URSI-RCRS 2024, ARIES/GEHU, Bhimtal, India, 22 - 25 October, 2024

A Novel Approach to Global Ionosphere Delay Estimation with Constrained Optimization

Gurpreet Singh*(1), Megha Maheshwari(1) and Nirmala Srini(1)
(1) U R Rao Satellite Center, ISRO, Bangalore, India, e-mail: *gurpreet.singhj29@gmail.com
megham@ursc.gov.in; nirmalas@ursc.gov.in

Abstract

The ionosphere is a section of the atmosphere consisting of free electrons and ions. Its ionized nature has an impact on the transmission of radio signals that pass through it. Consequently, in order to determine the user's position accurately, ionosphere error corrections are necessary for a satellite navigation system. The (NavIC) is a self-sufficient regional navigation system designed to provide user positions for both single- and dual-frequency users. To make sure that user positioning is as accurate as possible, NAVIC provides two types of ionosphere corrections: grid-based and ionosphere coefficient-based on the L5 frequency. The Kriging approach is utilized in NAVIC to create an ionosphere grid throughout the NAVIC service region. Nevertheless, researchers worldwide often utilize ionosphere grid data that is based on spherical harmonics. Kriging and spherical harmonics are both effective strategies for reconstructing global ionosphere data. In order to address this issue, the NavIC system offers real-time ionosphere error corrections across the Indian landmass, which is divided into 90 grids. By expanding the NavIC constellation beyond the Indian land mass, it will be possible to deliver services across the ocean and at crucial places where data insufficiency prevails. This necessitates the uninterrupted dissemination of ionosphere grid characteristics across the entire globe. Conventional estimating methods result in negative ionosphere delays in regions with limited measurement coverage areas. Using the values in such places can result in negative consequences for the users; hence, it is strongly recommended to rectify the issue of negative ionospheric estimates due to data unavailability and estimation by spherical harmonic fitting. To overcome the limitations, we have presented a novel methodology that uses the constraint- based approach included with the spherical harmonics technique to generate the grid parameters. Mathematically, one submatrix is identified and solved during estimation of coefficients in such a way that it eliminates the unwanted negative values and brings it to within the logical range of the parameters without affecting its accuracy around the measurements. In addition, the method gives a more accurate estimate of the bad flag area, where the root mean square deviation (RMSD) dropped from 6.11 m to 3.89 m when compared to the exact global data.

It has been observed that our approach reduces the number of unwanted bad flag grids to zero. In this way, this technique guarantees a better and more logical range of ionosphere delay throughout the day for the users.

1. Introduction

Global Navigation Satellite System (GNSS) is widely used for Positioning and Timing (PNT) services. The simplicity and affordability of single-frequency GNSS receivers make them accessible tools for applications ranging from consumer device to industrial and scientific uses. Despite providing lesser accuracy, single-frequency GNSS receivers are more popular commercially than dual frequency receivers due to its low cost. Single Frequency Receiver (SFR) plays crucial role in various applications, including navigation on earth and beyond, surveying, etc. Due to its extensive uses, all the GNSS control centers provides broadcast parameters for both single and dual frequency users.

For the computation of position, the ionosphere introduces the highest error for SFR users. It is one of the crucial parameters for getting accurate user positions from single frequency GNSS receivers. Therefore, for the single frequency user, the ionosphere correction or delay is broadcast in terms of Grid parameters or coefficients. The broadcasted parameter/coefficents are the modelled Total Electron Content (TEC) and is computed in near real time by using dual frequency reference receiver measurements. With the planning of Navigation with Indian Constellation's (NavIC) future global expansion, the need arises to generate real time global ionosphere error correction in terms of grid parameters. There are various techniques, such as B-spline, kriging, spherical harmonics, etc., that are available in the literature to generate the vertical grid based ionosphere corrections [1-3].

Due to its continuity characteristics for global systems, the International GNSS Service (IGS) chose the spherical harmonics model among many others [3]. They use various data processing techniques that use data from several days in an estimation model to provide a valid range of ionosphere delay. However, they removed a few grid points and computed values over the same grid points. Therefore, the generalization of the technique was not tested on the measurements.

In general, ionosphere delay is estimated using measurements from ground-based reference receiver stations and broadcasted as grid parameters or coefficients to the users. The precision of the output is related to the number of stations and the distribution of data from the stations. Due to poor distribution of data (receivers are only on land) and measurement noise, the estimated ionosphere delay using the traditional techniques can be negative for some areas. Currently, to overcome the data distribution limitations, various agencies use data from several days in postprocessed mode to generate the grid ionosphere delay, which cannot be used for real-time systems such as NavIC [1] [3].

2. Methodology

In the case of approximating ionospheric behavior, the popular techniques are kriging for regional and spherical harmonics (SH) for global data as the basis for building the functional relationship between TEC and spatial coordinates. The SH model is based on minimizing the error

$$\boldsymbol{J}(x) = \frac{||A\boldsymbol{x}-\boldsymbol{b}||^2}{2} \qquad \text{Eq. 1}$$

where A is design matrix of SH with order NxP, N is total number of measurements and P is equal to $(k+1)^2$, where k is order of SH

$$A = \begin{bmatrix} SH_{11}(\theta_1,\phi_1) & SH_{11}(\theta_1,\phi_1) & \dots & SH_{1P}(\theta_1,\phi_1) \\ SH_{21}(\theta_2,\phi_2) & SH_{22}(\theta_2,\phi_2) & \dots & SH_{2P}(\theta_2,\phi_2) \\ \vdots & \vdots & \vdots & \vdots \\ SH_{N1}(\theta_N,\phi_N) & SH_{N2}(\theta_N,\phi_N) & \dots & SH_{NP}(\theta_N,\phi_N) \end{bmatrix}_{NXP}$$

Eq. 2

b is vector of TEC measurements at IPP coordinates (θ, ϕ) with order Nx1

$$\boldsymbol{b} = \begin{bmatrix} TEC_1(\theta_1,\phi_1) \\ TEC_2(\theta_2,\phi_2) \\ \vdots \\ TEC_N(\theta_N,\phi_N) \end{bmatrix}_{NX1} \qquad \text{Eq. 3}$$

and x is unknown vector of coefficients to be estimated.

$$\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_P \end{bmatrix}_{\text{PX1}} \qquad \text{Eq. 4}$$

Also, to deal with the negative estimated values, the above unconstrained optimization problem in equation 1 is put under constraints. These constraints make the estimation algorithm give meaningful estimated TEC values that are not negative. The constrained optimization problem is formulated below, where a sub-matrix is identified corresponding to IPP locations where estimated values are negative after solving the optimization problem in equation 5.

$$\boldsymbol{J}(x) = \frac{||Ax-b\,||^2}{2}$$
$$\textit{subject to } \; Gx \;\geq\; c \qquad \text{Eq. 5}$$

where G is a matrix identified from the design matrix A of SH with negative estimated values and c is small non-negative vector.
Equation 5 defines a constrained optimization problem with one inequality constraint. Karush-Kahn-Tucker (KKT) conditions are the result of using Lagrange multipliers to solve such issues [6]. We introduce a penalty to the objective function **J**(x) to convert the constrained problem to an unconstrained optimization problem. The equation given below denotes and defines the new objective function, which goes by the name Lagrangian.

$$\boldsymbol{L}(x,\lambda) = J(x) \;+\; \lambda^T(Gx \;-c) \qquad \text{Eq. 6}$$

where λ is the vector of Lagrange multipliers.

The KKT conditions are obtained by minimizing the Lagrangian function defined in equation (13).
The following equations give the KKT conditions:

a) **Stationarity Condition**:

$$\frac{\partial \boldsymbol{L}(x,\lambda)}{\partial x} = 0 \qquad \text{Eq. 7}$$

a) **Feasibility Condition**:

$$\frac{\partial \boldsymbol{L}(x,\lambda)}{\partial \lambda} = 0 \qquad \text{Eq. 8}$$

To establish the minima, the corresponding sufficient conditions are Hessian matrices satisfying the following:

$$\frac{\partial^2 \boldsymbol{L}(x,\lambda)}{\partial x^2} \geq 0 \qquad \text{Eq. 9}$$
$$\frac{\partial^2 \boldsymbol{L}(x,\lambda)}{\partial \lambda^2} \geq 0 \qquad \text{Eq. 10}$$

These KKT conditions, when applied to the Lagrangian $\boldsymbol{L}(x,\lambda)$ defined in equation 6 give us the following set of equations:

a) Stationarity:

$$\nabla J(x) \;+\; \lambda^T \nabla(Gx \;-\; c) \;\;=\;\; 0$$

$$A^T(Ax-b) \;+\; \lambda^T G \qquad = 0 \qquad \text{Eq. 11}$$

b) Primal and Dual Feasibility:

$$Ax \;=\; b$$

$$Gx \;-c \;\geq 0 \qquad \text{Eq. 12}$$

c) Complementarity Condition:

$$\lambda \;\geq\; 0$$
$$\lambda^T\,(Gx \;-\; c) = \;0 \qquad \text{Eq. 13}$$

By recasting these conditions as the Linear Complementarity Problem (LCP), it is possible to find the KKT conditions' solution [7-8]. Then, these KKT conditions are transformed into standard form of linear complementarity problem. The standard form of linear complementarity problem (LCP) is defined as finding vectors **w** and **z** where

$$\mathbf{w} = \mathbf{Mz} + \mathbf{q} \text{ with } \mathbf{w} \geq 0 \perp \mathbf{z} \geq 0 \qquad \text{Eq. 14}$$

The KKT conditions can be put in standard form of LCP by identifying following matrices and vectors

$$M = \begin{bmatrix} A^T A & G \\ G & 0 \end{bmatrix} \qquad \text{Eq. 15}$$

$$z = \begin{bmatrix} x \\ \lambda \end{bmatrix} \qquad \text{Eq. 16}$$

$$q = \begin{bmatrix} -A^T b \\ -c \end{bmatrix} \qquad \text{Eq. 17}$$

The widely used algorithm known as Lemke's Method is used to solve the LCP in its standard form, found in Equation 14. Rearranged equations (15-17) can be organized into a tableau of the form.

Table 1 : Tableau to solve the LCP via Lemke's method.

w	z	
I	-M	q

3. Results and discussion

As explained in Section 2, the SH technique is applied to the measurements with respect to IPP points. Fig. 1 represents the root mean square difference (RMSD) between the measurements and the estimated VD for every hour of a day. It has been found that RMSD varies from 0.5 to 0.75 m at L1 frequency. The error quoted in various literatures when using IGS station measurements and SH to estimate the VD in post processed mode is 2-5 TECU [3]. One TECU is ~0.162 m at the L1 frequency. Therefore, our accuracy in estimation is within the desired range.

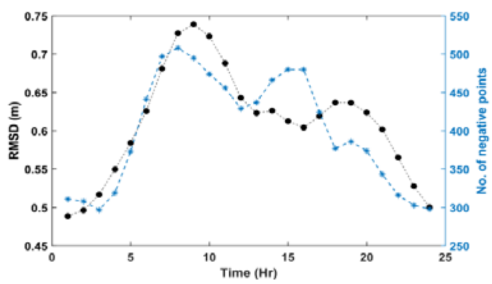

Fig. 1: Root mean square difference and the no. of negative points for every hour by using the spherical harmonics technique.

However, the right-side axis of the same figure shows the number of grid points that have negative ionosphere delay values when estimated using the SH algorithm. It varies from 297 to 508 over a day. In a normal scenario, these grid points will be flagged as bad, and users are advised not to use them through NavIC SIS ICD [5].

To avoid it, we included constraint-based algorithm along with SH. A detailed analysis and comparison are explained in the following paragraphs. For this purpose, a typical hour of 10 Universal Time (UT) is used to demonstrate the results, considering that the ionosphere variation is the maximum over the Indian region at this time.

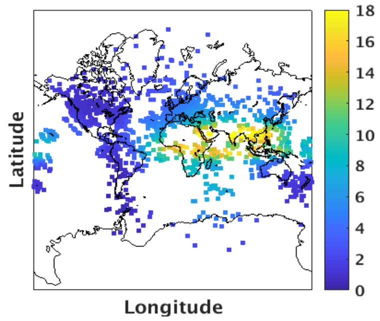

Fig. 2: Typical distribution of the simulated measurements at ionosphere pierce points computed using satellite – station geometry at 10 UT.

Fig. 2 shows the simulated global measurement distribution at IPP at 10 UT. The vertical delay in the data varies 0 to 18 m at L1 frequency. As seen in the figure, the measurement distribution at the Pacific Ocean is sparse. Fig. 3 shows the gridded vertical delay estimation using the SH technique. It is found that the SH VD also reaches up to 18 m. However, it is observed that the VD is negative at some places, which is shown by the white color plus sign in the same figure (Fig. 3). Therefore, the area covered with the white plus sign cannot be used by the users as it is. There are two ways to provide value to users. The first is to set all negative VD values to zero (SHZ) so that the user algorithm will consider no ionosphere correction in that area instead of a bad value flag. The other one is to generate VD using the method proposed in this paper. The VD estimated using the traditional SH method is called SHT, and the VD generated using constraint-based SH optimization algorithm is called as SHC throughout the paper.

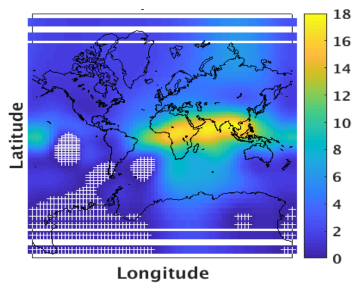

Fig. 3 : Typical ionosphere grid map using spherical harmonics method at 10 UT. The white colored plus sign is the area with negative values, which is called "bad flag" area.

Fig. 4 is the ionosphere delay map generated using the SHC technique. As shown in the figure, the SHC not only makes the number of bad flag points zero but also fill the area with logical values matching the nearby measurements.

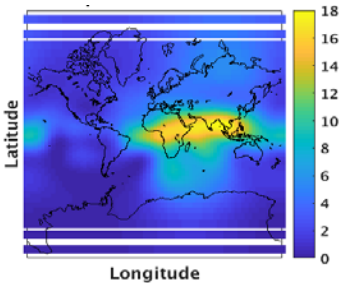

Fig. 4: Ionosphere grid map using optimization using spherical harmonics with constraints method at 10 UT. The map does not have any bad-flag area.

The close-up look of the bad flag area for the 10th hour is shown in Fig. 5.

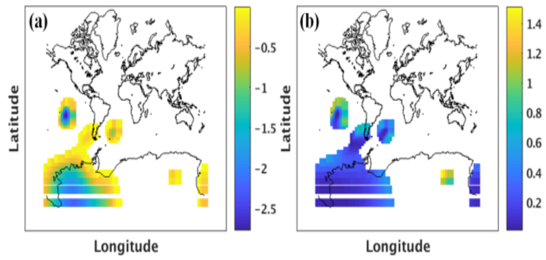

Fig. 5: Ionosphere delay values estimated using (a) SH and (b) SHC techniques over the bad flag area at 10 UT.

SHT provides the negative VD varies from -3m to -0.25 m. However, the SHC algorithm applied constraints during the estimation and converted it to the logical range of VD which varies from 0.1 to 1.5 m. The similar observations are there for full day data. The statistics in terms of RMSD and mean absolute error of the negative area VD are given in Table 2. The statistics is generated by comparing estimated VD with GIM post processed data. Table 2 provides data at 10th hour and full day

statistics. Moreover, statistics with SHZ is also shown in the table for better comparison. The RMSD reduces from 6.11 m to 3.89 m when SHC is applied instead of SHT. Not only is SHC better than SHT, our method gives better ionosphere correction as compared to the zero-ionosphere correction, which gives RMSD of 4.43 m. Further the SHC provides the lowest mean absolute error as compared to SHT and SHZ. Also, SHC promises to provide no bad flag which was around 9592 over a 24-hour period.

Table 2: Statistics of the model's performance over bad-flag area with respect to IGS Global Ionosphere Map data.

Time	RMSD (m)			Mean absolute error (m)			No. of negative points in SH	
	SHT	SHC	SHZ	SHT	SHC	SH-Z	SHT	SNC
Hr= 10	3.43	2.31	2.57	2.92	1.79	2.1	456	0
Full day	6.11	3.89	4.43	4.42	2.47	2.96	9592	0

In this way, it is proven that SHC is better technique to generate ionosphere corrections for NavIC as compared to the traditional estimation technique. Also, it provides the and assure seamless ionosphere delay service to real time NavIC users.

4. Conclusion

This study presents a new approach for optimization that takes into account positive TEC value restrictions. The use of the spherical harmonics technique in generating a global ionosphere map results in a negative ionosphere delay value, mostly caused by the inadequate distribution of observations. We have introduced an approach that limits the estimated value to a range that is possible for a physical system, such as the ionosphere. The constraint problem is resolved by employing the Linear Complementarity Problem (LCP). The root mean square error in the negative ionosphere delay area has been decreased from 6.11 m to 3.89 m. The provided solution completely eradicates the issue of negative ionosphere delay and significantly improves performance. This technique guarantees uninterrupted data flow, allowing users to smoothly access the data without encountering any default values or flags indicating the absence of values.

5. Acknowledgements

The authors would like to thank the Director, U.R. Rao Satellite Center (URSC), ISRO, Bangalore; Deputy Director, Communication and Power Area, URSC, ISRO; and Group Director, Space Navigation Group, URSC, ISRO, for their guidance, support, help, and encouragement. Members of the Space Navigation Group have also made helpful suggestions for this research work. The global ionosphere data from https://https://cddis.nasa.gov/ is also acknowledged.

References

1. S. Schaer, Mapping and predicting the earth's ionosphere using the Global Positioning System, 1999, Thesis, University of Berne.

2. R. Orus M. Pajares, J. M. Juan, and J. Sanz, Improvement of the global ionospheric VTEC maps by using kriging interpolation technique, 2005, J. Atmos. Sol. Terres. Phys.

3. Q. Zhang and Q. Zhao, Analysis of the data processing strategies of spherical harmonic expansion model on global ionosphere mapping for moderate solar activity, 2018, Advances in Space Research.

4. S. Gurpreet, M. Megha and S.Nirmala, Ionosphere Modelling using Spherical harmonics with in-equality constraints over IRNSS service area, 2022 URSI Regional Conference on Radio Science (USRI-RCRS), Indore, India, 2022.

5. A. S. Ganeshan, 2014, Indian Regional Navigation Satellite System, signal in space ICD for standard positioning service, ISRO-IRNSS-ICD-SPS-1.0.

6. K. G. Murty, Linear Complementarity, Linear and Nonlinear Programming, 1988, Helderman-Verlag.

7. R. W. Cottle, J. S. Pang, R. E. Stone, The Linear Complementarity Problem, 2009, Classics in Applied Mathematics, ISBN: 978-0-89871-686-3.