A MARS-based method for estimating regional 2-D ionospheric VTEC and receiver differential code bias

Szu-Pyng Kao [a,*], Yao-Chung Chen [a], Fang-Shii Ning [b,1]

[a] National Chung Hsing University, No.250, Guoguang Rd., South Dist., Taichung City 402, Taiwan, ROC
[b] National Cheng Chi University, No.64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei City 116, Taiwan, ROC

Abstract

The geometry-free linear combination of dual-frequency GNSS reference station ground observations are currently used to build the Vertical Total Electron Content (VTEC) model of the ionosphere. As it is known, besides ionospheric delays, there are differential code bias (DCB) of satellite (SDCB) and receiver (RDCB) in the geometry-free observation equation. The SDCB can be obtained using the International GNSS Service (IGS) analysis centers, but the RDCB for regional and local network receivers are not provided. Therefore, estimating the RDCB and VTEC model accurately and simultaneously is a critical factor investigated by researchers. This study uses Multivariate Adaptive Regression Splines (MARS) to estimate the VTEC approximate model and then substitutes this model in the observation equation to form the normal equation. The least squares method is used to solve the RDCB and VTEC model together. The research findings show that this method has good modeling effectiveness and the estimated RDCB has good reliability. The estimated VTEC model applied to GPS single-frequency precise point positioning has better positioning accuracy in comparison to the IGS global ionosphere map (GIM).

1. Introduction

Navigation and positioning accuracy is reduced as the GPS signal is influenced by the ionosphere. Ionospheric influence can reduce the accuracy by several meters to tens of meters, and more than a hundred meters during a violent ionospheric storm. This ionospheric effect has apparently become the largest error source in GNSS navigation and positioning after Selective Availability (SA) is turned-off for single-frequency users. Therefore, the ionospheric effects must be considered for high accuracy positioning. There are two common ways to reduce the effects. The first method

uses the ionosphere-free linear combination of observations and the second uses the ionospheric correction model. In this study, we focus on the development of a new, effective ionospheric correction model for single frequency receiver applications. The precise ionospheric model is also applicable to space weather studies and geoscience (Komjathy, 1997). The widely known correction model for real-time application is the Klobuchar model (1986). The model coefficients are transmitted to the GPS users in the navigation message. This is an approximate model, providing only about 50% ionospheric effects (Klobuchar, 1987). Estimating a more precise ionospheric correction model has been the focus of many scholars. The Continuously Operating Reference Stations (CORS) are popular and have become increasingly dense throughout the world. The ground-based dual-frequency GPS observations have been used extensively to study ionospheric models. The geometry-free linear combination of dual-frequency GPS observations can eliminate

---

* Corresponding author. Address: National Chung Hsing University (NCHU), Taichung, Taiwan, ROC. Tel./fax: +886 422858180.
  E-mail addresses: spkao@dragon.nchu.edu.tw (S.-P. Kao), seven680007@gmail.com (Y.-C. Chen), fsn@nccu.edu.tw (F.-S. Ning).
[1] Tel.: +886 2939309x50741; fax: +886 229390251.

the frequency independent terms in the observation equation. However, the geometry-free linear combination consists of three parameters, the differential code bias of the satellite (SDCB) and receiver (RDCB) and the difference between ionospheric delays on L1 and L2 frequencies. Therefore, ionospheric model estimation is dependent on the method used to estimate the aforesaid three parameters.

At present, the function-based method is the commonly used method for simultaneous RDCB, SDCB and 2-D ionospheric VTEC model estimation. The function-based method assumes a single layer model (Klobuchar, 1987) for ionospheric electron content representation. In single layer model, all charged particles in the ionosphere are considered to be concentrated in a tiny spherical shell where the electron content is represented as Vertical Total Electron Content (VTEC). The intersection point of the satellite emitted signal path through a single-layer model to the ground receiver is called Ionospheric Pierce Point (IPP). The distribution of VTEC over IPPs is assumed to be a known mathematical function, e.g. Cosine, Parabola and Polynomial functions, Spherical Harmonics, etc. (Georgiadou, 1994; Schaer, 1999; Paulo and Joao, 2000; Gao and Liu, 2002; Ping et al., 2002; Yuan and Ou, 2004; Azpilicueta et al., 2006; Zhang et al., 2006; Opperman et al., 2007; Kao et al., 2012). This method has always bothered researchers for the selection of mathematical function with appropriate degree and order over modeling region. The ionosphere not only varies periodically with the time and space, but also has short-term irregular disturbances because of solar activity and geomagnetic variations. The same function model is therefore unlikely to fit different ionospheric behaviors. In order to overcome the above problems, statistical learning and spatial interpolation techniques have been proposed to model the ionospheric VTEC with good performance in previous studies (Wielgosz et al., 2003; Orús et al., 2005; Leandro and Santos, 2007; Durmaz et al., 2010; Durmaz and Karslioglu, 2011; Habarulema et al., 2011). However, the aforesaid studies only modeled VTEC or did not regard RDCB as an unknown parameter and solved it together with the model during modeling. Software (e.g. Bernese) was used to assume that the VTEC corresponds to the SH or Taylor series functions to estimate the RDCB. The estimated RDCB was then substituted as a given value in the observation equation to work out the VTEC in each IPP position before modeling. The estimated RDCB value will be related to the selected VTEC function model (Kao et al., 2013). The model built using the above method results from the VTEC function model conditions selected for estimating the RDCB. Considering the above problem, the RDCB shall be regarded as an unknown parameter and estimated together when a statistical learning method is used to estimate the ionospheric model.

The function-based method will improve ionospheric VTEC and RDCB estimation effectiveness by finding a method that automatically constructs a best fitting mathematical function of ionospheric VTEC according to the observations without prior knowledge of the ionospheric VTEC distribution. Therefore, this paper presents a statistical learning-based method to model ionospheric VTEC and estimate the RDCB simultaneously. This study proposes using the Multivariate Adaptive Regression Splines (MARS) technique to estimate the approximate ionospheric VTEC model first. This model is then substituted in the observation equations to form the normal equation. The least squares method is used to simultaneously solve RDCB and the ionospheric VTEC model. Durmaz et al. (2010) used the MARS technique for VTEC estimation over the European region. They showed that the algorithm has the ability to efficiently model the regional distribution of the VTEC and MARS can provide similar RMSE values with a much smaller number of coefficients compared to the spherical harmonic modeling. The structure of this paper is as follows; Section 2 describes the observation data preprocessing from ground GPS reference stations and how the observation equation is formed. Section 3 contains a brief description of the MARS theoretical basis and the concept and process of this research method. Section 4 declares the research data sources and the model specifications during estimation and analyzes and compares the research findings. The last section concludes the research findings of this paper.

2. GPS observations and preprocessing

GPS is used extensively in various domains, such as surveying, geoscience, navigation and so on. Nearly every country has its own ground GPS network. The IGS provides global observation data, allowing researchers to obtain the data free of charge through networks. The easy acquisition of GPS data and the ultra-high time and space data sampling rate have allowed dual-frequency GPS observations to be used extensively by the present researchers to estimate the ionospheric VTEC. This method uses mostly the geometry-free linear combination of dual-frequency observations to estimate the ionospheric VTEC. The geometry-free linear combination can eliminate the frequency independent terms in the original observation equation completely, such as geometrical distance, tropospheric delays, satellite clock error and receiver clock error, leaving only the ionospheric delays, RDCB, SDCB and noise. It is known that the accuracy of GPS pseudo range observation is lower than that of carrier phase observation, but the ambiguity must be estimated when using carrier phase observations. In order to overcome the above problem, this study used phase-smoothed code to compose geometry-free linear combination of observations (Schaer, 1999):

$$\tilde{P}_1(t) = \Phi_1(t) + \bar{P}_1 - \bar{\Phi}_1 + 2 \bullet \frac{f_2^2}{f_1^2 - f_2^2}$$
$$\bullet \left((\Phi_1(t) - \bar{\Phi}_1) - (\Phi_2(t) - \bar{\Phi}_2) \right) \qquad (1)$$

$$\tilde{P}_2(t) = \Phi_2(t) + \bar{P}_2 - \bar{\Phi}_2 + 2 \bullet \frac{f_1^2}{f_1^2 - f_2^2}$$
$$\bullet \left((\Phi_1(t) - \bar{\Phi}_1) - (\Phi_2(t) - \bar{\Phi}_2) \right) \qquad (2)$$

$$\tilde{L}4 = \tilde{P}_1 - \tilde{P}_2 = RDCB + (dion1 - dion2) + SDCB \qquad (3)$$

where $\tilde{P}_F(t)$ is the smoothed code measurement of frequency F at epoch t; $\Phi_F(t)$ is the carrier phase measurement of frequency F at epoch t; f_F is the frequency of carrier F; $\bar{P}_F - \bar{\Phi}_F$ is the mean difference between all accepted code and phase measurements in the current observation arc on frequency F; $\tilde{L}4$ is the smoothed geometry-free linear combination in units of meters. $RDCB$ and $SDCB$ are the hardware delays of the receiver and satellite respectively, in units of meters; $dion1$ and $dion2$ are ionospheric delays on $L1$ and $L2$ respectively, in units of meters. Only the first order of ionospheric refraction is considered while estimating the ionospheric delays in GPS processing, due to the effect of first order accounts for 99.9% of total. Its conversion relation with the Slant Total Electron Content (STEC) is (Hernández-Pajares M. et al., 2011):

$$dionF = \frac{40.3}{f_F^2} STEC \qquad (4)$$

where $STEC$ is the slant total electron content in the satellite and receiver propagation path, usually in TECU (Total Electron Content Units), 1 TECU $= 10^{16}$ electrons/m^2, (Hofmann-Wellenhof et al., 2007).

The ionosphere is a region of ionized plasma that extends from roughly 50 km to 1000 km above the Earth's surface. The ionosphere can be divided into several layers in altitude according to electron density. For convenient ionosphere modeling, this study uses the single-layer model concept introduced by Klobuchar (1987). It is assumed that all electrons in the ionosphere are concentrated into a thin shell at altitude H km above the Earth's surface. The intersection point of the GPS signal path with ionospheric shell is defined as the Ionospheric Pierce Point (IPP) at which the slant ionospheric delays has a zenith distance of Z'. The projection of the IPP onto the Earth's surface is called sub-ionospheric point, as shown in Fig. 1: where Z, Z' are the zenith distances of the satellite at the station and IPP respectively. R is the mean radius of the Earth, H is the height of the single layer above the Earth's surface. Since the ionospheric model is based on VTEC, it is required to use a mapping function to convert the STEC on the GPS signal path into VTEC at the sub-ionospheric point (Schaer, 1999):

$$F_I(z) = \frac{STEC}{VTEC} = \frac{1}{\cos z'} \quad \text{with} \quad \sin z' = \frac{R}{R+H} \sin z \qquad (5)$$

where H is generally recognized as 350–450 km above the Earth's surface.

Eqs. (4) and (5) are substituted in Eq. (3) to obtain:

$$\tilde{L}4 = RDCB + SDCB + 40.3 \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right) \times F_I \times VTEC \qquad (6)$$

According to Eq. (6), the number of $RDCB$ parameters is the number of stations, the number of $SDCB$ parameters is the number of satellites, $VTEC$ is the Vertical Total Electron Content in each IPP position. Therefore, the number of unknown parameters will be larger than the number of observation equations leading to rank deficiency. In order to estimate the $RDCB$, $SDCB$ with $VTEC$, it is usually assumed that the VTEC distribution corresponds to a known mathematical function.

3. Methodology

3.1. MARS (multivariate adaptive regression splines)

MARS is an adaptive regression procedure from the statistical learning field proposed by Friedman (1991). MARS can be seen as an extension of linear models that automatically model non-linearity and the interactions between variables. It is defined as a multivariate, piecewise regression method that can be used to model complex relationships between inputs and outputs (Hastie et al., 2009). Using a piecewise regression method, MARS divides the space of inputs into multiple knots first and then fits a spline function between these knots using so-called basis functions in an adaptive way. The MARS basis functions has the form (Hastie et al., 2009):

$$(x-t)_+ = \begin{cases} x - t, & \text{if } x > t \\ 0, & \text{otherwise} \end{cases}$$
$$(t-x)_+ = \begin{cases} t - x, & \text{if } x < t \\ 0, & \text{otherwise} \end{cases} \qquad (7)$$

where x is the input variables; t is the knot location; "+" sign indicates the positive part of the function.

Eq. (7) is called a reflected pair. The idea of MARS is to form reflected pairs for each input space with knots at each observed location of that input value. Therefore, the collection of basis functions is (Hastie et al., 2009):

$$C = \left\{ \left(X_j - t \right)_+, (t - X_j)_+ \right\}$$
$$\text{with} \quad t \in x_{ij}, \ i = 1, 2, \ldots, N; \ j = 1, 2, \ldots, p \qquad (8)$$

where p is the dimension of the input space; N is the number of observations; C is the basis functions set, if all of the input values are distinct, there are $2 \times N \times p$ basis functions; X_j is the j-th component of the input space; x_{ij} is the input values; t is the univariate knot location from x_{ij}. The MARS algorithm builds a regression function using the functions from the C set and their products. A general MARS model is defined as follows (Hastie et al., 2009):

$$f(x) = a_0 + \sum_{m=1}^{M} a_m B_m(x) \qquad (9)$$

where M is the number of basis functions; a_0 is the intercept; a_m is the coefficient of basis functions and $B_m(x)$ is a basis function constructed from set C, or a product of two or more such functions, $m = 1, 2, \ldots, M$; x is the input variables; $f(x)$ is the regression function built by MARS.

The basic MARS concept is to build the model by adding basis functions in the forward stepwise procedure and eliminating the term whose removal causes the smallest increase in residual squared error (least contribution) from the model in the backward deletion procedure (Hastie et al.,

2009). In the forward stage (constructive phase) a large over-fitted model of the form Eq. (9) is built. A generalized cross-validation (GCV) statistic is employed to prune the model to overcome this problem in the backward stage (pruning phase). The GCV is defined as (Hastie et al., 2009):

$$GCV(\lambda) = \frac{\sum_{i=1}^{N}\left(y_i - \left(\hat{a}_0 + \sum_{m=1}^{M}\hat{a}_m B_m(x_i)\right)\right)^2}{(1 - M(\lambda)/N)^2} \qquad (10)$$

where λ is the tuning parameter; N is the number of observations; M is the number of basis functions; y_i is the observation, $i = 1,2,\ldots,N$; $\hat{a}_0$ is the estimated intercept; $\hat{a}_m$ is the estimated coefficient of basis functions and $B_m(x)$ is the basis function, $m = 1,2,\ldots,M$; x_i is the input variables, $i = 1,2,\ldots,N$; $M(\lambda)$ is the effective number of parameters in the model, which can also be given by $M(\lambda) = r + ck$, where r is the number of linearly independent basis functions in the model. k is the number of knots which are selected in the forward process, c is the penalty for selecting knots, which is recommended as 3 (Hastie et al., 2009). In the backward stage, the model, which has the minimum GCV value is selected as the optimal MARS model. The more detailed information in forward stage and backward stage, please see Friedman (1991), Hastie et al. (2009) and Durmaz et al. (2010).

$$Y = \frac{\bar{L}4 - SDCB}{FF_I} = \frac{RDCB}{FF_I} + VTEC\ (\beta, S)$$
$$\text{with}\quad FF_I = 40.3\left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right) \times F_I \qquad (11)$$

where Y is the vector of observations; β is the geographic latitude of IPPs; S is the sun-fixed longitude of IPPs. First, if the term $RDCB$ in Eq. (11) is considered as a part of the error. The optimum ionospheric VTEC approximate model can be worked out by adjusting the model using adaptive continuous learning and using β, S as the independent variables and Y as the dependent variable, using the MARS algorithm. The estimated VTEC model as Eq. (9) form:

$$MVTEC(\beta, S) = \hat{a}_0 + \sum_{m=1}^{M}\hat{a}_m B_m(\beta, S) \qquad (12)$$

where $MVTEC(\beta, S)$ is the ionospheric VTEC approximate model; $\hat{a}_0$ is the estimated intercept; $\hat{a}_m$ is the estimated coefficient of basis functions and $B_m(\beta, S)$ is the basis function, $m = 1,2,\ldots,M$. In this paper, the estimated ionospheric VTEC model from Eq. (12) is called an approximate model because the model has RDCB not yet separated out. The estimated basis functions are then substituted into Eq. (11) to obtain the observation equation:

$$
\overbrace{\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ y_N \end{bmatrix}}^{\mathbf{Y}}_{N \times 1}
=
\overbrace{\begin{bmatrix}
FF_I(1,1)^{-1} & 0 & \cdots & 0 & 1 & B_1(\beta_1,S_1) & \cdots & B_M(\beta_1,S_1) \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
FF_I(1,k_1)^{-1} & 0 & \cdots & & \vdots & \vdots & \cdots & \vdots \\
0 & FF_I(2,1)^{-1} & \cdots & & \vdots & \vdots & \cdots & \vdots \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\vdots & FF_I(2,k_2)^{-1} & \cdots & \vdots & 1 & \vdots & \cdots & \vdots \\
\vdots & 0 & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\vdots & \vdots & \cdots & 0 & \vdots & \vdots & \cdots & \vdots \\
\vdots & \vdots & \cdots & FF_I(r,1)^{-1} & \vdots & \vdots & \cdots & \vdots \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & FF_I(r,k_r)^{-1} & 1 & B_1(\beta_N,S_N) & \cdots & B_M(\beta_N,S_N)
\end{bmatrix}}^{\mathbf{A}}_{N \times (r+1+M)}
\overbrace{\begin{bmatrix} RDCB_1 \\ RDCB_2 \\ \vdots \\ RDCB_r \\ a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix}}^{\mathbf{X}}_{(r+1+M)\times 1}
+
\overbrace{\begin{bmatrix} e_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ e_N \end{bmatrix}}^{\mathbf{e}}_{N \times 1}
\qquad (13)
$$

3.2. Algorithm

In this study the DCB values for the satellites are obtained from IGS Final IONEX files through the internet and substituted in the equation in the SDCB part. Eq. (6) is then changed to:

where N is the number of observations; M is the number of basis functions; r is the number of receivers; Y is the $N \times 1$ vector of observations; A is the $N \times (r + 1 + M)$ design matrix; X is the $(r + 1 + M) \times 1$ vector of unknown parameters; e is the $N \times 1$ vector of observation errors; $FF_I(i,p)^{-1}$ is the calculated value from Eq. (5) and (11) for p-th obser-
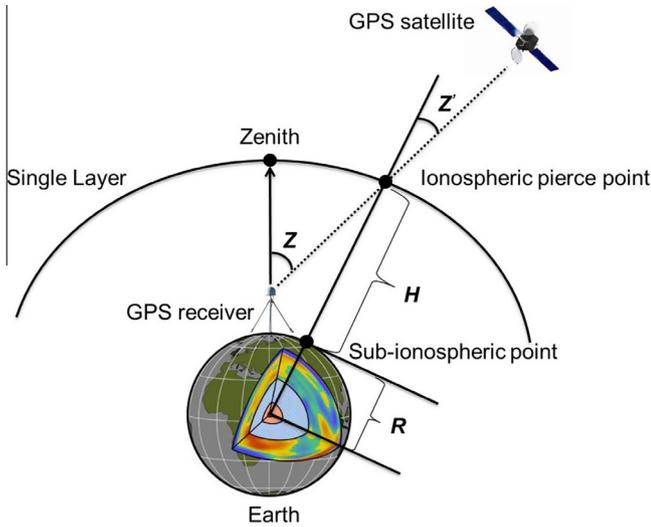
Fig. 1. Single-layer model of ionosphere.

vation of i-th receiver, $i = 1,2,\ldots,r$, $p = 1,\ldots,k_i$; $B_m(\beta_j, S_j)$ is the calculated value from m-th basis function with the geographic latitude and sun-fixed longitude for j-th IPP, $m = 1,2,\ldots,M$, $j = 1,2,\ldots,N$; $RDCB_i$ is the differential code bias of the receivers, $i = 1,2,\ldots,r$; a_k is the VTEC model coefficient of MARS need to be re-estimated, $k = 0,1,\ldots,M$. In this study it is assumed that the mean and RMS for the observation errors are zero and elevation dependent respectively (Schaer, 1999) and the observation errors are uncorrelated with each other. The variance–covariance matrix and weight matrix for the observations are defined as the following:

$$\mathbf{Q} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \cdots & 0 & 0 \\ \vdots & \vdots & \sigma_k^2 & \vdots & \vdots \\ 0 & 0 & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \sigma_N^2 \end{bmatrix}_{N \times N} \quad \text{with} \quad \sigma_i(z_i) = \frac{\sigma_0}{\cos(z_i)}$$
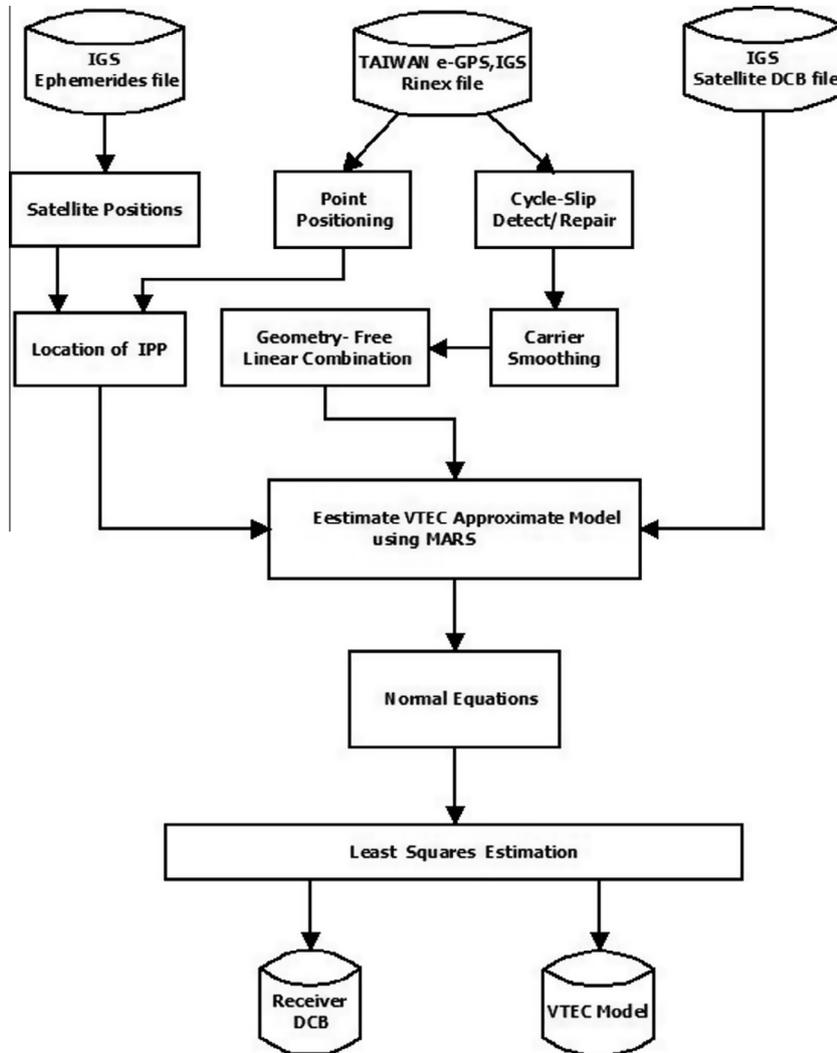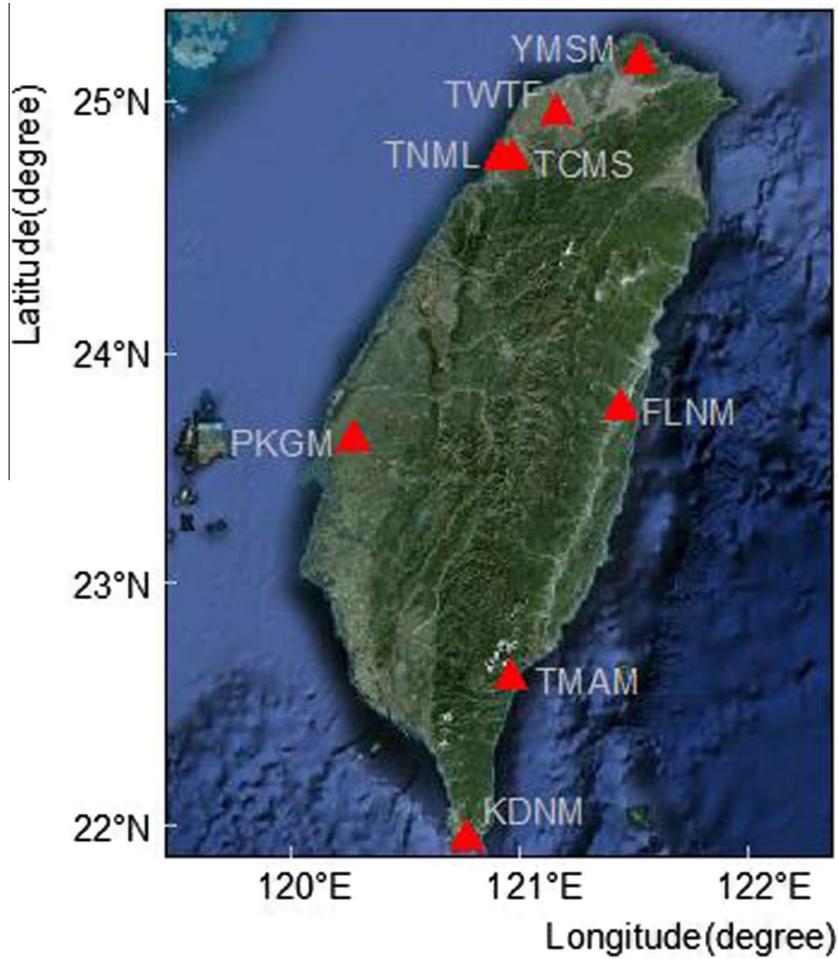
(14)

Fig. 2. Computing process.
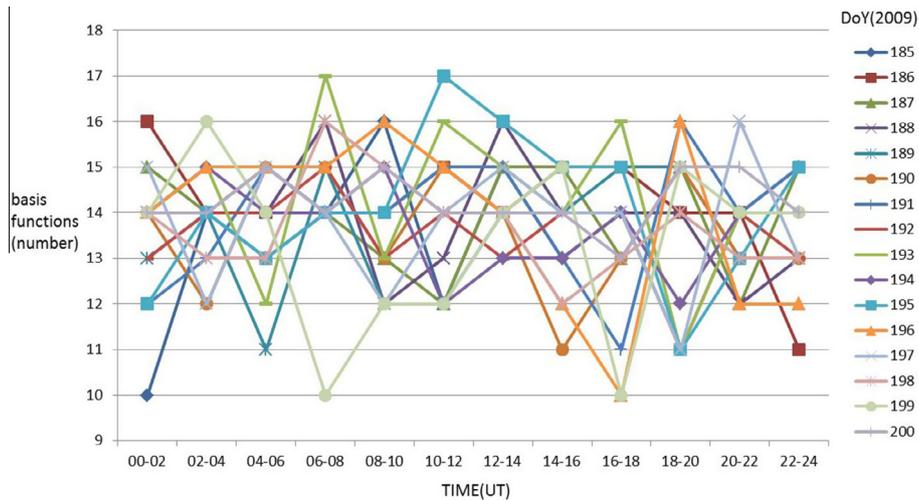
Fig. 3. The distribution of GPS reference stations.

Fig. 4. The number of basis functions for the estimated MARS models for each dataset on days from DoY 185 to 200 in 2009.

$$\mathbf{P} = \sigma_0^2 \mathbf{Q}^{-1}$$

where $\mathbf{Q}$ is the $N \times N$ diagonal variance–covariance matrix of observations; $\mathbf{P}$ is the weight matrix of observations; N is the number of observations; z_i is the zenith

angle of the satellite and σ_i is the standard deviation of the observations, $i = 1,2,...,N$; σ_0 is the standard deviation of unit weight corresponding to an observation at zenith. The Gauss–Markov model is built from mathematical model Eq. (13) and stochastic model Eq.
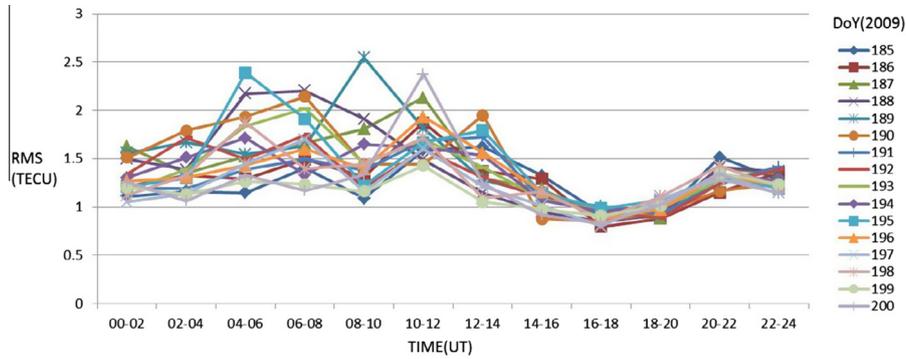
Fig. 5. RMS of residuals for the estimated models for each dataset on days from DoY 185 to 200 in 2009.

Table 1
The mean and standard deviation (std) for the RMS of residuals for the estimated models for each interval on days from DoY 185 to 200 in 2009 (TECU).

	Time (UT)										
	00–02	02–04	04–06	06–08	08–10	10–12	12–14	14–16	16–18	18–20	20–22	22–24
Mean	1.29	1.36	1.61	1.63	1.47	1.73	1.41	1.09	0.90	0.97	1.32	1.27
Std	0.17	0.22	0.35	0.31	0.37	0.25	0.26	0.12	0.06	0.08	0.09	0.08

Table 2
The daily mean and standard deviation (std) for the RMS of residuals for the estimated models on days from DoY 185 to 200 in 2009 (TECU).

	DoY (2009)															
	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
Mean	1.26	1.28	1.43	1.44	1.46	1.44	1.31	1.33	1.38	1.36	1.43	1.34	1.22	1.31	1.16	1.25
Std	0.24	0.27	0.36	0.46	0.45	0.44	0.25	0.28	0.33	0.25	0.42	0.28	0.25	0.28	0.15	0.39

Table 3
The number of basis functions and the RMS of residuals for the estimated models for each interval on DoY 69 in 2012.

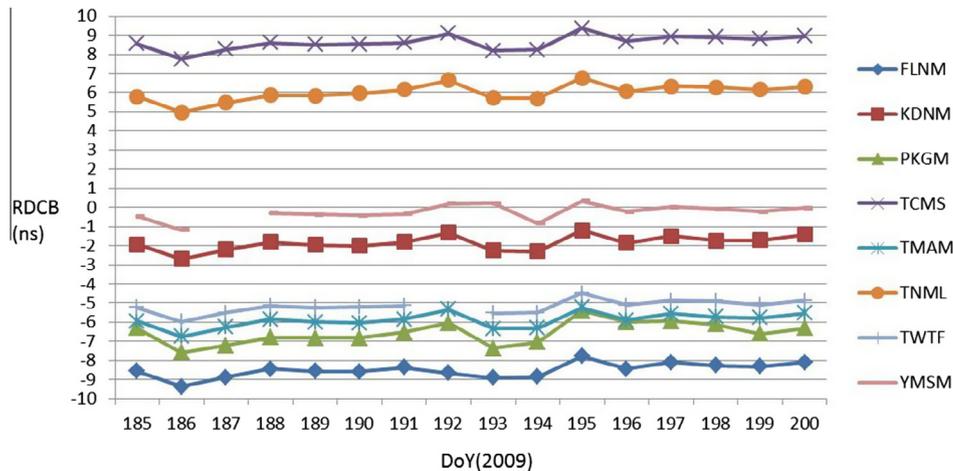|	Time (UT)										
	00–02	02–04	04–06	06–08	08–10	10–12	12–14	14–16	16–18	18–20	20–22	22–24
BF	14	15	13	12	14	14	14	16	13	13	13	14
RMS (TECU)	2.55	3.91	4.84	4.23	5.05	8.22	5.27	2.25	3.34	2.83	3.05	1.42

Fig. 6. The estimates of RDCB on days from DoY 185 to 200 in 2009.

Table 4
The mean and standard deviation (std) of receiver DCB (RDCB) estimates on days from DoY 185 to 200 in 2009 (ns).

	Station							
	FLNM	KDNM	PKGM	TCMS	TMAM	TNML	TWTF	YMSM
Mean	−8.51	−1.86	−6.55	8.62	−5.90	6.00	−5.19	−0.25
Std	0.38	0.39	0.59	0.40	0.39	0.44	0.36	0.40

Table 5
The receiver DCB (RDCB) estimates on DoY 69 in 2012 (ns).

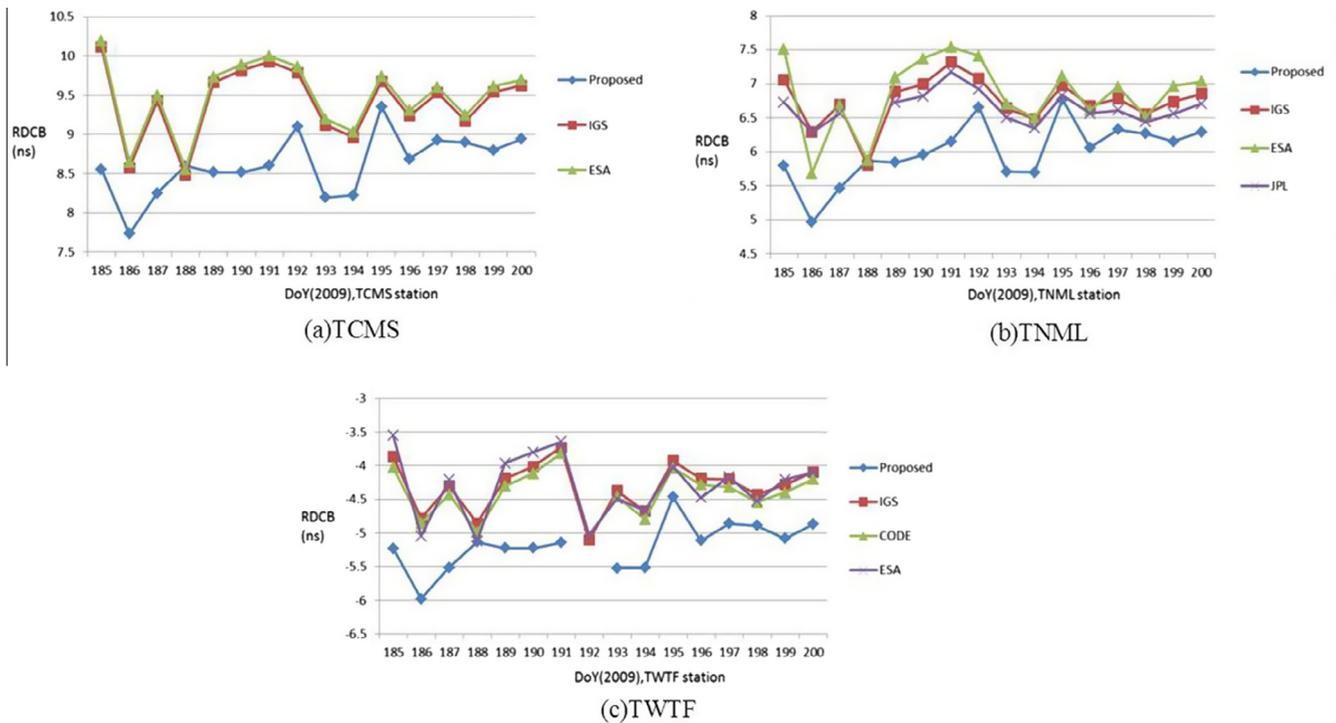|	Station							
	FLNM	KDNM	PKGM	TCMS	TMAM	TNML	TWTF	YMSM
RDCB	−8.52	0.08	8.03	10.20	−4.42	7.11	−2.01	−0.03

Fig. 7. Comparison of the receiver DCB (RDCB) estimated by IGS, CODE, ESA, JPL and the proposed method at TCMS (a), TNML (b) and TWTF (c) stations on days from DoY 185 to 200 in 2009.

(14). The RDCB and the coefficients of ionospheric VTEC model of MARS are estimated together using the least squares method:

$$\mathbf{X} = \begin{bmatrix} RDCB_1 \\ \vdots \\ RDCB_r \\ a_0 \\ \vdots \\ a_M \end{bmatrix} = (\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{P}\mathbf{Y}) \qquad (15)$$

Finally, the ionospheric VTEC model is created by the estimated basis functions from Eq. (12) and the re-estimated coefficients of MARS from Eq. (15). The calculation in this study is carried out using a program developed by the Matlab platform. The flow chart is shown in Fig. 2.

4. Data and results

The data used in this study were derived from 24-h dual-frequency GPS observations from five Taiwan e-GPS reference stations, including YMSM, FLNM, KDNM, PKGM and TMAM, and three IGS reference stations, including TCMS, TNML and TWTF. The experimental data is divided into two parts, the first part was extracted for 16

Table 6
Comparison of the receiver DCB (RDCB) estimated by IGS, CODE, ESA, JPL and the proposed method at TCMS, TNML and TWTF stations on days from DoY 185 to 200 in 2009 (ns).

	Station								
	TCMS			TNML			TWTF		
	Mean	Mean Δ_RDCB	RMS of Δ_RDCB	Mean	Mean Δ_RDCB	RMS of Δ_RDCB	Mean	Mean Δ_RDCB	RMS of Δ_RDCB
Proposed	8.62	–	–	6.00	–	–	−5.19	–	–
IGS	9.41	0.79	0.90	6.74	0.74	0.85	−4.31	0.88	0.98
CODE	–	–	–	–	–	–	−4.37	0.82	0.88
ESA	9.49	0.87	0.97	6.85	0.85	0.96	−4.31	0.88	1.02
JPL	–	–	–	6.65	0.65	0.74	–	–	–

Table 7
Comparison of the receiver DCB (RDCB) estimated by IGS, CODE, ESA, JPL and the proposed method at TCMS, TNML and TWTF stations on DoY 69 in 2012 (ns).

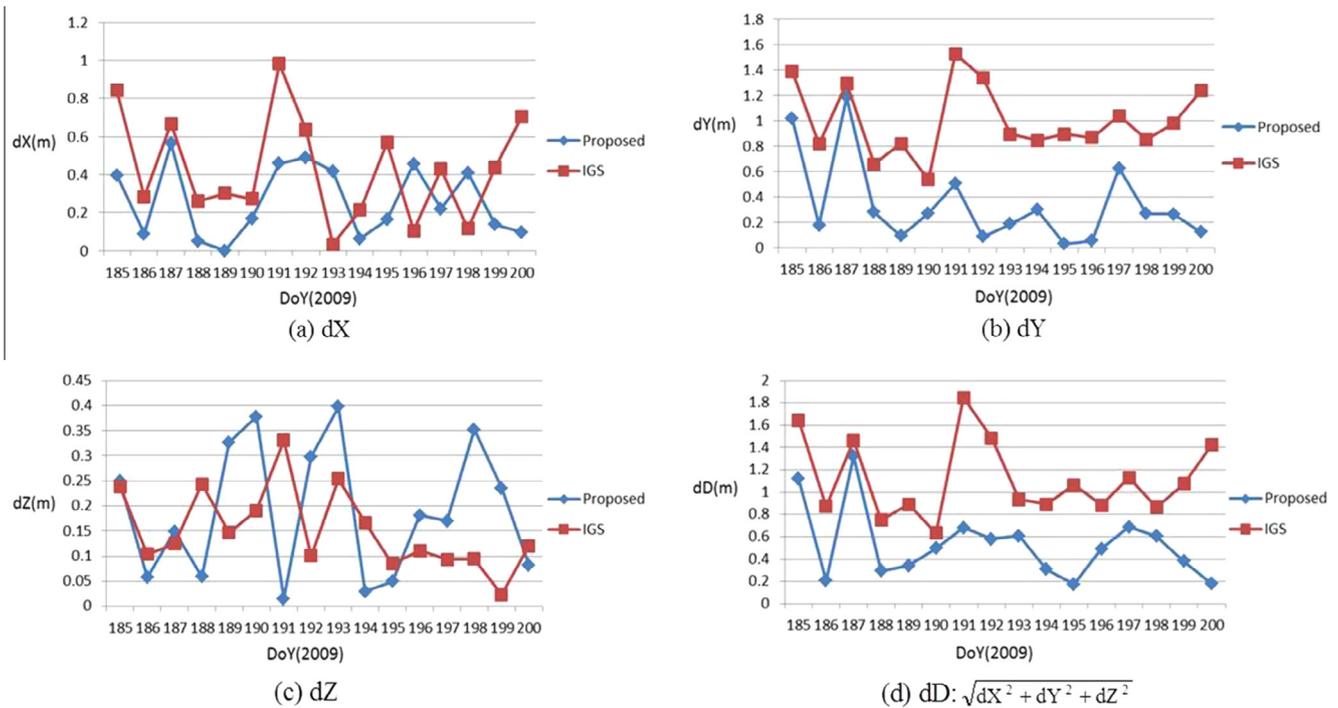|	Station					
	TCMS		TNML		TWTF	
	RDCB	Δ_RDCB	RDCB	Δ_RDCB	RDCB	Δ_RDCB
Proposed	10.20	–	7.11	–	−2.01	–
IGS	10.70	0.50	7.22	0.11	−2.24	−0.23
CODE	–	–	–	–	−2.17	−0.16
ESA	10.79	0.59	7.45	0.34	−1.83	0.18
JPL	–	–	7.06	−0.05	−2.56	−0.55

Fig. 8. The positioning errors on days from DoY 185 to 200 in 2009.

days, from July 4 to 19, 2009 (DoY 185–200), the ionospheric activity is relatively quiet during this period. The second part was extracted from March 9, 2012 (DoY 69), on this day a geomagnetic storm was observed. The GPS reference station distribution is shown in Fig. 3. The height

of the single-layer model was set to 450 km as well as the mean radius of the Earth was set to 6371 km for calculations. A sampling rate of 30 s and an elevation cut-off angle of 10° were used. The precise ephemeris was obtained from IGS. In order to compare the estimated ionospheric VTEC

Table 8
The mean positioning errors on days from DoY 185 to 200 in 2009 (m).

		dX	dY	dZ	$dD = \sqrt{dX^2 + dY^2 + dZ^2}$
Proposed	Mean	0.253	0.301	0.186	0.530
IGS	Mean	0.402	0.977	0.146	1.116

Table 9
The positioning errors on DoY 69 in 2012 (m).

	dX	dY	dZ	$dD = \sqrt{dX^2 + dY^2 + dZ^2}$
Proposed	0.119	−0.565	−0.161	0.599
IGS	−1.798	1.598	−0.345	2.430

of the proposed method with IGS GIM, this study estimated a set of solutions for 24-h daily data at an interval of every 2 h.

Fig. 4 shows the number of basis functions for the estimated MARS models for each dataset on days from DoY 185 to 200 in 2009. The number of basis functions for each dataset is between 10 and 18. Fig. 5 shows the RMS of residuals for the estimated models for each dataset on days from DoY 185 to 200 in 2009. The large RMS occurs at 0400–1200UT (local time: UT+08). This time interval is just when the VTEC in Taiwan is large. Table 1 shows the mean and standard deviation (std) for the RMS of residuals for the estimated models for each interval during 16 experiment days. The mean RMS in each time interval was less than 2 TECU, and the std was less than 0.4 TECU. Table 2 shows the daily mean and std for the RMS of residuals for the estimated models. The mean RMS on various experiment days was less than 1.5 TECU, and the std was less than 0.5 TECU. Table 3 shows the number of basis functions and the RMS of residuals for the estimated models for each interval on DoY 69 in 2012. The basis functions for each dataset are between 12 and 16 and the maximum and minimum values for the RMS are 8.22 TECU and 1.42 TECU, respectively. It is noted that the residuals described in this paragraph is calculated from Eq. (15).

Fig. 6 shows the daily RDCB estimate results at various reference stations. Fig. 6 shows a consistent trend in variation in the RDCB for eight reference stations. As YMSM cannot obtain GPS observations on DoY 187 and TWTF cannot obtain GPS observations on DoY 192, they fail to estimate RDCB. Table 4 shows the mean and std for daily RDCB estimates at various reference stations during 16 experiment days. The std for eight reference stations is less than 0.6 ns, meaning the aforesaid RDCB is quite stable with slight day-to-day variations during the experiment days. Table 5 shows the RDCB estimates on DoY 69 in 2012.

Fig. 7 and Table 6 compare IGS, CODE, ESA, JPL with TCMS, TNML and TWTF estimates of RDCB with the method proposed in this study on days from DoY 185 to 200 in 2009. Fig. 7 shows that the daily RDCB estimates from the method proposed in this study and the four

institutions have similar variation trends. Table 6 shows that the mean and RMS for the differences between the method proposed in this study and the four institutions are almost less 0.9 ns and 1.0 ns, respectively. Table 7 compares IGS, CODE, ESA, JPL with TCMS, TNML and TWTF estimates of RDCB with the method proposed in this study on DoY 69 in 2012. Table 7 shows that the differences between the method proposed in this study and the four institutions are less 0.6 ns.

The IGS global ionosphere map (GIM) and the regional ionospheric VTEC model were estimated using the proposed method for GPS single-frequency precise point positioning (PPP) to further analyze ionospheric VTEC model accuracy. The PPP was implemented based on GpsTools ver. 0.6.4, developed by T. Takasu. The phase-smoothed code was used for positioning with a sample rate of 30 s and an elevation cut-off angle of 15°. TCMS was used as the reference station and the weekly geocentric coordinates from SINEX files published by IGS were used as reference values. A set of solutions were estimated every day to obtain the positioning errors. Fig. 8 and Table 8 compare the absolute values and the mean positioning errors on days from DoY 185 to 200 in 2009, respectively. Table 9 compares the positioning errors on DoY 69 in 2012. Tables 8 and 9 show that the regional ionospheric VTEC model estimated using the proposed method has better positioning accuracy. Its three-dimensional positioning errors of 0.530 m on days from DoY 185 to 200 in 2009 and 0.599 m on DoY 69 in 2012 are better than 1.116 m and 2.430 m of IGS GIM model, respectively.

5. Conclusion

This study proposed a statistical learning-based method to estimate the regional 2-D ionospheric VTEC and RDCB simultaneously in Taiwan. The MARS was used in this method to develop an ionospheric VTEC approximate model from ground GPS reference station observations first. The least squares method was then used to simultaneously estimate the RDCB and the coefficients of ionospheric VTEC model. Compared to conventional function-based methods, such as polynomial functions, Taylor series and Spherical Harmonics, the proposed method can find the optimal approximate model flexibly and adaptively from the observations using MARS without a priori knowledge of the ionospheric VTEC distribution mathematical function. We tested the proposed method in this paper under different ionospheric conditions. The research findings showed that the proposed method has good modeling effectiveness and can solve the ionospheric

VTEC and RDCB simultaneously based on statistical learning techniques. The results showed that the estimated RDCB has a good reliability. The estimated ionospheric VTEC model applied to GPS single-frequency precise point positioning gives better positioning accuracy than using IGS GIM. Therefore, this method can serve as an attractive and alternative method for estimating 2-D ionospheric VTEC and RDCB. However, this method was first proposed and tested using regional observations in Taiwan. The proposed method's suitability for modeling global or single-station ionosphere TEC distributions will be studied in future studies.

References

Azpilicueta, F., Brunini, C., Radicella, S.M., 2006. Global ionospheric maps from GPS observations using modip latitude. Adv. Space Res. 38 (11), 2324–2331.

Durmaz, M., Karslioglu, M.O., 2011. Non-parametric regional VTEC modeling with multivariate adaptive regression B-splines. Adv. Space Res. 48 (9), 1523–1530.

Durmaz, M., Karslioglu, M.O., Nohutcu, M., 2010. Regional VTEC modeling with multivariate adaptive regression splines. Adv. Space Res. 46 (2), 180–189.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–67.

Gao, Y., Liu, Z.Z., 2002. Precise ionosphere modeling using regional GPS network data. J. Global Positioning Syst. 1, 18–24.

Georgiadou, Y. Modeling the ionosphere for an active control network of GPS stations. Delft Geodetic Computing Centre LGR-Series No. 7, Delft University of Technology, 1994.

Habarulema, J.B., McKinnell, L.A., Opperman, B.D.L., 2011. Regional GPS TEC modeling: attempted spatial and temporal extrapolation of TEC using neural networks. J. Geophys. Res. (116), 1–14. A04314.

Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in, Statistics, 2009.

Hernández-Pajares, M., Juan, J.M., Sanz, J., Aragón-A', A'., Garćia-Rigo, A., Salazar, D., Escudero, 2011. The ionosphere: effects, GPS modeling and the benefits for space geodetic techniques. J. Geod. (85), 887–907.

Hofmann-Wellenhof, B., Lichtenegger, H., Wasle, E., 2007. GNSS Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more. Springer Wien, New York.

Kao, S.P., Tu, Y.M., Chen, W., Weng, D.J., Ji, S.Y., 2013. Factors affecting the estimation of GPS receiver instrumental biases. Surv. Rev. 45 (328), 59–67.

Kao, S.P., Tu, Y.M., Ji, S.Y., Chen, W., Wang, Z., Weng, D.J., Ding, X., Hu, T. A 2-d ionospheric model for low latitude area-Hong Kong. Adv. Space Res (2012). http// dx.doi.org/10.1016/j.asr. 2012.11.031.

Klobuchar, J.A., 1987. Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. AES-23 (3), 325–331.

Komjathy, A. Global ionospheric total electron content mapping using the Global Positioning System (Ph.D. dissertation), Department of Geodesy and Geomatics Engineering, Technical Report No. 188, University of New Brunswick, Fredericton, New Brunswick, Canada, 248, 1997.

Leandro, R.F., Santos, M.C., 2007. A neural network approach for regional vertical total electron content modeling. Stud. Geophys. Geod. 51 (2), 279–292.

Opperman, B.D.L., Cilliers, P.J., McKinnell, L.A., Haggard, R., 2007. Development of a regional GPS-based ionospheric TEC model for South Africa. Adv. Space Res. 303 (39), 808–815.

Orús, R., Hernández-Pajares, M., Juan, J.M., Sanz, J. Improvement of global ionospheric VTEC maps by using kriging interpolation technique. J. Atmos. Sol. Terr. Phys. 67 (16):1598–1609, 2005.

Paulo, D., Joao, F., 2000. Application of ionospheric corrections in the equatorial region for L1 GPS users. Earth Planets Space 52 (5), 1083–1089.

Ping, J., Kono, Y., Matsumoto, K., Otsuka, Y., Saito, A., Shum, C., Heki, K., Kawano, N., 2002. Regional ionosphere map over Japanese Islands. Earth Planets Space 54 (12), 13–16.

Schaer, S. Mapping and predicting the earth's ionosphere using the global positioning system (Ph.D. thesis), University of Berne, Switzerland, 1999.

Wielgosz, P., Grejner-Brzezinska, D.A., Kashani, I., 2003. Regional ionosphere mapping with kriging and multiquadric methods. J. Global Positioning Syst. 2 (1), 48–55.

Yuan, Y., Ou, J. A generalized trigonometric series function model for determining ionospheric delay. Prog. Nat. Sci. 14 (11), 1010–1014, http://dx.doi.org/10.1080/10020070412331344711, 2004.

Zhang, H.-P., Ping, J.-S., Zhu, W.-Y., C., Huang, 2006. Brief review of the ionospheric delay models. Prog. Astron. 24 (1), 16–26 (in Chinese).