$

Equations (3-16a) and (3-16b) show that the phase advance and code delay in pseudorange measurements are equal in magnitude and opposite in sign.

3.2.1 Dual Frequency Estimation of Ionospheric Delay

A dual-frequency GPS receiver can estimate the ionospheric delay by exploiting the dispersive nature of the ionosphere. Inserting equation (3-16) into equation (2-2), the code pseudorange on L1 can be written as

$$P_{L1} = r + c(\delta t_u - \delta t^s) + T_P + \frac{40.3 TEC}{f_{L1}^2} + cT_{GD,r} - cT_{GD,s} + \varepsilon_{P,1} \qquad (3\text{-}17)$$

Similarly, the code pseudorange on L2 is

$$P_{L2} = r + c(\delta t_u - \delta t^s) + T_P + \frac{40.3 TEC}{f_{L2}^2} + c\frac{f_{L1}^2}{f_{L2}^2}T_{GD,r} - c\frac{f_{L1}^2}{f_{L2}^2}T_{GD,s} + \varepsilon_{P,2} \qquad (3\text{-}18)$$

where $\dfrac{f_{L1}^2}{f_{L2}^2}$ is a scale factor for converting $T_{GD,r}$ and $T_{GD,s}$ on L1 to $T_{GD,r}$ and $T_{GD,s}$ on L2

By differencing equation (3-17) from equation (3-18), the geometric range, clock biases and tropospheric delay terms are eliminated.

$$P_{L1} - P_{L2} = 40.3 TEC\left(\frac{1}{f_{L1}^2} - \frac{1}{f_{L2}^2}\right) + \varepsilon_{P,12} + b_r + b_s \qquad (3\text{-}19)$$

where

$$\varepsilon_{P,12} = \sqrt{\varepsilon^2_{P,1} + \varepsilon^2_{P,2}}$$

$$b_s \quad = -\left(1 - \frac{f_{L1}^2}{f_{L2}^2}\right)cT_{GD,s} \qquad \text{is the satellite inter-channel bias}$$

$$b_r = \left(1 - \frac{f_{L1}^2}{f_{L2}^2}\right) cT_{GD,r}$$
is the receiver inter-channel bias

From equation (3-19), TEC can be derived and written as

$$TEC_P = [(P_{L1} - P_{L2}) - \varepsilon_{P,12} - b_s - b_r] \frac{f_{L1}^2 f_{L2}^2}{f_{L2}^2 - f_{L1}^2} \frac{1}{40.3} \tag{3-20}$$

Usually TEC is measured in TEC units (TECU) and one TECU is defined as 10^{16} electrons/m^2. As shown in equation (3-20), the correct calibration of the satellite and receiver inter-channel biases is required to derive an absolute TEC. In addition to that, code-generated TEC is corrupted by a high level of noise $\varepsilon_{P,12}$. The code ionospheric delay on L1 can be derived from equations (3-20) and (3-16).

$$I_{P,L1} = \frac{40.3 TEC}{f_{L1}^2} = \frac{f_{L2}}{f_{L2}^2 - f_{L1}^2} [(P_{L1} - P_{L2}) - \varepsilon_{\rho,12} - b_s - b_r] \tag{3-21}$$

TEC can also be derived from the phase observables in a similar way as above. The result is

$$TEC_\Phi = -[(\Phi_{L1} - \Phi_{L2}) - \varepsilon_{\Phi,12} - \lambda_1 N_1 + \lambda_2 N_2] \frac{f_{L1}^2 f_{L2}^2}{f_{L2}^2 - f_{L1}^2} \frac{1}{40.3} \tag{3-22}$$

where

$$\varepsilon_{\Phi,12} = \sqrt{\varepsilon_{\Phi_{L1}}^2 + \varepsilon_{\Phi_{L2}}^2}$$

In equation (3-22), the inter-channel bias terms are missing because phase inter-channel biases are very small compared to code inter-channel biases. The TEC derived from

phase observables contains ambiguous integer biases (N_1 and N_2) but less noise due to smaller noise terms ($\varepsilon_{\Phi,1}$ and $\varepsilon_{\Phi,2}$). The phase ionospheric delay on L1 can be derived from equations (3-22) and (3-16)

$$I_{\Phi,L1} = [(\Phi_{L1} - \Phi_{L2}) - \varepsilon_{\Phi,12} - \lambda_1 N_1 + \lambda_2 N_2] \frac{f_{L2}^2}{f_{L2}^2 - f_{L1}^2} \tag{3-23}$$

3.2.2 Single-frequency Model of Ionospheric Delay

A single-frequency GPS receiver can make use of a simple broadcast ionospheric delay model called Klobuchar model, which represents the diurnal variation of the zenith ionospheric with a half-cosine function (Klobuchar, 1996). The zenith ionospheric delay in units of time is represented as

$$T_{zenith,iono} = \frac{I_{zenith}}{c} = DC + A\cos[\frac{2\pi(t-B)}{T}], \text{ if } |t-B| < P/4 \tag{3-24}$$

$$= DC \qquad\qquad , \text{ if } |t-B] > P/4$$

where

DC   is an offset constant (5ns)

A amplitude of the cosine function

B local time when the cosine function is at maximum (14h, local time)

T period of the cosine function

t local time of the ionospheric pierce point where the signal intersects with mean ionospheric height (350 km)

Amplitude A and period P are represented as third-order polynomials, with 4 coefficients that are contained in the navigation message.

$$A = \sum_{n=0}^{3} \alpha_n \phi_m^n, \quad P = \sum_{n=0}^{3} \beta_n \phi_m^n \tag{3-25}$$

where

ϕ_m is the magnetic latitude of the ionospheric pierce point

α_n, β_n are coefficients that are selected by Master Control Station.

The zenith ionospheric delay given above is converted to the slant delay by applying an obliquity factor

$$OF(E) = 1.0 + 16.0 \times (0.53 - E)^3 \tag{3-26}$$

where E is elevation angle in units of semi-circles. Therefore the slant ionospheric delay is given as

$$T_{slant,iono} = \frac{I_{slant}}{c} = OF(E)(DC + A\cos[\frac{2\pi(t-B)}{P}]), \text{ if } |t-B| < P/4 \tag{3-27}$$

$$= OF(E) \times DC \qquad\qquad , \text{ if } |t-B| > P/4$$

3.3 SMOOTHING OF TEC ESTIMATES

As noted above, TEC can be derived from both dual-frequency code and phase observables. The code-derived TEC, however, is corrupted by a high level of noise

factors including multipath effects and random measurement noise. The phase-derived TEC is also corrupted by these factors, but to a much lesser extent and can be derived with precision of on the order 0.10 TECU (Skone, 2002). However, due to the ambiguous integer bias terms, the phase-derived TEC is a relative value. One way to deal with the drawbacks inherent in TEC estimates derived separately from code and phase measurements is to combine these measurements. The basic idea of this, referred to as a smoothing technique, is to smooth the noisy code-derived absolute TEC with the precise phase-derived relative TEC. First, the offset between the code-generated TEC and the phase-generated TEC is calculated by subtracting equation (3-22) from equation (3-20)

$$Offset = TEC_\rho - TEC_\Phi = [-\lambda_1 N_1 + \lambda_2 N_2 - b_r - b_s] \frac{f_{L1}^2 f_{L2}^2}{f_{L2}^2 - f_{L1}^2} \frac{1}{40.3} \qquad (3\text{-}28)$$

And then the calculated offset is averaged over time

$$\hat{Offset}(t_k) = \frac{1}{N} \sum_{l=1}^{N} (TEC_\rho(t_l) - TEC_\phi(t_l)) \qquad (3\text{-}29)$$

where

k,l are time indices

N is the number of epoch between t_i and t_j

The final smoothed TEC can be derived as

$$TEC_{smoothed}(t_k) = TEC_\phi(t_k) + offset(t_k) \qquad (3\text{-}30)$$

In equation (3-29), all offset values are given the same weights, but in reality, the noise level of the offset values is dependent on the elevation angle of the satellite being tracked. Therefore it is reasonable to give different weights according to the elevation angle of each measurement. With this in mind, equation (3-29) can be rewritten in a recursive way using sequential least-squares estimation.

$$\hat{offset}(t_{k+1}) = \hat{offset}(t_k) + K(t_{k+1})[\hat{offset}(t_{k+1}) - \hat{offset}(t_k)] \tag{3-31}$$

$$K(t_{k+1}) = C_{est}(t_k)[C_{est}(t_k) + C_{obs}(t_k)]^{-1} \tag{3-32}$$

$$C_{obs}(t_k) = M^{-1}(E(t_k)), \; C_{est}(t_k) = [1 - K(t_{k-1})]C_{est}(t_{k-1}) \tag{3-33}$$

where

$\hat{offset}(t_k)$ is estimated offset at epoch t_k

$K(t_k)$ is gain factor at epoch t_k

$C_{est}(t_k)$ variance of the estimated offset at epoch t_k

$C_{obs}(t_k)$ variance of the observed offset at epoch t_k

$M(E(t_k))$ weighting function for a measurement at elevation angle E

3.4 CYCLE SLIP DETECTION AND CORRECTION

The integer ambiguity terms in the carrier phase observables remain fixed as long as the phase lock loop keeps track of the signal. But, sometimes loss of the phase lock can occur and results in a discontinuity in the integer cycle number, referred to as a cycle slip. After a cycle slip, all carrier phase observables are shifted by an integer amount. The smoothing of TEC in the previous section is possible only when continuous relative TEC measurements are available and for this reason, cycle slips in the phase observables must

be detected and corrected if possible. There are several algorithms for detection and correction of cycle slips. In this thesis, algorithms developed by Blewitt (1990) and Skone (2002) are used.

3.4.1 Cycle Slip Detection

For purposes of detecting cycle slips in phase observables, two linear combination of observables are formed: widelane combination and geometric-free combination.

3.4.1.1 Cycle Slip Detection in Widelane Combination

As shown in Chapter 2, the wavelength of the widelane phase observable is much larger than L1 and L2 wavelengths. Therefore the widelane observable has an advantage in detecting cycle slips. The widelane ambiguity term is derived by differencing equation (2-22) from equation (2-18)

$$\Phi_{wl} - P_{nl} = \lambda_{wl} N_{wl} \tag{3-34}$$

$$N_{wl} = \frac{1}{\lambda_{wl}} (\Phi_{wl} - P_{nl}) \ \text{(cycles)} \tag{3-35}$$

The cycle slip detection algorithm is based on the assumption that the widelane ambiguity derived above remains constant as long as there are no cycle slips. And an appropriate detection criterion is set by accounting for the magnitude of the multipath errors present in equation (3-34) for narrowlane code observables. A cycle slip detection criterion can be formulated as follows (Skone, 2002):

$$| N_{wl,k} - \langle N_{wl} \rangle_k | > F \times \sigma_{k1} \tag{3-36}$$

where

$\langle \cdot \rangle$ is mean value

F is a threshold factor

σ is variance

$$\langle N_{wl} \rangle_k = \langle N_{wl} \rangle_{k-1} + \frac{1}{k}(N_{wl,k} - \langle N_{wl} \rangle_{k-1}) \tag{3-37}$$

$$\sigma_k^2 = \sigma_{k-1}^2 + \frac{1}{k}[(N_{wl,k} - \langle N_{wl} \rangle_{k-1})^2 - \sigma_{k-1}^2] \qquad (k \leq 10) \tag{3-38}$$

$$\sigma_k^2 = 0.90 \times \sigma_{k-1}^2 + 0.10 \times (N_{wl,k} - \langle N_{wl} \rangle_{k-1})^2 \qquad (k > 10) \tag{3-39}$$

For survey-grade dual frequency GPS receivers with choke-ring antennas, the threshold factor is set to 4~6 (Skone, 2002).

3.4.1.2 Cycle Slip Detection in Geometry-free Combination

In a very rare case, the same amount of cycle slips can occur on L1 and L2. In this case, the cycle slips are not detected using widelane phase observables. The geometric-free combination can be a good alternative for such a rare case. The two geometric-free observations from both code and phase measurements are shown in equations (2-14) and (2-15. If the two equations are added,

$$B_{gf} = \Phi_{gf} + P_{gf} = \lambda_{L1} N_{L1} - \lambda_{L2} N_{L2} \tag{3-40}$$

The resulting B_{gf} consists of only integer ambiguity terms and therefore remains constant over time without any cycle slips. The cycle slip detection for the geometric-free observables is done in a similar way as that used for widelane observables.

$$| B_{gf,k} - \langle B_{gf} \rangle_k | > F \times \sigma_{k-1} \qquad (3\text{-}41)$$

where

$$\langle B_{gf} \rangle_k = \langle B_{gf} \rangle_{k-1} + \frac{1}{k}(B_{gf,k} - \langle B_{gf} \rangle_{k-1}) \qquad (3\text{-}42)$$

$$\sigma_k^2 = \sigma_{k-1}^2 + \frac{1}{k}[(B_{gf,k} - \langle B_{gf} \rangle_{k-1})^2 - \sigma_{k-1}^2] \qquad (k \le 10) \qquad (3\text{-}43)$$

$$\sigma_k^2 = 0.90 \times \sigma_{k-1}^2 + 0.10 \times (B_{gf,k} - \langle B_{gf} \rangle_{k-1})^2 \qquad (k > 10) \qquad (3\text{-}44)$$

For survey-grade GPS receivers, F is set to 4~6 (Skone, 2002).

3.4.2 Cycle Slip Correction

When a cycle slip occurs at an epoch, sometimes it is possible to correct discontinuities in TEC due to cycle slips by extrapolating from the previous epoch. For that, the TEC change rate at the previous epoch must be available. When no cycle slip occurs at t_j, the TEC change rate is derived as

$$\frac{\Delta TEC_k}{\Delta t} = \frac{TEC(t_k) - TEC(t_{k-1})}{t_k - t_{k-1}} \qquad (3\text{-}45)$$

If there is no cycle slip at t_{k+1}, the prediction accuracy is estimated by using the difference between the predicted TEC and the observed TEC.

$$\delta TEC_{k+1} = TEC_{k+1} - \{TEC_k + \frac{\Delta TEC_k}{\Delta t}(t_{k+1} - t_k)\} \qquad (3\text{-}46)$$

And an updated variance of the prediction accuracy is

$$\sigma_{k+1}^2 = \sigma_k^2 + \frac{1}{k+1}\{(\delta TEC_{k+1} - \langle \delta TEC_\phi \rangle_k)^2 - \sigma_k^2\} \tag{3-47}$$

If a cycle slip occurs at t_{k+1}, the following test is done.

$$\sigma_k^2 < 0.10 \text{ TECU} \tag{3-48}$$

If the above condition is satisfied, the TEC values from t_{k+1} are shifted by the offset calculated as the discontinuity between the predicted value and the observed value at t_{k+1}

$$offset_{k+1} = TEC_{k+1} - \{TEC_k + \frac{\Delta TEC_k}{\Delta t}(t_{k+1} - t_k)\} \tag{3-49}$$

CHAPTER 4

TWO-DIMENSIONAL IONOSPHERE MODEL

4.1 INTRODUCTION

The smoothed TEC derived in Chapter 3 is, by definition, the total electron content along the line-of-sight from the satellite to the receiver biased by satellite and receiver interchannel biases. Electron density profiles with height cannot be obtained reliably with ground-based GPS observations only. For this reason, a single layer ionosphere model is used in this thesis. As shown in Figure 4.1, the single layer model is based on the assumption that all electrons in the ionosphere are concentrated in an infinitesimally thin shell at altitude H (normally taken to be 350 km)-see Figure 4.1. Therefore TEC values from various line-of-sights, referred to as slant TECs, can be mapped to a two-dimensional plane. But, in order to eliminate the dependence on elevation angle, the slant TECs are converted to vertical TECs (VTECs) by applying a mapping function. The mapping function used in this thesis is

$$M(E) = [1 - (\frac{R_E}{R_E + H})^2 \cos^2 E]^{-\frac{1}{2}}$$
(4-1)

where

R_E is the earth radius

E is the satellite elevation angle

H is the altitude of the ionospheric shell
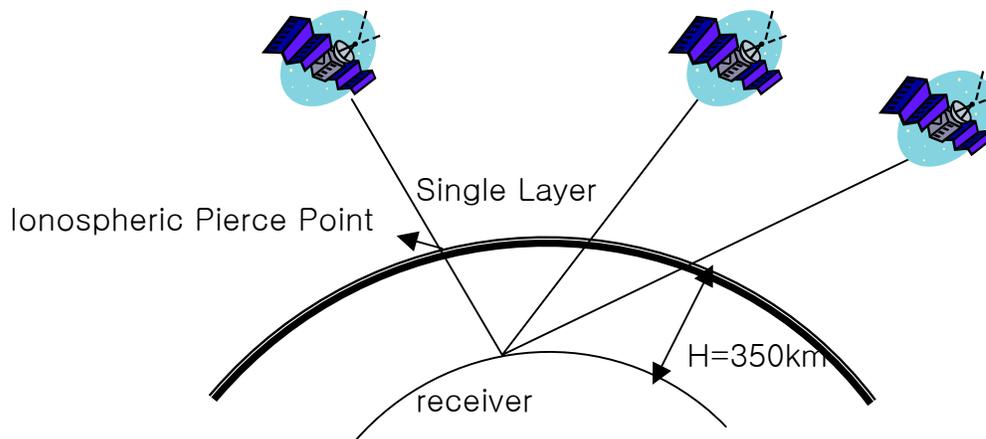
Figure 4.1 Single layer ionosphere model

The locations of VTEC values on the plane (IPPs), is calculated from the coordinates of the receivers and satellites-see Appendix B.

Using the VTEC values on the plane as raw observations, it is possible to develop two-dimensional ionosphere models. There are several algorithms available for ionosphere modeling. These can be divided into two groups: grid-based interpolation methods and mathematical fitting methods. In a grid-based interpolation algorithm, grid points of an appropriate spacing are designated throughout the region of interest and VTECs at the IGPs are interpolated from values at IPPs. For function based approaches, coefficients describing the spatial distribution of VTEC are derived, and VTEC at designated IGPs are extracted from the underlying function.

In this thesis, two algorithms are selected for evaluation and implementation: thin-plate smoothing spline and spherical harmonic fitting. These models are based on two-dimensional estimation on an ionospheric shell at 350 km altitude. The input observations are slant delays computed from dual frequency GPS observations. Corrections for both

satellite and receiver code interchannel biases are applied to the slant delay observations, and a mapping function (defined in (4-1)) is applied to derive VTEC measurements.

4.2 SPHERICAL HARMONICS MODEL

For function-based approaches, all network observations are included in estimates of the function coefficients. In regions of sparse observations the functional fit may be strongly influenced by observations outside the local region of interest. "It is possible that the local ionospheric variations will not be fully represented, or that higher-order variations will be assumed that do not actually exist" (Skone, 2002). Implementation of a functional approach can, however, be advantageous in networks with smoothly varying spatial distributions of TEC.

A function-based model using spherical harmonics is proposed here. This model is based on the Fourier series expansion in longitude and Legendre function in latitude as the orthogonal basis functions. The mathematical expression of vertical TEC using spherical harmonics is as follows (Schaer, 1999).

$$VTEC(\varphi, \lambda) = \sum_{n=0}^{n} \sum_{m=0}^{m} P_{nm}(\sin(\varphi))(C_{nm} \cos(m\zeta) + S_{nm} \sin(m\zeta)) \tag{4-2}$$

where

φ is the geomagnetic latitude of an IPP

ζ is the geomagnetic longitude of an IPP

n, m are integer degree and order of Legendre function, respectively

C_{nm}, S_{nm} are unknown spherical harmonic coefficients

P_{nm} are normalized associated Legendre functions

The radial term is omitted in the above equation since all TEC measurements are assumed to be in a fixed spherical thin shell. The degree and order of the Legendre function defines the resolution of the model. For sparsely distributed observations, it is necessary to employ a low-order fit.

The Legendre function formulated in equation (4-2) is appropriate for global representation of TEC. In regional TEC modeling, the conventional Legendre functions are not orthogonal over a regional area. A new technique, called Spherical Cap Harmonic Analysis (SCHA), was developed by Haines (1985) to model cap-like regional TEC data. In SCHA, Legendre functions with integer order but non-integer degree are used. These non-integer Legendre functions are orthogonal over a regional cap. Using SCHA, equation (4-2) can be rewritten as

$$VTEC(\varphi, \lambda) = \sum_{l=0}^{K} \sum_{m=0}^{l} P_{n_k(m)}^{m}(\sin \varphi)(C_{nm} \cos(m\zeta) + S_{nm} \sin(m\zeta)) \tag{4-3}$$

where n_l is non-integer degree.

When $l = K$, the number of coefficients is $(K+1)^2$. The non-integer Legendre functions can be derived as follows (Haines, 1985):

$$P_{n_l(m)}^{m}(\sin \varphi) = \sum_{j=0}^{\infty} A_j(m, n_l(m)) \sin^{2j}(\varphi/2) \tag{4-4}$$

where

$$A_j(m, n_l(m)) = \frac{(j+m-1)(j+m) - n_l(m)(n_l(m)+1)}{j(j+m)} A_{j-1}(m, n_l(m)) \tag{4-5}$$

$$A_0(m, n_l(m)) = K^m_{n_l(m)} \sin^m(\varphi) \tag{4-6}$$

and

$$K^m_{n_l(m)} = \frac{2^{-m}}{(m\pi)^{1/2}} \left(\frac{n_l(m)+m}{n_l(m)-m}\right)^{(n_l(m)/2+1/4)} p^{m/2} \exp(e_1 + e_2) \tag{4-7}$$

$$p = \left(\frac{n_l(m)}{m}\right)^2 - 1$$

$$e_1 = -\frac{1}{12m}\left(1+\frac{1}{p}\right), \quad e_2 = \frac{1}{360m^3}\left(1+\frac{3}{p^2}+\frac{4}{p^3}\right), \quad K^0_{n_l(m)} = 1 \tag{4-8}$$

The values of n_l for half angle $50º$ are summarized in Table 4.1 (Haines, 1989).

After calculation of non-integer Legendre functions, the remaining unknown coefficients in equation (4-3) are estimated using a Kalman filter approach.

Table 4.1 $n_l(m)$ **for Half angle 50º**

	M								
k	0	1	2	3	4	5	6	7	8
0	0.00								
1	2.24	1.78							
2	3.92	3.92	3.27						
3	5.82	5.66	5.50	4.71					
4	7.56	7.56	7.30	7.02	6.12				
5	9.41	9.31	9.21	8.88	8.52	7.51			
6	11.17	11.17	11.00	10.82	10.42	9.99	8.90		
7	13.01	12.94	12.86	12.63	12.40	11.94	11.44	10.27	
8	14.78	14.78	14.65	14.52	14.24	13.95	13.44	12.89	11.64

4.2.1 Kalman Filter Estimation

4.2.1.1 Measurement Model

The raw observations in the spherical harmonic functional fitting are the slant smoothed TECs derived from dual-frequency GPS data. The detailed derivation process of the smoothed TEC was addressed in Chapter 3. The measurement model relates the slant TECs to the unknown coefficients. In the case of $K=2$ in equation (4-2), the measurement model can be expressed as

$$STEC = M(E)VTEC + b_r + b_s \tag{4-9}$$

$$= M(E)\sum_{l=0}^{K}\sum_{m=0}^{l} P_{n_l(m)}^{m}(\sin\varphi)(C_{nm}\cos(m\zeta) + S_{nm}\sin(m\zeta)) + b_r + b_s$$

where

$M(E)$ is the vertical to slant mapping function

b_r, b_s are receiver and satellite code interchannel biases

The above equation can be written in a matrix form

$$\mathbf{z} = \mathbf{Hx} + \mathbf{v}$$ (4-10)

where

$$\mathbf{x} = \begin{bmatrix} b_{s1} \\ \vdots \\ b_{sNSAT} \\ b_r \\ a_1 \\ \vdots \\ a_9 \end{bmatrix}, \quad \mathbf{z} = (STEC)$$

$$\mathbf{H} = [1 \quad M(E)P_{n_0(0)}^0(\sin\varphi) \quad M(E)P_{n_1(0)}^0(\sin\varphi) \quad M(E)P_{n_1(1)}^1\sin(\varphi)\cos(\zeta) \quad M(E)P_{n_2(0)}^0(\sin\varphi)$$
$$M(E)P_{n_2(1)}^1(\sin\varphi)\cos(\zeta) \quad M(E)P_{n_2(2)}^2(\sin\varphi)\cos(2\zeta) \quad M(E)P_{n_1(1)}^1(\sin\varphi)\sin(\zeta)$$
$$M(E)P_{n_2(1)}^1(\sin\varphi)\sin(\zeta) \quad M(E)P_{n_2(2)}^2(\sin\varphi)\sin(2\zeta)]$$

$\mathbf{v}$ is measurement noise with the covariance matrix $\mathbf{R} = E(\mathbf{vv}^T)$

NSAT   is the number of satellites observed in one epoch

and $a_1 = C_{00}$, $a_2 = C_{10}$, $a_3 = C_{11}$, $a_4 = C_{20}$, $a_5 = C_{21}$, $a_6 = C_{22}$, $a_7 = S_{11}$, $a_8 = S_{21}$,

       $a_9 = S_{22}$

4.2.1.2 System Model

The unknown coefficients are assumed to vary stochastically in time and approximated with a first order Gauss-Markov process:

$$a_i(t_{k+1}) = e^{-\beta(\Delta t)} a_i(t_k) + w \tag{4-11}$$

where

a_i is the coefficient defined above ($i = 1\ldots9$)

k is time index

$\dfrac{1}{\beta}$ is correlation time

w is white noise

Δt $= t_{k+1} - t_k$

In equation (4-11), the temporal correlation of the coefficients is described with the normalized autocorrelation function $e^{-\beta(\Delta t)}$ and the remaining uncorrelated part is modeled by a white noise. The process noise of the white noise is given as

$$q(t_k) = \sigma^2 (1 - e^{-2\beta(\Delta t)}) \tag{4-12}$$

where σ^2 is spectral density divided by 2β.

In this research, the correlation time is set as 1/1800s and the values of σ^2 are set as follows:

a_1: $\sigma^2 = 1.0 \, \text{TECU}^2$

a_2, a_3, a_7: $\sigma^2 = 0.2 \, \text{TECU}^2$

$$a_4, a_5, a_6, a_8, a_9: \qquad \sigma^2 = 0.05 \text{TECU}^2 \tag{4-13}$$

4.2.1.3 Prediction and Observation Update

Based on the measurement and system model described above, the unknown parameters are predicted and updated in a Kalman filter procedure. Prediction is performed every 5 minutes for model coefficients, which are assumed to vary as first-order Gauss-Markov processes. The mathematical expression of parameter prediction is given as (Gelb, 1974).

$$\mathbf{x}^-(t_{k+1}) = \mathbf{\Phi}(t_k, t_{k+1}) \mathbf{x}^+(t_k) + \mathbf{w} \tag{4-14}$$

$$\mathbf{P}^-(t_{k+1}) = \mathbf{\Phi}(t_k, t_{k+1}) \mathbf{P}^+(t_k) \mathbf{\Phi}(t_k, t_{k+1})^{\mathrm{T}} + \mathbf{Q}(t_k) \tag{4-15}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{I}((\text{NSAT}+1)\times(\text{NSAT}+1)) & & & \mathbf{O} \\ & & e^{-\beta \, (\Delta t)} & \\ \mathbf{O} & & & \ddots \\ & & & & e^{-\beta \, (\Delta t)} \end{bmatrix} \tag{4-16}$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{O}((\text{NSAT}+1)\times(\text{NSAT}+1)) & & & \mathbf{O} \\ & & \sigma^2(1 - e^{-2\beta \, (\Delta t)}) & \\ \mathbf{O} & & & \ddots \\ & & & & \sigma^2(1 - e^{-2\beta \, (\Delta t)}) \end{bmatrix} \tag{4-17}$$

where

$-$	denotes prediction
$+$	denotes update at the previous epoch
$\mathbf{\Phi}$	is the transition matrix
$\mathbf{Q} = E[\mathbf{w}\mathbf{w}^\mathrm{T}]$	is the covariance matrix of the process noise

In the above equations, the satellite and receiver biases are assumed to remain constant with an identity transition matrix and zero process noise.

At each epoch, whenever new observations are available, the observation update is performed sequentially, assuming zero cross-correlation between each observation. The basic principle of the observation update is that the new estimates of the parameters are a linear combination of information from new observations and the most recent estimates of parameters. The weights (gains) attached to the new information and the most recent estimates are determined to minimize the covariance of the new estimates. The parameter update is given in a mathematical form as follows (Gelb, 1974):

$$\mathbf{x}^+(t_{k+1}) = \mathbf{x}^-(t_{k+1}) + \mathbf{K}[\mathbf{z}(t_{k+1}) - \mathbf{H}(t_{k+1})\mathbf{x}^-(t_{k+1})] \tag{4-18}$$

$$\mathbf{P}^+(t_{k+1}) = [\mathbf{I} - \mathbf{K}\mathbf{H}(t_{k+1})]\mathbf{P}^-(t_{k+1}) \tag{4-19}$$

where $\mathbf{K}$ is the gain matrix:

$$\mathbf{K} = \mathbf{P}^-(t_{k+1})\mathbf{H}^\mathrm{T}(t_{k+1})[\mathbf{H}(t_{k+1})\mathbf{P}^-(t_{k+1})\mathbf{H}^\mathrm{T}(t_{k+1}) + \mathbf{R}(t_{k+1})]^{-1} \tag{4-20}$$

4.2.2 U-D Filter

The time and observation update of the covariance matrices in equations (4-15) and (4-19) are susceptible to numerical errors. As a result, the Kalman filter may diverge and the covariance matrices may not keep the property of non-negative definiteness required. In order to reduce the numerical errors and instability associated with update of the covariance matrices, the U-D factorization algorithm was developed (Bierman, 1977). In this algorithm, the covariance matrix is factored into modified Cholesky factors.

$$\mathbf{P} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathbf{T}} \qquad (4\text{-}21)$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{U}$ is a upper triangular matrix with diagonal elements of 1.

The modified Cholesky factors are updated instead of the covariance matrix itself. Because the square-root factors are propagated, the numerical error propagated at each stage of the update is reduced in proportion to the square root.

4.2.2.1 Observation Update

The Bierman U-D observation update is used for scalar observation update. Equation (4-19) can be factored in terms of U-D factors, assuming scalar data and omitting time index:

$$\mathbf{P}^{-} = \mathbf{U}^{-}\mathbf{D}^{-}(\mathbf{U}^{-})^{\mathbf{T}}, \ \mathbf{P}^{+} = \mathbf{U}^{+}\mathbf{D}^{+}(\mathbf{U}^{+})^{\mathbf{T}} \qquad (4\text{-}22)$$

$$\mathbf{U}^{+}\mathbf{D}^{+}(\mathbf{U}^{+})^{\mathbf{T}} = \mathbf{U}^{-}\mathbf{D}^{-}(\mathbf{U}^{-})^{\mathbf{T}} - \frac{\mathbf{U}^{-}\mathbf{D}^{-}(\mathbf{U}^{-})^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}\mathbf{H}\mathbf{U}^{-}\mathbf{D}^{-}}{R + \mathbf{H}\mathbf{U}^{-}\mathbf{D}^{-}(\mathbf{U}^{-})^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}}(\mathbf{U}^{-})^{\mathbf{T}} \qquad (4\text{-}23)$$

$$= \mathbf{U}^{-}\mathbf{D}^{-}(\mathbf{U}^{-})^{\mathbf{T}} - \frac{\mathbf{U}^{-}\mathbf{v}\mathbf{v}^{\mathbf{T}}(\mathbf{U}^{-})^{\mathbf{T}}}{R + \mathbf{a}^{\mathbf{T}}\mathbf{D}^{-}\mathbf{a}} \qquad (4\text{-}24)$$

$$= \mathbf{U}^- [\mathbf{D}^- - \frac{\mathbf{v}\mathbf{v}^{\mathbf{T}}}{R + \mathbf{a}^{\mathbf{T}}\mathbf{D}^-\mathbf{a}}](\mathbf{U}^-)^{\mathbf{T}} \tag{4-25}$$

where

$$\mathbf{v} = \mathbf{D}^-(\mathbf{U}^-)^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}, \ \mathbf{a} = (\mathbf{U}^-)^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}$$

If the bracketed term in equation (4-25) is factored into $\overline{\mathbf{U}}\,\overline{\mathbf{D}}\,\overline{\mathbf{U}}^{\mathbf{T}}$, then equation (4-25) can be rewritten as

$$\mathbf{U}^+\mathbf{D}^+(\mathbf{U}^+)^{\mathbf{T}} = \mathbf{U}^-\overline{\mathbf{U}}\,\overline{\mathbf{D}}\,\overline{\mathbf{U}}^{\mathbf{T}}(\mathbf{U}^-)^{\mathbf{T}} \tag{4-26}$$

From equation (4-26) it follows that $\mathbf{U}^+ = \mathbf{U}^-\overline{\mathbf{U}}$, $\mathbf{D}^+ = \overline{\mathbf{D}}$. Therefore the observation update of U-D factors depends on the U-D factorization of the bracketed term in equation (4-25), that is

$$\mathbf{D}^- - \frac{\mathbf{v}\mathbf{v}^{\mathbf{T}}}{R + \mathbf{a}^{\mathbf{T}}\mathbf{D}^-\mathbf{a}} = \overline{\mathbf{U}}\,\overline{\mathbf{D}}\,\overline{\mathbf{U}}^{\mathbf{T}} \tag{4-27}$$

The above factorization is solved using Agee-Turner rank one modification algorithm, where $\overline{\mathbf{U}}$ and $\overline{\mathbf{D}}$ are derived recursively as

$$\overline{d}_j = d_j^- + c_j v_j^2 \ (j = n \text{ downward to } 1) \tag{4-28}$$

$$c_{j-1} = c_j d_j^- / \overline{d}_j \tag{4-29}$$

$$\overline{U}_{ij} = c_j v_j v_i / \overline{d}_j \ (i = 1 \text{ to } j\text{-}1) \tag{4-30}$$

where $d_j^-, \bar{d}_j$ are diagonal elements of $\mathbf{D}^-$ and $\bar{\mathbf{D}}$

$$c_n = -\frac{1}{R + \mathbf{a}^{\mathrm{T}} \mathbf{D}(-)\mathbf{a}}$$

Refining the above algorithm, Bierman (1977) devised a numerically efficient and stable recursive algorithm for calculating updated U-D factors and Kalman gain.

$$d_1^+ = d_1^- R / \alpha_1, \; \alpha_1 = R + v_1 a_1 \tag{4-31}$$

$$\alpha_j = \alpha_{j-1} + v_j a_j \; (j = 2 \text{ to } n) \tag{4-32}$$

$$d_j^+ = d_j^- \alpha_{j-1} / \alpha_j \tag{4-33}$$

$$\mathbf{u}_j^+ = \mathbf{u}_j^- + \lambda_j \mathbf{k}_j, \; \lambda_j = -a_j / \alpha_{j-1} \tag{4-34}$$

$$\mathbf{k}_{j+1} = \mathbf{k}_j + v_j \mathbf{u}_j^- \tag{4-35}$$

where

d_j^+ are diagonal elements of $\mathbf{D}^+$

$\mathbf{u}_j^+$ and $\mathbf{u}_j^-$ are jth columns of $\mathbf{U}^+$ and $\mathbf{U}^-$

$\mathbf{a} = [a_1 \cdots a_n]^{\mathrm{T}}$, $\mathbf{v} = [v_1 \cdots v_n]^{\mathrm{T}}$

$\mathbf{k}_{j+1} = \mathbf{U}^- \mathbf{v}^{(j)}$, $\mathbf{v}^{(j)} = [v_1 \cdots v_j \, 0 \cdots 0]^{\mathrm{T}}$

The Kalman gain is given by

$$\mathbf{K} = \mathbf{k}_{n+1} / \alpha_n \tag{4-36}$$

The original proof of the above algorithm is shown in Bierman (1977) and alternative derivation can be found in Kailath (2001).

4.2.2.2 Time Update

The covariance prediction equation (4-15) can be factored into U-D factors in a similar way.

$$\mathbf{U}_{k+1}^{-}\mathbf{D}_{k+1}^{-}(\mathbf{U}_{k+1}^{-})^{\mathbf{T}} = \mathbf{\Phi}_k\mathbf{U}_k^{+}\mathbf{D}_k^{+}(\mathbf{U}_k^{+})^{\mathbf{T}}\mathbf{\Phi}_k + \mathbf{G}_k\mathbf{D}_{Q_k}\mathbf{G}_k \tag{4-37}$$

$$= \begin{bmatrix} \mathbf{\Phi}_k\mathbf{U}_k^{+} & \mathbf{G}_k \end{bmatrix}\begin{bmatrix} \mathbf{D}_k^{+} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{Q_k} \end{bmatrix}\begin{bmatrix} (\mathbf{U}_k^{+})^{\mathbf{T}}(\mathbf{\Phi}_k)^{\mathbf{T}} \\ (\mathbf{G}_k)^{\mathbf{T}} \end{bmatrix} \tag{4-38}$$

$$= \mathbf{W}^{\mathbf{T}}\mathbf{D}_W\mathbf{W} \tag{4-39}$$

where

$\mathbf{D}_{Q_k}$ is a diagonal matrix satisfying $\mathbf{Q} = \mathbf{G}_k\mathbf{D}_{Q_k}\mathbf{G}_k$

$\mathbf{W}^{\mathbf{T}} = [\mathbf{\Phi}_k\mathbf{U}_k^{+} \quad \mathbf{G}_k]$

$\mathbf{D}_W = \begin{bmatrix} \mathbf{D}_k^{+} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{Q_k} \end{bmatrix}$

If the matrix $\mathbf{W}$ in equation (4-39) can be factored into $\mathbf{B}\hat{\mathbf{U}}^{\mathbf{T}}$ where the column vectors $\mathbf{b}_1\cdots\mathbf{b}_n$ of $\mathbf{B}$ are mutually orthogonal with respect to the weighting matrix $\mathbf{D}_W$:

$$\mathbf{B}^{\mathbf{T}}\mathbf{D}_W\mathbf{B} = \hat{\mathbf{D}} \tag{4-40}$$

then, equation (4-39) can be rewritten as

$$\mathbf{U}_{k+1}^{-}\mathbf{D}_{k+1}^{-}(\mathbf{U}_{k+1}^{-})^{\mathbf{T}} = \hat{\mathbf{U}}\mathbf{B}^{\mathbf{T}}\mathbf{D}_W\mathbf{B}\hat{\mathbf{U}}^{\mathbf{T}} = \hat{\mathbf{U}}\hat{\mathbf{D}}(\hat{\mathbf{U}})^{\mathbf{T}} \tag{4-41}$$

From equation (4-41), it follows that

$$\mathbf{U}_{k+1}^{-} = \hat{\mathbf{U}}, \ \mathbf{D}_{k+1}^{-} = \hat{\mathbf{D}} \tag{4-42}$$

Therefore, the time update of the U-D factors depends on the factorization of $\mathbf{W} = \mathbf{B}\hat{\mathbf{U}}^{\mathrm{T}}$ satisfying equation (4-40). This can be solved using weighted Gram-Schmidt orthogonalization (Thornton, 1976). Given linearly independent n vectors $\mathbf{w}_1, \cdots \mathbf{w}_n$ and a diagonal matrix $\mathbf{D}_W$, the weighted Gram-Schmidt algorithm generates n mutually orthogonal vectors $\mathbf{b}_1 \cdots \mathbf{b}_n$ defined by

$$\mathbf{b}_n = \mathbf{w}_n \tag{4-43}$$

$$\mathbf{b}_j = \mathbf{w}_j - \sum_{k=j+1}^{n} (\mathbf{w}_j^{\mathrm{T}} \mathbf{D}_W \mathbf{b}_k / \mathbf{b}_k^{\mathrm{T}} \mathbf{D}_W \mathbf{b}_k) \mathbf{b}_k \quad (j = n-1 \ldots 1) \tag{4-44}$$

Defining $\hat{\mathbf{U}}(k,j) = \mathbf{w}_k^{\mathrm{T}} \mathbf{D}_W \mathbf{b}_j / \hat{d}_j$, $\hat{d}_j = \mathbf{b}_j^{\mathrm{T}} \mathbf{D}_W \mathbf{b}_j$, equation (4-44) is rewritten as

$$\mathbf{w}_j = \mathbf{b}_j + \sum_{k=j+1}^{n} \hat{\mathbf{U}}(j,k) \mathbf{b}_k \tag{4-45}$$

Because equation (4-45) holds for $j = 1 \ldots n-1$, equation (4-45) is expressed in a matrix form

$$\mathbf{W} = \mathbf{B}\hat{\mathbf{U}}^{\mathrm{T}} \tag{4-46}$$

where $\hat{\mathbf{U}}(j,j)$ for $j = 1\ldots n$.

Since the conventional Gram-Schmidt algorithm is known to be numerically unstable, a modified Gram-Schmidt algorithm (Bjorck, 1967) is used in the real calculation of orthogonal vectors in equation (4-44). A recursive algorithm using modified Gram-Schmidt orthogonalization is presented as follows (Bierman, 1977).

$$\mathbf{b}_l^{(n)} = \mathbf{w}_l \quad (l = 1\ldots n) \tag{4-47}$$

For $j = n\text{-}1$ downward to 1

$$\hat{d}_{j+1} = (\mathbf{b}_{j+1}^{(j+1)})^{\mathbf{T}} \mathbf{D}_W \mathbf{b}_{j+1}^{(j+1)} \tag{4-48}$$

$$\hat{\mathbf{U}}(k, j+1) = (\mathbf{b}_k^{j+1})^{\mathbf{T}} \mathbf{D}_W \mathbf{b}_{j+1}^{(j+1)} / \hat{d}_{j+1} \quad (k = 1\ldots j) \tag{4-49}$$

$$\mathbf{b}_k^{(j)} = \mathbf{b}_k^{(j+1)} - \hat{\mathbf{U}}(k, j+1)\mathbf{b}_{j+1}^{(j+1)} \quad (k = 1\ldots j) \tag{4-50}$$

$$\mathbf{b}_j^{(j)} = \mathbf{b}_j \tag{4-51}$$

and $\hat{d}_1 = (\mathbf{v}_1^{(1)})^{\mathbf{T}} \mathbf{D}_W \mathbf{v}_1^{(1)}$ \tag{4-52}

4.3 THIN PLATE SPLINE INTERPOLATION

4.3.1 General Property of Thin Plate Spline

The cubic spline is based on fitting a cubic polynomial that is defined piecewise. At the intersection points between two piecewise functions the condition of continuity of both first and second derivatives of each function must be satisfied. This results in smooth and continuous local functions with no discontinuities or sudden changes throughout all pieces of the curve. The boundary conditions ensure that local functions are influenced

nonlocally, and this is an advantage in regions of sparse observations. The nonlocality guarantees global smoothness of the interpolated function.

Thin plate spline is a two dimensional extension of the cubic spline in one dimension. It finds a minimally bended surface that passes through all scattered data. In other words, thin plate splines minimize the surface curvature and can be thought of as a thin metal plate that is constrained not to move at data points. The curvature of a surface is defined for d dimensional vector $\mathbf{x}$ as

$$J_m(f) = \iint_{\Re^d} \sum \frac{m!}{\alpha_1! \cdots \alpha_d!} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right)^2 d\mathbf{x} \tag{4-53}$$

where α is all non-negative integers satisfying $\sum_{j=1}^{d} \alpha_j = m$ and f has m-1 continuous derivatives and a square integrable m^{th} derivative.

Duncan (1977) and Meinguet (1979) proved that if f is a smooth function, the minimization of equation (4-53) allows the following form

$$f(\mathbf{x}_j) = \sum_{i=1}^{p} \phi_i(\mathbf{x}_j) a_i + \sum_{l=1}^{n} E\left(\left\| \mathbf{x}_j - \mathbf{x}_l \right\|\right) b_l \tag{4-54}$$

with the condition $\sum_{j=1}^{n} \phi(\mathbf{x}_j) b_j = 0$ \hfill (4-55)

where

$\phi_i(\mathbf{x}_j)$ \qquad are p polynomial functions (of order $\leq m$-1)

$E(\|\mathbf{x}_j - \mathbf{x}_l\|)$ are n radial basis functions; $\|\cdot\|$ means the length of a vector

$a_i,\ b_l$ are coefficients of polynomial and radial basis functions

The radial functions mentioned above are defined as

$$E(r) = \begin{cases} a_{md}\, r^{(2m-d)} \log(r) & 2m-d \ \ \text{even} \\ a_{md}\, r^{(2m-d)} & 2m-d \ \ \text{odd} \end{cases}$$

where

$$a_{md} = \begin{cases} \dfrac{(-1)^{d/2+1+m}}{2^{2m-1} \pi^{d/2} (m-1)!(m-d/2)!} & 2m-d \ \ \text{even} \\[2ex] \dfrac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} & 2m-d \ \ \text{odd} \end{cases}$$

For $m=2$ and $d=2$, Equations (4-53) and (4-54) can be rewritten:

$$J_2(f) = \int_{\Re^2} [(\frac{\partial^2 f}{\partial x_1^2})^2 + 2(\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2})^2 + (\frac{\partial^2 f}{\partial x_2^2})^2] d\mathbf{x} \qquad (4\text{-}56)$$

$$f(\mathbf{x}_j) = a_1 + a_2(\mathbf{x}_j)_1 + a_3(\mathbf{x}_j)_2 + \sum_{l=1}^{n} E(\|\mathbf{x}_j - \mathbf{x}_l\|)b_l \qquad (4\text{-}57)$$

where

$$E(r) = a_{md} \log(r) r^2$$

4.3.2 Derivation of Thin Plate Spline Surface

Given a set of n distinct nodes $\mathbf{x}_j$ ($j = 1$ to n) and n scattered data points y_j, the thin plate spline surface $f(\mathbf{x}_j)$ (equation (4-57)) satisfies the interpolation condition:

$$y_j = \sum_{i=1}^{3} (\mathbf{x}_j)_i a_i + \sum_{l=1}^{n} E(\|\mathbf{x}_j - \mathbf{x}_l\|) b_l \tag{4-58}$$

where $(\mathbf{x}_j)_1 = 1$, $(\mathbf{x}_j)_2 = x_j$, and $(\mathbf{x}_j)_3 = y_j$

Equation (4-58) and the condition equation (4-55) can be written in a matrix form

$$\mathbf{y} = \mathbf{A}\mathbf{a} + \mathbf{B}\mathbf{b} \tag{4-59}$$

$$\mathbf{A}^T \mathbf{b} = \mathbf{0} \tag{4-60}$$

where

$$\mathbf{y} = [y_1 \cdots y_n]^T, \quad \mathbf{a} = [a_1 \; a_2 \; a_3]^T, \quad \mathbf{b} = [b_1 \cdots b_n]^T$$

$$\mathbf{A}_{n\times3} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix}, \quad \mathbf{B}_{n\times n} = E_{ij} = E(\|\mathbf{x}_i - \mathbf{x}_j\|)$$

The coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ are obtained by solving equations (4-59) and (4-60). A QR factorization can be used to enforce the $\mathbf{A}^T \mathbf{b} = \mathbf{0}$ condition. $\mathbf{A}$ is factored into

$$\mathbf{A} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \tag{4-61}$$

where $\mathbf{Q}_{n \times n}$ is an orthogonal matrix and $\mathbf{R}_{n \times 3}$ is an upper triangular matrix. There are several numerically stable algorithms for QR factorization including Housholder transformation and Givens rotation (Golub and Van Loan, 1996). If $\mathbf{Q}$ is partitioned into two parts

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \tag{4-62}$$

where $\mathbf{Q}_1$ is a $n \times 3$ matrix and $\mathbf{Q}_2$ is a $n \times (n-3)$ matrix, then it follows

$$\begin{bmatrix} \mathbf{Q}_1^{\mathrm{T}} \\ \mathbf{Q}_2^{\mathrm{T}} \end{bmatrix} \mathbf{A} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \tag{4-63}$$

From equation (4-63), it is evident that $\mathbf{Q}_1^{\mathrm{T}}\mathbf{A} = \mathbf{R}$ and $\mathbf{Q}_2^{\mathrm{T}}\mathbf{A} = \mathbf{0}$, which means that $\mathbf{Q}_1$ spans the column space of $\mathbf{A}$ and $\mathbf{Q}_2$ is orthogonal to the columns of $\mathbf{A}$. Therefore the condition $\mathbf{A}^{\mathrm{T}}\mathbf{b} = \mathbf{0}$ is enforced by letting $\mathbf{b} = \mathbf{Q}_2\mathbf{m}$, where $\mathbf{m}$ is a vector with the same dimension as $\mathbf{b}$. By inserting $\mathbf{b} = \mathbf{Q}_2\mathbf{m}$ into equation (4-59), it follows

$$\mathbf{y} = \mathbf{A}\mathbf{a} + \mathbf{B}\mathbf{Q}_2\mathbf{m} \tag{4-64}$$

By multiplying both sides of equation (4-64) with $\mathbf{Q}_2^{\mathrm{T}}$, $\mathbf{m}$ is derived as

$$\mathbf{m} = (\mathbf{Q}_2^{\mathrm{T}}\mathbf{B}\mathbf{Q}_2)^{-1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{y} \tag{4-65}$$

The coefficient vector $\mathbf{b}$ is derived from equation (4-65)

$$\mathbf{b} = \mathbf{Q}_2\mathbf{m} = \mathbf{Q}_2(\mathbf{Q}_2^\mathrm{T}\mathbf{BQ}_2)^{-1}\mathbf{Q}_2^\mathrm{T}\mathbf{y} \qquad (4\text{-}66)$$

Inserting equation (4-66) into equation (4-59),

$$\mathbf{Aa} = (\mathbf{y} - \mathbf{BQ}_2(\mathbf{Q}_2^\mathrm{T}\mathbf{BQ}_2)^{-1}\mathbf{Q}^\mathrm{T}\mathbf{y}) \qquad (4\text{-}67)$$

and using $\mathbf{Q}_1^\mathrm{T}\mathbf{A} = \mathbf{R}$, the coefficient vector $\mathbf{a}$ is derived as

$$\mathbf{a} = \mathbf{R}^{-1}(\mathbf{Q}_1^\mathrm{T})^{-1}(\mathbf{y} - \mathbf{BQ}_2(\mathbf{Q}_2^\mathrm{T}\mathbf{BQ}_2)^{-1}\mathbf{Q}^\mathrm{T}\mathbf{y}) \qquad (4\text{-}68)$$

CHAPTER 5

RESULTS AND ANALYSIS

In this chapter, the two algorithms introduced in Chapter 4 are evaluated and compared in two different GPS networks: CACS and WCDA. The CACS is a sparse network that covers the whole of Canada and WCDA is a medium-size regional GPS network that is distributed in the western part of British Columbia. The purpose of selecting those two networks is to evaluate the performance of the two algorithms in networks of different dimensions-for analysis of both wide area national and regional models. The algorithms are also evaluated for a range of ionospheric activity.

5.1 DATA DESCRIPTION

CACS is a Canada-wide GPS reference network developed and operated by NRCan. The objectives of CACS are to provide the Canadian spatial reference system and improve the precision of various GPS applications in Canada (Erickson *et al.*, 1995). Due to the size of Canada, the CACS is inherently a data sparse network. WCDA was established by NRCan for the purpose of crustal deformation monitoring. Since the first station DRAO was installed in 1990, 11 additional stations have been constructed in southwestern British Columbia. In this thesis, 12 stations of CACS and 8 stations of WCDA are selected for data processing. The locations of the stations that constitute the two GPS networks are illustrated in Figures 5.1 and 5.2.

The stations that make up the two networks are equipped with dual frequency geodetic quality GPS receivers. The types of receivers/antennas are summarized in Appendix A. The GPS data collected at each station are archived daily in a RINEX observation file format (Gurtner *et al.*, 1990) and transferred to the International GPS Service for Geodynamics (IGS). This data is freely available for scientific research puposes. A

RINEX GPS observation file contains dual frequency pseudorange and phase data, with a sampling interval of 30s. RINEX ephemeris files are also available from the IGS to compute the positions of the GPS satellites.

In the analyses presented here, 4 days of GPS data from May 28 to May 31 of 2003 are used. The main reason for selecting this time period is to assess the performance of the two algorithms in disturbed ionospheric conditions. Figure 5.3 shows the Kp index plot for 3 days from May 28 to May 30, 2003. This index is an indicator of the geomagnetic disturbances and is obtained by averaging the disturbances in the magnetic field at 13 stations in the sub-auroral region (Mayaud, 1980). The Kp values below 4 indicate low ionospheric condition, while Kp values in the range of 5-7 indicate moderate-major ionospheric activity, and Kp values greater than 7 indicate intense ionospheric activity. In Figure 5.3, the maximum Kp value of 8 was reached at the end of May 29 and early hours of May 30. This means that a significant ionospheric disturbance occurred during that time.
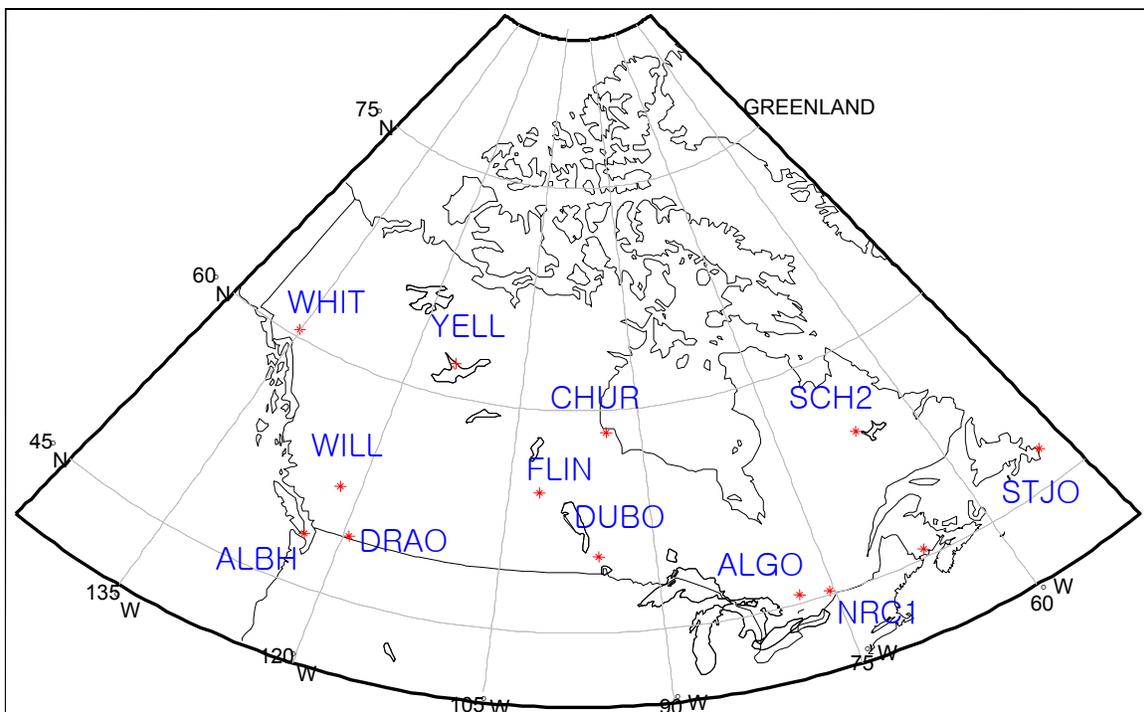
Figure 5.1 CACS network
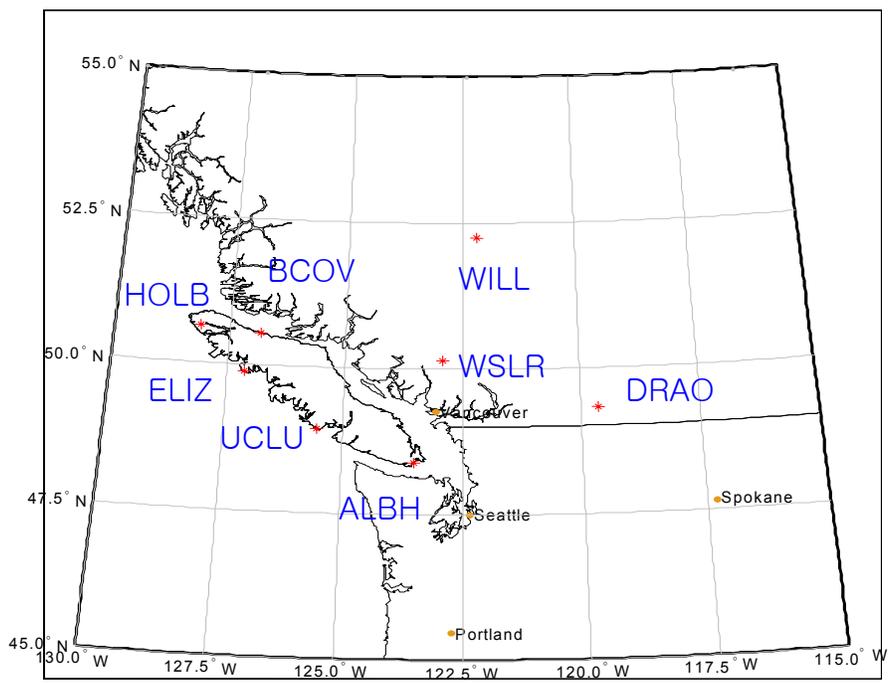
Figure 5.2 WDCA network.
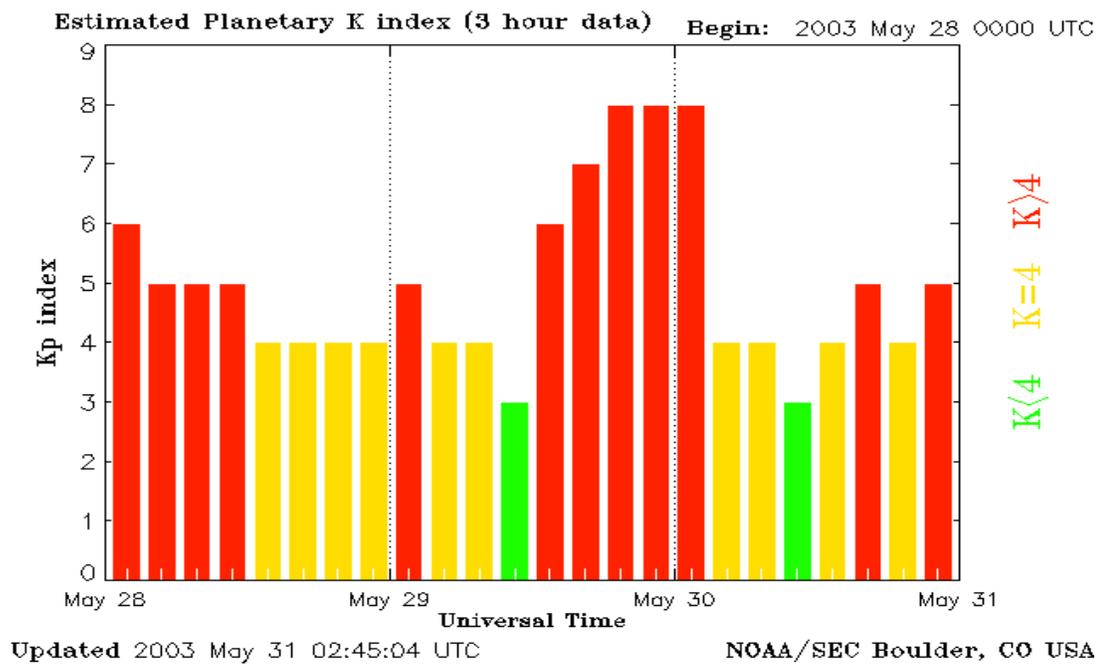
Figure 5.3 Kp index for May 28-30, 2003.

This storm event was discussed in Skone *et al.* (2003). In particular, the presence of a feature known as storm enhanced density (SED) was identified during the late hours (UT) of May 29, 2003. SED can be described as a narrow plume of greatly enhanced TEC values (~100 TEC units), with very large associated gradients. A typical plume extends from the New England coast across the Great Lake region and into the central Canada. This localized effect can be very difficult to resolve using ionosphere models for sparse networks. Figure 5.4 shows a VTEC map over Canada at one epoch (20:50 UT, May 29, 2003) during the storm event.
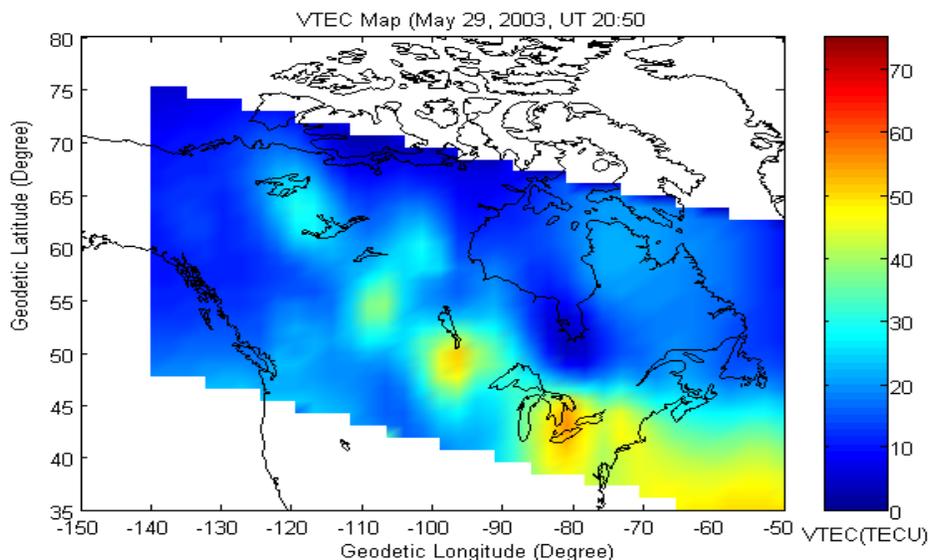
Figure 5.4 VTEC Map at 20:50 UT on May 29, 2003.

Figure 5.5 show time series of VTECs above station ALGO (CACS network). Diurnal variations in VTECs are evident, with increased values on May 29 (Day 149).
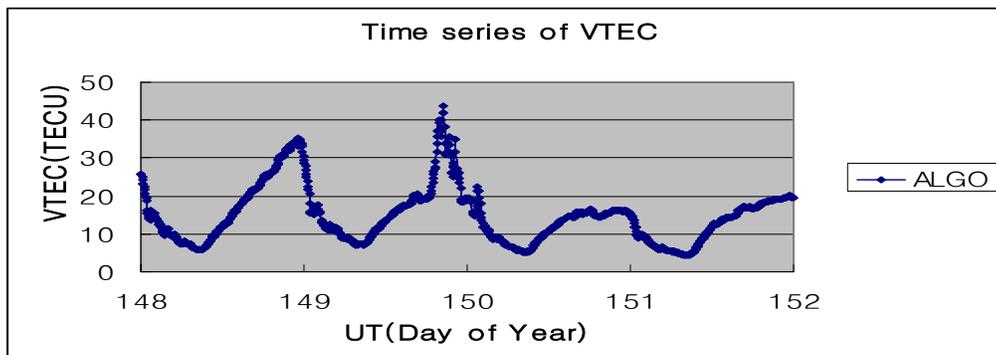
Figure 5.5 Time series of VTEC at station ALGO (where May 28 is Day 148)

At the later hours of May 29, much enhanced VTECs are shown and large temporal gradients of VTECs are also observed. The large spatial and temporal gradients observed during a storm event make accurate ionosphere modeling difficult.

5.2 METHODOLOGY

5.2.1 Data Preprocessing

Dual frequency GPS data are processed with **PREPROC** software developed at the University of Calgary to generate time series of the slant TEC. At each epoch, **PREPROC** software reads an observation file in RINEX format that contains dual frequency phase and code observables. After quality control that includes detection of cycle slips and estimation of multipath statistics using algorithms addressed in chapter 3, slant TECs are calculated both from phase and code observables. The phase-derived relative TECs are used to smooth the code-derived absolute TECs using methods described in Section 3.3. The smoothed slant TECs and other parameters including satellite elevation angles and the coordinates of IPPs, which are calculated using information contained in a RINEX navigation file, are saved in a binary output file. This process is repeated for all specified epochs and stations in the network. The binary output files created in this way are used as raw observation files for post-processing modeling using spherical harmonics model and thin plate spline interpolation.

5.2.2 Spherical Harmonics Model

The slant TECs derived from preprocessing are used as raw observations for spherical harmonics model. The coefficients that describe the spherical harmonics expansion are estimated every 5 minutes in a real time mode using a Kalman filter. Interchannel satellite and receiver biases can be estimated or read from an external file that contains bias information. In this research, satellite biases are fixed to the values obtained from the Center for Orbit Determination in Europe (CODE) and receiver biases at the individual stations are estimated in the adjustment along with the spherical harmonics coefficients. In this case, the receiver bias of one station is fixed to a predetermined value and the

receiver biases of the remaining stations are estimated relative to the fixed one. Figure 5.6 illustrates the processing algorithms in detail.

```
┌─────────────────────────────────────────┐
│      Read satellite biases from a file   │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  Initialize parameters (coefficients,    │
│  receiver biases) and covariance matrix  │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│               New epoch                  │ ◄────────┐
└─────────────────────────────────────────┘          │
                    │                                 │
                    ▼                                 │
┌─────────────────────────────────────────┐          │
│              New station                 │ ◄──────┐ │
└─────────────────────────────────────────┘        │ │
                    │                               │ │
                    ▼                               │ │
┌─────────────────────────────────────────┐        │ │
│  Read slant TEC, satellite elevation     │        │ │
│  angle, coordinates of IPP points        │        │ │
└─────────────────────────────────────────┘        │ │
                    │                               │ │
                    ▼                               │ │
┌─────────────────────────────────────────┐        │ │
│  Sequential observation update of        │        │ │
│  coefficients and receiver bias          │        │ │
└─────────────────────────────────────────┘        │ │
                    │                               │ │
                    ▼                               │ │
┌─────────────────────────────────────────┐   N    │ │
│   Finished processing of all stations?   │ ───────┘ │
└─────────────────────────────────────────┘          │
                 Y  │                                 │
                    ▼                                 │
┌─────────────────────────────────────────┐          │
│       Time update of coefficients        │          │
└─────────────────────────────────────────┘          │
                    │                                 │
                    ▼                                 │
┌─────────────────────────────────────────┐   N      │
│    Finished processing of all epochs?    │ ─────────┘
└─────────────────────────────────────────┘
                 Y  │
                    ▼
┌─────────────────────────────────────────┐
│                  Exit                    │
└─────────────────────────────────────────┘
```

Figure 5.6 Spherical harmonics model processing algorithm.

5.2.3 Thin Plate Spline Interpolation

Thin plate spline generates VTECs at IGPs using the VTEC measurements at IPPs. The raw data obtained from the observation file created by **PREPROC** software are slant TEC observations. Hence the slant TECs are converted to VTECs using equation (4-3). The satellite and receiver biases needed for the conversion are obtained from the estimated values in the spherical harmonics model. The VTECs at IPPs are substituted in equation (4-58) as raw observations. The coefficients that describes the thin plate spline surface are derived solving equations (4-59) and (4-60). The solution can be found in equations (4-66) and (4-68). Using the estimated coefficients, the VTECs at IGPs are generated. The information of VTECs at IGPs can be provided to the interested GPS users through a Wide Area DGPS service.

Given the above procedure for generating VTECs at IGPs, it is evident that the thin plate spline is a global interpolation scheme. The thin plane spline is particularly efficient for interpolating between scattered measurements, which is the case for the sparse CACS network. The resolution of the grid is limited by the size of the network and station spacing. For example, in case of the WAAS (FAA, 1999), the grid resolution is set to 5 degrees in latitude and longitude between 55N and 55S latitude, and 10 degrees between 55 and 75 degrees latitude. In this research, the spacing of CACS network is set to 5 degrees, while that for the WCDA is set to 2 degrees.

5.3 PERFORMANCE ANALYSIS OF THE ALGORITHMS

In this section, the performance of the two algorithms is evaluated in CACS and WCDA networks. Accuracies are measured by computing the residuals-the differences between the VTECs predicted by the model at observation IPPs and the actual VTEC observations at the IPPs. In the case of the spherical harmonics model, the VTEC predictions at IPPs are from estimated coefficients. In the grid-based thin plate spline algorithm, predicted VTEC values at the IPPs are interpolated from the VTEC values at IGPs surrounding the

IPPs. In this analysis, bilinear interpolation is used because it is a basic algorithm that can be easily implemented in user GPS receivers. The residuals are calculated by subtracting the estimated VTECs from the observed ones at the IPPs.

5.3.1 CACS Network Results

As mentioned in Section 5.1, the CACS is a data sparse national GPS network. Here the performance of the two algorithms is evaluated in the CACS network and the impact of the storm event on the algorithms is investigated.

5.3.1.1 Derivation of Optimal Index for Spherical Harmonics Fit

Prior to presenting the results of the two algorithms in the CACS network, the K index of SCHA, referred to as Haines's index, is determined. As mentioned in Chapter 4, the number of spherical harmonics coefficients is the square of the K index, which means that there would be 9 and 16 coefficients in the case that K is 2 and 3. Estimated coefficients are substituted in equation 4.3 to calculate the residuals as described above. The rms and mean values of the residuals are derived from

$$\text{rms} = \sqrt{\frac{\sum_{i=1}^{N}(residual)^2}{N}} \qquad (5\text{-}1)$$

$$\text{mean} = \frac{\sum_{i=1}^{N}residual_i}{N} \qquad (5\text{-}2)$$

where N is the number of observations and the elevation cutoff angle for residuals is set to 20°.

Table 5.1 shows the comparison results for different K values. The mean and rms values were derived using all residuals between VTEC observations and predicted VTEC values calculated at 12 CACS stations in Figure 5.1.

Table 5.1 Comparison of Spherical Harmonic *K* index

Date	Mean (TECU)		RMS (TECU)	
	K=2	*K*=3	*K*=2	*K*=3
May 28	-0.063	-0.546	2.986	2.991
May 29	-0.100	-0.064	3.547	3.529
May 30	-0.103	-0.118	2.657	2.714
May 31	-0.047	-0.026	2.437	2.504

The case where *K* is 2 shows slightly better results in rms and mean values than results where *K* is 3. This shows that higher order and degree expansion than 2 does not improve the performance of the algorithm in the sparse CACS network. For this reason, only the case of *K*=2 is considered for further processing to derive CACS network results.

5.3.1.2 Overall Results

Figure 5.7 shows the distribution of all IPPs observed at 12 CACS stations for a 24-hour period (May 28, 2003). The slant TEC measurements at these points are used as raw data for estimation of spherical harmonics model and the thin plate spline interpolation. The goodness of the fit and interpolation may be inferred through analysis of the residuals. Figures 5.8-5.11 show time series of the rms of the residuals between VTEC observations and predicted VTEC values for all 12 CACS stations for both spherical harmonics model and the thin plate spline. The rms values were generated every 5 minutes in a real-time mode. Irrespective of the algorithm used, all stations show maximum rms values at the end of May 29 and early hours of May 30 when the Kp index reached the maximum value of 8. The large rms values during that time period are due to the large ionospheric gradients caused by a geomagnetic storm. It is also notable in the figures that there exist time delays in the peak rms values between stations. This time shift will be discussed in a later section (Section 5.5) with a focus on the temporal variations in the VTEC maps over Canada.

As shown in Figures 5.8-5.11, thin plate spline shows much smaller rms values for all stations. The overall rms value of the thin plate spline was around 1.5 TECU on a highly disturbed day (May 29, 2003) and 0.8 TECU on a relatively quiet day (May 31, 2003). In contrast, the mean rms values of the spherical harmonics model were 3.5 TECU on May 29, 2003 and 2.4 TECU on May 31, 2003. Because 1 TECU amounts to 0.162m on L1, this means that the overall rms error of thin plate spline is from 0.13m to 0.24 m and that of spherical harmonics model is from 0.39m to 0.57m. Overall, the performance of the thin plate spline interpolation is better than that of the spherical harmonics model. Figures 5.12 and 5.13 show histograms with the distribution in residuals for May 29 and May 31, 2003. These days are selected because they represent the most disturbed (May 29) and quiet days (May 31) of the 4 days processed. The Kp indices for May 31, which are omitted in Figure 5.3, are 5 (00-03 hours, UT), 5 (03-06 hours, UT), 3 (06-09 hours, UT), 1 (09-12 hours, UT), 3 (12-15 hours, UT), 3 (15-18 hours, UT), 2 (18-21 hours, UT) and 2 (21-24 hours, UT). Table 5.2 summarizes the mean and rms values of the residuals at 12 stations for May 29 and May 31, 2003. The relatively larger rms values of stations ALGO, DUBO and NRC1 on May 29, 2003 are due to the very large increase in TEC that occurred over these stations, associated with the geomagnetic storm development at the end of May 29. The details will also be discussed in Section 5.5.
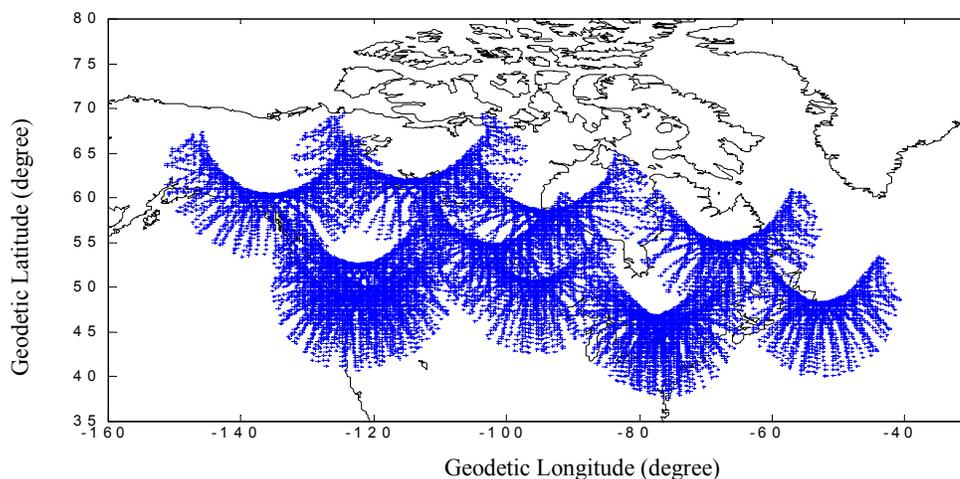
Figure 5.7 Distribution of IPPs in CACS network.
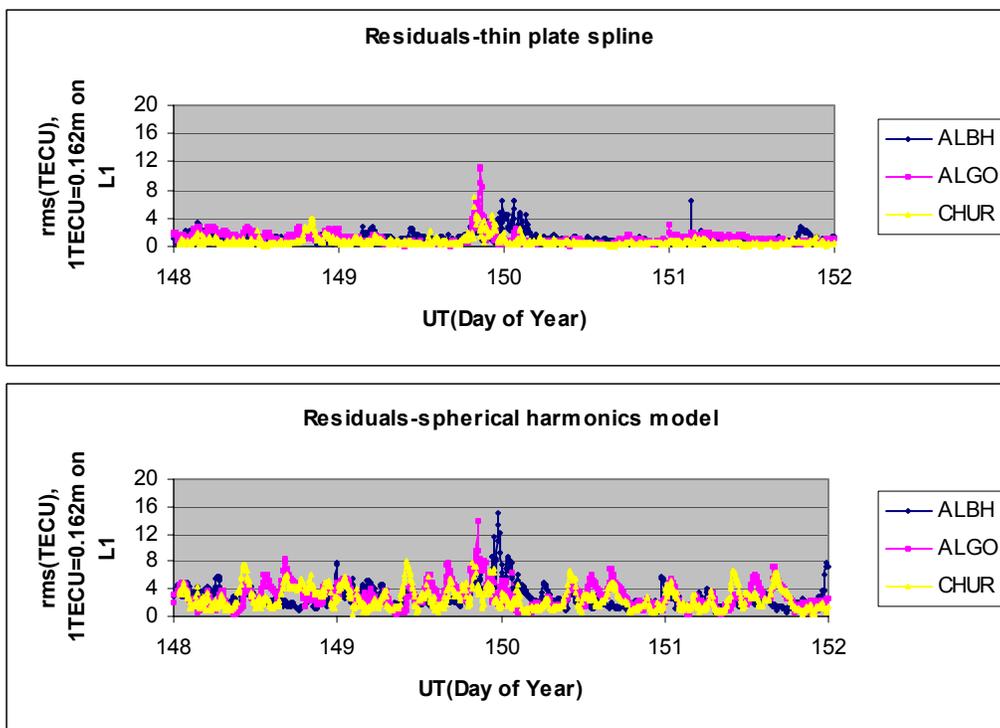
Figure 5.8 Time series of rms of residuals for stations ALBH, ALGO, and CHUR for May 28-31, 2003 (where May 28 is Day 148).
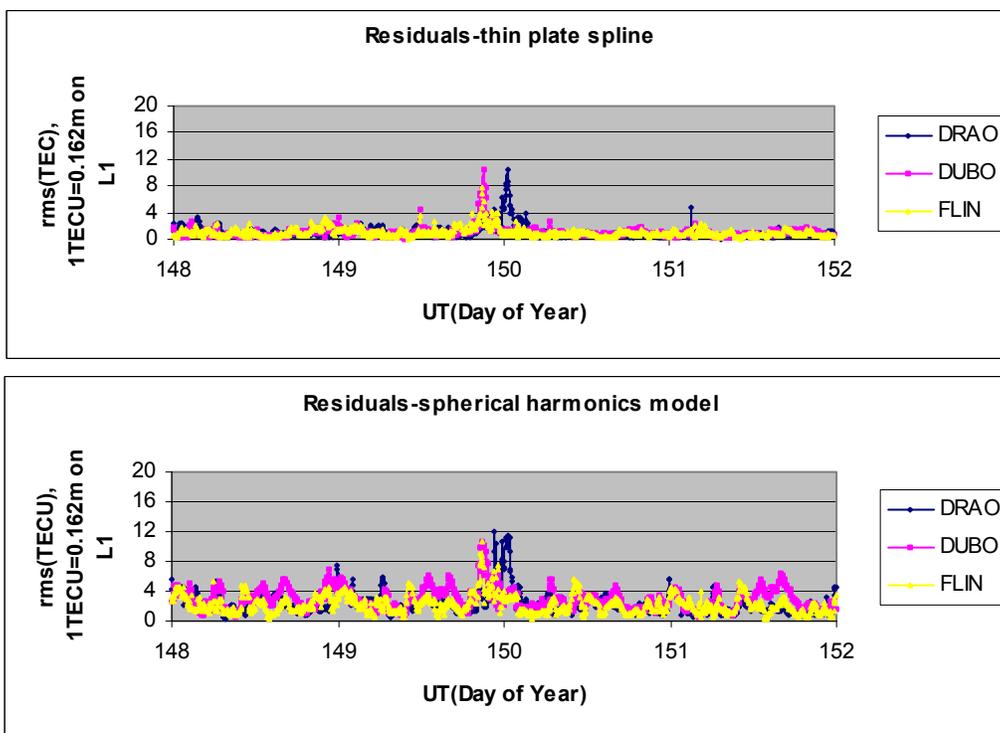
Figure 5.9 Time series of rms of residuals for stations DRAO, DUBO, and FLIN for May 28-31, 2003 (where May 28 is Day 148).
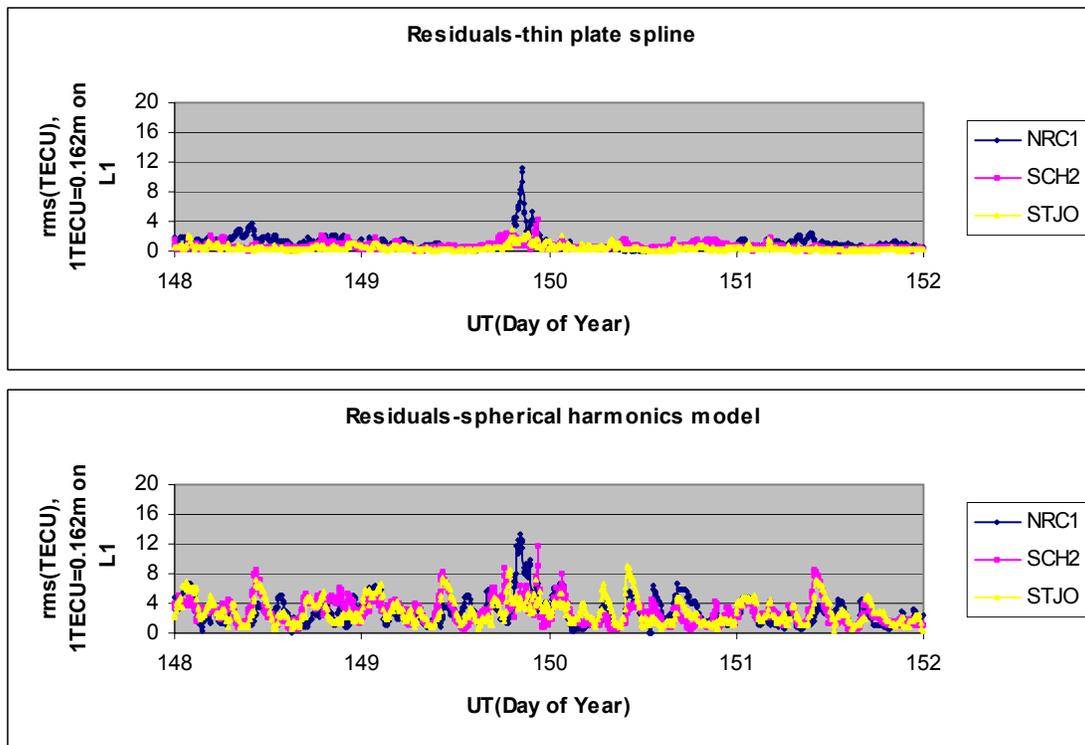
Figure 5.10 Time series of rms of residuals for stations NRC1, SCH2, and STJO for May 28-31, 2003 (where May 28 is Day 148).
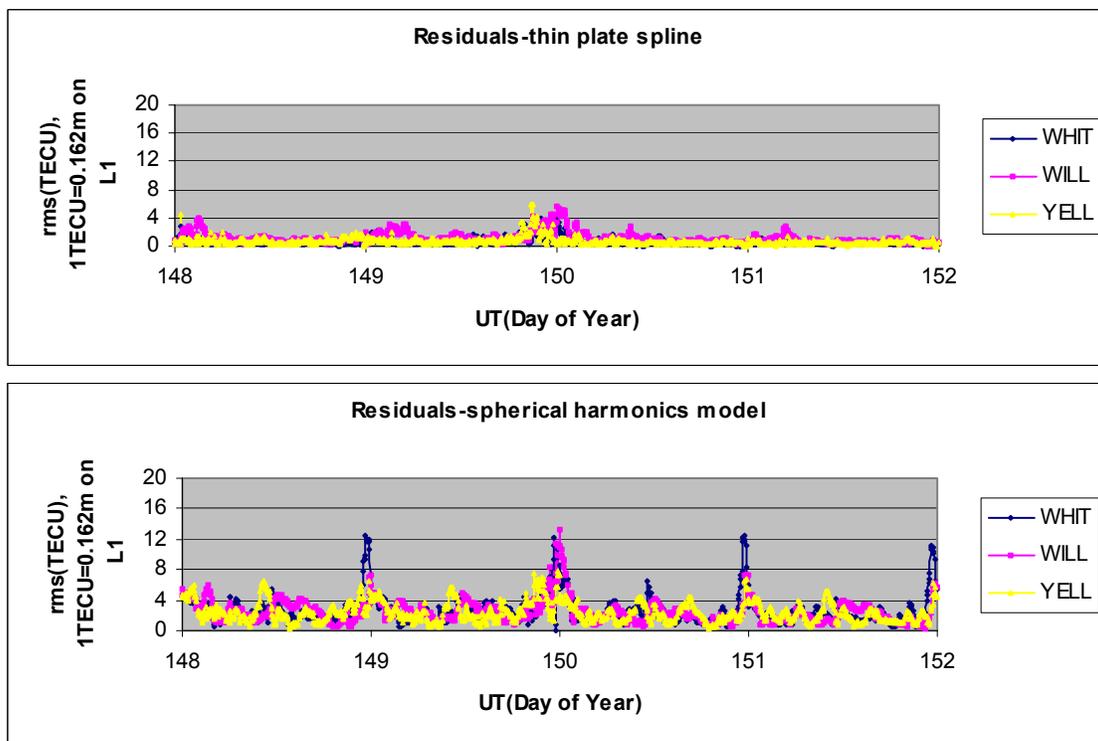
Figure 5.11 Time series of rms of residuals for stations WHIT, WILL, and YELL for May 28-31, 2003 (where May 28 is Day 148).
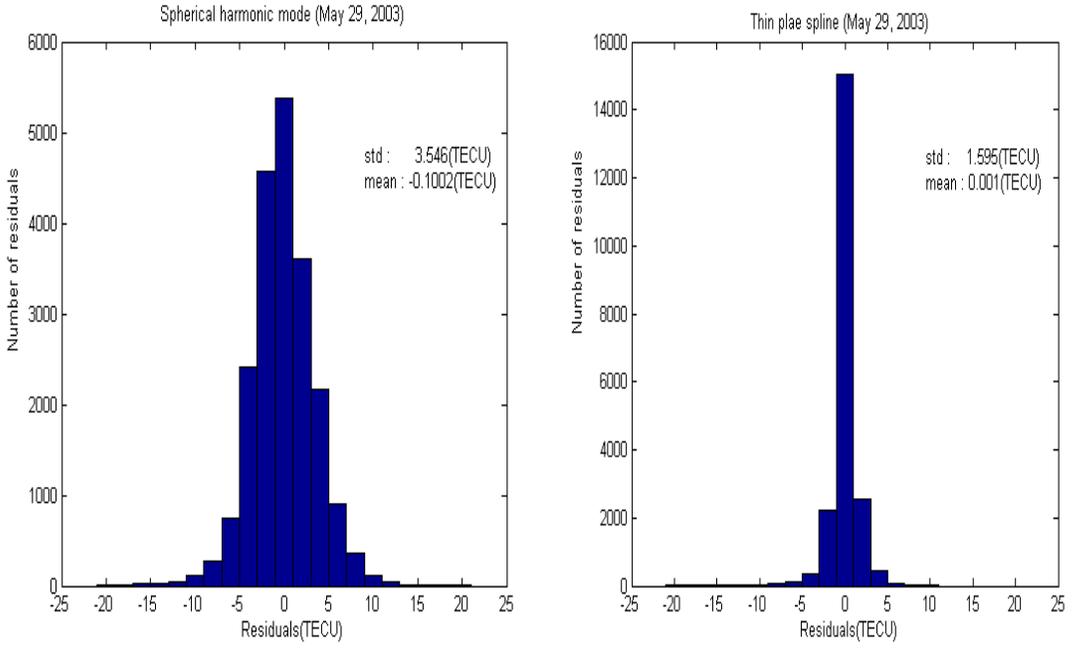
Figure 5.12 Distribution of residuals for all CACS stations combined (May 29, 2003).
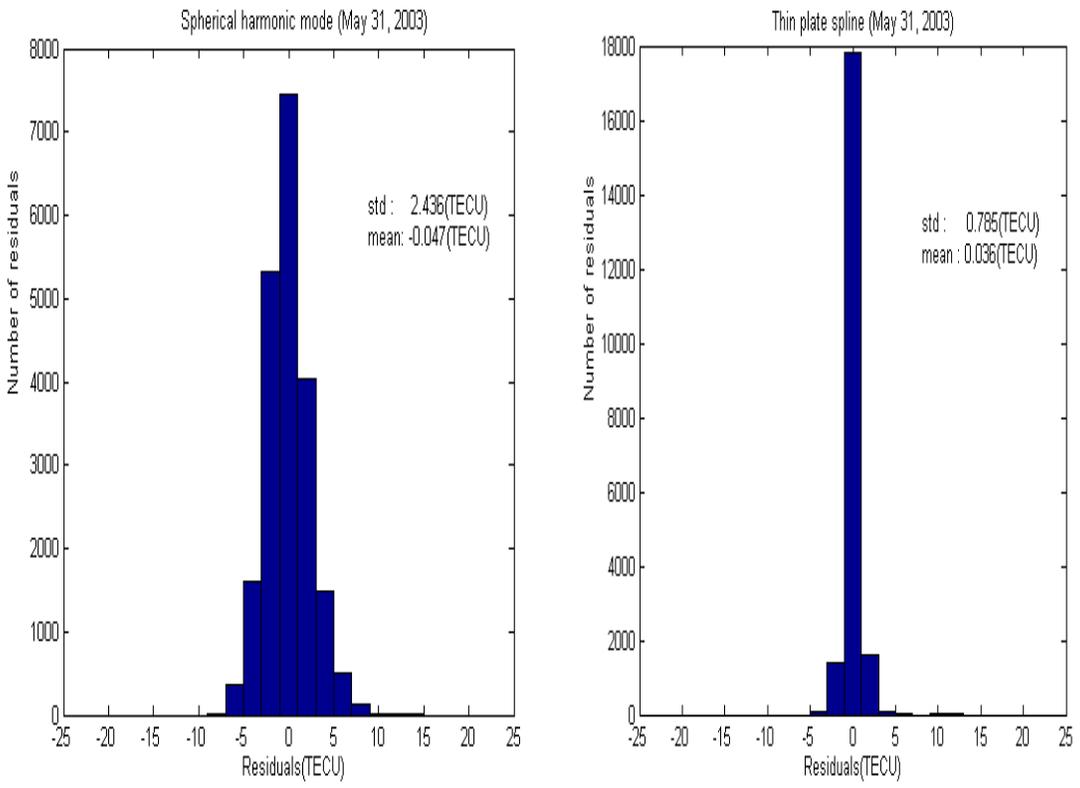
Figure 5.13 Distribution of residuals for all CACS stations combined (May 31, 2003).

Table. 5.2 Comparison of Residuals at 12 stations in CACS Network (May 29, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
ALBH	-0.666	-0.099	3.608	1.564
ALGO	-1.043	-0.184	4.149	2.093
CHUR	-0.297	-0.117	3.402	1.473
DRAO	0.292	-0.182	2.949	1.469
DUBO	-0.588	-0.249	3.879	2.023
FLIN	-0.002	0.409	3.177	1.808
NRC1	-0.263	-0.004	4.479	2.223
SCH2	0.082	0.032	3.879	0.931
STJO	0.658	0.136	3.714	0.816
WHIT	-0.087	-0.126	2.915	1.109
WILL	0.198	0.151	2.985	1.660
YELL	0.421	0.223	3.118	1.334

Table. 5.3 Comparison of Residuals at 12 stations in CACS Network (May 31, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
ALBH	-0.084	-0.139	2.224	1.129
ALGO	-1.082	0.787	2.758	1.119
CHUR	-0.569	-0.178	2.597	0.526
DRAO	0.407	0.016	1.904	0.782
DUBO	-0.704	-0.091	2.922	0.865
FLIN	0.195	0.192	2.089	0.895
NRC1	-0.043	-0.573	2.255	0.995
SCH2	0.528	-0.086	2.639	0.430
STJO	0.448	0.105	2.889	0.304
WHIT	0.129	0.005	2.396	0.453
WILL	-0.288	0.166	2.119	0.846
YELL	0.410	0.232	2.248	0.552

5.3.2 WCDA Network Results

5.3.2.1 Derivation of Optimal Index for Spherical Harmonics Fit

In a similar way as in the CACS network, the optimal K index for SCHA is determined in the WCDA network. Table 5.4 shows the comparison results for different K values. The mean and rms values were derived using all residuals between VTEC observations and predicted VTEC values calculated at 8 WCDA stations in Figure 5.2.

Table 5.4 Comparison of Spherical Harmonic K index

Date	Mean (TECU)		RMS (TECU)	
	$K=2$	$K=3$	$K=2$	$K=3$
May 28	0.142	0.137	3.138	3.181
May 29	0.173	0.197	3.358	3.346
May 30	0.115	0.116	3.161	3.201
May 31	0.127	0.114	2.836	2.913

As in the CACS network, the case where K is 2 shows slightly better results in rms values than results where K is 3. For this reason, the case of $K=2$ is used for further processing to drive WCDA network results.

5.3.2.2 Overall Results

In this section, the performance of the two algorithms is evaluated in the WCDA network using the same procedure as described for the CACS network. As mentioned above, WCDA is a medium size regional GPS network. Due to the station spacings in the WCDA, the grid resolution for the thin plate spline interpolation is set to 2 degrees in longitude and latitude. Figure 5.14 shows the distribution of IPPs observed at 8 stations in the WCDA for a 24-hour period (May 28, 2003).

Distribution of IPP-WCDA (May 28, 2003)
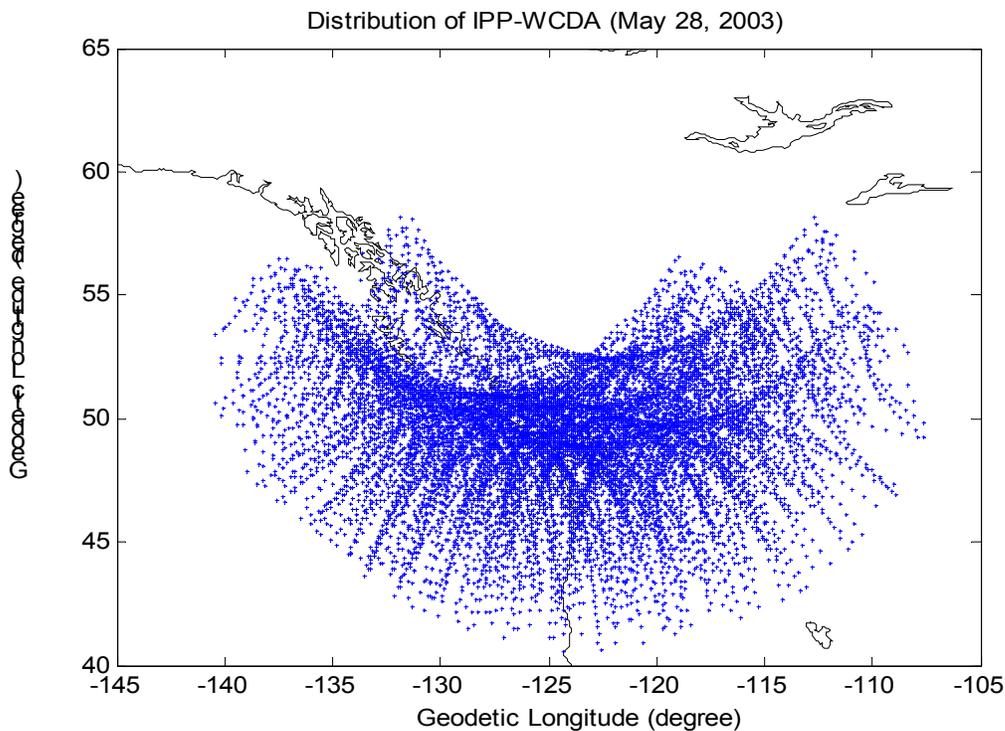
Figure 5.14 Distribution of IPPs in WCDA network.

Figures 5.15-5.17 show time series of the rms of residuals at 8 stations in the WCDA for the spherical harmonics model and the thin plate spline. The thin plate spline outperforms the spherical harmonics model in the WCDA, similar to the results for the CACS. The overall rms of the thin plate spline is approximately 1.0 TECU and overall rms for the spherical harmonics model ranges from 2.8 TECU on May 31 to 3.3 TECU on May 29, 2003. Figures 5.18-5.19 show the distribution of residuals for each each modeling technique for May 29 and May 31, 2003. Tables 5.5 and 5.6 summarize the mean and rms values of the residuals for the 8 stations in the WCDA network for May 29 and May 31, 2003.
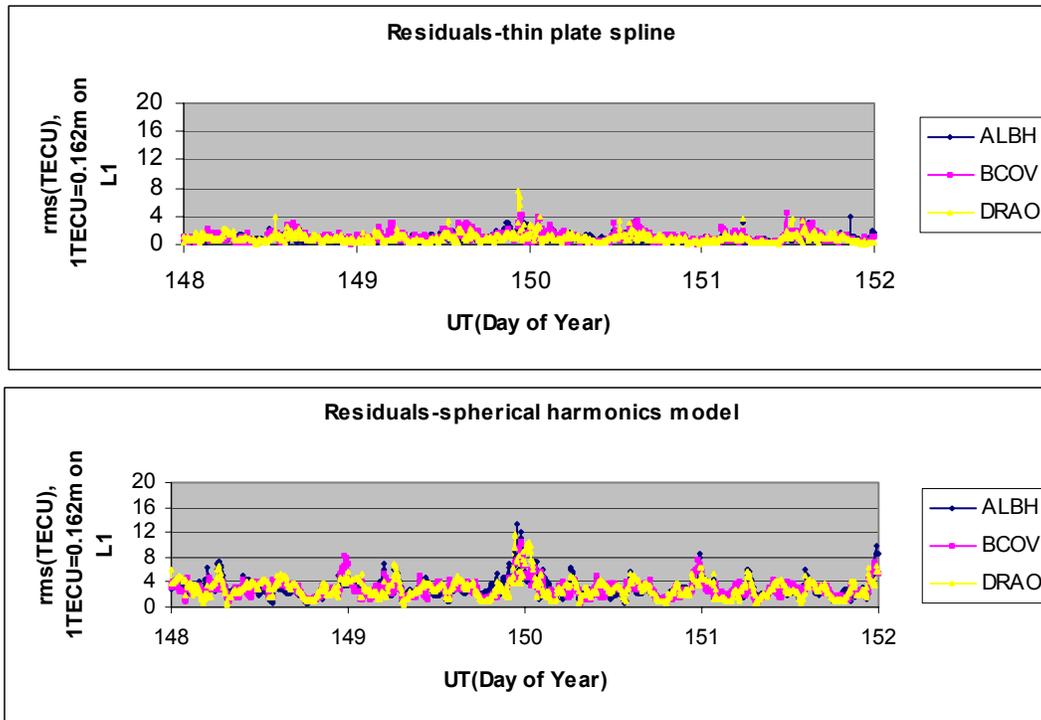
Figure 5.15 Time series of rms of residuals for stations ALBH, BCOV, and DRAO for May 28-31, 2003 (where May 28 is Day 148).
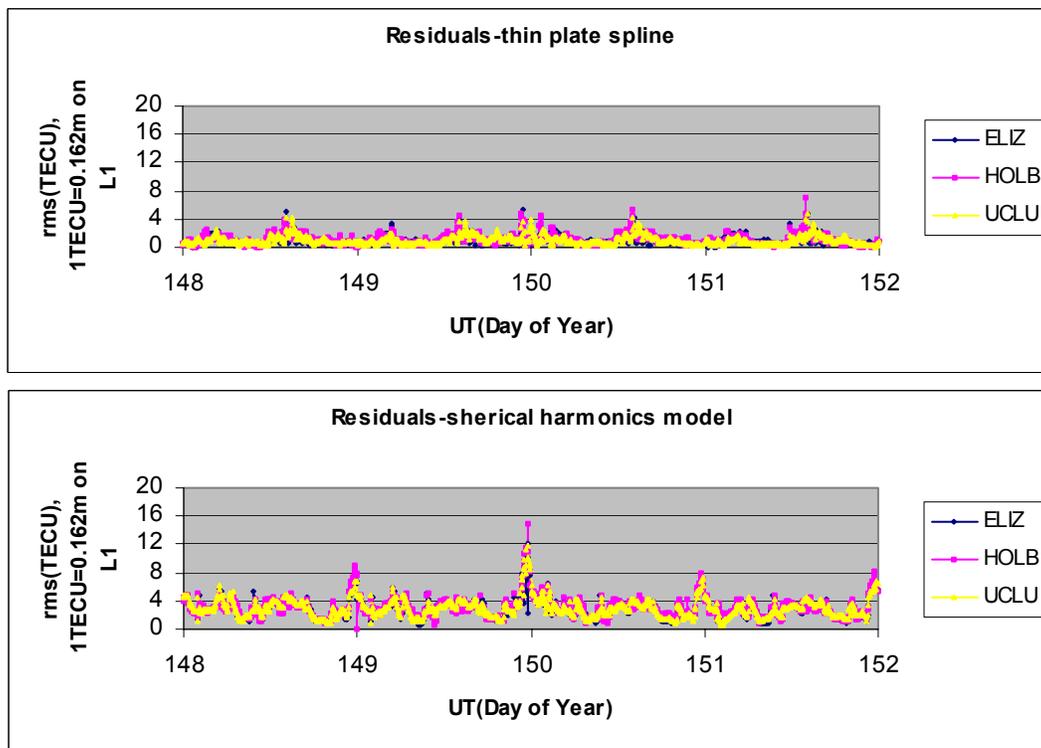
Figure 5.16 Time series of rms of residuals for stations ELIZ, HOLB, and UCLU for May 28-31, 2003 (where May 28 is Day 148).
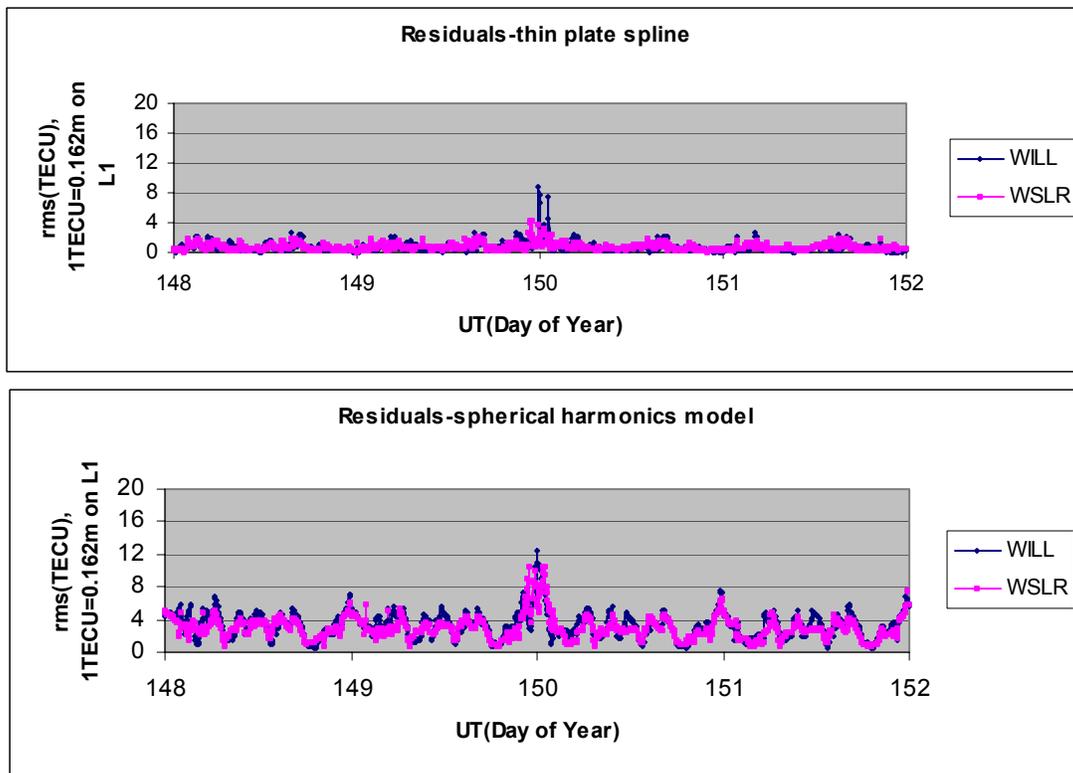
Figure 5.17 Time series of rms of residuals for stations WILL and WSLR for May 28-31, 2003 (where May 28 is Day 148).
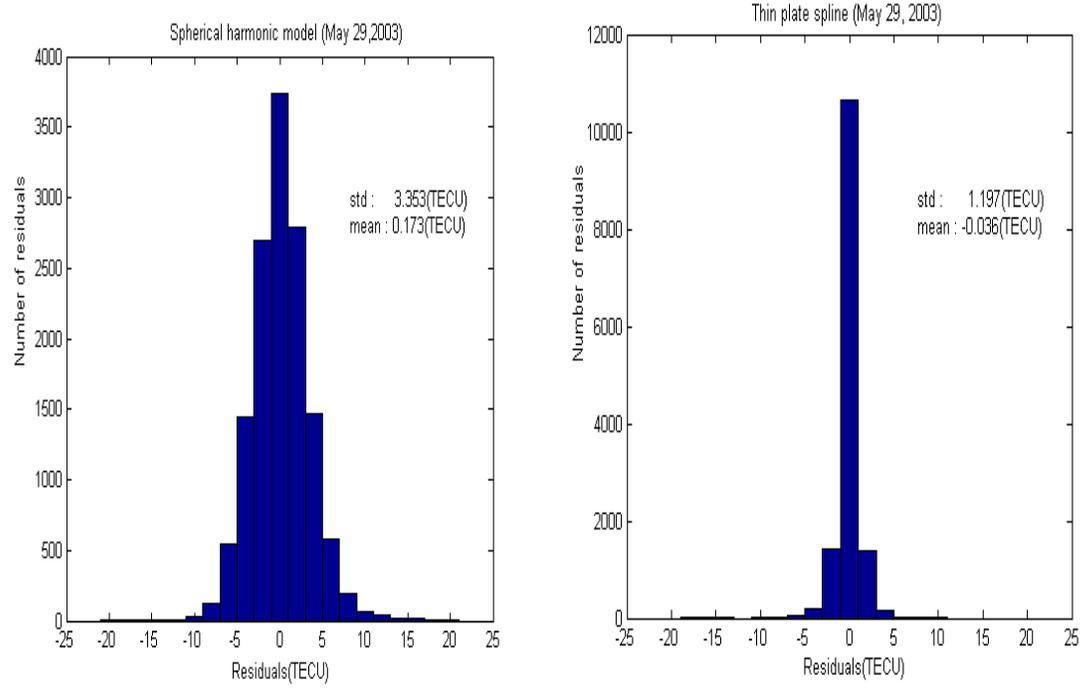
Figure 5.18 Distribution of residuals for all WCDA stations combined (May 29, 2003).
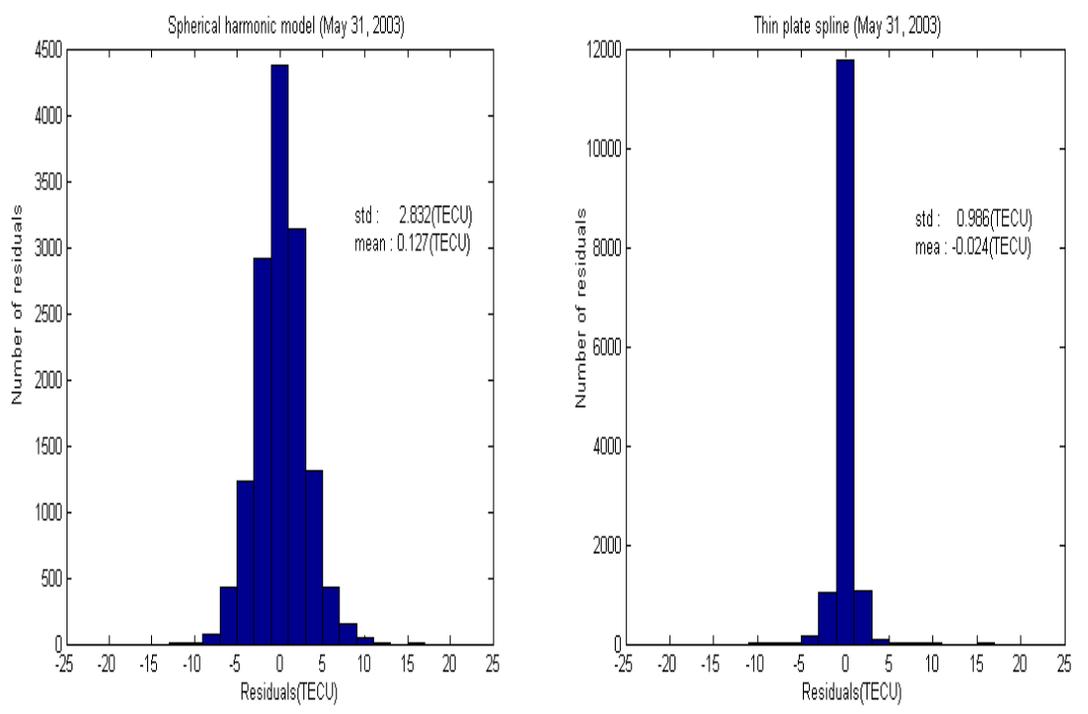
Figure 5.19 Distribution of residuals for all WCDA stations combined (May 31, 2003).

Table. 5.5 Comparison of Residuals at 8 stations in WCDA (May 29, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
ALBH	-0.397	-0.174	3.582	1.216
BCOV	0.399	-0.612	3.161	1.377
DRAO	0.136	0.220	3.382	1.177
ELIZ	0.404	-0.113	3.183	1.141
HOLB	0.155	-0.064	3.346	1.297
UCLU	0.277	-0.231	3.314	1.225
WILL	0.222	0.044	3.595	1.129
WSLR	0.215	0.168	3.255	0.991

Table. 5.6 Comparison of Residuals at 8 stations in WCDA (May 31, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
ALBH	0.112	-0.037	2.849	0.915
BCOV	0.214	-0.300	2.766	1.125
DRAO	0.261	0.266	2.863	0.959
ELIZ	0.129	-0.331	2.729	1.102
HOLB	0.319	-0.145	2.830	1.186
UCLU	0.096	0.160	2.805	1.073
WILL	-0.066	0.089	3.103	0.643
WSLR	0.054	0.106	2.718	0.760

5.3.3 Impact of Network Station Spacing

In this section, the impact of network station spacing on the performance of the algorithms is assessed by comparing the rms values of the residuals at the stations that are included both in the CACS and WCDA networks. Stations ALBH, DRAO and WILL are included in both networks. Tables 5.7 and 7.8 show the comparison results of the rms of the residuals at three stations on May 29 and May 31, 2003. These values were extracted from Tables 5.2, 5.3, 5.5 and 5.6.

Table 5.7 Comparison of RMS of Residuals for Different Network Spacing (March 29, 2003).

Station	RMS (TECU)			
	Spherical harmonics model		Thin plate spline	
	CACS	WCDA	CACS	WCDA
ALBH	3.608	3.582	1.564	1.216
DRAO	2.949	3.382	1.469	1.117
WILL	2.985	3.595	1.660	1.129

Table 5.8 Comparison of RMS of Residuals for Different Network Spacing (March 31, 2003).

Station	RMS (TECU)			
	Spherical harmonics model		Thin plate spline	
	CACS	WCDA	CACS	WCDA
ALBH	2.224	2.849	1.129	0.915
DRAO	1.904	2.729	0.782	1.102
WILL	2.119	3.103	0.846	0.643

In Tables 5.7 and 5.8, it is observed that the thin plate spline shows better results for the smaller network spacing (WCDA network). This is consistent with the fact that the thin plate spline is basically an interpolation scheme. The spherical harmonics model, however, shows worse results for the WCDA network versus CACS network. This is due to the fact that the half angle for SCHA, which was set to 50º, is more realistic to the CACS network than the WCDA network. The number and distribution of observations can be another factor. At each epoch, more observations are available in the CACS network over a larger spatial region, while the number of variables to be estimated is the same for both networks. Note that the spherical harmonics model estimates the variables at each epoch in a real time mode.

5.4 PERFORMANCE ANALYSIS AT INDEPENDENT STATIONS

In the previous section, residuals of the model adjustments were analyzed to determine the accuracies of the two modeling techniques. Residuals are comparisons between model predictions and input observables, however, and these values do not provide truly independent measure of the model accuracies (model estimates being defined by the input observables). In order to evaluate the performance of the two algorithms more rigorously, independent measures of accuracy are derived. The TEC observations at one station in the network are excluded as input observations in the model estimation and interpolation process. Independent accuracies are then measured by computing the residuals between VTEC measurements at IPPs observed at the independent station and VTEC values predicted by the given model at the observation IPPs.

In this thesis, independent stations are divided into two groups: inner stations and outer stations. Inner stations are defined as ones located in the inner part of the network, with maximum overlap of IPPs with neighboring stations, while outer stations are at the boundaries or outside of the network with minimum IPP overlap.

5.4.1 CACS Network Results

Considering the location of the stations in the network and the distribution of IPPs shown in Figure 5.6, DRAO and FLIN are selected as inner stations while SCH2 and WHIT are considered to be outer stations. Figures 5.20-5.21 show time series of the accuracies computed at IPPs for independent stations. The overall rms values are larger than the residuals computed for stations included in estimation and interpolation process (Figures 5.9-5.11). Tables 5.9 and 5.10 summarize the mean and rms values of the accuracies at inner and outer stations for May 29 and May 31, 2003. On May 29, when the ionosphere was highly disturbed, at inner stations, the thin plate spline shows slightly better results than the spherical harmonics model. On the other hand, at outer stations, both algorithms perform poor, giving large rms values and biases.
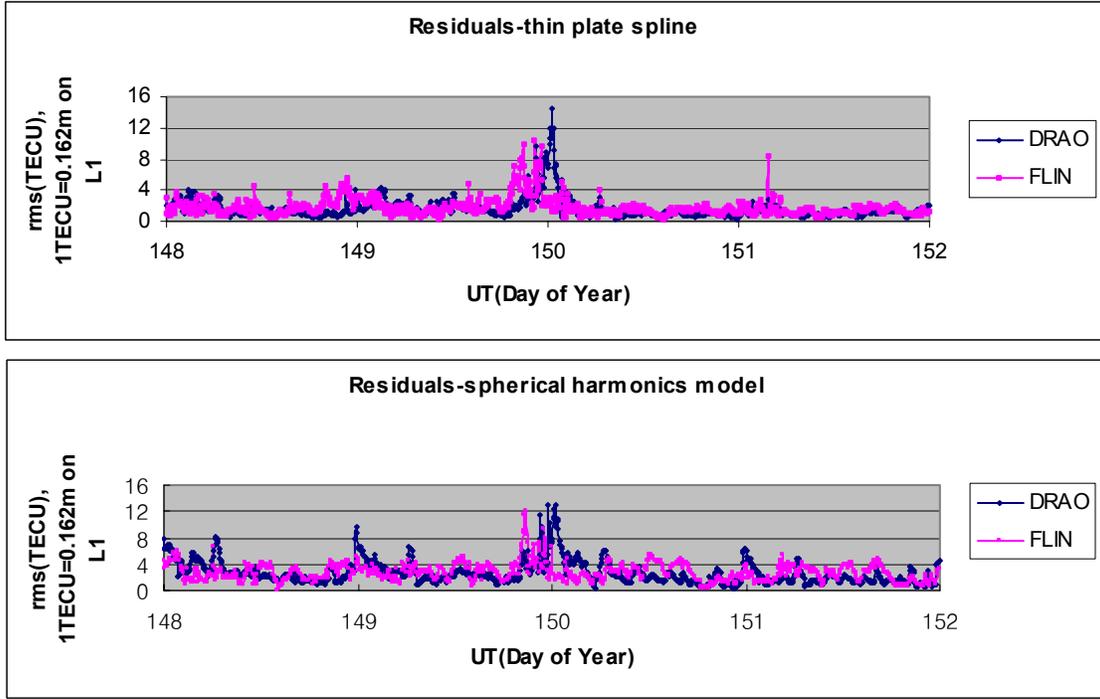
Figure 5.20 Time series of rms of residuals for CACS inner stations DRAO and FLIN for May 28-31, 2003 (where May 28 is Day 148).
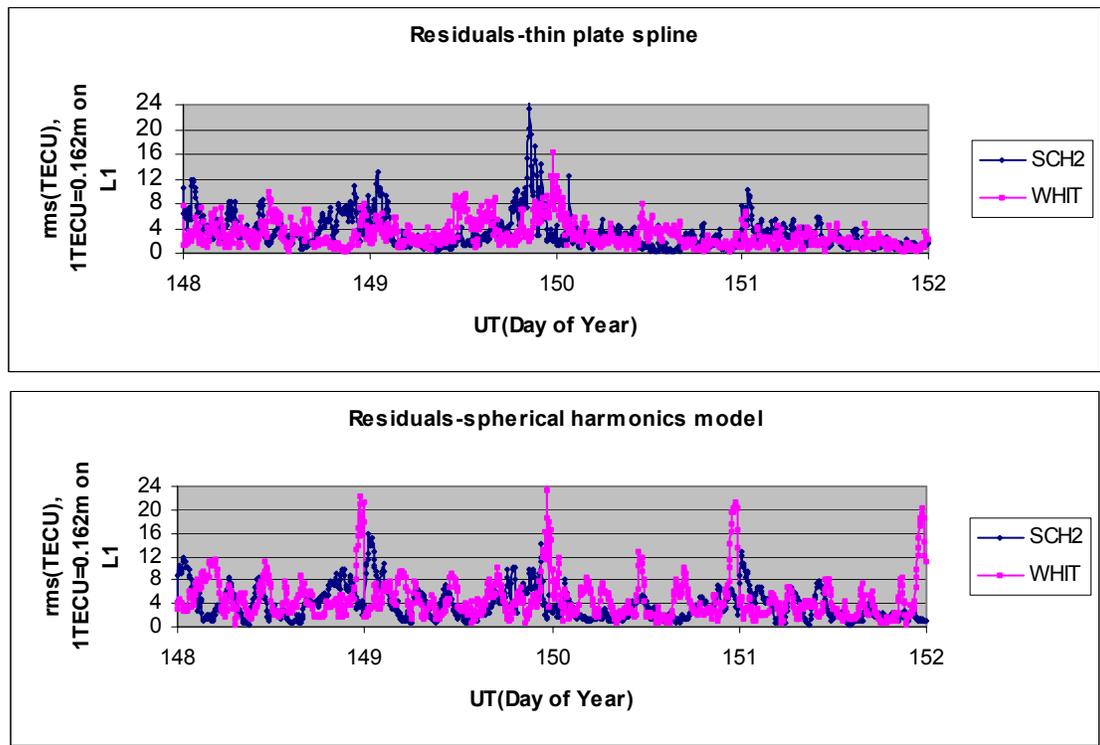
Figure 5.21 Time series of rms of residuals for CACS outer stations SCH2 and WHIT for May 28-31, 2003 (where May 28 is Day 148).

Table 5.9 Statistics of Residuals at Independent Stations (May 29, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
DRAO (Inner)	0.667	-0.457	3.309	2.474
FLIN (Inner)	-1.136	0.977	3.706	3.160
SCH2 (Outer)	-2.222	-1.543	5.648	6.099
WHIT (Outer)	1.287	-1.559	5.333	4.538

Table 5.10 Statistics of Residuals at Independent Stations (May 31, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
DRAO (Inner)	0.506	-0.036	2.024	1.213
FLIN (Inner)	-1.268	0.377	2.573	1.533
SCH2 (Outer)	-1.251	-2.037	3.743	2.946
WHIT (Outer)	0.636	0.345	4.309	2.087

This indicates that under highly disturbed ionospheric conditions, both interpolation and functional fitting is not reliable at the boundary or beyond the sparse network. On a relatively quiet day, May 31, the thin plate spline performs slightly better at both inner stations and outer stations. The thin plate spline shows particularly good results for the outer stations in comparison to the spherical harmonics model. The rms value at outer stations for thin plate spline was below 3.0 TECU, which corresponds to 0.48 m range delay on L1.

Figure 5.21 shows the residuals of VTECs versus elevation for the spherical harmonics model and thin plate spline at station DRAO on May 31, 2003. In the case of the

spherical harmonics model, large residuals are observed more frequently in low elevation angles. This is related to the mapping function used for converting slant TECs to VTECs. As described in Section 4.1, slant TECs are mapped to VTECs in the 2-dimensional ionosphere shell by applying a mapping function (equation 4-1) on the assumption that all electrons in the ionosphere are concentrated in the thin shell at fixed altitude (nominally 350km). But this assumption may not be valid, and the error in the altitude of the thin shell results in the error of the mapping function (see equation 4.1) and IPP coordinates (see APPENDIX B). These error factors are more significant in lower elevation angles. But in the case of the thin plate spline, the residual-elevation relation observed in the spherical harmonics model is not shown. This is due to the fact that the thin plate spline is based on the exact interpolation scheme.

(a)   (b)
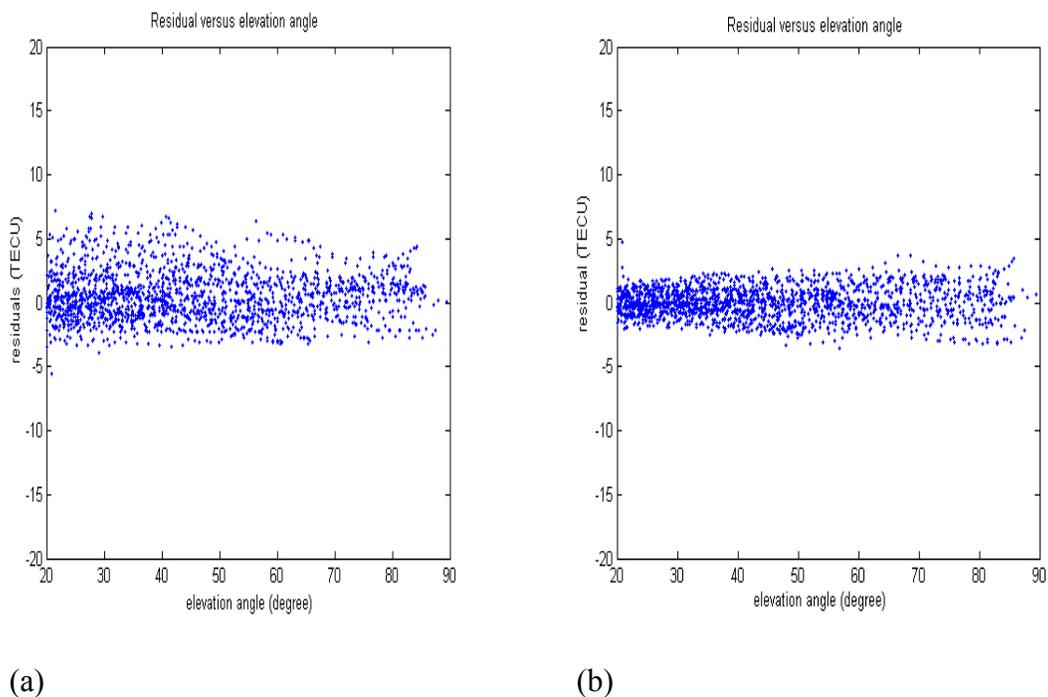
Figure 5.22 Residuals of VTEC versus elevation angle for a) the spherical harmonics model and b) thin plate spline at station DRAO on May 31, 2003.

5.4.2 WCDA Network Results

In this network, two stations WSLR and WILL are selected as inner and outer stations, respectively. Time series of the rms values are presented in Figures 5.23 and 5.24. The general trend of rms temporal variations is similar to that shown in Figures 5.15-5.17, but with increased rms values overall. The statistics of the accuracy estimate for May 29 and May 31 are presented in Tables 5.11 and 5.12. One notable feature of results in Table 5.9 is that, in comparison to results in the CACS network, the thin plate spline shows much improved rms values at the outer station, WILL. This is due to the denser grid configuration in the WCDA network. On May 31, the general performance of the spherical harmonics model and the thin plate spline at outer and inner stations is almost same in the WCDA network as in the CACS network.
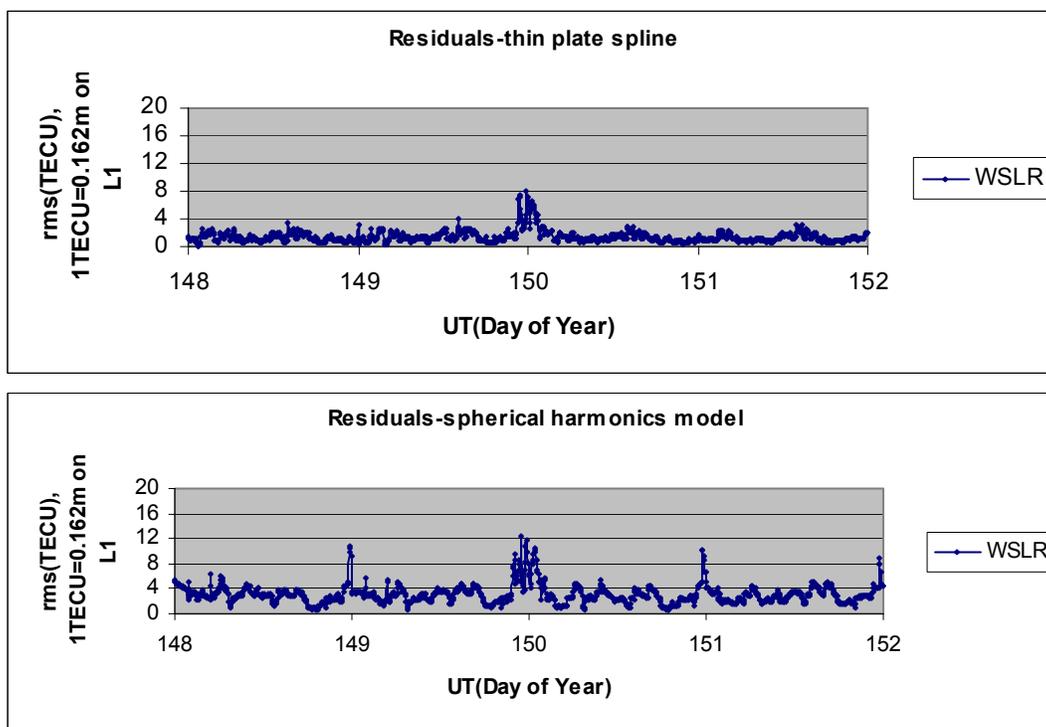
Figure 5.23 **Time series of rms of residuals for a WCDA inner station WSLR for May 28-31, 2003 (where May 28 is Day 148).**
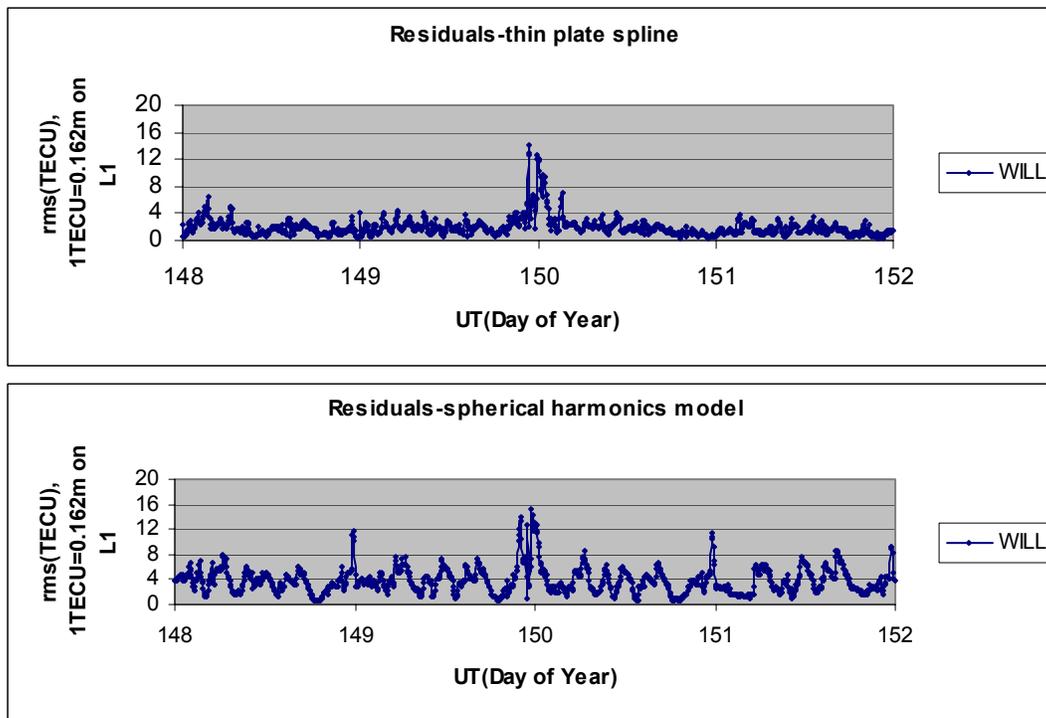
Figure 5.24 **Time series of rms of residuals for a WCDA outer station WILL for May 28-31, 2003 (where May 28 is Day 148).**

Table 5.11 Statistics of Residuals at Independent Stations-WCDA (May 29, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
WSLR (Inner)	-0.278	0.259	3.528	1.751
WILL (Outer)	-0.590	-0.118	4.731	2.685

Table 5.12 Statistics of Residuals at Independent Stations-WCDA (May 31, 2003)

Station	Mean (TECU)		RMS (TECU)	
	Spherical harmonics model	Thin plate spline	Spherical harmonics model	Thin plate spline
WSLR (Inner)	-0.984	0.120	3.016	1.265
WILL (Outer)	-1.699	0.007	3.987	1.599

Figure 5.25 shows the residuals of VTECs versus elevation angle for the spherical harmonics model and thin plate spline at station WSLR on May 31. For the spherical harmonics model, the same residual-elevation angle relation as observed at station DRAO in the CACS network is found. But in the case of the thin plate spline, no significant residual-elevation relation is observed.

a)   b)
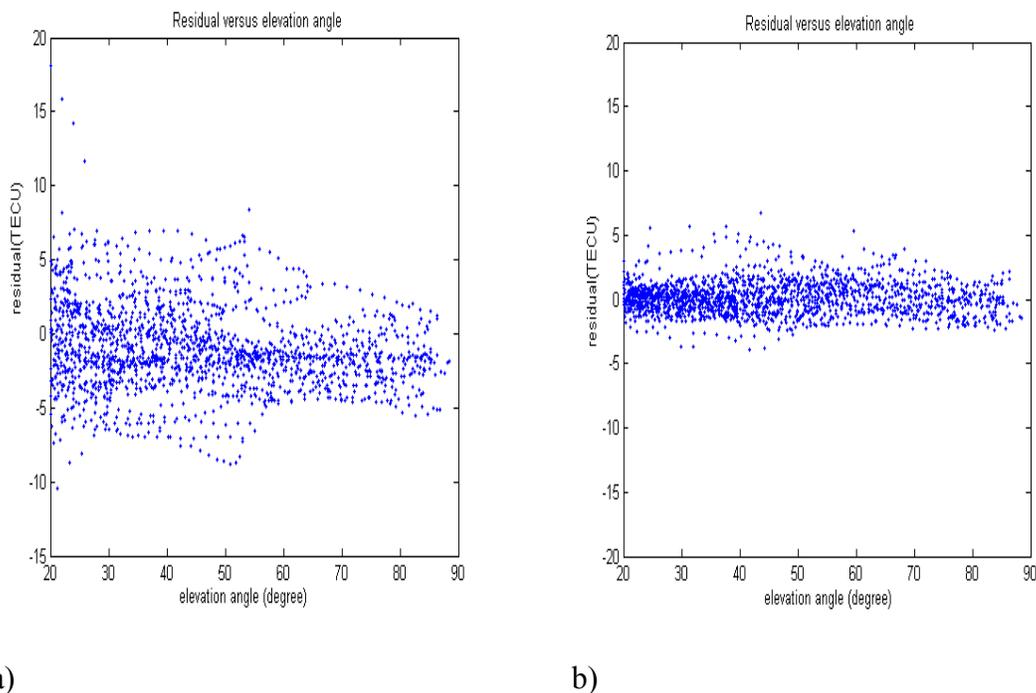
Figure 5.25 Residuals of VTEC versus elevation angle for a) the spherical harmonics model and b) thin plate spline at station WSLR on May 31, 2003.

5.5 VTEC MAPS

In this section, the spatial variation of VTEC over Canada and its impact on the performance of the algorithm are analyzed. The thin plate spline is selected for this analysis because it shows better performance in accuracy than the spherical harmonics model. The spatial variation of VTEC is represented with VTEC maps for the whole of Canada. In this thesis, three epochs are selected from the geomagnetic storm period, and time snapshots of VTEC maps at those epochs are plotted. The TEC maps are interpolated from VTEC values at the grid points generated over Canada with the

resolution of 2 degree in latitude and longitude respectively. The performance of the thin plate spline at these epochs is evaluated from the rms values of residuals for 12 stations in the CACS network.

Figure 5.26 shows rms values of residuals derived at the 12 stations using a thin plate spline at 20:50 UT on May 29, 2003. The stations ALGO, DUBO, FLIN and NRC1 show much larger rms values compared to typical accuracies shown Table 5.2. The station YELL also has an increased rms value.
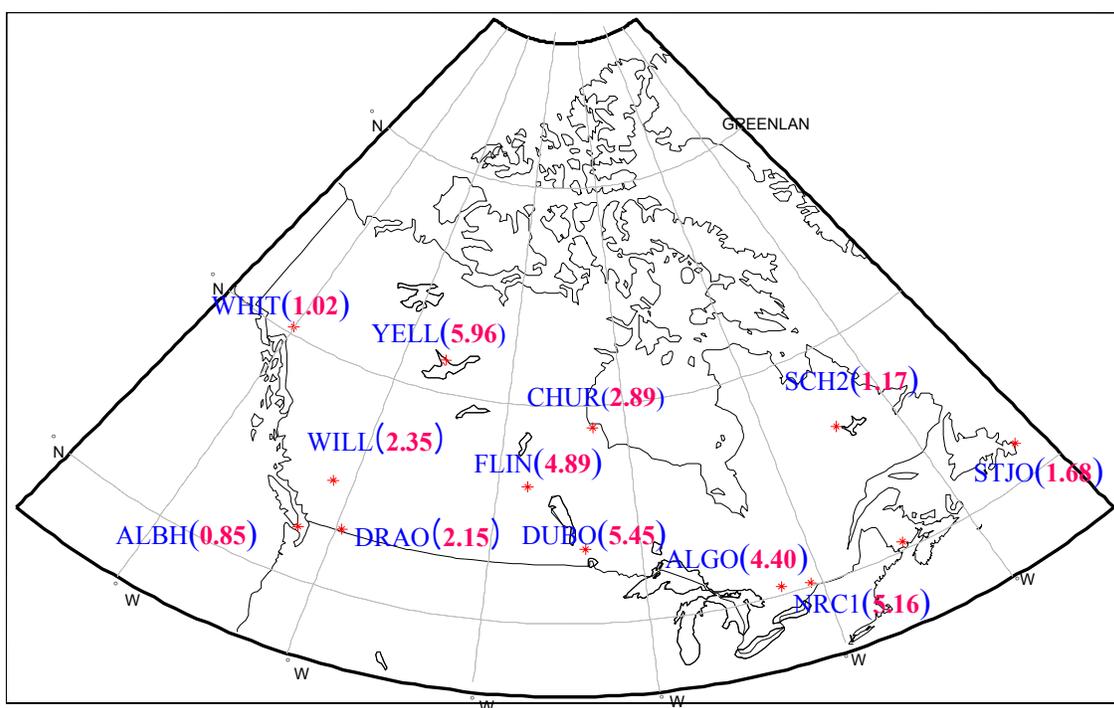
Figure 5.26 Spatial distribution of rms of residuals for 12 CACS stations at 20:50 UT on May 29.

Figure 5.27 shows the VTEC map at the same epoch. Regions with large localized enhancement of VTEC in Figure 5.27 coincide with the locations of stations with larger rms values (compare Figure 5.26 and Figure 5.27). This demonstrates the impact of large ionospheric gradients on the performance of the thin plate spline algorithm.
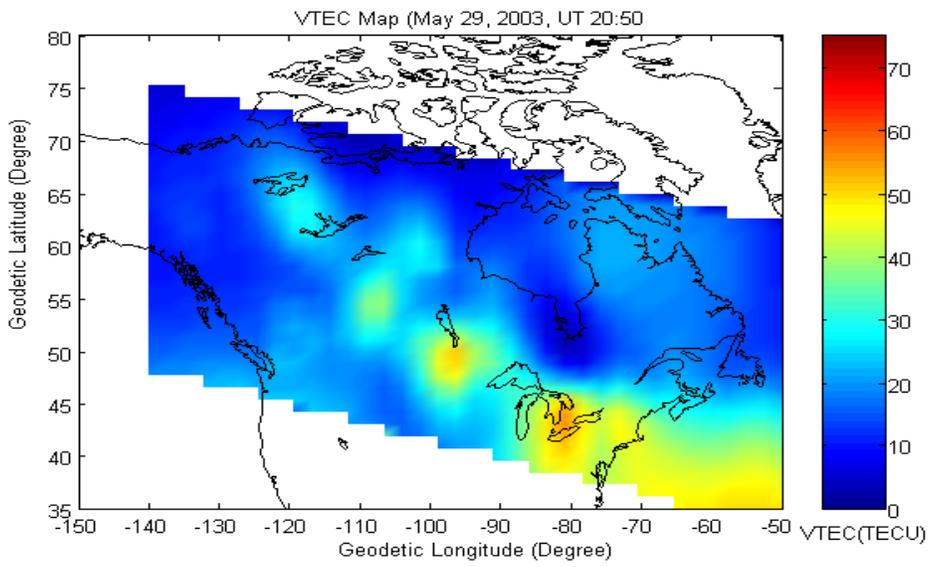
Figure 5.27 VTEC Map at 20:50 UT on May 29.

Figure 5.28 shows the spatial distribution of rms of VTEC residuals at a different epoch, 22:45 UT on May 29.
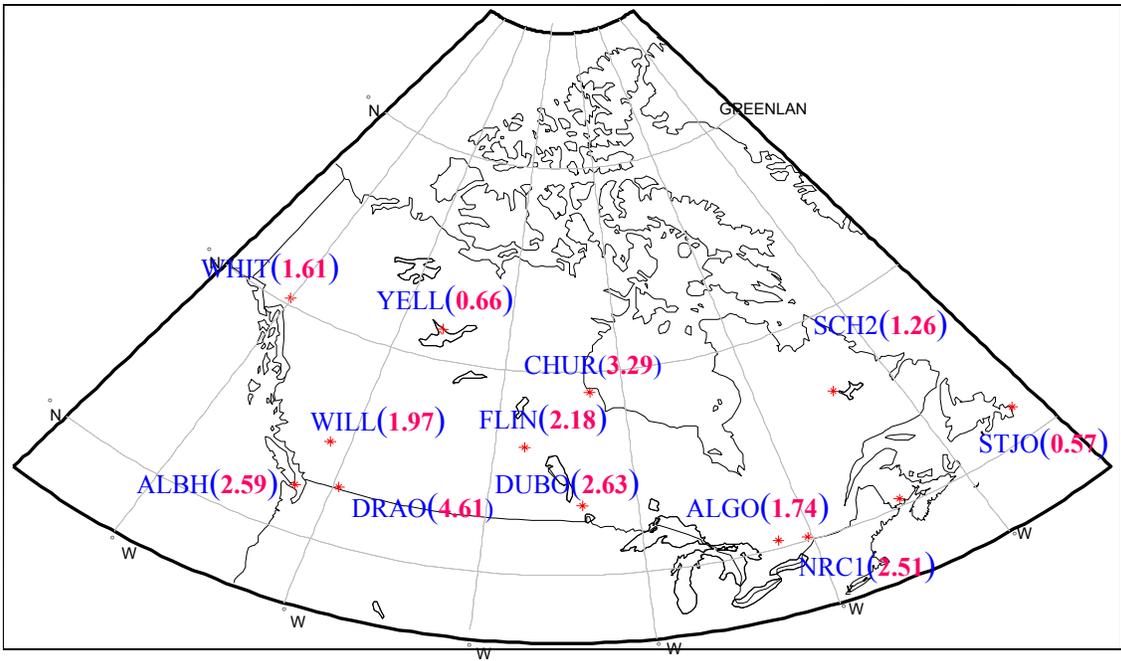
Figure 5.28 Spatial distribution of rms of residuals for 12 CACS stations at 22:35 UT on May 29.

Compared to the previous epoch (Figure 5.26), the rms values at stations ALGO, DUBO, FLIN, NRC1 and YELL are reduced. On the other hand, stations ALBH, DRAO and WILL show increased rms values. Figure 5.29 is the VTEC map for the same epoch. As shown in the map, the region of enhanced VTEC has moved to the west, with larger ionospheric gradients in western Canada. This is consistent with the increased rms values at stations ALBH, DRAO and WILL.
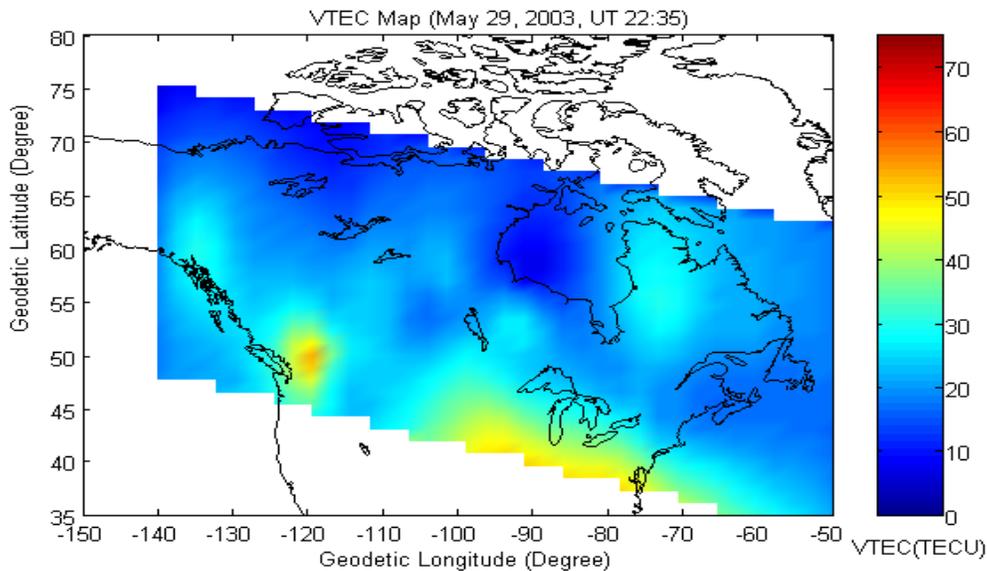
Figure 5.29 VTEC Map at 22:35 UT on May 29.

Figure 5.30 illustrates the spatial distributions of rms of residuals at epoch 23:45 UT on May 29. One noticeable change in comparison to the previous epochs is the increased rms values at stations ALBH, DRAO and WHIT. Alternatively, the rms values at stations ALGO, DUBO, FLIN and NRC1 are much reduced. Figure 5.31 shows the VTEC map at the same epoch. Larger ionospheric gradients are observed in western Canada, consistent with the larger model residuals for western Canadian stations in Figure 5.30. VTEC values are decreased over most parts of eastern Canada and increased over northern British Columbia.
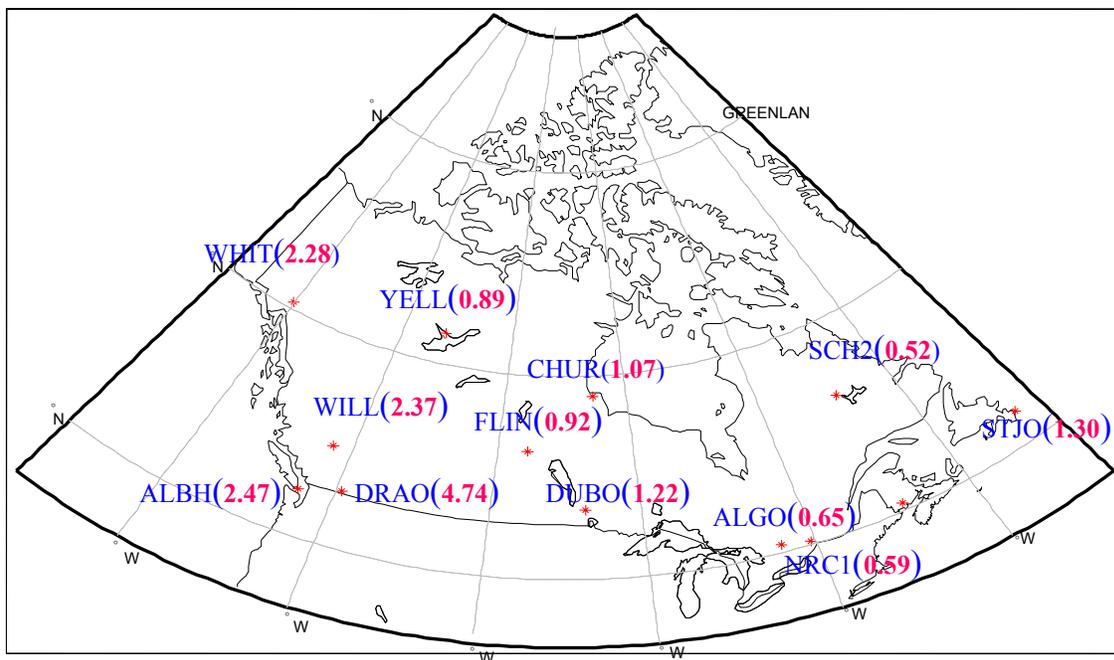
Figure 5.30 Spatial distribution of rms of residuals for 12 CACS stations at 23:45 UT on May 29.
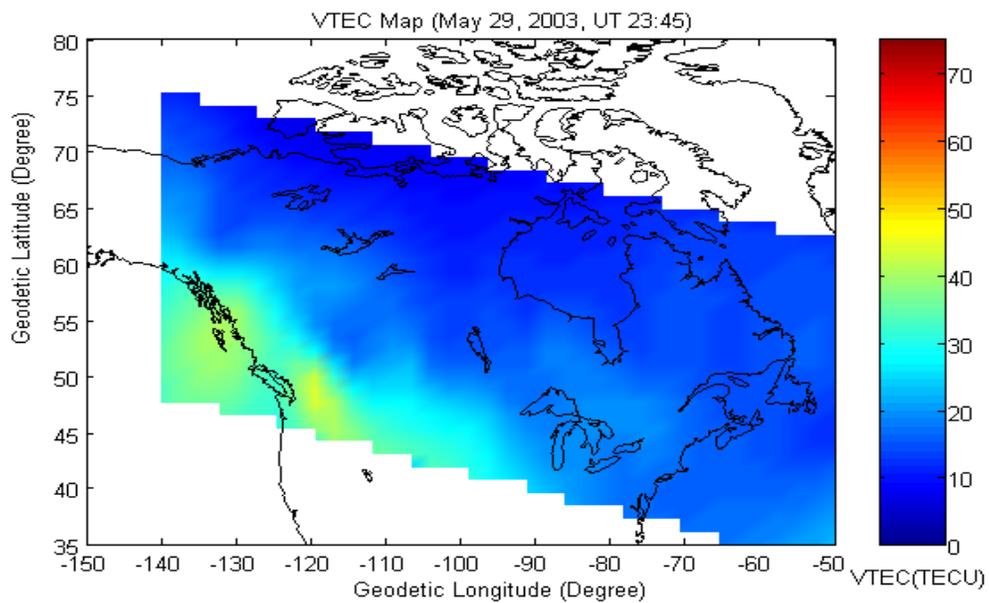
Figure 5.31 VTEC Map at 23:45 UT on May 29

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this thesis was to find an optimal ionosphere model applicable for real time WADGPS applications in Canada. Due to the sparse network of reference stations in Canada and the geographical location of Canada that is more susceptible to ionospheric disturbances, a reliable ionosphere model that is optimal to Canadian region is required. In this thesis, two ionosphere modeling algorithms, spherical harmonics model and thin plate spline interpolation, were selected, evaluated and compared in terms of accuracy in two permanent GPS networks of different dimensions in Canada: CACS and WCDA. The accuracies of the algorithms were measured by computing the residuals between the VTEC observations and predicted VTECs at IPPs. The accuracies of the algorithms were also evaluated for a range of ionospheric activity. The following conclusions were drawn from the research performed in this thesis.

6.1 CONCLUSIONS

The thin plate spline interpolation showed better results in both networks. On a highly disturbed day (May 29, 2003), the overall rms values for the thin plate spline and spherical harmonics model were 1.5 and 3.5 TECU, respectively. On a relatively quiet day (May 31, 2003), the overall rms values for the thin plate spline and spherical harmonics model were 0.8 and 2.4 TECU. Note that 1 TECU amounts to 0.162m range delay on L1. In the CACS network, both algorithms showed much increased rms values (up to 16 TECU) at the end of May 29 and early hours of May 30 when the Kp index reached the maximum value of 8. This indicates that both algorithms could not resolve the large ionospheric gradients caused by a geomagnetic storm in the sparse CACS network. In the WCDA network, the thin plate spline showed reduced rms values at epochs, when the Kp index reached the maximum value of 8.

The effect of the network station spacing on the algorithms was investigated by comparing the rms values of the residuals at stations that are included in both networks. The thin plate spline showed improved results in the smaller network spacing (WCDA network), which is consistent with the fact that the thin plate spline is an interpolation-based algorithm. The spherical harmonics model showed worse results in the WCDA network. This is due to the unrealistic half angle for SCHA and smaller observation number.

The performance of the algorithms was analyzed more rigorously by excluding the VTEC observations at one station from model adjustment and computing independent statistics. During disturbed ionospheric conditions, both model algorithms showed poor accuracy (rms values of over 5 TECU) at the boundary of the CACS network. At the boundary of the WCDA network, however, the thin plate spline showed relatively good accuracy (rms value of 2.6 TECU) even during disturbed ionospheric conditions.

The spatial variation of VTEC over Canada and its impact on the performance of the algorithm were analyzed. In this case, the thin plate spline was selected for the analysis. In considering the VTEC maps generated for the storm period, it was observed that regions with large enhancements of vertical TEC coincided with the locations of stations with larger rms values.

Given the accuracy analysis of the two algorithms in two networks, the grid-based thin plate spline method is recommended for both national and regional GPS networks in Canada. Typical accuracies of the thin plate spline under average ionospheric conditions are approximately 0.32m for the CACS network and 0.23m for the WCDA network. One drawback of the thin plane spline is that it cannot estimate the satellite and receiver interchannel biases, which are required to convert slant TECs to vertical TECs. Accurate values of the interchannel biases must be provided from external sources.

6.2 RECOMMENDATIONS

In this research, the algorithms were evaluated during a geomagnetic storm, which occurred on May 29, 2003. The accuracy of the algorithms is degraded in the presence of large ionospheric gradients, and the extent of the accuracy degradation is dependent on the spatial and temporal variation of VTEC over the region of interest. For more rigorous evaluation of the performance of the ionosphere model, data from several geomagnetic storm events, with different characteristics of spatial and temporal VTEC variations, should be processed. In accordance with the actual ionosphere gradients, different grid sizes can be applied for different regions.

In this thesis, the coefficients for the spherical harmonics model were estimated using a Kalman filter in a real time mode. If this requirement is relaxed so that all observations for a short time period (several minutes) are processed in a batch mode, higher K index for SCHA would be possible, and this may result in improved accuracy for the spherical harmonics model.

The thin plate spline implemented in this thesis is based on the exact interpolation scheme. But in disturbed ionospheric conditions, it would be more realistic to introduce a smoothing parameter that smoothes noisy data. The selection of the appropriate value of the smoothing parameter would be dependent on the characteristics of the ionospheric condition in hand.

Further research would be to validate and compare the ionosphere algorithms investigated in this thesis with the existing WADGPS ionosphere models. Especially the comparison with the WAAS ionosphere algorithm in the CONUS and Canadian region would be an interesting topic.

The two ionosphere models proposed in this thesis were evaluated in the Canadian GPS networks. Further research can be done to evaluate the performance of the algorithms in different geographic region, especially lower latitude region, where different ionospheric features are exhibited.

REFERENCES

Bierman, G. J., Factorization Methods for Discrete Sequential Estimation, Academic Press, INC, New York, 1977.

Bjorck, A., Solving least squares problems by Gram-Schmidt orthogonalization, BIT, 7, pp 1-21, 1967.

Blewitt, G, An automatic editing algorithm for GPS data, Geophysical Research Letters, 17, no 3, pp 199-202, 1990.

Brunner, F. K., and M. Gu, An improved model for the dual frequency ionospheric correction of GPS observations, Manuscripta Geodaetica, 16, pp 205-214, 1991.

Davies, K., Ionospheric Radio, Peter Peregrinus Ltd, London, 1990.

Duchon, J., Spline minimizing rotation-invariant semi-norms in Sobolev spaces, Constructive theory of functions on several variables (W. Schempp and K. Zeller, eds.), Lecture Notes in Math. **571**, Springer-Verlag, Berlin, 1977.

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 Institute of Navigation, 41, no. 4, pp. 393-413, winter 1994-95.

Erickson, C., and P. Heroux, GPS locations for GIS: getting them right the first time, EOM, 4, no 6, 1995.

FAA, FAA-E-2892B, Change 1: Wide Area Augmentation (WAAS) specification, 1999.

Gelb, A., Applied Optima Estimation, MIT Press, Cambridge, 1974.

Golub, G. H., and C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, 1996.

Grewal, M. S., L. R. Weil and A. P. Andrews, Global Positioning Systems, Inertial Navigation, and Integration, John Wiley and Sons, New York, 2001.

Gurtner, W., and G. Mader, Receiver Independent Exchange Format Version , CSTG GPS Bulletin, 3, no. 3, National Geodetic Survey, Rockville, 1990.

Haines, G. V., Spherical Cap Harmonic Analysis, Journal of Geophysical Research, 90, B3, pp 2583-2591, 1985.

Hopfield, H. S., two quartic tropospheric refractivity profiles for correcting satellite data, Journal of Geophysical Research, 74, no 18, pp 4487-4499, 1969.

Janes, H. W., R. B. Langley and S. B. Newby, Analysis of tropospheric delay prediction models: comparison with ray tracing and implication for GPS relative positioning, Bulletin Geodesique, 65, no. 3, pp 151-161, 1991.

Kailath, T., Linear Estimation, Prentice Hall, Upper Saddle River, 2001.

Kaplan, D., Understanding GPS: Principles and Applications, Artech House Publishers, Boston, 1996.

Kassam, A., B. Hlasney, V. Vogt, K. Lochhead, P. Panther, and N. Carter, The Canada-wide differential GPS (CDGPS) service: new infrastructure launched for GPS-based geo-referencing and navigation, Proceedings of the ION GPS/GNSS 2002, Portland, pp 1798-1805, September 24-27, 2002.

Kee, C. and B. W. Parkinson, Wide Area Differential GPS (WADGPS): Future Navigation System, IEEE Transaction on Aerospace and Electronic Systems, 32, no 2, 1996.

Kolbuchar, J. A., Ionospheric effects on GPS, in Global Positioning Sytem: : Theory and Applications, Vol. 1, B.W. Parkinson and J. J. Spilker, P. Axelrad and P. Enge (Eds), pp 485-513, AIAA, Washington D.C., 1996.

Leick, A., GPS Satellite Surveying, John Wiley & Sons, Inc, Hoboken, 2004.

Mayaud, P. N., Derivation, meaning and use of geomagnetic indices, AGU, Washington D.C., 1980.

Meinguet, J., An intrinsic approach to multivariate spline interpolation at arbitrary points, in Polynomial and Spline Approximation, ed. B. N. Sahney, Reidel, Dordrecht, Holand, 1979.

Misra, P., and P. Enge, Global Positioning System: Signals, Measurements, and Performance, Ganga-Jamuna Press, Lincoln, 2001.

Rishbet, H., and O. K. Garriot, Introduction to Ionospheric Physics, Academic Press, New York, 1969.

Saastamoinen, J., Atmospheric correction for troposphere and stratosphere in radio ranging of satellites, Geophysical Monograph, 15, American Geophysical Union, Washington D. C., pp 247-252, 1972.

Schaer, S., Mapping and predicting the earth's ionosphere using the Global Positioning System, Ph.D dissertation, Astronomical Institute, University of Berne, Switzerland, March 25, 1999.

Seeber, G., Satellite Geodesy: Foundations, Methods, and Applications, de Gruyter, Berlin, 2003.

Skone, S., TECANALYS^{TM}, Operating Manual, Department of Geomatics Engineering, University of Calgary, 2002.

Skone, S., A. Coster, V. Hoyle, and C. Laurin, WAAS availability and performance at high altitudes, Proceedings of the ION GPS/GNSS 2003, Portland, pp 1279-1287, September 9-12, 2003.

Spilker, J. J., Signal Structure and Theoretical Performance, In Global Positioning System: Theory and Applications, Vol. 1, B.W. Parkinson and J. J. Spilker, P. Axelrad and P. Enge (Eds), pp 89-121, AIAA, Washington D.C., 1996.

Thornton, C. L., Triangular Covariance Factorization for Kalman Filtering, Ph.D. Thesis, University of California at Los Angeles, School of Engineering, 1976.

Townsend, B. P., P. Fenton, K. Van Dierendonck and R. Van Nee, Performance evaluation of the multipath eliminating delay lock loop, NAVIGATION: Journal of the Institute of Navigation, 42, no 3, pp 503-514, 1995.

Van Dierendonck, A. J., P. Fenton and T. Ford, Theory and performance of narrow correlator spacing in a GPS receiver, NAVIGATION: Journal of the Institute of Navigation, 39, no 3, pp 265-283, 1992.

Van Dierendock, A.J., GPS Receivers, In Global Positioning System: Theory and Applications, Vol. 1, B.W. Parkinson and J. J. Spilker, P. Axelrad and P. Enge (Eds), pp 329-406, AIAA, Washington D.C., 1996.

Ward, P., Satellite Signal Acquisition and Tracking, in Understanding GPS: Principles and Applications, E. Kaplan (Ed), Artech House Publishers, Boston, 1996.

Weil, L. and B. Fisher, Method of mitigation of multipath effects in radio ranging systems, U.S Patent No. 6,031,881, Feb 29, 2000.

APPENDIX A

RECEIVER/ANTENNA TYPES FOR CACS AND WCDA NETWORKS

Table A.1 CACS Network

Station	Location		Receiver Type	Antenna Type
	Latitude (deg min sec)	Longitude (deg min sec)		
ALBH	48 23 23.28	-123 29 14.64	AOA BENCHMARK ACT	AOAD/M_T
ALGO	45 57 31.68	-78 04 17.04	AOA BENCHMARK ACT	AOAD/M_T
CHUR	58 45 32.76	-94 05 19.32	AOA SNR-12 ACT	AOAD/M_T
DRAO	49 19 21.36	-119 37 30.00	AOA BENCHMARK ACT	AOAD/M_T
DUBO	50 15 31.68	-95 51 58.32	ROGUE SNR-8000	AOAD/M_T
FLIN	54 43 32.16	-101 58 40.80	AOA BENCHMARK ACT	AOAD/M_T
NRC1	45 27 15.12	-75 37 25.68	AOA SNR-12 ACT	AOAD/M_T
SCH2	54 49 55.56	-66 49 57.36	AOA SNR-12 ACT	AOAD/M_T
STJO	47 35 42.86	-52 40 39.89	AOA SNR-12 ACT	AOAD/M_T
WHIT	60 45 01.80	-135 13 19.56	AOA SNR-8000 ACT	AOAD/M_T
WILL	52 14 12.84	-122 10 04.08	AOA BENCHMARK ACT	AOAD/M_T
YELL	62 28 51.24	-114 28 50.52	AOA BENCHMARK ACT	AOAD/M_T

Table A.2 WCDA Network

Station	Location		Receiver Type	Antenna Type
	Latitude (deg min sec)	Longitude (deg min sec)		
BCOV	50 32 39.48	-131 50 33.36	LEICA CRS1000	LEIAT504
ELIZ	49 52 22.80	-132 07 21.72	LEICA CRS1000	LEIAT504
HOLB	50 38 25.44	-128 08 06.00	LEICA CRS1000	LEIAT504
UCLU	48 55 32.16	-130 32 29.76	LEICA CRS1000	LEIAT504
WSLR	50 07 35.40	-122 55 16.32	AOA SNR-8000 ACT	AOAD/M_T

Stations ALBH, DRAO and WILL are included in both CACS and WCDA network.

APPENDIX B

DERIVATION OF IONOSPHERE PIERCE POINT

The coordinate of the ionosphere pierce point (IPP) can be derived if the coordinates of the receiver and satellite are available (Skone, 2002). Figure B.1 shows the geometry of the derivation. Here the thin shell is assumed to be located at a fixed altitude (nominally 350km).
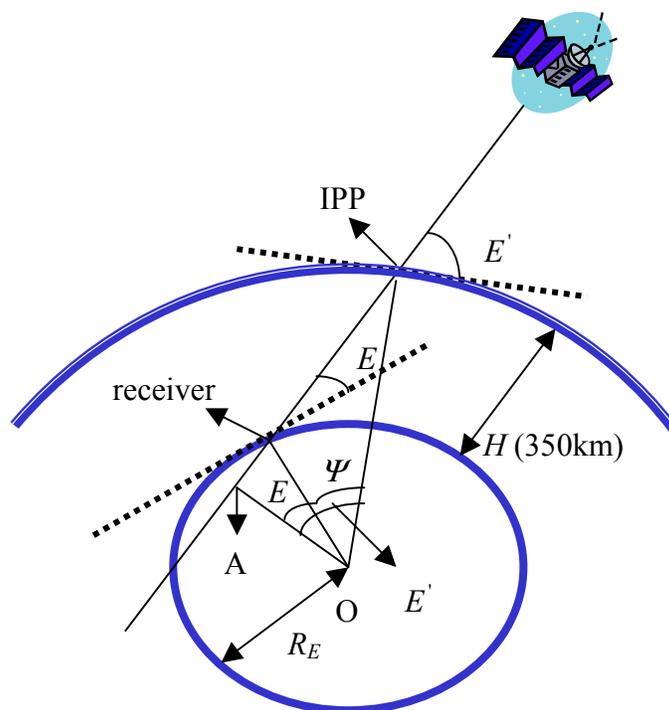
Figure B.1 Geometry of ionosphere pierce point derivation

In the figure, E and E' is elevation angles at the user receiver and IPP. The angle Ψ, which is the offset angle between the IPP and receiver, is calculated as follows:

$$\Psi = E' - E \tag{B.1}$$

From the geometry in the figure, the following relations is satisfied:

$$OA = R_E \cos E = (R_E + H)\cos E^{'} \tag{B-2}$$

From equation B-2, the elevation angle $E^{'}$ is derived as follows:

$$E^{'} = \cos^{-1}((\frac{R_E}{R_E + H})\cos E) \tag{B-3}$$

where R_E is the mean radius of a spherical earth, and H is the height of IPP.

Using equation B-3, equation B-1 can be rewritten as

$$\Psi = \cos^{-1}((\frac{R_E}{R_E + H})\cos E) - E \tag{B-4}$$

If the receiver's coordinate (ϕ_r, λ_r) is provided, then the coordinate of the IPP ($\phi_{IPP}, \lambda_{IPP}$) is derived as follows:

$$\phi_{IPP} = \phi_r + \Psi \cos A \tag{B-5}$$

$$\lambda_{IPP} = \lambda_r + \frac{\Psi \sin A}{\cos \phi_{IPP}} \tag{B-6}$$

where A is the azimuth angle of the satellite.