Multifunctional Operation and Application of GPS

Edited by Rustam B. Rustamov and Arif M. Hashimov

MULTIFUNCTIONAL OPERATION AND APPLICATION OF GPS

Edited by **Rustam B. Rustamov**
and **Arif M. Hashimov**

Multifunctional Operation and Application of GPS
http://dx.doi.org/10.5772/intechopen.71221
Edited by Rustam B. Rustamov and Arif M. Hashimov

Contributors

Arul Elango, Venkatramanan C, Senthil Kumar B, Pavan Kumar Ch.V.M.S.N, Daniel Okoh, Yang Dongkai, Wang Feng, Mohamed Tamazin, Malek Karaim, Aboelmagd Noureldin, Takuya Kawabata, Yoshinori Shoji, Korkut Yeğin, Sridevi Jade, T. S. Shrungeshwara, Ahmed El-Mowafy, Davide Imparato, Mohamed Elsheikh

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

3,450+
Open access books available

110,000+
International authors and editors

115M+
Downloads

151
Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Meet the editors

Associate Prof. Dr. Rustam B. Rustamov was born on May 25, 1955, in Azerbaijan. He is a freelance expert on space science and technology. In the past he was in charge of Azerbaijan National Aerospace Agency activities as an Acting Director General. Rustam B. Rustamov graduated with a PhD at the Russian Physical-Technical Institute (St. Petersburg). He was invited to work at the European Space Agency within the Framework of the United Nations Program on Space Applications at the European Space Research and Technology Center, Netherlands. He is a Cambridge Scholars Publishing International Editorial Advisory Board member on Astronomy, Astrophysics and Space Science, London, UK. He is the author of 12 books published by the European and United States noted publishers and more than 120 scientific papers.

Arif M. Hashimov was born on September 28, 1949, in Shahbuz, Nakchivan, Azerbaijan. Since 1993 he has been Professor of Power Engineering and from 2007 he was made First Vice-President of the Azerbaijan Academy of Sciences. He is laureate of the Azerbaijan State Prize, an Honored Scientist of Azerbaijan, and co-chairman of the International Conferences "Technical and Physical Problems of Power Engineering" research area: theory of non-linear electrical chains with distributed parameters, neutral earthing and ferroresonance processes, alternative energy sources, high-voltage physics and technology, and electro physics. He is the author of 280 articles, one patent, and five monographs.

Contents

Preface

Global Positioning System (GPS) is a satellite constellation expanded for the use of navigation, positioning of the objects on Earth or outer space, and timing. There is a wide range of applications of the use of GPS within the above framework, e.g., land surveying, cartography/geodesy/mapping, land transportation, telecommunications, scientific research activities, and many others.

This book addresses the general aspects of GPS applications with a reflection of the fundamental bases of the technology structure, segments, and elements. It has been demonstrated that satellite systems use orbits for this purpose, and the facilities of receivers for gathering and collection of required data.

There is no doubt that it is always challenging to provide detailed classification of information using advanced methods and technologies to embrace expectations in processing. In this book the use of computer neural networks with the integration of GPS data is demonstrated for ionosphere modeling. This has been performed by applying GPS technology to study the ionosphere based on the development of artificial neural networks.

An assessment of GPS Total Electron Content (TEC) covers the spatial equatorial ionization anomaly (EIA) region with two solar cycles. It reflects ionosphere variability in space, time, and geographical location. It has been presented by different phases of the solar cycle, EIA, and annual, daily, diurnal, and seasonal variability of the ionosphere in the Northern Hemisphere. This book demonstrates comprehensive research into high and random variability of TEC associated with the changes in solar activity, intensity of the sun's radiation, zenith angle at which they impinge Earth's atmosphere, equatorial electrojet, and plasma flow.

One more aspect of GPS is the successful selection and application of appropriate software utilization. This depends on what kind of task and data need to be solved and collected for expected outcomes within conducted measurements. It is a vital option and a subject of GPS technology, which has been studied and undertaken in this book.

The book estimates the error originating at the receiver due to multiple paths taken by the satellite transmitted radiofrequency (RF) signal. It is a very important subject in urban areas, and is the major error among other GPS error sources taking place at the receiver. It has been calculated as a multipath error using code range, carrier phase range, and carrier frequencies. The proposed algorithm considers the random nature of the multipath error and avoids complex calculations involved in the error calculations. Based on such an approach a valuable aid in precise navigation, surveying, and ground-based geodetic studies has been achieved.

This book also presents review materials dedicated to the Global Navigation Satellite System (GNSS) with a description of its structure control, space, and use elements. In the mean-

time, other existing facilities for navigation purposes such as the Russian GLONASS, its signal characteristics, and described modernization program are demonstrated.

Obviously, the vital option of GNSS and the main segment of the system is the antenna needed to achieve low-height, low-cost, and relatively good narrowband performance. It is presented as a ceramic antenna patch that meets indicated requirements. The book provides an analysis of a considered antenna system with existing counterparts. There is one more option related to the appropriate software application during data processing. Also presented are front-end architectures from a traditional super-heterodyne to zero/low-intermediate frequency configuration.

The most serious reasons affecting the accuracy of GNSS are also discussed. It is obvious that the circumstance of signal error needs to be undertaken and avoided for the achievement of a high quality of measurement. It drives the need to define and point out the nature of segments negatively facilitating the resulting information.

The methods RAM and advanced receiver autonomous integrity monitoring have been applied to land applications for detection and exclusion of faults and to achieve an alarm system in case of unsafe object positioning. An integrity monitoring system is used, which is generally employed in aviation. This book demonstrates integrity monitoring for land applications.

It is desirable to use facilities with a multipurposeful task. There is an effort to use GNSS for Earth study, which has a vital place when used for remote sensing with the development of geographical information system (GIS). Attempts have been demonstrated to use GNSS to monitor Earth's physical parameters, such as sea wind speed, sea surface height, sea ice, and soil moisture with application of GNSS Reflectometry (GNSS-R).

There is one more application of GNSS presented in this book. It has been demonstrated that GNSS can fix the signals from thunderstorms or rainfalls. The approach of weather prediction with the use of vapor distribution data collected by GNSS has a significant place in the forecasting of rainfall.

In the meantime, I am pleased to note that it was a pleasure to work with colleagues from around the word who achieved all expectations during the publication stages of this book.

I would like to express my personal gratitude to the book co-editor and editor assistant for kind support and encouragement in all aspects of book development.

It is an honor to take the opportunity to thank personally to Mr. Julian Virag, Publishing Process Manager—Author Acquisitions, for exceptional endorsement and support at all stages of book publication.

Assoc. Prof. Dr. Rustam B. Rustamov
Freelance expert on space science and technology
Azerbaijan

Navigation System – Advances of Software Application

Review on Sparse-Based Multipath Estimation and Mitigation: Intense Solution to Counteract the Effects in Software GPS Receivers

Ganapathy Arul Elango, B. Senthil Kumar,
Ch.V.M.S.N. Pavan Kumar and C. Venkatramanan

Additional information is available at the end of the chapter

Abstract

Multipath is the major concern in GPS receivers that fade the actual GPS signal causes positioning error up to 10 m so special care need to be taken to mitigate the multipath effects. Numerous methods like hardware based antenna arrays technique, receiver based narrow correlator receiver, double -delta discriminator, Adaptive Multipath Estimator, Wavelet Transformation and Particle filter, Kalman filter based post receiver methods etc. used to resolve the problem. But some of the methods can only reduce code multipath error but not effective in eliminating carrier multipath error. Most of these techniques are based on the assumption that the Line-of-Sight (LOS) signal is stronger than the Non-Line of-Sight (NLOS) signals. However, in the scenarios where the LOS signal is weaker than the composite multipath signal, this approach may result in a bias in code tracking. In this chapter, different types of multipath mitigation and its limitation are described. The recent development in sparse signal processing based blind channel estimation is investigated to compensate the multipath error. The Rayleigh and Rician fading model with different multipath parameters are simulated to test the urban scenario. The inverse problem of finding the GPS signal is addressed based on the deconvolution approach. To solve linear inverse problems, the suitable kind of appropriate objective function has been formulated to find the signal of interest. By exploiting this methods, the signal is observed and the carrier and code tracking loop parameters are computed with minimal error.

Keywords: GPS, multipath, Rayleigh, Rician, sparse, de-convolution, ISTA

1. Introduction

Global Positioning System (GPS) is a satellite based navigation system designed and developed by US Department of Defense (DoD) to provide instantaneous 3D position, velocity and time information almost anywhere on or above the surface of the earth at any time, and in any weather. The GPS receiver receives right-hand circularly polarized signals from minimum four satellites to find the user position. The commercial GPS receivers operate at L1 (1575.42 MHz) and L2 (1227.6 MHz) which are modulated on to 50-bps data stream [1]. The positional accuracy provided by GPS is deteriorated by various errors originating at the satellites, Clock error, Ephemeris, Ionospheric, Tropospheric, orbital errors, satellite clock errors, Selective Availability, Receiver Noise and multipath errors. With the use of differential techniques it is possible to remove many of the common-mode error sources, but the error effects of multipath have proven much more difficult to mitigate.

Multipath effect is one of the prominent problems in Wireless communication environment that effects in radio signals reaching the receiving antenna by two or more versions of the transmitted signal arrive at the receiver at slightly different times cause severe degradation in signal reception. Multipath propagation occurs in GPS receivers caused by reflection, refraction, atmospheric ducting, and reflection from nearby objects, water bodies, other reflecting surfaces etc. [2]. The reflecting surface may be buildings, hills, ground, water, or any object that happens to be a radio reflector. The Multipath error result when the receiver receives the direct or line-of-sight (LOS) satellite signal via multiple paths that can be constructively or destructively combined at the receiver antenna to give a resultant signal which can vary widely in amplitude and phase, depending on the distribution of the intensity and relative propagation time of the waves and bandwidth of the transmitted signal. A generic multipath propagation scenario diagram is shown in **Figure 1**.

1.1. Overview of GPS signal

The nominal signal strength of a GPS signal would be around 45–55 dB-Hz. The GPS signal power level lies approximately 15 dB under the noise background level. The GPS falls in the category of spread spectrum signal having processing gain of 45 dB. By consequence, if an interfering signal is introduced in the receiver with power 45 dB higher than the noise floor, then the receiver is completely jammed [3]. Interference signals may be in the form of Narrowband and wideband interference. Narrowband can be modeled as continuous wave or pulsed interference at a specified frequency that can be characterized by a pulse duty cycle. Similarly the wideband interference can be modeled as additive white Gaussian noise having flat power spectral density over a wide range of frequencies.

In case, the GPS signal is severely degraded due to Multipath, the signal should be carefully processed by different Multipath algorithms to counteract the effect of diffraction, scattering, Reflection, Refraction, Shadowing etc. The spectrum of the undisturbed (noise free) GPS signal is plotted for a sampling frequency of 5.714 MHz with an IF frequency of 1.6205 MHz in **Figure 2**.
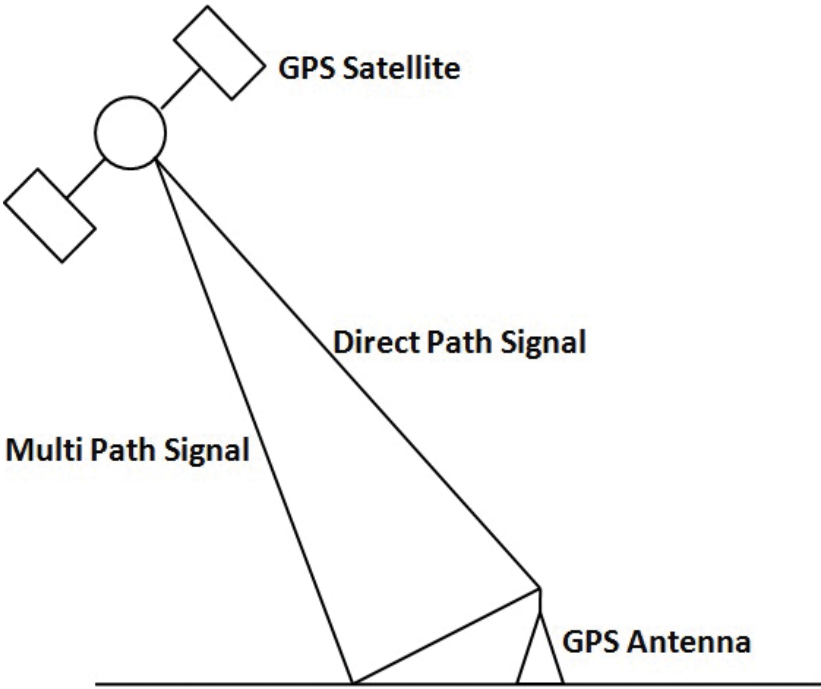
Figure 1. Typical multipath scenario.
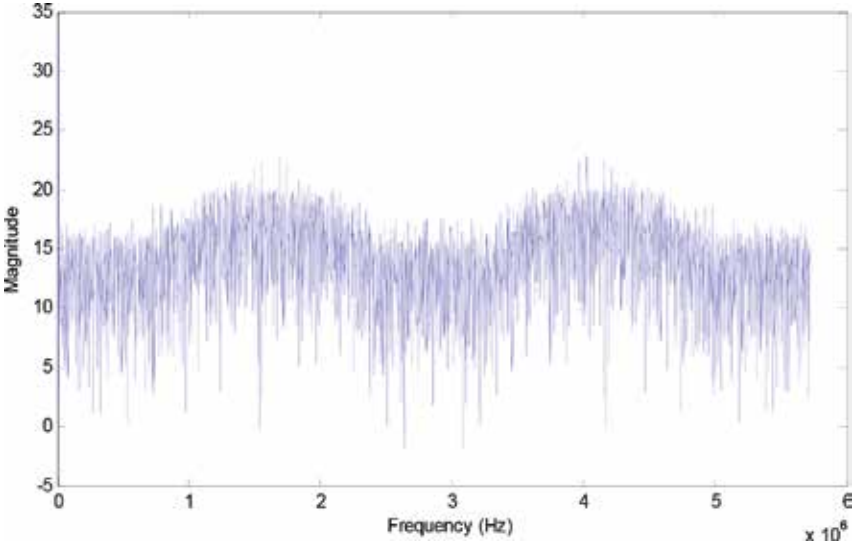
Figure 2. Spectrum of GPS signal with nominal power level.

1.2. Overview of multipath channel model

In mobile (outdoor) radio channels, the Rician distribution is commonly used to describe the statistical time varying nature of the received envelope of a flat fading signal, or the envelope of an individual multipath component. In the channel model, incoming signal is delayed due to different types of obstacles and reached at the receiver side with different time delays with attenuated amplitude and change in phase for each path is shown in **Figure 3(a)**. The complex baseband received signal is given by

$$y(t) = \sum_{i=0}^{L-1} \alpha_i x(t - \tau_i) e^{-j2\pi f_c \tau_i} \tag{1}$$

α_i-attenuation in amplitude, τ_i- phase.

As a signal is transmitted, a series of attenuated and delayed versions of the original signal is received leading to a typical multipath channel response. Furthermore, this channel response changes over time.

On the other hand, the indoor propagation channels are characterized by severe multipath propagation. In past two decades, classical Jakes fading model is widely used. In the Jake's Doppler spectrum, the receiver (or transmitter) is assumed to move at certain speed to model the Rayleigh channel. However, in fixed wireless communication systems, both the transmitter and the receiver are stationary and time-variations are actually due to moving scatterers. Filtering White Gaussian Noise (FWGN), AR Model and Sum of Sinusoidal (SOS) exhibits the property of the Rayleigh model [4]. The typical FWGN and AR Model are shown in **Figure 3(b)** and **(c)**.

From **Figure 4(a)**, the faded envelope is obtained by considering the carrier frequency of 1.6205 MHz and number of multipath components N = 10, the Doppler value is kept around
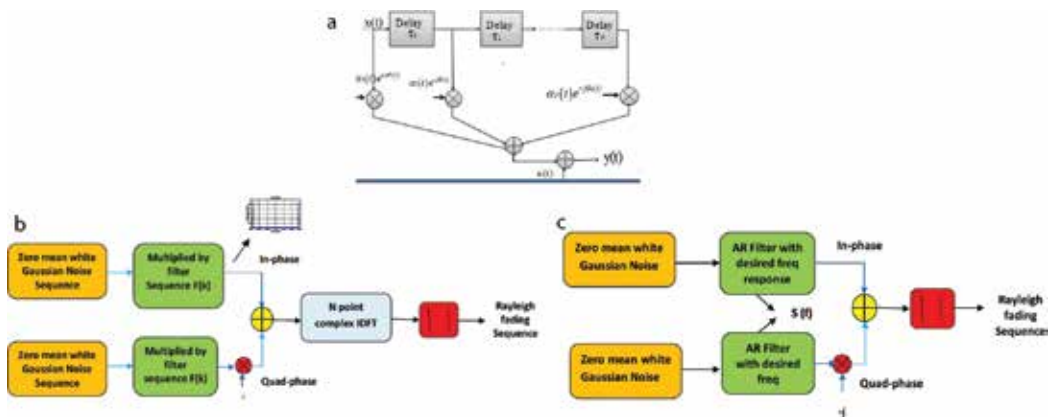
Figure 3. (a) Tapped delay line Multipath Channel model. (b) IDFT model fading simulators. (c) AR model fading simulators.

(slow moving receiver) 500 Hz and it is observed that the fading severely degrades the amplitude of the GPS signal and mean variation of the signal is around 0.22 with respect to time when compared to the actual value. To show the effectiveness of the fading the spectrum of the faded signal with SNR of −15 dB is plotted in **Figure 4(b)**. A code discriminator in a tracking loop is used to estimate the arrival time of the satellite code. The discriminator function, which is known as the S-curve as shown in **Figure 5** and it is given by

$$D\left(\tau\right) = R_E - R_L \tag{2}$$

where τ is the time of the reference signal. R_E and R_L are the samples of the early and late correlation functions respectively. The estimated arrival time is the time at which the discriminator is zero. However, in the presence of multipath signals, the autocorrelation function (ACF) will be distorted so that the discriminator will fail to estimate the true arriving time, resulting in pseudo-range estimation error. Due to multipath, the ideal triangular function
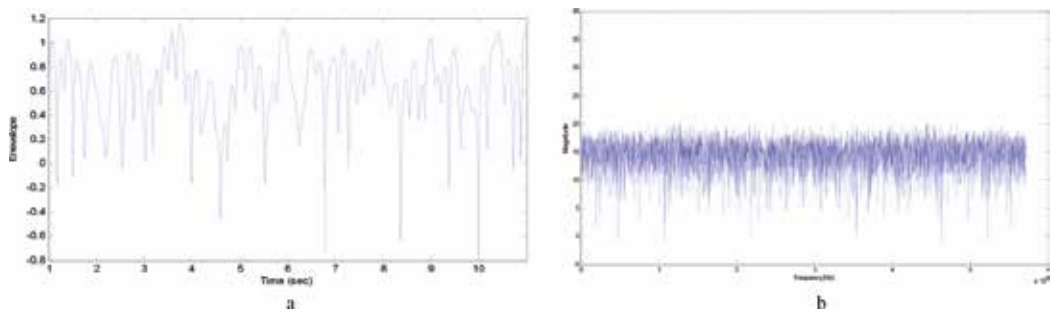
Figure 4. (a) Faded envelope. (b) Spectrum of distorted GPS signal with SNR -15 dB.
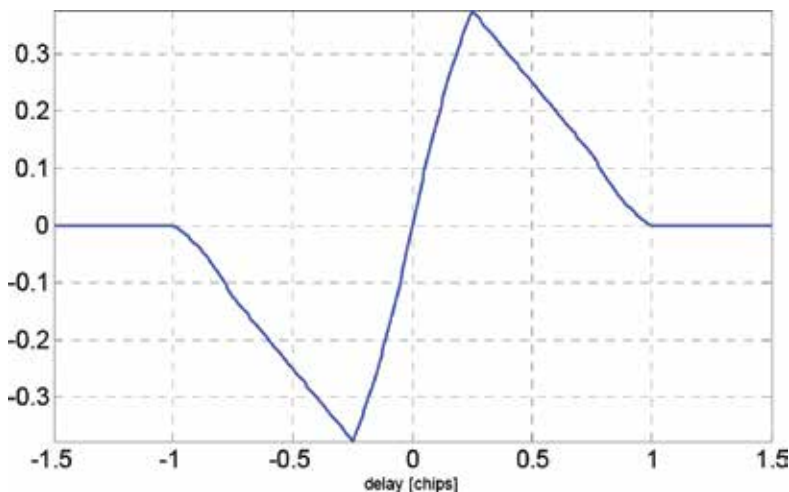
Figure 5. Multipath free discriminator function (s curve) used in tracking stage.
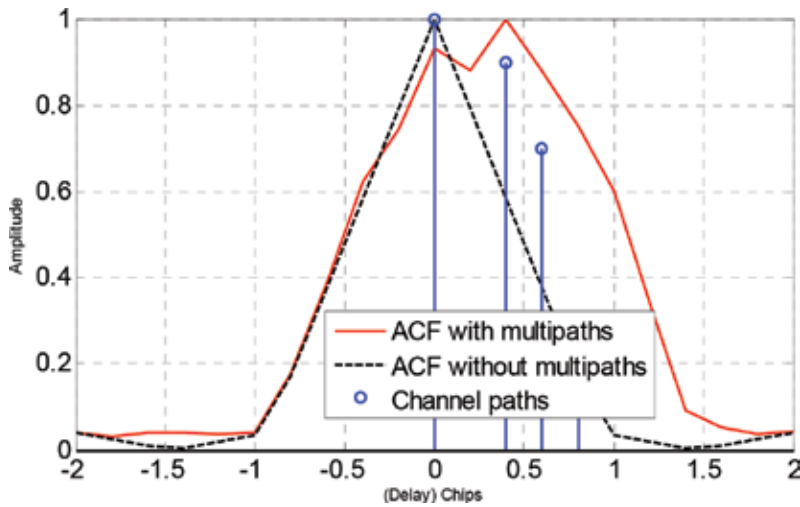
Figure 6. Auto correlation plot with and without multipath component.

loses its symmetry. The distortion of the correlation function is illustrated in **Figure 6**. When one path is a LOS receives at the receiver which is in phase with the direct signal that follows the ideal triangular shape but in the case of composite 10 path scenario, the correlation function is distorted by 0.4 chips as denoted in the dotted line.

2. Existing multipath mitigation techniques

To alleviate multipath problem several pre-filtering and post correlation based methods are introduced. In Many literatures, the problem of multipath is treated inside i.e. signal processing chain of the receiver especially at the stage of tracking and also before signal arrives at the RF front end i.e., at the antenna side, on the other hand some methods describe the effect of multipath is reduced even in the position calculation stage.

2.1. Miscellaneous methods

The sparse channel estimation can be estimated by using any one of the estimation techniques like sparse like blind channel estimation or least square based estimation. Once impulse response of the channel is estimated the inverse filter (channel equalizer) is designed to compensate the multipath error. The equalizer output is a delayed version of an impulse response positioned anywhere on the time, finally the LOS signal is observed by subtracting the strongest component from the composite signal. This method combined both estimation and mitigation techniques which is used to compensate the code and carrier tracking error. Some of the other methods deal with mitigation is also given in this section.

2.2. Antenna-based mitigation

Microstrip antennas are frequently used antenna type in GPS receivers because of its added advantage for airborne application, materialization of GPS receiver and easy construction. However, for geodetic needs, antennas are designed to receive both carrier frequencies L1 and L2 [5]. Also they are protected against multipath by extra ground planes or by using choke rings. A choke ring consists of strips of conductor which are concentric with the vertical axis of the antenna and connected to the ground plate which in turns reduces the multipath effect [5]. This involves improving the gain pattern of antenna to counter the effects of multipath. These antenna-based methods include the use of special antennas, processing in spatial domain with multi-antenna arrays, antenna location strategies and long-term signal observation for inferring multipath parameters [6, 7]. The circularly polarized antenna facilitates the rejection of multipath signals.

2.3. Receiver code tracking loop

The methods include all receiver technologies that are used to mitigate multipath. Usage of Narrow Correlators, Multipath Elimination Technique (MET), Edge Correlator, Strobe Correlator, and Multipath Estimation Delay Lock Loop (MEDLL) and simulation of multipath error in DLL [16] are some of the examples under this category and they will be discussed in detail this section. These techniques, however, are not very effective for short delay multipath [17], due to close-by reflectors. These methods cannot be operated in conjunction with all existing receivers and would need manipulation at the receiver hardware end to work. This remains as one of the major issues with using receiver related techniques [7].

a. Early-minus-late delay lock loop: GPS receiver uses classical correlation-based code tracking structure based on a feedback delay estimator implemented via a feedback loop. The well-known feedback delay estimator is the Early-Minus-Late (EML) DLL, where two correlators spaced one chip apart are used in the receiver in order to form a discriminator function, whose zero crossings determine the path delays of the received signal [8]. The classical EML usually fails to cope with multipath propagation. Therefore, several enhanced EML-based techniques have been introduced in the literature for last two decades in order to mitigate the impact of multipath, especially in closely spaced path scenarios. A first approach to reduce the influences of code multipath is based on the idea of narrowing the spacing between the early and late correlators, i.e., nEML or narrow correlator spacing depends on the receiver's available front-end bandwidth along with the associated sampling frequency.

b. Adaptive Filtering: Multipath Mitigation in GPS/Galileo Receivers with different Signal Processing Techniques has been introduced by Benachenhou et al. [10] efficiently minimize the code and carrier tracking error. Yedukondalu et al. [18] used an adaptive filtering method of estimation and mitigation of Multipath interference in GPS receivers. In this chapter, to estimate the effect of multipath interference at the receiver antenna, a technique based on both code and carrier phase measurements using Code minus Carrier (CMC), is carried out to mitigate multipath for static applications. Different adaptive filters using

algorithms such as Least Mean Squares (LMS) and various Recursive Least Squares (RLS) are considered to mitigate the error [12]. The estimated multipath error for a typical signal is 0.8 and 2.1 m on L1 and L2 carriers, respectively.

2.4. Other filtering methods

In a simulated multipath environment, the reflection geometry is used in combination with a special GPS antenna arrangement to detect and track multipath. In the highly non stationary environment, Researchers also used Kalman Filter, particle filters and multiple differential GPS receivers to remove multipath errors in final positioning [13]. Code multipath is calibrated and estimated using spherical harmonics in static applications, similarly for kinematic applications, the multipath error mitigation is carried out by Mozaviet et al. [14] using wavelet transform. The estimation of frequency components of multipath error signal using spectral analysis and its effective mitigation using time varying digital filters are designed by Yedukondalu et al. [11]. The four types of filters, namely, Butterworth, Type I and II Chebyshev and Elliptic filters, are examined for mitigation of multipath and their performance are compared. It is observed that by applying digital filters of different cut-off frequencies over the spectrum of the multipath, one can significantly reduce the multipath errors. It was found that Butterworth filter reduced the error most effectively.

3. Works related to sparse signal processing-based multipath mitigation

Sparse Signal Processing otherwise known as compressive Sensing (CS), is a classical Signal Processing technique efficiently acquiring and reconstructing a signal completely from reduced number of measurements, by exploiting its compressibility. CS has become a very interesting research area in recent years due to its theoretical and practical utility to capture a wide range of signals at a rate significantly lower than the Nyquist rate representing signal with lesser number of coefficients.

Optimal demodulation and decoding in wireless communication systems often requires accurate knowledge of the channel impulse response. Typically, this is accomplished by searching the channel with a known training sequence and linearly processing the channel with sparse impulse response. On the other hand, conventional linear channel estimation schemes, such as the least-squares method, fail to take advantage of on the anticipated sparsity of the channels. In contrast, it is observed that a CS channel estimate obtained as a solution significantly outperforms a least-squares based channel estimate in terms of the mean squared error (MSE) when it comes to learning sparse (or approximately sparse) channels.

This section highlights some of the most prominent state-of-the-art techniques, which have gained a lot of interest in the research community. The Critical review of literature indicates that exhaustive research has been done by several researchers to develop techniques to improve the performance of software GPS receivers under multipath environment. Researchers have

concentrated the methods based on compressive sensing implemented in software based GPS receivers for accurate undisturbed reception and positioning.

Dragunas and Borre et al. [19] proposed the sparse deconvolution based Projection onto Convex Sets (POCS) method which is used to mitigate the multipath in indoor environments. The author compared the several multipath mitigation techniques suitable for the indoor environments. By using the proposed method the author chooses one of the secondary paths as LOS signal. In this method, the author achieves better resolution than the conventional methods. An extension to this work, Dragunas et al. [20] presented a modified Projection onto Convex Sets (POCS) that optimizes the Coarse/Acquisition codes employed in Global Positioning Systems. The author deals with the problem of joint LOS code delay and carrier phase estimation of GPS signals in a multipath environment. The modified POCS algorithm acts as the most resistant in closely-spaced multipath static channels both when LOS code delay and carrier phase estimation are concerned. Another sparse based modified iterative Projection onto convex sets (POCS) method proposed by Negin Sokhandan and Ali Broudman [21] is used to reduce the multipath error in harsh environment. The algorithm estimates the channel impulse response (CIR) and removes the spurious noise peaks at each iteration. This method is carried out to estimate the LOS time of arrival from the position of its first non-zero element that passes a certain threshold. The modified POCS algorithm correctly estimates the code delay and carrier phase for GPS signals with few iterations. Hence, faster performance has been achieved when compared to conventional POCS.

Kumar and Lau et al. [22] implemented the deconvolution approach for the code phase and carrier phase estimation. The deconvolution approach shows that it is very different from POCS approach where each path can be estimated. The deconvolution approach can accurately estimate the Line of Sight (LOS) signal. Initially the channel impulse response is computed and by getting the deconvolution filter coefficients, multipath can be removed by convolving the measurements with deconvolution filter coefficients and the code and carrier phase can be estimated and finally the LOS is found.

The novel sparse reconstruction method for mitigating the multipath induced code delay estimation has been implemented by Fei and Liao et al. [23] in GPS receivers. The author exploited to enhance the direct signal without affecting the accuracy of the GPS code delay estimates. The coherent accumulation of received GPS signals and by transforming it into frequency domain and parameters of multipath signals are estimated by sparse reconstruction algorithm. The author estimates the code delay without affecting the accuracy of the GPS by sparse reconstruction method. Tian and Li et al. [24] proposed a novel method based on nonnegative matrix factorization (NMF) spectral unmixing for land seismic additive random noise attenuation. In this method, the noisy seismic signal is first decomposed into a collection of intrinsic mode functions (IMFs) instead of being directly processed. Then, a sparse NMF is used to unmix the STFT spectrum of each IMF. By separating the sub-spectrums by the inverse STFT, the sub-signals can be easily acquired. Finally, the desired signal is reconstructed from the sub-signals by K-means clustering algorithm. Bostan and Kamilov et al. [25] proposed a novel statistically-based discretization paradigm and derive a class of maximum a posterior (MAP) estimators for solving ill-conditioned linear inverse problems. It proposes the theory of

sparse stochastic processes, which specifies the continuous –domain signals as solutions of linear stochastic differential equations. It provides the algorithm that handles the nonconvex problems and by applying it to the reconstruction algorithm and finally compares the performance of estimators, associated with the models of increasing sparsity.

Broumandan and Lin et al. [15] established a way to enhance the performance of GNSS time arrival estimation techniques in multipath environments by determining the multipath channel estimation using equivalent discrete-time linear time-invariant system method which is modeled as a Moving Average system. It modeled the multipath channel as a sparse channel by describing the number of parameters of the channel is less than the number of unknowns in the Moving Average model. The author compares the performance of the sparse estimation with the Cramer-Rao Lower Bound (CRLB) of the parameter estimation problem and the least square estimate. It provides the better sparse signal recovery method to estimate the channel impulse response.

3.1. Sparse signal deconvolution

Sparse de-convolution finds variety of application in accurate estimation of multipath channels with sparse impulse response of a channel is calculated by degradation version of convolution matrix. After down conversion to baseband, the signal from all the satellites can be represented in complex baseband representation as.

$$y(t) = \sum_{s=0}^{S-1} \alpha_s \, c_s\left(t - \tau\right) e^{j2\pi(f_c + f_d)t} \tag{3}$$

where $\alpha^{(s)}$ is the channel attenuation from the sth satellite to the receiver, τ is the time delay or code phase of the C/A code and f_d is the Doppler frequency for the sth satellite.

We assume that the observed GPS signal y from a multipath channel can be written as

$$y = Hx + n \tag{4}$$

where x is the signal of interest which is to be estimated, n is additive noise, and H is a matrix representing the degradation process. The estimation of actual GPS signal x from the faded version y can be treated as a linear inverse problem. An appropriate objective function, J(x) has been formulated to solve linear inverse problems and to find the signal x, by minimizing J(x).

Generally, the chosen objective function is the sum of two terms:

$$J(x) = D(y, H_x) + \lambda R(x) \tag{5}$$

where.

D(y, H_x) measures the discrepancy between y and x.

R(x)—Regularization term (or penalty function).

λ—Regularization parameter (positive value).

To find a signal x, so that H_x is very similar to y, i.e., here it is needed to find a signal x which is consistent with the observed data y. For D (y, H_x), the mean square error can be calculated as

$$D(y, H_x) = \|y - H_x\|^2_2 \tag{6}$$

The squared error between y and H_x is minimized by finding the norm difference of D (y, H_x) that will give a signal x, which is as consistent with y as possible, according to the square error criterion. To minimize D (y, H_x) by setting x = H^{-1}y; however, H may not be invertible. Let convolution filter be {1,-1, 1,-1....M} and signal be of length M. Convolution sum will have length equal to N + M-1. So H in this case will have N × M-1dimension

$$H = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & \vdots \\ \vdots & 0 & -1 & 1 & 0 \\ 0 & \vdots & \cdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}$$

which is definitely invertible. Even if H were invertible, it may be very ill-conditioned, in which case, this solution amplifies the noise, sometimes to such an extent that the solution is useless. The role of the regularization term R(x) is exactly to address this problem. The regularizer R(x) should be chosen so as to penalize the undesirable/unwanted behavior in x.

3.1.1. L1-norm regularized linear inverse problem

By assuming that the GPS signal of interest after acquisition x, is known to be sparse. i.e., x has relatively few non-zero values, i.e., x consists of a few impulses and is otherwise zero. In this case, the R(x) may be defined to be the number of non-zero values of x. R(x) is not a convex function of x, which is not differentiable then the objective function J(x) is very difficult to minimize and therefore J(x) will have many local minima. To minimize J(x), it is better to choose J(x) to be a convex function of x that measures sparsity, but which is also convex. For this reason, when x is known to be sparse, the regularization function R(x) is often chosen to be the L_1-norm. Hence, the approach is to estimate x from y by minimizing the objective function,

$$J(x) = \|y - Hx\|^2_2 + \lambda \|x\|_1 \tag{7}$$

3.1.2. Soft-thresholding algorithm (ISTA)

The requirement for development of fast algorithm is to minimize the equation and its related functions. This is carried out by another significant algorithm called iterated soft-thresholding algorithm (ISTA), also referred as Thresholded Landweber (TL) algorithm. ISTA is a combination of the Landweber algorithm and soft-thresholding. To minimize J(x), consider first the minimization of the simpler objective function

$$J(x) = \|y - Hx\|_2^2 = (y - Hx)^T(y - Hx) \tag{8}$$

$$J(x) = y^Ty - 2y^THx + x^TH^THx \tag{9}$$

Because J(x) in Eq. (8) is differentiable and convex, thus one can obtain its minimizer by setting the derivative with respect to x to zero. The derivative of J(x) is given by

$$\frac{\partial}{\partial x}J(x) = \text{-}2H^T + 2H^THx$$

Setting the derivative to zero gives a system of linear equations,

$$\frac{\partial}{\partial x}J(x) = 0 \text{ implies } (H^TH)x = H^Ty.$$

So the minimizer of J(x) in Eq. (9) is given by

$$x = (H^TH)^{-1}H^Ty \tag{10}$$

3.1.3. Majorization-minimization (MM) approach

However, it is not able to solve these equations easily. Since GPS data is a very long, then H will be very large matrix and solving the system of equations may require huge memory and computation time. Moreover, the matrix H^TH is not invertible, or ill-conditioned. By using the Majorization-minimization (MM) approach to minimize J(x) in Eq. (10), solving a system of linear equations can be avoided. At each iteration k of the MM approach, a function $G_k(x)$ that coincides with J(x) at x_k has been found. A majorizer $G_k(x)$ has introduced that can be minimized more easily without solving a system of Eqs.

A function $G_k(x)$ that majorizes J(x) by adding a non-negative function to J(x),

$$G_k(x) = J(x) + \text{Non} - \text{negative function of x} \tag{11}$$

When $G_k(x)$ coincides with J(x) at $x = x_k$, the non-negative function added to J(x) should be equal to zero at x_k then $G_k(x)$ to be

$$G_k(x) = J(x) + (x - x)^T(\alpha I - HH^T)(x - x_k) \tag{12}$$

The function which is added to J(x) is clearly zero at x_k so that $G_k(x)$ equals to $J(x_k)$ as required. To ensure the function added to J(x) is non-negative, for all x, the scalar parameter α must be chosen to be equal to or greater than the maximum eigenvalue of H^TH, i.e., $\alpha \geq$ max eig (H^TH). Then the matrix αI- H^TH is a positive semi-definite matrix, meaning that v^T (αI- H^TH) $v \geq 0$. Now, using MM procedure, to obtain x_{k+1}, function $G_k(x)$ is minimized. Expanding $G_k(x)$ in (12) gives

$$G_k(x) = y^Ty - 2y^THx + x^THx + (x - x_k)^T(\alpha I - H^TH)(x - x_k) \tag{13}$$

Note that the quadratic term in Eq. (12) is simply x^Tx instead of x^TH^THx. Therefore, we can minimize $G_k(x)$ more easily

$$\frac{\partial}{\partial x}G_k(x) = -2H^Ty - 2(\alpha I - H^TH)x_k + 2\alpha x, \text{Setting}, \frac{\partial}{\partial x}G_k(x) = 0$$

$$x = x_k + \frac{1}{\alpha}H^T(y - Hx_k) \tag{14}$$

Hence, by using MM procedure to obtain x value at each iteration is given by Landweber update equation as

$$x_{k+1} = x_k + \frac{1}{\alpha}H^T(y - Hx_k) \tag{15}$$

4. Results and discussion

In this simulation, four multipath components are considered with time varying amplitude and the phase. Initially the GPS signal needs to be framed in the form of sparse signal. This can be done in the acquisition stage only, the sparse representation of this signal easily decomposed in the form of basis function and the coefficient term. Then one can easily reconstruct the sparse coefficient of minimum number of non-zero coefficient by random by l_1 minimization. The code and carrier tracking loop of the software GPS receiver has to be synchronized if and only if the lock is achieved. Due to multipath error, the code loop error may be varied more than 1 chip delay and the carrier loop (Costas) is also intercoupled with this, so error may be introduced in the carrier tracking loop also hence, both the tracking errors should be carefully minimized to certain extent to achieve the lock.

The objective function (J) and the 2000 samples of the recovered GPS signal after ISTA algorithm is plotted in **Figures 7** and **8** respectively. The recovered signal is further given to the acquisition stage to find the visible satellites (SVN's) and allocate those SVN to initiate the tracking stage. The code and carrier tracking error is observed after recovering the GPS signal using MM method. The significant improvement in carrier tracking is achieved within 50 msec
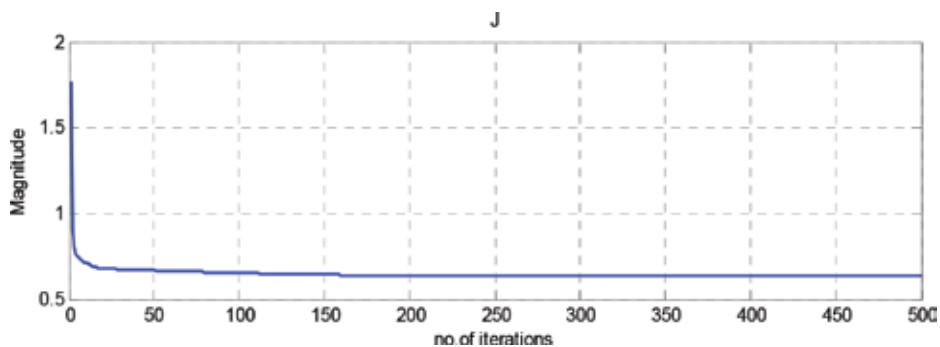
Figure 7. Objective function.
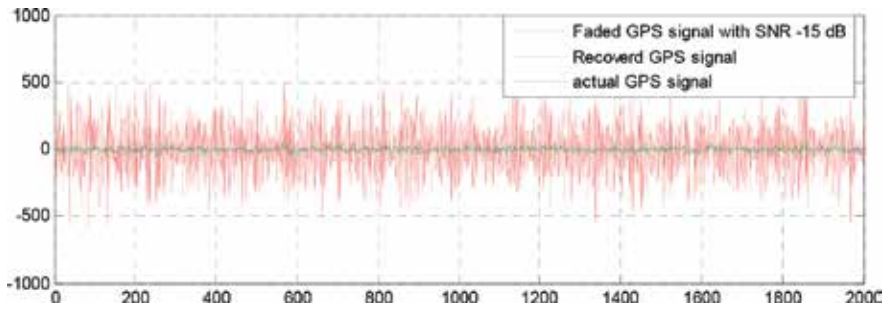
Figure 8. Recovered GPS signal through ISTA algorithm.
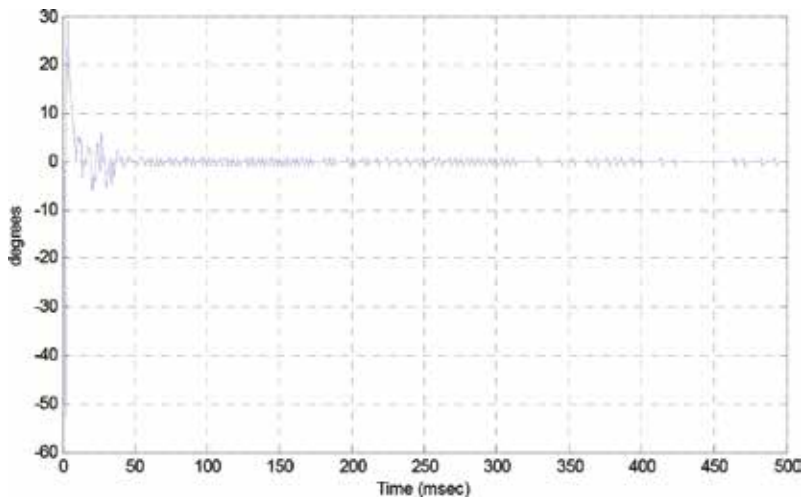
Figure 9. Carrier loop tracking error for SVN-12.
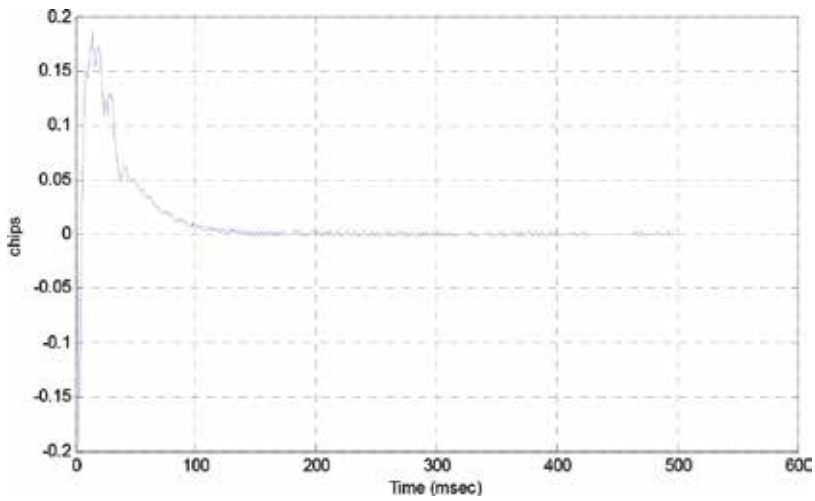
Figure 10. Code loop tracking error for SVN-12.

period where as in the case of code tracking, error is settled down quickly within 100 msec of GPS data as shown in **Figures 9** and **10**. The lock has been achieved with 0.5 chips spacing of early, late and prompt code replicas, hence the navigation data can easily be demodulated and the pseudorange is calculated for each satellite.

5. Conclusion

In this Chapter, various techniques have been discussed to nullify the effect of the multipath, we have provided an in depth review of existing multipath mitigation techniques. These techniques were classified in categories according to the involved process before and after correlation with the C/A code. Compressive sensing is a promoting tool for the next generation communication systems. However, it still faces a number of challenges in the real time implementation. In multipath applications, compressive sensing exploits the GPS signal need to be converted to a sparse equivalent structure then the channel impulse response of the filter is determined from the convolution matrix. For reconstruction, the challenge resides in how to separate the LOS signal from composite signals in multi-channel environments, where the channel powers and behaviors evolve over time. A comparison of several compressive techniques was given and discussed. The sparse recovery of the signal is obtained from unconstrained optimization algorithms.

Author details

Ganapathy Arul Elango*, B. Senthil Kumar, Ch.V.M.S.N. Pavan Kumar and C. Venkatramanan

*Address all correspondence to: arulelango2012@gmail.com

Department of Electronics and Communication Engineering, Sree Vidyanikethan Engineering College, Tirupathi, India

References

[1] Michel SB, Van Dierendonck AJ. GPS receiver architecture and measurements. Proceedings of the IEEE. 1999;**87**:48-64

[2] Borre K, Akos DM, Bertelsen N, Rinder P, Jensen SH. A software defined GPS and Galileo receiver. New York: Birkhäuser Bostonin, Springer Science & Business Media; 2007. ISBN-10 978–0–8176-4390-4

[3] Wildemeersch M. Fortuny-Guasch J. Radio Frequency Interference Impact Assessment on Global Navigation Satellite Systems. Security Technology Assessment Unit, EC Joint Research Centre. In: EUR 24242 EN. 2010

[4] Arul Elango G, Sudha GF. Design of complete software GPS signal simulator with low complexity and precise multipath channel model. Journal of Electrical Systems and Information Technology. 2016;**3**(2):161-180

[5] Raju PLN. Fundamentals of GPS, satellite remote sensing and GIS applications in agricultural meteorology. In: Proceedings of a Training Workshop; 7–11 July 2003; Dehradun. pp. 121-150

[6] Tranquilla JM, Carr JP, Al-Rizzo H. Analysis of a choke ring ground plane for multipath control in global positioning system (GPS) applications. IEEE Transactions on Antennas and Propagation; **42**(7):905-991

[7] Vagle N, Broumandan A, Jafarnia-Jahromi A, Lachapelle G. Mitigation using antenna arrays. The Journal of Global Positioning Systems. 2016;**14**:4

[8] Townsend BR, Fenton PC, Van Dierendonck KJ, van Nee DJR. L1 carrier phase multipath error reduction using MEDLL technology. In: Proceedings of ion GPS. Vol. 8. Institute of Navigation; 1995. pp. 1539-1544

[9] Betz JW, Kolodziejski KR. Extended theory of early-late code tracking for a bandlimited GPS receiver. Navigation, Journal of the Institute of Navigation. 2000;**47**(3):211-226

[10] Benachenhou K, Sari E, Hammadouche M. Multipath mitigation in Gps/Galileo receivers with different signal processing techniques. In: Proceedings of 5th International Conference: Sciences of Electronic Technologies of Information and Telecommunications; March 22–26; Tunisia; 2009. pp. 1-8

[11] Yedukondalu K, Sarma AD, Kumar A, Satyanarayana K. Spectral analysis and mitigation of GPS multipath error using digital filtering for static applications. IETE Journal of Research. 2013;**59**(2)

[12] Yedukondalu K, Sarma AD, Kumar A. Mitigation of GPS multipath error using recursive least squares adaptive filtering. IEEE Asia Pacific Conference on Circuits and Systems. Kuala Lumpur, Malaysia, 2010:104-107

[13] Bertozzi T, Le Ruyet D, Panazio C, Thien HV. Channel tracking using particle filtering in unresolvable multipath environments. EURASIP Journal on Advances in Signal Processing; **2004**(15):705208

[14] Mosavi MR, Azarbad MR. Multipath error mitigation based on wavelet transform in L1 GPS receivers for kinematic applications. AEU—International Journal of Electronics and Communications. 2013;**67**(10):875-888

[15] Broumandan A. Lin T. Performance of GNSS time of arrival estimation techniques in multipath environments. In: Proceedings of ION GNSS08. Savannah, GA; September; 2008:632-643

[16] Cheng L, Jie C, Gang X. Model and simulation of multipath error in DLL for GPS receiver. Chinese Journal of Electronics, 3. July, 2014;**23**

[17] Sleewaegen JM, Boon F. Mitigating short-delay multipath: A promising new technique. In: Proceedings of the International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS '01); September 2001; Salt Lake City, Utah, USA. 2001; pp. 204-213

[18] Yedukondalu, A. D. Sarma, And V. Satya Srinivas. Estimation and mitigation of GPS multipath interference using adaptive filtering. Progress in Electromagnetics Research M. 2011;**21**:133-148

[19] Dragunas K, Borre K. Multipath mitigation based on deconvolution. Journal of Global Positioning Systems. 2011;**10**(1):79-88

[20] Dragunas K. Indoor multipath mitigation. In: International Conference on Indoor Positioning and Indoor Navigation; September, 2010. pp. 578-584

[21] Sokhandan N, Broumandan A, Dehghanian V, Lachapelle G. GNSS multipath error reduction in harsh environments. In: ION proceedings of GNSS 2011; Portland; September; 2011: 20-23

[22] Kumar R, Lau K. Deconvolution approach to carrier and code multipath error elimination in high precision GPS. In: Proceedings of the 1996 National Technical Meeting of the Institute of Navigation. Santa Monica, CA; 1996:729-737

[23] Xiang F, Liao G, Zeng C, Wang W. A multipath mitigation discriminator for GPS receiver. International Journal of Electronics and Communications. October, 2013;**67**(10):839-847

[24] Tian Y, Li Y, Lin H, Ma H. A sparse NMF-SU for seismic random noise attenuation. IEEE Geosciences and Remote Sensing Letters. May, 2013;**10**(3)

[25] Bostan E, Kamilov US. Sparse stochastic processes and discretization of linear inverse problems. IEEE Transactions on Image Processing. July, 2013;**22**(7):2699-2710

GPS Technology Applications

Integrity Monitoring: From Airborne to Land Applications

Davide Imparato, Ahmed El-Mowafy and Chris Rizos

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75777

Abstract

Safety-critical applications in transportation require GNSS-based positioning with high levels of continuity, accuracy and integrity. The system needs to detect and exclude faults and to raise an alarm in the event of unsafe positioning. This capability is referred to as integrity monitoring (IM). While IM was considered until recently only in aviation, it is currently a key performance parameter in land applications, such as Intelligent Transport Systems (ITS). In this chapter the IM concepts, models and methods developed so far are compared. In particular, Fault Detection and Exclusion (FDE) and bounding of positioning errors methods borrowed from aviation (i.e. Weighted RAIM and ARAIM) are discussed in detail, in view of their possible adoption for land applications. Their strengths and limitations, and the modifications needed for application in the different context are highlighted. A practical demonstration of IM in ITS is presented.

Keywords: ITS, C-ITS, integrity monitoring, RAIM, FDE, SBAS

1. Introduction

Integrity is a key performance parameter in positioning for ITS safety applications [1, 2]. To provide absolute positioning in safety-critical and mission-critical applications, satellite navigation shall maintain a very high level of service. Correctness — within tight bounds — of the position solution, shall be guaranteed to extremely high levels of probability. In aviation, the risk for so-called Hazardous Misleading Information (HMI) due to the navigation system is typically budgeted at the 10^{-7} to 10^{-9} level, and a similar level of safety is expected to be required in land applications in the era of fully automated transportation. In more formal terms, integrity is about the trust that a user (or the AI in charge of the vehicle) can have in the indicated position information. The trust is measured by the probability of HMI (or

integrity risk), which is the probability that the position error exceeds a certain tolerance, without being detected and an alert being raised in time. The given position information will then be misleading, as it is not correct within specified bounds, and the user is not aware of the potentially hazardous situation.

While IM was considered until recently only in aviation, it is currently a key performance parameter in safety-critical land applications. Even though integrity requirements in vehicular transport have not been defined yet, the demand for higher levels of automation in an increasing number of applications is pushing the relevant authorities to fill this gap.

1.1. Integrity monitoring in aviation

Today, integrity monitoring in aviation is implemented in two different ways, at system level or at user level. At system level, two types of external augmentation systems can be distinguished, Space-Based Augmentation Systems (SBAS), see [3, 4], and Ground-Based Augmentation Systems (GBAS), see [5]. Both are Differential GPS systems (DGPS). SBAS and GBAS develop corrections that improve the accuracy of the measurements and generate real-time error bounds. These bounds are called Protection Levels (PL) and must exceed the actual error under all conditions with very high probability [6]. SBAS and GBAS are both very powerful means of guaranteeing integrity, but they present the drawback of needing a very complex and costly infrastructure.

At user level the GNSS integrity can be monitored by exploiting the redundancy of the GNSS signals as collected at the receiver. This is done by performing calculations within the user equipment itself to check the measurements consistency. This method is known as Receiver Autonomous Integrity Monitoring (RAIM). RAIM is possible as long as a number of observations larger than the minimum necessary for a position fix are available. RAIM strictly relies on the strength of the satellite geometry. With the deployment of the new GNSS constellations many more satellite signals will soon be available: this will increase the redundancy of measurements and the RAIM power.

1.2. Integrity monitoring on land

Both SBAS/GBAS and RAIM methods can in principle be adopted for IM in land applications, since the fundamental positioning problem is the same. However, some important differences in the applications may make the task not straightforward. GNSS positioning in aviation is generally restricted to Single Point Positioning (SPP), based on code observations on the civil frequencies, L1 (E1 for Galileo) and soon L5 (E5). With SPP, accuracy of few meters is attainable. However, most current and future land applications (such as ITS) require lane-level accuracy, i.e., sub-meter accuracy [7]. As such level of accuracy is considered unattainable with SPS, ITS applications are foreseen to be relying on Satellite Based Augmentation Systems (SBAS), RTK or Precise Point Positioning (PPP) techniques [7].

The different positioning methods and the corresponding higher precisions involved bring with them a new set of specific vulnerabilities. For instance, anomalies that would create positioning errors of too small magnitude in an SPP context, and could therefore been

neglected, would now need to be taken into account. In case carrier phase observations are to be used, cycle-slip monitoring shall be included, as well as IM for ambiguity resolution. Another difference from the aviation case that shall be taken into account is the environment in which positioning is to take place. Land users are often located in urban environment, which is characterized by the presence of high-rise buildings: as a result, GNSS observations are highly more likely to be affected by multipath and Non-Line-of-Sight (NLOS). Furthermore, the urban environment brings extra vulnerabilities linked to the higher risk of interference.

As the SBAS integrity monitoring concept has not been defined yet for ITS applications, this chapter focus is on the RAIM concept. This is in fact the most versatile integrity monitoring approach, generally applicable to any estimation problem. The chapter is organized as follows: in Section 2 the integrity as a navigation performance parameter is introduced and the focus moves to the RAIM approach. The RAIM problem is defined and the most important performance parameters of RAIM algorithms (PL, Probability of HMI, etc.) are introduced. In Section 3, a number of possible approaches to deal with the RAIM problem are introduced, whereas in Section 4 the most popular RAIM methods developed in aviation are described. In Section 5 the challenges related to the adaptation of current aviation RAIM methods to land applications are illustrated, and in Section 6 an example of preliminary results of an IM prototype method in ITS is shown. Finally, in Section 7 conclusions on the state-of-art in IM and directions of present and future work are given.

2. Integrity and RAIM

2.1. Navigation performance parameters

The navigation system's role is to collect and process measurements or other input data and deliver a position/state estimation, and guide the user to reach their destination. Based on the input data, called observables[1], the parameters of interest are estimated. In the GNSS case, the model for the estimation problem is non-linear, but it is standard practice to transform it into a linear form, such as:

$$\underline{y} = Ax + \underline{e} \tag{1}$$

where $\underline{y}$ is a vector of m observables, x is the state vector (n components) of the parameters on which the observables depend, among which are the parameters of interest, the $m \times n$ matrix A is the design matrix and $\underline{e}$ is a vector of measurement errors. $\underline{y}$ and $\underline{e}$ are random variables (indicated by an underscore).

In the GNSS case, the observable $\underline{y}$ is constituted by the range measurements (code and carrier phase) from each visible satellite, and in some cases by the Doppler observations, to determine

---

[1] the term observable is used to refer to the random variable, while the term observation refers to its realization.

the velocity of the user. Such observable can be further augmented with external measurements/ information, such as estimates of the ionosphere, troposphere, corrections for biases, or by other navigation systems, such as INS. The design matrix A in (1) is determined by the geometrical configuration of the satellites in view, which links the range measurements to the unknowns, and by all the linear relations that link the eventually available additional information (e.g. INS or external corrections) to the unknowns.

In ITS, integrity is listed among the navigation key performance parameters (KPP), which have been identified [1, 2] as: nolistsep [noitemsep].

- *Accuracy.* Accuracy defines how well the estimated or measured position agrees with the true position. It is usually measured by the 95% confidence level for the position error, or by the Root Mean Square Error. Accuracy is computed assuming that the system is working in *fault-free* conditions, with standard performance.

- *Integrity.* Integrity defines the level of trust that can be given to the system. It is the ability of the positioning system to identify when a pre-defined Alert Limit (a bound to the position error) has been exceeded and to then provide timely warnings to drivers. Integrity is measured by either: a) the Probability of Hazardous Misleading Information, P_{HMI}, which is the probability that a position error larger than an Alert Limit (AL) occurs without a warning being timely raised, or b) the Protection Levels, which are the largest position error that may occur without any warning being timely raised, with probability smaller than the maximum allowed P_{HMI}.

- *Continuity.* Continuity is the capability of the navigation system to provide a navigation output with the specified level of accuracy and integrity throughout the intended period of operation (POP). Continuity is expressed as the probability that during the POP the system is providing trustworthy navigation information, without any disruption or Alert being raised.

- *Availability.* Availability is the fraction of time the navigation function is usable, as determined by its compliance with accuracy, integrity and continuity requirements. At any epoch of time, the navigation system is deemed either available or unavailable, depending on whether the availability, integrity and continuity requirements are satisfied.

The KPPs are inter-related. In particular, integrity is tightly connected with continuity, since raising an Alert constitutes a disruption to the continuity of the operations.

2.2. RAIM problem definition

Assume a single epoch scenario in which a user at an unknown position receives signals from the GNSS satellites, and eventual positioning information from other augmentation systems/ external linkage. In this scenario, the RAIM problem is defined as: for any satellite geometry, to which corresponds a certain statistical distribution of the observable $\underline{y}$, find an 'acceptance' region $\Omega \in R^m$ (sub-domain of R^m) *and* an estimation/detection function $F\left(\underline{y}\right)$ that to the observable $\underline{y} \in \Omega$ assigns a position estimator $\underline{\hat{x}}$:

$$\underline{y} \in \Omega \rightarrow \widehat{\underline{x}} = F\left(\underline{y}\right) \tag{2}$$

such that:

$$P\left(\widehat{\underline{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega\right) = P_{\text{HMI}} \leq \overline{P}_{\text{HMI}} \quad \forall x \tag{3}$$

and

$$P\left(\underline{y} \notin \Omega\right) = P_{\text{FA}} \leq \overline{P}_{\text{FA}} \quad \forall x \tag{4}$$

where:

- $\overline{P}_{\text{FA}}$ is the requirement of False Alert probability, the maximum allowable probability that an Alert is raised by the algorithm and the continuity of the operation is interrupted, without any actual reason. $\overline{P}_{\text{FA}}$ is a sub-allocation of the full continuity requirement $\overline{c}$, which has to account also for justified Alert (e.g. in the occurrence of an actual hazardous anomaly).

- Ω_{AL} is the 'integrity region' around the true position which boundaries are the Alert Limits (AL). Fundamentally the position error is required to lie within the boundaries defined by the ALs (therefore inside Ω_{AL}) with an extremely high probability, $1 - \overline{P}_{\text{HMI}}$. While in aviation this region is cylindrical, with the radius of the cylinder defined by the Horizontal Alert Limit (HAL) and height defined by the Vertical Alert Limit (VAL), in ITS the shape of this region has not been defined yet, and possibly will be dependent on the specific application. It is expected that in most land applications the vertical error will not need to be monitored, and only limits in the horizontal plane will be considered. On the horizontal plane, distinction shall be made between along-track (AT) and cross-track (CT) directions of motion. A rectangular integrity region could be used, defined by the ALs in the two directions, AL_{AT} and AL_{CT} respectively. Alternatively, an ellipsoidal region could be adopted with semi-axes AL_{AT} and AL_{CT}. **Figure 1** shows the different types of integrity regions.

- $\overline{P}_{\text{HMI}}$, the (maximum allowed) Probability of Hazardous Misleading Information P_{HMI}, is the integrity requirement per epoch. This is the probability that the information on the vehicle position is wrong by an amount larger than the ALs, without any alert or warning on possibly present anomaly being provided along. In aviation $\overline{P}_{\text{HMI}}$ values range from 10^{-7} to 10^{-9} per operation (e.g. approach), whereas for ITS there are yet no candidate values apart from those for aviation.
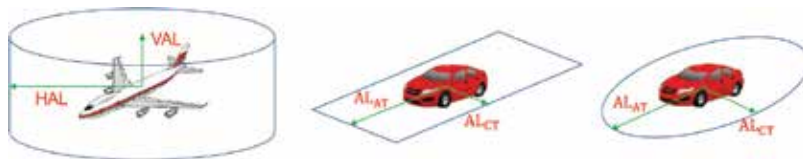

Figure 1. Integrity regions in aviation and in ITS.

The acceptance region Ω fundamentally defines the set of all the measurements y from which it is possible to determine a safe position estimate $\widehat{x}$, i.e., for which the requirement on the P_{HMI} is satisfied.

In any geometry, the rule can be optimized in different ways. The two extreme approaches would be: 1) minimizing the P_{HMI} given the requirement on the continuity is satisfied, or viceversa 2) minimizing the P_{FA} (maximizing the continuity c) given the requirement on the P_{HMI} is satisfied. The first is usually the preferred approach.

2.3. Protection levels (PL)

To define the PLs the total requirement on the P_{HMI}, the $\overline{P}_{\mathrm{HMI}}$, must be split into the different position components. In aviation, it is to be split into horizontal and vertical allocations, $\overline{P}_{\mathrm{HMI}}^{\mathrm{hor}}$ and $\overline{P}_{\mathrm{HMI}}^{\mathrm{ver}}$. In ITS instead, it is to be split between the horizontal along-track (AT) and cross-track (CT) components, $\overline{P}_{\mathrm{HMI}}^{\mathrm{AT}}$ and $\overline{P}_{\mathrm{HMI}}^{\mathrm{CT}}$, whereas the vertical component is (generally) not of concern. $\mathrm{PL}_{\mathrm{AT}}$ and $\mathrm{PL}_{\mathrm{CT}}$ are defined as the maximum position error size (in the AT direction and in the CT direction) that can pass undetected with a probability smaller or equal to the probability requirements, $\overline{P}_{\mathrm{HMI}}^{\mathrm{AT}}$ and $\overline{P}_{\mathrm{HMI}}^{\mathrm{CT}}$, i.e.,

$$\mathrm{PL}_{\mathrm{AT}} = \arg\min_{\delta} P\left(\left|\widehat{\underline{x}}_{\mathrm{AT}} - x_{\mathrm{AT}}\right| > \delta | \mathrm{No\ Alert}\right) \leq \overline{P}_{\mathrm{HMI}}^{\mathrm{AT}}$$

$$\mathrm{PL}_{\mathrm{CT}} = \arg\min_{\delta} P\left(\left|\widehat{\underline{x}}_{\mathrm{CT}} - x_{\mathrm{CT}}\right| > \delta | \mathrm{No\ Alert}\right) \leq \overline{P}_{\mathrm{HMI}}^{\mathrm{CT}}$$

(5)

with $\overline{P}_{\mathrm{HMI}}^{\mathrm{AT}} + \overline{P}_{\mathrm{HMI}}^{\mathrm{CT}} = \overline{P}_{\mathrm{HMI}}$. To satisfy the navigation availability requirement it has to be:

$$\mathrm{PL}_{\mathrm{AT}} \leq \mathrm{AL}_{\mathrm{AT}} \quad and \quad \mathrm{PL}_{\mathrm{CT}} \leq \mathrm{AL}_{\mathrm{CT}}$$

(6)

If those equations are satisfied integrity is maintained for the epoch under consideration. Instead of computing the PLs, the integrity monitoring system can simply compute the actual P_{HMI} or an upperbound for it, and then compare it to the requirement $\overline{P}_{\mathrm{HMI}}$. If $P_{\mathrm{HMI}} \leq \overline{P}_{\mathrm{HMI}}$, integrity is maintained.

2.4. RAIM input, output and performance parameters

In this Section the input and output parameters of a RAIM algorithm are summarized. **Figure 2** shows a schematic representation of a RAIM algorithm. A RAIM algorithm is constituted of two blocks: the first one assesses the geometry or model strength the second one processes the real time observations and assesses their coherency.

The model strength assessment takes as input the design matrix A and the distribution function of the observable $f_{\underline{y}}$, i.e., the observation model, at each epoch. Output of this first assessment are the PLs and/or the P_{HMI}, and consequently the availability prediction for that epoch: if any PL > AL, or equivalently $P_{\mathrm{HMI}} > \overline{P}_{\mathrm{HMI}}$, the navigation service is declared
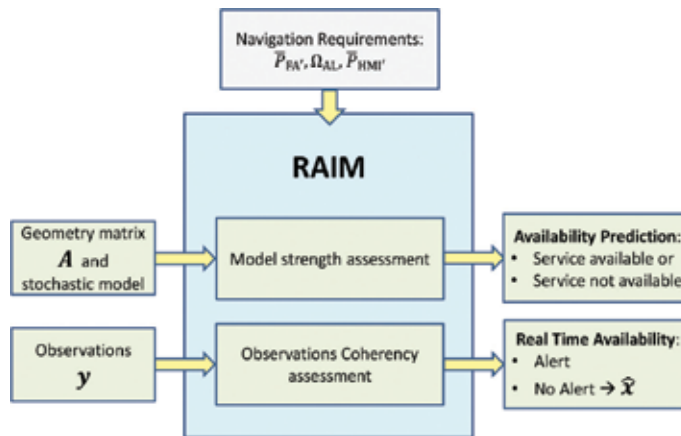
Figure 2. RAIM scheme. Integrity can be assessed first on the basis of the model strength only, and next in real time after an observation is taken.

unavailable. Such availability assessment can be made at each epoch on the basis of the model strength, before the actual measurements are taken.

The observation coherency assessment takes as input the observations y at each epoch. The output of the observations processing is the issue of a state of either Alert or No Alert for that epoch; in case of No Alert, a position solution is provided to the user. In this step, a real time check of the observations is performed. Alert is declared in case the sample measurement taken is too inconsistent: the case $y \notin \Omega$ introduced in the definition of RAIM problem.

Both blocks of the RAIM structure require as input the navigation requirements on integrity and continuity, i.e., Ω_{AL}, the integrity region, $\overline{P}_{HMI}$, the maximum allowed P_{HMI}, and $\overline{P}_{FA}$, the False Alert (or continuity) requirement. The performance of a RAIM algorithm can be measured over time by computing (estimating) the actual P_{FA}, P_{HMI} and PLs.

3. RAIM approaches

Since RAIM is linked with the estimation method, two approaches to RAIM can be distinguished:

- *Fault Detection and Exclusion (FDE) procedure*: one adopts a standard estimation rule, for instance the Best Linear Unbiased Estimation (BLUE [8], characterized by highest accuracy in fault-free conditions); in case the BLUE is not satisfying the integrity requirements (i.e., too large P_{HMI}, because a fault is suspected), one can switch to a different estimator, e.g., a BLUE applied on a subset of the original measurements set. In this way the suspected fault is excluded, and the associated bias in the estimation removed.

- *Robust estimation*: one adopts an estimation rule tailored to integrity. Instead of employing the BLUE, one can sacrifice on some accuracy in fault-free conditions to gain in integrity.

  A combination of both methods listed above is also possible. Here only FDE procedures are analyzed.

3.1. FDE procedure

In an FDE procedure one assumes the possible occurrence of different hypotheses, the fault-free case (null hypothesis $\mathcal{H}_0$) and the occurrence of fault/anomalies (alternative hypotheses $\mathcal{H}_i$). An FDE procedure is applied to detect whether an anomaly is affecting the system, and, in case of detection, exclude the anomalous observations. In a common FDE procedure, typically the BLUE is applied to the model corresponding to the hypothesis $\mathcal{H}_i$ that is more likely (or safer to use). Once it has been decided which hypothesis is most likely to hold true (this decision is made through a statistical testing procedure), the estimator to be used is the BLUE for the model corresponding to that hypothesis. The BLUE for the unknown x in the linear model (1), assuming known dispersion of $\underline{e}$, i.e., $D(\underline{e}) = Q_y$, reads:

$$\widehat{\underline{x}} = S\underline{y} \tag{7}$$

with $S = \left(A^T Q_y^{-1} A\right)^{-1} A^T Q_y^{-1}$ the pseudo-inverse of matrix A in the metric defined by Q_y.

Fundamentally $\widehat{\underline{x}}\left(\underline{y}\right)$ in this approach will be constituted by different linear functions of the observable: it will be in the form of (7) when the null hypothesis $\mathcal{H}_0$ is considered most likely, or conversely different forms $\widehat{\underline{x}}_i$ when they one of the alternative hypothesis is designated to be most likely.

3.2. Statistical hypothesis testing

FDE procedures are based on statistical hypothesis testing [9]. In an FDE procedure statistical tests are performed to determine which hypothesis (fault-free/faulty) on the system state is most likely to hold, and determine the observable domain subdivision discussed in Section 3.1. In this chapter only linear models are analyzed, therefore a special attention shall be given to statistical hypothesis testing in linear models. The aim is to decide between competing linear models that could describe the observed phenomenon or process, once an observation has been made. Furthermore the observables are assumed to have *normal* distributions, and different hypotheses differ only in the specification of the functional model. The models considered are thus Gauss-Markov models [10].

Given the linear model of Eq. (1), we assume the random noise distribution to be known, Gaussian and zero mean:

$$\underline{e} \sim N\left(0, Q_y\right) \tag{8}$$

The linear system in (1) represents the state of standard or nominal operations, that is the case in which the system is working properly without any fault. This state is considered as the null hypothesis $\mathcal{H}_0$. The case of a fault affecting the system constitutes instead a different state, described by an alternative hypothesis $\mathcal{H}_a$, under which the linear model assumes a different form. Therefore:

$$\begin{aligned}\mathcal{H}_0 : \quad &\underline{y} = Ax + \underline{e} \\ \mathcal{H}_a : \quad &\underline{y} = Ax + C_y\nabla + \underline{e}\end{aligned} \tag{9}$$

where C_y is a $m \times q$ matrix which represents the 'signature' of the errors in the measurements and ∇ is a q-sized vector that contains the sizes of the biases in each degree of freedom (q) of C_y.

To test $\mathcal{H}_a$ against $\mathcal{H}_0$, the Uniformly Most Powerful Invariant (UMPI) test statistic (through application of the Generalized Likelihood Ratio (GLR) criterion) reads:

$$\underline{T}_q = \widehat{\underline{e}}_0^T Q_y^{-1} C_y \left(C_y^T Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \right)^{-1} C_y^T Q_y^{-1} \widehat{\underline{e}}_0 \tag{10}$$

where $\widehat{\underline{e}}_0 = \underline{y} - A\widehat{\underline{x}}_0$ is the vector of residuals computed considering the null hypothesis holding true ($\widehat{\underline{x}}_0$ being the position estimator under the null hypothesis, obtained by Eq. (7)).

The test statistic $\underline{T}_q$ is χ^2 distributed:

$$\mathcal{H}_0 : \underline{T}_q \sim \chi^2(q, 0) \quad \text{and} \quad \mathcal{H}_a : \underline{T}_q \sim \chi^2(q, \lambda) \tag{11}$$

with non-centrality parameter:

$$\lambda = \nabla^T Q_{\hat{\nabla}}^{-1} \nabla \tag{12}$$

where $Q_{\hat{\nabla}}^{-1} = C_y^T Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y$, $Q_{\hat{e}_0} = P_A^\perp Q_y P_A^{\perp T}$ and $P_A^\perp = I - A\left(A^T Q_y^{-1} A\right)^{-1} A^T Q_y^{-1}$.

Knowing (though only partially in case of $\mathcal{H}_a$) the distributions of the test statistic under the different hypotheses, one can define a critical region K (to reject the null hypothesis) on the basis of type I and type II error probabilities. The critical region is one sided, of the type:

$$K : T_q > k \tag{13}$$

with k the test threshold (or critical value).

The theory above constitutes the basis of statistical hypothesis testing in linear models, that allows to build the specific test in the simple binary case of null versus one alternative hypothesis. In case one has to choose among multiple alternative hypotheses, one option is to employ a set of binary tests. However, a number of different methods exist in statistics, aiming to answer this more complex problem. Hypothesis testing based methods are known as Multiple Comparisons methods [11], while other methods that do not recur to hypothesis testing are known as Subset Selection methods [12].

4. The aviation legacy

In this section, first the observation models and the typical assumptions adopted in civil aviation applications are described, and next the two most popular RAIM algorithms developed for such

applications, i.e., the Weighted RAIM [13] and the Advanced RAIM (ARAIM) [14] algorithms, are introduced.

4.1. GNSS anomalies and their models

The main categories of High Dynamics Threats (HDTs) to be monitored in aviation applications, which rely on code-based SPP, are here listed. The HDTs are threats that cannot be monitored by the GNSS ground control system, as opposed to the Low Dynamics Threats (LDTs) [14]. They are categorized into: [noitemsep].

- Clock and ephemeris estimation errors, see [15];

- Signal deformations, see [16];

- Code-carrier incoherency, see [17];

From the snap-shot perspective (considering a single epoch of time), and working with carrier-phase smoothed code measurements, an outlier in a single satellite is believed to be the main threat (in terms of probability of occurrence). Simultaneous outliers on multiple satellites (wide failure errors) can occur, but with a much lower likelihood [14]. Among these are the constellation faults (e.g. upload of incorrect navigation messages that may impact a full constellation).

Errors/anomalies in signal propagation, as ionosphere, troposphere and multipath, shall not be considered hazardous for aviation when the new civil frequency L5/E5, and new certified receivers, are available: the tropospheric delay has typically a small effect (and one can correct sufficiently well for this error source), ionosphere gradients/fronts effects are supposed to cancel out with the use of ionosphere-free combination, and multipath depends on the local satellite-receiver geometry and can be considered on a per satellite basis (typically outlier-like).

4.2. General distribution of the observable

In the previous section the main threats possibly affecting the positioning system in aviation applications have been described: on this basis a model to describe the distribution of the observable, able to take into account the possible occurrence of anomalies, shall be formulated. The pdf of $\underline{y}$ is generally supposed to be known in standard fault free conditions, but it cannot be fully defined in the presence of anomalies. However, it is assumed that anomalies in the systems will occur with a low failure rate.

Different hypotheses can be defined to represent the state of the system: a fault free (null) hypothesis $\mathcal{H}_0$ and N_a alternative hypothesis $\mathcal{H}_i$, representing the different possible types of anomalies affecting the system, with $i = 1, ..., N_a$. Here only linear models are considered, and hypotheses of the type of Eq. (9), i.e., $\mathcal{H}_i : \underline{y} = Ax + \nabla y_i + \underline{e}$.

Single satellite faults and constellation faults can be modeled by different C_i matrices: in case of single satellite faults, or combinations of independent single satellite faults, the main C_i's to consider shall be the canonical unit vectors of R^m or $m \times q$ matrices made up of different canonical unit vectors of R^m, respectively; in case of constellation faults, a matrix C_i of $m - n$ columns, fully complementing A in the vector space R^m, shall be used.

The distribution of the observable $\underline{y}$ depends on the state of the system. Under each hypothesis, $\underline{y}$ is assumed to be distributed as a multivariate normal distribution (Eqs. (1), (8) and (9)). It is possible to associate prior probabilities to the occurrence of the different hypotheses, in such a way that the variable $\underline{\mathcal{H}}$, representing the state of the system, has a prior Probability Mass Function (PMF), with discrete values p_i for each realization. Thus $\underline{\mathcal{H}}$ and $\underline{y}$ marginal distributions are:

$$\underline{\mathcal{H}} \sim \begin{cases} P(\underline{\mathcal{H}} = \mathcal{H}_0) = p_0 \\ P(\underline{\mathcal{H}} = \mathcal{H}_1) = p_1 \\ \vdots \\ P(\underline{\mathcal{H}} = \mathcal{H}_{N_a}) = p_{N_a} \end{cases} \quad \Rightarrow \quad \underline{y} \sim p_0 \cdot f_{\underline{y}|\mathcal{H}_0} + \sum_{i=1}^{N_a} p_i \cdot f_{\underline{y}|\mathcal{H}_i} \tag{14}$$

At this point the uncertainty about the $\underline{y}$ distribution is expressed by its dependence on the unknown variable ∇_i beside x. To tackle this uncertainty, most RAIM algorithms assume worst-case bias scenarios or compute bounds for the worst-case risk that could result, see for instance [18].

4.3. Weighted RAIM

In [13] a Weighted RAIM implementation is described. This constitutes one of the first relevant RAIM algorithms conceived and is still in use today, typically implemented in aviation grade GPS receivers, to provide low-precision lateral integrity only. The method consists of the two steps defined in Section 2.4, the model strength assessment and the real time observation coherency assessment. Even though not theorized in the original paper, the method is based on the assumption of the observable distribution described in previous section, with the constraint that only single satellite faults are possibly occurring.

A single test, the OMT, is used to judge the quality of the observations at each epoch [13]. The OMT, also known as χ^2 test, is a UMPI test that employs a test statistic of the form of Eq. (10), and addressing a most generic anomaly, i.e., with $q = m - n$. Such test statistic coincides with the Weighted Sum of Squared Errors (WSSE), defined as:

$$\underline{\text{WSSE}} = \widehat{\underline{e}}^T Q_y^{-1} \widehat{\underline{e}} \tag{15}$$

If this statistic exceeds a certain threshold k, the estimated position is assumed significantly biased; otherwise, it is assumed acceptable. This threshold is chosen to meet the probability of False Alert requirement, $\overline{P}_{\text{FA}}$, knowing that in the fault-free hypothesis, the $\underline{\text{WSSE}}$ is distributed as a central χ^2 with $m - 4$ degrees of freedom (using GPS only).

4.4. Model strength assessment

If a range error from one measurement occurs, the expected value of the test statistic grows, along with, proportionally, the expected position error. The satellite geometry determines how

the error in the range domain propagates into the position domain. The original Weighted RAIM algorithm focuses on monitoring only the vertical component of the position solution, but the same reasoning can be made for the other components. In a simple two-dimensional graph, plotting $\sqrt{\text{WSSE}}$ on the horizontal axis and the vertical position error on the vertical axis, their relation can be represented by a straight line (see **Figure 3**), with a steepness (slope), for satellite i given by:

$$V_{\text{slope}_i} = \frac{|S_{[3,i]}|\sigma_i}{\sqrt{1 - P_{A_{[i,i]}}}} \tag{16}$$

with $S = \left(A^T Q_y^{-1} A\right)^{-1} A^T Q_y^{-1}$, $\sigma_i = \sqrt{Q_{y_i}} = \sigma_{y_i}$ and where the subscripts in square brackets indicate the indexes of the matrix elements'. The Vertical Protection Level (VPL) is computed as:

$$VPL \equiv \max_i \left(V_{\text{slope}_i}\right) k + k_{\text{MD}} \sigma_{\hat{x}_3} \quad i = 1, 2, \ldots, m \tag{17}$$

where k and k_{MD} are obtained as:

$$k = \sqrt{\text{inv-}\chi^2_{\text{CDF}}\left(\overline{P}_{\text{FA}}, m - n\right)}; \qquad k_{\text{MD}} = \Psi^{-1}\left(\frac{\overline{P}_{\text{HMI}}}{mp}\right) \tag{18}$$

with inv-$\chi^2_{\text{CDF}}(\cdot, m - n)$ the inverse of a central χ^2 CDF function with $m - n$ degrees of freedom, $\Psi(\cdot)$ the tail probability of the cumulative distribution function of a zero mean unit Gaussian distribution, and p the a-priori probability of hazardous fault in one satellite. The above formulas for the VPL are based on the following expression of the integrity risk under an alternative hypothesis:

$$P_{\text{HMI}|\mathcal{H}_i} = P_{\text{MD}_i} \cdot P\left(|\hat{x}_3 - x_3| > \text{VAL}|\mathcal{H}_i\right) \tag{19}$$

which assumes that an integrity event corresponds to the simultaneous occurrence of an MD and a positioning error larger than the Vertical AL (VAL), and is justified by the fact that test
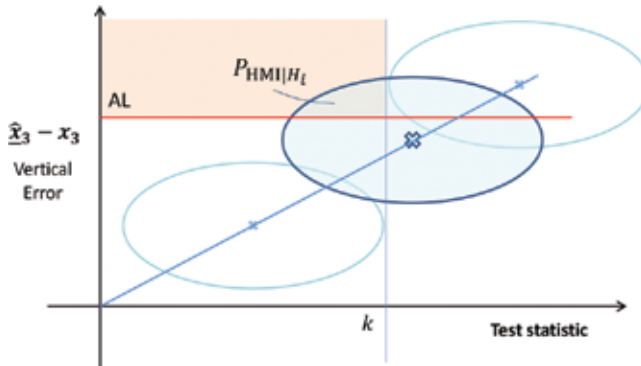

Figure 3. Representation of the weighted RAIM's V_{slope} concept.

statistic and positioning error are uncorrelated. The VPL is the measure of the observation model strength: if VPL > VAL, integrity is not available for the geometry considered.

4.5. Real time availability

The real time availability assessment is performed if the model strength assessment was passed successfully (VPL < VAL). At each epoch, once the observations are collected, the WSSE is computed and compared with the threshold. As in standard hypothesis testing, we have the following decision rule:

$$\text{If WSSE} > k, \text{ reject the fault-free hypothesis and declare Alert} \tag{20}$$

else standard operations continue.

4.6. ARAIM

The Weighted RAIM presented in the previous section was developed for the single GPS constellation and has been found generally suboptimal, even though presenting a very practical and efficient approach. An enhanced approached, known as ARAIM, provides the following improvements [14]:

- in addition to single satellite faults, multi-dimensional faults (affecting multiple satellites at a time) are accounted for [14, 18];

- the potential of the multi-constellation GNSS is fully exploited, instead of GPS only [14, 18];

- rather than using only single-frequency observations, use of dual-frequency observations, to remove the first order ionospheric delay, is foreseen, [14, 18];

- a proof of safety is given [14, 18]. Weighted RAIM is not proven to be always conservative;

- different statistical tests, more tailored to detecting faults that have sensible impact on the position estimate [19], are employed.

The basic concepts of ARAIM are here outlined. For more details, see [14, 18, 19]. **Figure 4** shows a block diagram representation of the ARAIM algorithm. From a statistical point of view, ARAIM is based on the following concepts:

- Multiple Hypothesis approach with a-priori probabilities: the system is supposed to be in one out of a set of different possible states described by multiple hypotheses, to each of which is assigned an a-priori probability of occurrence (Section 4.2). The P_{HMI} is computed by the sum of the P_{HMI} under the different hypotheses, weighted on the base of their prior probabilities.

- Solution Separation (SS) as test statistics: to discriminate between hypotheses, to eventually exclude faulty measurements, the difference between the position solutions under the different alternative hypotheses and the null hypothesis is computed and used as a test statistic. For each alternative hypothesis considered a difference vector (SS) is computed and a test is run for each of the position components of the vector.
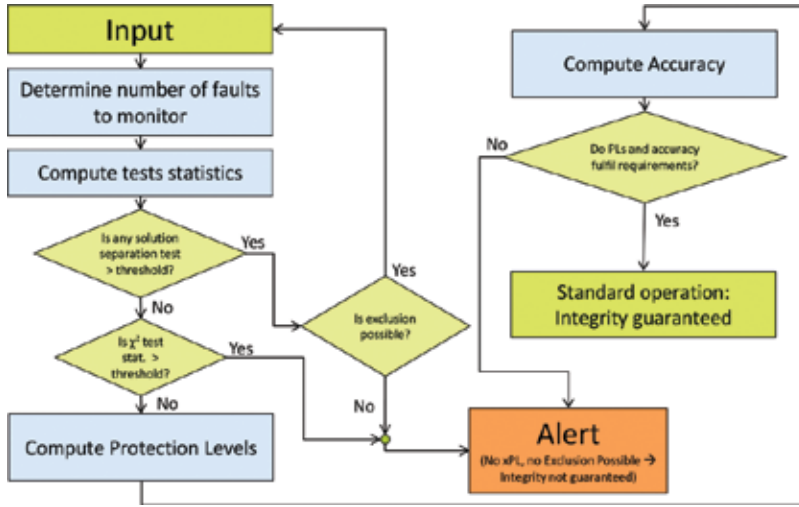
Figure 4. ARAIM baseline architecture. The algorithm checks the coherency of the observations by means of the solution separation tests, evaluates the possibility of excluding corrupted observations with exclusion specific tests, and computes the PLs. Integrity is guaranteed if PLs < ALs.

If one characterizes each alternative hypothesis by a different subscript i, the i th Solution Separation vector can be written as:

$$\underline{T}_{SS_i} = \widehat{\underline{\nabla}}\widehat{\underline{x}}_i = \widehat{\underline{x}}_0 - \widehat{\underline{x}}_i \tag{21}$$

where $\widehat{\underline{x}}_0$ and $\widehat{\underline{x}}_i$ are the position solutions obtained employing the null and the alternative model respectively, i e.:

$$
\begin{aligned}
\widehat{\underline{x}}_0 &= \left(A^T Q_y^{-1} A\right)^{-1} A^T Q_y^{-1} \underline{y} = S\underline{y} \\
\widehat{\underline{x}}_i &= \left(A^T Q_{y_i}^{-1} A\right)^{-1} A^T Q_{y_i}^{-1} \underline{y} = S_i \underline{y}
\end{aligned}
\tag{22}
$$

where $Q_{y_i}^{-1}$ is obtained from Q_y^{-1} replacing the diagonal elements corresponding to the faulty satellites in hypothesis $\mathcal{H}_i$ with 0 (this means giving zero weight to such observations). In practice, these tests have similar performance to the UMPI tests (see [20, 21]).

4.7. Model strength assessment

The PLs are computed on the basis of the model strength (satellite geometry and stochastic model), and compared to the AL to determine the integrity availability. The computation of the PLs is based on an iterative procedure: the PLs are determined in such a way that the sum of the P_{HMI_i} under each alternative hypothesis is equal to the full $\overline{P}_{HMI}$ requirement:

$$\overline{P}_{HMI} = \sum_i^{N_a} P_{HMI_i} = \sum_i^{N_a} \left[P_{MD_i} \cdot P(\widehat{\underline{x}} - x \notin \Omega_{AL} | \mathcal{H}_i) \right] \tag{23}$$

where in the last equality the relation (19) is applied. As a result, the VPL must satisfy the following equation [18]:

$$2\Psi\left(\frac{\text{VPL}}{\sigma_{\widehat{x}_{0,3}}}\right) + \sum_{i=1}^{N_a} p_i \Psi\left(\frac{\text{VPL} - k_{i,3}}{\sigma_{\widehat{x}_i}, 1}\right) = \zeta \overline{P}_{\text{HMI}} \tag{24}$$

where $k_{i,3}$ is the test threshold for the i th SS test, 3 rd component (vertical), and ζ is the fraction of the full $\overline{P}_{\text{HMI}}$ allocated to the vertical direction ($0 < \zeta < 1$). The HPL instead is computed with $\text{HPL} = \sqrt{\text{PL}_{x_1}^2 + \text{PL}_{x_2}^2}$, where $\text{PL}_{x_1}^2$ and $\text{PL}_{x_2}^2$, the PLs for the two horizontal components, are computed with formulas equivalent to (24). The thresholds $k_{i,j}$ are computed with

$$k_{i,j} = -\Psi^{-1}\left(\frac{P_{\text{FA}}}{4N_a}\right)\sigma_{\widehat{\nabla}\widehat{x}_{i,j}} \tag{25}$$

where $\sigma_{\widehat{\nabla}\widehat{x}_{i,j}}$ is the standard deviation of the corresponding SS test statistic (see [18]). If any PL > AL, integrity is not available for the geometry considered.

4.8. Real time availability

In ARAIM the testing is subdivided in two steps:

1. Detection tests: the SS tests are computed and compared to their thresholds; if none of the tests exceeds the threshold, the fault-free hypothesis is confirmed and standard operations continue, otherwise the algorithm proceeds to the next step.

2. Exclusion confirmation tests: extra tests are run to determine if it is safe to exclude some observations and continue to provide navigation service. These tests are meant to minimize the risk of wrong identification. More details are given in [18].

After detection and eventual exclusion of observations, the PLs are re-computed (as post-observations PLs) and compared with the thresholds. If any $\text{PL}_{\text{post}} > \text{AL}$, an Alert is raised.

5. The ITS challenge

As mentioned in Section 1.2, when moving from aviation to land applications, a number of issues have to be taken into account in the context of integrity monitoring. The main two issues are:

- Positioning has to be performed in urban environment: additional vulnerabilities are to be taken into account, i.e., multipath, NLOS, interference and spoofing.

- Higher precision/smaller PLs are required: this may lead to the use of precise positioning techniques (PPP, RTK) with their additional vulnerabilities, as well as additional navigation sensors/technologies (INS,V2I and V2I communication, camera, etc.).

The main assumptions on which the FDE procedures and RAIM algorithms described so far rely on are (Section 4.2): 1) linear estimation problem, 2) Gaussianity of the observables and 3)

mean-shift model for the anomalies. In land applications, these assumptions are likely to hold, though multipath and NLOS may challenge the second one, while the large number of observations available and vulnerabilities increases the computational complexity of FDE procedures. These aspects are addressed in more detail in the following.

5.1. Urban environment: multipath, NLOS and interference

Multipath is the most significant source of measurement errors in ITS applications, as it is dependent on the environment surrounding the antenna and is especially intense in dense urban areas. Buildings and other obstacles degrade the signal reception in three ways: 1) signals are completely blocked and unavailable for positioning, 2) signals are blocked in their direct path, but are still received via a reflected path, with the NLOS reception, 3) both direct Line-Of-Sight (LOS) and reflected signals are received, i.e., the case of multipath. NLOS code signals can exhibit positive ranging errors of tens of meters magnitude in dense urban areas.

Numerous innovative techniques have been developed in the recent years to address the multipath and NLOS threats in urban environment. Interest was raised by 3D-map-aided (3DMA) GNSS, a range of different techniques that use 3D mapping data to improve GNSS positioning accuracy in dense urban areas. 3D models of the buildings can be used to predict which signals are blocked and which are directly visible at any location [22, 23]. A technique that determines position by comparing the measured signal availability and strength with predictions made using a 3D city model over a range of candidate positions is known as the shadow matching technique [24]. Such techniques may possibly be integrated with RAIM algorithms for ITS in the near future.

5.2. Precise positioning techniques and multi-sensor integration

The use of precise positioning techniques rather than SPP and the need of integration with other sensors bring a number of complications to the IM for land applications. Some of the main challenges are summarized in the following.

5.2.1. PPP and RTK: Carrier phase observations vulnerabilities

Precise positioning techniques employ carrier phase observations next to code observations. Even though the estimation problem is characterized by a much larger number of observations and unknowns to solve for, it is still a linear estimation problem. The same hypothesis testing theory applies, and therefore, the same RAIM concepts developed for aviation can be implemented, with appropriate adjustments. However, one drawback of the ARAIM is the associated heavy computational burden, due to the need of running a test for each possible combination of simultaneously biased observations. When multi-systems, multi-frequency and carrier phase based positioning is in use, the total number of combinations of possibly biased observations increases dramatically — so does the computational load for the algorithm. It is thus possible that the current ARAIM approach will not be optimal.

Another issue is constituted by the additional vulnerabilities that affect the precise positioning techniques, mainly carrier phase multipath and cycle-slips. Multipath affects carrier-phase observations with the same mechanism as code observations [25]. Carrier-phase multipath is

one of the critical elements in determining the Time to Ambiguity Resolution (TAR), which can become of concern in regards of timeliness requirements. Furthermore, cycle-slips, which constitute the main RTK-specific threat, as they can cause wrong ambiguity fixing and result in large errors in the positioning, require specific additional monitoring. There is a vast literature on cycle-slip detection, e.g., by [26, 27]. Most cycle-slip detection methods are based on hypothesis testing, but exploit the multi-epoch data processing to increase their detection power.

5.2.2. Multi-sensor integration and recursive data processing

Use of multiple sensors for navigation means that extra observations shall be integrated with the GNSS observations. If the extra observations are linear in the unknown parameters, they can be simply stacked together in the same linear estimation problem. Integration with INS is a complex problem on which a large literature exist [28]. Finally, while the focus of this chapter was only on snapshot RAIM (single epoch), RAIM techniques for multi-epoch recursive data processing are under development [29].

5.2.3. Cooperative integrity monitoring (CIM) concept

Section 2.4 shows that an integrity assessment can be made before the observations are taken, when only satellite geometry and environment are known or partially known. New IM concepts intend to exploit the fact that satellite geometry and satellite visibility can be reasonably predicted at any time and location (for instance with the use of city models), and that the same observability conditions repeat periodically over time. Beside the environment nearby the receiver in its nominal conditions, these new concepts plan to exploit also the potentialities offered by a Vehicular Ad-hoc Network (VANET) infrastructure [30]. The potential availability of multiple observations of GNSS signals, taken by different vehicles participating to a VANET, can be shared and combined in order to implement a collaborative spatial/temporal characterization and prediction of the local degradation of the GNSS signals.

6. An example

In this section, the results of a first attempt to perform IM in urban environment employing the RTK positioning method with a short baseline, and applying a prototype ARAIM algorithm, are shown. Such results are only indicative, since most of the assumptions behind the use of ARAIM in an RTK set-up are yet to be justified.

A kinematic test is conducted for practical demonstration of IM for ITS. A small vehicle is fitted with a Trimble multi-GNSS geodetic receiver and a survey-grade antenna. The test is carried out in a dense urban area in Tokyo, Japan. The RTK system uses GPS, GLONASS and BeiDou dual-frequency observations with a sampling rate of 10 Hz. A prototype RAIM algorithm derived from ARAIM is implemented. Due to the lack of common standards, the PLs are computed in the test using different values of $\overline{P}_{\mathrm{HMI}}$ ranging from 10^{-3} to 10^{-6} in order to track empirically the impact of $\overline{P}_{\mathrm{HMI}}$ on the obtained results. A false alert probability ($\overline{P}_{\mathrm{FA}}$) of 0.01 is applied.

Figure 5 shows the PL for the along-track and cross-track directions (shown as PL_{AT} and PL_{CT}) and the absolute values of the positioning errors along these directions (denoted as err_{AT} and err_{CT}) using an integrity risk of 10^{-4} and 10^{-6} as examples. The figure shows that the RTK with correct ambiguity fixing gives positioning errors within a few centimeters. The average absolute value of the AT and CT positioning errors are 0.058 and 0.054 m, respectively. The FDE method detected 15 code observations with severe irregularities, which are attributed to high multipath in this environment. These observations were excluded from further processing. There were a few cases where the ambiguity fixing seemed to be incorrect by one or two cycles, which were not detected by the FDE procedure. However, the PL adapted to these situations and bounded this error as illustrated in the **Figure 5**. Inspection of the Figure also shows that when using RTK with correct ambiguity fixing, an Alert Limit (AL) can be safely chosen as 1 m. The sub-decimeter positioning accuracy of RTK is bounded by a tight protection level. The positioning errors in the test were always bounded by the PLs, and PLs < ALs for the whole period, with an integrity availability of 100%. The medians of the PL for the AT and CT using different integrity risk ($\overline{P}_{HMI}$) values are given in the **Table 1**. Both table and **Figure 5** show that the PLs increase with the decrease of the allowed integrity risk.
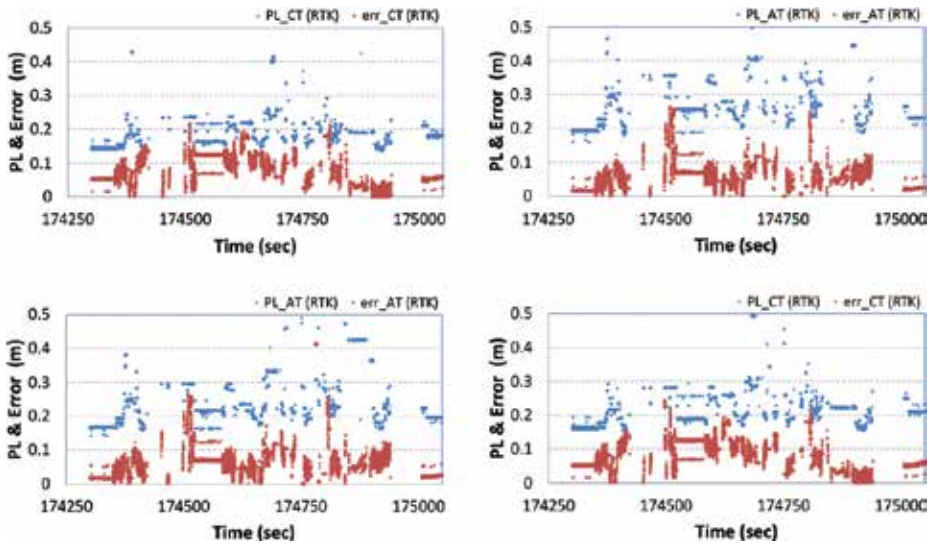
Figure 5. PL_{AT} and PL_{CT} and positioning errors in the AT and CT directions for the integrated positioning systems using $\overline{P}_{HMI} = 10^{-4}$ (top panel) and $\overline{P}_{HMI} = 10^{-6}$ (bottom panel).

Integrity risk ($\overline{P}_{HMI}$)	10^{-3}	10^{-4}	10^{-5}	10^{-6}
PL_{AT}	0.176	0.197	0.215	0.232
PL_{CT}	0.148	0.164	0.177	0.188

Table 1. Median PL_{AT} & PL_{CT} in meters for different values of integrity risk ($\overline{P}_{HMI}$).

7. Concluding remarks

In this chapter the concepts of integrity and IM have been introduced, and the main RAIM methods currently in use or under development have been presented. As these methods were developed in the aviation context, their adoption in land applications has been discussed. The positioning methods used in land applications still satisfy the assumptions made by current RAIM algorithms, though great care shall be taken in addressing the larger number of vulnerabilities affecting the positioning system, in particular multipath and the carrier-phase specific vulnerabilities. Some preliminary but promising results of the application of a RAIM algorithm in urban environment were shown. Further research and practical experiments are necessary to strengthen the confidence in the models.

Acknowledgements

Author details

Davide Imparato[1], Ahmed El-Mowafy[1]* and Chris Rizos[2]

*Address all correspondence to: a.el-mowafy@curtin.edu.au

1 Curtin University, Bentley, WA, Australia

2 UNSW, Kensington, NSW, Australia

References

[1] Research and Innovative Technology Administration. Connected Vehicle Technology: Safety Pilot Driver Acceptance Clinic Overview. RITA. Washington, DC, US: US Department of Transportation; 2011

[2] ETSI. Intelligent Transport Systems (ITS). Vehicular communications; basic set of applications; part 1: Functional requirements. In: European Telecommunications Standards Institute, ETSI TS 102 637–1 V1.1.1. 2010

[3] Enge P, Walter T, Pullen S, Kee C, Chao YC, Tsai YJ. Wide area augmentation of the global positioning system. Proceedings of the IEEE. 1996;**84**(8):1063-1088

[4] Roturier B, Chatre E, Ventura-Traveset J. The SBAS integrity concept standardised by ICAO. Application to EGNOS. Navigation-Paris. 2001;**49**:65-77

[5] Murphy T, Imrich T. Implementation and operational use of ground-based augmentation systems (GBASs) – A component of the future air traffic management system. Proceedings of the IEEE. 2008;**96**(12):1936-1957

[6] DeCleene B. Defining pseudorange integrity-overbounding. In: Proceedings of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2000); 2000. pp. 1916-1924

[7] Crash Avoidance Metrics Partnership. Vehicle safety communications — Applications (VCS-A): Final report. DOT HS 811 492A, National Highway Traffic Safety Administration, Washington DC, US; 2011

[8] Arnold SF. The Theory of Linear Models and Multivariate Analysis. Vol. 2. New York: Wiley; 1981

[9] Neyman J, Pearson ES. On the Problem of the Most Efficient Tests of Statistical Hypotheses. Springer; 1992

[10] Teunissen PJG, Simons DG, Tiberius CCJM. Probability and Observation Theory. Faculty of Aerospace Engineering, Delft University of Technology: Delft University; 2005

[11] Miller RG. Simultaneous Statistical Inference. Second ed. Berlin: Springer-Verlag New York Heidelberg; 1981

[12] Miller A. Subset Selection in Regression. CRC Press; 2002

[13] Walter T, Enge P. Weighted RAIM for precision approach. Proceedings of the ION GPS-95, Palm Springs, September 1995. 1995. p. 1995-2004

[14] Blanch J, Walter T, Enge P, Wallner S, Fernandez FA, Dellago R, et al. A proposal for multi-constellation advanced RAIM for vertical guidance. Proceedings of the 24th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2011). 2001:2665-2680

[15] Heng L, Gao GX, Walter T, Enge P. Proceedings of the international technical meeting of the satellite division of the Institute of Navigation (ION GNSS 2010). In: GPS Signal-In-Space Anomalies in the Last Decade. Portland: OR; 2010

[16] ICAO. GNSS Standards and Recommended Practices (SARPS). Annex 10, aeronautical telecommunications, volume 1 (radio navigation aids), amendment 84, published 20 July 2009, effective 19 November 2009; 2009

[17] Montenbruck O, Hauschild A, Steigenberger P, Langley R. Three's the challenge: A close look at GPS SVN62 triple-frequency signal combinations finds carrier-phase variations on the new L5. GPS World. 2010;**21**(8):8-19

[18] Blanch J, Walter T, Enge P, Lee Y, Pervan B, Rippl M, et al. Baseline advanced RAIM user algorithm and possible improvements. IEEE Transactions on Aerospace and Electronic Systems. 2015;**51**(1):713-732

[19] Brown RG, PW MB. Self-contained GPS integrity integrity checks using maximum solution separation as the test statistic. In: Proceedings of the First Technical Meeting of the Satellite Division of the Institute of Navigation, Colorado Springs. USA: Colorado; 1987. pp. 263-268

[20] Imparato D. Detecting multi-dimensional threats: A comparison of solution separation test and uniformly most powerful invariant test. In: Proceedings of the European Navigation Conference (ENC)-GNSS. Vol. 2014. Instituut voor Navigatie: Nederlands; 2014. pp. 1-13

[21] Imparato D. GNSS Based Receiver Autonomous Integrity Monitoring for Aircraft Navigation. Ph.D. [Thesis]. TU Delft; 2016

[22] Suh Y, Shibasaki R. Evaluation of satellite-based navigation services in complex urban environments using a three-dimensional GIS. IEICE. 2007;**E-90**(B):1816-1825

[23] Wang L, Groves PD, Ziebart MK. Multi-constellation GNSS performance evaluation for urban canyons using large virtual reality City models. Journal of Navigation. 2012;**65**(3): 459-476

[24] Groves PD, Wang L, Adjrad M, Ellul C. GNSS shadow matching: The challenges ahead. In: Proceedings of the ION GNSS+ 2015. Tampa, Florida: The Institute of Navigation; 2015

[25] Rost C, Wanninger L. Carrier phase multipath mitigation based on GNSS signal quality measurements. Journal of Applied Geodesy. 2009;**3**(2):81-87

[26] Kim D, Langley RB. Proceedings of the international symposium on kinematic systems in geodesy, geomatics and navigation. In: Instantaneous Real-Time Cycle-Slip Correction of Dual Frequency GPS Data. Banff Alberta; 2001. pp. 255-264

[27] Lee HK, Wang J, Rizos C. Effective cycle slip detection and identification for high precision GPS/INS integrated systems. The Journal of Navigation. 2003;**56**(3):475-486

[28] Groves PD. Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems. Artech house; 2013

[29] Joerger M, Pervan B. Integrity risk of kalman filter-based RAIM. Proceedings of the 23rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS). 2011:3856-3867

[30] Margaria D, Falletti E. A novel local integrity concept for GNSS receivers in urban vehicular contexts. In: Proceedings of the IEEE/ION PLANS 2014. Monterey, CA: USA, 5–8 May; 2014. pp. 413-425

GPS in Artificial Neural Networks

GPS Modeling of the Ionosphere Using Computer Neural Networks

Daniel Okoh

Additional information is available at the end of the chapter

Abstract

This chapter presents a detailed description to modeling the ionosphere using the method of computer neural networks and data from the GPS (Global Positioning System). The chapter essentially motivates the use of artificial neural networks for ionospheric modeling, and it presents a detailed description of the processes and considerations involved in using artificial neural networks to model the ionosphere. Specific illustration was done using vertical total electron content (VTEC) data from 14 GPS stations in Nigeria that cover the period from years 2011 to 2016, to develop a neural network model of the GPS-TEC over Nigeria in space and in time. Sample simulations from the developed model shows that the model was accurate in predicting the VTEC variation patterns in terms of diurnal, seasonal, annual, longitudinal and latitudinal variations. A comparative analysis between the neural network predictions and those of other models like the IRI-Plas and NeQuick showed that predictions from the neural network model were predominantly more accurate.

Keywords: GPS, GNSS, TEC, neural network, ionosphere, model

1. Introduction

The ionosphere is important to our existence as it affects our radio communication systems, especially our satellite communication systems. Particularly, the ionosphere poses the greatest natural challenge for our global navigation satellite systems (GNSS) when it comes to precise position measurement by ground-based receivers. There are a couple of satellite navigation systems, e.g. the United States' GPS (Global Positioning System), Russia's GLONASS (Global Navigation Satellite System), European Union's GALILEO, China's

BEIDOU (or COMPASS), etc. The GPS is the most common and the most popular of the GNSS systems, and so in this chapter, we will carefully use the two words interchangeably. The evolution of our navigation requirements into satellite-based systems is therefore adding a rapid stair to interest in ionospheric research. The greatest efforts in ionospheric research have been directed towards ionospheric modeling, and related studies that tend to understand how the ionosphere changes in time and space. Several ionospheric models have been developed (e.g. Ref. [1–7]).

To understand exactly how the ionosphere influences our satellite-based navigation systems, it is important to understand how satellite-based navigation systems work. A more detailed introduction to the GNSS is presented by Ref. [8], but the core ideas are briefly and elegantly presented here. A satellite-based navigation system basically consists of some satellites in space. The satellites know their positions in space through the help of ground-based control stations and some internal programming. Through on-board transmitters on the satellites, each satellite continuously transmits radio signals. Each radio signal contains information about the 3-D position in space of the satellite from which it is transmitted, and the time in which the signal is transmitted. GPS receivers on ground (like the ones you and I own in cell phones and other devices) can receive these signals and automatically be able to compute the receivers' 3-D positions.

Exactly how does this happen? How does a GPS receiver know its position by merely receiving position and time stamped radio signals from the satellites? The GPS receivers use in-built computer programs to compute their own positions from the positions of the satellite they receive signals from. The computer programs are based on quite simple geometric calculations. The geometric calculations are based on the premise that if we know the exact 3-D positions of any three objects and the exact range to each of them, then we would be able to determine our own 3-D position. It is emphasized here that the positions of three objects are required because we are interested in 3-D positions. If we are interested in knowing our position in 2-D space, then we will require the positions of only two objects. In the case of the GPS, the interest is to know the 3-D position of the receiver as well as the time the signal is received (this makes 4-D), so we need four objects. A GPS receiver will therefore be able to compute its exact position and time if it receives signals from at least four satellites. From the satellite radio signals they receive, GPS receivers retrieve information on the 3-D positions and times of the satellites as well as the exact range to each of them.

As explained earlier, we know that each of the signals already contain information on the 3-D positions and times of the satellites, but how do the receivers know the ranges to the satellites? GPS receivers estimate ranges to the satellites by using the formula

$$Range = speed\ of\ radio\ signal \times travel\ time \tag{1}$$

The travel time is how long the radio signals have traveled between their transmission and their reception (That is the time difference between when the signals were transmitted from the satellite and when they were received). The signals already contain information on when they were transmitted from the satellite, and the receiver time is one of the four parameters the receiver will compute.

The computation in Eq. (1) is based on the assumption that radio signals (which are electro-magnetic) travel at the constant speed of about c = 2.998×10^8 ms^{-1} in vacuum. And this is where the problem of the ionosphere comes in. The space between the satellites and receivers is not entirely vacuum; there is an intervening region containing ionized matter known as the ionosphere. Because of ionized matter contained in the ionosphere, electromagnetic waves (e.g. the transmitted radio signals) do not travel through the ionosphere at the constant speed of about c = 2.998×10^8 ms^{-1}. The signals are delayed, and this delay is interpreted in Eq. (1) as part of the travel time. This introduces an error into the computed range (the computed range is greater than the actual or true range) which therefore subsequently manifests as an error in the computed receiver position.

An obviously intelligent thing to do is to remove this effect of the delay introduced by the ionosphere, but this is only possible if we know how much the delay is. To make the situation worse, the ionosphere is highly dynamic; it changes appreciably over space and time. We therefore need to know the extent of ionospheric ionization at any given time along the radio route so as to be able to correct for the effect of the ionosphere on the radio signal. This is where ionospheric models are useful. Ionospheric models can be used to now-cast (and even fore-cast) the extent of ionospheric ionization over space and time. And by so doing, ionospheric models are useful and usually applied in GPS error correction for single frequency receivers.

Single frequency receivers are GNSS receivers that can receive radio signals from the satellites in only one frequency. These are the most common types of GNSS receivers we see in everyday usage. They are cost-effective but incapable of estimating the ionospheric delay. On the other hand, there are dual or more frequency receivers which can receive GNSS radio signals at two or more frequencies. In the explanation that follows (on the Data and Methods section), dual or more frequency GNSS receivers are capable of estimating the ionospheric delays, and therefore capable of internally removing the effects of such delays. These types of receivers are mainly used for research and other specialized usages. It is from these types of receivers that data used in this chapter was obtained. There is general intuition that dual-frequency receivers are better than single frequency ones, but in highlighting the tradeoffs between the two, Ref. [9] explained that, asides cost effectiveness of the single frequency receivers, a single frequency receiver may actually outperform the more advanced dual-frequency receiver in terms of accuracy during the first 10 minutes or so, and also in places associated with frequent loss of lock on GNSS signals. Rather than using dual-frequency receivers, some applications therefore prefer using ionospheric models on single frequency receivers to correct for the effects of the ionosphere. The accuracy obtained from this practice however depends on the accuracy of the model used; more accurate models will give more accurate GNSS positions. The development of a regional GPS model of the ionosphere (with improved accuracy) is presented in this chapter. The modeling technique used is the method of computer neural networks.

Computer neural networks (also commonly referred to as neural networks or just NNs for short) have capability for machine learning as well as pattern recognition, and they have been demonstrated to be powerful tools for predictive modeling. NNs operate in a manner

that is similar to the human brain; the networks are composed of simple elements operating in parallel and inspired by the biological nervous system. NNs can learn trends and patterns in particular data they are given and consequently be able to correctly predict unseen and future trends for the data. A neural network can be trained to perform a particular function by adjusting the value of connections (also called weights) between elements [7]. The true power and advantages of neural networks lies in the ability to represent both linear and non-linear relationships directly from the data being modeled. Traditional linear models are simply inadequate when it comes for true modeling data that contains non-linear characteristics [8]. Recent explosion of ionospheric data from the GNSS is spurring interest in using computer neural networks for ionospheric modeling. A number of works have shown that neural networks (NNs) are good candidates for ionospheric modeling [6, 7, 10–13]. In this chapter, neural networks have been used to develop a regional model of the ionosphere over Nigeria. Predictions from the model have also been demonstrated to be more improved in terms of accuracy when compared to predictions from global ionospheric models like the IRI-Plas (International Reference Ionosphere—extended to the Plasmasphere) and the NeQuick.

2. Development of the GPS-based neural network model

2.1. Data and methods

Three main sets of data were used in this chapter, these include: (i) GPS data, (ii) sunspot number (SSN) data, and (iii) disturbance storm time (DST) data. The next section will dwell on GPS data which is of major interest in this chapter.

2.1.1. GPS data

The GPS data used in this chapter were derived from dual-frequency receivers on the NIGNET (Nigerian Permanent GNSS Network, www.nignet.net). A brief description of how ionospheric information is usually obtained from dual-frequency GPS receivers is presented.

How are dual-frequency GPS receivers able to estimate ionospheric delays? The delays introduced on radio signals by the ionosphere are frequency-dependent; the lower frequency signals are more delayed while the higher frequency signals are less delayed. More precisely, the delay (t) is inversely proportional to the radio frequency (f) as shown in Eq. (2a) [14].

$$t = 40.3 \times \frac{TEC}{cf^2} \qquad (2a)$$

c = 2.998×10^8 ms^{-1} is the speed of electromagnetic waves in vacuum, and TEC is the Total Electron Content. TEC is a parameter of the ionosphere that represents the total number of free electrons contained in a 1 m squared column, along the path of the signal through the ionosphere. It is this parameter of the ionosphere that is modeled in this chapter. Eq. (2a) shows that the ionospheric delay is directly proportional to the TEC, therefore the radio signals are more delayed when they travel through a route in the ionosphere with more number of free electrons.

The proportionality expressed in Eq. (2a) forms the underlying principle for deriving iono-spheric information (precisely TEC) using dual-frequency GPS receivers. This is because two radio signals (having frequencies, f_1 and f_2) transmitted at the same time from the same satel-lite will be delayed differently by the ionosphere so they arrive the same receiver at different times. The delays that will be experienced by the two radio signals are, respectively, given by Eq. (2b) and Eq. (2c).

$$t_1 = 40.3 \times \frac{TEC}{cf_1^2} \tag{2b}$$

$$t_2 = 40.3 \times \frac{TEC}{cf_2^2} \tag{2c}$$

Subtracting Eq. (2b) from Eq. (2c), we get the time delay between arrivals of the two signals as in Eq. (3a).

$$\Delta t = t_2 - t_1 = \frac{40.3\,TEC}{c}\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) \tag{3a}$$

Dual-frequency GPS receivers compute the TEC using Eq. (3b) which is obtained by making TEC subject of the formula from Eq. (3a).

$$TEC = \frac{c\,\Delta t}{40.3\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right)} \tag{3b}$$

The TECs computed in this manner using the pseudo-range measurements alone are usually noisy; differential carrier phase measurements are used to obtain precise measures of the relative TECs, and a combination with the pseudo-range measurements provide the abso-lute slant TEC values (STECs) along the receiver-satellite path [15–17]. The computed TECs are referred to as slant, to distinguish them from the unique TEC that will be obtained for a particular location when the satellite is exactly overhead the location (that is, satellite eleva-tion = 90°). This unique TEC is called the vertical TEC (VTEC). VTECs are usually derived from the STECs using Eq. (4).

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

where br and bs are, respectively, the receiver and satellite biases, S(E) is the mapping func-tion defined by Eq. (5) [18].

$$S(E) = \frac{1}{\cos(z)} = \left[1 - \left(\frac{R_E \times \cos(E)}{R_E + h_s}\right)^2\right]^{-\frac{1}{2}} \tag{5}$$

z and E are, respectively, the zenith and elevation angles in degrees; R_E and h_s are, respectively, the mean Earth radius and the ionosphere (effective) height above the Earth surface in km. The value of h_s used for this chapter is 350 km.

GPS Data obtained from the NIGNET are in RINEX (Receiver Independent Exchange) format. The RINEX format is the standard data interchange format for raw satellite navigation system data. RINEX format data obtained from the NIGNET were processed into VTEC data using software developed by Dr. Gopi Seemala (seemala.blogspot.in). The software works basically on the principles highlighted above, and as expressed in Refs. [15, 19].

GPS Data used were from the 14 stations illustrated in **Figure 1** and in **Table 1**. All available data coving the periods from years 2011 to 2016 were used. To obtain instantaneous values of VTEC for a given location, VTEC values from the various satellites that are visible over the location at the time were averaged excluding those from satellites with elevation angles less than 25°. The reason for excluding data associated with low elevation angles is usually to minimize multipath errors. Multipath errors are errors associated with signals that bounce off (or reflected from) nearby buildings, trees, or other structures before they reach the receiver antenna. The problem with these signals is that the resulting range will be greater than the actual straight path range between the satellite and receiver, because the signal first has to bounce off other structures before they reach the receiver antenna. The multipath problem is typical of signals coming from low elevation satellites; the lower the satellite elevation angles (especially satellites close to the horizon), the more likely signals from them are to bouncing
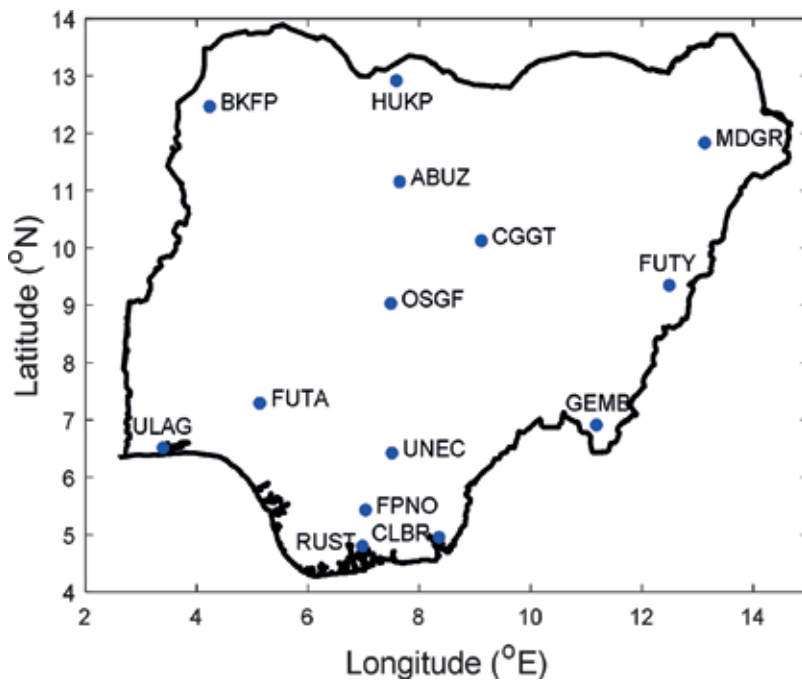

Figure 1. Map of Nigeria showing locations of GPS stations used in this work.

Station Code	City	State	Latitude (deg)	Longitude (deg)	Elevation (m)
ABUZ	Zaria	Kaduna	11.1517	7.6486	706.1
BKFP	Birnin-Kebbi	Kebbi	12.4684	4.2292	251.0
CGGT	Toro	Bauchi	10.1231	9.1181	917.4
CLBR	Calabar	Cross-River	4.9503	8.3514	61.5
FPNO	Owerri	Imo	5.4345	7.0331	92.6
FUTA	Akure	Ondo	7.2986	5.1364	416.0
FUTY	Yola	Adamawa	9.3497	12.4978	248.4
GEMB	Gembu	Taraba	6.9170	11.1839	1796.6
HUKP	Katsina	Katsina	12.9211	7.5909	565.0
MDGR	Maiduguri	Borno	11.8381	13.1309	351.8
OSGF	Abuja	Federal-District	9.0275	7.4861	533.6
RUST	Port-Harcourt	Rivers	4.8017	6.9784	46.6
ULAG	Lagos	Lagos	6.5172	3.3975	45.5
UNEC	Enugu	Enugu	6.4247	7.5047	255.4

Table 1. Description of NIGNET stations used in this work.

off other structures before they reach the receiver antenna. This problem is mostly the reason why research-class GNSS receivers are installed such that their antennas are raised above nearby structures/buildings (or away from the structures/buildings), and the antennas are built in such a shape that the receiving surface faces the sky. In this way, radio signals that are reflected from structures beneath the antenna do not get received by the antenna even when they hit the bottom surface of the antenna. The resulting VTEC data were further averaged in 1-hour intervals to reduce data and to lessen spikes on the data profiles.

2.1.2. Other data

The other two set of data used in this chapter are the DST and SSN data. The DST is a measure of the disturbances in the Earth's magnetic field, it is an index often used to describe the level of geomagnetic activity during storms. On the other hand, the SSN is a count of the number of sunspots present on the surface of the Sun. It is a measure of the Sun's activeness (the level of activity going on in the Sun), and is found to be cyclical, reaching its peak in about every 11 years.

The idea in formulating the input layer structure of a neural network is to consider parameters/factors that affect the output parameter (which is VTEC in this chapter). VTEC has been convincingly proven to be affected by both geomagnetic storm activity [20] and solar activity [21]. The practice in neural networks is to supply parameters like DST and SSN which are

well established to affect VTEC as inputs during the training of a network that will predict VTEC. For this reason, the DST and SSN parameters corresponding to instances of the VTEC data used in this chapter were used as inputs during the training of the neural networks in this chapter.

DST indices were obtained from the World Data Center (WDC) for Geomagnetism (http://wdc.kugi.kyoto-u.ac.jp/dstdir/index.html), while data on SSN were obtained from the WDC-SILSO (Sunspot Index and Long-term Solar Observations, http://www.sidc.be/silso/datafiles), Royal Observatory of Belgium, Brussels.

2.1.3. Neural network training and testing

The Levenberg-Marquardt back-propagation algorithm [22] as implemented in MATLAB was used in this chapter. A couple of other algorithms exist [23] but the Levenberg-Marquardt algorithm is admired for its speed and efficiency in learning [24, 25]. NNs typically have input layers, output layers, and intermediary hidden layers. Each layer could consist of one or more units or nodes (also called neurons).

As explained in the previous section, the idea in formulating the input layer structure of a neural network is to consider parameters/factors that affect the output parameter. In the previous section, the inclusion of DST and SSN as inputs was justified. Other factors that have been established to affect VTEC are time and space; VTEC is known to vary with time and space.

Particularly, VTEC changes with time in the forms of diurnal, seasonal, and long-term yearly variations [7]. For the neural networks to learn long-term yearly variations, the year for each of the GPS VTEC data was included as input for the training. To learn seasonal variations, the day of the year for each of the data was included, and to learn diurnal variations, the hour of day for each data was included.

Spatially, VTEC changes with longitude and latitude of the GPS receiver location, and so the longitudes and latitudes of the GPS receivers were included for each of the GPS VTEC data so that the networks will learn spatial variations of the VTEC. Geomagnetic longitudes and latitudes (rather than geographic longitudes and latitudes) were used since the ionospheric properties are based mainly on the interactions between the solar radiation and the Earth's geomagnetic field [6, 26]. Conversion of geographic to geomagnetic coordinates was done using the Apex Coordinate Conversion Utility Software [27].

In summary, a total of the following seven input nodes were used for the neural network training:

1. Hour of Day (to learn diurnal variations of the VTEC)

2. Day of Year (to learn the seasonal variations)

3. Year (to learn the long-term yearly variations)

4. Longitude (to learn the spatial variations longitude-wise)

5. Latitude (to learn the spatial variations latitude-wise)

6. DST index (to learn variations of the VTEC with geomagnetic storm activity)

7. SSN (to learn variations of the VTEC with solar activity)

The output layer is clearly known to have one neuron which is the GPS-VTEC to be modeled, but deciding the number of neurons in the hidden layer is an intricate aspect of neural network trainings. This is an aspect that conspicuously affects the performance of the trained networks. The most credible practice to deciding an appropriate number of hidden layer neurons has been to train several networks that vary in the number of hidden layer neurons, and then selecting the best of them using a performance index.

In this chapter, 20 neural networks were simulated, varying the number of hidden layer neurons in integer steps from 1 to 20. The main performance index used is the root-mean-squared-errors (RMSEs). RMSEs were computed using the formula in Eq. (6).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\text{GPSVTEC}_i - \text{NNV TEC}_i)^2}{n}} \tag{6}$$

where GPSVTEC$_i$ and NNVTEC$_i$ NNVTEC$_i$ are, respectively, the GPS-VTEC values and the NN-predicted VTEC values, n is the number of samples predicted.

The criteria for deciding the best network is to choose the one that gives the least RMSE on the test dataset. Testing of the networks was done using 15% dataset that was randomly selected from the entire data and which were not used for the training. Another randomly selected 15% of the data was used for validation during the training, and the remaining 70% was used for the actual training. **Figure 2** illustrates outcomes of the RMSEs when different number of hidden layer neurons were used on the networks.

Figure 2 shows that the network that gave the least RMSE is the network that has 6 hidden layer neurons. The RMSE for this network is 5.03 TECU. It is this network that has been adopted as the optimal network in this study. A detailed and elementary treatment on how to train neural networks using MATLAB is contained in a more elementary book [28].

2.2. Results and discussions

2.2.1. Sample simulations

Using the Neural Network model developed in this chapter, sample simulations were made to assess predictions from the model in terms of known ionospheric variation patterns.

2.2.1.1. Diurnal variations

Diurnal variations of the ionosphere are variations in the ionosphere that are observed as the Earth makes a complete rotation about its axis. That is, the changes that are observed within an entire day as we go from morning to night. **Figure 3(a)–(d)** are constructed to visualize
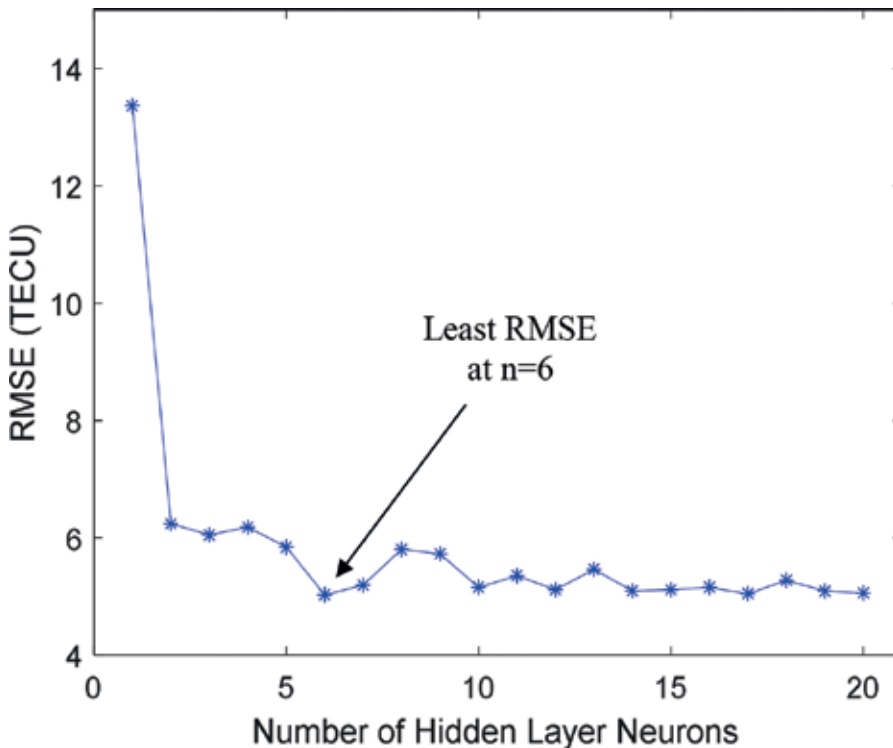
Figure 2. Plot of the RMSEs for varied number of hidden layer neurons.

diurnal variations in the ionosphere over the Nigerian region. The figures are, respectively, images of the VTEC over Nigeria for 05:00 UT (06:00 local Nigerian time, which is around local sunrise), 11:00 UT (12:00 local Nigerian time, which is around local midday), 17:00 UT (18:00 local Nigerian time, which is around local sunset), and 23:00 UT (24:00 local Nigerian time, which is around local midnight) of 1st July 2014. The day was arbitrarily chosen for this illustration. Local time in Nigeria is UT + 1. In the color scheme used for the figure (and for all other figures in this chapter), the blue colors indicate lower VTECs, the red colors indicate higher VTECs, and the green-yellow colors indicate moderate VTECs (see the associated color bars for exact VTEC values in each case).

Figure 3 shows that within a day the VTEC values are greatest around local midday. Since the Sun is the major source of ionospheric ionization, the level of ionospheric ionization (and hence VTEC value) is usually higher during the daytime (when the solar-zenith angle is low) than at nights (when the solar-zenith angle is high). The VTEC values are also relatively high around sunset because the ionizations produced by the Sunlight do not instantly disappear (it takes about 2 hours for the ionized particles to substantially recombine when the Sun goes below horizon).

Figure 3(a) and **(c)** also reveals the interplay between the Sun and the ionosphere during sunrise and sunset. At sunrise (**Figure 3(a)**), the VTECs are higher eastwards than westwards.
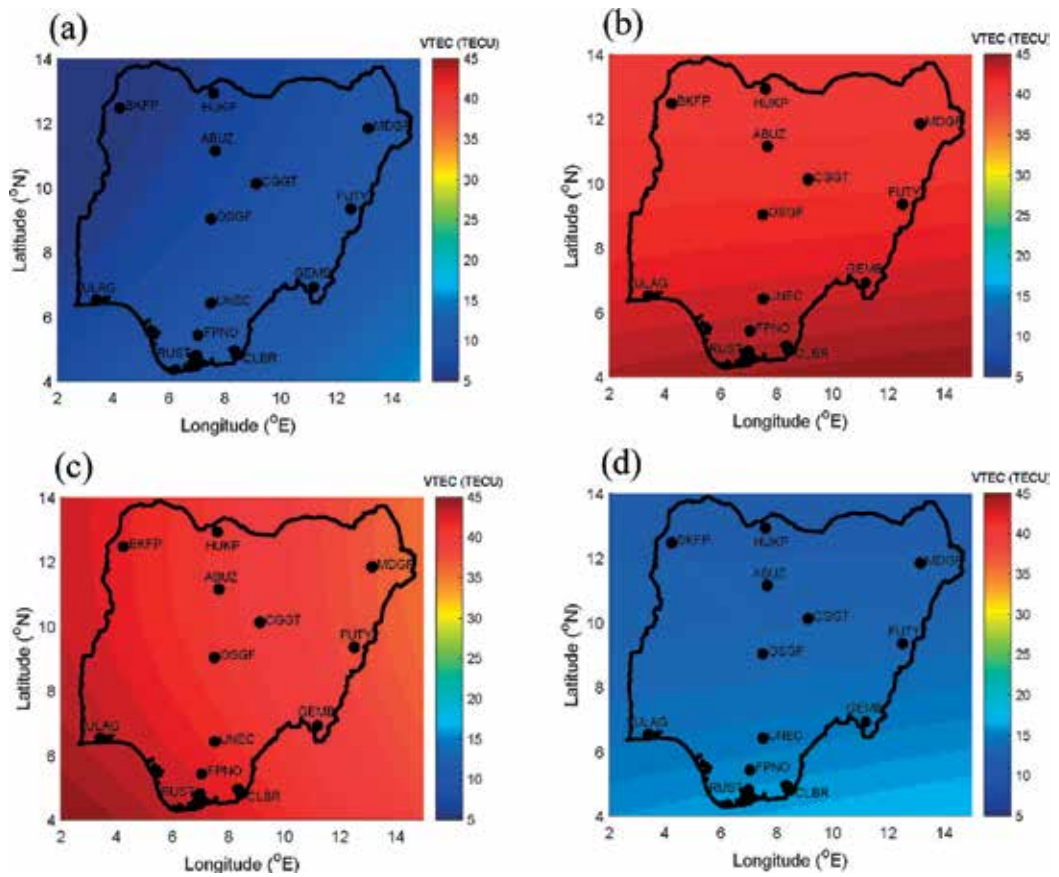
Figure 3. VTEC maps over Nigeria for (a) 06:00 LT, (b) 12:00 LT, (c) 18:00 LT, and (d) 24:00 LT, of 1st July 2014.

This is because the Sun rises from the east. At sunset (**Figure 3(c)**), the VTECs are higher west-wards than eastwards. This is because the Sun sets to the west.

2.2.1.2. Seasonal variations

Seasonal variations have to with variations that are observed as the Earth makes a complete revolution about the Sun. That is, the changes that are observed within an entire year as we go through the seasons. **Figure 4(a)–(d)** are constructed to illustrate seasonal variations over Nigeria during year 2012. The figures are, respectively, VTEC maps of the Nigerian region for 11:00 UT (local midday) of 20th March 2012 (the March equinox day), 21st June 2012 (the June solstice day), 22nd September 2012 (the September equinox day), and 21st December 2012 (the December solstice day). The year 2012 was arbitrarily chosen for the illustration.

Figure 4(a) and **(c)** illustrates that the VTECs are relatively high during the equinoxes. This is because Nigeria is located close to the equator, and as such receives much sunlight during the

equinoxes. During the equinoxes, the solar-zenith angle is lower at the equator as compared to the solstices. It is also conspicuous that the VTEC values are high during the December solstice (**Figure 4(d)**), even higher than at the September equinox (**Figure 4(c)**). This is because Nigeria is mostly located on the geomagnetic southern hemisphere. The December solstice is the summer solstice in the southern hemisphere, and so the solar-zenith angle is lower in the southern hemisphere during this season than at other seasons.

2.2.1.3. Long-term solar cycle variations

The ionosphere has also been established to vary with the level of solar activity. As explained earlier, the sunspot number is a good measure of the level of solar activity. The solar activity is known to have a time series cycle of about 11-years during which the activity level goes from peak to peak or trough to trough.
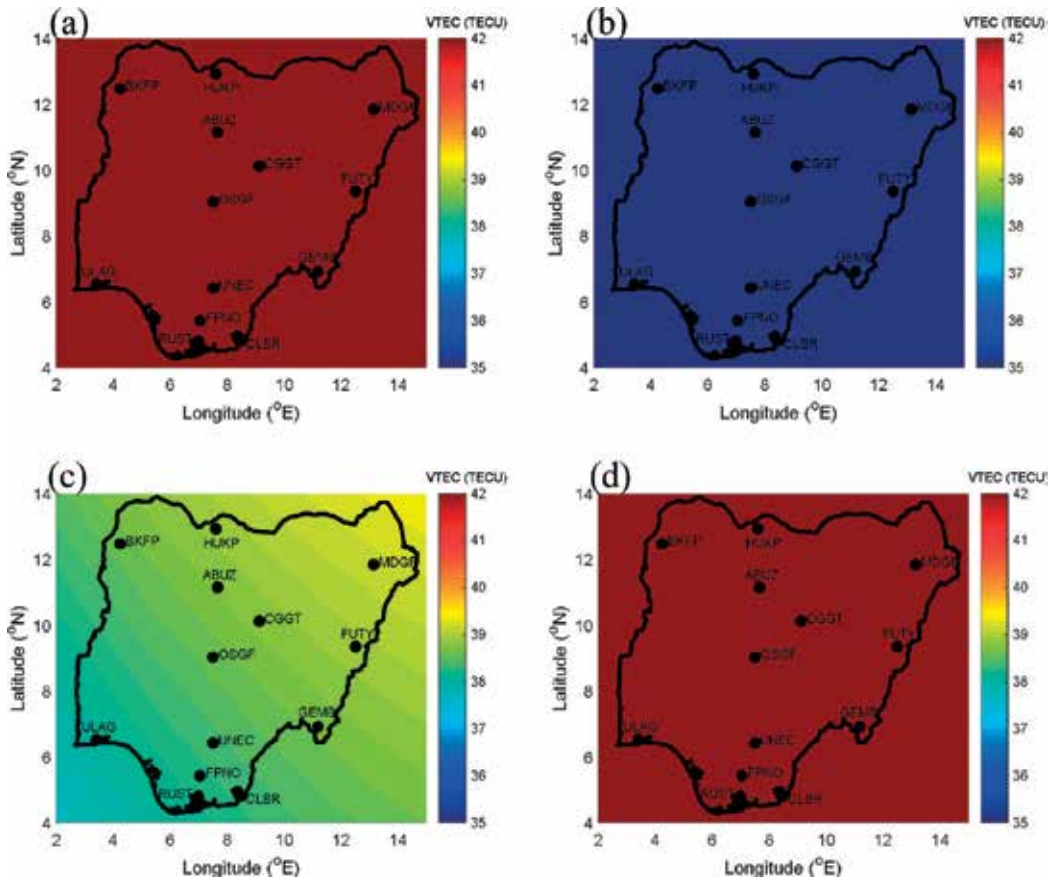

Figure 4. VTEC maps over Nigeria for local midday of (a) 20th March, (b) 21st June, (c) 22nd September, and (d) 21st December, of year 2012.

Figure 5(a)–(d) was constructed to illustrate how the ionosphere over Nigeria varies with the solar activity. The figures are, respectively, the local midday VTECs over Nigeria for the same day (1st July) of years 2011, 2012, 2013, and 2014.

A look at the solar activity cycle shows that the solar activity level was on the rise from year 2011 to year 2014. Comparing with the VTEC values for those years (**Figure 5(a)–(d)**), it is observed that the VTEC values are also on the increase; the VTECs are least during year 2011 (**Figure 5(a)**) which also has the least solar activity level, and greatest during year 2014 (**Figure 5(d)**) which also has the greatest solar activity level. **Figure 5(a)–(d)** clearly indicates that the neural network was able to learn/capture the long-term variations associated with the solar activity.

2.2.2. Comparison of neural network predictions with IRI-Plas and NeQuick predictions

Two of the most popular global ionospheric models (the IRI-Plas and the NeQuick) have been selected to make a comparative assessment of the neural network model developed in this chapter.
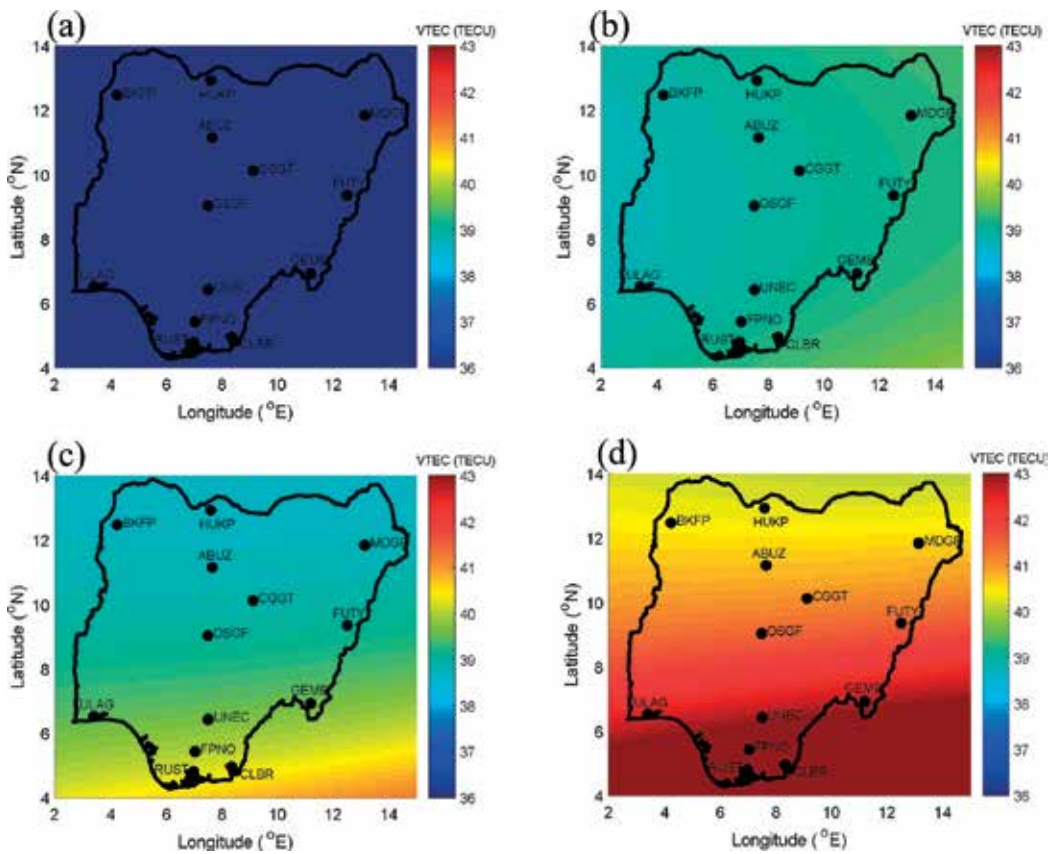

Figure 5. VTEC maps over Nigeria for local midday of 1st July, year (a) 2011, (b) 2012, (c) 2013, and (d) 2014.

The IRI-Plas is the IRI (International Reference Ionosphere) extended to the plasmasphere [29]. The IRI model [3] has been widely accepted as a defector standard for specifying ionospheric parameters across the globe. The IRI-Plas model (rather than the IRI model) is selected for use in this chapter because TEC computed by the IRI-Plas model involves electron density integrations up to the GPS satellite altitudes of about 20,200 km, whereas for the IRI model, it only gets up to a maximum of 2000 km. Since, this chapter concentrates on TEC derived from the GPS, a more comprehensive comparison is therefore obtained using the IRI-Plas model rather than the IRI model. The IRI-Plas model has also been proposed for extension of the IRI model to the plasmasphere [30]. The most recent version of the IRI-Plas (the IRI-Plas 2017) was used for this comparison. The windows executable program of the IRI-Plas used was obtained from the website of the IZMIRAN Institute (http://ftp.izmiran.ru/pub/izmiran/SPIM/).

The NeQuick [31–33] is another popular global ionospheric model which has been severally compared with GNSS TEC measurements and shown to be a good representation of the
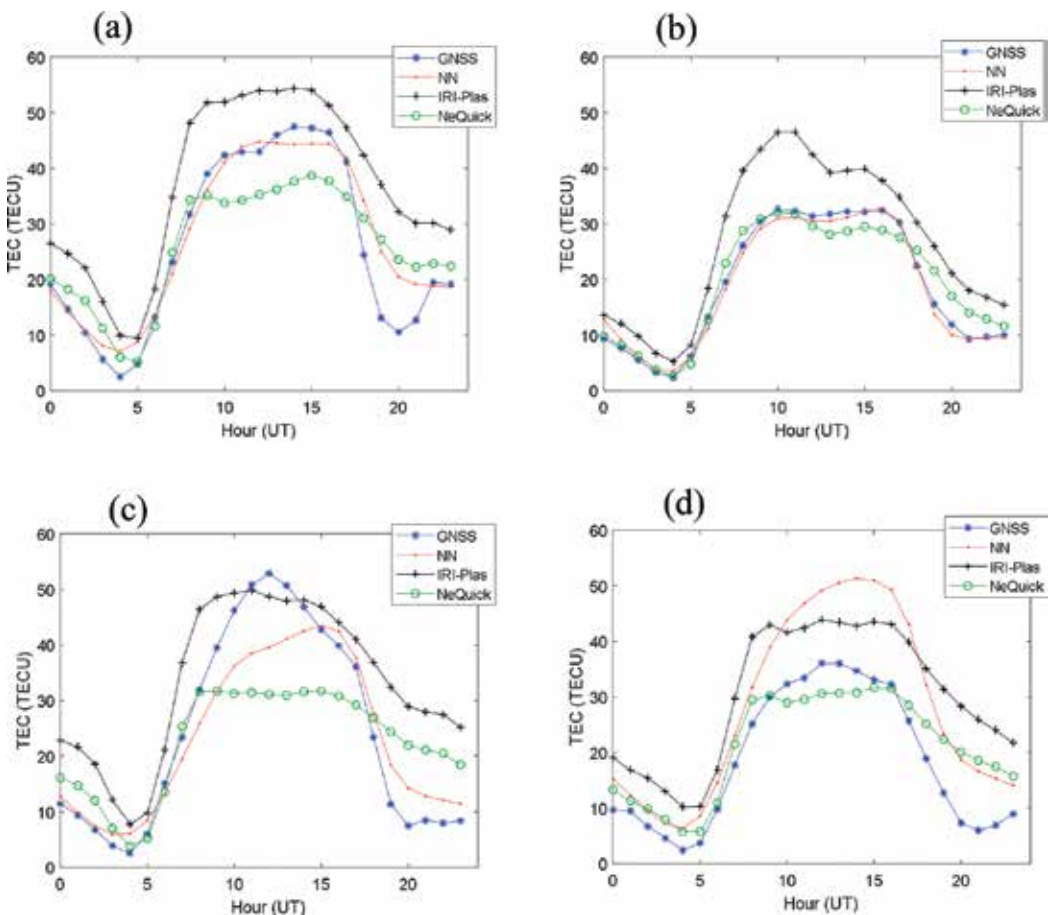

Figure 6. Diurnal VTEC plots over the OSGF stations for days of (a) March Equinox, (b) June Solstice, (c) September Equinox, and (d) December Solstice, in year 2012.

	NN	IRI-Plas	NeQuick
March Equinox	4.4544	12.0488	7.1809
June Solstice	1.2941	8.5256	2.7717
September Equinox	5.9616	11.6634	11.0269
December Solstice	10.7685	12.3848	5.8034

Table 2. Diurnal RMSEs of the 3 models for the days illustrated in **Figure 6**.

ionosphere. The NeQuick is admired because of its improved performance in predicting the topside ionosphere, and consequently versions of the IRI model from 2007 and later have included the topside formulation of the NeQuick, and has adopted it as the most mature of the different proposals to compute the topside part of the IRI electron density profile [33, 34]. The NeQuick includes routines that compute the electron density along any ray-path from ground to GPS satellite altitudes of about 20,200 km, and so also makes for a comprehensive comparison with observations from the GPS. The latest version of the NeQuick (the NeQuick-2, which is currently recommended by the ITU [35] is the one used for this comparison. The NeQuick-2 used in this chapter is the windows executable program created from the FORTRAN source code, and was obtained from the Ionosphere Radio propagation Unit of the T/ICT4D Laboratory (https://t-ict4d.ictp.it/nequick2/source-code).

For the purpose of visual illustration, the diurnal VTEC profiles from GPS observations for four selected days, over the OSGF station, are illustrated in **Figure 6(a)–(d)** alongside corresponding VTEC predictions from the NeQuick, the IRI-Plas model, and the neural network (NN) model developed in this chapter. **Figure 6(a)–(d)**, respectively, represents diurnal VTEC profiles over the OSGF station for 20th March 2012 (the March equinox day), 21st June 2012 (the June solstice day), 22nd September 2012 (the September equinox day), and 21st December 2012 (the December solstice day).

Figure 6 clearly indicates that the VTEC predictions from the NN model developed in this chapter were closer to the GPS VTEC observations in most of the times than the VTEC predictions of the IRI-Plas and NeQuick. **Table 2** summarizes the RMSEs (computed using Eq. (6)) for each of the days and models illustrated in **Figure 6**. The RMSEs for each of the models were computed with reference to the GPS observations. **Table 2** shows that the prediction errors for the NN model were predominantly lower than for the other two models, except for the December solstice day when the NeQuick prediction error was lower.

Asides the demonstrated capability of neural networks to very accurately learn and predict variations in the ionosphere, the better performance of the NN model could also be linked to the fact that more volume of regional GPS data (GPS data from the Nigerian region) were used in the NN model than the volume used in either of the NeQuick or IRI-Plas models.

3. Conclusion

A regional VTEC model over Nigeria was developed using the method of computer neural networks and GPS-VTEC data from 14 stations spanning the period from years 2011 to 2016. A total of seven input layer neurons (namely, Year, Day of Year, Hour of Day, Geomagnetic Longitude, Geomagnetic Latitude, SSN, and DST indices) were used to learn the studied output (GPS-TEC). By simulating 20 different networks that differed in their number of hidden layer neurons, the network with 6 hidden layer neuron was determined to be the best in terms of minimizing the prediction errors (using the RMSE as criterion for measuring the prediction error).

The neural network model was demonstrated to be proficient in predicting the VTEC variation patterns in terms of diurnal variations, seasonal variations, long-term solar cycle variations, and spatial variations across Nigeria.

When compared with two popular global ionospheric models (the NeQuick and the IRI-Plas), predictions from the neural network model was observed to be more accurate in terms of closeness to the GPS-VTEC values. Typical RMSEs for the neural network model predictions were between 1.3 and 10.8 TECU, the mean RMSE was 5.6 TECU. For the IRI-Plas model, the RMSEs were between 8.5 and 12.4 TECU, and the mean was 11.2 TECU. For the NeQuick, the RMSEs were between 2.8 and 11.0 TECU, and the mean was 6.7 TECU. The work done in this chapter further validates neural networks as excellent candidates for modeling of ionospheric parameters.

Acknowledgements

Author details

Daniel Okoh

Address all correspondence to: okodan2003@gmail.com

Space Environment Research Laboratory, Centre for Atmospheric Research, National Space Research and Development Agency, Abuja, Nigeria

References

[1] Chiu YT. An improved phenomenological model of ionospheric density. Journal of Atmospheric and Terrestrial Physics. 1975;**37**:1563-1570

[2] Roble RG, Ridley EC, Richmond AD. A coupled thermosphere/ionosphere general circulation model. Geophysical Research Letters. 1988;**15**(12):1325-1328

[3] Bilitza D. International Reference Ionosphere 1990. Greenbelt Maryland: National Space Science Data Center 90-22; 1990

[4] Bailey GJ, Balan N, Su YZ. The Sheffield University plasmasphere ionosphere model-a review. Journal of Atmospheric and Solar - Terrestrial Physics. 1997;**59**(13):1541-1552

[5] Huba JD, Joyce G, Fedder JA. Sami2 is another model of the ionosphere (SAMI2) - a new low-latitude ionosphere model. Journal of Geophysical Research. 2000;**105**(A10): 23035-23053

[6] Habarulema JB. A contribution to TEC modelling over Southern Africa using GPS data [PhD Thesis]. Rhodes University; 2010

[7] Okoh D, Owolabi O, Ekechukwu C, Folarin O, Arhiwo G, Agbo J, Bolaji S, Rabiu B. A regional GNSS-VTEC model over Nigeria using neural networks: A novel approach. Geod Geodyn. 2016;**7**(1):19-31

[8] NovAtel, An Introduction to GNSS [Internet]. 2010. Available from: http://www.borealisprecision.com/pdf/An_Introduction_to_GNSS.pdf [Accessed: 2017-12-07]

[9] Marel HVD, Bakker PFD. Single versus Dual Frequency Precise Point Positioning. GNSS Solutions, in InsideGNSS [Internet]. 2012. Available from: www.insidegnss.com/auto/julyaug12-Solutions.pdf [Accessed: 2017-12-11]

[10] Leandro RF, Santos MC. A neural network approach for regional vertical total electron content modeling. Studia Geophysica et Geodaetica. 2007;**51**:279-292

[11] Hernandez-Pajares M, Juan J, Sanz J. Neural network modeling of the ionospheric electron content at global scale using GPS. Radio Science. 1997;**32**:1081-1090

[12] Tulunay E, Senalp ET, Cander LR, Tulunay YK, Bilge AH, Mizrahi E, et al. Development of algorithms and software for forecasting, nowcasting and variability of TEC. Annales de Geophysique. 2004;**47**:1201-1214

[13] Senalp ET, Tulunay E, Tulunay Y. Total electron content (TEC) forecasting by Cascade modeling, a possible alternative to the IRI-2001. Radio Science. 2008;**43**:RS4016. DOI: 10.1029/2007RS003719.2008

[14] Klobuchar JA. Ionospheric effects on GPS. In: Parkinson BW, Spilker JJ, editors. Global Positioning System: Theory and Applications, Vol. 2. Progress in Astronautics and Aeronautics. 1996. p. 164

[15] Rao PVS, Gopi KS, Niranjan K, Prasad SVVD. Temporal and spatial variations in TEC using simultaneous measurements from the Indian network of receivers during the low solar activity period of 2004-2005. Annales de Geophysique. 2006;**24**:3279-3292

[16] Jin SG, Occhipinti G, Jin R. GNSS ionospheric seismology: Recent observation evidences and characteristics. Earth Science Reviews. 2015;**147**:54-64

[17] Zhu FY, Wu Y, Lin J, Zhou Y. Temporal and spatial characteristics of VTEC anomalies before Wenchuan Ms8.0 earthquake. Geodesy and Geodynamics. 2010;**1**(1):23-28

[18] Mannucci AJ, Wilson BD, Edwards CD A new method for monitoring the earth's ionospheric total electron content using the GPS global network. In: Proc. of ION GPS-93. Inst Navigation; 1993. pp. 1323-1332

[19] Seemala GK, Valladares CE. Statistics of total electron content depletions observed over the south American continent for the year 2008. Radio Science. 2011;**46**:RS5019. DOI: 10.1029/2011RS004722

[20] Malik R, Sarkar S, Mukherjee S, Gwal AK. Study of ionospheric variability during geomagnetic storms. Journal of Indian Geophysical Union. 2010;**14**(1):47-56

[21] Correia E, Paz AJ, Gende MA. Characterization of GPS total electron content (GPS-TEC) in Antarctica from 2004 to 2011. Annals of Geophysics. 2013;**56**(2):R0217

[22] Levenberg K. A method for the solution of certain nonlinear problems in least squares. Quarterly of Applied Mathematics. 1944;**2**:164-168

[23] Quesada AA. Five Algorithms to train a neural network [Internet]. 2017. Available from: https://www.neuraldesigner.com/blog/5_algorithms_to_train_a_neural_network [Accessed: 2017-12-07]

[24] Demuth H, Beale M. Neural Network Toolbox for Use with MATLAB. Natick, MA, USA: Mathworks Inc; 2002

[25] Kisi O, Uncuoglu E. Comparison of three back-propagation training algorithms for two case studies. Indian Journal of Engineering and Materials Science. 2005;**12**:434-442

[26] Komjathy A. Global ionospheric total electron content mapping using the Global Positioning System [PhD Thesis]. New Brunswick, Canada: Department of Geodesy and Geomatics Engineering Technical Report No. 188, University of New Brunswick, Fredericton; 1997

[27] Richmond A, Emmert J, Nair M, Maus S, Woods A, Wickwar V, Barnes R, Maute. Apex Coordinate Conversion Utility [Internet]. 2011. Available from: http://www.ngdc.noaa.gov/geomag/geom_util/apex.shtml [Accessed: 2016-11-20]

[28] Okoh D. Computer Neural Networks on MATLAB. North Charleston, SC, USA: Createspace; 2016. p. 52. ISBN-13: 978-1539360957

[29] Gulyaeva TL, Huang X, Reinisch BW. Plasmaspheric extension of topside electron density profiles. Advances in Space Research. 2002;**29**(6):825-831

[30] Gulyaeva TL, Bilitza D. Towards ISO standard earth ionosphere and Plasmasphere model. In: Larsen RJ, editor. New Developments in the Standard Model. Hauppauge, New York, USA: Nova Science Publishers; 2012. pp. 1-39

[31] Hochegger G, Nava B, Radicella SM, Leitinger R. A family of ionospheric models for different uses. Physics and Chemistry of the Earth. 2000;**25**(4):307-310

[32] Radicella SM, Leitinger R. The evolution of the DGR approach to model electron density profiles. Advances in Space Research. 2001;**27**:35-40. DOI: 10.1016/S0273-1177(00)00138-1

[33] Nava B, Coïsson P, Radicella SM. A new version of the NeQuick ionosphere electron density model. Journal of Atmospheric and Solar - Terrestrial Physics. 2008;**70**(15):1856-1862

[34] Bilitza D, Reinisch BW. International reference ionosphere 2007: Improvements and new parameters. Advances in Space Research. 2008;**42**(4):599-609

[35] Nava B, Radicella SM. The NeQuick Model: Characteristics and Uses. In: IGS Workshop. Pasadena; 23-27 June 2014

Technological Progress of Navigation System

GNSS Error Sources

Malek Karaim, Mohamed Elsheikh and
Aboelmagd Noureldin

Additional information is available at the end of the chapter

Abstract

This chapter discusses the most serious sources of error affecting global navigation satellite systems (GNSS) signals, classifying these in a new way, according to their nature and/or effects. For instance, errors due to clock bias or drift are grouped together. Errors related to the signal propagation medium, too, are treated in the same way. GNSS errors need to be corrected to achieve accepted positioning and navigational accuracy. We provide a theoretical description for each source, supporting these with diagrams and analytical figures where possible. Some common metrics to measure the magnitude of GNSS errors, including the user equivalent range error (UERE) and the dilution of precision (DOP), are also presented. The chapter concludes with remarks on the significance of the sources of error.

Keywords: GNSS, errors, ionosphere, troposphere, multipath, clock, jamming, noise

1. Introduction

The services provided by global navigation satellite systems (GNSS) are used in a massive number of applications, both civilian and military. All GNSS systems comprise many satellites orbiting the Earth at very high elevations. At a single point in time, there will be several satellites from which a receiver may have a clear line of sight to receive signals and build its own navigation solution. However, these signals are prone to several sources of disturbance, causing errors in the measurements that are generated inside the receiver, which in turn degrades positioning accuracy.

Most of the discussions here apply to all GNSS systems, but in some instances, we use the US GNSS system—the Global Positioning System (GPS)—as an example to explain our ideas.

Figure 1 depicts the structure of a typical GNSS system—GPS—with its three primary segments. In all GNSS systems, the signal makes a journey of thousands of kilometers between the satellite antenna and its destination, the receiver. The first and longest part of this trip is through space where the signal attains its characteristics. Nonetheless, when the signal travels through the atmosphere, this medium imposes some undesirable effects. The layers of the atmosphere add delays to signal propagation time, causing some errors in the measurements.

Once the signal nears the receiver antenna, it usually experiences some reflections and echoes, i.e. the signal often bounces off objects surrounding the receiver, potentially hitting the antenna multiple times—this phenomenon is known as multipath. Multipath is one of the major error sources that can be very harmful to GNSS signals in many applications [1]. All the abovementioned signal disturbances result from the nature of the signal or the propagation medium and are unintentional. Intentional signal degradation or replacement could be in many cases a tougher source of GNSS errors. One major type of deliberate errors is signal jamming. Signal jamming is deliberate interference caused by the broadcasting of radio frequency (RF) signals near the receiver with the aim of preventing the tracking of GNSS signals. Some other less harmful error sources are discussed in this chapter, including system (circuit) errors and satellite orbital errors.

In general, this chapter discusses thoroughly the major sources of GNSS error sources, their causes, consequences, and scales. Each error source or factor is explained in depth, with supporting figures whenever possible. Another contribution of this chapter is the presentation of a new scheme for categorizing GNSS errors.
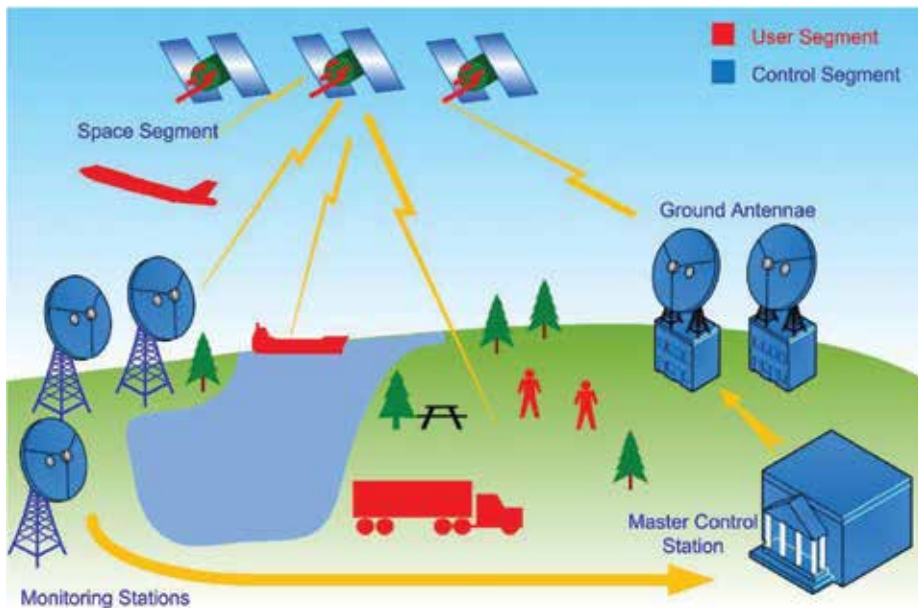

Figure 1. The three GPS segments. Courtesy of Noureldin et al. [2].

It is worth mentioning that in this work, we address the error sources that affect the standard point positioning (SPP) accuracy level where the receiver uses the broadcast ephemeris information and single-frequency measurements to estimate its position with a meter-level accuracy. However, there are other error sources that affect accuracy within the centimeter and millimeter level such as antenna phase center, phase wind-up, and site displacement errors. To achieve this degree of accuracy, the receiver needs to work in either differential GNSS mode or precise point positioning (PPP) mode, both of which are beyond the scope of this chapter.

2. Error sources and consequences

GNSS signals have very low power, and hence they are prone to several sources of noise and errors. The range measured by the GNSS receiver is contaminated by these errors, which is why it is called the *pseudorange*. The general pseudorange observation equation is expressed as follows:

$$P_r^s = \rho_r^s + c\left(dt_r - dT^s\right) + I_r^s + T_r^s + \varepsilon_r^s \tag{1}$$

where P_r^s is the pseudorange between the satellite s and the receiver r. ρ_r^s is the true geometric range, c is the speed of light, and dt_r and dT^s represent the receiver and satellite clock errors in seconds. The symbol I denotes the ionospheric delay, while T is the tropospheric delay in meters. Finally, ε combines the multipath and receiver noise errors.

In this section, GNSS errors will be categorized based on the nature of the error itself. Timing-related errors in both the satellite and receiver are grouped as clock-related errors. Signal propagation errors combine atmospheric errors, multipath errors, and the effect of the relative motion between the satellite and receiver. Satellite orbit parameters needed to calculate satellite position and velocity are estimated at the control segment. These parameters are sent to GNSS satellites to be broadcast in the navigation message. This estimation error is combined with the receiver noise effect as system errors. The last type of GNSS error is intentional errors. Those errors are, however, deliberate and can be harmful; these include signal jamming and spoofing.

2.1. Clock-related errors

Receivers generate measurements based mainly on measuring time [3]. Indeed, time is central to GNSS systems; therefore, GNSS satellites are equipped with very precise, and hence very expensive, clocks [4]. Despite their accuracy, satellite clocks still drift slightly from GNSS time. For reasons of affordability and size, receiver clocks are usually much cheaper; consequently, they drift from GNSS time rapidly. This drift translates into dramatic range errors in receiver measurements. Accordingly, it is significant to correct or compensate for timing errors in the GNSS signal. These clock errors can be summarized as follows:

2.1.1. Satellite clock errors

There are three factors that can affect GNSS satellite clocks: stability, relativistic effects, and timing group delay.

Clock stability: The stability of a satellite clock is about 1 to 2 parts in 10^{13} a day, which is approximately 8.64 to 17.28 ns a day. This is equivalent to a range error of about 2.59 m to 5.18 m [5]. This instability $dT^{s'}$ is modeled using the quadratic function:

$$dT^{s'} = a_{f0} + a_{f1}(t - t_{oc}) + a_{f2}(t - t_{oc})^2 \tag{2}$$

where t is the receiver GPS time, t_{oc} is the reference epoch time, a_{f0} is the clock offset, a_{f1} is the clock drift coefficient, and a_{f2} is the clock drift rate coefficient. The values of t_{oc}, a_{f0}, a_{f1}, and a_{f2} are obtained from the broadcasted navigation message.

Relativistic effects: A clock aboard a satellite will be affected by both the general and special relativity theories. The net result is that this clock will appear to run faster than the same clock on Earth by approximately 38.4 μs/day. Scaled by the speed of light, this is equivalent to a range error of about 11,512 m. To compensate for this effect, a proper offset is introduced to the satellite clock rate before launching [4]. However, there is still a residual effect because of the noncircular satellite orbit, which should be compensated for at the user side. This relativistic correction Δt_r is calculated by [6].

$$\Delta t_r = -\frac{2}{c^2}\sqrt{\mu a}(e\sin E) \tag{3}$$

where c is the speed of light, $\mu = 3.986005 \times 10^{14}$ m³/s² is the Earth's universal gravitational parameter for GPS, a is the Earth's semimajor axis, e is the eccentricity of the satellite orbit, and E is the eccentric anomaly of the satellite orbit. If the orbit was a perfect circle, this effect is zero as the eccentricity is zero. For instance, for an eccentricity of 0.015, the maximum value will be 16.8 ns, which corresponds to around 5 m. Another alternative equation to Eq. (3) is to use the satellite position and velocity to calculate the relativistic correction using the following formula [7]:

$$\Delta t_r = -\frac{2 r^s \cdot v^s}{c^2} \tag{4}$$

where $r^s \cdot v^s$ is the dot product of the satellite position and velocity vectors.

Timing group delay (TGD): the satellite clock corrections in the navigation message are referred to one GNSS signal or signal combination. In the case of GPS, this signal is the ionospheric-free combination of the codes at L1 and L2 frequency bands. In the case of a single-frequency operation, a correction should be made to compensate for the bias offset between L1 and the ionospheric-free combination signals. This correction is also provided in the navigation message, named as timing group delay (TGD) [7].

The total satellite clock error is now calculated by the sum of the three terms as

$$dT^s = dT^{s'} + \Delta t_r + TGD \tag{5}$$

2.1.2. Receiver clock errors

GNSS receivers are equipped with inexpensive crystal clocks, which have low accuracy compared to satellite clocks [8]. As a result, the receiver clock error is much larger than that of the GNSS satellite clock. There are two ways to fix this issue. One is to use external precise, usually cesium or rubidium, clocks which have superior performance, but the problem is that they are very expensive, as they cost between a few thousand dollars to about $20,000 [4].

The other solution, which is much more common, is to remove this error through differencing between satellites or by estimating the error as an additional unknown parameter in the position estimation process. This latter solution is meant to make receiver prices affordable [2]. Adding the receiver clock bias to the set of unknowns, in addition to three position parameters, sets the limitation to a minimum of four visible satellites, instead of three, for obtaining a solution from the receiver.

To prevent the receiver clock error from becoming too large, receiver manufacturers apply a clock-steering mechanism. Two main approaches are used for this [9]. The first method is continuous steering to keep the clock error within the acceptable range. The other method is clock jumping, where clock bias is adjusted only when the error reaches a certain threshold. Although the clock bias is estimated as an unknown parameter in the estimation filter, it should still be kept within a certain limit. The reason for this is that the receiver clock is used to time tag the receiver output. This time tag must have a minimum level of accuracy for time synchronization between different systems to occur.

2.1.3. Intersystem biases

One way to enhance the accuracy and the availability of the GNSS receiver solution is to use all the observations from all available GNSS constellations. GPS and the Russian global navigation satellite system, GLONASS, are currently fully operational systems with global coverage, while other systems are now evolving to achieve the worldwide coverage such as the European Galileo and the Chinese BeiDuo systems.

Each GNSS has its own timing system, and hence, there are some intersystem clock biases that should be considered when dealing with a multi-constellation system. This can be achieved by introducing new unknowns, which represent the time difference between the added GNSS constellation time and GPS time [10]. For example, if GLONASS measurements are to be used, then the receiver clock bias in Eq. (1) can now be represented as $dt_r = dt_{r,GPS} + dt_{r,GPS-GLONASS}$. As the number of unknowns is increased to five, this will require a minimum of five visible satellites from both constellations.

2.2. Signal propagation errors

During signal propagation time, the Earth would have rotated, causing a relative shift between the satellite and receiver locations at signal transmission time and signal reception time. If not accounted for, this relative distance, known as the Sagnac effect, will cause an extra error in the measured range.

Furthermore, the GNSS signal travels a long trip between the satellite and the receiver. The first and longest part of the GNSS signal journey is through space where the signal preserves its original characteristics, foremost of which is its constant speed. At lower altitudes, however, the signal will experience some disturbances, e.g., ionosphere and troposphere effects. Moreover, during the final part of the signal path, the GNSS signal arrives directly at the receiver or via single or multiple reflections from the surrounding objects. This multipath effect is not deterministic and can degrade the signal dramatically. This section covers the factors that affect the signal throughout its journey between the satellite and the receiver.

2.2.1. Sagnac effect

The Sagnac effect is a relativistic error caused by the Earth's rotation during signal propagation time between the satellite and the receiver [11]. Ephemeris parameters obtained from the navigation message provide information about the satellite position expressed in the Earth-centered Earth-fixed (ECEF) frame at signal transmission time. However, during signal transit time, the Earth would have rotated (see **Figure 2**) and, hence, the ECEF frame; consequently, a correction is needed to express the satellite position in the ECEF frame at signal reception time instead of transmission time [2]. The amount of frame rotation during the signal transit time is $w_e(t_r\text{-}t_t)$, where w_e is the Earth rotation rate, t_r is the signal reception time, and t_t is the signal transmission time.

Although this error is not directly observable in Eq. (1), it is inherent in calculating the geometric range ρ_r^s. The geometric range is calculated as the difference between the receiver position and the satellites' position, and by adding the Sagnac correction, it can be written as

$$\rho_r^s = \left\| r_r(t_r) - R_z\big(w_e(t_r - t_t)\big) \cdot r_s(t_t) \right\| \tag{6}$$

where r_r is the receiver position vector and r_s is the satellite position vector, both in ECEF frame. $\|\cdot\|$ is the norm, operator and $R_z(\theta)$ is the coordinate rotation matrix around the z-axis of ECEF frame by an angle θ which is defined as

$$R_z(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

If left uncompensated for, this effect could cause a position error of about 30 m [12].

2.2.2. Ionosphere errors

When the signal reaches an altitude of about 1000 km above the Earth's surface, it penetrates the upper layer of the atmosphere, namely, the ionosphere (see **Figure 3**). This layer of atmosphere includes various types of gases that are readily ionized by the sun's radiation [4]. The intensity of solar activity is the key factor determining the condition of the ionosphere, but it is also affected by season and time of day. Accordingly, these three parameters define the level of ionization, thereby changing the refractive indices of the layers of the ionosphere, therefore, influencing the signal transit time measured by the receiver [8].
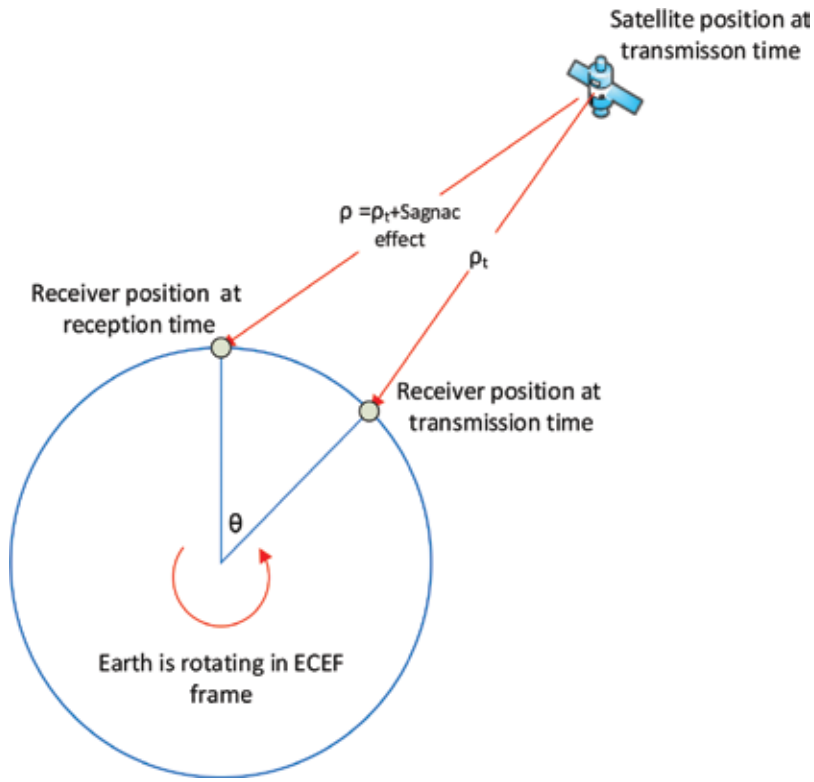
Figure 2. The Sagnac effect.

The ionosphere acts as a dispersive medium, meaning that the ionospheric delay is frequency dependent. This delay represents one of the significant ranging errors in GNSS positioning and can reach a value of 300 ns (100 m) in some situations [13]. The first-order ionospheric delay I, in meters, is represented by the equation:

$$I = \frac{40.3 \cdot TEC}{f^2} \tag{8}$$

where TEC is the total electron content which is defined as the number of electrons in a tube of 1 m^2 cross section in the signal propagation direction and f is the signal frequency.

For dual-frequency receivers, using the ionospheric-free signal combinations, this first-order error can be removed and with it 99.99% of the ionospheric delay [14]. On the other hand, in single-frequency receivers, the ionospheric delay must be modeled or estimated. The simplest way is to use the broadcast models transmitted in the satellite navigation message, such as GPS Klobuchar model [13] and Galileo NeQuick model [15]. Nevertheless, these models can correct for approximately 50% rms of the ionospheric error; even the most accurate theoretical model can only correct up to 80% of this error [13].
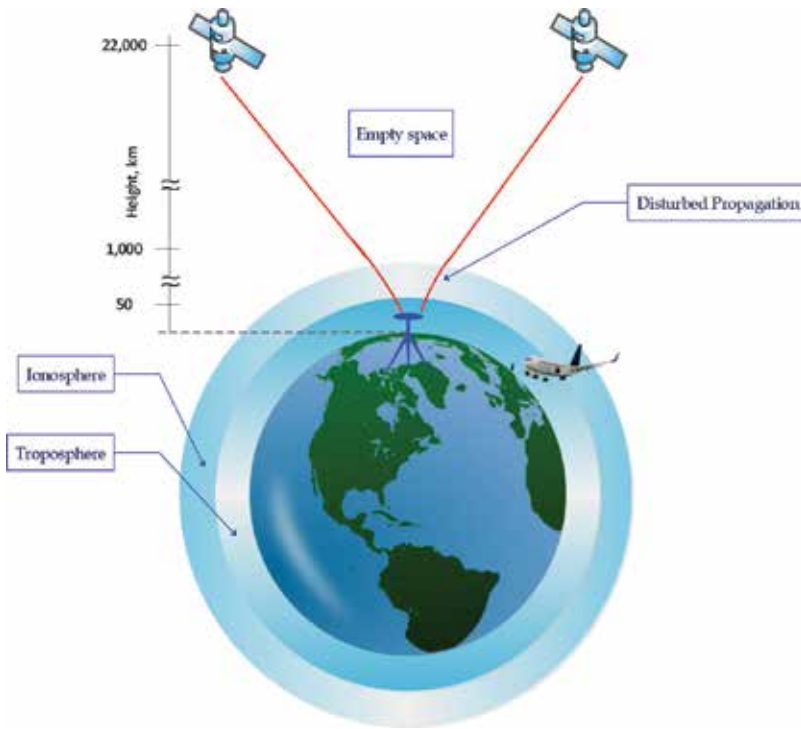
Figure 3. The GPS signal's propagation mediums [18].

The other approach is to use a network of global or local dual-frequency receivers to esti-mate ionospheric corrections in a grid model. This network usually estimates the vertical TEC (VTEC) and sends these corrections to the users. The satellite-based augmentation network (SBAS) corrections provided by the American Wide Area Augmentation System (WAAS) are an example of such corrections. The VTEC can be used to obtain the total slant TEC through an obliquity factor that accounts for the effect of the satellite elevation angle [16]. If the single-frequency receiver has the capability of receiving these corrections, ionospheric error model-ing will be more accurate than using broadcast models.

2.2.3. Troposphere errors

The next step is for the signal to move through the troposphere, the lowest part of the atmo-sphere, extending from the Earth's surface up to a maximum height of 20 km above sea level (see **Figure 3**). This part of the atmosphere is composed of dry gases and water vapor [16]. Since it is a refractive layer, the troposphere, too, delays GNSS signals; however, being electrically neutral, this layer is nondispersive for some GNSS frequencies [10]. The tropospheric delay has two components: wet and dry. The wet one is difficult to model, but luckily, it accounts for only 10 percent of the delay. The dry one, which is responsible for the rest of the delay, can be more easily modeled. The tropospheric delay is frequency independent; therefore, unlike

the ionospheric delay, it cannot be removed by combining measurements from L1 and L2 GPS signals. Depending on satellite elevation, the tropospheric delay adds up about 2.5 m to 25 m to range measurements [4].

For meter-level accuracy, several models can be used to mitigate the total tropospheric error, such as Hopfield model and Saastamoinen model. These models usually calculate the zenith delay (for elevation angle = 0) and then use a mapping function to obtain the total slant delay, depending on the satellite elevation angle [17]. For applications that need a higher level of accuracy in tropospheric error estimation, the dry component is modeled, while the zenith wet component is estimated as an additional unknown in the navigation filter.

2.2.4. Multipath errors

As the signal nears the receiver antenna, it can often be further degraded. In several scenarios, the signal may reach the receiver's antenna via more than one path (see **Figure 4**), owing to signal reflections from surrounding structures or the ground [19]. Usually, one of the received signals would be the direct line-of-sight (LOS) signal, along with one or more of its echoes, which are delayed versions of the original signal. Those delayed versions are superimposed on the LOS signal, which can significantly distort the desired LOS signal. The multipath effect depends on the surrounding environment and the relative satellite-receiver motion. Moreover, in general, this effect cannot be canceled through differential positioning—even for closely spaced receivers. Therefore, the multipath error can limit positioning accuracy even if the other error sources have been removed. In the most severe conditions, the multipath error can cause a pseudorange error of up to 100 m [3].

One solution to avoid this source of error is to place the receiver antenna in a reflection-free location; however, this is not always practical, particularly when the GNSS receiver is on a moving platform. Another way to mitigate multipath error is through the receiver or antenna design. The "choke ring" antenna is one of the best-known antennas that mitigates multipath [20]. Other designs were made to keep the same high performance of the "choke ring" with lighter weight and smaller size [21]. Some modern receivers use techniques relying on multiple antennas or what is known as an antenna array. With such technology, the receiver can tune itself to track only the LOS signal and block all other replicas of the signal [22]. The multipath effect can also be mitigated at the measurement level while processing data. The simple way is by weighting the measurements according to the elevation angle, since the multipath error increases at lower elevation angles [1]. A more advanced approach is to detect the multipath effect using code-phase information, such as the code minus carrier observation. This data can be used to adjust satellite weighting or even to reject some measurements with severe multipath effects [23].

2.3. System errors

Some GNSS errors result from the overall nature of the system, e.g., the shape of orbital planes and receiver structure. These error sources are discussed in this section.
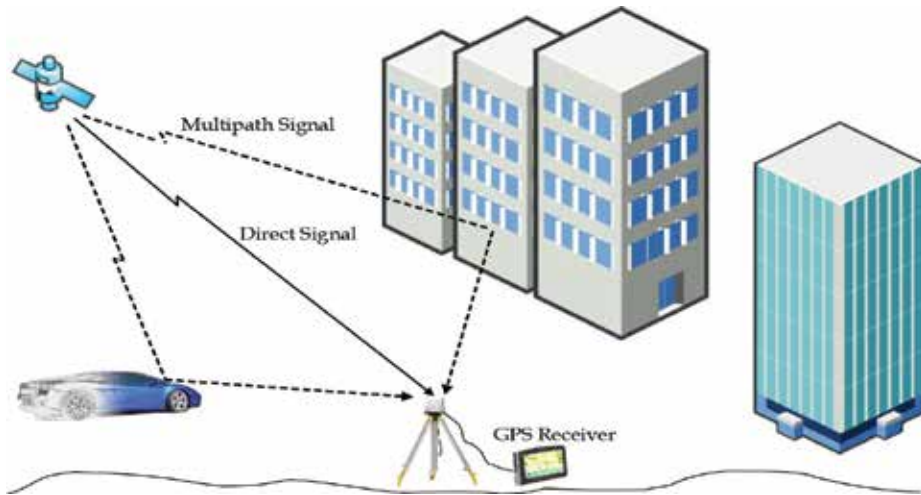
Figure 4. Line-of-sight (direct) and multipath (indirect) signals [18].

2.3.1. Satellite orbital errors

Receivers calculate satellite position based on information contained in the navigation message known as satellite ephemeris. These ephemeris parameters are estimated at the control segment and then uploaded to satellites. Satellites broadcast updated ephemeris data every 2 hours; however, these parameters are estimated using a curve fit to predict the satellite orbit, which leaves residual errors relative to the actual orbit [2]. This error source introduces a root mean square (RMS) error of about 2 m [4]. This error can be mitigated if global or local network corrections for the satellite position are available. These corrections are used to refine the broadcast ephemeris corrections and, hence, improve accuracy. For post-processing, a more precise ephemeris, available from IGS [24], can be used if centimeter-level precision is required and a dual-frequency receiver was used.

2.3.2. Receiver noise

Receiver noise is a complex error generated at the receiver's side while measuring satellite signals. It covers a broad spectrum of noise types, including but not limited to microwave radiations sensed by the antenna in the band of interest unrelated to the signal; noise introduced by system components such as the antenna, cables, and amplifiers; and signal quantization noise [25]. Receiver noise is considered white noise; therefore, it cannot be avoided entirely. However, with modern receiver technology, this term is lessened to about 0.1–1% of a cycle in the carrier phase and d of centimeters in pseudorange measurements. The contradiction here is that receiver noise increases by $\sqrt{2}$ for single-differenced observations, while double-differenced ones have a noise amplification of two [26]. Observation differencing is sometimes used to cancel the common between-receiver errors, between- satellite errors, or both.

2.4. Intentional error sources

Some GNSS error sources are deliberate, i.e., imposed by the service provider or an attack on the system. These are discussed in the following subsections.

2.4.1. Selective availability

Selective availability is associated with only the GPS system among all the GNSS systems. Selective availability (SA) was an intentional degradation of GPS performance by the US government for national security reasons. Satellite clock corrections in the broadcast ephemeris were deliberately degraded to reduce the accuracy for civilian use of GPS to an accuracy level of 100 m for the horizontal position [20]. However, on 2 May 2000, this feature was discontinued, and the USA announced that it would no longer impose this. Furthermore, the new generation of GPS satellites (GPS III) will not have this feature, meaning that SA cannot be used by the US government anymore [27].

2.4.2. Signal jamming

Intentional interference is, in many cases, a significant source of GNSS signal degradation. Intentional interference, known as signal jamming, is caused by the broadcast of malicious radio frequency (RF) signals to prevent GNSS receivers in the area from tracking GNSS signals [11]. The typical direct consequences of jamming are signal frequency shifts in Hertz (Hz) and a drop in signal power in decibels (dB). These effects, in turn, have the potential to cause severe errors in position, velocity, and time calculations and even completely freeze the receiver causing a denial of service condition. Attacking a GNSS signal through jamming requires neither sophisticated knowledge nor complex equipment: all that is needed is a signal of a higher power in the same frequency to defeat the target signal [28]. **Figure 5** shows the visibility of several satellites in an open sky simulation scenario. **Figure 6**, on the other hand, shows the discontinuity (the highlighted rectangle) in satellite availability for the same scenario but when a jamming signal is inserted. The jamming signal lasted for about 1 minute with a power of around −70 dBm and a bandwidth of 10 MHz around the central GPS L1 signal frequency. A slightly higher power jamming signal can completely block signals from all satellites in view.

One option to fight this problem is to use the military (M-Code) receivers or multi-constellation receivers. Another option is to completely switch to any other available navigation solutions [29]. Among these is the long-range navigation system (Loran-C) which is not active now, but there is a noteworthy argument by the US Department of Defense to reactivate it for its significance as an alternative for GPS-based navigation. Furthermore, a modernized version of the system, known as enhanced Loran (e-Loran-C), has been already established and tested. What is special about this system is that its signal power is about a thousand times greater than the GPS signal power. Moreover, it uses an entirely different frequency range from GPS. This makes it safe from the intentional GPS jamming signals. Another alternative is the satellite-based augmentation systems (SBAS) and ground-based augmentation systems
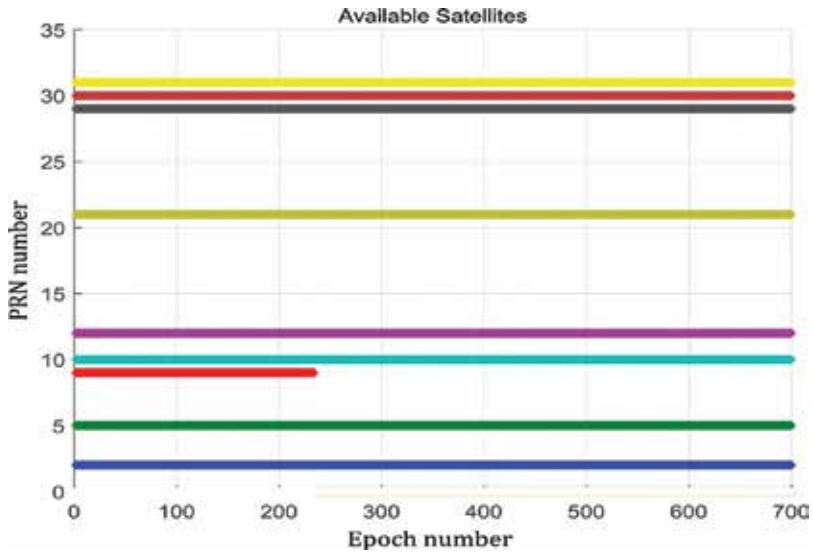
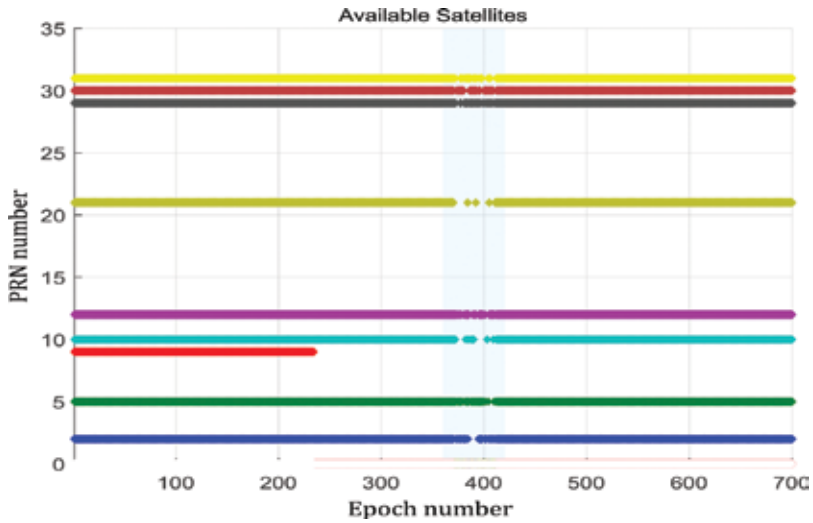**Figure 5.** Satellite availability in a clean scenario.

Figure 6. Discontinuity in satellite availability during the presence of a jamming signal.

(GBAS), which are approved by the US Federal Aviation System (FAA). These systems, though local, efficiently help in mitigating GPS signal outages. Moreover, they are robust against GPS signal jamming. Jamming-free navigation systems, e.g., Inertial Navigation Systems (INS), are an excellent alternative to rely on under jamming conditions.

2.4.3. Signal spoofing

GNSS signal spoofing is the creation of a faked GNSS signal that looks authentic to the GNSS receiver. Signal spoofing is more harmful than jamming because it is not readily detected.

The receiver can be fooled by the spoofing signal, which in turn affects its navigation solution. Furthermore, using correlation techniques to detect the spoofing is not feasible because the received signal is statistically correlated with the authentic GNSS signal, unlike the signal jamming case [30]. The effect of signal spoofing in degrading the navigation solution can have serious impacts in both military and civilian applications, especially those related to safety-of-life services.

Research is ongoing to find reliable techniques for mitigating the effects of spoofing attacks [31]. One example of such techniques is based on the signal direction of arrival (DOA). If the GNSS receiver and its antenna can detect the signal DOA, this can be used to reject the spoofing signal. This depends on the fact that, in most cases, the fake signal will be coming from a ground transmitter and therefore has a low elevation angle. On the other hand, the elevation angle of authentic signals can be predicted from the broadcast ephemeris [31].

2.5. User equivalent range error

After applying the appropriate models and the data in the navigation message to mitigate for the errors, one can use the so-called user equivalent range error (UERE) to quantify the total effect of the remaining errors on pseudorange measurements [2]. The metric, defined as the root sum square of the "unintentional" errors discussed above, is used to analyze the accuracy of the GNSS positioning solution under two assumptions. First, the measurement errors for all the satellites are uncorrelated; second, the independent errors are affecting the pseudorange measurement equivalently [4]. It is worth mentioning that the UERE is typically combined with the dilution of precision (DOP) to meaningfully express the expected accuracy of the GNSS positioning solution. The DOP measure is discussed in the next section.

2.6. Dilution of precision

One parameter that is independent of the cleanliness of measurements but plays a role in the accuracy of position accuracy is the DOP. This factor depends on the geometry of visible satellites; the better the geometry is, the lower the DOP, and, hence, the better the position solution. **Figure 7** visually depicts the concept of DOP. **Figure 7(a)** shows ideal case where signals from two satellites would form circles that intersect at the receiver position assuming the receiver has perfect measurements for the signal which is never true due to GNSS errors. **Figure 7(b)** represents a practical scenario in which uncertainty in measurement makes the virtual circuits radii a little ambiguous. The intersection region characterizes the area of possible receiver positions.
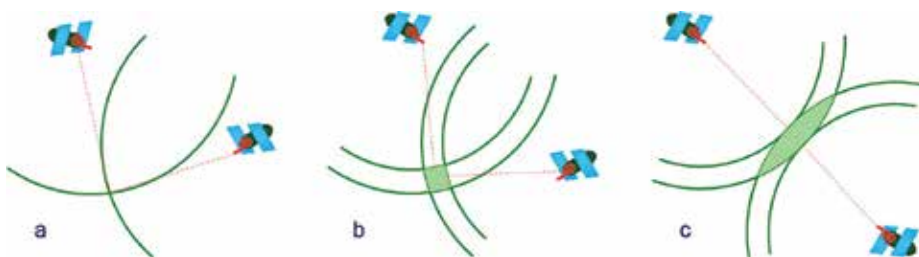

Figure 7. Dilution of precision with range measurements in 2D. Courtesy of Noureldin et al. [2].

This region could have a totally different shape as in **Figure 7(c)**. This solely depends on the geometry of seen satellites. DOP is used to select which satellites should be included in position calculations. An ideal receiver would select only the set of satellites with the minimum DOP [32]. The DOP number is unit-less, and calculating it requires knowing only the receiver and satellites' positions, i.e., no measurements are needed [4]. Hence, DOP could be computed before the journey to plan for trajectory data collection [2].
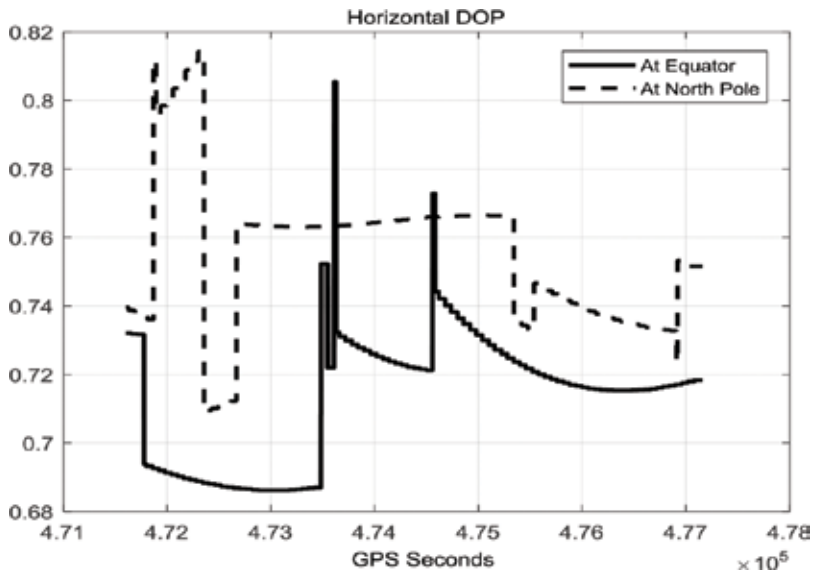

Figure 8. Horizontal DOP values at low versus high latitudes.
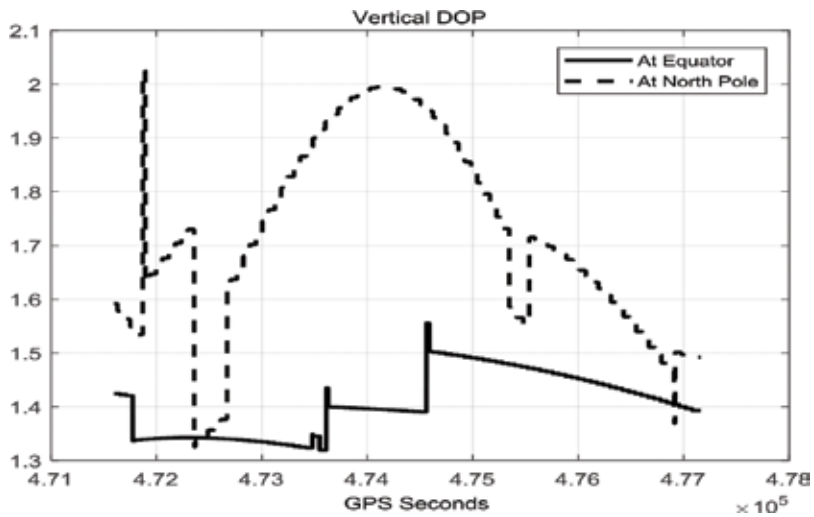

Figure 9. Vertical DOP values at low versus high latitudes.

DOP or the geometric DOP (GDOP) is the general term to describe the geometry of satellites; however, there are subcategories of this. Horizontal dilution of precision (HDOP), vertical dilution of precision (VDOP), and position dilution of precision (PDOP) are examples, to name a few. Simulation DOP values, using GPS only constellation, are shown in **Figures 8** and **9**, respectively, for low latitude (Equator) versus high latitude (North Pole) areas. It can be seen from both figures that DOP values at the Equator are always lower due to better satellite geometry. It can also be noted that GPS provides better HDOP against VDOP due to the arrangement of satellites and their orbits. The accuracy of the obtained/expected GPS solution is expressed as the product of the pseudorange error factor (i.e., UERE) and the geometry factor (i.e., DOP) [11]:

$$\textit{Error in GPS solution } = \textit{ pseudorange error factor} \times \textit{geometry factor} = \textit{UERE} \times \textit{DOP} \qquad (9)$$

As an example on this, a UERE value of 9 m and an HDOP value of 1.4 will indicate a horizontal position accuracy of 12.6 m at the two-sigma level.

3. Conclusion

GNSS signals have low power levels, and hence they are prone to many errors. These errors have various causes, scales, and, hence, consequences. This chapter discusses and classifies GNSS error sources according to their nature and effects. Errors related to the receiver and satellite clocks form one category—clock errors. Signal propagation errors explore a wide range of factors impacting the signal throughout its journey between the satellite and the receiver. Intentional error sources are grouped together. Whenever possible, diagrams and figures are used to explain the error type and/or size of the effect. Common error measure terms, including the user equivalent range error (UERE) and the dilution of precision (DOP), are also presented. Some of the GNSS errors could be as small as a fraction of a signal cycle, e.g., receiver noise error, whereas other errors can be in the order of dozens of meters, e.g., ionosphere and multipath. Receiver clock bias can grow up to thousands of meters and, thus, needs to be modeled. Intentional error sources can completely deny the GNSS services. Regardless their scale, GNSS errors need to be mitigated to achieve accepted navigation accuracy. In addition to exploring each error type, this chapter mentions the best ways to address them.

Author details

Malek Karaim[1]*, Mohamed Elsheikh[1] and Aboelmagd Noureldin[1,2]

*Address all correspondence to: malek.karaim@queensu.ca

1 Electrical and Computer Engineering Department, Queen's University, Canada

2 Electrical and Computer Engineering Department, Royal Military College of Canada, Canada

References

[1] Braasch M. Multipath. In: Teunissen P, Montenbruck O, editors. Springer Handbook of Global Navigation Satellite Systems. Cham: Springer; 2017. pp. 443-468

[2] Aboelmagd N, Karmat T, Georgy J. Fundamentals of Inertial Navigation, Satellite-Based Positioning and their Integration. 1st ed. New York: Springer-Verlag Berlin Heidelber; 2013

[3] Borre K, Akos DM, Bertelsen N, Rinder P, Jensen SH. A Software-Defined GPS and Galileo Receiver a Single-Frequency Approach. Boston: Birkha¨user Boston; 2007

[4] Misra P, Enge P. Global Positioning System: Signals, Measurements and Performance. Massachusetts: Ganga-Jamuna; 2010

[5] El-Rabbany A. Introduction to GPS: The Global Positioning System. Norwood: Artech House Inc; 2002

[6] Hauschild A. Basic observation equations. In: Teunissen P, Montenbruck O, editors. Springer Handbook of Global Navigation Satellite Systems. Cham: Springer; 2017. pp. 561-582

[7] IS-GPS-200H. Interface specification. Navstar GPS Space Segment/Navigation User Interfaces. 2013 Sept:24

[8] Farrell J. Aided Navigation GPS with High Rate Sensors. New York: McGraw-Hill; 2008

[9] Kim H, Lee H. Compensation of time alignment error in heterogeneous GPS receivers. In: Proceedings of the 13th IAIN World Congress. 2009. pp. 27-30

[10] Jin S. Global Navigation Satellite Systems Signal. InTech: Theory and Applications Croatia; 2012

[11] Kaplan E, Hegarty C. Understanding GPS Principles and Applications. Norwood, MA, USA: ARTECH HOUSE, Inc; 2006

[12] Ashby N, Weiss M. Global Positioning System Receivers and Relativity. Boulder, CO, USA: National Institute of Standards and Technology (NIST); 1999;**1385**:1-46

[13] Klobuchar J. Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Transactions on aerospace and electronic systems. 1987;(3):325-331

[14] Juan J, Hernández-Pajares M, Sanz J, Ramos-Bosch P, Aragón-Àngel A, Orús R, et al. Enhanced precise point positioning for GNSS users. IEEE transactions on geoscience and remote sensing. 2012;**50**(10):4213-4222

[15] Nava B, Coisson P, Radicella S. A new version of the NeQuick ionosphere electron density model. Journal of Atmospheric and Solar-Terrestrial Physics;**70**(15):1856-1862

[16] Hobiger T, Norbert J. Atmospheric signal propagation. In: Teunissen P, Montenbruck O, editors. Springer Handbook of Global Navigation Satellite Systems. Cham: Springer; 2017. pp. 165-193

[17] Schuler T. On Ground-Based GPS Tropospheric Delay Estimation: Univ. der Bundeswehr Munchen; 2001.

[18] Karaim M. The Long Journey of a GPS Signal. Inside Signal Processing Newsletter. 2017 October.

[19] Meguro J, Murata T, Takiguchi J, Amano Y, Hashizume T. GPS multipath mitigation for urban area using omnidirectional infrared camera. IEEE transactions on intelligent transpoertation systems. 2009;**10**(1):22-30

[20] Hofmann-Wellenhof B, Lichtenegger H, Herbert W, Wasle E. GNSS—Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More. Springer Science & Business Media; 2007

[21] Kunysz W. High performance GPS pinwheel antenna. In: Proceedings of the ION GPS. 2000. pp. 19-22

[22] Daneshmand S, Broumandan A, Sokhandan N, Lachapelle G. GNSS multipath mitigation with a moving antenna array. IEEE Transactions on Aerospace and Electronic Systems. 2013;**49**(1):693-698

[23] Xu G, Xu Y. GPS: Theory, Algorithms and Applications. Springer; 2016

[24] IGS products. [Online]. [cited 2017 December 3. Available from: http://www.igs.org/products

[25] Tsui J. Fundamentals of Global Positioning System Receivers: A Software Approach. New York: John Wiley & Sons, Inc; 2000

[26] Wu H. On-The-Fly GPS Ambiguity Resolution with Inertial Aiding. Calgary, Canada: Department of Geomatics Engineering, University of Calgary; 2003. [M.S. Thesis]

[27] Selective Availability. [Online]. [cited 2017 December 10. Available from: https://www.gps.gov/systems/gps/modernization/sa/

[28] Curran J, Bavard M, Closas P, Navarro M. A look at the threat of systematic jamming of GNSS. InsideGNSS. 2017;**12**(5):46-53

[29] National PNT Advisory Board comments on Jamming the Global Positioning System. A National Security Threat: Recent Events and Potential Cures; 2010

[30] Humphreys T. Interference. In: Teunissen P, Montenbruck O, editors. Springer Handbook of Global Navigation Satellite Systems. Cham: Springer; 2017. pp. 469-503

[31] Ioannides R, Pany T, Gibbons G. Known vulnerabilities of global navigation satellite systems, status, and potential mitigation technique. Proceedings of the IEEE. 2016 June;**104**(6):1174-1194

[32] Jeffrey C. An Introduction to GNSS. GPS, GLONASS, Galileo and Other Global Navigation Satellite Systems. Calgary: NovAtel Inc.; 2010

GPS in Remote Sensing and Geographical Information System

GNSS Application in Retrieving Sea Wind Speed

Dongkai Yang and Feng Wang

Additional information is available at the end of the chapter

Abstract

In traditional Global Navigation Satellite System (GNSS) application, the reflected GNSS signals from Earth's surface generally are considered as an interference source to be suppressed or removed. Recently, a new idea which treats the reflected GNSS signal as opportunity source of remote sensing has been proposed to monitor Earth's physical parameters. This technique is called as GNSS-Reflectometry (GNSS-R) which has the advantages of low-power, -mass and -cost. With the development and modernization of GPS, Galileo, GLONASS, and BeiDou system, spaceborne GNSS could significantly improve the temporal-spatial resolution by receiving and processing the reflected signal from multiple satellites. This chapter mainly describes this new bi-static remote sensing technique. First, basic theories of GNSS-R including spatial geometry, polarization, and scattering model of reflected signal are discussed; second, spaceborne receivers and fast-response processing methods are reviewed and analyzed; finally, the empirical models retrieving wind speed are proposed and demonstrated using the DDM data from the UK-TechDomeSat-1 satellite. Based on the discussion of this chapter, it could be concluded that although GNSS-R still has some key challenges which have to be addressed, it could be an optimal choice of remote sensing in some special conditions, such as the tropical cyclone.

Keywords: GNSS-Reflectometry, remote sensing, delay-Doppler maps, wind speed

1. Introduction

Sea wind speed is an important parameter which impacts the ocean circulation and global climate. Moreover, tropical cyclones, as one of the most serious natural disasters, interrupt infrastructure and endanger life safety [1]. For these reasons, it is important to monitor sea wind speed to study and forecast some complex weather. Traditional tools, such as buoys,

could provide long-term observation, however, have low spatial sampling. The spaceborne microwave remote sensing devices such as radar altimeter [2], scatterometer [3], and radiometer [4], provide all weather and all day observation of global wind speed. However, high-mass, -power, and -cost equipment limit their application in commercial remote sensing based on the single-function microsatellite. GNSS not only provides the positioning, velocity, and timing (PVT), but also could be considered as the opportunity sources of remote sensing to observe Earth physical parameters. This observation is known as GNSS-Reflectometry which uses GNSS signal reflected off Earth's surface to observe Earth's physical parameters. It has been demonstrated that GNSS-R could have been utilized to measure sea wind speed [5], sea surface height [6], sea ice [7], and soil moisture [8] from ground-based, airborne, and spaceborne platforms in past 20 years. Compared to traditional remote sensing ways, GNSS-R has the advantages of the low-cost and low-power, because only a receiver is needed. Spaceborne GNSS-R could improve the global temporal-spatial resolution because there are more than 100 GNSS satellites in orbit or planned. In addition, the temporal-spatial resolution could be further improved by microsatellite network.

The idea of using reflected GNSS signal from Earth's surface for remote sensing was discussed by Hall and Cordey in 1988 [9]. In 1993, Martin-Neria proposed the concept of PAssive Reflectometry and Interferometry System (PARIS) to use GNSS signal reflected off the ocean for the measurement of ocean altimetry [10]. In 1994, Auber et al. detected GNSS signal reflected off the ocean from an aircraft platform [11]. In 1998, Garrison and Katzberg demonstrated that reflected global positioning system (GPS) signal could measure sea wind speed through an aircraft experiment [12]. The first collection of reflected GPS signal from spaceborne platform was reported in 2002 [13]. The use of reflected GPS signal acquired by UK-DMC satellite is recognized as the first spaceborne experiment to measure wind speed [14]. After the success of the UK-DMC, UK TechDemoSat-1 [15], PARIS-IoD [16], GEROS-ISS [17], and CYGNSS [18] were planned and developed to further demonstrate the feasibility of using reflected GNSS signal to observe Earth parameters. Especially, CYGNSS mission aims to improve intensity forecasting by sensing sea wind speed in the inner core of tropical cyclones using a constellation of eight microsatellites.

The basic product of GNSS-R is the so-called Delay Waveform or delay-Doppler map (DDM) of reflected GNSS signal, which describes the distribution of power level in delay or delay-Doppler domain. Based on the bi-static radar equation and the geometric optics limit of the Kirchhoff approximation, Zavorotny and Voronovich developed the theoretical power model of the scattered GNSS signal as a function of the geometrical and environmental parameters [19]. To simulate DDM, the Z-V model can be implemented by defining the scenario in a reference system and evaluating the functions inside the integrand to compute the integral for every delay and Doppler bin. However, such approach is extremely time- and resource-consuming, especially for spaceborne scenario. In [20], a new and efficient method to compute DDM was proposed based on the derivation of explicit expressions of the space coordinates as a function of the delay offset and Doppler shift. The realistic problem that the noisy DDM measurements should be generated motivated the model and the simulator to produce DDM time series based on a multiple-input-multiple-output filter bank [21]. Other approaches of simulating DDM generate a synthetic sea surface with accurate temporal-spatial correlations and model the time evolution of the scattered signal. In [22], Clarizia et al. presented a new

facet-based approach to model the scattering of GNSS signal from the synthetic sea surface and used it to compute DDM for the spaceborne scenario. In practice, the DDM is generated by correlating the reflected GNSS signal with the locally generated replicas at different delay offset and Doppler shift. In [23], a delay/Doppler-mapping hardware/software receiver was presented. A real-time instrument called as GPS open-loop differential real-time receiver (GOLD-RTR) was designed and developed to gather the complex correlations of reflected GPS signal in [24]. To conform the development of spaceborne GNSS-R mission, spaceborne GNSS-R receivers have been designed and developed, such as Space GNSS Receiver Remote Sensing Instrument (SGR-ReSI) [25] and GNSS-REflectometry, Radio Occultation, and Scatterometry (GEROS) [26] for UK-TDS-1 satellite and International Space Station.

The purpose of remote sensing is to determine the Earth's physical parameters, for which it is the key technology to develop the link between GNSS-R observable and the Earth's physical parameters. Some algorithms have been proposed to retrieve sea wind speed using GNSS-R. The first type of method is to fit the measured Delay Waveform or DDM to theoretical one. In [27], the trailing edge slope and the complete shape of the Delay Waveform were used to fit the analytical models to estimate wind speed. A 2-D least-squares fitting approach was given to retrieve wind field by fitting the simulated DDM to space-based measured one from UK-DMC satellite [28]. These methods based on fitting are time-and resource-consuming. The second type of method is to directly link wind speed and different Delay Waveform or DDM observable by the regression. In 2013, Rodrigues-Alvarez et al. reported their airborne experiment results in which four DDM observable including weighted area, 2-norm Euclidean distance, distance from the geometric center to the maximum value of DDM, and 1-norm Euclidean distance were defined and linked with the wind speed [29]. First results of the retrieved wind speed using the data from UK TechDemoSat-1 satellite were presented in [30], where in-situ wind from Advanced Scatterometer (ASCAT) and the bi-static scattering coefficient derived from peak DDM using bi-static radar equation are utilized to develop a wind speed retrieval algorithm and the accuracy of 2.2 m/s could be obtained in range from 3 to 18 m/s. To improve retrieval accuracy of high wind speed and spatial resolution for CYGNSS mission, Rodriguez-Alvarez defined the generalized linear DDM observable based on maximum signal-to-noise ratio (SNR), minimum variance of wind speed, and principal component analysis (PCA) to develop relationships with wind speed and found that the observable based on PCA had the best performance [31]. The retrieval algorithm proposed by M. P. Clarizia in 2014 was used to produce the Level 2 wind speed data product of CYGNSS mission in [32], in which several additional processing steps were added to correct the influence of incident angle and adaptively select the parameters of the algorithm to improve the retrieving accuracy [32]. The algorithms for CYGNSS mission have limited the spatial resolution in the region of 25 km × 25 km. These retrieval algorithms above all provide a single wind speed descriptor in the overall observation region. The third type of method is based on the mapping between the physical space and the delay-Doppler domain. In [33], an algorithm to retrieve the bi-static scattering coefficient distribution over the observation region from measured DDM using the deconvolution was proposed. D. Schiavulli demonstrated the validity and feasibility of this approach using UK-TechDemoSat-1 data [34].

This chapter will illustrate the GNSS application in retrieving wind speed using spaceborne GNSS-R. Section 2 will discuss the basic theories of GNSS-R including the observation

geometry, the polarization, and the model of reflected GNSS signal. Section 3 will present some typical spaceborne GNSS-R receivers and processing methods of reflected GNSS signal. The explicit retrieval algorithms based on the regression will be developed in Section 4. Finally, the conclusion of this chapter will be addressed in Section 5.

2. Basic theory of GNSS-R

2.1. Geometry

GNSS-R works as a bi-static radar in which the transmitter and the receiver are separated in the spatial distribution. It is noted that when GNSS-R receivers allow to track the scattered signal from different GNSS satellites, the system is called as multi-static radar as shown in **Figure 1**, in which for each GNSS satellite, bi-static observation occurs at different azimuth angle, and creates a glistening zone. To simplify observation geometry of GNSS-R, a bi-static geometry could be developed as **Figure 2**, in which (1) the coordinate origin is at Earth's core; (2) the YOZ plane is in the incident plane of GNSS signal; (3) Z axis has the same direction with the normal of the tangent plane of the specular point; and (4) it is assumed that Earth, the orbits of GNSS and Low Earth Orbit (LEO) satellite are circular. In **Figure 2**, h_r and h_t are the heights of LEO and GNSS satellite; R_e is the Earth radius; θ is the incident angle of GNSS signal. In the coordinate system, the positioning of GNSS and LEO satellite are given as

$$\mathbf{T} = (0 \quad D\sin\theta \quad R_e + D\cos\theta) \tag{1}$$

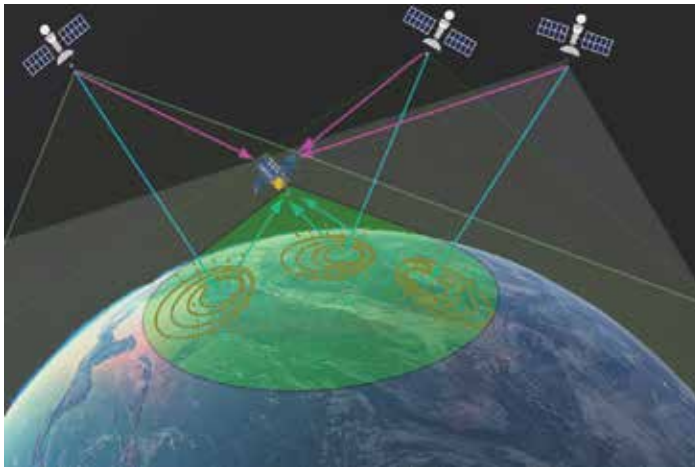$$\mathbf{R} = (0 \quad -d\sin\theta \quad R_e + D\cos\theta) \tag{2}$$

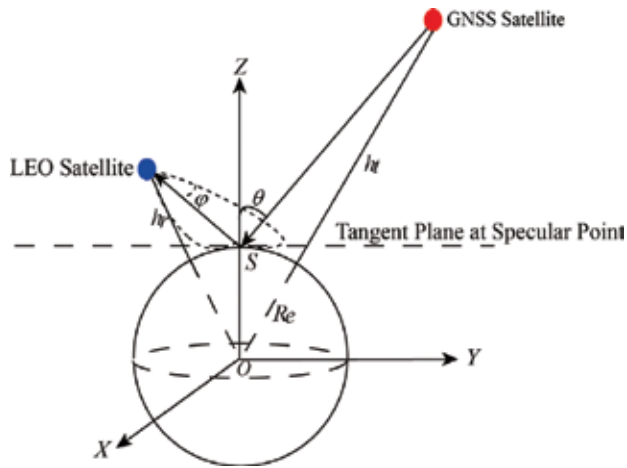**Figure 1.** Illustration of multi-static GNSS-R.

Figure 2. Scattering scenario and local coordinate system of GNSS-R.

where D and d are the distance from GNSS and LEO satellite to the specular point, and are expressed as

$$D = -R_e\cos\theta + \sqrt{(h_t + R_e)^2 - R_e^2\sin^2\theta} \tag{3}$$

$$d = -R_e\cos\theta + \sqrt{(h_r + R_e)^2 - R_e^2\sin^2\theta} \tag{4}$$

The coordinate of scattering unit $\boldsymbol{\rho}$ could be represented as

$$\boldsymbol{\rho} = \begin{pmatrix} \delta_x & \delta_y & 1 \end{pmatrix} \cdot R_e \tag{5}$$

where δ_x and δ_y are the included angles between the projection of the connection line of $S_{x,y}$ and the coordinate origin at the XOZ and YOZ plane. From the above discussion, when the height of GNSS and LEO satellite, and the incident angle of GNSS signal are known, the geometry of GNSS-R could be determined.

2.2. Polarization of reflected GNSS signal

When Right-Hand Circular Polarization (RHCP) GNSS signal occurs reflection phenomenon, the polarization is changed. The Fresnel coefficients of RHCP and LHCP (Left-Hand Circular Polarization) are expressed as the combination of the vertical and horizontal polarization as

$$\mathfrak{R}_{RR} = \mathfrak{R}_{RR} = \frac{1}{2}(\mathfrak{R}_{VV} + \mathfrak{R}_{HH}) \tag{6}$$

$$\mathfrak{R}_{RL} = \mathfrak{R}_{LR} = \frac{1}{2}(\mathfrak{R}_{VV} - \mathfrak{R}_{HH}) \tag{7}$$

The Fresnel coefficients of vertical and horizontal polarization are computed as

$$\Re_{VV} = \frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} \tag{8}$$

$$\Re_{HH} = \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} \tag{9}$$

Figure 3 gives the simulated Fresnel coefficients of RHCP and LHCP for GPS L1 signal, when sea surface temperature is 25° and sea surface salinity is 35 psu. From the figure, it could be seen that as the incident angle increases, the Fresnel coefficients of RHCP and LHCP show decreasing and increasing trends, respectively. This illustrates that at nadir and small incident angle, the LHCP signal dominates GNSS signal reflected off the sea surface. Therefore, the combination of LHCP and RHCP antenna or only LHCP antenna is used to receive reflected GNSS signal from the sea surface.

2.3. Model of reflected GNSS signal

The signal reflected from the sea surface is described as [21]

$$u_r(\mathbf{R}, t) = A_R \exp(-2\pi j f_0 t) \iint G(\boldsymbol{\rho}) a\left(t - \frac{D+d}{c}\right) g(\mathbf{R}, t) d^2\boldsymbol{\rho} \tag{10}$$

where f_0 is the carrier frequency; A_R is the amplitude of reflected signal; G is the pattern of receiving antenna; $a(t)$ is the baseband spreading signal; c is the speed of light; $g(\mathbf{R}, t)$ is the function as
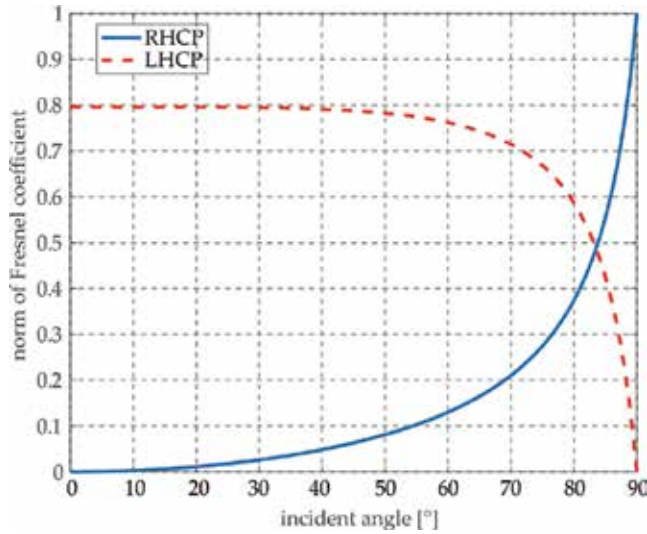

Figure 3. Relationship between the amplitude of RHCP and LHCP Fresnel coefficient and incident angle.

$$g(\mathbf{R},t) = -\frac{\mathfrak{R}}{4\pi j D(\boldsymbol{\rho})d(\boldsymbol{\rho})}\exp[jk(D(\boldsymbol{\rho})+d(\boldsymbol{\rho}))]\frac{\mathbf{q}^2(\boldsymbol{\rho})}{\mathbf{q}_\perp^2(\boldsymbol{\rho})} \tag{11}$$

where $\mathfrak{R}$ is the Fresnel coefficient discussed in Section 2.2; $D(\boldsymbol{\rho})$ and $d(\boldsymbol{\rho})$ are the distance from GNSS and LEO satellite to scattering unit $\boldsymbol{\rho}$; $\mathbf{q}$ is the scattering vector defined as

$$\mathbf{q}=\begin{pmatrix}\mathbf{q}_\perp & q_z\end{pmatrix}=k\left(\frac{\mathbf{d}(\boldsymbol{\rho})}{d(\boldsymbol{\rho})}+\frac{\mathbf{D}(\boldsymbol{\rho})}{D(\boldsymbol{\rho})}\right) \tag{12}$$

where k is the wavenumber of GNSS signal; $\mathbf{D}(\boldsymbol{\rho})$ and $\mathbf{d}(\boldsymbol{\rho})$ are the vectors from scattering unit $\boldsymbol{\rho}$ to GNSS and LEO satellite. The receiver processes reflected GNSS signal by cross-correlating with local replicas over a range of delay and Doppler as

$$Y(\tau,f,t) = \int_t^{t+T_I} u_r(\mathbf{R},t)p(t-\tau)\exp\left[2\pi j(f_0+f)t\right]dt \tag{13}$$

where T_I is the coherent integration time. The output of spaceborne GNSS-R receiver is the power level as a function of delay offset and Doppler shift. To reduce the influence of thermal and speckle noise, the successive snapshots are averaged as

$$\left\langle|Y(\tau,f)|^2\right\rangle = \frac{1}{\mathrm{T_{incoh}}}\int_0^{\mathrm{T_{incoh}}}|Y(\tau,f,t)|^2 dt \tag{14}$$

where $\mathrm{T_{incoh}}$ is the incohernt integration time. Zavorotny and Voronovivh first derived the expression (14) as [19]

$$\left\langle|Y(\tau,f)|^2\right\rangle = A_R^2 T_I^2 \iint \frac{G^2(\boldsymbol{\rho})\Lambda^2(\tau-\tau(\boldsymbol{\rho}))\cdot\mathrm{sinc}^2(f-f(\boldsymbol{\rho}))}{4\pi D(\boldsymbol{\rho})^2 d(\boldsymbol{\rho})^2}\sigma_0^2 d^2\boldsymbol{\rho} \tag{15}$$

where $\tau(\boldsymbol{\rho})$ and $f(\boldsymbol{\rho})$ are the delay and Doppler frequency of scattering unit $\boldsymbol{\rho}$; σ_0 is the normalized bi-static cross section and it is as follows:

$$\sigma_0 = \pi|\mathfrak{R}|^2\frac{q_z^4}{\mathbf{q}_\perp^4}P_{\mathrm{pdf}}\left(-\frac{\mathbf{q}_\perp}{q_z}\right) \tag{16}$$

where $P_{\mathrm{pdf}}(\cdot)$ is the probability density function of mean square slopes of the sea surface, and could be usually assumed as a 2-D Gaussian distribution. **Figure 4** shows the normalized DDM distribution for the wind speed of 5 m/s and 15 m/s, in which both DDMs show clear horseshoe shapes; moreover, with the increase in wind speed, DDM spreads to larger delay and Doppler. Different scattering units have different delay and Doppler frequency; hence, DDM not only represents the distribution in delay-Doppler domain, but also indicates the distribution in sea surface as illustrated in **Figure 5**. In practice, the one cell in delay-Doppler domain associates to two spatial solutions, i.e., the mapping between the space and the delay-Doppler domain exists with ambiguity. To reduce this mapping ambiguity, the one approach is to tilt the downward antenna beam away from the specular reflection point.
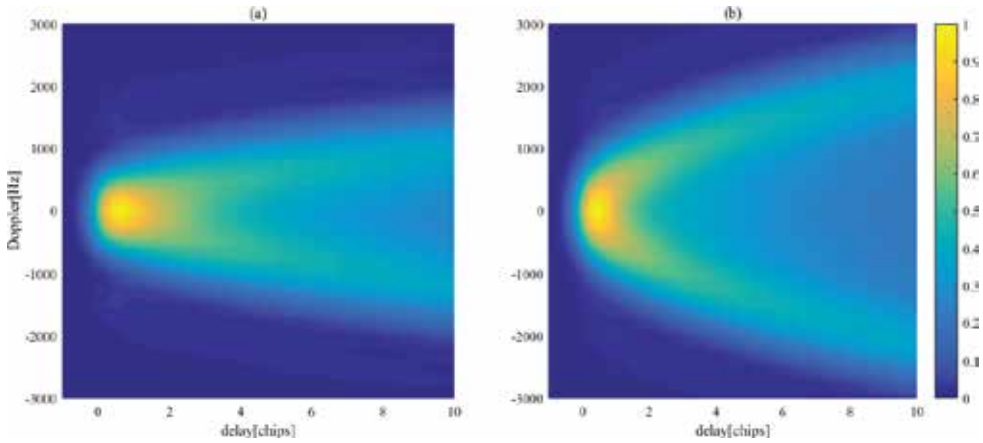
Figure 4. Normalized DDM for the wind speed of (a) 5 m/s and (b) 15 m/s, when the height of LEO satellite is 657 km and the incident angle of GNSS signal is 60°.
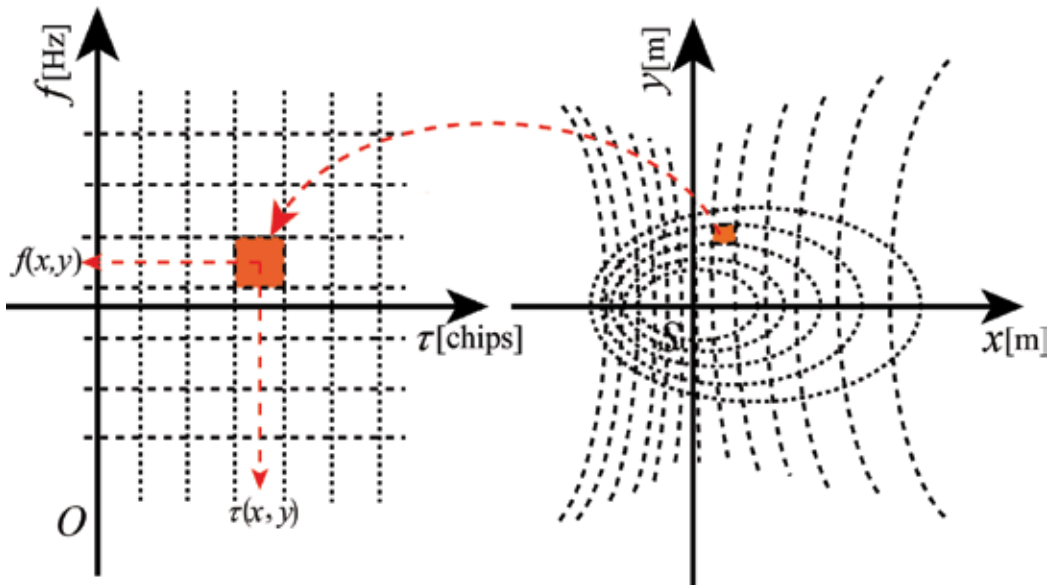

Figure 5. Mapping from space to the delay-Doppler domain.

The other approach is using two separate antennas from two different viewing angles to observe an ocean surface region.

3. Processing of GNSS reflected signal

The GNSS-R receiver is a processing unit to cross-correlate reflected GNSS signal with locally generated replicas. Spaceborne GNSS-R receiver not only requires the capacity of

processing reflected GNSS signal in real-time, but also with low-mass, -power and -cost. At present, some research groups have developed spaceborne GNSS receivers to meet their spaceborne missions.

3.1. Existing receivers

3.1.1. UK-DMC

To promote the development of spaceborne GNSS-R and explore its prospect of commercial application, Surrey Satellite Technology Ltd. (SSTL) designed and developed spaceborne GPS-R receiver which would be carried on UK-DMC. This receiver which was implemented based on spaceborne GPS receiver of SSTL consisted of a zenith antenna to receive direct GPS signal, a LCHP antenna with the gain of 11.8 dBi to downward point to receive reflected GPS signal, a real-time processor unit to perform delay-Doppler mapping of reflected GPS signal as well as log and download raw sampled IF data to the ground for more intensive postprocessing together with a solid state data recorder (SSDR) and high rate downlink (HRD) [14]. Moreover, the processor unit supported to track the direct signal provides the positioning, velocity and timing (PVT). When the receiver was configured as the collection mode, 20 s raw sampled data of the direct and reflected GPS signal could be saved in SSDR to download to the ground.

3.1.2. SGR-ReSI

Following the success of the UK-DMC, SSTL embarked on the new generation spaceborne GNSS-R receiver Space Receiver Remote Sensing Instrument (SGR-ReSI) which had been equipped with UK-TechDemoSat-1 satellite launched in July, 2014 [25]. Compared to the receiver carried on UK-DMC, SGR-ReSI supported to process in real-time and save sampled IF data of multi-frequency GNSS signal. The core component of the receiver is a 24-channel navigation receiver of GPS L1 signal and reprogrammable to a co-processor unit based on Actel ProASIC3 FPGA. The reprogrammable capacity of SGR-ReSI also enables more scientific and remote sensing application. Two types of RF front-ends including Max2769 optimized for GPS L1 signal and Max2112 which is reconfigured to other GNSS bands have been adopted in the receiver. In operation, the receiver could be configured as fixed gain mode (FGM) and unmonitored automatic gain control (uAGC). To allow the storage of both sampled and processed data, a bank of DDR2 memory with a capacity of 1 GB was used. From September 2014 on, the GNSS-R basic observable DDM data formatted as TIFF were downloaded from UK-TechDemoSat-1 to the ground.

3.1.3. GEROS

The main objectives of GEROS mission are to measure the sea surface height and mean square slope, and the secondary ones are to further explore the potential of monitoring vertical profiles of atmospheric using GNSS-RO and to assess the feasibility of GNSS-R for land application [17]. The receiving antenna has 4π steradian field of view capability to receive GNSS signal arriving from any direction and is distinguished into four distinct parts (FoV-1, FoV-2, FoV-3, FoV-4), in which FoV-1 is for altimetry and scatterometry, FoV-2 is for altimetry,

and FoV-3 is for atmosphere/ionosphere sounding. The bandwidth of the GEROS payload is 47 MHz at L1 and 64 MHz at L2 frequency band so that GEROS can perform GNSS-R and GNSS-RO using the signal from different GNSS systems, such as GPS, Galileo, GLONASS, and BeiDou. GEROS can in parallel track three direct and corresponding reflected signals in both RHCP and LHCP at L1 band, only LHCP at L5 band using three antenna beams. The fourth dual-frequency beam could be used to perform either GNSS-R or GNSS-RO. The main Level 1 data of the first three beams are 1 Hz DDM for which the delay and Doppler domain is 1500 m × 8 KHz and the delay resolution could be configured as 3.75 m and 11.25 m. The fourth beam outputs different Level 1 data depending on the application. GEROS payload has a total mass of 376 kg, a power of 395 W, a 2 GB memory, a 1.2 Mbps output data rate, and the overall dimension of 1.55 m × 1.17 m × 0.86 m.

3.1.4. PAU/GNSS-R

Passive Advanced Unit (PAU) is a new instrument concept which aims to monitor ocean parameters. It consists of the PAU/RAD which is an L-band radiometer to measure the brightness temperature of the sea surface, the RAU/GNSS-R which is a reflectometer to measure the roughness of the sea surface and the PAU/IR which is two infrared radiometers used to observe the temperature of the sea surface. PAU/GNSS-R was designed based on FPGA to synchronously process the reflected GPS signal from different satellites in real-time and output corresponding to 2-D DDM [35]. To reduce the hardware resources of the payload, hardware reuse technique was adopted based on two RAM-like registers that change their respective input and output connections to allow the DDM generator to be implemented at a higher clock rate. A software running on MicroBlaze soft processor mainly performs functions including sending and receiving data from the GPS receiver, selecting the available satellites, and computing the DDM parameters transferred to the DDM generator. The computed DDM are transferred to a terminal computer using an USB and an external master based on FPGA.

3.1.5. PYCARO

According to the research and demonstration mission planned by the Remote Sensing Lab and the NanoSat Lab at the Universitat Politècnica de Catalunya-Barcelona Tech, PYCARO will be carried on a 3 × 2 unit Cube satellite to perform the observation of Earth surface and atmosphere. Compared to the receivers above, the difference of PYCARO is a P(Y) and C/A ReflectOmeter, which adds the processing of encrypted L1 and L2 P(Y) signals by using semicodeless technique [36, 37]. The zenith antenna of PYCARO is a single microstrip patch, and the nadir-looking antenna is a 3 × 2 array of microstrip patches with 13 dB gain. A dual-channel Software Defined Radio (SDR) is used to sample and collect the signals from the up- and down-looking antennas, and a Gumstix Overo IronStorm OBC manages the payload, configures the SDR, and computes the DDM. PYCARO supports two reconfigurable operation modes including the nadir-pointing mode to perform the Earth surface observation of GNSS-R and the limb-pointing to implement the atmosphere observation GNSS-RO.

3.2. General architecture of receiver

According to the description on the spaceborne GNSS-R receivers above, the general architecture of the receiver could be given as **Figure 6**, which consists of the antennas, the RF front-end, the processor of baseband signal, the computer on-board, and some external components. The antennas include zenith- and nadir-pointing antennas to receive direct and reflected GNSS signal. Usually, the zenith-pointing antenna is an omnidirectional RHCP GNSS antenna, and nadir-pointing antenna is a narrow-beam and high-gain LHCP antenna. However, it is able to be design the different combination of the zenith- and nadir-pointing antennas to perform special spaceborne mission and application. For example, the zenith- and nadir-pointing antennas of PARIS are both the narrow-beam and high-gain antennas to make the incoming signal from the single satellite be received; to significantly measure the soil moisture of the land, the nadir-pointing antenna of GEROS is the combination of the LHCP and RHCP antenna. RF front-ends convert the analog RF signal to the sampled IF one, consisting of Low Noise Amplifier (LNA), bandpass filter, mixer, gain control, and A/D converter. According to the requirement of the mission, the gain control could have different configuration. For altimetry, the Automatic Gain Control (AGC) mode is utilized to dynamically adjust the power level of incoming signal to the most effective input range of A/D converter; however, the scatterometry of GNSS-R generally selects the fixed gain or monitoring AGC to correctly receiving gain to obtain accuracy power level of reflected signal. To improve the generality of the receiver, the gain control could be designed as reconfigurable mode, such as SGR-ReSI of which the gain control supports the fixed gain and unmonitoring AGC mode. The core of the receiver is a baseband processor which consists of controller and processor unit. The controller unit controls the acquirement
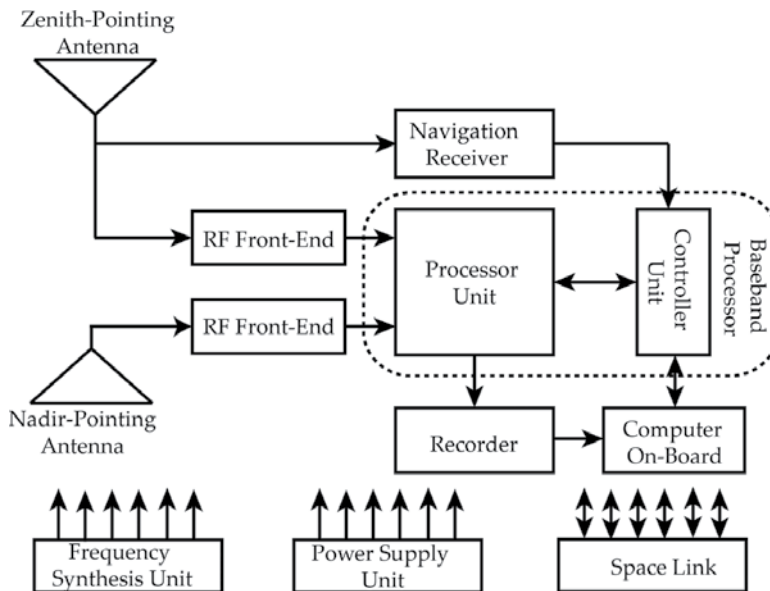

Figure 6. Architecture of Spaceborne GNSS-R receiver.

and tracking of direct signal, estimates the difference of delay and Doppler frequency between direct and reflected signal, and configures the DDM parameters, such as delay and Doppler frequency window range and resolution and others. The estimations of the differences of delay and Doppler frequency between direct and reflected signal are given as

$$\Delta\tau = \frac{|\mathbf{T} - \mathbf{S}| + |\mathbf{R} - \mathbf{S}| - |\mathbf{T} - \mathbf{R}|}{c} \tag{17}$$

$$\Delta f = \frac{\mathbf{v}_r \cdot \mathbf{u}_{rs} - \mathbf{v}_t \cdot \mathbf{u}_{ts} - (\mathbf{v}_t \cdot - \mathbf{v}_r) \cdot \mathbf{u}_{tr}}{\lambda} \tag{18}$$

where $\mathbf{T}$ and $\mathbf{v}_t$ are the positioning and velocity of GNSS satellite; $\mathbf{R}$ and $\mathbf{v}_r$ are the positioning and velocity of LEO satellite; $\mathbf{S}$ is the positioning of the specular point; $\mathbf{u}_{rs}$ and $\mathbf{u}_{ts}$ are the unit vectors from LEO and GNSS satellite to the specular point; $\mathbf{u}_{tr}$ is the unit vector from the LEO to GNSS satellite; c is the speed of light; and λ is the wavelength of GNSS signal. To make the DDM in the window of delay and Doppler frequency, it is necessary to provide high-accuracy positioning and velocity of GNSS, LEO satellite and specular point to meet the estimated accuracy of expression (17) and (18). The processor unit performs the auto/cross-correlation operation, in which the auto-correlation mode correlates the reflected signal with locally generated replicas at the different delay and Doppler frequency, and the cross-correlation mode correlates the reflected signal with the direct signal after proper delay and Doppler adjustment. Compared to the auto-correlation mode, the cross-correlation mode overcomes the bandwidth limitation to improve the range resolution and decreases the computational complexity. However, the main shortcoming of the cross-correlation is that the large-size and narrow-beam antennas are required for both direct and reflected signal to separate signal from different satellites. To overcome the bandwidth limitation and decrease the antenna size, a new approach is to use semicodeless technique to locally reconstruct encrypted P(Y) code and then cross-correlate them with reflected signal. The computer on-board performs the control and management of satellite mission to receive and route ground commands for the configuration of the DDM parameters and the operation modes, assemble and packet useful data for downloading to the ground. External components include power supply unit, frequency synthesis unit and space link to provide the required power supplies, working clocks for the units of the whole system, and the spatial interface linking the satellite and ground station.

3.3. Processing method

At present, the processing methods of reflected GNSS signal have serial and parallel correlation. The serial correlation correlates the reflected GNSS signal with the locally generated replicas at the different delay and Doppler frequency one by one in time domain. The parallel correlation first transforms the reflected and local signal to the frequency domain using fast Fourier transform (FFT) and performs a multiplication operation, then transforms multiple results to time domain using Inverse FFT (IFFT). The serial processing is lossless and can be considered as the

reference for the assessments of parallel method. The assessment metric is defined as normalized mean squared error (MSE) between serial and parallel correlation as

$$\sigma^2 = \frac{\sum\limits_{i=1}^{N_\tau}\sum\limits_{j=1}^{N_f}\left|\left\langle\left|Y_S\left(\tau_i,f_j\right)\right|^2\right\rangle - \left\langle\left|Y_P\left(\tau_i,f_j\right)\right|^2\right\rangle\right|^2}{\sum\limits_{i=1}^{N_\tau}\sum\limits_{j=1}^{N_f}\left|\left\langle\left|Y_S\left(\tau_i,f_j\right)\right|^2\right\rangle\right|^2} \tag{19}$$

where $\left\langle\left|Y_S\left(\tau_i,f_j\right)\right|^2\right\rangle$ and $\left\langle\left|Y_P\left(\tau_i,f_j\right)\right|^2\right\rangle$ are the mean correlation power computed by serial and parallel method when the delay and Doppler frequency of local replica are τ_i and f_j; N_τ and N_f are the number of delay and Doppler bin.

3.3.1. Serial correlation

Serial correlation in time domain is expressed as

$$Y_S\left(\tau_i,f_j\right) = \frac{1}{N_S}\sum_{i=1}^{N_S} u_r(t)p(t-\tau)\exp\left\{-2\pi j\left(f_0+f_j\right)t\right\} \tag{20}$$

where N_S is the sampling number in a GNSS-code period. The block diagram of the serial processing is shown in **Figure 7**, in which the channel is $N_\tau \times N_f$. Each channel requires multiplier $3N_S$ and adder $2N_S$. The total numbers of the multipler and the adder are $3N_S N_\tau N_f$ and $2N_S N_\tau N_f$, respectively.

3.3.2. Parallel correlation in delay domain

Parallel correlation in delay domain parallelly performs correlation operation for all code delay based on Fourier transform in each Doppler channel as

$$Y_P\left(:,f_j\right) = \text{IFFT}\left\{\text{FFT}\left\{u_r(t)\exp\left(-2\pi j\left(f_0+f_j\right)\right)\right\}\{\text{FFT}\{p(t)\}\}^*\right\} \tag{21}$$

where $\text{FFT}\{\cdot\}$ and $\text{IFFT}\{\cdot\}$ respresent the FFT and IFFT operator; and $\{\cdot\}^*$ is the conjugate operator. **Figure 8** is the block diagram of the parallel correlation in delay domain. The number of the processing channel is N_f. As it is known, FFT operation with N_S sampled ponits requires multiplier and adder $2N_S\log_2 N_S$ and $3N_S\log_2 N_S$. From **Figure 8**, it is seen that each channel performs 2 FFT and 1 IFFT with N_S sampled ponits. The total numbers of the required multipler and adder are $6N_f N_S\log_2 N_S$ and $9N_f N_S\log_2 N_S$, respectively.

3.3.3. Parallel correlation in Doppler domain

Parallel correlation in Doppler domain performs spectrum estimation for the signal removed the modulation of the code at each delay channel as

$$Y_P(\tau_i, :) = \text{FFT}\{u_r(t)p(t - \tau)\exp\{-2\pi jf_0 t\}\} \tag{22}$$

The block diagram is illustrated as in **Figure 9**, in which the number of processing channel is N_τ. In each channel, correlation operation is divided into two steps: decimation and spectrum estimation. To prevent the overlapping of the spectrum, it is necessary to add an anti-aliasing filter before resampling. In engineering, one of the most efficient anti-aliasing filters is Cascaded Integrator Comb (CIC) filter which requires adder about $2N_S$. The spectrum of decimated signal is estimated using FFT, which needs multipler and adder
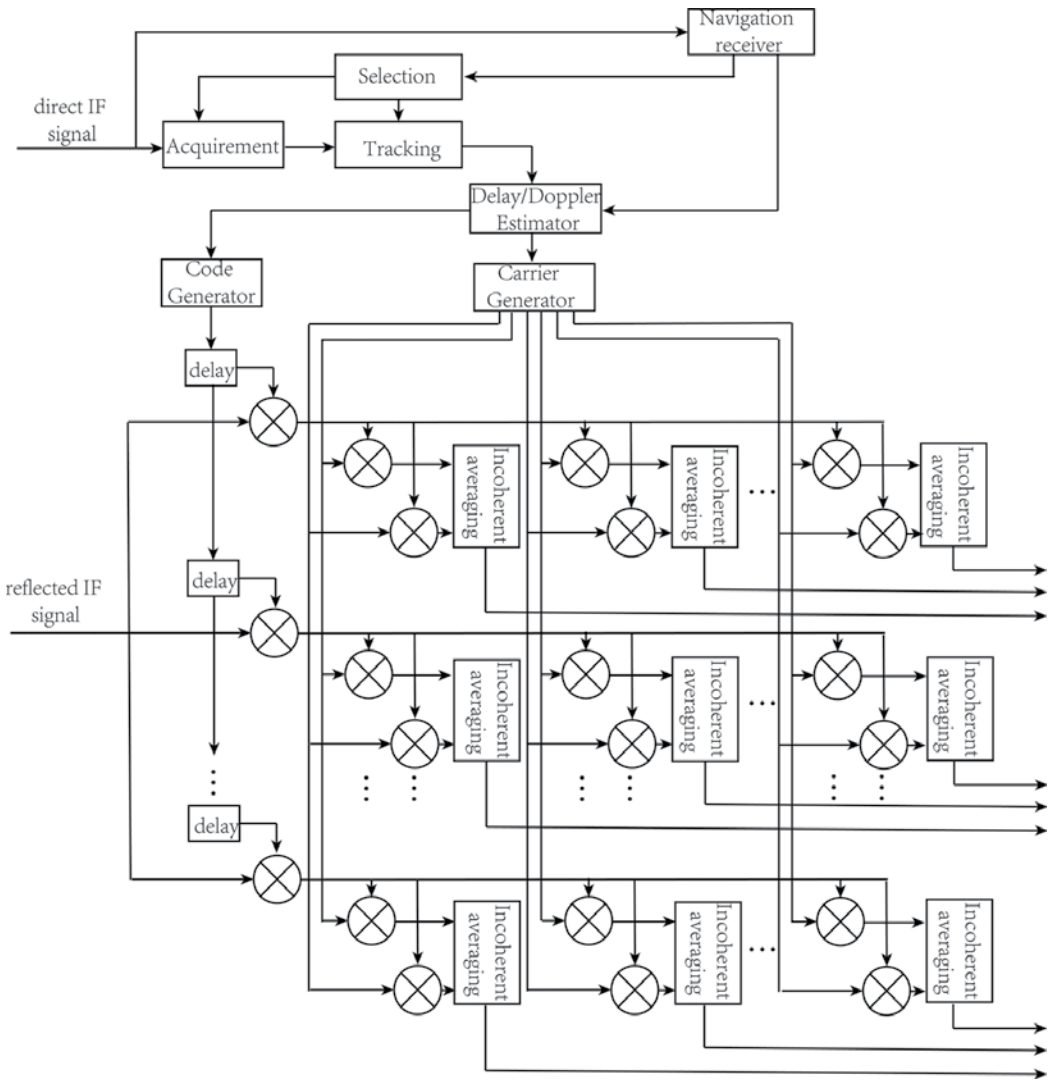

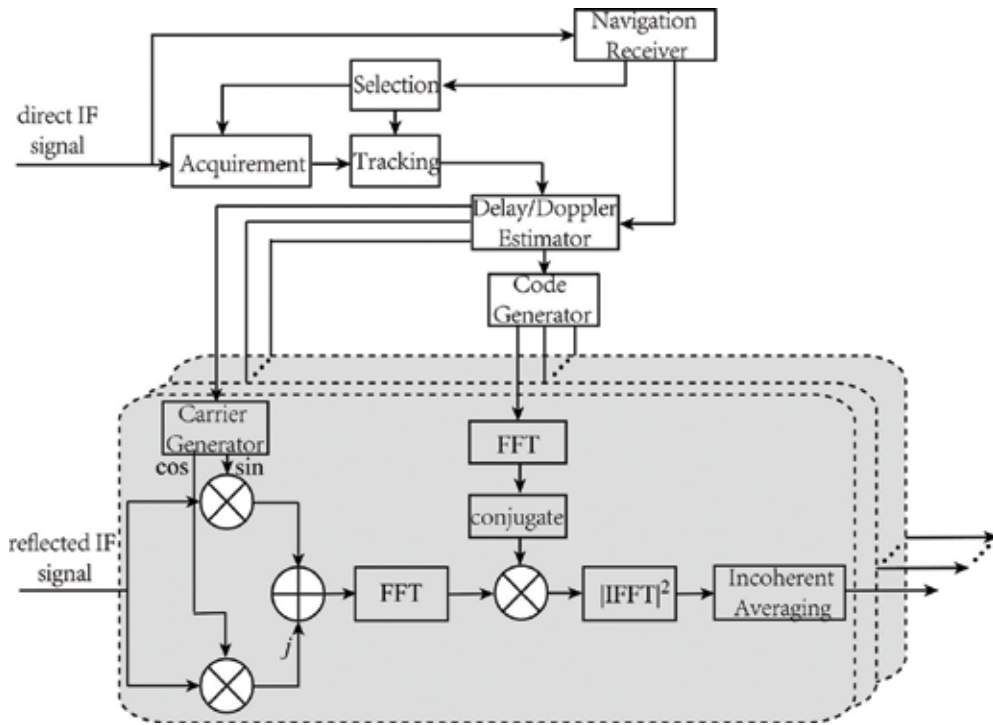**Figure 7.** Block diagram of serial correlation method.

Figure 8. Block diagram of parallel correlation in delay domain.

about $2N_f\log_2 N_f$ and $3N_f\log_2 N_f$. The total numbers of the multipler and adder are $2N_\tau N_f\log_2 N_f$ and $N_\tau(3N_f\log_2 N_f + 2N_S)$.

3.3.4. Parallel correlation in delay-Doppler domain

Parallel correlation in delay-Doppler domain [38], as given in **Figure 10**, is a block processing technique called as double-block zero-padding (DBZP) expressed as

$$Y_P(:,:) = \text{FFT}_c\left\{\begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_{N_b} \end{array}\right\} \tag{23}$$

$$A_i = \text{IFFT}_r\{\text{FFT}_r\{[U_i, U_{i+1}]\}\cdot\{\text{IFFT}_r[P_i, \mathbf{0}]\}^*\} \tag{24}$$

where U_i and P_i are the ith blocks of the reflected signal and local replica; $\mathbf{0}$ is the zero vector; $\text{FFT}_r\{\cdot\}$ and $\text{IFFT}_r\{\cdot\}$ respresent FFT and IFFT operation for the row of the matrix; and $\text{FFT}_c\{\cdot\}$ is FFT operation for the column of the matrix. Expression (23) is the parallel correlation in delay domain for the block, and expression (24) could be considered as the spectrum estimation. To simplify the analysis of the computational complexity, it is assumed that $N_b = N_f/2$
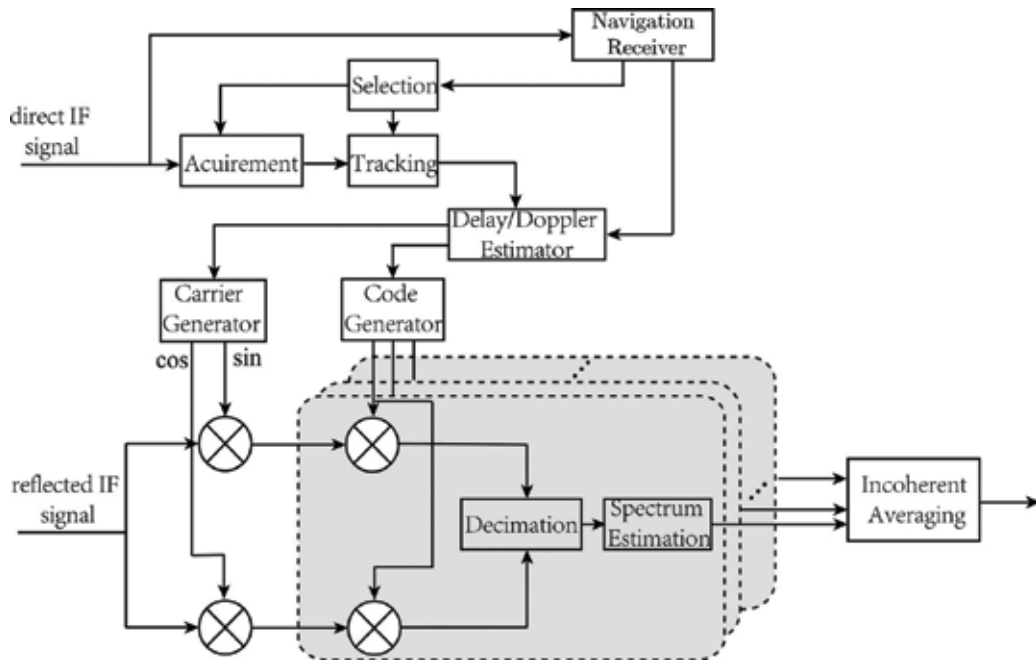
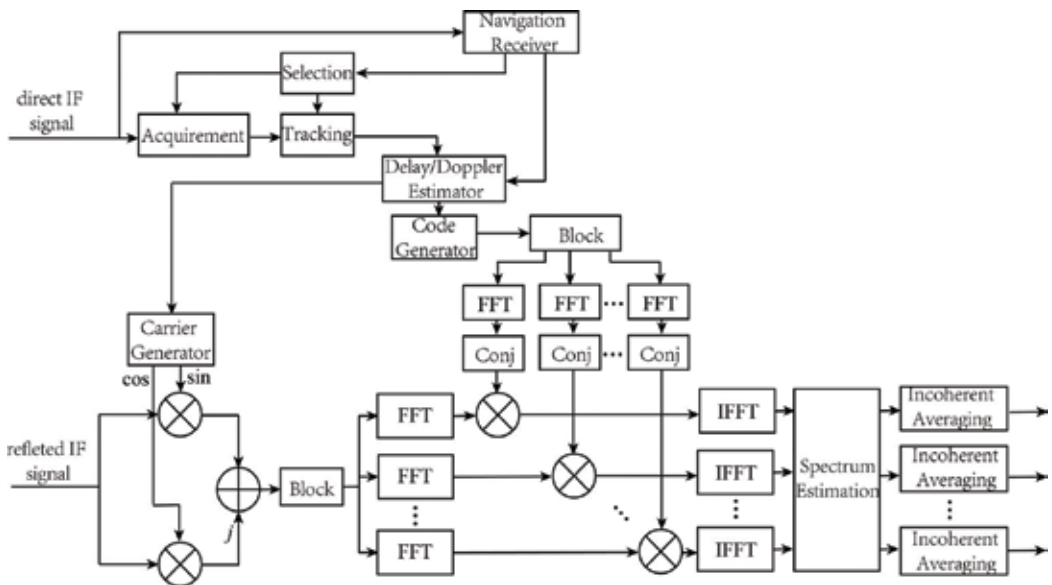**Figure 9.** Block diagram of parallel in Doppler domain.

Figure 10. Block diagram of parallel correlation in delay-Doppler domain.

and the element number of the block N is $[2N_s/N_f]$ ($[\cdot]$ is rounding operator). The numbers of the multipler and adder for each block is $12N\log_2 2N$ and $18N\log_2 2N$. The numbers of multiplier and adder required by the spectrum estimation are $4N_b\log_2 2N_b$ and $6N_b\log_2 2N_b$. The total numbers of multipler and adder for DBZP are $12N_bN\log_2 2N + 2N_f\log_2 N_f$ and $18N_bN\log_2 2N + 3N_f\log_2 N_f$.

3.3.5. Comparison of computational complexity

Although the computational complexities of the correlation processing have been analyzed above, here, the simulated comparison is conducted. It is assumed that the range and the resolution of Doppler frequency are $[-8, 8]$ KHz and 500 Hz, respectively, and the numbers of delay bin and the DBZP block are same. **Figure 11** gives the simulated computational complexities when the numbers of the sampling are 2048, 4096, 8192, and 16,384, from which it is seen that (1) serial correlation needs most multiplier and adder and consumes most hardware resources so that it is unsuitable for spaceborne receiver; (2) parallel correlation methods significantly reduce the number of required multiplier and adder, especially parallel correlation in Doppler domain. It should be noted FFT and IFFT on hardware is a 2-based operation which needs to padding zero to make the number of operating point to be the power of 2. Large number of padding zero causes the increasing of computational complexity; therefore, it has to be considered to choose optimal DDM parameters.

3.4. Processing of UK-TDS/UK-TDS-1 data

The raw IF data from UK-DMC and UK-TechDemoSat-1 satellite are used to demonstrate the correlation approaches above. UK-DMC satellite which first carried an experimental GNSS reflectometry aimed to monitor the disaster using optical imaging. In 2003, UK-DMC satellite was launched into the sun-synchronous orbit with 680 km altitude. Some
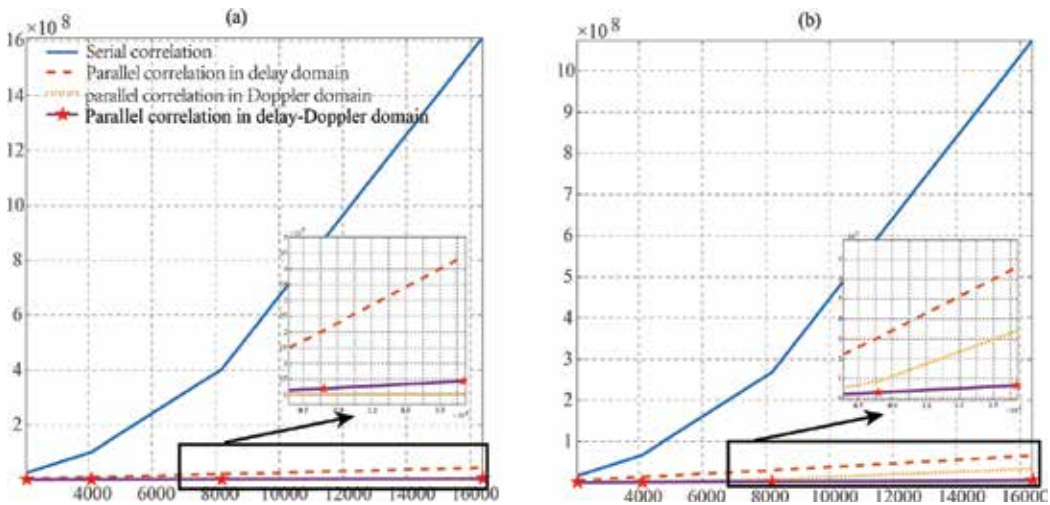

Figure 11. Simulated number of multiplier and adder required correlation processing.

researchers have used GNSS-R data from UK-DMC satellite to demonstrate the feasibility of remotely sensing sea wind speed, sea ice, and others. Following the end of UK-DMC's operational life in 2011, SSTL launched UK-TDS-1 satellite equipped with new generation spaceborne GNSS-R receiver SGR-ReSI in 2014. The orbit of UK-TDS-1 satellite is also sun-synchronous orbit with the altitude and the inclination of 635 km and 98.4°, respectively. UK-TDS-1 satellite is controlled and operated with a 8-day duty circle shared between 8 payloads. SGR-ReSI can be operated for only 2 days in every 8-day period. Both UK-DMC and UK-TDS-1 satellite support the collection of raw IF data, however, because of the limitation of the storage capacity, only a few raw IF data are available. Here, the data collected by UK-DMC on November 11, 2004 and UK-TDS-1 on January 27, 2015 are used. The corresponding ground track of GNSS, UK-DMC, and UK-TDS-1 satellite, and the positioning of the specular pointing are shown in **Figure 12**, in which both specular pointings lie to the ocean surface, i.e., the signals collected by UK-DMC and UK-TDS-1 satellite are both reflected from the ocean surface. **Figure 13** shows the DDM distribution of reflected GNSS signal collected by UK-DMC and UK-TDS-1 satellite, in which from left to right, the DDMs are computed using serial correlation, parallel correlation in delay domain, parallel correlation in Doppler domain, and parallel correlation in delay-Doppler domain. From **Figure 13**, it could be seen that (1) the DDMs from UK-DMC and UK-TDS-1 satellites both show the distribution of the horseshoe shape as simulated in **Figure 4** of Section 2.3; (2) the DDMs computed by four correlation methods have similar distribution in delay-Doppler domain. The assessment metrics defined by expression (19) are computed for the DDMs in **Figure 13**. The results are given in **Table 1**, which indicate that compared to the serial correlation, parallel correlations have negligible MSEs, and in addition, one-dimensional parallel methods have lower MSEs than two-dimensional ones. The results above illustrate that parallel correlations, especially the parallel correlation in Doppler domain, are better chosen to compute the DDM than serial one for spaceborne GNSS-R receiver.
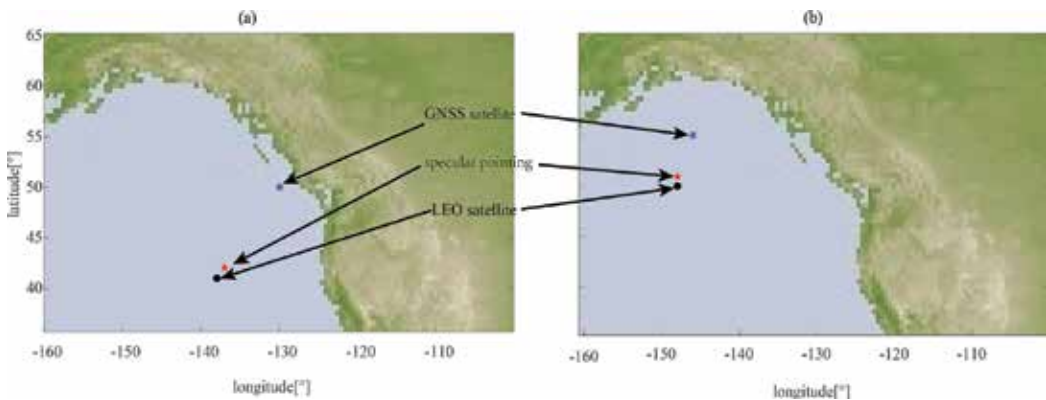

Figure 12. Ground track of satellites and the positioning of the specular pointing when (a) UK-DMC and (b) UK-TDS-1 data were collected.
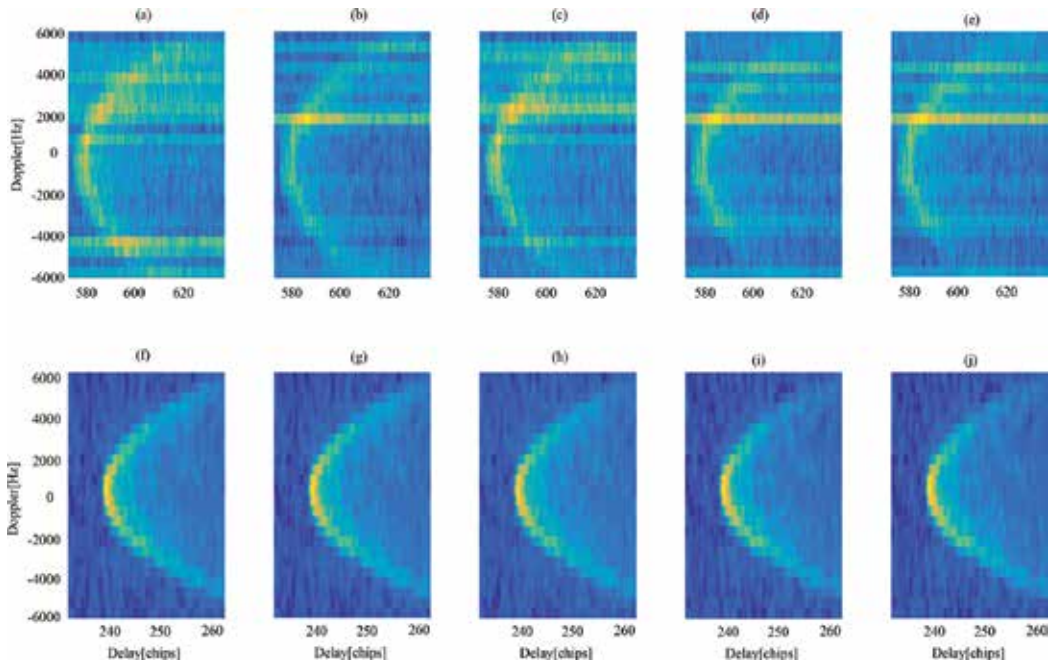
Figure 13. DDM of (a)–(e) UK-DMC and (f)–(j) UK-TDS-1 data.

Satellite	Serial	Parallel ($\times 10^{-3}$)		
		Delay domain	**Doppler domain**	**Delay-Doppler domain**
UK-DMC	0	5.81	2.62	6.94
UK-TDS-1	0	0.09	0.32	1.17

Table 1. Comparison of normalized MSE between correlation powers obtained by serial and parallel processing.

4. Retrieving of wind speed

Retrieving wind speed is an inversion problem of remote sensing, of which the key is to find and extract the sensitivity parameters to wind speed from the basic observables of the remote sensing sensors. In fact, L-band signals are less optimal frequency for the measurement of wind speed compared to scatterometers, however, are not significantly attenuated by rainfall so that they are available to observe wind speed in tropical cyclones. As mentioned in Section 1, at present, three types of retrieving approaches for spaceborne GNSS-R have been proposed. Here, the approaches based on the direct mapping from DDM metrics to wind speed are discussed.

4.1. Retrieving method

An expression for the distribution of DDM in delay-Doppler domain has been given in (15), in which the parameter related to sea wind speed is bi-static scattering coefficient σ_0^2. Therefore, it is able to observe wind speed using bi-static scattering coefficient derived from the DDM. As bi-static scattering coefficient is ununiform over the sea surface, the integrated area used in calculating bi-static scattering coefficient should be chosen to optimize the accuracy of calculation. The ideal area in which bi-static scattering coefficient could be assumed as the constant is a small region around the specular point. The corresponding power which usually is nearby the peak DDM is expressed as [30, 39]

$$
\left\langle \left| Y\left(\widehat{\tau},\widehat{f}\right) \right|^2 \right\rangle = A_R^2 T_I^2 \sigma_0^2 \iint \frac{G^2(\boldsymbol{\rho})\Lambda^2(\widehat{\tau}-\tau(\boldsymbol{\rho}))\cdot \mathrm{sinc}^2\left(\widehat{f}-f(\boldsymbol{\rho})\right)}{4\pi D(\boldsymbol{\rho})^2 d(\boldsymbol{\rho})^2} d^2\boldsymbol{\rho} \tag{25}
$$

The expression computing bi-static scattering coefficient is derived as

$$
\sigma_0^2 = \frac{\left\langle \left| Y\left(\widehat{\tau},\widehat{f}\right) \right|^2 \right\rangle}{\left[A_R^2 T_I^2 \iint \frac{G^2(\boldsymbol{\rho})\Lambda^2\left(\widehat{\tau}-\tau(\boldsymbol{\rho})\right)\cdot \mathrm{sinc}^2\left(\widehat{f}-f(\boldsymbol{\rho})\right)}{4\pi D(\boldsymbol{\rho})^2 d(\boldsymbol{\rho})^2} d^2\boldsymbol{\rho} \right]} \tag{26}
$$

In expression (26), A_R is an unknown parameter related to the transmit power and antenna gain of GNSS satellite, the atmospheric loss. Generally, direct signal and some atmospheric attenuation models are used to estimate it. In addition, the integral operation in expression (26) is time-consuming; therefore, the expression (26) is further simplified by assuming the changes of $G(\boldsymbol{\rho})$, $\tau(\boldsymbol{\rho})$, and $f(\boldsymbol{\rho})$ in the small region around the specular pointings as unity. The simplified expression of σ_0^2 could be given as

$$
\widehat{\sigma}_0^2 = \frac{4\pi D^2 d^2 \left\langle \left| Y\left(\widehat{\tau},\widehat{f}\right) \right|^2 \right\rangle}{A_R^2 T_I^2 G_{\mathrm{SP}} A_0} \tag{27}
$$

where G_{SP} is the gain of the nadir-looking antenna at the specular point; and A_0 is the size of scattering area around the specular point. To reduce the influence of the noise, $\left\langle \left| Y\left(\widehat{\tau},\widehat{f}\right) \right|^2 \right\rangle$ in expressions (26) and (27) could be replaced by the averaged DDM around the peak as

$$
P_{\mathrm{avg}} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left\langle \left| Y\left(\tau_i, f_j\right) \right|^2 \right\rangle \tag{28}
$$

As illustrated in **Figure 5**, the DDM not only gives the distribution of correlation power in delay-Doppler domain, but also presents the spatial distribution of power. The given

delay-Doppler interval in expression (28) determines the geometric resolution of spaceborne GNSS-R as [40]

$$R_{\mathrm{geo}} = \sqrt{\int \Pi\left(\frac{\tau(\boldsymbol{\rho}) - \tau_L/2}{\tau_L}\right)\cdot\Pi\left(\frac{f(\boldsymbol{\rho})}{2f_M}\right)d\boldsymbol{\rho}} \tag{29}$$

where τ_L and f_M are the maximum delay and Doppler frequency in expression (28).

4.2. Demonstration of UK-TDS-1 and ASCAT data

The gain modes of RF front-end in SGR-ReSI have unmonitoring AGC and fixed gain. After June 2015, SGR-ReSI was configured as the fixed gain of 36 dB. The UK-TDS-1 Level 1 data collected from November 26, 2015 to April 19, 2016 are used to demonstrate the feasibility of retrieving wind speed using the approach above. Wind speed monitored by ASCAT-A/B which have a resolution of 25 km are selected as in-situ data. The matchup between the specular pointing of UK-TDS-1 and the ground track of ASCAT-A/B are set to allow a maximum space and time separation of 0.25° and 1 h. To improve the accuracy of retrieved wind speed, it is necessary to select high-quality L1 data to retrieve wind speed. The first is spatial selection that the latitude range of the specular point is limited from −50°S to −50°N to remove the data reflected off the sea ice, and the distances between the specular pointing and the coastline are controlled to be over 0.5° to remove the impact of the land on the DDM. The second one is the selection based on the antenna gain at the specular point. The nadir-looking antenna gains of SGR-ReSI rapidly attenuate when the incident angles of incoming signal are over 40°. Therefore, the incident angles of GNSS signal are limited within the range of 30°. The reflected GNSS signal arriving at the nadir-pointing antenna is exceedingly weak. The coherent integration of 1 ms and incoherent averaging of 1000 significantly improve the signal-to-noise ratio (SNR); however, some DDMs still are buried in the noise floor. In addition, the SNR is one of the factors influencing on the accuracy of wind speed. Therefore, it is needed to remove DDMs with the low SNR. The SNR of DDM could be defined as

$$\mathrm{SNR} = \frac{P_{\mathrm{peak}}}{N_{\mathrm{floor}}} - 1 \tag{30}$$

where P_{peak} is the peak DDM; and N_{floor} is the noise floor of DDM, which can be computed by averaging the DDM in the region where reflected signal is absent.

After the DDMs are selected, the delay and Doppler interval in expression (28) should be determined. The delay and Doppler interval are limited by the geometric resolution. **Figure 14** shows the relationship between the geometric resolution and the incident angle for different delay interval when Doppler interval is [−1, 1] KHz, from which it is seen that when delay interval is [−0.25, 0.25] chips, the geometric resolution is lower than 25 km which is required resolution of the traditional spaceborne remote sensing. Based on the discussion, the delay and Doppler interval in expression (28) are configured as [−0.25, 0.25] chips and [−1, 1] KHz.
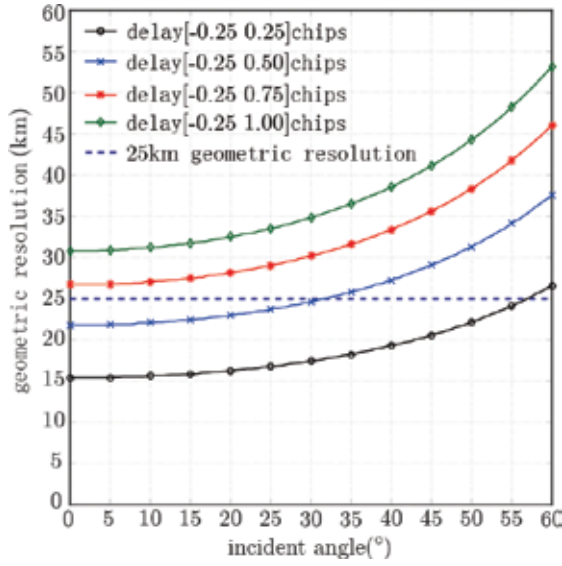
Figure 14. Relationship between geometric resolution and incident angle.

The retrieval model is the mapping linking wind speed with the bi-static scattering coefficient computed by expression (26) or (27). Note that because of the unknown transmit power and antenna gain, the influence of A_R^2 on the bi-static scattering coefficient ignored by assuming them as the constants. Moreover, when the incident angle of GNSS signal is lower than 35°, $1/\cos^2\theta$ could be used to replace A_0 in the expression (27) because of the proportional relationship between them [41]. Therefore, the observable used to develop the link with wind speed is not a true bi-static scattering coefficient. The selected data are divided into two groups: train and test set. The relationship between the observable and wind speed is fitted as

$$U_{10} = Ae^{B\sigma_{obs}^2} + C \tag{31}$$

where σ_{obs}^2 is the obervable computed by expressions (26) and (27) when A_R^2 are ignored and A_0 is replaced using $1/\cos^2\theta$; and A, B, and C are fitted parameters. **Figure 15** shows the inversely proportional relationship between the observable computed by expressions (26) and (27) and wind speed. The fitted parameters in expression (31) are obtained by fitting the observables of train set with wind speed measured by ASCAT-A/B. The compared results between ASCAT-A/B-retrieved wind speed using test set are presented in **Figure 16** and **Table 2**. The biases between ASCAT-A/B and retrieved wind speed are ignorable, 0.024 m/s and 0.021 m/s, for the observables derived by expressions (26) and (27). The RMSEs are 2.11 m/s and 2.15 m/s, respectively. In addition, it is clear that the retrieving accuracy for high wind speed is larger than the low wind speed as illustrated in **Figure 17**, in which when wind speed is lower than 11 m/s, the RMSE is less than 2 m/s. The reasons of bad RMSE for high wind speed are that (1) the low percentage of high wind speed in train set causes the poor fitting of expression (31); and (2) the sensitivity of the observable on wind speed decreases as wind speed increases. Moreover, it is pointed that the approach based on the expression (27) can take less time about 3000 times than the approach based on the expression (26).
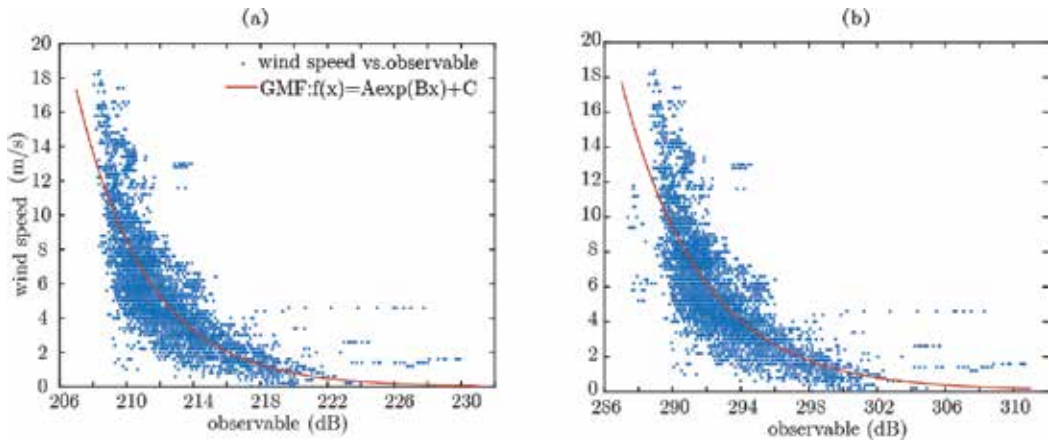
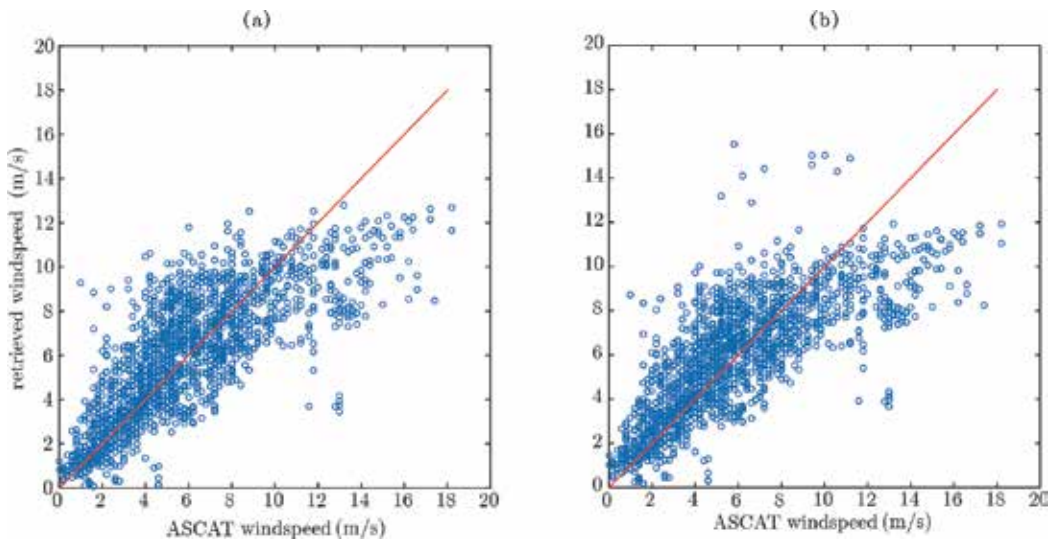**Figure 15.** Scatter plot of observable computed by expression (a) (26) and (b) (27) vs. wind speed.

Figure 16. Scatter plot of ASCAT-A/B and retrieved wind speed.

Observable	Train set			Test set	
	A	**B**	**C**	**Bias (m/s)**	**RMSE (m/s)**
Expression (26)	3.506×10^{22}	-0.237	-0.0115	0.024	2.11
Expression (27)	4.299×10^{27}	-0.2177	-0.0665	0.021	2.15

Table 2. Trained parameters of Eq. (31) using train set and corresponding test performance using test set.

The SNR of DDM has an important impact on retrieving result of wind speed as shown in **Figure 18**. The accuracy of retrieved wind speed is improved by selecting the DDMs with higher SNR. However, it should be noted that the larger the threshold of the SNR is, the lesser
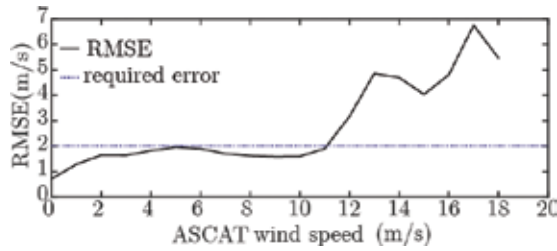
Figure 17. Relationship of RMSE of retrieved wind speed vs. wind speed.
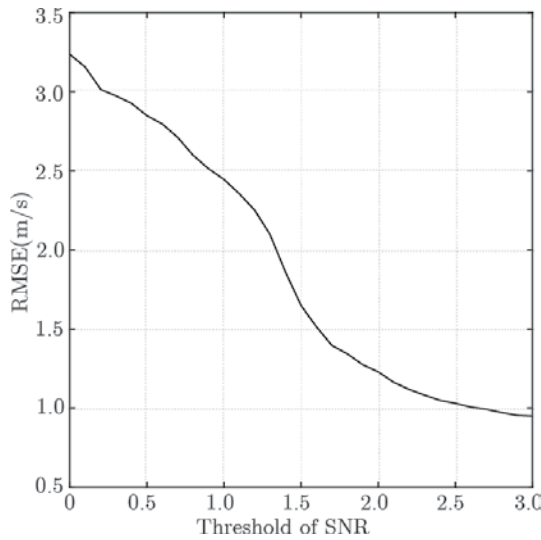

Figure 18. Trend plot of RMSE as the threshold of SNR changing.

the DDMs used to the retrieved wind speed are, i.e., the coverage of the UK-TDS-1 is lower for the higher threshold of the SNR. It has to be noted that no precise calibration and correction for the DDM measurement in SGR-ReSI also is one of the reasons causing the low performance of retrieving wind speed.

5. Conclusion

GNSS-R has been a new tool of remote sensing to monitor Earth's physical parameters, such as sea wind speed, sea surface height, sea ice, soil moisture and others. Following the success of UK-DMC satellite which was carried on the first GNSS-R receiver, the spaceborne GNSS-R which could provide the global and all-days observation has been an hot focus of the research groups. Some spaceborne missions have been proposed and planned, such as CYGNSS of NASA, GEROS-ISS of ESA, [3]CAT-2 of the Remote Sensing Lab and the NanoSat Lab at the universitat Politècnica de Catalunya-Barcelona Tech. This chapter was focused on the discussion of spaceborne GNSS-R to monitor sea wind speed. First, the basic theories

of spaceborne GNSS-R, including the geometry, polarization, and scattering model were discussed. The geometry of GNSS-R mainly is determined by the incident angle of GNSS signal, and the height of LEO and GNSS satellite. When GNSS signal occurs reflection on the sea surface, the polarization of GNSS signal is changed. At nadir and small incident angle, the LCHP component dominates the reflected GNSS signal. For the diffuse scattering of the GNSS signal, the correlation power level called as DDM spreads to large delay and Doppler frequency and presents the horseshoe shape in delay-Doppler domain. Then, the spaceborne receivers and processing methods were discussed. The results showed that parallel correlation in Doppler domain has lowest computational complexity and optimal performance so that it is optimal to choose to process the reflected GNSS signal in real-time for spaceborne receiver. Finally, the approach based on DDM metric was used to retrieve wind speed. Proposed DDM metrics have inversely proportional relationship with wind speed. By latitude mask and the threshold of SNR to remove low-quality data, proposed approaches could obtain the RMSE of 2.11 m/s and 2.15 m/s. For further improvement of retrieving accuracy, it is necessary to set higher SNR threshold which results in lower spatial coverage. In one word, GNSS-R will become an effective way to observe global sea wind speed, especially low attenuation of GNSS signal for the rainfall will be an optimal choice to monitor tropical cyclone.

Author details

Dongkai Yang* and Feng Wang

*Address all correspondence to: wangf.19@163.com

School of Electronic and Information Engineering, Beihang University, Beijing, China

References

[1] Adler RF. Estimating the benefit of TRMM tropical cyclone data in saving lives. In: American Meteorological Society, 15th Conference on Applied Climatology; 2005; Savannah, USA

[2] Draper DW, Mccabe PJ. A review of satellite altimeter measurement of sea surface wind speed: With a proposed new algorithm. Journal of Geophysical Research. 1985;**90**(C3): 4707-4720. DOI: 10.1029/JC090iC03p04707

[3] Stoffelen A, Anderson DLT. Wind retrieval and ERS-1 scatterometer radar backscatter measurements. Advances in Space Research. 1993;**13**(5):53-60. DOI: 10.1016/0273-1177(93)90527-I

[4] Chang PS, Li L. Ocean surface wind speed and direction retrievals from the SSM/I. IEEE Transactions on Geoscience and Remote Sensing. 1998;**36**(6):1866-1871. DOI: 10.1109/36.729357

[5] Komkathy A, Armatys M, Masters D, Axelrad P, Zavorotny VU, Katzberg SJ. Retrieval of ocean wind speed and wind direction using reflected GPS signals. Journal of Atmospheric and Oceanic Technology. 2004;**21**(3):515-526

[6]   Lowe ST, Zuffada C, Chao Y, Kroger P, et al. 5-cm precision aircraft ocean altimetry using GPS reflections. Journal of Geophysical Research. 2002;**29**(10):13(1)-13(4). DOI: 10.1029/ 2002GL0-14759

[7]   Rivas MB, Maslanik JA, Axelrad P. Bistatic scattering of GPS signals off Arctic sea ice. IEEE Transactions on Geoscience and Remote Sensing. 2009;**48**(3):1548-1553. DOI: 10.1109/TGRS.2009.2029342

[8]   Katzberg SJ, Torres O, Grant MS, Masters D. Utilizing calibrated GPS reflected signals to estimate soil reflectivity and dielectric constant: Results from SMEX02. Remote Sensing of Environment. 2005;**100**(1):17-28. DOI: 10.1016/j.rse.2005.09.015

[9]   Hall C, Cordy R. Multistatic scatterometry. In: Proc. IEEE Int. Geoscience Remote Sensing Symp; 1988; Edinburgh, Scotland

[10]   Martin-Neira M. A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry. ESA Journal. 1993;**17**(4):331-355

[11]   Auber JC, Bibaut A, Rigal MJ. Characterization of multipath on land and sea at GPS frequencies. In: ION GPS-94; Sept. 1994; Salt Lake City, UT. p. 1155-1171

[12]   Garrison JL, Katzberg SJ, Hill MI. Effect of sea roughness on bistatically scattered range coded signals from the global positioning system. Geophysical Research Letters. 1998; **25**(13):2257-2260. DOI: 10.1029/98GL51615

[13]   Lowe S, LaBrecque JL, Zuffada C, Romans LJ, Young LE, Hajj GA. First spaceborne observation of an earth-reflected GPS signal. Radio Science. 2002;**37**(1):1-28. DOI: 10.1029/ 2000RS002539

[14]   Gleason ST, Hodgart S, Yiping S, Gommenginger C, Mackin S, Adjrad M, Unwin M. Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. IEEE Transactions on Geoscience and Remote Sensing. 2005;**43**(6):1229-1241. DOI: 10.1109/TGRS.2005.845643

[15]   Unwin M, Duncan S, Jales P, Blunt P, Brenchley M. Implementing GNSS reflectometry in space on the TechDemoSat-1 mission. In: Proc. Institute Navigation; 2014; Tampa, Florida. p. 1222-1235

[16]   Martin-Neira M, D'Addio S, Buck C, Floury N, Prieto-Cerdeira R. The PARIS ocean altimeter in-orbit demonstrator. IEEE Transactions on Geoscience and Remote Sensing. 2011;**49**(6):2209-2237. DOI: 10.1109/TGRS.2010.2092431

[17]   Wickert J, Andersen O, Beyerle G, Cardellach E, Chapron B, Gommenginger C, Hatton J, Høeg P, Jäggi A, Jakowski N, Kern M, Lee T, Martin-Neira M, Montenbruck O, Pierdicca N, Rius A, Shum CK, Zuffada C. Innovative remote sensing using the International Space Station: GNSS reflectometry with GEROS. In: Proc. IEEE Int. Geoscience Remote Sensing Symp; 2014; Québec, Canada

[18]   Ruf CS, Gleason S, Jelenak Z, Katzberg S, Ridley A, Rose R, Scherrer J, Zavorotny V. The CYGNSS nanosatellite constellation hurricane mission. In: Proc. IEEE Int. Geoscience Remote Sensing Symp; 2012; Munich, Germany. pp. 214–216

[19] Zavorotny VU, Voronovich AG. Scattering of GPS signals from the ocean with wind remote sensing application. IEEE Transactions on Geoscience and Remote Sensing. 2000; **38**(1):951-964. DOI: 10.1109/36.841977

[20] Marchan JF, Alvarez NR, Valencia E, Ramos-Perez I. An efficient algorithm to the simulation of delay–Doppler maps of reflected global navigation satellite system signals. IEEE Transactions on Geoscience and Remote Sensing. 2009;**47**(8):2733-2740. DOI: 10.1109/TGRS.2009.2014465

[21] Garrison JL. A statistical model and simulator for ocean-reflected GNSS signals. IEEE Transaction on Geoscience and Renote Sensing. 2016;**54**(10):6007-6019. DOI: 10.1109/TGRS.2016.2579504

[22] Clarizia MP, Gommeninger C, Di Bisceglie M, Galdi C, Srokosz MA. Simulation of L-band bistatic returns from the ocean surface: A facet approach with application to ocean GNSS reflectometry. IEEE Transactions on Geoscience and Remote Sensing. 2012;**50**(3): 960-971. DOI: 10.1109/TGRS.2011.2162245

[23] Lowe ST, Kroger P, Franklin G, LaBrecque JL, Lerma J, Lough Ml, Marcin MR, Muellerschoen RJ, Spitzmesser D, Young LE. A delay/Doppler-mapping receiver system for GPS-reflection remote sensing. IEEE Transactions on Geoscience and Remote Sensing. 2002;**40**(5):1150-1163. DOI: 10.1109/TGRS.2002.1010901

[24] Nogués-Correig O, Galí El C, Campderrós JS, Rius A. A GPS-reflections receiver that computes Doppler/delay maps in real time. IEEE Transactions on Geoscience and Remote Sensing. 2007;**45**(1):156-174. DOI: 10.1109/TGRS.2006.882257

[25] Steenwijk R, Unwin M, Jales P. A next generation spaceborne GNSS receiver for navigarion and remote sensing. In: Satellite Navigation Technoloies and Ruropean Workshop on GNSS Signals and Signal Processing; 2010; Noordwijk, Netherlands

[26] Wickert J, Cardellach E, Martin-Neria M, et al. GEROS-ISS: GNSS REflectometry, radio Occulation and Scatterometry Onboard the international Space Station. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016;**9**(10):4552-4581. DOI: 10.1109/JSTARS.2016.2614428

[27] Garrison JL, Komjathy A, Zavorotny VU, Katzberg SJ. Wind speed measurements using forward scattered GPS signals. IEEE Transactions on Geoscience and Remote Sensing. 2002;**40**(1):50-65. DOI: 10.1109/36.981349

[28] Li C, Huang W. An algorithm for sea-surface wind field retrieval from GNSS-R delay_doppler map. IEEE Geoscience and Remote Sensing Letters. 2014;**11**(12):2110-2114. DOI: 10.1109/LGRS.2014.2320852

[29] Rodriguez-Alvarez N, Akos DM, Zavorotny VU, Smith JA, Camps A, Fairall CW. Airborne GNSS-R wind retrievals using delay-Doppler maps. IEEE Transactions on Geoscience and Remote Sensing. 2013;**51**(1):626-641. DOI: 10.1109/TGRS.2012.21-96437

[30] Foti G, Gommenginger C, Jales P, Unwin M, Shaw A, Robertson C, Roselló J. Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission. Geophysical Research Letters. 2015;**42**:5435-5441. DOI: 10.1002/2015GL064204

[31] Rodriguez-Alvarez N, Garrison JL. Generalized linear observables for ocean wind retrieval from calibrated GNSS-R delay-Doppler maps. IEEE Transactions on Geoscience and Remote Sensing. 2016;**52**(2):1142-1155. DOI: 10.1109/TGRS.2015. 2475317

[32] Clarizia MP, Ruf CS. Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) misson. IEEE Transactions on Geoscience and Remote Sensing. 2016;**54**(8):4419-4432. DOI: 10.1109/TGRS.2016.2541343

[33] Valencia E, Camps A, Marchan-Hernandez JF, Park H, Bosch-Lluis X, Rodriguez-Alvarez N, Ramos-Perez I. Ocean surface's scattering coeggicient retrieval by delay-Doppler map inversion. IEEE Geoscience and Remote Sensing Letters. 2011;**8**(4):750-754. DOI: 10.1109/ LGRS.2011.2107500

[34] Schiavulli D, Nunziata F, Migliaccio M, Frappart F, Ramilien G, Darrozes J. Reconstruction of the radar image from actual DDMs collected by TechDemoSat-1 GNSS-R mission. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; **9**(10):4700-4708. DOI: 10.1109/JSTARS.2016.2543301

[35] Marchan JF, Camps A, Rodriguez Alvarez N, Bosch X, Ramos-Perez I, Valencia E. PAU/GNSS-R: Implementation, performance and first results of a real-time delay-Doppler map reflectometer using global navigation satellite system signals. Sensors. 2008;**8**:3005-3019. DOI: 10.3390/s8053005

[36] Carrneo-Luengo H, Camps A, Ramos-P'erez I, Rius A. Experimental evaluation of GNSS-reflectometyr altimetric precision using P(Y) and C/A signals. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2014;**7**(5):1493-1500. DOI: 10.1109/JSTARS.2014.2320298

[37] Carreno-Luengo H, Camps A, Perez-Ramos I, Rius A. Pycaro's instrument proof of concept. In: Proc. Workshop ReflectometryUsing GNSS Other Signals Opportunity (GNSS+R); West Lafayette, IN; 2012. pp. 1–4

[38] Han L, Meng Y, Wang Y, Han X. A fast algorithm of GNSS-R signal processing based on DBZP. In: China Satellite Navigation Conference (CSNC); 2017; Shanghai, China. pp. 187-197

[39] Gleason S. Space-based GNSS Scatterometry: Ocean wind sensing using an empirically calibrated model. IEEE Transactions on Geoscience and Remote Sensing. 2013;**51**(9):4853-4863. DOI: 10.1109/TGRS.2012.2230401

[40] Clarizia MP, Ruf CS. On the spatial resolution of GNSS reflectometry. IEEE Geoscience and Remote Sensing Letters. 2016;**13**(9):1064-1068. DOI: 10.1109/LGRS. 2016.2565380

[41] Soisuvarn S, Jelenak Z, Said F, et al. The GNSS reflectometry response to the ocean surface winds and waves. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing. 2016;**9**(10):4678-4699. DOI: 10.1109/JSTARS. 2016.2602703

GPS, GNSS and GLONASS

GNSSs, Signals, and Receivers

Mohamed Tamazin, Malek Karaim and
Aboelmagd Noureldin

Additional information is available at the end of the chapter

Abstract

This chapter describes Global Navigation Satellite Systems (GNSSs) and their signal characteristics, beginning with an overview of Global Positioning System (GPS) architecture and describing its three primary segments: control, space, and user segments. After that, it addresses the GPS modernization program including the new civilian and military signals and their significance. It continues by outlining the GPS signal characteristics and the sources of GPS measurement error. GPS receivers as well are briefly described. Then, it gives an overview of the GLONASS and describes its modernization program. Additionally, it delves into many aspects the GLONASS, including GLONASS signal characteristics, the GLONASS radio frequency (RF) plan, pseudorandom (PR) ranging codes, and the intra-system interference navigation message. Finally, GPS and GLONASS are compared to highlight the advantages of combined GPS and GLONASS measurements over the GPS-only measurements.

Keywords: GNSS, GPS, GLONASS, signals, GNSS modernization

1. Introduction

Navigation solutions have become part of our daily life due to their widespread use in a range of applications including agriculture, navigation by land vehicles, and pedestrian navigation. A key navigation technology used in such applications is Global Navigation Satellite Systems (GNSSs), and several such systems currently provide this service. The US Global Positioning System (GPS) was this first such fully functional system. GLONASS, the Russian system, was the second to be active, and it also has global coverage. Similarly, the European Union satellite navigation system, Galileo, is scheduled to be fully operational in 2018.

While each of these systems has unique characteristics, all have major aspects in common. Each has a space segment, control segment, and user segment. What is more, all are based on transmitting radio frequency (RF) signals in a one-way fashion from satellites to receivers on and near the Earth's surface. Using measurements obtained from these signals, a GNSS receiver can find its position, velocity, and time (PVT) solution. Moreover, all GNSS systems use the notion of time-of-arrival (TOA) ranging. This requires measuring the signal transit time and the time interval the signal takes to travel between the satellite and the receiver to calculate the receiver-to-satellite range [1]. The transmitter-to-receiver distance can then be obtained by multiplying the signal transit time by the speed of light.

This chapter provides an overview of Global Positioning System (GPS) and GLONASS and their signals. First, it describes the system architecture in terms of the three main segments: control, space, and user. Then, it addresses the new civilian and military GPS signal characteristics, highlighting their significance. Following that, it briefly discusses the GPS measurement error sources. The chapter also covers essential aspects of the GLONASS system, including GLONASS signal characteristics, the GLONASS modernization program, the GLONASS Radio Frequency (RF) plan, pseudorandom (PR) ranging codes, and the intra-system interference navigation message. Finally, advantages of combining both GPS and GLONASS are listed to give the reader insight into the benefits of such integration.

2. Overview of GPS

GPS provides three-dimensional positioning and navigation services for both civilian and military users [2]. The GPS receivers use the TOA ranging to generate code pseudorange to determine the user's position. They also monitor changes in signal frequency to produce a rate of change of range measurements to determine velocity [3]. The time between the transmission of the signal and its arrival at the receiver is measured. The transmitter-to-receiver distance can then be obtained by scaling the signal transit time by the speed of light. Using the concept of trilateration, a GPS receiver can determine its position using the measured travel time along with the satellites' locations that are obtained from the navigation message carried by the signal. Though three satellites can be used to determine the user's position, at least four are required owing to an additional estimation of the receiver clock offset.

Figure 1 illustrates the concept of position fixing by trilateration by using the range to three satellites. Using four satellites to find the position improves the accuracy of the solution by eliminating the receiver clock offset. The first and second user-to-satellite range measurements define two spheres on two different satellites, and the intersection of these two spheres defines a circle of possible receiver positions. A third range measurement, intersecting with the first two, narrows those receiver positions to an ambiguous pair, while the fourth measurement resolves this ambiguity and determines the clock bias. The GPS positioning equations are found in [1–6]. Military GPS signals are more robust against interference and spoofing than civilian signals [3]; hence, the position determined by military signals is more precise than the position determined using civilian signals.
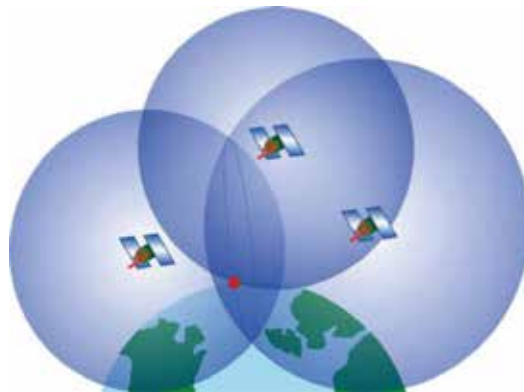
Figure 1. The concept of position fixing by trilateration using signals from three satellites. The user's position is indicated by the red dot [4].

3. The GPS structure

As mentioned earlier, the GPS is composite of three segments [7]: the space segment, a constellation of satellites orbiting the Earth at very high altitudes; the control segment, made up of a group of ground control stations; and the user segment, a user's equipment or simply the variety of military and civilian receivers. **Figure 2** illustrates the three segments, which are discussed in greater detail in this section.

3.1. The space segment

The GPS space segment is made up of a constellation of satellites that continuously broadcasts RF signals to users. In recent years, the US Air Force has operated 32 GPS satellites, of which 24 are available 95% of the time [4]. GPS satellites travel in medium Earth orbit (MEO) at an altitude of approximately 20,200 km, and each circles the Earth twice a day, meaning that the orbital period is approximately 12 h [7]. These satellites are distributed among six equally-spaced orbital planes, each having a target inclination of 55° [6], a satellite distribution that improves the visibility of satellites to GPS users across the globe, thereby enhancing navigation accuracy. GPS satellites broadcast RF signals containing coded information and navigation data, enabling a receiver to calculate pseudoranges and Doppler measurements to estimate position, velocity, and time.

In June 2011, the US Air Force successfully expanded its GPS constellation, known as the "Expandable 24" configuration [9]. Three of the 24 slots were upgraded, and six satellites were repositioned; thus, three additional satellites were added to the constellation. With a 27-slot constellation, GPS improved satellite visibility across the globe. **Table 1** summarizes the features of the current and future generations of GPS satellites, including Block IIA (second generation, "Advanced"), Block IIR ("Replenishment"), Block IIR (M) ('Modernized"), Block IIF ("Follow-on"), and GPS III [10].
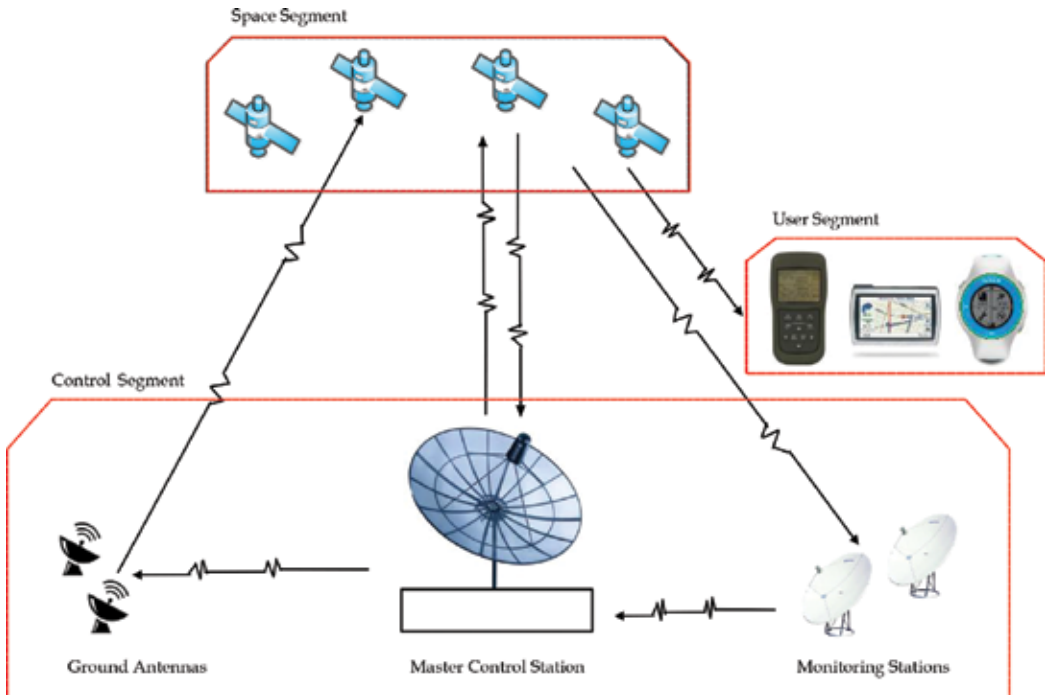
Figure 2. The GPS segments [8].

3.2. The control segment

Made up of a global network of ground facilities that track GPS satellites, the GPS control segment's main tasks are the control and maintenance of the system through monitoring and analyzing signal transmissions and sending commands and data updates to the GPS constellation.

Referring to [7], the current operational control segment includes a Master Control Station (MCS), an alternate master control station, 12 command and control antennas, and 16 monitoring sites. The locations of these facilities are shown in **Figure 3**.

3.3. The user segment

The user segment is represented by the wide array of types of GPS receivers. These capture and track satellite signals and process signals transmitted by GPS satellites, estimate the user-to-satellite ranges and range rates, and compute a PVT solution [12]. A GPS receiver had cost more than $100,000 in the mid-1980s; nowadays, an on-chip receiver is available in the market for less than $20, and it is estimated that more than 1 million receivers have been produced each year since 1997 [1]. As GPS is available at no direct charge to users, they can use receivers at any time and any place across the globe to determine their position [6].
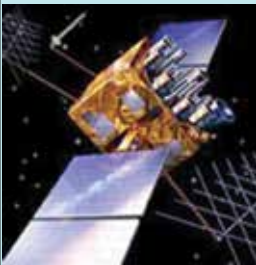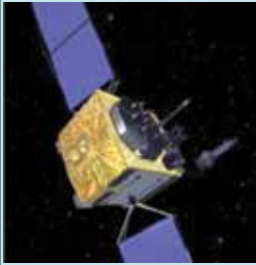
Legacy satellites		Modernized satellites		
Block IIA	Block IIR	Block IIR (M)	Block IIF	GPS III
6 Operational	12 Operational	7 Operational	6 Operational	Now in production
- Coarse acquisition (C/A) code on L1 frequency for civil users - Precise P(Y) code on L1 and L2 frequencies for military users 7.5-year design lifespan - Launched in 1990–1997	- C/A code on L1 - P(Y) code on L1 and L2 - On-board clock monitoring 7.5-year design lifespan - Launched in 1997–2004	- All legacy signals - Second civil signal on L2 (L2C) - New military M code signals for enhanced jam resistance - Flexible power levels for military signals 7.5-year design lifespan - Launched in 2005–2009	- All Block IIR (M) signals - Third civil signal on L5 frequency (L5) - Advanced atomic clocks - Improved accuracy, signal strength, and quality 12-year design lifespan - Launched since 2010	- All Block IIF signals - Fourth civil signal on L1 (L1C) - Enhanced signal reliability, accuracy, and integrity - No Selective Availability - Satellites 9+: laser reflectors; search and rescue payload 15-year design lifespan - Begins launching in 2016

Table 1. The features of the current and future generations of GPS satellites [10].
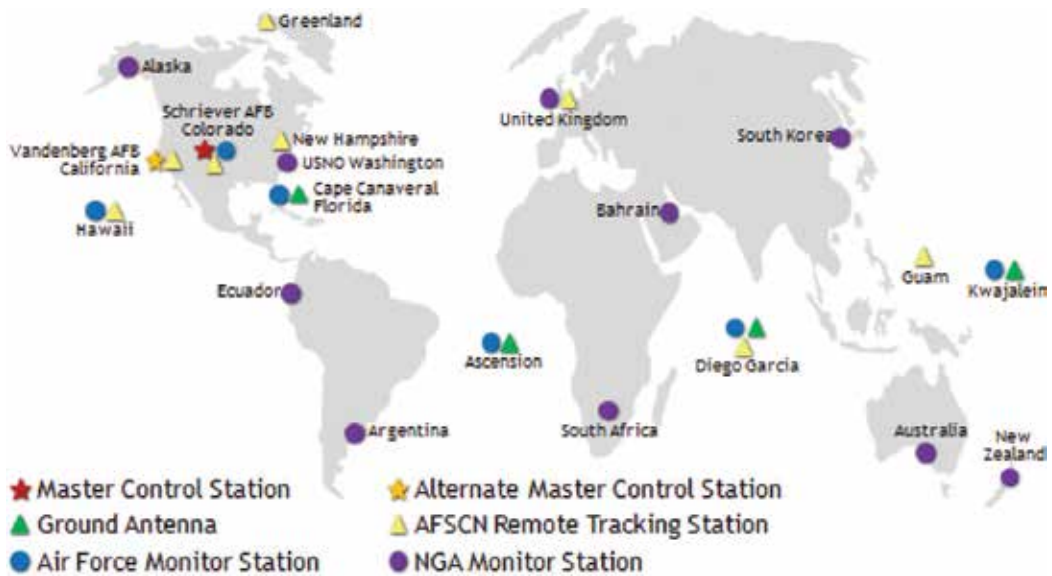
Figure 3. The locations of the GPS Master Control Station, an alternate Master Control Station, 12 command and control antennas, and 16 monitoring sites [11].

4. GPS signal characteristics

GPS satellites produce a central L-band frequency of 10.23 MHz using very stable clocks. Satellites then multiply this frequency by 154 and 120 to generate two carrier frequencies at L1 = 1575.42 MHz and L2 = 1227.60 MHz [13]. GPS signals consist of a carrier signal with frequency L1 or L2, a unique code assigned to each satellite, and a data message conveying information about satellite position, velocity, and clock bias. The two carrier frequencies are modulated by a combination of the data message and the unique code to carry required information to the user. The L1 frequency is modulated by two ranging code signals: the coarse/acquisition code (C/A) and the precise (P) code [2].

Each satellite has a unique C/A PRN code, and all these PRN codes are nearly orthogonal to each other, enabling a GPS receiver to differentiate among the satellites even though the satellites are broadcasting on the same two carrier frequencies, L1 and L2 [14]. Each C/A code repeats every millisecond and has a length of 1023 bit. The duration of each chip in a C/A code is about 1 ms, and the code rate is 1.023 MHz (or megachips/second (Mcps)) with a wavelength of about 300 m. The duration of the P code is about 7 days, and it modulates both L1 and L2. Used only by the military, this code has a rate of 10.23—10 times than that of a C/A code. The P code wavelength is about 30 m, making it much shorter and consequently much more precise than the C/A code [2].

The last key part of the GPS signal is the navigation message. It takes 12.5 min to receive the entire message, which is downloaded at a rate of 50 bit/s [6]. Its most important parts are the ephemeris, almanac data, and satellite clock bias parameters.

To prepare the GPS signal for transmission by the satellite, first, an XOR operation is applied to combine the binary navigation message with the code. If the message bit and the code chip are the same, the result is 0; if they are different, the result is 1. Second, the combined signal is merged with the carrier using binary phase shift keying (BPSK) modulation: a "0" bit leaves the carrier signal intact, whereas a "1" bit causes the signal to be multiplied by −1 and shifts the carrier by 180°. **Figure 4** illustrates this process.

As mentioned above, the PRN code patterns are nearly orthogonal, an important property that makes the satellite identification process much easier [2]. Two codes are orthogonal when the sum of their term products shifted arbitrarily against each other is nearly zero. The cross correlation function for satellites m and n, with PRN codes $C^{(k)}$ and $C^{(l)}$, is expressed as

$$\sum_1^{1023} C^{(k)}(i) \cdot C^{(l)}(i+n) \approx 0, \text{ for all } k \neq l \tag{1}$$

This orthogonality makes the cross satellite interference small [14].

Another important property of PRN codes is that each PRN pattern is almost uncorrelated with itself:

$$\sum_1^{1023} C^{(k)}(i) \cdot C^{(k)}(i+n) \approx 0, \text{ for all } |n| \geq 1 \tag{2}$$

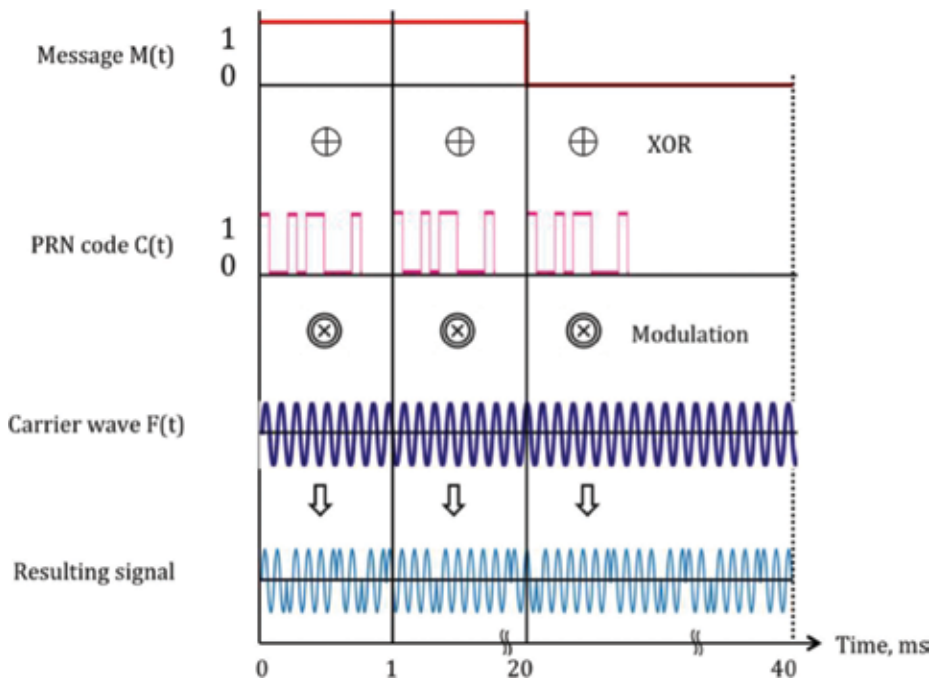**Figure 4.** GPS signal structure [15].

Autocorrelation of a PRN pattern is nearly zero for any shift |n| ≥ 1. When n is zero, however, the function reaches a peak. Using this feature, the receiver compares the PRN code on the received signal against a locally generated replica of the same code to identify which satellite has generated the corresponding signal.

5. GPS receiver architecture

Figure 5 shows the high-level architecture of a GPS receiver. GPS receivers are made up of the antenna, RF front end, local oscillator, and navigation processor. The first element of the receiver architecture is the antenna, which must be able to receive right-hand circularly polarized (RHCP) signals because this is the type of signal transmitted by GPS satellites [1]. Also important is the antenna gain pattern, which indicates how well the antenna performs at various center frequencies, polarizations, and elevation angles.

The preamplifier is the first active component that comes after the antenna. It is often housed in the same enclosure as the antenna element. Because the antenna can receive multiple frequency bands, typically, there is one preamplifier per band; nonetheless, a single preamplifier may cover multiple bands. The main function of the preamplifier is to amplify the signal at the antenna's output [3]. Preamplifiers generally have three components: (1) a preselector filter that removes out-of-band interference and limits the noise bandwidth, (2) burnout protection that prevents possible high-power interference with the electronic components of the receiver, and (3) a low-noise amplifier (LNA). GPS signals are typically very weak, around −160 dBw or 10–6 W; thus, an LNA amplifies the signals by 20 to 35 dB to increase them to levels suitable for processing [17].
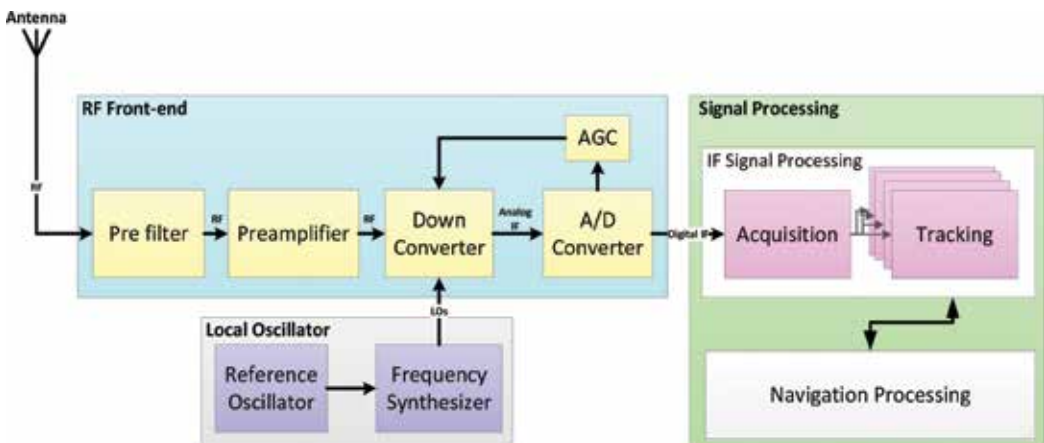

Figure 5. High-level architecture of GPS receivers [16].

After the antenna and LNA comes the RF front end. This unit generates a clean sampled signal for the signal-processing block [12]. Indeed, the front-end pre-filters amplify, downconvert, and digitize the received signal.

Filtering is crucial for several reasons: it rejects out-of-band signals, reduces noise in the received signal, and lessens the impact of aliasing. Wide bandwidth signals can provide high-resolution measurements in the time domain but demand higher sampling rates, causing the receiver to consume much more power [18]. A filter can mitigate this by allowing narrower band signals.

Down-conversion is the process performed by the front end to lower the RF signal frequency to either an intermediate frequency or directly to baseband [3]. This is necessary to facilitate the sampling and filtering processes. The down-conversion is often done using a mixer which multiplies the received signal by a locally generated replica and, then, filters the output signal to remove double-frequency terms [1], as depicted in **Figure 6**. The filtering and down-conversion of the signal frequencies are typically achieved in multiple, consecutive, stages due to the difficulty in implementing a stable band-pass filter with a high central frequency.

The last stage in the processing of the signal inside the RF front end is the conversion of the analogue signal to a digital signal. The band-pass sampling completes both discretization and down-conversion of the signal [12].

GPS receivers make their measurements using the estimates of the signal TOA and received carrier phase and frequency. A single local reference oscillator (see **Figure 4**) forms all frequency references in the receiver [19]. Because the oscillator is critical to receiver performance, particular attention needs to be given to its size, power consumption, stability (both short and long terms), and its temperature and vibration sensitivity [3]. In some cases, GPS receivers have multiple frequency references for down-conversion. In these instances, each mixer requires a precise reference frequency. The process of producing reference frequencies in the receiver from the local oscillator is called frequency synthesis, which uses a combination of integer and fractional frequency multiplications [20].

Figure 4 illustrates that the final stage of a GPS receiver is the navigation processor. This unit receives the conditioned signal (the output of the front end). This filtered and down-converted signal should contain all the necessary information carried by the signal when it
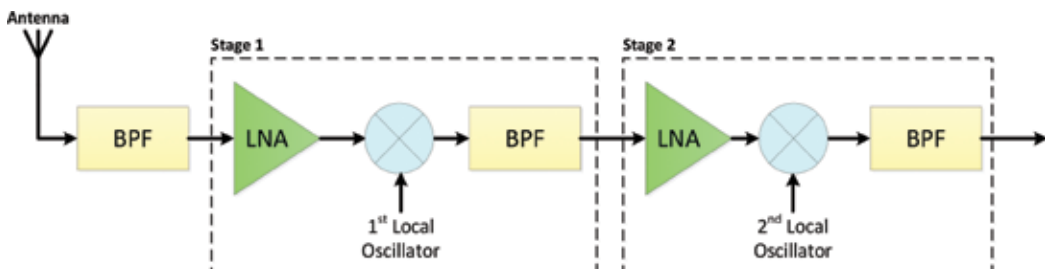

Figure 6. Block diagram of two-cascaded-stage down-conversion.

was received by the antenna. At the navigation-processing stage, the receiver extracts the measurements for pseudorange and rate-of-change of pseudorange to all satellites in view, and using these, it estimates the PVT solution for the antenna.

The navigation process usually happens in two stages: first, the pseudorange and pseudorange rates to each satellite are estimated; second, the user's position, velocity, and time information are estimated using these measurements. Signal processing at this level can be, in turn, divided into the following stages [12]:

- Signal acquisition: This involves detection of the signals from satellites in view and provides a rough estimation of the code delay and the Doppler frequency of the incoming signal from each satellite.

- Signal tracking: This is a recursive estimation process that continuously updates estimates of time-varying signal parameters.

- Signal monitoring: This is simultaneous with tracking and involves estimation of several parameters, including the carrier-to-noise ratio (C/N_0). The receiver uses signal monitoring to decide when loss of lock of signal occurs, for example.

- Navigation message extraction: This process, too, happens in parallel to signal tracking. The navigation message extraction includes satellite ephemerides' decoding.

- Measurement generation: Uses the tracking parameters to estimate ranges and range rate of change for all visible satellites.

- PVT solution: Uses the range and range rate of change estimates to compute the desired navigational solution.

6. GPS measurements

While tracking a satellite signal, a GPS receiver monitors three parameters: pseudoranges, carrier phase, and Doppler [7, 11]. A pseudorange is calculated by measuring the signal transit time from a satellite to the receiver and is described as "pseudo" ranges because these measurements are corrupted by satellite and receiver clock biases [6]. Carrier phase measurements track the difference between the carrier phases for the received and a locally generated replica of the signal. The Doppler measurement reflects the rate of change of the carrier phase [12].

7. GPS errors

GPS signals and measurements are prone to many disturbance factors commonly known as GPs errors. The first error source is due to the drift of both the satellite and receiver clocks.

Despite their high level of accuracy, satellite clocks still drift slightly from GPS time. For affordability reasons and size, receiver clocks are usually much cheaper; consequently, they drift from GPS time rapidly. This drift translates into significant range errors in receiver measurements.

Once it departs the satellite antenna, the GPS signal needs to travel thousands of kilometers to reach to the receiver antenna and then the receiver circuitry. The first and longer part of this trip is by space where the signal maintains its characteristics. However, when the signal enters the atmosphere, this medium causes some unwanted effects. The two primary layers of the atmosphere, namely, ionosphere and troposphere, respectively, will add delays to the signal transit time and, hence, cause some errors in the measurements.

Once it nears the receiver antenna, the signal usually experiences reflections and echoes, i.e., it often bounces off objects near the receiver causing it to hit the antenna from different directions—a phenomenon known as multipath. Multipath is one of the major sources of errors, which harms GPS signals [6]. All aforementioned disturbances are a result of the nature of the signal or the propagation medium and are considered unintentional. Intentional signal degradation or replacement is, in many cases, a more problematic source of GPS errors. One major type of intentional errors is signal jamming. Signal jamming is deliberate interference caused by broadcasts of radio frequency (RF) signals around the receiver neighborhood with the aim of preventing the tracking of true GNSS signals.

8. Overview of GLONASS

Like GPS, GLONASS offers three-dimensional positioning and navigation services for both civilian and military users. In this system too, users determine their position and velocity using pseudorange and carrier phase measurements. Both systems use time-of-arrival (TOA) ranging to determine user position and velocity [21]. The GLONASS includes three components: a constellation of satellites (equivalent to the GPS space segment), ground control stations (also equivalent to the GPS control segment), and user's equipment (as well, equivalent to the GPS user segment) [22]. The ground segment consists of a master control station (MCS). The user segment consists of all the military and civilian receivers.

8.1. GLONASS space segment

The full GLONASS constellation consists of 24 satellites [21]. According to [23], 26 functional GLONASS-M satellites are in orbit, and 22 of them are in service, with four more having reserve status. With the launches of several GLONASS-M satellites and the GLONASS-K satellites, a full constellation of 24 satellites is now available.

GLONASS satellites circle the earth in three orbital planes evenly spaced by 120°. Each plane has eight satellites that are separated by an argument of latitude of 45°, and those satellites have a target inclination of 64.8°—considerably higher than that of GPS satellites. GLONASS

orbits are highly circular with eccentricities smaller than those of GPS and closer to zero [24]. GLONASS satellites have a radius of 25,510 km, which gives an altitude of 19,130 km [22]. Compared to GPS, GLONASS has a shorter orbital period (11 h 15 min 40 s) due to its lower altitude. A comparison of the main differences between GLONASS and GPS is given in later sections.

8.2. GLONASS control segment

A key task of the GLONASS control station is to synchronize the satellite clocks with GLONASS time and calculate the time offset between GLONASS time and UTC [3]. It also uploads clock corrections, predicted ephemeris, and almanac data to GLONASS satellites. Moreover, this segment monitors the status of the current GLONASS constellation and corrects the orbital parameters accordingly. GLONASS uploads its navigation data to the satellites twice a day, while this is done once a day by the GPS [25].

GLONASS's ground control segment has two main parts: the system control center (SCC), located in Moscow, and a network of command tracking stations (CTS), located throughout the former Soviet Union (SU). The roles of the SCC and CTS are similar to those of the GPS Master Control Station and its monitoring stations [22].

8.3. GLONASS user segment

Like that for GPS, the GLONASS user segment contains the end user receiver equipment, which tracks and receives satellite signals. Similar to GPS receivers, these also process signals transmitted by the seen satellites, estimate pseudorange and rate of change of pseudorange from these signals, and calculate a position, velocity, and time (PVT) solution.

9. GLONASS modernization

The design of the GLONASS satellite has been improved several times, resulting in three satellite generations: the original GLONASS (started in 1982), GLONASS-M (started in 2003), and GLONASS-K (started in 2011). There are two types of GLONASS spacecraft in the constellation: the GLONASS-M satellite and GLONASS-K satellite. A brief description of each type is as follows.

9.1. First generation (GLONASS)

The first generation of GLONASS satellites (Uragan) was launched in 1982. Each satellite weighed approximately 1250 kg and was equipped with a basic propulsion system enabling it to relocate within its orbit [26]. Initially, GLONASS's main role was to control the formulation of the navigation signal and obtain the satellite ephemeris and almanac data. This generation is no longer in use.

Satellite series	Launch	Current status	Clock error (s)
GLONASS	1982	Out of service	5×10^{-13}
GLONASS-M	2003	In service	1×10^{-13}
GLONASS-K1	2011	In service	5×10^{-14}
GLONASS-K2	2013	Design phase	1×10^{-14}

Table 2. Roadmap of GLONASS modernization.

9.2. Second generation (GLONASS-M)

GLONASS-M, the modernized version of the former constellation, was launched in 2003, boasting a longer design lifespan of about 7 years and a civil modulation to its L2 frequency band. These changes improved navigation performance, provided updated navigation radio signals, and increased the stability of those signals [27].

9.3. Third generation (GLONASS-K)

Significant improvements came in 2011 with the launch of the third generation, GLONASS-K. Among these changes was the increase of its satellites' lifespan to a decade and the reduction of their weight by half [22]. The accuracy was improved as well, with each satellite transmitting five navigation signals instead of two. These new satellites were intended to transmit four military signals on the L1 and L2 carriers and one civilian signal on the L3 frequency. The GLONASS-K satellites broadcast other signals; two of them are compatible with GPS/Galileo navigational signals. Adding the CDMA signals improved compatibility and enabled interoperability with services provided by other GNSSs, which paved the way for the production of receivers usable with all GNSSs [23]. **Table 2** shows how the system was upgraded over the years.

10. GLONASS signal characteristics

The GLONASS Interface Control Document (ICD) held by the Russian Institute of Space Device Engineering provides the detailed information about the structure of the GLONASS radio signals [22]. In contrast to GPS, GLONASS uses frequency division multiple access (FDMA) for signal modulation. This technique uses the same pseudorandom noise (PRN) code for all satellites to produce a spread spectrum signal. GPS, on the other hand, uses code-division multiple access (CDMA) to identify each individual satellite. FDMA provides better interference rejection for narrow-band interference signals compared to CDMA. In CDMA a single source of narrow-band interference source can disrupt all GPS satellite signals simultaneously, such interference only affects one FDMA GLONASS signal at a time. A shortcoming

of FDMA, however, is that it requires more spectrum than CDMA systems. GLONASS uses L1 in the range of 1602.0–1615.5 MHz and L2 in the range of 1246.0–1256.5 MHz to transmit C/A code and P code.

10.1. GLONASS RF frequency plan

The nominal values of L1 and L2 carrier frequencies are expressed as [22]

$$f_{k1} = f_{01} + K\Delta f_1 \tag{3}$$

$$f_{k2} = f_{02} + K\Delta f_2 \tag{4}$$

where K is frequency channel number of the signals transmitted by GLONASS satellites in the L1 and L2 sub-bands:

$$f_{01} = 1602 \text{ MHz}; \Delta f_1 = 562.5 \text{ kHz, for sub} - \text{band L1}$$

$$f_{02} = 1246 \text{ MHz}; \Delta f_2 = 437.5 \text{ kHz, for sub} - \text{band L2}$$

Each satellite has a standard nominal frequency, with a value of 5.0 MHz that generates the carrier frequencies L1 and L2. The system uses 12 channels to switch among its 24 operational satellites. Antipodal satellites in the same orbit plane are separated by an argument of latitude of 180°, as illustrated in **Figure 7** [26].

10.2. GLONASS signal structure

GLONASS satellites, too, transmit two PRN codes: a coarse acquisition (C/A) code and a precise (P) code. The C/A code is transmitted only on the L1 frequency, while the P code is transmitted on both L1 and L2 frequencies. GLONASS uses bi-phase modulation to merge the carrier signal with a modulo-2 summation of the PRN code at a rate of 511 kHz, the navigation message at a rate of 50 bps, and a 100 Hz auxiliary meander sequence [21].

GLONASS-K satellites also broadcast new CDMA signals in the L3-band at a carrier frequency of 1202.025 MHz [23]. The chipping rate for the ranging code is 10.23 Mcps, and it repeats every 1 ms. The new signal, however, uses a quadrature phase shift keying (QPSK) technique with an in-phase channel dedicated for data and a quadrature channel for pilot information. This signal spectrum is depicted in **Figure 8**.

10.3. Standard accuracy ranging code (C/A code)

The C/A code is a 511-bit binary sequence that is modulated onto the carrier frequency at a chipping rate of 0.511 MHz and thus repeats every millisecond [3]. It is derived from the
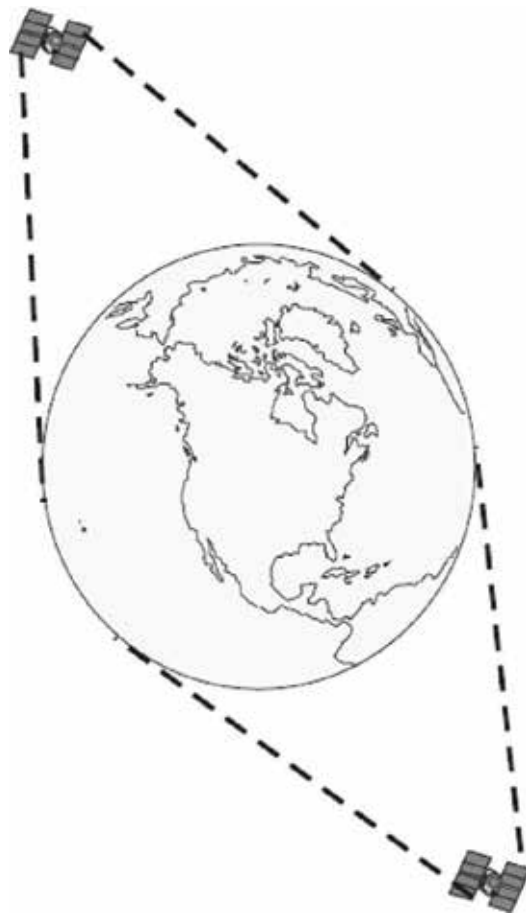
Figure 7. GLONASS antipodal satellites [3].

seventh bit of a nine-bit shift register. The code is described by the irreducible polynomial $1 + x^5 + x^7$. The initial state is defined as each bit containing the value "1" [22].

10.4. High accuracy ranging code (P code)

The GLONASS P code is a 5.11-million-bit-long binary sequence. It is modulated onto the carrier signal at a rate of 5.11 MHz and, hence, repeats every 1 s [3].

10.5. Intra-system interference

The intra-system interference in GLONASS is due to the intercorrelation properties of the ranging codes and the used FDMA technique [22]. The interference, indeed, happens inside the receiver between the signals transmitted on frequency channel K = n and signals transmitted on neighbor channels K = n + 1 and K = n − 1. In other words, this interference occurs when satellites with adjacent frequencies are visible at the same time.
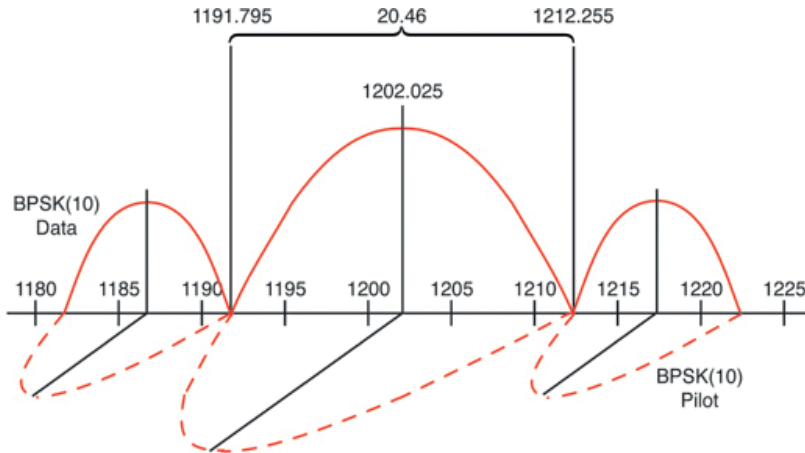
Figure 8. L3 CDMA signal spectrum [23].

10.6. GLONASS navigation message

The navigation message contains immediate and non-immediate data. It is broadcast from GLONASS satellites at a rate of 50 bps to provide users with necessary data for positioning, timing, and planning observations [22].

The immediate data contains information about GLONASS satellites. It is broadcast of a navigation signal which includes mainly the enumeration of the satellite time and the difference between the onboard time scale of the satellite and GLONASS time. The difference between the carrier frequency of the satellite signal and its nominal value is also included in this data along with ephemeris and other parameters.

The non-immediate data, on the other hand, contain information about almanac of the system. Almanac data provides information about the status of all satellites in the current constellation, coarse corrections of the onboard timescale for each satellite with respect to GLONASS time. Almanac data also have information about the orbital parameters of all satellites (orbit almanac) and correction to GLONASS time with respect to UTC (SU) and some other parameters [22].

11. Comparison between GPS and GLONASS

This section gives a brief comparison of GPS and GLONASS. It is essential to understand the similarities and differences between both GPS and GLONASS in particular when combining them in one navigation service or solution. Major differences between both systems are related to the constellation structure, the reference time system, the coordinates system, and the signal modulation or multiplexing technique. The following subsections briefly describe the GPS and GLONASS time and coordinate reference systems.

11.1. Time reference systems

Both GPS and GLONASS have their own time systems; thus, it is not straight forward to make time transformation from GLONASS time into GPS time or vice versa. The most important factor one must account for when processing data from a combined GPS and GLONASS is the difference between the two time scales.

11.2. GLONASS time system

As can be seen in **Table 2**, the daily satellite clock stability for GLONASS, GLONASS-M, and GLONASS-K is better than 5×10^{-13}, 1×10^{-13}, and 5×10^{-14}, respectively. The time shift between GLONASS time and the National Reference Time UTC (SU) is 3 h ICD (2008):

$$t_{GLONASS} = t_{UTC(SU)} + 03h \ 00mins \tag{5}$$

The following expression is used to align GLONASS satellite ephemeris at one instance with measurements given in UTC(SU):

$$t_{GLONASS} = t + \tau_c + \tau_n(t_b) - \gamma_n(t_b)(t - t_b) \tag{6}$$

where

t time of transmission of the navigation signal in the onboard time scale,

τ_c GLONASS time scale correction to UTC (SU) time,

t_b index of a time interval within current day,

$\tau_n(t_b)$ correction to nth satellite time relative to GLONASS time at time t_b,

$\gamma_n(t_b)$ relative deviation of the predicted carrier frequency value of n-satellite from nominal value at time t_b.

GLONASS-M satellites transmit the difference between the GPS and GLONASS time scale (which is never more than 30 ns) [22].

11.3. Time transformation

GLONASS time could be transformed into GPS time using the following formula [27]:

$$t_{GPS} = t_{GLONASS} + \tau_c + \tau_u + \tau_g \tag{7}$$

where

$$\tau_c = \tau_{UTC(SU)} - t_{GLONASS} \tag{8}$$

		GLONASS	GPS
Constellation	Number of satellite	24	32
	Number of orbits	3	6
	Orbital inclination	64.8°	55°
	Orbital radius	25,510 km	26,560 km
	Orbital altitude	19,130 km	20,200 km
	Orbit period	11 h 15.8 min	11 h 58 min
Signal characteristics	Multiplexing	FDMA	CDMA
	Carrier frequencies	1602 + k × 0.5625 MHz	1575.42 MHz
		1246 + k × 0.4375 MHz	1227.60 MHz
	Code frequencies	C/A code: 0.511	C/A code: 1.023
		P code: 5.11	P code: 10.23
	Broadcast ephemerides	Position, velocity, acceleration	Keplerian elements
Coordinates system		PZ-90.02	WGS-84
Time system		GLONASS time	GPS time

Table 3. Comparison between GPS and GLONASS.

$$\tau_u = t_{UTC} - t_{UTC(SU)} \tag{9}$$

$$\tau_g = t_{GPS} - t_{UTC} \tag{10}$$

In combined GPS/GLONASS data processing, the differences between these time scales must be accounted for. Otherwise, systematic errors are introduced that will affect the combined positioning solution.

Table 3 summarizes vital parameters of GPS and GLONASS that must be considered when combining GPS/GLONASS data processing.

12. Advantages of combined GPS and GLONASS

In many cases, such as navigating in urban or mountainous areas, during aircraft high-dynamic scenario, or under the effect of interference, satellite visibility becomes an issue. In such situations, incorporating both GPS and GLONASS constellations in the navigation system may significantly improve the accuracy of the navigational solution. Merging both systems in one navigation solution provides the next significant advantages:

- Increased satellite observability

- Remarkable increased spatial distribution of visible satellites

- Reduced horizontal and vertical dilution of precision (DOP) factors

On the other hand, the next considerations should be accounted for when combining GLONASS and GPS:

- The different aspects of the GLONASS and GPS navigation data

- The differences between the reference coordinate systems used in GLONASS and GPS

- The time scale offset between GLONASS and GPS

13. Conclusion

The demand for GNSS services and applications has been rapidly increasing. Luckily, we have more accessible GNSSs providing better functionality and broader coverage. Among these, GPS and GLONASS are fully functional at the time of writing. In this chapter, we gave a general overview of both systems, discussing the systems structure and signal characteristics, and provided an overview of the new features of GLONASS that intended to rectify the shortcomings of the GPS. The chapter is tailed with a short comparison between GLONASS and GPS highlighting advantages of combining both systems together.

Author details

Mohamed Tamazin[1], Malek Karaim[2]* and Aboelmagd Noureldin[2,3]

*Address all correspondence to: malek.karaim@queensu.ca

1 Electronics and Communications Engineering Department, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Egypt

2 Electrical and Computer Engineering Department, Queen's University, Canada

3 Electrical and Computer Engineering Department, Royal Military College of Canada, Canada

References

[1] Bradford P, James S. Global Positioning System: Theory and Applications, Volume I. The American Institute of Aeronautics and Astronautics, Inc: Washington; 1996

[2]   Pratap M, Per E. Global Positioning System: Signals, Measurements and Performance. 2nd ed. Massachusetts: Ganga-Jamuna Press; 2006

[3]   Elliott K, Christopher H. Understanding GPS: Principles and Applications. 2nd ed. Artech House: Norwood; 2006

[4]   Aboulmagd N, Tashfeen K, Jacques G. Fundamentals of Inertial Navigation, Satellite-Based Positioning and their Integration. Springer; 2013

[5]   Kai B, Dennis A, Nicolaj B, Peter Ri, Søren J. A Software-defined GPS and Galileo Receiver: A Single-Frequency Approach. Birkhauser Boston: Springer; 2007

[6]   Ahmed E. Introduction to GPS: The Global Positioning System. Artech House: Norwood; 2002

[7]   ICD G. Navstar GPS Space Segment/Navigation User Interfaces, Interface Specification. IS-GPS-200E El Segundo, CA, USA2010

[8]   Malek K. The Long Journey of a GPS Signal. Inside Signal Processing Newsletter. [Internet]. 2017. Available from: https://signalprocessingsociety.org/newsletter/2017/10/long-journey-gps-signal [Accessed: 2017-17-12]

[9]   Camacho-Lara S. Current and Future GNSS and their Augmentation Systems in Handbook of Satellite Applications. New York: Springer; 2013. pp. 617-654

[10] USA Force. Space Segment [Internet]. Available from: http://www.gps.gov/systems/gps/space/ [Accessed: 2017-01-12]

[11] USA Force. Control Segment [Internet]. Available from: http://www.gps.gov/systems/gps/space/ [Accessed: 2017-01-12]

[12] Dan D. Fundamentals of GPS Receivers: A Hardware Approach. Springer; 2012

[13] Hofmann-Wellenhof B, Herbert L, James C. GPS Theory and Practice. New York: Springer; 1993

[14] Jay F. Aided Navigation GPS with High Rate Sensors. New York: McGraw-Hill; 2008

[15] Malek K. Real-Time Cycle-Slip Detection and Correction for Land Vehicle Navigation Using Inertial Aiding [Thesis]. Kingston: Queen's University; 2013

[16] Mohamed T. High Resolution Signal Processing Techniques for Enhancing GPS Receiver Performance [Thesis]. Kingston: Queen's University; 2015

[17] Sheffer DK, Lee TH. A 1.5-V, 1.5-GHz CMOS low noise amplifier. IEEE Journal of Solid-State Circuits. 1997;**32**:745-759. DOI: 10.1109/4.568846

[18] Jonas T, Per-Ludvig N, Christian S. A High-performance real-time GNSS software receiver and its role in evaluating various commercial front end ASICs. In: Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2002); 24 September 2002; Portland: ION;2002. p. 2554-2560

[19] Braasch MS and Van Dierendonck AJ. GPS Receiver Architectures and Measurements. In: Proceedings of the IEEE; January 1999. p. 48-64

[20] Ralph E, Arthur W. Dual down Conversion GPS Receiver with Single Local Oscillator. Google Patents (US5108334 A), 1992

[21] Lachapelle G. GPS Theory and Applications. ENGO 625 Course Lecture Notes. Department of Geomatics Engineering, University of Calgary, 2009

[22] GLONASS ICD [Internet]. 2008. Available from: http://www.GLONASS-ianc.rsa.ru [Accessed: 2011-20-04]

[23] Yuri U, Valeriy S, Grigory S, Vyacheslav D, Alexander P, Sergey K. GLONASS developing strategies for the future. In: GPS World Magazine. April 2011. p. 42-49

[24] Keong J. Determining Heading and Pitch Using a Single Difference GPS/GLONASS Approach. UCGE Reports No. 20134. Department of Geomatics Engineering, University of Calgary, 1999

[25] GPS ICD [Internet]. 2010. Available from: http://www.gps.gov/technical/icwg [Accessed: 2011-09-04]

[26] Abbasiannik S. Multichannel Dual Frequency GLONASS Software Receiver in Combination with GPS L1 C/A. UCGE Reports No. 20215. Department of Geomatics Engineering, University of Calgary, 2009

[27] Cai C. Precise Point Positioning Using Dual-Frequency GPS and GLONASS Measurements. UCGE Reports No. 20291. Department of Geomatics Engineering, University of Calgary, 2009

Navigation System – Principles and Functions

Applications of GNSS Slant Path Delay Data on Meteorology at Storm Scales

Takuya Kawabata and Yoshinori Shoji

Additional information is available at the end of the chapter

Abstract

This chapter focuses on applications of Global Navigation Satellite Systems (GNSS) slant path delay data (SPD) to obtain signals from thunderstorms or rainbands. Current operational numerical weather prediction systems (NWPs) use water vapor distributions derived by GNSS technology as vital information for predicting convective rainfall. Mostly, zenith total delay or integrated water vapor data are used at horizontal scales of several tens of kilometers for this purpose. Beyond such operational use, SPD can be used to obtain information on storms (cumulonimbus) at horizontal scales of less than 10 km. For instance, found that SPD represents very small-scale phenomena of less than 10 km and can be used to estimate water vapor distribution around a thunderstorm with a strong tornado, and succeeded in improving the forecast skill of a rainband at 10 km scale. This chapter reviews SPD, which is invaluable for predicting thunderstorms and/or rainbands.

Keywords: slant path delay, storm scale, data assimilation

1. Introduction

The Earth's atmosphere, including the ionosphere, affects navigation signals transmitted by Global Navigation Satellite Systems (GNSS), which causes positioning errors. It is able to eliminate the ionospheric effect using a pair of GNSS carrier waves, and then GNSS analysis estimates the signal delay by the atmosphere as an unknown parameter [1–4]. The atmospheric delay is obtained by integrating the refractivity along the ray path, which is calculated with the variables of temperature, pressure, and water vapor pressure in the atmosphere. In the 1980s, several studies showed the feasibility of sensing the atmosphere using GNSS signal delay. For instance, Askne and Nordius [5] found a relationship between signal delay in the

zenith direction (zenith total delay; ZTD) and the precipitable water vapor (PWV; vertically integrated water vapor) in a formulation. Studies such as these led to the establishment of a new interdisciplinary research field (Bevis et al. [6], Businger et al. [7]). The Global Positioning System (GPS) maintained by the United States is the world's first operational GNSS, and has contributed significantly to the progress of the meteorological applications of GNSS. One of the most important contributions of GPS/GNSS meteorology is an innovative application of GNSS observations as a water vapor sensor.

Water vapor in the atmosphere plays an important role in convective rainfall systems. Particularly in the summer in Japan, convective clouds form frequently during the evening and develop into thunderstorms. Weather radar has been used in several observational studies to investigate the evolution of such convection. However, since weather radar only detects the distribution and movement of raindrops, it is still difficult to predict the precise position and time of the initiation of heavy precipitation. GNSS tropospheric delays are sensitive to water vapor, and can be used with radar to monitor the early stages of the convergence of water vapor before severe precipitation. Since the mid-1990s, procedures for retrieving precipitable water vapor (PWV) using the GNSS have advanced significantly.

In Japan, the Geospatial Information Authority of Japan (GSI) operates a nationwide permanent ground-based GNSS observation network called the Earth Observation Network (GEONET), which covers the entire Japanese archipelago with an average spacing of 17 km. It is regarded as one of the densest GNSS networks in the world. On July 13, 2012, the GSI began to provide data from the Russian GLONASS (Globalnaya Navigatsionnaya Sputnikovaya Sistema) and Japan's QZSS (Quasi Zenith Satellite System) along with GPS data. Also, from March 2016, the GSI started providing GALILEO data.

Several studies have been shown GNSS PWV as a useful tool for monitoring heavy rainfall. Kanda et al. [8] found that periods of maximum PWV tended to be a precursor of the initiation of heavy rain, and that larger hourly increases in PWV led to higher frequency of precipitation.

Niimura et al. [9] statistically investigated the relationship between precipitation, PWV, and temperature, and found the correlation between the observed PWV and surface temperature. They also found a threshold of PWV depending on the temperature; for instance, when PWV exceeds the threshold, frequency of precipitation rapidly increases. After their statistical study, it was concluded that this threshold is connected with the humidity averaged vertically above the observation point, and that the saturation level of the entire atmosphere can be estimated with PWV. Inoue and Inoue [10] also statistically studied the two-dimensional PWV field derived by GNSS in connection with thunderstorms in summer. They found that cloud-to-ground (CG) strokes were observed 1 hour after large amounts of PWV and its 30-min increments.

These studies have illustrated that observing PWV and its variation are useful for monitoring heavy rainfall. However, monitoring PWV alone is not always enough to capture information on the precursor of a severe storm. For instance, Niimura et al. [9] illustrated a good agreement between the level of saturation of the entire atmosphere given by PWV and stratiform rain, but the relationship was unclear in localized heavy rain cases. Inoue and Inoue [10] statistically showed that the maximum CG stroke only in 40% of the observed thunderstorms was observed 15–30 min after the maximum PWV. Therefore, it is necessary to find other method for the remaining 60%.

Since the development of the GPS/GNSS meteorology, the applications of slant path delays (SPDs; delays along paths from GNSS satellites to GNSS receivers) have been studied for water vapor variation in local and its role in the development of hazardous thunderstorms. One such approach is the tomographic method, which estimates structures of water vapor in 3D. For this purpose, several GPS/GNSS observational campaigns with dense networks have been conducted, and obtained successful results [11–13]. However, it is difficult for operational GNSS networks to retrieve 3D structures of water vapor using only GNSS SPDs, due to the low density of the network, small number of SPDs, and slow movement of GNSS satellites.

Another application is assimilation of GNSS-derived water vapor information for improving the initial fields of numerical weather prediction (NWP) models. The impact of this information on mesoscale NWP systems has been thoroughly investigated (Nakamura et al. [14], Koizumi and Sato [15], Seko et al. [16]). These studies mainly used PWV data retrieved by GNSS analysis using IGS's precise final ephemerides. Shoji [17] developed a near real time (NRT) PWV analysis system. This system provides the PWV data by the time when the Japan Meteorological Agency (JMA) assimilates the data in the operational mesoscale data assimilation (DA) system. Using PWV by the NRT system, Shoji et al. [18] succeeded to predict heavy rainfalls, which JMA's mesoscale NWP model failed operationally. Further, an impact assessment of the local-scale atmospheric phenomena on GNSS positioning was also carried out using mesoscale NWP outputs (Seko et al. [13], Shimada et al. [19], Ishikawa et al. [20]).

Other studies have investigated the assimilation of SPD data. The advantage of SPD for PWV and ZTD is that it includes information about several atmospheric parameters (pressure, temperature, and humidity) in several directions from each receiver. Therefore, the assimilation of SPD data is expected to improve the water vapor field with the thermodynamics field of the model above and around the observation points.

Ha et al. [21] assimilated virtual GNSS slant water vapor observations, which are accumulated water vapor along slant paths (SW data), as an Observation System Simulation Experiment (OSSE). The SW data is advantageous to PWV observations in its various directions from each receiver. They succeeded to simulate a realistic squall line after assimilation of the SW data with a fifth-generation mesoscale model (MM5)-4D-Var system (Zou and Kuo, [22]) at a 27-km grid spacing. Järvinen et al. [23] assimilated SPD data with a High Resolution Local Area Modeling-3D-Var system (HIRLAM; Gustafsson et al. [24]) at a 9-km grid spacing, and showed that the analysis increments of the SPD assimilation were larger that of the ZTD assimilation, and that the horizontal distribution of the SPD analysis increments was different from that of ZTD assimilations. Bauer et al. [25] tried to assimilate SPD data with the MM5-4D-Var system and succeeded to improve quantitative precipitation forecasting skills. Though they demonstrated the superior QPF scores by assimilating SPD data over a month, their horizontal grid spacing was 18 km; this is not enough to represent thunderstorms directly in the model. In the past few years, the German Weather Service (DWD) has developed an assimilation method for SPD data with their operational data assimilation system (Kilometer-Scale Ensemble Kalman Filter; KENDA) and obtained a 20% improvement on prediction of a convective rainfall system (Potthast 2017; personal communication).

The SPD data represent both vertical and horizontal atmospheric conditions, whereas the ZTD and PWV observations contain only vertical information. Thus, it is advantageous to assimilate

the SPD observations at storm scales. For instance, a GNSS signal with a 30° elevation angle at a GNSS receiver travels from the top of the troposphere to the receiver with a horizontal distance of 17 km. Thus, the SPD data cover only two model grid cells, when an assimilation system with 20-km grid spacing is applied. As a result, it is expected that the assimilation effect of SPD data would be similar to the ZTD assimilation. Therefore, it is important for SPD assimilations to use assimilation systems with high grid spacings, hopefully less than 5 km.

In this chapter, the use of SPD data for the monitoring of hazardous convection is described in Section 2, and the first assimilation of SPD data with a DA system with 2-km grid spacing (storm scale) is given in Section 3. The chapter is summarized in Section 4.

2. Using SPD to monitor hazardous convection

GNSS-derived PWV is, so to speak, a representative value of PWV within an inverted cone-shaped space above each GNSS antenna. GNSS-derived PWV is inherently unsuitable for monitoring water vapor variation under severe storms where several-kilometer-scale phenomena prevail. Is SPD suitable in this case? In this section, we first divide SPD into three components based on an SPD model in GNSS analysis, and discuss the horizontal scale of each component. Next, we propose new indices that represent finer water vapor distribution than the GNSS PWV. Finally, we introduce a procedure to analyze a several-kilometer-scale PWV distribution around each GNSS antenna using SPD.

Following MacMillan [26], the SPD between a GNSS satellite and a receiver at an elevation angle and a direction angle measured clockwise from north can be written in the following form:

$$\mathrm{SPD}(\theta, \phi) \, = \, m(\theta)\big[\mathrm{ZTD} + \cot\theta\big(G_n\cos\phi + G_e\sin\phi\big)\big] + \varepsilon, \tag{1}$$

where ZTD, $m(\theta)$, G_n (G_e) and ε are the total delay in the zenith direction, the isotropic mapping function that describes the ratio of SPD to ZTD, the delay gradient parameters in the north (east) directions, and the postfit phase residual, respectively. ZTD is vertically integrated refractivity (N) of the atmosphere in the zenith direction. Refractivity is expressed by temperature (T), partial dry air pressure (P_d), and partial water vapor pressure (P_w):

$$N \, = \, (n-1)\,10^6 \, = \, K_1\frac{P_d}{T} + K_2\frac{P_w}{T} + K_3\frac{P_w}{T^2}, \tag{2}$$

where n is refractive index, and K_1, K_2, and K_3 are constants that have been determined theoretically or by fitting observed atmospheric data. Several studies have evaluated GPS-derived gradient parameters by comparison with those observed by water vapor radiometers (WVR), and found good agreement (e.g., [27–29]).

The gradient parameters represent the horizontal first-order gradients of water vapor. The postfit residuals are expected to contain information on higher order water vapor inhomogeneity (HI). However, other errors that do not originate from the atmosphere are also included (e.g., antenna phase center variation (PCV), signal scattering, multipath, errors in satellite

orbit, and errors in clocks of both satellite and receiver). Therefore, to estimate HI, it is necessary to remove all errors not related to atmospheric inhomogeneity. Shoji et al. [1] performed a procedure to eliminate multi-path and satellite clock error-induced residuals to reconstruct HI components. The correlation coefficient of each component retrieved at a different GNSS station and sorted by distance demonstrated that the horizontal scale of the ZTD can be considered as 644 ± 120 km, the gradient parameter G_n, G_e as 62 ± 23 km, and the HI as 2–3 km. This result suggests that ZTD, G, and HI relate to atmospheric motion of the meso- α, meso- β, and meso- γ scales, respectively (**Figure 1**).

Sato et al. [30] compared zenith-scaled SPD using a mapping function and that retrieved from radiosonde observations. They found that the zenith-scaled SPD, in which the path is closest to a radiosonde path, exhibited better agreement than zenith total delay (ZTD) retrieved by standard GNSS analysis (i.e., a representative value of the inverted-cone-shaped space above the GNSS antenna).

Although it requires some careful effort to retrieve, GNSS SPD possesses information on local-scale atmospheric activity. Shoji [31] proposed procedures for retrieving two indices indicating the degree of inhomogeneity of water vapor using the GNSS SPDs. One index (WVC) describes the spatial concentration of water vapor (Eq. 3), whereas the other (WVI) indicates higher order water vapor inhomogeneity (Eq. 4). The horizontal scales of the two indices are considered to be approximately 60 km and 2–3 km, respectively.

$$WVC = -\nabla^2 PWV_G, \tag{3}$$

where ∇PWV_G is the horizontal gradient of PWV estimated from the atmospheric gradient parameter (G).

$$WVI = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(HI_{PWV}^i - \overline{HI_{PWV}} \right)^2}, \tag{4}$$
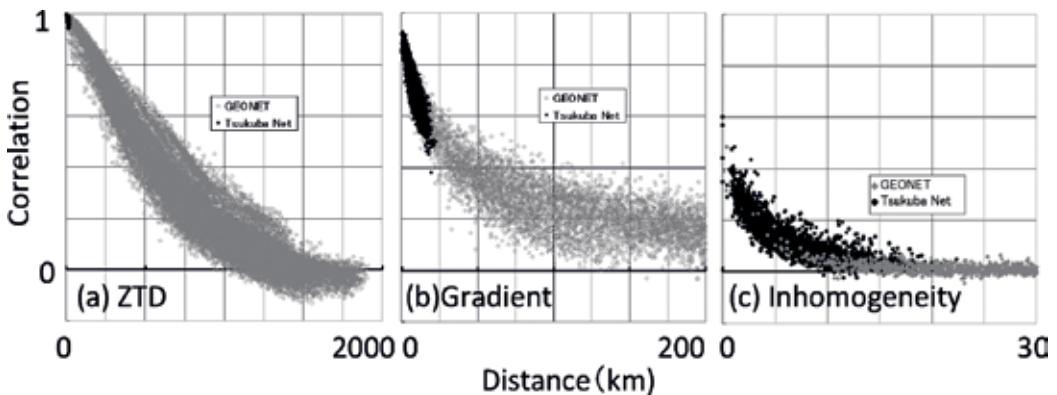

Figure 1. Correlation coefficients of (a) ZTD, (b) gradient component, and (c) postfit residuals as a function of distance. Gray-filled circles are from GEONET, and black-filled circles are from the Tsukuba GPS dense net campaign. Each data point is based on 51 days of data from July 14 to September 2, 2001. (Modified from **Figures 9–11** of Shoji et al. [1]).

where HI_{PWV} is the inhomogeneity component of SWV normalized as a vertical value using a mapping function. To extract HI_{PWV} from the postfit phase residual (ε), it is essential to eliminate the effects of GNSS antenna phase center variation (PCV), multipath effects, and errors in satellite orbits and clocks.

The statistical examination between these indices and precipitation (**Figure 2**) illustrate that the inhomogeneity indices show stronger correlation with rainfall amount than with PWV. It seems that PWV related to precipitation of less than 10 mm h^{-1}, but was not connected to precipitation greater than 10 mm h^{-1}. It is also true for both present and imminent precipitation. Furthermore, the spatiotemporal variations of the indices were also examined on a particular thunderstorm on August 11, 2011. Both the WVC and WVI indices increased ahead of the convective initiation (**Figure 3**). The WVI index is based on the standard deviation of the SPDs measured at a GNSS station, so the directional information for each SPD is neglected.

Shoji et al. [2] proposed a new method for estimating PWV distribution around each ground-based GNSS station on a scale of several kilometers (**Figure 4**).
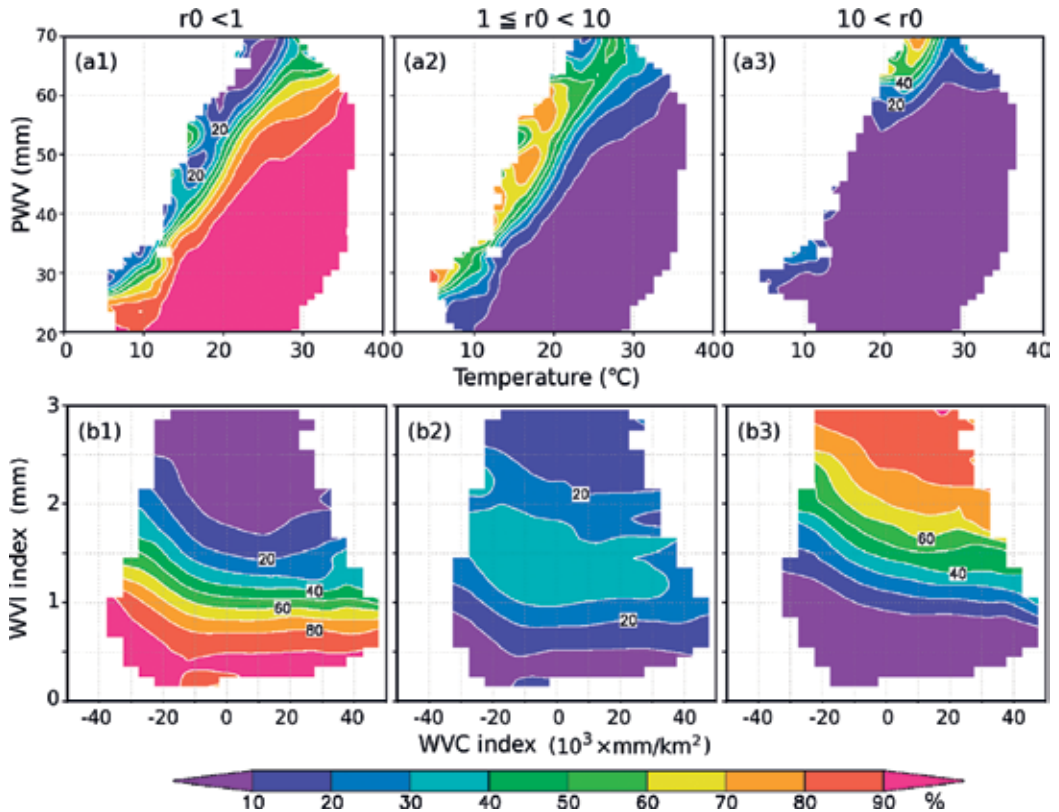

Figure 2. Frequency of precipitation in the last hour against PWV, WVC, and WVI index. (a) Frequency of 1-h precipitation (FD1P) against temperature (abscissa) and PWV (ordinate). Precipitation is less than 1 mm (a1), between 1 and 10 mm (a2), and more than 10 mm (a3). (b) FD1P against WVC (abscissa) and WVI indexes (ordinate). Modified from **Figure 10** of Shoji 2013 [31].
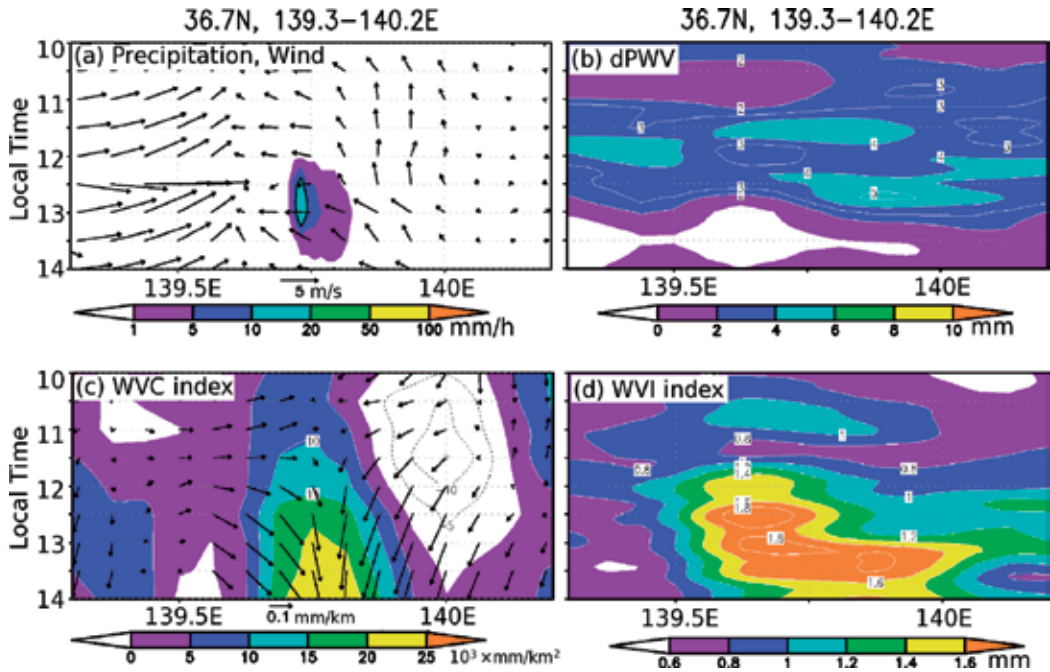
Figure 3. Longitude–time cross sections on August 11, 2011. (a) 1-h accumulated precipitation, (b) PWV deviation from 1-month average, (c) WVC index, and (d) WVI index. Modified from **Figure 14** of Shoji 2013 [31].
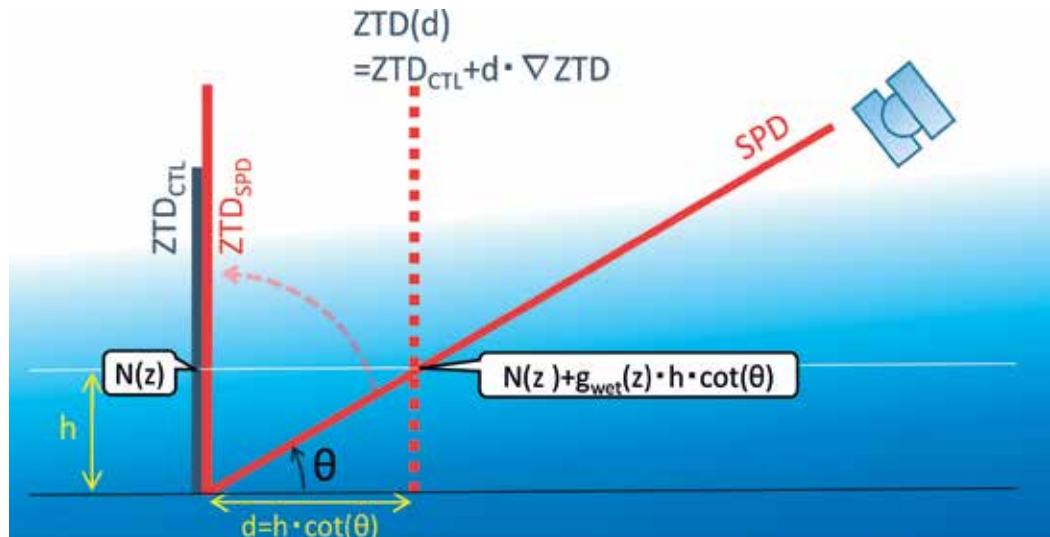

Figure 4. Schematic illustrating the difference in integrated refractivity (N) at height (Z) between the zenith direction and signal direction. The difference between ZTD_{SPD} and ZTD_{EST} is assumed to be caused by the horizontal gradient of wet refractivity. Shoji et al. 2014 [2].

In this approach, the following three assumptions were set.

1. The horizontal gradient of dry refractivity (N_{dry}) is small enough to be negligible.

2. The difference between ZTD_{EST} (retrieved value through standard GNSS analysis) and ZTD_{SPD} (normalized SPD into the zenith direction using a mapping function) is caused by the several-kilometer-scale horizontal gradient of water vapor refractivity (N_{wet}) alone.

3. The horizontal N_{wet} gradient (g_{wet}) decreases exponentially with height.

$$g_{wet}(Z) = g_{wet}(0) \exp\left(-\tfrac{z}{H}\right), \tag{5}$$

where $g_{wet}(z)$ is the horizontal water-vapor gradient at altitude Z, and H is the scale height of g_{wet}.

Under these assumptions, the horizontal PWV gradient can be expressed as the following equation.

$$\nabla PWV = \Pi \, \nabla ZTD \approx \Pi \, \nabla ZWD = \Pi \frac{ZTD_{SPD} - ZTD_{EST}}{H \cot(\theta)}, \tag{6}$$

where Π is the proportionality coefficient. The horizontal gradient of PWV is expressed as a function of the ZTD difference, elevation angle, and scale height.

For practical monitoring of severe convection, we must carefully assess the accuracy of PWV distribution using the new method. Shoji et al. [3] quantitatively evaluated the method of Shoji et al. [2] using the simulation results of a high-resolution NWP model performed by Mashiko [32] for a tornadic supercell case, which generated an F3 tornado.

Figure 5 plots the evaluation results at a particular moment in time (1208 JST, May 6, 2012). The thick gray lines represent the root mean square (RMS) difference of PWV_{EST} against the true PWV field around each virtual GNSS station. PWV_{EST} is the GNSS-derived PWV (not equal to the true PWV at virtual GNSS stations). In the figure, the RMS difference of PWV_{EST} is 0.5 mm at the GNSS site. The RMS increased with distance and the value reached 2.5 mm at a distance of 3 km, when PWV_{EST} was extrapolated as the PWV value near a virtual GNSS station. The thin colored lines in **Figure 5**, which illustrate distances at where the RMS difference was the smallest, show the distance increases as the elevation angle decreases. The RMS was the smallest within 1 km range from a virtual GNSS station for PWV_{SPD} with a 77.6° elevation angle. In case of PWV_{SPD} with a 17.6° elevation angle, RMS was the smallest of 1.5 mm at a 4.5 km distance. The distance and elevation dependency of minimum RMSE are illustrated by the red dots (**Figure 5**). Overall, the conventional procedure causes about RMS of 0.5 mm at the GNSS site, and the error by simple extrapolation increases with distance, reaching 1.5 mm at 1 km. The distance dependency of PWV errors in PWV_{SPD} differs in each elevation angle. From this result, we can estimate PWV with RMSE of less than 2 mm within 6 km from a GNSS station by using SPD with an elevation angle over 15°. Essentially, with PWV_{SPD}, it is able to estimate the PWV distribution around each GNSS station with better than half the RMSE of that obtained by the conventional GNSS PWV retrieval method (PWV_{EST}).
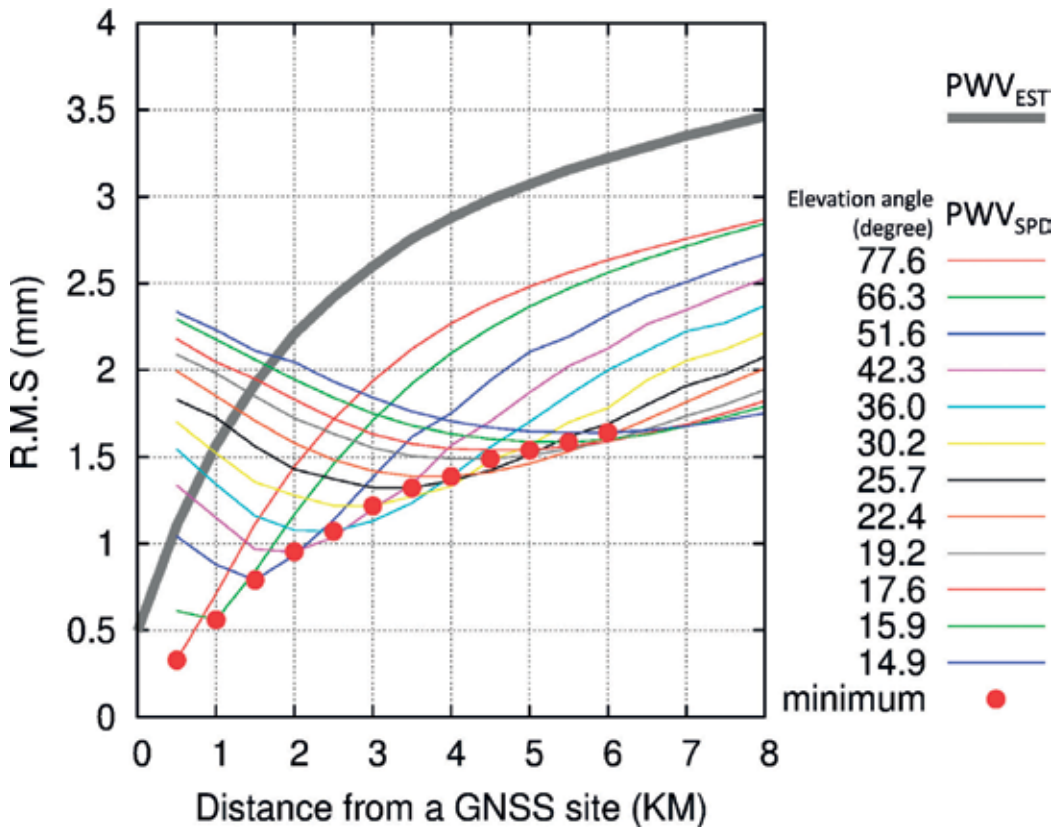
Figure 5. Variation in RMS difference as a function of distance from virtual GNSS stations. The gray thick line is PWVEST, the colored thin lines are PWVSPD at various elevation angles, and the red dots are the minimum RMS difference at each elevation angle. The number of samples is different according to satellite elevation, which ranges from 326,836 to 1,307,344. (Modified from **Figure 3** of Shoji et al. [3]).
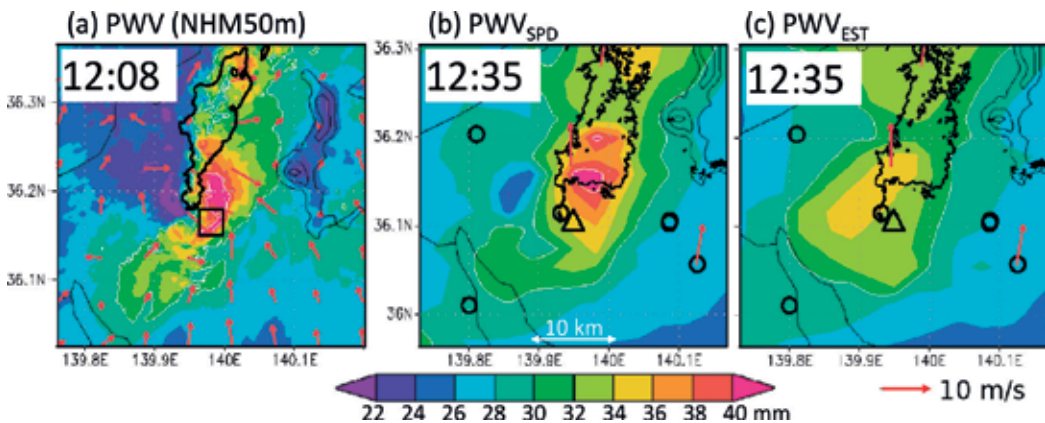

Figure 6. PWV fields around a tornadic supercell. (a) NHM50m. (b) Both PWV$_{EST}$ and PWV$_{SPD}$. (c) Only PWV$_{EST}$. Black lines in (a) are the rainfall intensity of 40 mm h^{-1} reproduced by NHM50m; those in (b) and (c) are the reflectivity intensity of 40dBZ at the 0.5° elevation angle of the MRI C-band radar. Although the timing differs by about 30 minutes, the NHM simulation succeeded in reproducing the movement and development of the parent storm. (Modified from **Figure 7** of Shoji et al. [2]).

The new method demonstrated the capability to capture a strong PWV gradient associated with the parent storm of the F3 tornado that struck Tsukuba City in Ibaraki Prefecture, Japan, on May 6, 2012, with a numerical model simulation at a super-high resolution of 50 m (NHM50m; **Figure 6**). An area of large PWV contrast centered on strong precipitation (**Figure 6a**) implies a strong upward wind in front of and a strong downdraft behind the parent storm. The PWV contrast toward the tornado is also seen in **Figure 6b**, whereas, no such PWV gradient is seen at all in **Figure 6c**.

3. Data assimilation of SPD at storm scale

3.1. Data assimilation system

Data assimilation (DA) techniques optimize differences between first guess fields provided by numerical models and observations. One advanced DA theory is a variational method (Sasaki [33, 34]), and the NHM-4DVAR is a variational DA system for predicting thunderstorms (Kawabata et al. [35, 36]). The NHM-4DVAR is based on the Japan Meteorological Agency Nonhydrostatic Model (JMANHM; Saito et al. [37]), which includes three-ice bulk cloud microphysics, as the forward model. In the first version of NHM-4DVAR, the adjoint model considered only dry dynamics and advection of water vapor. Kawabata et al. [36] implemented an additional warm rain process for assimilating radar reflectivity. The horizontal resolution of NHM-4DVAR is 2 km. This 2-km spacing is known as "storm scale" because in this range it is possible to explicitly represent cumulonimbus and thunderstorms by numerical models. The control variables are the three wind components, potential temperature, surface pressure, nonhydrostatic pressure, total water (water vapor and cloud water), the relative mixing ratio of rainwater, and the pseudo-relative humidity (only for lateral boundary conditions).

This method optimizes the difference between observations and numerical models using their error statistics. To measure the difference, we need observation operators to convert the numerical model fields to observation spaces. For SPD DA, Eq. (2) is used as the operator and provides delays of GNSS radio waves from GNSS antennas to the model top boundary as

$$\Delta L = \int_{Model\_surface}^{Model\_top} (n-1)ds. \tag{7}$$

Here, ΔL denotes the atmospheric delay (m), n denotes the refractive index along the path in a grid cell, and ds is the path length (m) in each model grid cell. Because the top boundary of numerical models is limited to a certain height level (usually 20–40 km), we need an assumption to add a delay from the top boundary to a GNSS satellite.

The authors assumed that the delay decreased to 1/e every 10 km from the top boundary to 200 km height (Bean and Thayer [38]) and that above 200 km the amount of delay was zero. In addition to this assumption, a straight-line assumption was adopted; the bending effect was thus eliminated. To account for these assumptions, a relatively large observational error was implemented. To calculate SPD in the model, first, the linear path is determined from the receiver to the GNSS satellite. Then, the middle point of the path within the model grid cells are set down. Delays are interpolated and averaged from eight points surrounding the middle point,

with each weight set according to the distance. Finally, the slant path delays are calculated by integrating each delay in model grid cells from the receiver to the top boundary of the model.

The observations over elevation angles of 5° or more were assimilated in their study. The operator works on the World Geodetic System 1984 (WGS84; National Imagery and Mapping Agency 1997) [39].

3.2. Assimilation experiment

3.2.1. Single set of SPD observations at a single site

First, three experiments were performed using NHM-4DVAR with a 2-km horizontal grid spacing in which three SPD observations (SO_SPD), one ZTD observation (SO_ZTD), or one PWV observation (SO_PWV) from a single observation site were assimilated. Note that the ZTD and PWV observations were derived from the SPD observations originally. These experiments were conducted to confirm the effects of SPD assimilation on a single analysis and to examine the differences between SPD, ZTD, and PWV assimilations. The assimilation window was set at 10 min, and the observations were assimilated every 5 min (at 0, 5, and 10 min in the assimilation window). The observational data set was chosen by considering the horizontal distributions, elevation angles, and the first guess field from an experimental data set.

The propagation paths of GNSS signals from three satellites to a receiver in the model atmosphere in both the horizontal (**Figure 7a**) and the vertical plane (projected from the south; **Figure 7b**) are illustrated. Path I with the smallest elevation angle is also the longest, while path II at a near 90° elevation angle is the shortest. The large amount of delays is illustrated mostly in low altitudes area by warm colors.

The distributions of the increments (analysis minus first guess) of PWV in SO_SPD (**Figure 8a**) are different from that in SO_ZTD (**Figure 8b**) at the end of the assimilation window. In SO_ZTD, the increment distribution (**Figure 8b**) is axisymmetric and elliptical mostly (i.e., isotropic). Although 4D-Var captures analysis increments shaped flow-dependent (anisotropic) in general, it is not
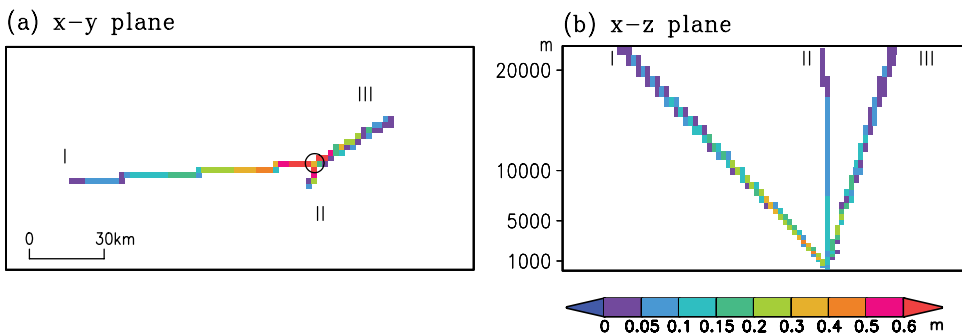

Figure 7. Propagation paths of GNSS signals from satellites (I, II, and III) to a receiver in the simulated atmosphere, (a) viewed in the horizontal plane and (b) the vertical plane. Colors illustrate the values of delay in each model grid cell (each cell is shown by one pixel). The open circle in (a) displays the GNSS receiver (the observational site). After Kawabata et al. [4].
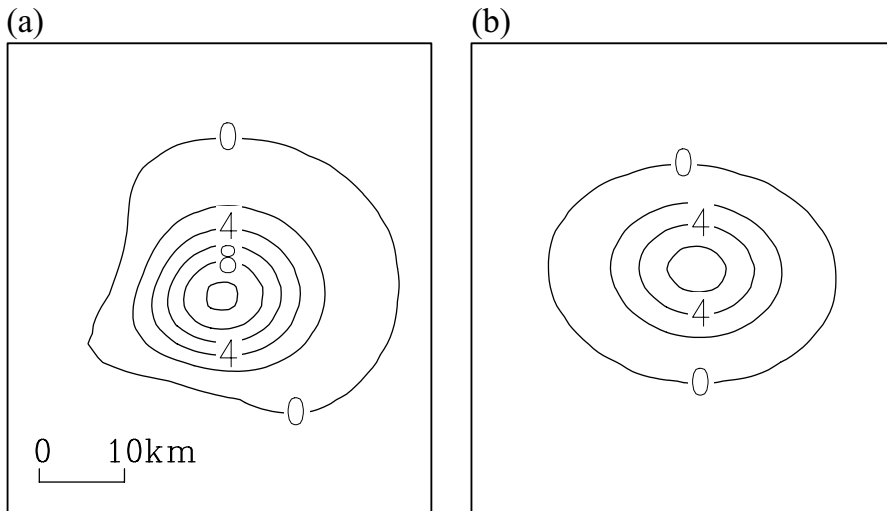
Figure 8. Analysis increments of precipitable water vapor (mm) at the end of the assimilation window for (a) SO_STD and (b) SO_ZTD. After Kawabata et al. [4].

long enough for the 10-min assimilation window to capture such flow dependency. As a result, the shape of the analysis increment seen in SO_ZTD is isotropic mostly. In contrast, the inhomogeneous distribution of the analysis increment in SO_SPD (**Figure 8a**) is given by the various distribution of the slant paths. Moreover, the maximum value of the increments is much larger in SO_SPD (10 mm; **Figure 8a**) than that in SO_ZTD (6 mm, **Figure 8b**). The increment distributions in SO_PWV and SO_ZTD were close to each other (not shown).

Seeing vertical cross sections of the mixing ratio of water vapor (Qv) along path III (**Figure 9**), the distributions of analysis increment in SO_ZTD and SO_PWV are close, and the magnitudes of the increment are quite similar: The distributions of Qv reach at high altitudes from 1 to 5 km in vertical and extend to 5–8 km away in horizontal. The increment along path III in SO_SPD is distributed along a distance of 15 km in horizontal and over 8 km in height; furthermore, the magnitudes of the increment are much larger than others, at low altitude especially (around 3 km), because all of the slant paths are within a narrow area of the lower troposphere above the observation site. The total weights in the cost function show good agreement among the assimilations of SO_SPD, SO_ZTD, and SO_PWV, but their effects are seen at different places.

3.2.2. Actual observational data assimilation

Next, an actual **observation assimilation** and forecast experiments were conducted using NHM-4DVAR with the assimilation window of 30 min. Three-hour forecasts were performed after the 4D-Var analysis with multiple observations. The assimilation was started at 1100 JST on August 19, and the forecast was performed from 1100 to 1400 JST. The SPD, ZTD, or PWV data were assimilated every 10 min (hereafter, A_SPD, A_ZTD, and A_PWV, respectively). Only the observations listed above, and no other kinds, were assimilated. No cumulus parameterization was applied in the experiments. The experimental domain and the distribution of GNSS receivers are illustrated in **Figure 10**. Unlike the PWV and ZTD data, it can be said that SPDs in the model atmosphere (**Figure 10a**) contain a great deal of horizontal information.
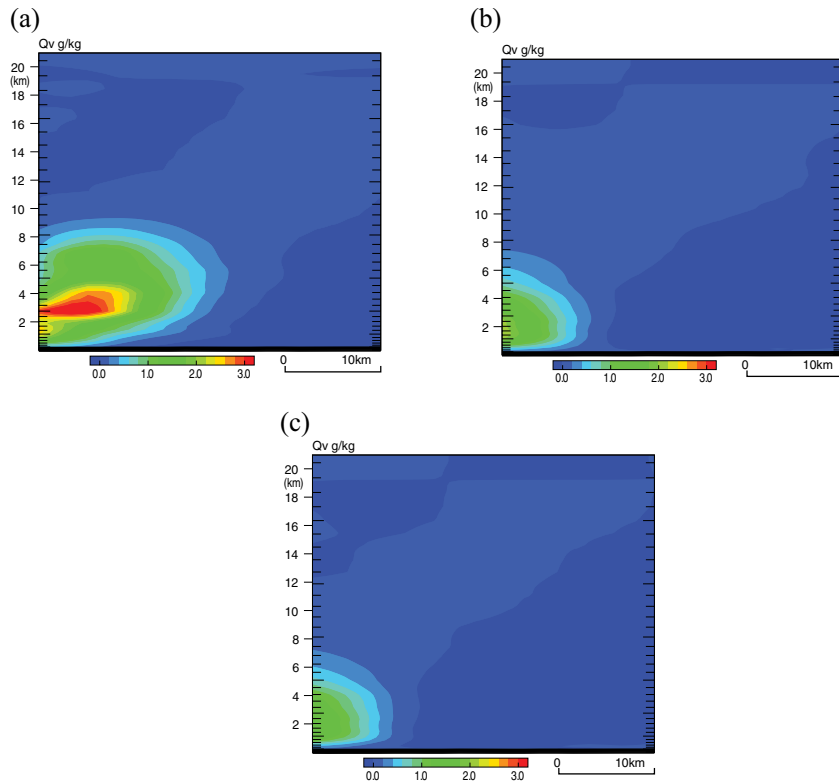
Figure 9. Vertical cross sections along path III. Analysis increments of the mixing ratio of water vapor (g kg⁻¹). (a) SO_SPD, (b) SO_ZTD and (c) SO_PWV. The observation site is at the lower left corner of each panel, and path III leaves the model top at the upper right corner. Modified Figure 11 of Kawabata et al. [4].
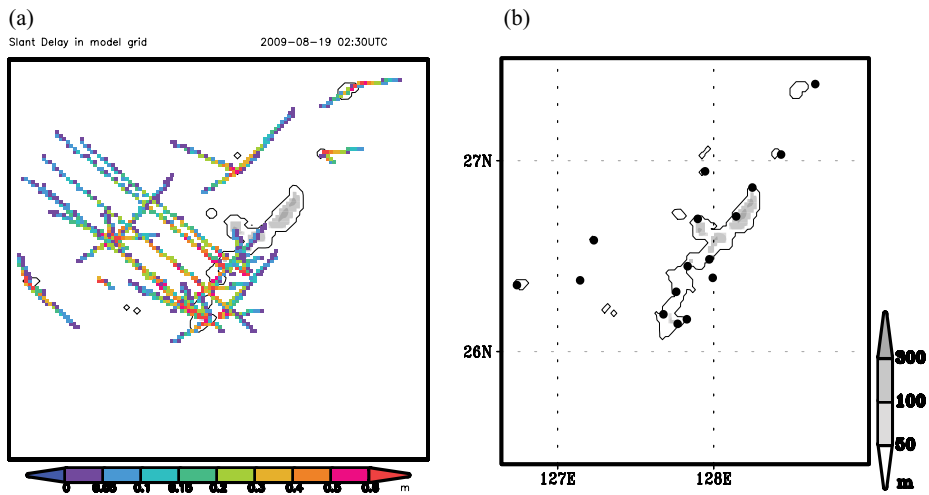
Figure 10. (a) Propagation paths of radio waves from GNSS satellites to receivers in the model atmosphere, viewed in the horizontal plane. Actual SPD observations obtained over Okinawa Island at 1130 JST. (b) the assimilation and forecast domain, topography and GNSS stations. After Kawabata et al. [4].

The horizontal distributions of rainfall in A_ZTD and A_PWV (**Figure 11c** and **d**) were close to each other. In A_SPD (**Figure 11b**), the initiation of intense rainfall enhanced over southern Okinawa at 1200 JST, and the maximum intensity reached 47 mm h^{-1} (not shown). By 1300 JST, a rainband had formed over the island, and the distribution and intensity resembled the radar observation (**Figure 11a**). Therefore, it can be concluded that the assimilation of SPD data showed improvement of the rainfall forecast with respect to both timing and intensity compared with the assimilation of PWV and ZTD data. This improvement was obtained through modifications in the atmospheric profile in the simulation after the SPD assimilation (not shown).
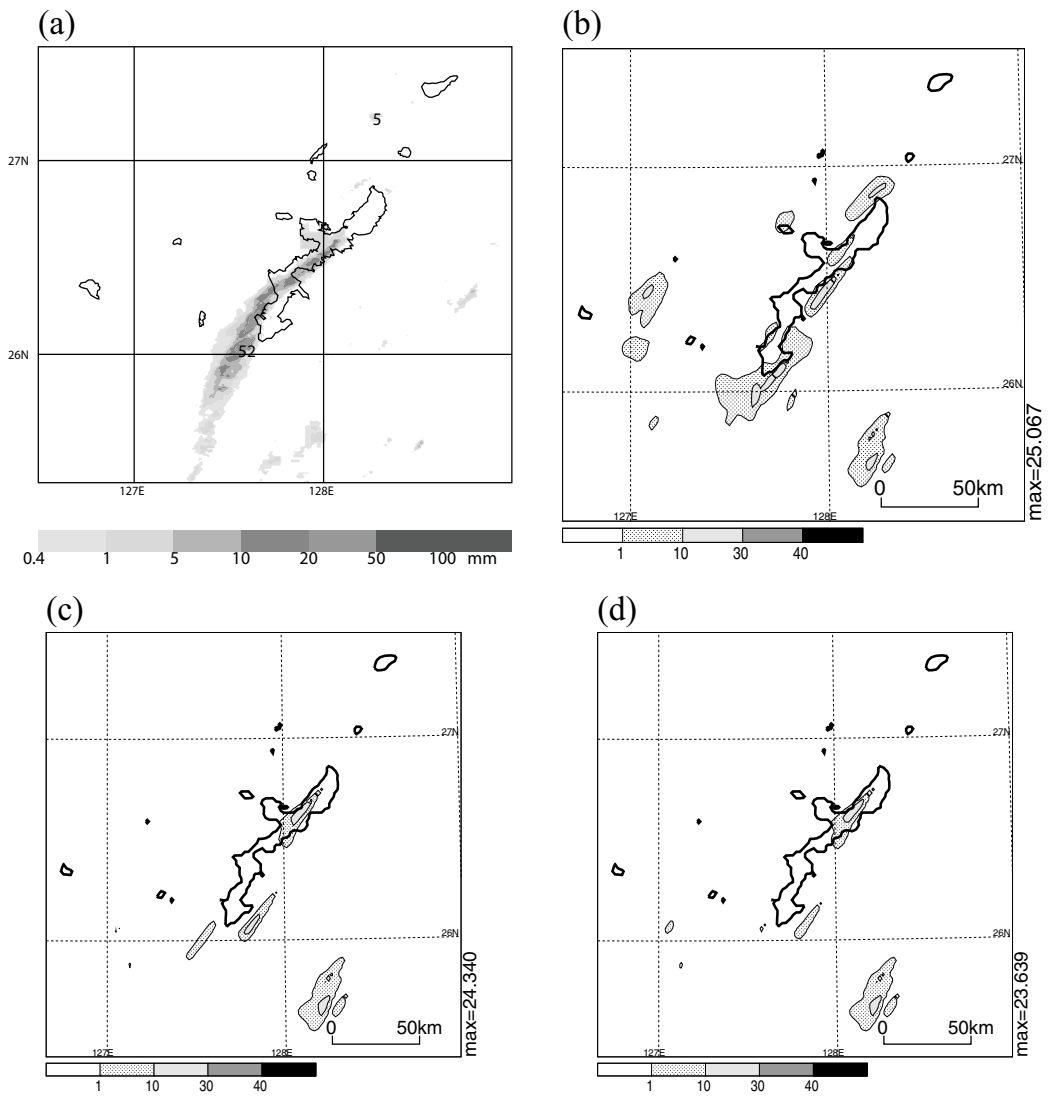
Figure 11. Horizontal distribution of 1-h accumulated rainfall amount between 1300 and 1400 (FT = 02) JST. (a) Radar observation, (b) A_SPD, (c) A_ZTD and (d) A_PWV. Modified Figures 2 and 15 of Kawabata et al. [4].

4. Summary

GNSS analysis is performed with a number of models for solid earth tide, ocean tide, earth rotation, and atmospheric delay. However, conventional GNSS analysis inevitably has errors caused by local-scale variations in the atmosphere, even if estimates of the gradient parameters are applied. These errors affect positioning as well as atmospheric delays. Spatial correlation in screened post-fit phase residuals (i.e., the higher order inhomogeneity part) must be caused by local-scale variations in the atmosphere, which are not resolved by ordinary ZTD and gradient parameter analysis. Thus, GNSS postfit residuals can be used to detect the inhomogeneous distribution of water vapor on a local scale of less than 10 km caused by weather disturbances like convective thunderstorms. In this chapter, we introduced two new indices and one experimental method to retrieve a several-kilometer-scale PWV distribution. These approaches may be a preliminary way to serve operational monitoring of cumulus convection.

Currently, multiple GNSSs (e.g., the Russian GLONASS, EU's GALILEO, China's BeiDou, and Japan's Quasi-Zenith Satellite System (QZSS)) can serve to more precisely resolve local-scale water vapor variation. In addition, the advancement of real-time GNSS analysis has been progressing rapidly. Real-time orbit and clock corrections have offered officially by the International GNSS Service (IGS) since April 2013. Moreover, a Multi-GNSS orbit and clock estimator called MADOCA (Multi-GNSS Advanced Demonstration tool for Orbit and Clock Analysis) has been developed by the Japan Aerospace Exploration Agency (JAXA) and they started to provide real-time ephemerides via the Internet (https://ssl.tksc.jaxa.jp/madoca/public/public_index_en.html) in September 2015. These rapid advancements in GNSS technology should speed the reality of new usage of GNSS-derived SPD.

Another application of SPD data at a local scale is data assimilation (DA). As shown in this chapter, a promising avenue for DA applications is to improve the initial conditions of a numerical weather model in high-resolution simulations. NHM-4DVAR with a 2-km horizontal grid spacing was used for assimilating GNSS slant path delay data. Compared with simulations after assimilating PWV and ZTD data, the assimilation of SPD showed inhomogeneous increments. Moreover, the analysis increments in the assimilation of SPD at low altitudes were larger than in the others. Conducting actual observation assimilations, the SPD assimilation clearly improved the forecast of both the timing and intensity of the rainband. This improved forecast was given through the decreased atmospheric stability over Okinawa Island.

The SPD data include information on temperature, dry atmospheric pressure, and water vapor pressure. In addition, the data contain both horizontal and vertical information on those atmospheric parameters. Because atmospheric inhomogeneity is greatest in the lower troposphere, assimilating SPD data is a promising way to improve forecast of a rainband. A GNSS signal along a path with a 15° elevation angle propagates about 11 km in horizontal and travels 3 km in vertical. Therefore, it is concluded that data assimilation systems at storm scales like used in this study are necessary to assimilate SPD data for taking advantages.

Author details

Takuya Kawabata[1]* and Yoshinori Shoji[2]

*Address all correspondence to: tkawabat@mri-jma.go.jp

1 Forecast Research Department, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

2 Meteorological Satellite and Observation System Research Department, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

References

[1] Shoji Y, Nakamura H, Iwabuchi T, Aonashi K, Seko H, Mishima K, Itagaki A, Ichikawa R, Ohtani R. Tsukuba GPS dense net campaign observation: Improvement in GPS analysis of slant path delay by stacking one-way postfit phase residuals. Journal of the Meteorological Society of Japan. 2004;**82**:301-314

[2] Shoji Y, Yamauchi H, Mashiko W, Sato E. Estimation of local-scale precipitable water vapor distribution around each GNSS station using slant path delay. Scientific Online Letters on the Atmosphere: SOLA. 2014;**10**:29-33

[3] Shoji Y, Mashiko W, Yamauchi H, Sato E. Estimation of local-scale precipitable water vapor distribution around each GNSS station using slant path delay: Evaluation of a severe tornado case using high-resolution NHM. Scientific Online Letters on the Atmosphere: SOLA. 2015;**11**:31-35

[4] Kawabata T, Shoji Y, Seko H, Saito K. A numerical study on a mesoscale-convective system over a subtropical island with 4D-var assimilation of 4 GPS slant total delays. Journal of the Meteorological Society of Japan. 2013;**91**:705-721

[5] Askne J, Nordius H. Estimation of tropospheric delay for microwaves from surface weather data. Radio Science. 1987;**22**:379-386

[6] Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH. GPS meteorology: Remote sensing of atmospheric water vapor using or estimation of the wet delay. Journal of Geophysical Research. 1992;**97**:15.787-15.801

[7] Businger S, Chiswell SR, Bevis M, Duan J, Anthes RA, Rocken C, Ware RH, VanHove T, Solheim FS. The promise of GPS in atmospheric monitoring. Bulletin of the American Meteorological Society. 1996;**77**:5-18

[8] Kanda M, Ishida T, Kashima M, Oishi S. Analysis of temporal and spatial change of a convective thunder storm in Tokyo metropolitan using GPS precipitable water: Case study on 23rd August in l997. Tenki. 2000;**47**:7-15 (in Japanese)

[9] Niimura N, Sasaki T, Kimura F. Statistical relation between precipitation and precipitable water vapor obtained from the global positioning system (GPS) in Tokyo metropolitan area. Tenki. 2000;**47**:635-642 (in Japanese)

[10] Inoue H, Inoue T. Characteristics of the water-vapor field over the Kanto District associated with summer thunderstorm activities. Scientific Online Letters on the Atmosphere: SOLA. 2007;**3**:101-104

[11] Hirahara K. Local GPS tropospheric tomography. Earth Planets and Space. 2000;**52**:935-939

[12] Noguchi W, Yoshihara T, Tsuda T, Hirahara K. Time-height distribution of water vapor derived by moving cell tomography during Tsukuba GPS campaigns. Journal of the Meteorological Society of Japan. 2004;**82**:561-568

[13] Seko H, Nakamura H, Shoji Y, Iwabuchi T. Meso-γ scale water vapor distribution associated with the thunderstorm on 1 august 2001 estimated by the tomography method. Journal of the Meteorological Society of Japan. 2004;**82**:569-586

[14] Nakamura H, Koizumi K, Mannoji N. Data assimilation of GPS precipitable water vapor into the JMA mesoscale numerical weather prediction model and its impact on rainfall forecast. Journal of the Meteorological Society of Japan. 2004;**82**:441-452

[15] Koizumi K, Sato Y. Impact of GPS and TMI precipitable water data on mesoscale numerical weather prediction model forecasts. Journal of the Meteorological Society of Japan. 2004;**82**:453-457

[16] Seko H, Kawabata T, Tsuyuki T, Nakamura H, Koizumi K, Iwabuchi T. Impacts of GPS-derived water vapor and radial wind measured by Doppler radar on numerical prediction of precipitation. Journal of the Meteorological Society of Japan. 2004c;**82**:473-489

[17] Shoji Y. A study of near real-time water vapor analysis using a nationwide dense GPS network of Japan. Journal of the Meteorological Society of Japan. 2009a;**87**:1-18

[18] Shoji Y, Kunii M, Saito K. Assimilation of nationwide and global GPS PWV data for a heavy rain event on 28 July 2008 in Hokuriku and Kinki, Japan. Scientific Online Letters on the Atmosphere: SOLA. 2009b;**5**:45-48

[19] Shimada S, Seko H, Nakamura H, Aonashi K, Herring TA. The impact of atmospheric mountain lee waves on systematic geodetic errors observed using the global positioning system. Earth, Planets and Space. 2002;**54**:425-430

[20] Ishikawa, Y., Application of GPS signal delay observed by ground-based receivers, validation and upgrading of analysis-forecast system, Suuchi Yohou Kenshuu Text (Text for NWP Training Course). 2008;**41**:53-57 (in Japanese)

[21] Ha S, Kuo Y-H, Guo Y-R, Lim G-H. Variational assimilation of slant-path wet delay measurements from a hypothetical ground-based GPS network. Part I: Comparison with precipitable water assimilation. Monthly Weather Review. 2003;**131**:2635-2655

[22] Zou X, Kuo Y-H. Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. Monthly Weather Review. 1996;**124**: 2859-2882

[23] Järvinen H, Eresmaa R, Vedel H, Salonen K, Niemelä S, de Vries J. A variational data assimilation system for ground-based GPS slant delays. Quarterly Journal of the Royal Meteorological Society. 2007;**133**:969-980

[24] Gustafsson N, Berre L, Hörnquist S, Huang X-Y, Lindskog M, Navascués B, Mogensen KS, Thorsteinsson S. Three-dimensional variational data assimilation for a limited area model. Part I: General formulation and the background error constraint. Tellus. 2001;**53A**:425-446

[25] Bauer H-S, Wulfmeyer V, Schwitalla T, Zus F, Grzeschik M. Operational assimilation of GPS slant path delay measurements into the MM5 4DVAR system. Tellus. 2011;**63A**:263-282

[26] MacMillan DS. Atmospheric gradients from very long baseline interferometry observations. Geophysical Research Letters. 1995;**22**:1041-1044

[27] Bar-Sever YE, Kroger PM, Borjesson JA. Estimating horizontal gradients of tropospheric path delay with a single GPS receiver. Journal of Geophysical Research. 1998;**103**(B3):5019-5035

[28] Ruffini G, Kruse LP, Rius A, Burki B, Cucurull L, Flores A. Estimation of tropospheric zenith delay and gradients over the Madrid area using GPS and WVR data. Geophysical Research Letters. 1999;**26**:447-450

[29] Aonashi K, Shoji Y, Ichikawa R, Hanado H. Estimating spatial variation of humidity from GPS and WVR observation in Tsukuba. Earth Planets and Space. 2000;**52**:907-912

[30] Sato K, Realini E, Tsuda T, Oigawa M, Iwaki Y, Shoji Y, Seko H, High-resolution A. Precipitable water vapor monitoring system using a dense network of GNSS receivers. Journal of Disaster Research. 2013;**8**:37-47

[31] Shoji Y. Retrieval of water vapor anisotropy using the Japanese nationwide GPS array and its potential for prediction of convective precipitation. Journal of the Meteorological Society of Japan. 2013;**91**:43-62

[32] Mashiko W. A numerical study of the 6 may 2012 Tsukuba City supercell tornado. Part I: Vorticity sources of low-level and midlevel mesocyclones. Monthly Weather Review. 2016;**144**:1069-1092

[33] Sasaki Y. An objective analysis based on the variational method. Journal of the Meteorological Society of Japan. 1958;**36**:77-88

[34] Sasaki Y. Some basic formalisms in numerical variational analysis. Monthly Weather Review. 1970;**98**:875-883

[35] Kawabata T, Seko H, Saito K, Kuroda T, Tamiya K, Tsuyuki T, Honda Y, Wakazuki Y. An assimilation and forecasting experiment of the Nerima heavy rainfall with a cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. Journal of the Meteorological Society of Japan. 2007;**85**:255-276

[36] Kawabata T, Kuroda T, Seko H, Saito K. A cloud-resolving 4D-Var assimilation experiment for a local heavy rainfall event in the Tokyo metropolitan area. Monthly Weather Review. 2011;**139**:1911-1931

[37] Saito K. The Japan Meteorological Agency Nonhydrostatic Model and its Application to Operation and Research. Atmospheric Model Applications. Vienna: InTech; 2012. pp. 85-110

[38] Bean BR, Thayer GD. Central radio propagation laboratory exponential reference atmosphere. Journal of Research of the National Bureau of Standards. 1959;**64D**:315-317

[39] National Imagery and Mapping Agency, Department of Defense World Geodetic System 1984. Its Definition and Relationships with Local Geodetic Systems, 3rd ed. 4 July 1997. NIMA Technical Report TR8350.2

Satellite Navigation System Advances

Antennas and Front-End in GNSS

Korkut Yegin

Abstract

Antenna and front-end play a key role in global navigation satellite system (GNSS) receivers where multi-frequency and multi-constellation services are used simultaneously to produce high-precision position, navigation, and timing information. Being the first element on the receiver system, specifications on the antenna for multi-constellation GNSS applications can be challenging. Especially, integration of the antenna into the target platform, either mobile or stationary, may severely affect antenna performance. This is usually an issue for small-size antennas where measured stand-alone antenna performance in ideal conditions is usually not descriptive of actual performance on the platform. Furthermore, carrier phase tracking has become popular among algorithm developers to obtain high accuracy and anti-spoofing at the same time which demand minimal phase centre variation of the antenna within the intended GNSS band. Spoofing and jamming of GNSS receivers is a growing concern especially for aerial vehicles with ever-increasing applications of drones. These requirements demand different characteristics on the antenna and front-end than traditional applications. One of the most utilized forms of GNSS antenna is ceramic patch, due to its low height, low cost, and relatively good narrow band performance. Simulations of this particular antenna in terms of axial ratio and impedance bandwidths, axial ratio variation over elevation, and half-power beam width are carried out and discussed with comparison to its counterparts. Another critical part of the receiver is its front-end where huge amount of signal amplification with minimal distortion takes place. Long integration times (>1 ms) in GNSS signal processing also puts severe requirements on the software and temperature-compensated crystal oscillator. For mass production, the front-end should be implemented in the form of an integrated circuit. Front-end architectures from traditional superheterodyne to zero/low-intermediate frequency configurations are presented. Advantages and disadvantages of each configuration are outlined in view of multi-band and multi-standard GNSS receivers.

Keywords: GNSS antenna, carrier phase tracking, GNSS low-noise amplifier, GNSS front-end, low-IF receiver

1. Introduction

Global navigation satellite systems (GNSSs) are indispensible in positioning, navigation, and timing (PNT) and have become an integral part of many outdoor positioning applications such as surveying, vehicle localization, parcel and container tracking, precision timing, synchronization of communication networks and radars, atmospheric observation, and meteorology. Although GNSS has started with US Global Satellite System (GPS) only, with the addition of new GNSS services introduced by Russia, China, and European Union, GNSS has evolved to multi-constellation and multi-band system. Regional satellite navigation system by India and GNSS assistance services such as QRZZ also complement global satellite navigation system. All global service providers of GNSS offer worldwide positioning for mobile and stationary platforms and assets.

While multiple GNSS services at different frequency bands offer tremendous advantages for the user which were not possible with single service provider, multi-band and multi-constellation receivers and antennas possess new challenges in the system design. For precise positioning, multiple satellites, at least four for each service provider, must be tracked simultaneously. One of the key components of the system is the GNSS passive antenna, which is vital to establish a good carrier-to-noise ratio for seamless positioning with minimum acquisition time. The antenna beam width must be broad to cover as much as possible the sky view while its axial ratio must be low throughout its coverage. The antenna must maintain these features throughout the target frequency bands. The receiver, on the other hand, must be able to handle multi-constellation GNSS signals. Instead of classical receiver architectures, software-defined, user-configurable GNSS architectures are much more in demand due to the flexibility in software they offer.

One of the key challenges in any GNSS is the susceptance of receiver to interference. The signals transmitted through satellites are at low power such that the received signals are very weak on earth and usually under the thermal noise floor of the receiver. Intentional and unintentional jamming of GNSS signals is common and still presents the biggest problem in GNSS applications. Especially, in urban environment where tall buildings block clear view of the antenna and multipath propagation is dominant, the receiver performance deteriorates significantly. Unintentional blockage of GNSS due to other communication systems is also common. Strong out-of-band signals or signal bleeding from nearby frequency bands can cause interruptions of GNSS service. Due to very weak signal levels on earth, intentional jamming with inexpensive hardware has also proven to be harmful for GNSS service. Jamming mitigation at the hardware and software are essential components of the mission-critical GNSS receivers.

2. GNSS passive antenna

Since the inception of GPS, satellite navigation antenna, maybe, is one of the most studied antenna structures in the literature. It is difficult to give a comprehensive list, but microstrip

patch antennas and antenna performance on the platform [1–7], wideband antennas [8, 9], wide beam width antenna [10], miniature and multi-functional antennas [11–18], dual/triple GNSS band antennas [19–25], conformal and missile application antennas [26–29], array antennas for anti-jam and anti-spoofing applications [30–35], cellular phone isolation [36, 37], metamaterial and plasma-supported antennas [38–41] are reported in the literature.

Depending on the antenna location inside the platform and available antenna space, antenna designers routinely face challenges to meet acceptable performance. Most utilized forms of GNSS antennas are microstrip antennas, helical antennas, slot-based antennas and miniature (chip-scale) antennas. GNSS antenna arrays are often essential for critical applications where precise positioning is required along with counter measures for jamming and spoofing.

2.1. GNSS passive antenna requirements

GNSS passive antenna performance is usually quantified in terms of operational frequency band, gain pattern, half-power beam width (HPBW), polarization, axial ratio, cross-polarization discrimination or multipath discrimination, and phase centre stability.

2.1.1. Operational frequency band

The passive antenna must be functional within the desired GNSS service band. The operation frequencies of current GNSS services are tabulated in **Table 1**. A passive antenna that is capable of receiving entire GNSS services must be operational from 1164 to 1610 MHz (32.1% fractional bandwidth), covering either entire band or multi-band within lower L-band (1164–1300 MHz) and upper L-band (1559–1610 MHz).

2.1.2. Polarization

L-band satellite navigation systems utilize right-hand circular polarization (RHCP) signals. Two orthogonal components of circular polarization signal at high elevations undergo same level of Faraday rotation when passing through ionosphere which does not degrade the

Service	Lower L-band	Upper L-band
GPS	L5: 1164–1189 MHz	L1: 1567–1587 MHz
	L2: 1215–1239.6 MHz	
Galileo	E5: 1164–1215	E1: 1559–1591
	E6:1260–1300 MHz	
GLONASS	G3: 1189–1214 MHz	G1: 1593–1610 MHz
	G2: 1237–1254 MHz	
BeiDou/compass	B2I: 1179–1203 MHz	B1I: 1553–1569 MHz
	B3: 1256–1280	

Table 1. GNSS frequency bands.

polarization purity of the signal. For linearly polarized signals, Faraday rotation causes the signal to a different tilt angle than the original which should be compensated at the receiving antenna on earth either by rotating the antenna to correct polarization or utilizing both orthogonal components of the received signal for polarization compensation.

2.1.3. Axial ratio and multipath rejection

The purity of a circular polarization is stated in terms of axial ratio (AR), which is defined as the ratio of two orthogonal components of electric field on polarization ellipse traced by the electric field vector in time domain. Axial ratio changes with elevation and azimuth, but it is usually stated at zenith as a single value. Good GNSS antennas usually have less than 1 dB axial ratio at zenith, and moderate ones have less than 3 dB. Although it is possible to obtain low axial ratios at zenith, it is relatively difficult to achieve the same performance at low elevation angles. Most circularly polarized antenna structures are linearly polarized at low elevation angles, which make reception somewhat difficult but jamming and spoofing easy. Furthermore, it is also difficult to maintain the same axial ratio over a large bandwidth. Thus, axial ratio bandwidth is often the limiting factor than the impedance bandwidth for circularly polarized antennas.

Cross-polarization (*Xpol*) of the antenna is related to the axial ratio through the following equation:

$$Xpol = (AR-1)^2/(AR+1)^2 \tag{1}$$

Thus, specification of axial ratio is sufficient to estimated cross-polarization level of the directly received signal. Upon reflection from a surface, the sense of polarization changes from right hand to left hand, thus good cross-polarization at the antenna leads to better rejection of multi-path signals. Multi-path rejection ratio (*MPRR*) of the antenna for ground reflection can be formulated as

$$MPRR_{dB} = 20\log\left(\frac{E_{RHCP}(\theta)}{E_{RHCP}(180-\theta)+E_{LHCP}(180-\theta)}\right) \tag{2}$$

where 180-θ corresponds to the angle for the ground reflected signal.

2.1.4. Beam width and gain

HPBW is a good measure of antenna gain roll-off and sky coverage. It is usually desired to be in excess of 85°. For wider beam width (>120°), the antenna gain must be compromised to lower values. This trade-off, sometimes, pays off depending on the platform where the antenna is mounted. However, larger beam width may be a disadvantage for spoofing countermeasures and antenna noise. Desired gain roll-off depends on the particular application and the platform where the antenna is situated.

Passive antenna gain is largely dependent on the choice of antenna structure which is usually dictated by the available space for the antenna. In space-constrained applications, the form factor antenna is so small that the expected gain can be as low as −5 dBic. Most fade margin

calculations assume 0 dBic antenna gain, but this can be difficult to achieve for a printed circuit board (PCB) or chip antenna. Moreover, these small, integrated antennas suffer non-uniform reception in the azimuth, making antenna reception dependent on the platform or user orientation.

2.1.5. Phase-centre stability

Electrical phase centre of the antenna is described as the geometrical point where all rays converge or emanate from it, i.e. incident rays add up in phase. Measured signals are all relative to this position, thus antenna phase centre plays a key role in achieving millimetre resolution in positioning. Although it is described as a single point in space, the phase centre changes with frequency. Within the respective GNSS band, a phase centre variation adds up to inaccuracies in pseudo-range calculation. Especially in an array configuration where the main lobe of the array is tilted to desired angle, phase centre may change depending on the tilt angle. Phase centre offset must be specified for every direction and frequency for accurate PNT estimation. In recent years, spoofing or interference detection based on received signal phase variation relies on phase-centre variation of the passive antenna. Less than 1 cm phase centre variation in transverse plane is usually acceptable, though smaller is better.

General technical specifications for a GNSS passive antenna are summarized in **Table 2**. It should be noted that these specification are by no means strict and can vary largely depending on the antenna platform and the specific application. For instance, mobile phone GNSS antenna specifications largely vary from that of a revolving missile guidance antenna or from a land-surveying antenna.

2.2. Microstrip antennas

Microstrip patch antennas are low-profile, easy to integrate, and relatively low cost. In most applications, the height of the antenna becomes the limiting factor in platform design. Although helical antennas, for instance, perform better than microstrip counterparts, they are rarely used in vehicular and mobile applications. Typical heights of microstrip antennas

Specification	Value
Frequency band	1164–1610 MHz
Polarization	RHCP
Input impedance	50 Ohm
VSWR	<2.5 (typical)
Gain at 0° (zenith)	Min 0 dBic
HPBW	85°–100° (typical)
Axial ratio (zenith)	<3 dB
Phase-centre stability	<10 mm

Table 2. GNSS antenna specifications.

for GNSS are 2–5 mm and can be designed on low or high dielectric substrates. Depending on substrate choice, the transverse dimensions range from 15 to 35 mm. One of the most commonly used form of microstrip antennas is ceramic dielectric with relative permittivity around 20, and overall size is typically 25 × 25 mm. This particular size and antenna performance match very well for most vehicle tracking and navigation applications. Many automobiles utilize this size for their information and navigation unit. The design of microstrip antennas is well documented in text books. Resonating patch dimension is made half of the guided wavelength and TM_{01} and TM_{10} modes are both excited with 90° phase difference with proper excitation and slight difference between electrical lengths that correspond to respective modes. Fractional and axial ratio bandwidths are usually small but tolerable for L1 band GPS and Galileo. For entire upper L-band, axial ratio is usually compromised for impedance bandwidth and most ceramic patches are linearly polarized than RH circular for 51 MHz bandwidth of upper L-band.

Typical ceramic patches and dimensions of 25 × 25 mm ceramic patch for GPS L1 band are shown in **Figure 1**. This particular patch is modeled on 70 × 70 mm ground plane and simulated using CST Microwave Studio [42]. The results are displayed in **Figures 2–4**. Although the impedance bandwidth (VSWR < 2.5) is 35 MHz, its axial ratio bandwidth is only 8 MHz. Elevation cut along phi = 0 plane also indicates that the axial ratio becomes larger than 3 dB at low elevations, in fact, it becomes linear vertically polarized, which makes it more susceptible to terrestrial jammers. Nevertheless, its peak gain of 5 dBic and HPBW of 100° at 1575.4 MHz, are quite acceptable for automobile and mobile applications, which is why it has been so popular among automotive OEMs. Measured performance is expected to be slightly worse than the simulations. The patch antenna is also ground plane size-dependent. For smaller ground planes, although peak gain is higher, HPBW gets narrower, and axial ratio performance especially at low elevations deteriorates considerably.

2.3. Helical antennas

Helical antenna, since its invention by J.D. Krauss, has become the prime choice for circular polarization due its excellent axial ratio bandwidth and ease of construction. When the
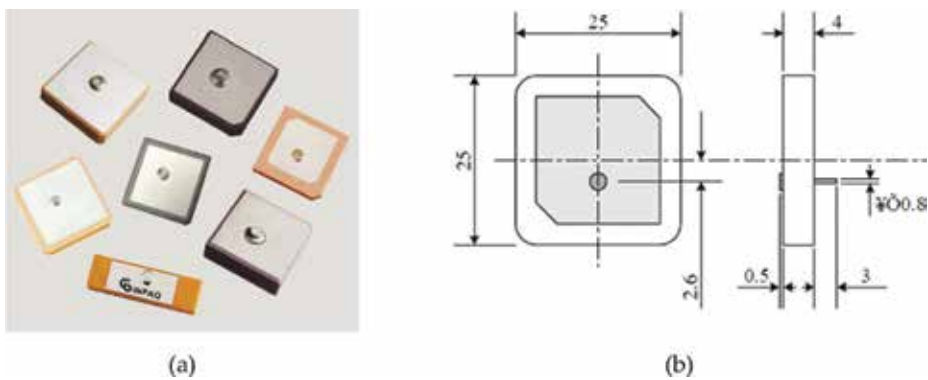

Figure 1. Ceramic GNSS patch antennas. (a) Inpaq [43] and (b) 25 × 25 mm patch by Amotech [44].
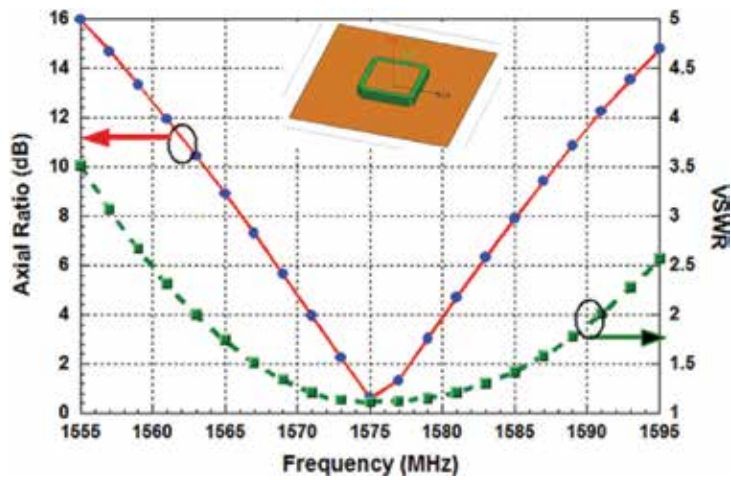
Figure 2. Axial ratio and VSWR of 25 × 25 mm ceramic patch on 70 × 70 mm ground plane.

diameter of the helix is in the order of operation wavelength, the antenna becomes circularly polarized with broadside radiation, and the rotation sense of helix provides the sense of polarization either left hand or right hand. This mode of operation is designated as axial mode. The axial ratio becomes smaller as the number of turns is increased. However, at GNSS bands, the wavelength is too large for the helical antenna to be of practical use for integration into a user platform. If smaller diameter is used, the antenna acts like a wire antenna, and this mode of operation is termed as normal mode. To obtain circular polarization, four helices with 90° apart geometrically are fed with 90° successive phase differences. This particular form of the antenna
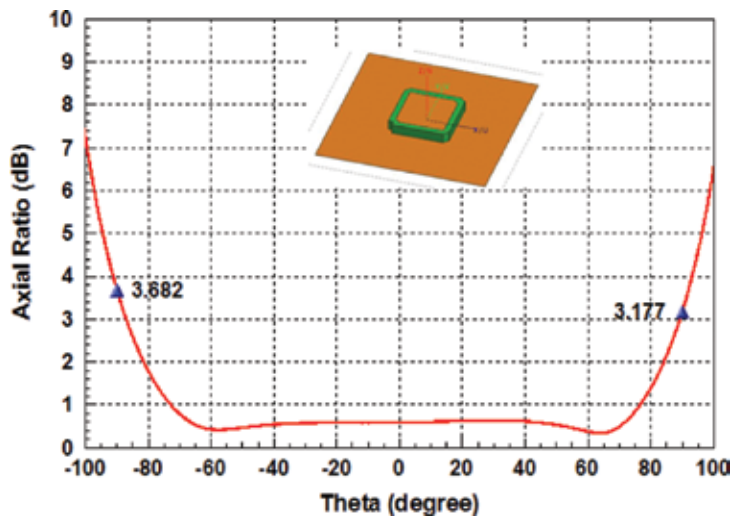

Figure 3. Axial ratio on phi = 0 cut for 25 × 25 mm ceramic patch on 70 × 70 mm ground plane at 1575.4 MHz.
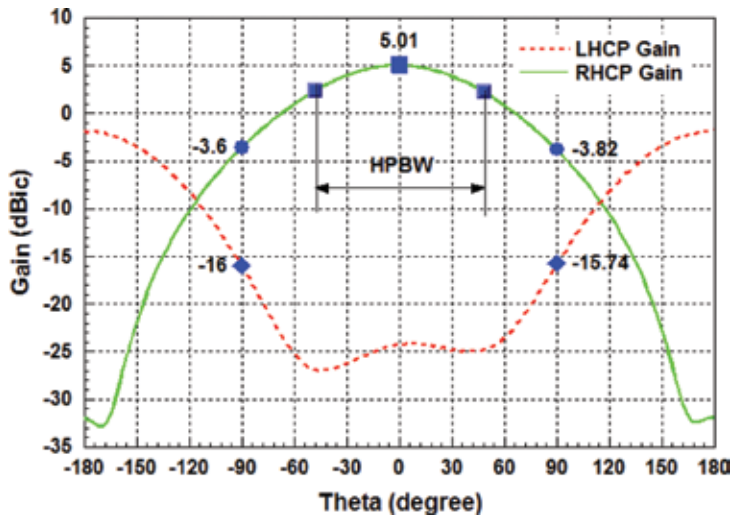
Figure 4. RHCP and LHCP gain of 25 × 25 mm ceramic patch on 70 × 70 mm ground plane at 1575.4 MHz.

is known as quadrifilar helical antenna (QHA) and has been heavily used due to its superior performance compared to patch antennas, especially wider axial ratio bandwidth. The size of the antenna can be reduced by printing helical arms on high dielectric constant ceramic substrate. The helical arms can also be implemented on flexible thin substrates. These two forms of helical antennas are shown in **Figure 5**. Most QHA designs utilize a quadrature hybrid coupler to feed the four arms of the antenna. In these designs, the bandwidth of the quadrature coupler is usually the limiting factor in achieving wide axial ratio bandwidth rather than the antenna structure itself. For instance, upper L-band GNSS can be entirely covered with QHA if an appropriate quadrature coupler is used. Quadrature couplers designed on low-temperature co-fired ceramic (LTCC) substrate provide chip-size dimensions but have narrow bandwidth due to high dielectric constant used.
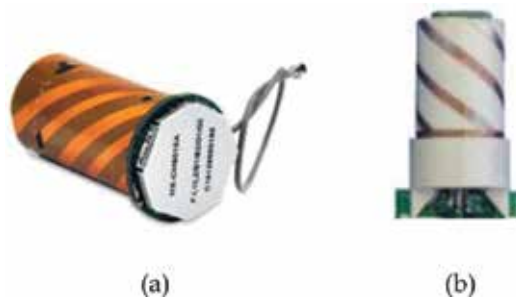

Figure 5. Quadrifilar helical antenna: (a) printed on foil (Harxon HX-CH6017A [45] and (b) printed on dielectric ceramic (Serantel [46]).

Figure 6. Chip antenna. (a) Monopole and helix by Abracon [47] and (b) Taoglass MAT.12A dual-chip antenna fed with a quadrature coupler [48].

2.4. Miniature antennas

Miniature antennas in several millimetre scales are often used in highly integrated consumer products. Their polarization is linear and average gains are usually less than −2 dBi even with large ground planes underneath. Different types of chip antennas in the form of monopoles and helix are shown in **Figure 6a**. In an attempt to improve their performance, Taoglass offered two chip antennas oriented perpendicular to each other and fed by a quadrature coupler as illustrated in **Figure 6b**.

2.5. High-performance antennas

High-precision GNSS systems require best antenna performance in terms of good axial ratio at low elevation angles, high rejection to multipath signals and minimum phase centre variation, and very high front-to-back ratio. Measured characteristics of these antennas are also integrated into the calibration of the GNSS receiver system. To achieve sub-centimetre accuracy, these antennas often supplemented with augmentation systems which provide local correction factors in pseudo-range estimates. These antennas also play an important role for rover applications as two-way communication outside the GNSS band is established between the units. Choke rings and artificial magnetic conductors are used to increase multipath rejection by maintaining good axial ratio even below the horizon of the antenna plane. Two different choke rings are illustrated in **Figure 7** by Leica Systems [49].
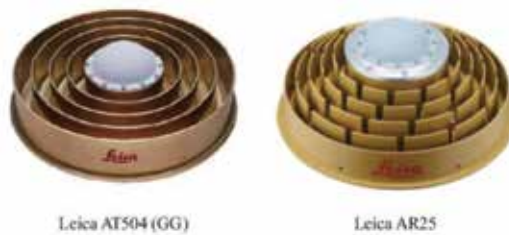

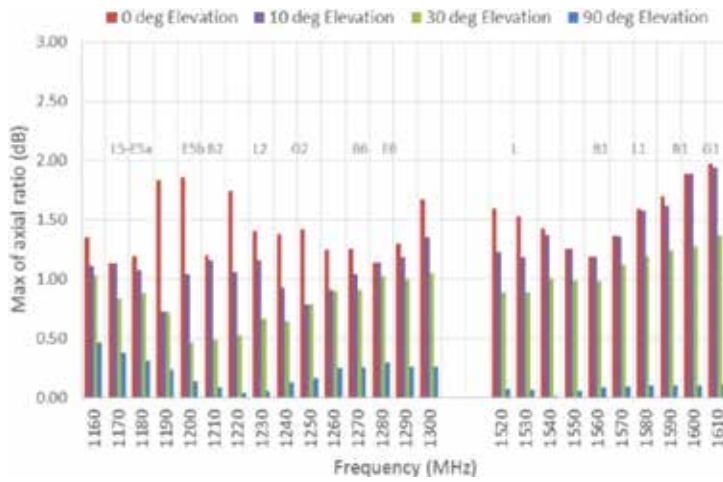**Figure 7.** Leica choke-ring antennas [49].

Figure 8. Axial Ratio of VeraPhase™ 6000 antenna by Tallysman Wireless Inc. [50].
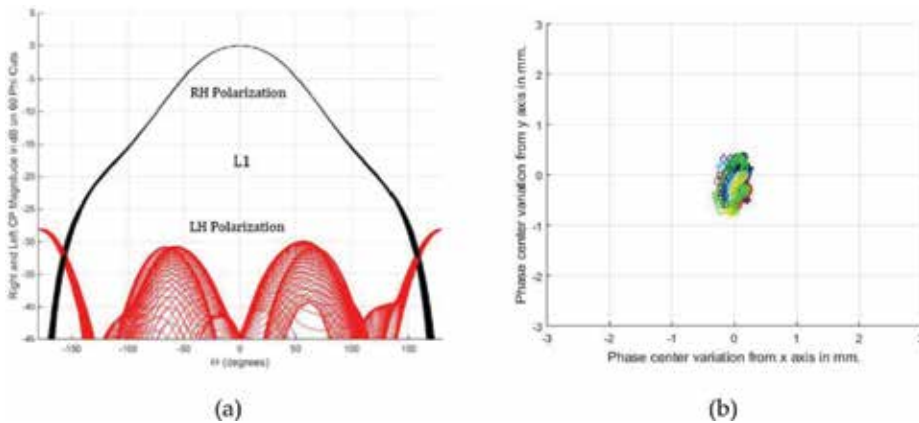

Figure 9. VeraPhase™ 6000 antenna by Tallysman Wireless Inc. (a) normalized RHCP and LHCP gain patterns and (b) phase centre variation [50].

Performances of Tallysman VeraPhase 6000 antenna [50] are shown in **Figures 8** and **9**. Axial ratio difference between 90° and at 0° elevation angles are about 1.5 dB, cross-polarization is less than −30 dB on antenna plane, and phase centre variation in horizontal plane is around 1 mm.

3. GNSS antenna LNA

LNA is an essential component of any satellite system [51, 52]. Typical strength of received GNSS signal is around 0.14 μV peak (−127 dBm for 50 Ohm), as mentioned earlier well below the thermal noise floor (−101 dBm for 20 MHz antenna bandwidth, i.e. 2.82 μV-peak). Received signal buried in noise, must be amplified all the way up to 1 V peak-to-peak for analog-digital

conversion (ADC) with minimal distortion. Most GNSS receivers utilize a low noise amplifier (LNA) connected to the passive antenna. This LNA is critical in determination of system noise figure, which is a measure of signal-to-noise ratio (SNR) degradation when the signal passes through a component or device. Total noise figure of the receiver chain is calculated as:

$$NF = NF_1 + \frac{NF_1 - 1}{G_1} + \frac{NF_2 - 1}{G_1 G_2} + \dots \tag{3}$$

Hence, first amplifier's gain and noise figure are vital to the system noise figure. The gain of the first amplifier should be kept as high as possible but the noise figure as low as possible which are contradictory in nature. Usually manufacturers of low-noise transistors provide optimum source reflection coefficient (Γopt) for the best noise figure, which is different than the maximum transducer gain. LNA design with feedback topologies target achieving best of both metrics at the expense of additional components and circuit complexity. LNA gain should be carefully selected to overcome the cable loss between the antenna LNA and receiver input LNA. Too much gain at the antenna LNA may overload and compress the receiver LNA for degraded performance. If the antenna is close to the receiver, passive antenna can be directly connected to the receiver with noise figure of the first stage being only the antenna cable loss. This configuration is not preferred in practical applications simply because prefilter for out-of-band rejection is usually added before the receiver LNA, and this, combşned with the antenna cable loss, increases the noise figure. Alternative configuration is to have a 10–15 dB LNA gain for short and moderately long cables and 25–30 dB gain for long cables.

Although GNSS receiver exploits processing gain to increase received signal SNR, in-band and out-band interference can degrade the receiver performance. Out-of-band interference can be handled at the antenna LNA using highly selective bandpass filter. The location of this filter either before or after the first gain stage impacts the noise figure and system performance. Sources of out-of-band interferers vary but they are often attributed to cellular phone base stations, terrestrial broadcast towers, radars, where second- and third-order products of transmitted signals fall into GNSS band. For instance, terrestrial broadcast of satellite digital audio system (SDARS) at S band (2320–2345 MHz) mixed 800 MHz 2G/3G cellular band falls into upper L-band, or L-band secondary surveillance radar at airport traffic control can easily mix up with 350/433 MHz or trunked radio services produce second-order products at GNSS bands. In addition, strong out-of-band signals may compress the LNA and de-sense the receiver. Closely packed cellular phone antennas, if not carefully designed, may easily compress the input LNA at the antenna.

In-band interferers due to third-order products should also be taken into account to avoid non-linear operation of the LNA. Too much gain at the antenna LNA can easily produce third-order products that compress the receiver LNA.

Typical passive antenna and LNA configurations are shown in **Figure 10** where pre-filter is located before or after the first-stage LNA. Having filter before the first-stage amplifier increases selectivity of the receiver by suppressing out-of-band signals but at the same time increases the noise figure, which may become critical, especially reception from low elevation angles or in an urban setting. For nearby radiators to GNSS antenna, this configuration
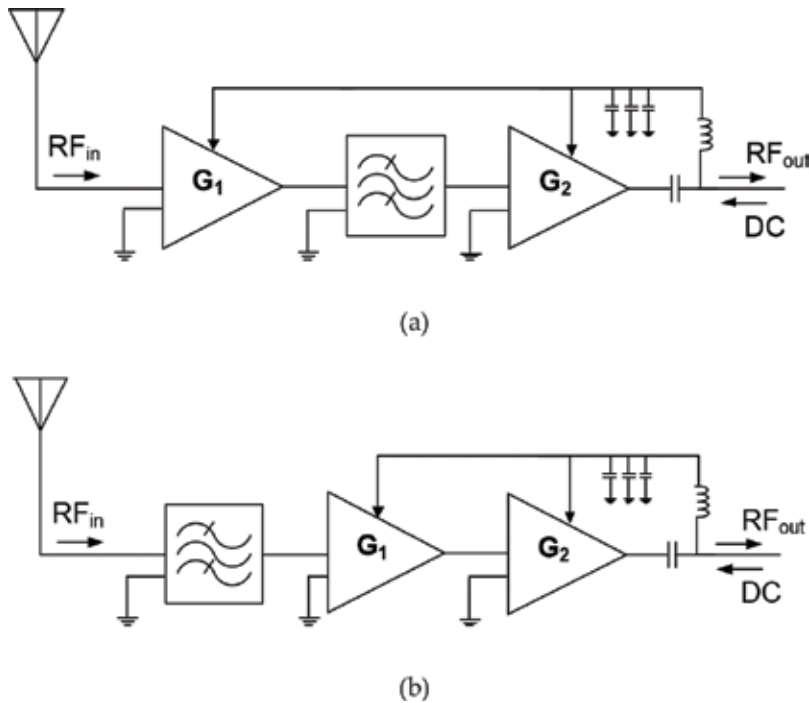
Figure 10. Antenna LNA: (a) filter after first stage and (b) filter before first stage.

is inevitable. Filter after the first-stage LNA does not degrade the noise figure much and provides good selectivity and rejection for out-of-band signals. However, in the presence of a strong interferer, antenna LNA can be overloaded and signal acquisition can be lost.

4. GNSS receiver front-end

GNSS receivers generally utilize RF down-conversion to an intermediate frequency (IF), using one or two conversions, followed by analog-to-digital conversion (ADC). The output of ADC is interfaced to a general purpose processor (GPP) for digital signal processing operations such as correlation, acquisition, tracking, and PNT extraction. The section where the signal remains analog, i.e. up to ADC, is usually termed as front-end. Receiver architectures based on how they process signal can be classified as superheterodyne, low IF, zero IF (homodyne), and direct-digital (bandpass sampling). Receiver performance metrics can be quite detailed but the most important ones are sensitivity, selectivity, inter-modulation characteristics, non-linearities, and spur-free dynamic range.

Superheterodyne receiver is the most classical architecture, and has been utilized in many communication systems due to its excellent sensitivity, selectivity, and dynamic range. Typical configuration is shown in **Figure 11**. However, the architecture is not flexible for
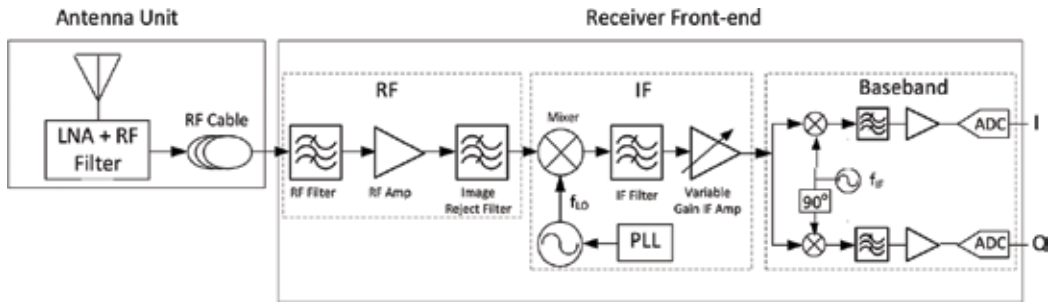
Figure 11. Antenna unit and superheterodyne receiver front-end.

multi-standard systems, not well suited for integrated circuits due to filter requirements for image-reject and non-linear products, and consumes substantial power.

To overcome the drawbacks of superheterodyne receiver, low-IF and zero-IF receiver configurations are proposed. Typical architecture of these receivers is displayed in **Figure 12**. In zero-IF configuration, image problem is completely removed, hence sharp and IC unfriendly image reject filters are not needed. Since gain is shared between RF and baseband amplifiers, requirements on these amplifiers become complicated to have sufficient SFDR. But, the most troubling problem of zero-IF is that the leakage of LO mixes with itself, putting severe requirements on second-order intermodulation products of the receiver. This leakage causes DC offsets at the baseband, and high baseband gain amplifies these offsets together with flicker noise to degrade receiver performance. Also, Doppler shifts of received signals can be lost. Low-IF configuration overcomes these problems but image issue comes up again and usually resolved by image-reject mixer design, which is not so easy in IC topology. Despite severe requirements on gain, noise, and linearity, low power budget and flexibility in DSP made these architectures very popular, especially among receiver ICs.

Today's multi-band and multi-constellation receivers mostly utilize software-defined radio architecture to process digital I and Q signals. When two or more frequencies of GNSS are targeted, the GPP unit of the receiver becomes critical for signal quality and high data throughput, which may require parallel processing of correlations [53]. In contrast to communication systems
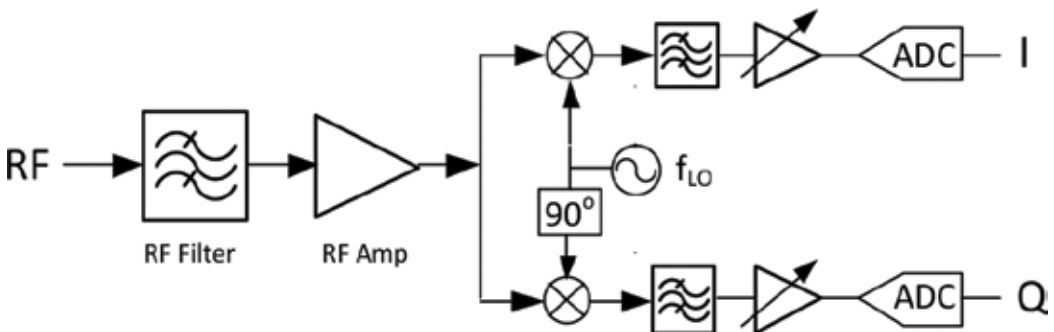

Figure 12. Zero-IF and low-IF receiver front-end. IF filters are either low-pass (zero-IF) or bandpass (low-IF).

where received signal SNR is high, GNSS receivers rely on long coherent integration (>1 ms) to exploit processing gain. Clock stability becomes an issue especially during carrier phase tracking. Instead of quartz oscillators, temperature-compensated crystal oscillators (TCXO) with an accuracy better than one part per million with very low phase noise are used in high-end receivers.

5. Conclusions

GNSS performance relies on the antenna, front-end receiver, and DSP algorithms utilized in the software. Antenna, being the first element in the reception chain, plays a key role in retrieving PNT information. Antenna specifications based on the expected performance of the GNSS receiver were reviewed and discussed in detail. Common forms of antennas ranging from high end to low-cost applications were presented. Front-end receiver configurations and elements were outlined for GNSS performance. Low-IF digital configuration was recommended for best receiver performance. This review should especially help algorithm developers in choosing the right antenna and receiver configuration.

Acknowledgements

Conflict of interest

None.

Author details

Korkut Yegin

Address all correspondence to: yegink@gmail.com

Ege University Electrical and Electronics Engineering, Izmir, Turkey

References

[1] Huang C-Y, Lin M-H. Ceramic GPS antenna for remote sensing. IGARSS 2000. In: IEEE 2000 International Geoscience and Remote Sensing Symposium. 2000. pp. 2182-2184. DOI: 10.1109/IGARSS.2000.858349

http://dx.doi.org/10.5772/intechopen.74971

[2] Abraham O, Leib A, Matzner H. Designing and building of a circular GPS antenna. In: IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems. 2011. pp. 1-3. DOI: 10.1109/COMCAS.2011.6105835

[3] Nasimuddin, Chen ZN, Qing X. Dual-band circularly polarized S-shaped slotted patch antenna with a small frequency-ratio. IEEE Transactions on Antennas and Propagation. 2010;**58**:2112-2115. DOI: 10.1109/TAP.2010.2046851

[4] Bilotti F, Vegni C. Design of high-performing microstrip receiving GPS antennas with multiple feeds. IEEE Antennas and Wireless Propagation Letters. 2010;**9**:248-251. DOI: 10.1109/LAWP.2010.2046874

[5] Boccia L, Amendola G, Di Massa G. A shorted elliptical patch antenna for GPS applications. IEEE Antennas and Wireless Propagation Letters. 2003;**2**:6-8. DOI: 10.1109/LAWP.2003.810767

[6] Yegin K. On-vehicle GPS antenna measurements. IEEE Antennas and Wireless Propagation Letters. 2007;**6**:488-491. DOI: 10.1109/LAWP.2007.907056

[7] Yegin K. Instrument panel mount GPS antenna. Microwave Optical Technology Letters. 2007;**49**:1979-1981

[8] Scire-Scappuzzo F, Makarov SN. A low-multipath wideband GPS antenna with cutoff or non-cutoff corrugated ground plane. IEEE Transactions on Antennas and Propagation. 2009;**57**:33-46. DOI: 10.1109/TAP.2008.2009655

[9] Yigit O, Yegin K. All band GNSS antenna with artificial magnetic conductor. In: 2017, Progress in Electromagnetic Research Symposium. 2017. pp. 2211-2214. DOI: 10.1109/PIERS.2017.8262118

[10] Hong-mei Z, Yan-juan N, Teng L. A novel GPS antenna with wide beamwidth. In: 2010 2nd International Conference on Future Computer and Communication. 2010. pp. 49-52. DOI: 10.1109/ICFCC.2010.5497838

[11] Brzezina G, Roy L. A miniaturized GPS antenna in LTCC with linear polarization suitable for SoP integration. In: 2010 14th International Symposium on Antenna Technology and Applied Electromagnetics & The American Electromagnetics Conference. 2010; pp. 1-4

[12] Yue L, Chen CC, Psychoudakis D, Volakis JL. Miniaturized 1: Dual-band GPS antenna element. In: 2010 IEEE Antennas and Propagation Society International Symposium. 2010. pp. 1-4. DOI: 10.1109/APS.2010.5561178

[13] Tang TC, Tsai CH, Lin KH, Huang YT, Chen CY. Fractal GPS antenna design on piezoelectric substrate. In: 2010 Asia-Pacific Microwave Conference. 2010. pp. 991-994

[14] Park J, Choi J, Lee TH, Kim YS. Small GPS antenna with lumped elements and interdigital structure. IEEE Radio and Wireless Symposium. 2012:55-58. DOI: 10.1109/RWS.2012.6175393

[15] Vallozzi L, Vandendriessche W, Rogier H, Hertleer C, Scarpello ML. Wearable textile GPS antenna for integration in protective garments. In: Proceedings of the Fourth European Conference on Antennas and Propagation. 2010. pp. 1-4

[16] Mariottini F, Albani M, Toniolo E, Amatori D, Maci S. Design of a Compact GPS and SDARS integrated antenna for automotive applications. IEEE Antennas and Wireless Propagation Letters. 2010;**9**:405-408. DOI: 10.1109/LAWP.2010.2049632

[17] Yap MS, Ng L, Aditya S. A triple band antenna for GSM and GPS application. In: Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia. 2003. pp. 1119-1123. DOI: 10.1109/ICICS.2003.1292634

[18] Bilgic MM, Yegin K. Modified annular ring antenna for GPS and SDARS automotive applications. IEEE Wireless and Propagation Letters. 2016;**15**:1442-1445. DOI: 10.1109/LAWP.2015.2512558

[19] Wang M, Deng X, Gu C. Tri-band GPS antenna with low-multipath ground plane. In: 2010 International Symposium on Signals, Systems and Electronics. 2010. pp. 1-4

[20] Basilio LI, Chen RL, Williams JT, Jackson DR. A new planar dual-band GPS antenna designed for reduced susceptibility to low-angle multipath. IEEE Transactions on Antennas and Propagation. 2007;**55**:2358-2366. DOI: 10.1109/TAP.2007.901818

[21] Lee Y, Ganguly S, Mittra R. Multi-band L5-capable GPS antenna with reduced backlobes. In: IEEE Antennas and Propagation Society International Symposium. 2005. pp. 438-441

[22] Zhou Y, Chen CC, Volakis JL. A single-fed element antenna for tri-band anti-jamming GPS arrays. In: IEEE Antennas and Propagation Society International Symposium. 2008. pp. 1-4

[23] Zhou Y, Chen CC, Volakis JL. Dual band proximity-fed stacked patch antenna for tri-band GPS applications. IEEE Transactions on Antennas and Propagation. 2007;**55**:220-223. DOI: 10.1109/TAP.2006.888476

[24] Zhou Y, Koulouridis S, Kiziltas G, Volakis JL. A novel 1.5 quadruple antenna for tri-band GPS applications. IEEE Antennas and Wireless Propagation Letters. 2006;**5**:224-227. DOI: 10.1109/LAWP.2006.875282

[25] Boccia L, Amendola G, Di Massa G. A high performance dual frequency microstrip antenna for global positioning system. IEEE Antennas and Propagation Society International Symposium. 2001. pp. 66-69

[26] Zhang X, Ma X, Lai Q. Two kind of conical conformal GPS antenna arrays on projectile. In: 2009 3rd IEEE International Symposium On Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. 2009. pp. 659-662

[27] Zhang X, Ma X, Guan X. Design of cylindrical conformal microstrip GPS antenna arrays. In: International Conference on Microwave Technology and Computational Electromagnetics. 2009. pp. 105-108

[28] Bang JH, Enkhbayar B, Min DH, Ahn BC. A compact GPS antenna for artillery projectile applications. IEEE Antennas and Wireless Propagation Letters. 2011;**10**:266-269. DOI: 10.1109/LAWP.2011.2135830

[29] Lingling Z, Jinghui Q, Bo S. Design of GPS antenna array on revolving cylinder. In: 2007 Asia-Pacific Microwave Conference. 2007. pp. 1-4

[30] Wu K, Zhang L, Shen Z, Zheng B. An anti-jamming 5-element GPS antenna array using phase-only nulling. In: 6th International Conference on ITS Telecommunications. 2006. pp. 370-373

[31] Kuester C, Chaloupka H, Knauth J. A high temperature superconducting anti-jam GPS antenna array. In: MILCOM 1999. 1999. pp. 671-674. DOI: 10.1109/MILCOM.1999.822768

[32] Kasemodel JA, Chen CC, Gupta IJ, Volakis JL. Miniature continuous coverage wideband GPS antenna array. In: IEEE Antennas and Propagation Society International Symposium. 2008. pp. 1-4. DOI: 10.1109/APS.2008.4619966

[33] Fernandez Prades C, Closas Gomez P, Fernandez Rubio JA. New trends in global navigation systems: Implementation of a GPS antenna array receiver. In: Proceedings of the Eighth International Symposium on Signal Processing and Its Applications. 2005. pp. 903-906. DOI: 10.1109/ISSPA.2005.1581085

[34] Gupta IJ. Non-planar adaptive antenna arrays for GPS receivers. IEEE Antennas and Propagation Magazine. 2010;**52**:35-51. DOI: 10.1109/MAP.2010.5687504

[35] Sisman I, Yegin K. Reconfigurable antenna for jamming mitigation of legacy GPS receivers. International Journal of Antennas and Propagation. 2017; Article ID 4563571. DOI: 10.1155/2017/4563571

[36] Hsu WH, Pan SC, Huang YT. A well placement design to improve GSM/DCS and GPS antenna isolation. In: Proceedings of the 9th International Symposium on Antennas, Propagation and EM Theory. 2010. pp. 8-11. DOI: 10.1109/ISAPE.2010.5696381

[37] Yegin K. AMPS/PCS/GPS active antenna for emergency call systems. IEEE Antennas and Wireless Propagations Letters. 2007;**6**:255-258. DOI: 10.1109/LAWP.2007.897510

[38] Zhu N, Ziolkowski RW. Metamaterial-inspired, near-field resonant parasitic GPS antennas: Designs and experiments. In: IEEE International Symposium on Antennas and Propagation. 2011. pp. 658-660

[39] Pigeon M, Morlaas C, Aubert H. A dual-band high impedance surface mounted with a spiral antenna for GNSS applications. In: IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications. 2011. pp. 994-997

[40] Bao XL, Ruvio G, Ammann MJ, John M. A novel GPS patch antenna on a fractal hi-impedance surface substrate. IEEE Antennas and Wireless Propagation Letters. 2006;**5**:323-326. DOI: 10.1109/LAWP.2006.878900

[41] Yigit O, Akkaya İ, Oztürk Y, Yegin K. Plasma frequency selective surface for GPS application. In: 2017, TELFOR Conference; Belgrade, Serbia

[42] CST Microwave Studio AG, Germany. 2017. www.cst.com [Accessed: 26.12.2017]

[43] Inpaq. Ceramic GNSS Patches [Internet]. 2017. www.inpaq.com [Accessed: 26.12.2017]

[44] Amotech. Ceramic GNSS Patches [Internet]. 2017. www.amotech.com [Accessed: 26.12.2017]

[45] Harxon. HX-CH6017A [Internet]. 2017. www.harxon.com [Accessed: 26.12.2017]

[46] Serantel. QHA Antenna 2017. www.farnell.com [Accessed: 26.12.2017]

[47] Abracon. Chip antennas [Internet]. 2017. www.abracon.com [Accessed: 26.12.2017]

[48] Taoglass. MAT.12A dual-chip antenna [Internet]. 2017. www.taoglass.com [Accessed: 26.12.2017]

[49] Leica. AT 504 and AR 25 [Internet]. 2017. www.leica.com [Accessed: 26.12.2017]

[50] Tallysman Wireless Inc. VeraPhase™ 6000 [Internet]. 2017. www.tallysman.com [Accessed: 26.12.2017]

[51] Yegin K. Design an ultra low-noise S-band amplifier. Electronic Design News. June 7, 2012

[52] Yegin K. On-glass automotive diversity antenna and LNA design for S-band satellite digital radio. International Journal of Electronics. 2015;**102**:1804-1817. DOI: 10.1080/00207217.2014.996783

[53] Curran JT, Fernández-Prades C, Morrison A, Bavaro M. Innovation: The continued evolution of the GNSS software-defined radio. GPS World. 2018

GPS in Monitoring Operation

Ionosphere Variability in Low and Mid-Latitudes of India Using GPS-TEC Estimates from 2002 to 2016

Sridevi Jade and Shrungeshwara T.S.

Additional information is available at the end of the chapter

Abstract

Continuous Global Positioning System (cGPS) observations spanning 14 years at 24 cGPS sites located in low and mid-latitudes (5–35° N) in the Indian subcontinent are analyzed to extract the ionosphere delay from one-way residuals for each satellite. Absolute integrated electron content (IEC) is the integral of electrons per m^2 along the line of sight between the satellite and receiver. Total electron content (TEC) is determined from IEC using elevation mapping function to normalize the variation of the ray path length through the ionosphere based on the GPS satellite elevation angle. In this study, GPS TEC estimates temporally cover two solar cycles (23 and 24) and spatially cover equatorial ionization anomaly (EIA) region and beyond, thus depicting the ionosphere variability in space, time and geographical location. Results capture different phases of solar cycle, EIA, annual, daily, diurnal and seasonal variability of ionosphere in northern hemisphere. This chapter gives significant insights in to the high and random variability of TEC associated with the changes in solar activity, intensity of the sun radiation, zenith angle at which they impinge the earth's atmosphere, equatorial electrojet (EEJ) and plasma flow.

Keywords: global positioning system (GPS), ionosphere, total electron content (TEC), solar radiation, equatorial ionization anomaly (EIA)

1. Introduction

Ionosphere consists of layers of earth's atmosphere containing free electrons as a result of ionization of the atoms in this region by high energy from sun and cosmic rays. These layers of free electrons surrounding the earth from 60 to 1100 km altitude influence the GPS signal propagation, causing errors in positioning. Total electron content (TEC) is estimated from the dual frequency GPS receiver signals by extracting the phase advances and code delays caused

by ionosphere. Precise TEC estimates give significant insights into the variability of ionosphere in space, time, geographical location and solar and cosmic activity. GPS based ionosphere research was initiated globally [1–6] for large scale studies, local earthquakes, mine blasts, and so on. Spatial and temporal variability of ionosphere based on GPS-TEC estimates was studied by several researchers [7–18] using GPS data in different regions of the world giving insights into the response of ionosphere due to the variations in solar activity, geomagnetic storms, and so on. Ionosphere maps for few regions were prepared from the GPS-TEC estimates from a network of stations.

In India, GPS based ionosphere studies were initiated after the establishment of dual frequency GPS stations in 2003 by Indian Space Research Organization and Airport Authority of India as a part of the GAGAN (Geo And GPS Augmented Navigation) program. For the first time in India, spatial and temporal variability [19] of equatorial ionosphere is studied using GPS-TEC estimates for a 16-month period (March 2004–June 2005) with low sunspot activity (LSSA) using 18 GPS station data covering a geomagnetic range of 1° S to 24° N. Using the same GAGAN network GPS data [20], estimates of GPS TEC were compared with the International Reference Ionosphere (IRI) predicted TEC values. They have also investigated diurnal, seasonal and annual variability of ionosphere over Indian subcontinent during the 16-month LSSA period. For a low latitude station Rajkot located near the equatorial ionization anomaly crest region in India [21], ionosphere variability during LSSA period (2005–2007) was investigated to give insights into solar activity dependence and effects of geomagnetic storm on GPS-TEC. Variability of GPS-TEC at a single station Udaipur in Rajasthan for a period of 2005–2010 was studied [22] and the result of seasonal variations are compared with IRI-2007 Model. Similarly diurnal and seasonal variation of GPS-TEC at a single station Agra, for the LSSA period (2006–2009) was studied [23]. GPS-TEC estimates for Surat GPS station [24, 25] were compared with model predictions from IRI-2007 and IRI-2012 and the ionosphere variability was investigated. GPS-TEC derived [26, 27] from a chain of Indian stations for a 1 year period (2011–2012) was used to study the diurnal, seasonal and latitude variability and its relation to geomagnetic storms, solar eclipse, and so on. They gave comparison of GPS-TEC with IRI-2012, Standard Plasmasphere-Ionosphere Model (SPIM) and Global Ionospheric Maps (GIM).

All the above studies so far reported in the Indian subcontinent were for a period of 1–2 years over single and network of GPS stations. For the first time we report the GPS-TEC estimates for a period spanning 14 years (2002–2016) covering solar cycle 23 (1996–2008) and 24 (2008–2019) from a network of about 24 cGPS stations (**Figure 1**; **Table 1**) with geodetic latitude ranging from 5 to 35° N and longitude ranging from 70 to 95° E in Indian subcontinent. New set of cGPS station data is used for the present study compared to majority of earlier ionosphere studies which use GAGAN network data and hence give an independent estimate of ionospheric TEC in this region. The geomagnetic latitude and longitude of these GPS stations (**Table 1**) is 0–26° N and 145–168° E which is very important for the study of ionosphere variability as equatorial region has the high ionosphere activity compared to the rest of the regions in the world. In addition, for the first time TEC estimates are reported for region beyond the EIA region in the Indian subcontinent using cGPS data. Annual, spatial, seasonal, diurnal variability of ionosphere is presented using these TEC estimates and its relation to solar activity, EEJ, EIA is investigated.
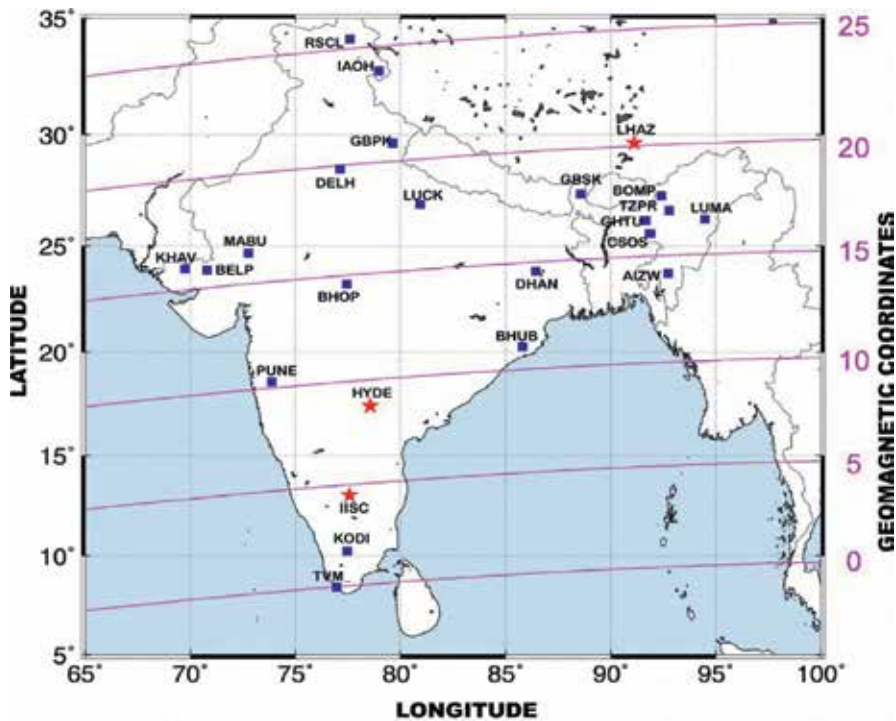
Figure 1. cGPS stations used in the study with geomagnetic latitude lines. Geomagnetic equator (0° N) passes through the bottom tip of Indian subcontinent and northern crest of EIA is located at geomagnetic latitude of 15° N. Red stars are IGS (International GNSS service) sites and blue squares are cGPS sites.

Station	Site code	Data span	Geodetic latitude (° N)	Geodetic longitude (° E)	Geomagnetic latitude (° N)	Geomagnetic longitude (° E)
Trivandrum	TVM_	2002–2005	8.42	76.97	−0.02	149.95
Kodaikanal	KODI	2002–2015	10.23	77.47	1.73	150.59
Bengaluru	IISC	2002–2015	13.02	77.57	4.49	150.93
Hyderabad	HYDE	2002–2015	17.42	78.55	8.76	152.24
Pune	PUNE	2002–2005	18.56	73.88	10.29	147.85
Bhubaneswar	BHUB	2002–2012	20.26	85.79	11.09	159.39
Aizwal	AIZW	2003–2006	23.72	92.73	14.18	166.17
Dhanbad	DHAN	2004–2005	23.82	86.44	14.58	160.24
Bhopal	BHOP	2004–2005	23.21	77.45	14.59	151.68
Bela Temple	BELP	2010–2011	23.87	70.80	15.84	145.46
Khavda	KHAV	2010–2011	23.92	69.77	15.98	144.48
Shillong	CSOS	2002–2008	25.57	91.86	16.05	165.43
Mount Abu	MABU	2010–2011	24.65	72.78	16.42	147.41
Lumami	LUMA	2003–2015	26.22	94.48	16.60	167.92

Station	Site code	Data span	Geodetic latitude (° N)	Geodetic longitude (° E)	Geomagnetic latitude (° N)	Geomagnetic longitude (° E)
Guwahati	GHTU	2003–2012	26.15	91.66	16.64	165.28
Tezpur	TZPR	2002–2013	26.62	92.78	17.06	166.35
Bomdilla	BOMP	2004–2013	27.27	92.41	17.72	166.04
Lucknow	LUCK	2002–2005	26.89	80.94	17.98	155.29
Panthang	GBSK	2003–2014	27.37	88.57	17.99	162.45
Delhi	DELH	2003–2005	28.48	77.13	19.85	151.87
Lhasa	LHAZ	2002–2015	29.66	91.10	20.15	164.94
Almora	GBPK	2002–2014	29.64	79.62	20.81	154.30
Hanle	IAOH	2002–2015	32.78	78.97	23.97	153.99
Leh	RSCL	2002–2012	34.13	77.60	25.42	152.91

Table 1. Details of cGPS stations used in the present study.

2. Data and methods

cGPS data, during 2002–2016, of about 24 cGPS stations (**Table 1**; **Figure 1**) located in the Indian subcontinent with data span ranging from 2 to 13 years with sampling interval of 30 s has been used. The dataset span is more than any previous study till date and spatially covers the length and breadth of the Indian subcontinent. Details of the cGPS sites and the data used are listed in **Table 1** and the data is analyzed using GAMIT software [28]. Data sampling interval of 30 s and elevation cut-off angle of 20° is used for the analysis. The quality check of GPS data at each station was done using TEQC software [29] to remove data with several cycle slips, multipath and span of less than 18 h. The daily data of all the stations after quality check is analyzed using GAMIT to extract the ionospheric delays suffered by GPS signals in L-band with frequency f_1 (1.57542GHz) and f_2 (1.2276GHz) from the one-way residuals of each satellite at 30 s interval. These one-way residuals for each satellite are output after cleaning the observables for cycle slips, multipath, outliers during the analysis. The *IEC* along the line of sight (LOS) between the satellite and receiver are calculated from the carrier phase delays L_1 and L_2 and the group delays P_1 and P_2 (code pseudo-ranges) in range units obtained from GAMIT analysis as detailed earlier. The absolute *IEC* [30] is given by

$$IEC = \frac{\lambda_2}{A} \frac{\left(f_1^2 f_2^2\right)}{\left(f_1^2 - f_2^2\right)} \left(B - L_G\right) \tag{1}$$

where, ambiguity constant B is given by

$$B = \frac{1}{n} \sum_{i=1}^{n} \left(P_{Gi} - L_{Gi}\right) \tag{2}$$

$$L_G = \frac{(L_1 - L_2)}{\lambda_2} \qquad (3)$$

$$P_G = \frac{(P_2 - P_1)}{\lambda_2} \qquad (4)$$

where, n is number of phase measurements in a given arc [1]. L_G and P_G are linear combinations as given above whereas L_G is precise and smooth with unknown phase ambiguity constant and P_G is noisy and less precise and not ambiguous [2, 31]. To estimate the absolute and precise estimate of IEC, we fit L_G on P_G using ambiguity constant B. λ_1 and λ_2 are the wavelengths of the L-band GPS signals. IEC is given in TEC units where 1 TECU = 10^{16} electrons/m^2. To calculate vertical equivalent TEC along the satellite receiver path, elevation mapping function ($emf\theta$) is used to account for the variation in the ray path length (L_θ) based on GPS satellite elevation angle θ varies with the orbital pass of each satellite as given below:

$$emf\theta = \frac{H_{ion}}{L_\theta}. \qquad (5)$$

$$L_\theta = \sqrt[2]{(R + H_t)^2 - R^2 \cos^2(\theta)} - \sqrt[2]{(R + H_b)^2 - R^2 \cos^2(\theta)} \qquad (6)$$

where, H_{ion} is the mean ionosphere thickness, R is earth's radius and H_t and H_b are the top and bottom altitudes of the ionosphere layer. Vertical equivalent TEC along the satellite receiver path in TEC units is given by

$$\text{TEC} = IEC \times emf\theta \qquad (7)$$

TEC is computed at 30 s interval during the orbital pass of the each satellite at the each GPS station for all the 24 cGPS stations during 2002–2016. Two-sigma iterated average of TEC at 30 s interval is computed from the TEC of all the visible satellites at that epoch. GPS-TEC thus estimated is used to discuss the ionosphere variability over Indian subcontinent.

3. Ionosphere variability

Ionosphere is highly variable in space (geographical location) and time (solar cycle, seasonal, diurnal) and with solar-related ionospheric disturbances and earthquakes. About 15 cGPS sites are located in the equatorial ionization anomaly (EIA) region from geomagnetic equator to northern crest of EIA region (17° N) where the low latitude ionosphere exhibits annual, spatial, seasonal and diurnal variability. Nine cGPS sites are located in mid-latitude region beyond the EIA region in northern India and Himalaya. Using the GPS-TEC estimates, variability of ionosphere is discussed in the subsequent sections.

3.1. Inter-annual

There are about 12 sites with data span covering the two solar cycles. Daily mean value of GPS-TEC is plotted for these stations from 2002 to 2016 in **Figure 2** to study the annual variability of ionosphere over the 14-year period. The last solar cycle 23 lasted for 12.3 years starting in August 1996 and ending in December 2008 with peak solar activity between 2001 and 2005 at low- and mid-latitude regions. The current solar cycle 24 began on January 4, 2008 with minimal solar activity till early 2010 and had two peaks in 2011 and early 2014. TEC at almost all the stations indicate the peak (2002–2005) and descending phase (2005–2008) of solar cycle 23. This is followed by low values of TEC during 2008–2010 consistent with minimum solar activity of current solar cycle 24 and high values of TEC during 2011 to 2014 consistent with peak solar activity. The TEC variation is higher for the sites (KODI, IISC, HYDE, BHUB, CSOS) which are located in the EIA region with a peak TEC value of 100–120 TECU in 2002 and 60–80 TECU in 2003. For the rest of the sites (GBPK, IAOH, RSCL, GBSK, BOMP, TZPR, LHAZ) located beyond the EIA region, the TEC variability is low and depends upon the geographic location of these sites which is discussed in detail in the subsequent sections. Daily mean value of TEC plotted in **Figure 2** indicates significant semi-annual and annual cycles. Sensitivity of TEC to solar activity is stronger at low latitudes when compared to mid-latitudes in the northern hemisphere.

3.2. Spatial variability

For the Indian subcontinent, geomagnetic equator passes through the southern bottom tip and the northern crest of EIA (15° N geomagnetic latitude) lies in the middle (**Figure 1**) providing a unique opportunity for studying the ionosphere variability. Since we have several stations and a larger spread of data, we give detailed in-depth study of spatial variation of ionosphere from 5 to 35° N latitude (0–26° N geomagnetic latitude) and 70 to 95° E longitude. Also since the data covers different phases of two solar cycles, the results are given separately for each solar cycle.

3.2.1. Solar cycle 23

TEC variability for all the available days of December 2004, March, June, September and December 2005 representing the solstice, equinox, summer and winter seasons are plotted in **Figures 3**–**5** for cGPS stations with geomagnetic latitude of 0–10° N, 10–17° N, >17° N to study the spatial, diurnal and seasonal variability in EIA region, northern crest of EIA and beyond. To study in detail the variability along the latitude, 10 sites were chosen with approximately same longitude (TVM_, KODI, IISC, HYDE, BHOP, DELH, GBPK, IAOH, RSCL) starting from geomagnetic equator covering the EIA region and beyond for the descending phase of solar cycle 23 from December 2004 to 2005. The peak value of diurnal TEC increases with latitude from geomagnetic equator (Trivandrum, 50 TECU) to Northern crest of EIA region (Bhopal, 80 TECU) and then decreases gradually from Delhi (50 TECU) to Leh (40 TECU) station located beyond the EIA region. Moreover, the maximum value of diurnal TEC (**Figure 3**) is for a longer duration for the stations (TVM_, KODI, IISC, HYDE) close to the geomagnetic equator when compared to the
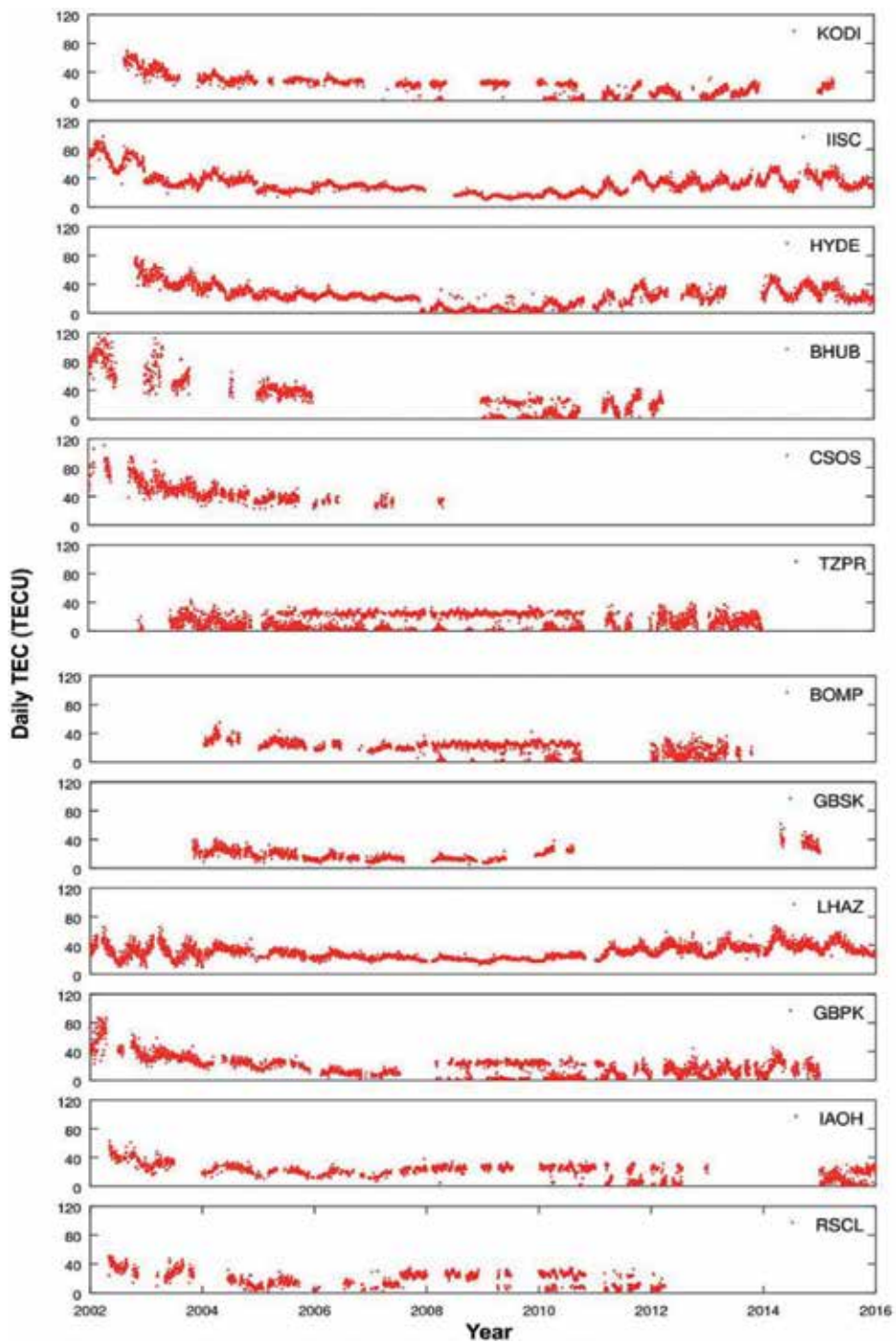
Figure 2. Daily mean GPS TEC values for 12 cGPS stations with increasing latitude.
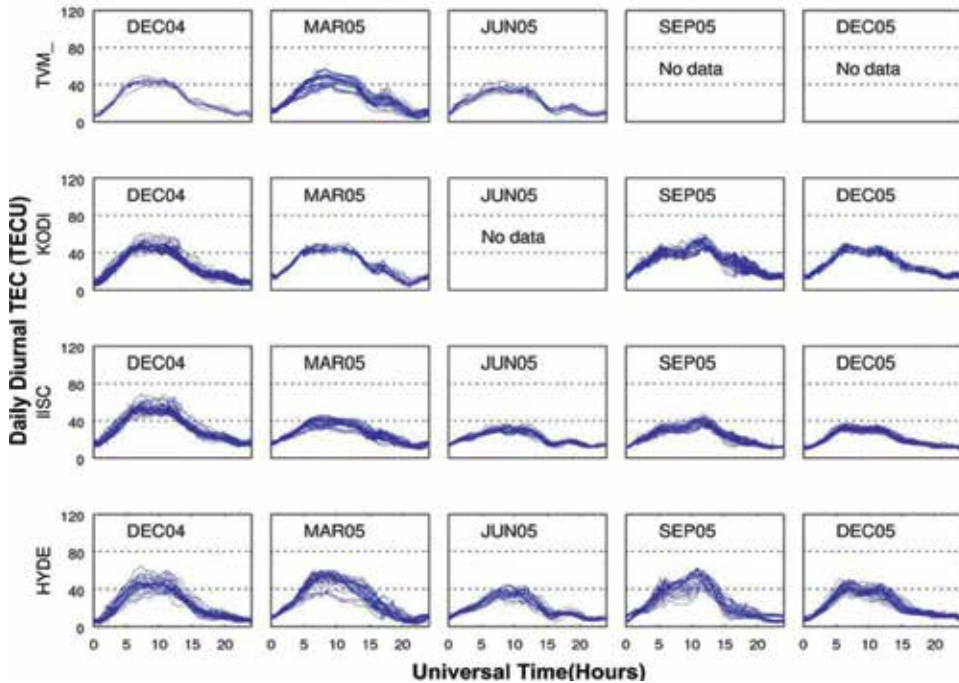
Figure 3. Daily diurnal GPS TEC during solar cycle 23 (2004–2005) for cGPS stations located between 0 and 10° N geomagnetic latitude.

stations located in crest of EIA region and beyond (**Figures 4** and **5**) which have more pronounced diurnal peaks. Conversely, the diurnal minima have longer duration and broad spread for the stations located in crest of EIA region and beyond as compared to the stations located close to geomagnetic equator. Day-to-day variability of diurnal TEC is more pronounced for the stations located in EIA region (**Figure 4**) when compared to the rest of the stations.

Pune and Bhubaneswar, Bhopal and Aizwal stations with 12 and 15° E longitude difference located in the northern crest of EIA region do not indicate any consistent variability related to longitude. Beyond EIA region DELH and BOMP, GBPK and LHAZ with 12 and 15° E longitude difference also do not show any significant longitude related variability during this period.

3.2.2. Solar cycle 24

The current solar cycle started in January 2008 and had minimum solar activity between the year 2008 and early 2010 followed by ascending solar activity. GPS TEC for March, June, September and December for 3 years (2009, 2010, 2011) for all available station data are plotted in **Figures 6–8** to study the spatial and temporal variability of the ionosphere during the low and ascending phase of current solar cycle. During the minimum solar activity year of 2009 stations (**Figure 6**)with increasing latitude (IISC, GHTU, GBSK, LHAZ) indicate marginal increase of TEC value (35–40 TECU) in EIA region from IISC to GHTU and marginal decrease beyond EIA region (GBSK, LHAZ). In 2010 (**Figure 7**), TEC value gradually increases from 40
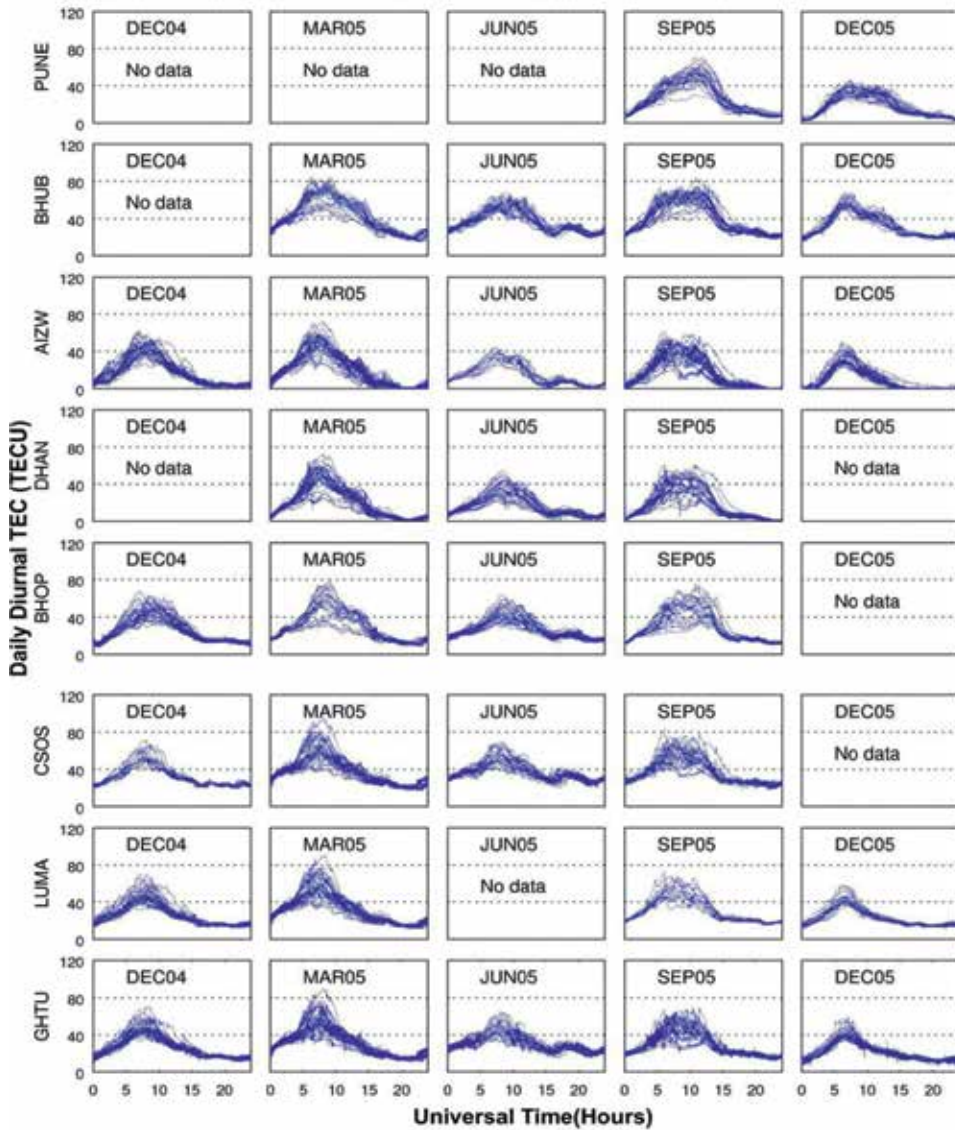
Figure 4. Daily diurnal GPS TEC during solar cycle 23 (2004–2005) for cGPS stations located between 10 and 17° N geomagnetic latitude.

to 60 TECU with increase in latitude in EIA region and marginal decrease beyond EIA region. In 2011 (**Figure 8**), peak TEC values ranging from 40 to 80 TECU are observed at IISC and the rest of the stations located in EIA with no consistent variation with latitude. Day-to-day variability of TEC is more pronounced in the EIA region during 2010 and 2011 when compared to 2009. Diurnal peak TEC value is for longer duration at Bengaluru and Hyderabad whereas pronounced peaks are observed in EIA region and beyond. Diurnal minima are for longer duration for the stations located in EIA region and beyond. Spatial variability of ionosphere between
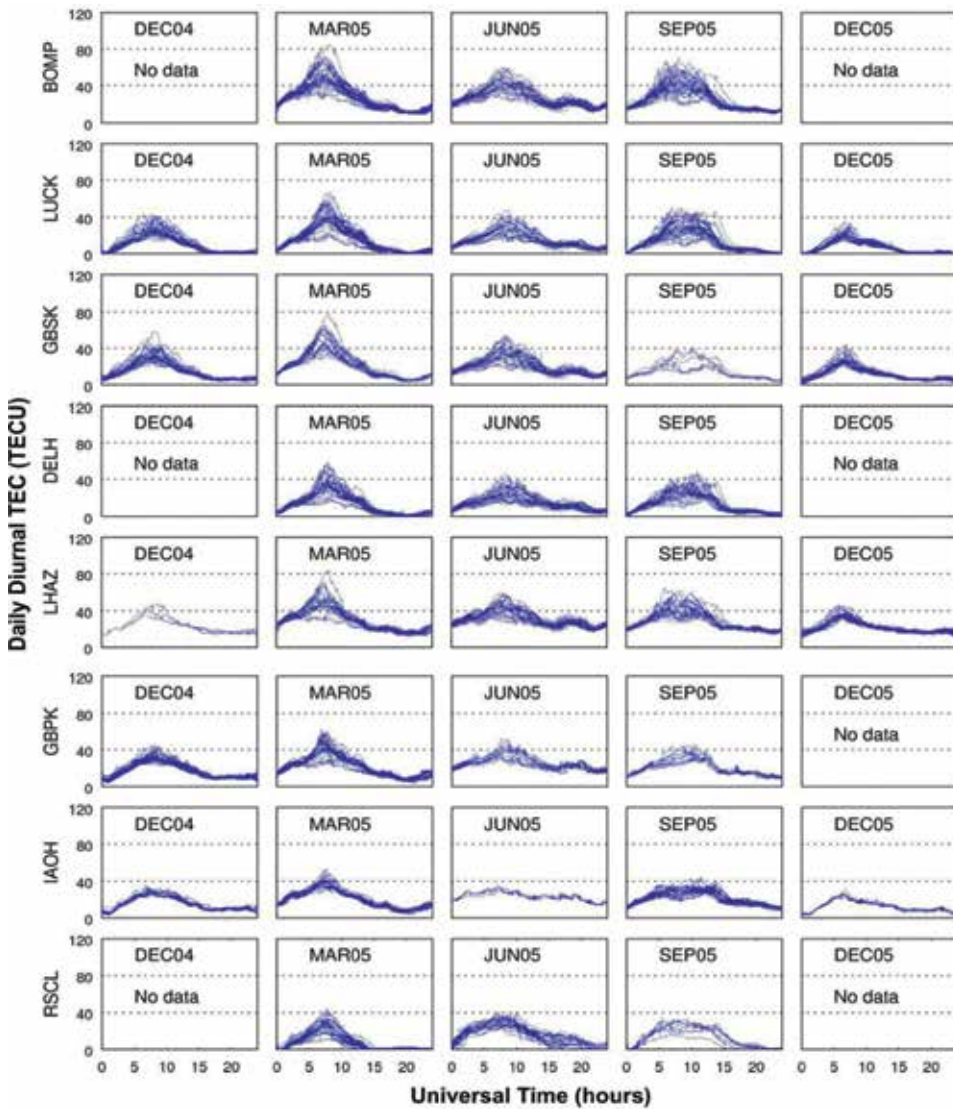
Figure 5. Daily diurnal GPS TEC during solar cycle 23 (2004–2005) for cGPS stations above 17° N geomagnetic latitude.

geomagnetic latitude of 0–17° N is not very pronounced during the current solar cycle when compared to the duration of diurnal peak and minimum TEC values.

During the ascending solar activity period of 2011, diurnal peak TEC values of 60 TECU are observed at KHAV, MABU and BELP stations (about 70° E longitude) when compared to diurnal peak TEC values of 80 TECU observed at GHTU and LUMA stations (90° E longitude). Distinct peaks are observed with increasing longitude. This indicates that during ascending phase of solar cycle the ionosphere increases with longitude difference of 19–25° E in the EIA region.
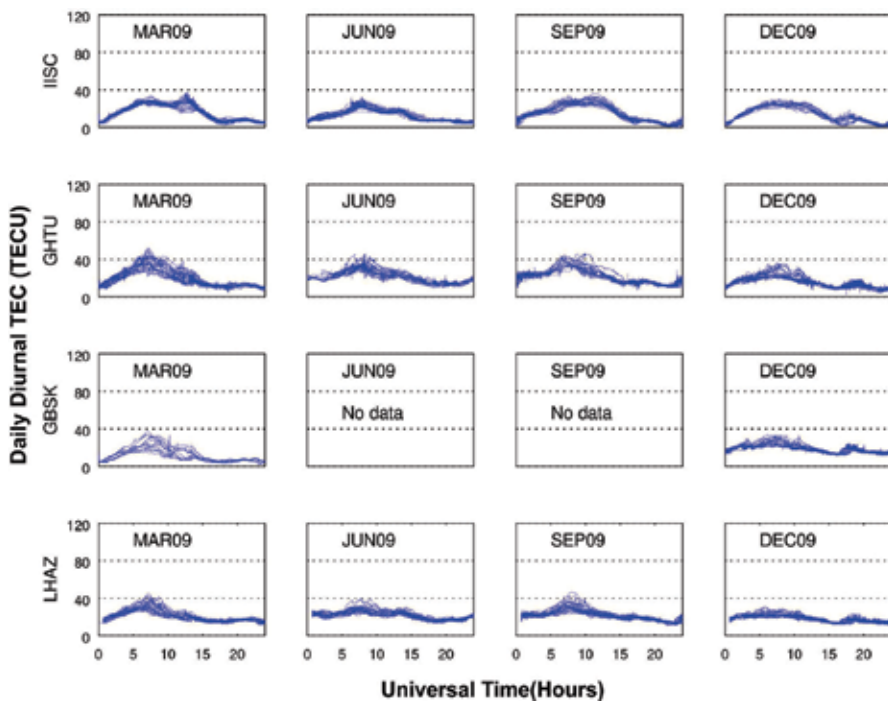
Figure 6. Daily diurnal GPS TEC during solar cycle 24 for cGPS stations with increasing latitude during minimum solar activity year 2009.

3.3. Diurnal variability

Diurnal variability of TEC depends on the Sun's orbit, changes in solar activity and intensity of radiance, earth's magnetic field and dynamics of neutral winds (diffusion of transported electrons from the equator). Plasma flow associated with the EIA effects the day-to-day variability of diurnal TEC for the stations located in EIA region. Geomagnetic and seismo-ionosphere disturbances also effect the day-to-day variability of diurnal TEC. Results of diurnal variability for two solar cycles is given below.

3.3.1. Solar cycle 23

Diurnal variability of TEC at all the stations in **Figures 3–5** shows the minima during the night hours between 17 and 24 h UT and increasing TEC from 0 h UT to peak at midday between 8 and 11 h UT. The highest peak value of diurnal TEC (80 TECU) is observed at stations (BHOP, BHUB, CSOS, LUMA, GHTU) located in northern crest of EIA region and the lowest peak (20 TECU) is recorded at RSCL and IAOH located in Ladakh Himalaya beyond the EIA region. Moreover, the maximum value of diurnal TEC has longer duration (5–12 h UT) for the stations (TVM_, KODI, IISC, HYDE) close to geomagnetic equator when compared to the stations located in EIA region and beyond which have more pronounced diurnal peaks. Conversely, the diurnal minima have longer duration (15–24 h UT) and broad spread for the stations
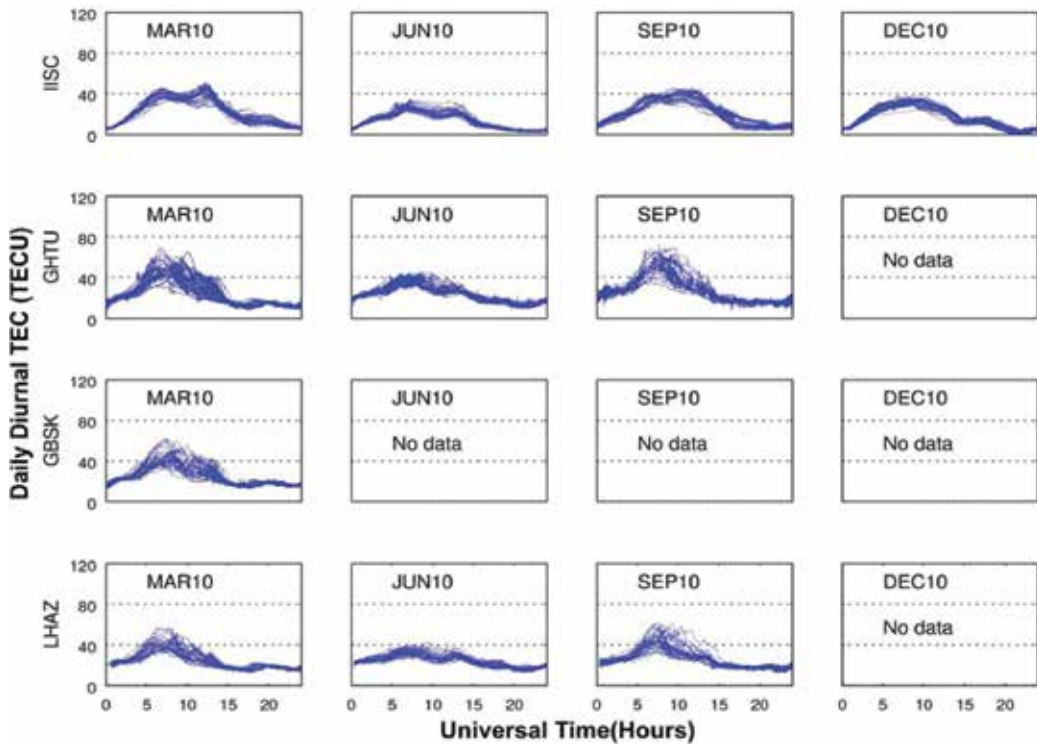
Figure 7. Daily diurnal GPS TEC during solar cycle 24 for cGPS stations with increasing latitude during low solar activity year 2010.

located in EIA region and beyond as compared to the stations located close to the geomagnetic equator. Day-to-day variability of diurnal TEC is more pronounced for the stations located in EIA region when compared to the rest of the stations.

Seasonal variability of diurnal TEC is clearly depicted in **Figures 3–5** with low TEC values during summer solstice (June) and high values for winter solstice and equinoxes for all the stations from geomagnetic equator to the northern crest EIA region. Beyond the EIA region, higher values of diurnal TEC with pronounced peaks were observed during the equinox month of March and lower TEC values during equinox month of September. For Trivandrum located on the geomagnetic equator, diurnal TEC values are the highest for March equinox whereas for Kodaikanal and Bengaluru the higher values of TEC are observed for December 2004 in winter. The lowest TEC was observed at Bengaluru during the solstice month of June. For the rest of the stations located in EIA region, the highest TEC value was observed during March and September equinox. Maximum spread of day-to-day variability of diurnal TEC is observed in the EIA region during the equinox months of March and September. Higher TEC values during December 2004 is due to the increase of electrons in winter caused by the transport of neutral constituents from summer to the winter hemisphere. This in turn increases the anomaly crest development in the winter. In addition, this may be also due to the seismo-ionospheric disturbance caused by Mw of 9.2, 26 December 2004 Sumatra earthquake [32] which affected the cGPS stations in southern India.
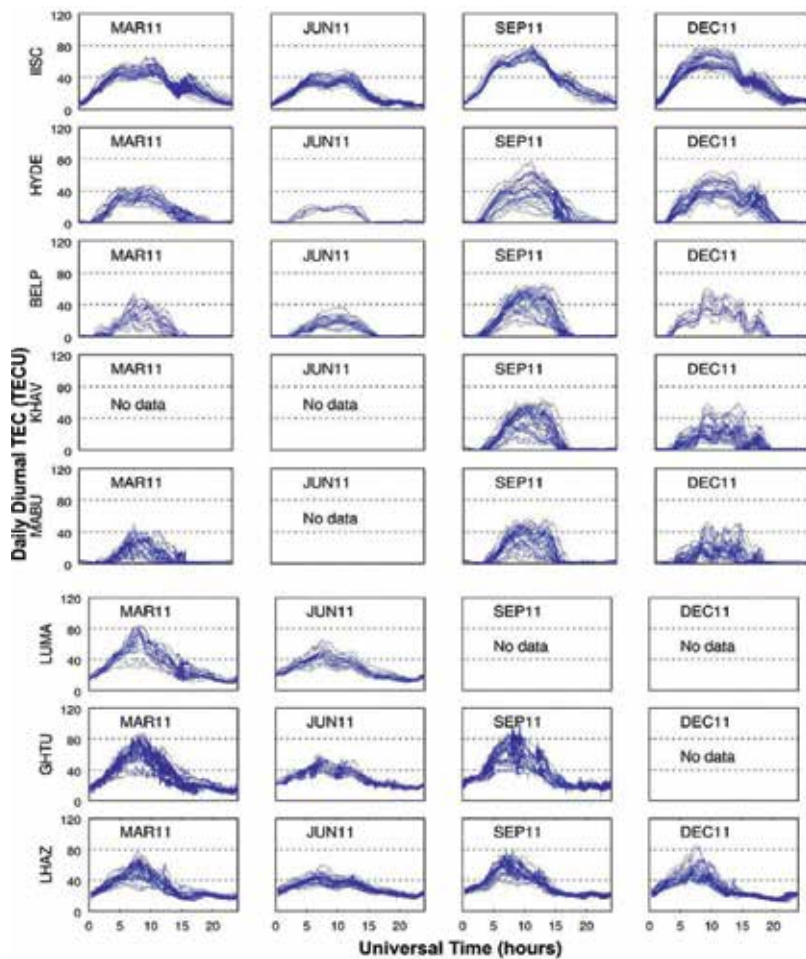
Figure 8. Daily diurnal GPS TEC during solar cycle 24 for cGPS stations with increasing latitude during ascending solar activity year 2011.

3.3.2. Solar cycle 24

Diurnal peak TEC values from 2009 to 2011 (**Figures 6–8**) increase with the increase in the solar activity from a minimum peak value of 30–40 TECU during 2009; 30–60 TECU in 2010; 40–80 TECU during 2011. Diurnal peak values occur for a long duration of 7–13 h UT for IISC located in the trough of EIA whereas for station located in the crest of EIA region pronounced diurnal peaks are observed between 7 and 10 h UT. Diurnal peak values are marginally higher for stations located in the crest of EIA regions. Diurnal minima for IISC occur during 20–24 h UT and for stations located in the crest of EIA region diurnal minima occur during 15–24 h UT. Gujarat stations (KHAV, BELP, MABU) recorded anomalous daily diurnal variation with very high day-to-day variability. For these three stations, diurnal peaks are not very distinct and the diurnal minima suddenly drops in the night hours and remains constant (15–24 and 0–2 h UT). Diurnal peak values are the highest with the maximum spread during equinox months of March and September for stations located in the northern crest (17° N geomagnetic latitude)
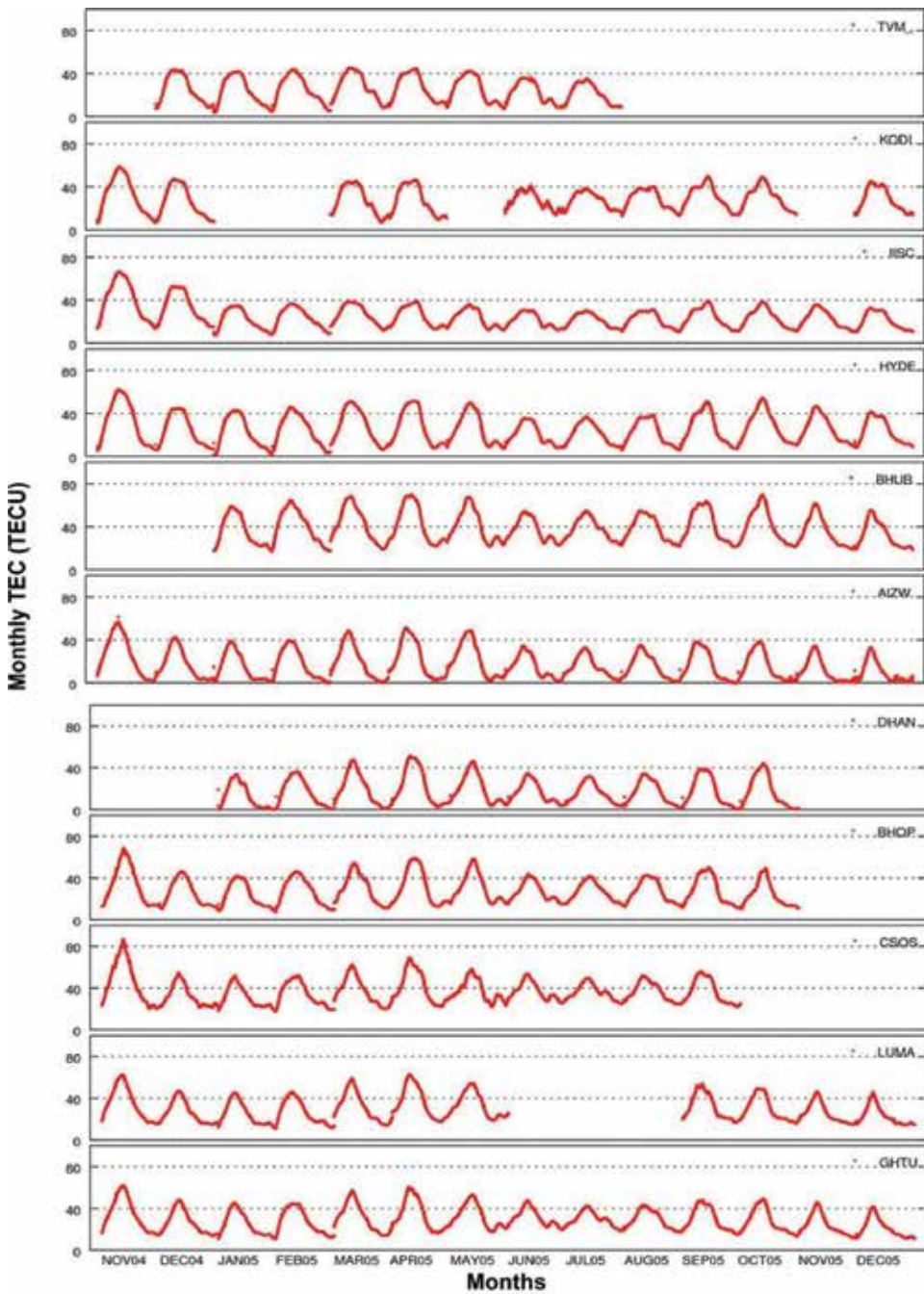
Figure 9. Monthly mean diurnal GPS TEC with increasing latitude for cGPS stations between 0 and 17° N geomagnetic latitude during solar cycle 23 from November 2004 to December 2005.
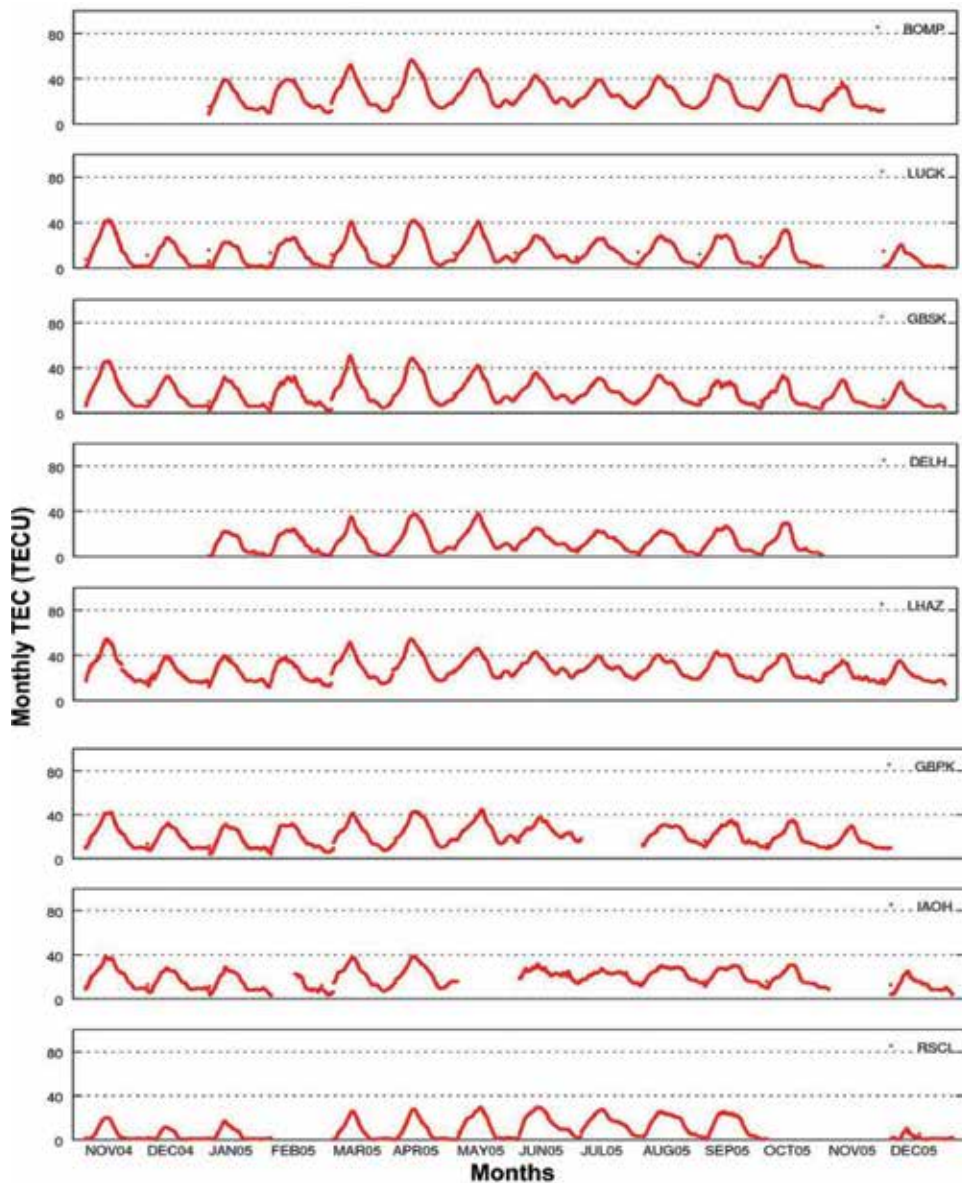
Figure 10. Monthly mean diurnal GPS TEC with increasing latitude for cGPS stations above 17° N geomagnetic latitude during solar cycle 23 from November 2004 to December 2005.

of EIA region. For Bengaluru and Hyderabad (5–9° N geomagnetic latitude) station high diurnal peak values are observed during equinox month of September and solstice month of December. Daily diurnal TEC values at LHAZ and GHTU stations (**Figure 8**) indicate very
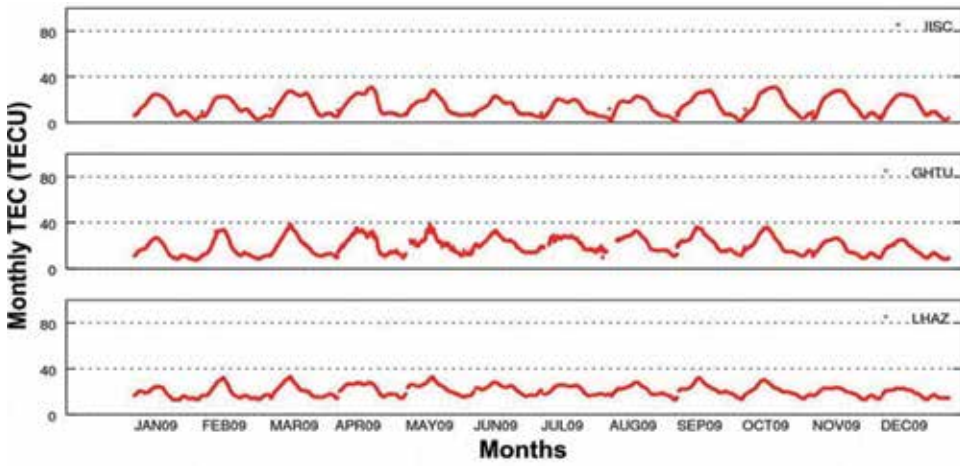
Figure 11. Monthly mean diurnal GPS TEC for cGPS stations with increasing latitude during solar cycle 24 for minimum solar activity year 2009.
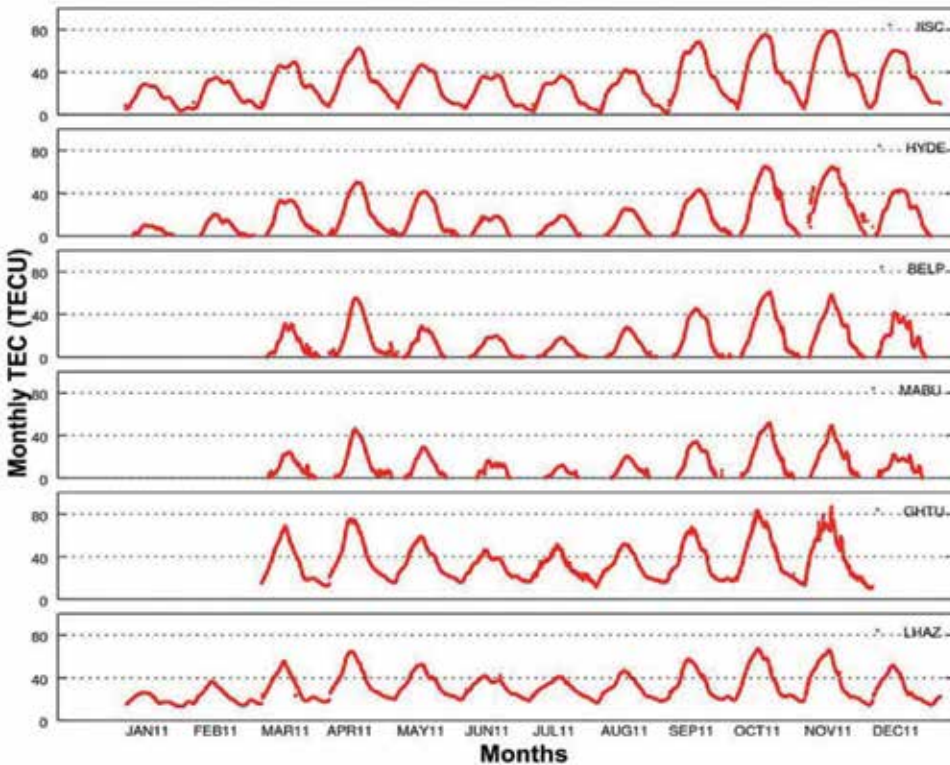

Figure 12. Monthly mean diurnal GPS TEC for cGPS stations with increasing latitude during solar cycle 24 for ascending solar activity year 2011.

high and random variability related to seismo-ionospheric disturbance due to Mw 6.9, 18 September 2011 Sikkim earthquake. It can be observed that the diurnal variability depends on the solar activity, solar radiance, geomagnetic field, latitude, longitude and plasma flow related to EIA effects.

4. Monthly and seasonal variability

4.1. Solar cycle 23

Monthly diurnal mean values of TEC are plotted from November 2004 to December 2005 for all the stations between geomagnetic equator and northern crest of EIA (0–17° N geomagnetic latitude) in **Figure 9** and beyond EIA region in **Figure 10**. The highest peak TEC values for all the sites in EIA region distinctly occur in November 2004 with a value of about 86 TECU for CSOS, 70 TECU for Bhopal to about 58 TECU for Kodaikanal. The lowest peak TEC values occur during the months of June, July, August for these sites with about 50 TECU at CSOS, 43 TECU for Bhopal to 35 TECU for Bengaluru. The highest peak TEC value during winter solstice and the lowest peak TEC value during summer solstice are due to seasonal anomaly prevalent in winter hemisphere due to increase in electrons caused by meridional neutral winds. For the sites beyond the EIA region (**Figure 10**), the highest peak TEC values (35–50 TECU) occur in the summer equinox months (March, April, May) and winter solstice month of November 2004 and the lowest peak TEC values (20–30 TECU) occur in January.

4.2. Solar cycle 24

Monthly diurnal mean TEC values are plotted from January to December for low solar activity period of 2009 (**Figure 11**) and ascending solar activity period of 2011 (**Figure 12**). Monthly and seasonal cycle is not very pronounced during the low solar activity period of 2009 with marginally higher peak values recorded during October for IISC (35 TECU) and March for GHTU (40 TECU) and lower peaks during January (20–30 TECU) in the EIA region. For the ascending solar activity period of 2011, the monthly and seasonal variation is distinct with the highest (80 TECU) in EIA region during October and November and the lowest during January (20–30 TECU) consistent with the winter anomaly observed in the northern hemisphere.

5. Summary

For the first time an in-depth study of ionosphere variability in low and mid-latitude region using TEC estimates from 24 cGPS stations for a 14-year (2002–2016) period is carried out. The cGPS data covers peak (2002–2004), descending phase (2005–2008) of solar cycle 23 and minimum (2008–2010), ascending phase (2011–2016) of current solar cycle 24.

Inter-annual variability of GPS-TEC depicts the peak, descending phase of solar cycle 23 and minimum, ascending phase of solar cycle 24. Maximum TEC values are observed during 2002–2004 and minimum TEC values are observed during 2008–2010. GPS TEC indicates a distinct daily, monthly, semi-annual and annual cycle. Sensitivity of TEC to solar activity is prominent in EIA region compared to mid-latitudes in northern hemisphere. TEC values recorded are consistent with large-scale electrodynamics associated with the equatorial electrojet (EEJ), plasma fountain, equatorial ionization anomaly (EIA), equatorial wind and temperature anomaly, which affect the ionosphere variability at equatorial and low latitude regions. The high variability of equatorial and low latitude ionosphere are due to the perfect horizontal alignment of the geomagnetic field lines at the dip equator and the shifting between the geographic and geomagnetic equator.

GPS-TEC values increase from geomagnetic equator to the crest of EIA region (17° N geomagnetic latitude) after which they gradually decrease toward mid-latitudes in the northern hemisphere. Latitude variability of ionosphere is more pronounced during the high solar activity years (2002–2004) when compared to low solar activity years (2008–2010). Diurnal peak TEC value has longer duration between 0 and 9° N geomagnetic latitude. Diurnal maxima have pronounced peaks and diurnal minima is observed for longer duration in the northern crest of EIA region and beyond. Ionosphere variability with longitude is observed for longitude difference of 19° E and above during the ascending phase of current solar cycle 24. Normally, solar radiation strikes the atmosphere more obliquely with increasing latitude decreasing its intensity and production of free electrons, whereas near the geomagnetic equator its strikes horizontally with eastward electric field during day and westward during night. This causes plasma diffusion along magnetic field lines at approximately ±15° geomagnetic latitudes forming crests on both the hemispheres (EIA region). Hence, TEC increases gradually from geomagnetic equator to the EIA crest, beyond which it decreases toward the mid-latitude regions. Intensity of EIA and its latitude of crest development vary with the strength of EEJ, season and solar activity. Our study indicates that the northern crest of EIA region extends up to about 17–18° N geomagnetic latitude in Indian region.

Diurnal variability of ionosphere depends on the intensity of solar activity, season and strength of geomagnetic field with high TEC values recorded in 2004 and 2011 and low values in 2009. Day-to-day variability is more pronounced for the high solar activity years when compared to low solar activity years. Maxima occurs during midday (7–13 h UT) with longer duration for geomagnetic latitudes between 0 and 9° N and pronounced peaks for greater than 9° N. *Minima* occurs after midnight (20–24 h UT) between 0 and 9° N geomagnetic latitude whereas it is for longer duration (15–24 h UT) for northern crest of EIA region and beyond. Day-to-day variability of maxima is more pronounced in the crest of EIA regions (9–18° N geomagnetic latitude). Day-to-day variability of diurnal TEC is high and random during December 2004 and September 2011 due to seismo-ionosphere disturbance caused by 2004 Sumatra and 2011 Sikkim earthquake. Also anomalous day-to-day TEC variability is observed for Gujarat stations (MABU, KHAV, BELP) in 2011 which needs further detailed study. Diurnal maxima and minima vary significantly during the equinox and solstice of summer and winter seasons with lower values during summer solstice in EIA region and higher values during equinox and winter solstice. Beyond EIA (>18° N), maxima with pronounced peak occurs in the equinox

month of March. Maximum spread of diurnal maxima is observed in crest of EIA during the equinox month of September. The results indicate that the variability of diurnal TEC in low-latitude region is highly random as it is caused by several factors as detailed earlier.

Monthly diurnal mean TEC values are the highest in November and the lowest in the months June to August for solar cycle 23 and increase with latitude in the EIA region. This is due to the winter anomaly observed in the EIA region of northern hemisphere and is consistent with previous studies. Beyond EIA region, the high values are observed in the summer equinox months and November and minimum values occur during January. For the current solar cycle 24, the monthly and seasonal variability is marginal for the low solar activity year (2009) when compared to 2011. In the EIA region, the highest values are recorded during October-November and the lowest during January for ascending phase (2011) of current solar cycle 24. The seasonal and monthly variation is random depending upon the intensity of solar cycle and seasons in each year.

In summary, the temporal and spatial variability of equatorial, low and mid-latitude ionosphere reported using the GPS-TEC estimated from new GPS data during 2002–2016 are broadly consistent with previous studies globally and specific to the Indian subcontinent. When compared to previous studies, present study with longer data span and spatial spread gives significant insights into the randomness of day-to-day variability of ionosphere as detailed above. This high and random variability of TEC is due to the changes associated with solar activity, intensity of the sun radiation and zenith angle at which they impinge the earth's atmosphere. TEC variability on quiet days depends on the changes in Earth's magnetic field and EEJ strength. In equatorial and low-latitude region of Indian subcontinent there is intense east–west electric current (EEJ) due to neutral winds and the plasma flow associated with the EIA plays a significant role in the day-to-day variability of diurnal TEC. Ionosphere is also affected by solar and geomagnetic storms, solar eclipse, seismic disturbances, volcanic eruptions, tsunamis, and so on. Indian Space Research Organisation in collaboration with Airports Authority of India developed a model to predict TEC in the Indian region which can be used to provide TEC maps. They have used GAGAN (GPS Aided Geo Augmented Navigation) ground network of 18 stations for this model and predict TEC between 8 and 30° N latitude and 60–100° E longitude. Since the present study uses a new set of cGPS data for a 14 year period, benchmarking ISRO ionosphere model with the current data and combining with the current TEC estimates would give an opportunity to develop precise ionosphere models and maps for this region. In addition these GPS-TEC estimates can be used to model the spatial and temporal variability of the low and mid latitude ionosphere specific to Indian subcontinent. GPS TEC study has several applications in varied fields such as precise positioning, navigation, seismo-ionosphere coupling, propagation of radio waves and solar-terrestrial events.

Acknowledgements

Conflict of interest

Authors declare that there is no conflict of interest.

Notes/thanks/other declarations

Notes: We used total of 55 cGPS stations data during 2002–2016 to estimate TEC. We have chosen about 24 cGPS data with common epochs to given the comparative study of ionosphere variability in this chapter.

Author details

Sridevi Jade* and Shrungeshwara T.S.

*Address all correspondence to: sridevi@csir4pi.in

CSIR-4PI, CSIR Fourth Paradigm Institute (Formerly CSIR-CMMACS), Bangalore, India

References

[1] Mannucci AJ, Wilson BD, Edwards CD. A new method for monitoring the Earth's ionospheric total electron using the GPS global network. In: Proceedings of the Institute of Navigation GPS-93. Alexandria, Va: Inst. of Navig; 1993. p. 1323

[2] Sylvander M, Feigl KL, Souriau A, Blelly PL. Absolute ionospheric vertical electron content inferred from transmissions of the global positioning system. Comptes rendus de l'Académie des Sciences. Serie II a. 1995;**320**:793-799

[3] Juan JM, Rius A, Hernandez-Pajares M, Sanz J. A two-layer model of the ionosphere using Global Positioning System data. Geophysical Research Letters. 1997;**24**:393

[4] Calais E, Minster JB. GPS detection of ionospheric perturnbations following the January 17, 1994. Northridge earthquake. Geophysical Research Letters. 1995;**22**:1045-1048

[5] Calais E, Minster JB. GPS detection of an ionospheric perturbation following a space shuttle ascent. Geophysical Research Letters. 1996;**23**:1897-1900

[6] Davis K, Hartmann GK. Studying the ionosphere with the global positioning system. Radio Science. 1997;**32**:1695-1703

[7] Goodwin GL, Silby JH, Lynn KJW, Breed AM, Essex EA. GPS satellite measurements: Ionospheric slab thickness and total electron content. Journal of Atmospheric and Terrestrial Physics. 1995;**57**(14):1723-1732

[8] Goodwin GL, Silby JH, Lynn KJW, Breed AM, Essex EA. Ionospheric slab thickness measurements using GPS satellites in southern Australia. Advances in Space Research. 1995;**15**(2):125-135

[9] Ho CM, Mannucci AJ, Lindqwister UJ, Pi X, Tsurutani BT. Global ionosphere perturbations monitored by the worldwide GPS network. Geophysical Research Letters. 1996;**23**: 3219-3222

[10] Mannucci AJ, Wilson BD, Yuan DN, Ho CH, Lindqwister UJ, Runge TF. A global mapping technique for GPS-derived ionospheric total electron content measurements. Radio Science. 1998;**33**:565-585

[11] Afraimovich EL, Ding F, Kiryushkin VV, Astafyeva EI, Jin S, Sankov VA. TEC responce to the 2008 Wenchuan earthquake in comaparison with other strong earthquakes. International Journal of Remote Sensing. 2010;**31**(13):3601-3613

[12] Wu C, Fry CD, Liu JY, Liou K, Tseng CL. Annual TEC variation in the equatorial anomaly region during the solar minimum: September 1996-August 1997. Journal of Atmospheric and Solar—Terrestrial Physics. 2004;**66**:199-207

[13] Wu S, Zhang K, Yuan Y, Wu F. Spatio-temporal characteristics of the ionospheric TEC variation for GPSnet-based real-time positioning in Victoria. Journal of Global Positioning Systems. 2006;**5**(1–2):52-57. DOI: 10.5081/jgps.5.1.52

[14] Jee G, Schunk RW, Scherliess L. On the sensitivity of total electron content (TEC) to upper atmospheric/ionospheric parameters. Journal of Atmospheric and Solar—Terrestrial Physics. 2005;**67**:1040-1052. DOI: 10.1016/j.jastp.2005.04.001

[15] Fayose RS, Babatunde R, Oladosu O, Groves K. Variation of total electron content (TEC) and their effect on GNSS over Akure, Nigeria. Applied Physics Research. 2011;**4**(2):105-109. DOI: 10.5539/apr.v4n2p105

[16] Purohit PK, Bhawre P, Mansoori AA, Khan PA, Gwal AK. GPS derived total electron content (TEC) variations over Indian Antarctic Station, Maitri. World Academy of Science, Engineering and Technology. 2011;**59**:597-599

[17] Andrade D, Lopez E. Studies of TEC in Ecuador using global positioning system (GPS) data. Sun and Geosphere. 2014;**9**(1–2):37-40

[18] Guo J, Li W, Liu X, Kong Q, Zhao C, Guo B. Temporal-spatial variation of global GPS-derived total electron content, 1999-2013. PLoS ONE. 2015;**10**(7):e0133378. DOI: 10.1371/journal.pone.0133378

[19] Rama Rao PVS, Gopi Krishna S, Niranjan K, Prasad DSVVD. Temporal and spatial variations in TEC using simultaneous measurements from the Indian GPS network of receivers during the low solar activity period of 2004–2005. Annales de Geophysique. 2006;**24**:3279-3292

[20] Bhuyan PK, Borah RR. TEC derived from GPS network in India and comparison with the IRI. Annales de Geophysique. 2007;**39**:830-840

[21] Bagiya MS, Iyer KN, Joshi HP, Iyer KN, Aggarwal M, Ravindran S, Pathan BM. TEC variations during low solar activity period (2005–2007) near the equatorial ionospheric anomaly crest region in India. Annales Geophysicae. 2009;**27(3)**:1047-1057

[22] Galav P, Dashora N, Sharma S, Pandey R. Characterization of low latitude GPS-TEC during very low solar activity phase. Journal of Atmospheric and Solar-Terrestrial Physics. 2010;**72**:1309-1317. DOI: 10.1016/j.jastp.2010.09.017

[23] Chauhan V, Singh OP. A morphological study of GPS-TEC data at Agra and their comparison with the IRI model. Advances in Space Research. 2010;**46**:280-290

[24] Karia S, Pathak K. Change in refractivity of the atmosphere and large variation in TEC associated with some earthquakes, observed from GPS receiver. Advances in Space Research. 2011;**47**:867-876. DOI: 10/1016/j.asr.2010.09.019

[25] Patel NC, Karia SP, Pathak KN. Comparison of GPS-derived TEC with IRI-2012 and IRI-2007 TEC predictions at Surat, a location around the EIA crest in the Indian sector, during the ascending phase of solar cycle 24. Advances in Space Research. 2016;**8**:2

[26] Panda SK, Gedam SS, Jin S. In: Jin S, editor. Ionospheric TEC Variations at Low Latitude Indian Region, Satellite Positioning—Methods, Models and Applications. Intech Open. 2015;**8**:149-174. DOI: 10.5772/59988. Available from: https://www.intechopen.com/books/satellite-positioning-methods-models-and-applications/ionospheric-tec-variations-at-low-latitude--indian-region

[27] Panda SK, Gedam SS, Rajaram G. Study of ionospheric TEC from GPS observations and comparisons with IRI and SPIM model predictions in the low latitude anomaly Indian subcontinental region. Advances in Space Research. 2015;**55**:2948-1964

[28] Herring TA, King RW, McClusky SC. Documentation of the GAMIT GPS Analysis Software Release 10.4. Cambridge, MA: Department of Earth, and Planetary Sciences, Massachusetts Institute of Technology. 2010a

[29] Estey LH, Meertens CM. TEQC: The multi purpose toolkit for GPS/GLONASS data. GPS Solutions. 1999;**3(1)**:42

[30] Calais E, Bernard Minster J, Hofton MA, MAH H. Ionospheric signature of surface mine blasts from Global Positioning System measurements. Geophysical Journal International. 1998;**132**:191-202

[31] Lanyi GE, Roth T. A comparison of mapped and measured total ionospheric electron content using global positioning system and beacon satellite observations. Radio Science. 1988;**23(4)**:483-492

[32] Jade S, Vijayan MSM, Gupta SS, Kumar PD, Gaur VK, Arumugam S. Effect of the M 9.3 Sumatra-Andaman islands earthquake of 26 December 2004 at several permanent and campaign GPS stations in the Indian continent. International Journal of Remote Sensing. 2007;**28**(13–14):3045-3054. DOI: 10.1080/01431160601094526

Edited by Rustam B. Rustamov
and Arif M. Hashimov

Today, Global Positioning System (GPS) has taken a significant place in human life with wide-scale applications. It is a multi-use, space-based radio-navigation system, embracing defense and security, civil, commercial, and scientific research needs. Taking into account the importance of the system, it has been suggested to present current advances of GPS with attention focused on vital aspects of technology.

This book provides a general description of GPS, GNSS, and GLONASS with reference to the improvement of characteristics of the segments, including software applications and the equipment itself containing GPS. The book also provides information on application areas. Equally, a number of common shortcomings and errors are described, and elimination and mnimization of such consequences are demonstrated.

IntechOpen