

# Quality Control and Its Impacts on TEC Modelling With GPS Radio Occultation Data and Ground-Based GPS Data

Jinling Wang and Gary Ouyang

*School of Surveying and Spatial Information Systems, University of New South Wales*

## Abstract

The accuracies of TEC models derived from ground-based GPS data are affected by several factors including assumed ionospheric shell height, satellites geometry, the chosen satellite cut off angles, the quantity of the GPS measurement data and etc.. Potential outliers within the measurements can have a significant impact on the quality of the estimated TEC models. However, there have been no any discussions on the outlier detection procedures in the literature on TEC modelling.

This paper for the first time extends the commonly used Truncated Singular Value Decomposition (TSVD)-based TEC modelling method with an outlier detection and exclusion procedure. Furthermore, the retrieved TEC is transferred as vertical electron density (VED) profiles using the Shape Function model for the outlier impact analysis. In the paper, the biases from the outlier and the Shape Function model error can be separated and evaluated with the use of ionosonde data. To test the performance of this proposed quality control procedure, a day-time data set was used to produce the VED profiles at 3 different locations in Australia where both GPS and ionosonde data are available. The real-data tests have shown that the proposed outlier detection and exclusion procedure can numerically evaluate the impact of outliers and the Shape Function model errors, which are time and location dependent, in terms of the relative RMS measure commonly used in the ionospheric modelling studies. For example, the above two influences are 17.95% and 45.95% at Townsville, and 13.13% and 13.31% at Canberra, respectively. Besides, the simulated data were also used in the analysis. The result has demonstrated that as the magnitude of the outlier increases, the estimated VED can be biased by over 50%. Therefore, it is necessary to include a quality control procedure in retrieving an accurate and reliable TEC distribution.

**Keywords:** Outliers, Quality Control, TEC Modelling, Radio Occultation

## 1. Introduction

To represent temporal/spatial variations of the ionosphere, many TEC models based on grids or functions have been developed with ground-based GPS data, and widely used in the fields related to electromagnetic wave transmission because the TEC along the radio signal ray path can decide the distorted magnitude of the phase and amplitude of the signals. In addition, through combining the TEC distribution and Low Earth Orbit (LEO) satellite Radio Occultation (RO) data, the VED profiles can be retrieved through the Shape Function model (Fjeldbo and Eshleman 1968; Hajj and Romans 1998; Hajj et al. 2002, 2004; Garcia-Fernandez 2004).

The accuracy and reliability of the above TEC models are strongly affected by the model assumptions and the quality of the GNSS data used. Because the estimation procedure is based on the least squares concept, both stochastic and functional models should be realistically established. For example, there should be no outliers within the measured TEC values, and the condition number of the coefficient (design) matrix for the functional model should be reasonable. Any misspecifications for the least-squares models will generate big effects on the accuracy and reliability of the TEC modelling results (Strang and Borre 1997; Liao 2000; Liu 2001; Hartl 2007).

To analyse the outlier effect on the TEC model, the “real” vertical TEC (VTEC) value which is from ground to GPS satellite orbit height in vertical direction, must be acquired and is used as a reference standard. But these values are not available unless the GPS receivers just have  $0^{\circ}$  zenith angle to the GPS satellites. Because, at this moment, the VTEC at the receiver position, which is calculated by the GPS geometric free combination equation (mentioned in Section 2), can be regarded as the “real” one. Therefore, we need to introduce another different reference standard for the comparison. This reference is the VED profile retrieved by the ionosonde.

In the paper, a proposed quality control method for TEC modelling is discussed, and the encouraging results generated from the proposed quality control procedure are also numerically validated by using the VED profiles from the ionosonde data.

## 2. Methodology for TEC modelling and quality control

The proposed quality control method focuses on analysing the bias from the outlier impact based on the retrieved VED profiles. Meanwhile, the bias from the Shape Function model error also is presented. Therefore, in this section, firstly, the procedure of generating VED profile is briefly described as: (1) modelling VTEC by a spherical harmonic function; (2) solving unknown parameters of the VTEC model using the combined Truncated Singular Value Decomposition (TSVD) and  $W$ -test method; (3) generating VED profiles based on the retrieved VTEC using the Shape Function model. Secondly, the method of separating the biases from the outlier and the Shape Function model error is presented.

### 2.1 Functional and stochastic models for TEC estimation with the least squares algorithm

The VTEC distribution can be described by a spherical harmonic function as follows:

$$VTEC(\theta, \phi) = \sum_{n=0}^{n_{\max}} \sum_{m=0}^n P_{nm}(\sin \theta) * (a_{nm} \cos(m\phi) + b_{nm} \sin(m\phi)) \quad (1)$$

where  $\theta$  is the solar fixed longitude of the IPP;  $\phi$  is the geomagnetic latitude of the IPP;  $n_{\max}$  is the maximum degree of the spherical harmonic function expansion;  $P_{nm}$  are the normalized associated Legendre functions;  $a_{nm}$  and  $b_{nm}$  are the coefficients to be solved (Ping et al. 2002).

The slant TEC (STEC) can be expressed as:

$$STEC(\theta, \phi) = M(e) \cdot VTEC(\theta, \phi) \quad (2)$$

where  $M(e)$  is the mapping function from vertical TEC to slant TEC:

$$M(e) = \left( 1 - \left( \frac{R}{R+h} \cos(e) \right)^2 \right)^{-\frac{1}{2}} \quad (3)$$

where  $R$  is the Earth radius;  $h$  is the assumed ionospheric shell height;  $e$  is the satellite elevation angle.

The STEC based on GPS pseudo-range measurements is expressed as:

$$STEC = \frac{f_1^2 f_2^2}{40.3(f_2^2 - f_1^2)} (P_1 - P_2) + DCB_R + DCB_S \quad (4)$$

where  $f_1$  and  $f_2$  are 1575.42MHz and 1227.60MHz respectively;  $P_1$  and  $P_2$  are pseudoranges (in meters);  $DCB_R$  is receiver differential code biases;  $DCB_S$  is satellite differential code biases. In this study, the  $DCB_R$  and the  $DCB_S$  are directly obtained from the IGS products, and therefore, are not estimated as unknown parameters in the least squares equations.

Given the limited number of STEC measurements over a short period of time, saying 30 seconds, during which the TEC distribution is considered as invariable, let the maximum degree of the spherical harmonic function as three ( $n_{\max} = 3$ ). Combining the above four equations, the linearized functional model can be expressed as:

$$l + v = Ax \quad (5)$$

where  $l$  is the  $m \times 1$  observation vector;  $v$  is the  $m \times 1$  residual vector;  $x$  is the  $t \times 1$  vector of the unknown parameters within the spherical harmonic function:

$$x = \begin{bmatrix} a_{00} & a_{10} & a_{11} & b_{11} & a_{20} & a_{21} & b_{21} & a_{22} \\ & b_{22} & a_{30} & a_{31} & b_{31} & a_{32} & b_{32} & a_{33} & b_{33} \end{bmatrix}^T \quad (6)$$

And the  $i^{\text{th}}$  row of the design matrix  $A$  is:

$$A_i = \begin{bmatrix} M(e)P_{00}(\sin \theta) & M(e)P_{10}(\sin \theta) \\ M(e)P_{11}(\sin \theta)\cos \phi & M(e)P_{11}(\sin \theta)\sin \phi \\ M(e)P_{20}(\sin \theta) & M(e)P_{21}(\sin \theta)\cos \phi \\ M(e)P_{21}(\sin \theta)\sin \phi & M(e)P_{22}(\sin \theta)\cos(2\phi) \\ M(e)P_{22}(\sin \theta)\sin(2\phi) & M(e)P_{30}(\sin \theta) \\ M(e)P_{31}(\sin \theta)\cos \phi & M(e)P_{31}(\sin \theta)\sin \phi \\ M(e)P_{32}(\sin \theta)\cos(2\phi) & M(e)P_{32}(\sin \theta)\sin(2\phi) \\ M(e)P_{33}(\sin \theta)\cos(3\phi) & M(e)P_{33}(\sin \theta)\sin(3\phi) \end{bmatrix} \quad (7)$$

In the least-squares estimation procedure, a realistic stochastic model for the observations should be established. In this case here, the ‘‘observations’’ are the STEC values. Normally, when this GPS signal is projected to the vertical direction at the given IPP, one will expect that the smaller the elevation angle for a GPS satellite, the bigger the STEC noise (bias). Thus, the covariance matrix is defined as below:

$$D = \sigma_0^2 Q = \sigma_0^2 P^{-1} = \text{diag} \left( \left( \sigma \cdot \left( \frac{H}{h_i} \right)^{-1} \right)^2 \right) \quad i = 1, 2, \dots, m \quad (8)$$

where  $Q$  is the cofactor matrix; and  $P$  is the weight matrix;  $m$  is the number of observation equations;  $\sigma$  is the standard TEC bias from GPS pseudorange calculation with a range from 1 to 5 TECU. In this study, the magnitude of 2.5 was used for  $\sigma$ ;  $\sigma_0^2$  is the priori variance factor, and can be equal to 1 for convenience;  $H$  is the whole ionospheric height (the distance from the bottom to the upper boundaries of the ionosphere in vertical direction) i.e. 3000km;  $h_i$  ( $i=1, 2, \dots, m$ ) is the distance from the bottom to the upper boundaries of the ionosphere along the  $i^{\text{th}}$  signal ray path (in slant direction) (Kaplan 1996; Liu et al. 2005). In some applications the covariance matrix  $D$  may be defined as a fully distributed matrix, reflecting the potential correlations between the STEC observations.

## 2.2 Estimating the unknown parameters and outlier detection

Based on Equation (5), the unique and optimal solution for the unknown parameters can be obtained using the least squares algorithm. In fact, many observation equations in Equation (5) are closely correlated and thus, matrix  $A$  has a large condition number. Therefore, the normal matrix in the least squares  $(A^T PA)$  does not have a normal inverse matrix, where matrix  $P$  is previously defined in Equation (8). To overcome this difficulty, the truncated singular value decomposition (TSVD) method may be adopted. The matrix  $(A^T PA)$  can be rewritten as (e.g., Xue et al., 2000; Hartl 2007).

$$A^T PA = U \Sigma V \quad (9)$$

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \quad (10)$$

$$\Sigma_r = \text{diag}(\sigma_i) \quad i = 1, \dots, r \quad (11)$$

where  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , they are the singular values of matrix  $(A^T PA)$ ;  $U$  and  $V$  are orthogonal matrices. Then an inverse matrix for  $(A^T PA)$  can be obtained as:

$$(A^T PA)^- = V \Sigma_h U^T \quad (12)$$

$$\Sigma_h = \text{Diag}(d_k), \quad d_k = \begin{cases} 1/\sigma_k & \text{for } \sigma_k \geq t_0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where  $t_0$  is the threshold value. Combining equations (5) and (12), the coefficients of the spherical harmonic function can be solved as:

$$\hat{x} = (A^T PA)^- A^T P l \quad (14)$$

The observation residuals can be computed as:

$$\hat{v} = A \hat{x} - l = A(A^T PA)^- A^T P l - l \quad (15)$$

The variance-covariance matrix for the residuals is derived as:

$$Q_{\hat{v}} = [A(A^T PA)^- A^T P - E] Q [A(A^T PA)^- A^T P - E]^T \\ = Q - A(A^T PA)^- A^T \quad (16)$$

It is noted that in the above formula, the matrix  $(A^T PA)^-$  is a general inverse matrix defined by Equation (12), instead of a regular inverse matrix in the case of the normal least-squares procedure. Similar to the normal least squares procedure, the redundancy numbers for the measurements can be defined as:

$$r_i = (Q_{\hat{v}} P)_{ii} = (E - A(A^T PA)^- A P)_{ii} \quad (17)$$

However, the sum of the redundancy numbers here is not the difference between the number of observations  $m$  and the number of the unknown parameters  $t$  due to the fact that the matrix  $(A^T PA)^-$  is rank deficient. Therefore,

$$\sum_{i=1}^m r_i = \sum_{i=1}^m (Q_{\hat{v}} P)_{ii} \\ = \text{trace} [E - A(A^T PA)^- A P] = m - \text{rank} [(A^T PA)^- A^T P A] \\ = m - r \quad (18)$$

Equation (18) can be used as a check for the computations within the TSVD software.

## 2.3 W-Test for outlier detection within the TSVD

Although the TSVD method has been widely used in TEC modelling, there have been no any discussions on the outlier detection procedure. Due to the singularity of the normal matrix  $(A^T PA)$ , the traditional outlier detection method  $W$ -test proposed by Baarda (1968) should be modified. Assume there is one outlier existing

in the  $i^{th}$  observation ( $i=1, 2, \dots, n$ ), which can be treated as a mean shift  $\nabla\hat{s}_i$

$$l + v_m = A\hat{x}_m + h_i\nabla\hat{s}_i \quad (19)$$

where  $h_i$  is the  $n \times 1$  unit vector in which  $i^{th}$  component holds the value one, indicating that the observation  $l_i$  contains the outlier  $\nabla\hat{s}_i$ . This extended function model (19) can be combined with the stochastic model given by Equation (8) to estimate the value of the suspected outlier as follows:

$$\nabla\hat{s}_i = (h_i^T P Q_{\hat{v}} P h_i)^{-1} h_i^T P \hat{v} \quad (20)$$

and the associated co-factor as:

$$\begin{aligned} Q_{\nabla\hat{s}_i} &= (h_i^T P Q_{\hat{v}} P h_i)^{-1} h_i^T P Q_{\hat{v}} P (h_i^T P Q_{\hat{v}} P h_i)^{-1} \\ &= (h_i^T P Q_{\hat{v}} P h_i)^{-1} \end{aligned} \quad (21)$$

where  $\hat{v}$  and  $Q_{\hat{v}}$  are defined by Equations (15) and (16), respectively. Then, to test the zero hypothesis  $H_0: E(\nabla\hat{s}_i)=0$ , against the alternative hypothesis  $H_a: E(\nabla\hat{s}_i) \neq 0$ , the following test statistic can be established as:

$$W_i = \frac{\nabla\hat{s}_i}{\sigma_0 \sqrt{Q_{\nabla\hat{s}_i}}} = \frac{h_i^T P \hat{v}}{\sigma_0 \sqrt{h_i^T P Q_{\hat{v}} P h_i}} \quad (22)$$

Therefore, the statistic  $W_i$  has a standard normal distribution (Wang and Chen 1999). The threshold of this new  $W$  test is set as  $W_0$ , for a given confidence level for the test. For example, if the confidence level is chosen as 99%, the test threshold is 2.579. As any single big outlier may increase the sizes for several residuals due to the correlations between the residuals, in any single iteration, only the biggest testing statistic is to be identified and compared with the preset threshold. If

$$\text{Max}(|W_i|) \geq W_0 \quad (23)$$

then, the observation equation with the biggest testing statistic will be removed and the coefficients of the spherical harmonic function will be re-calculated according to the methods in Sections 2.2. Such a procedure will be iterated until no more outlier is to be removed.

Assuming that the observations are uncorrelated, i.e., the covariance matrix  $D$  is diagonal; the test statistic to detect outliers can then be simplified as:

$$W_i = \frac{\hat{v}_i}{M_{\hat{v}_i}} \quad i=1, \dots, m \quad (24)$$

where  $\hat{v}_i = (A\hat{x} - l)_i$ ,  $i=1, \dots, m$  are the TSVD residuals and  $M_{\hat{v}_i} = \sigma_0 \sqrt{(Q_{\hat{v}_i})_{ii}}$ ,  $i=1, \dots, m$  are the standard deviations for the residuals.

As is shown above, this new procedure fortunately happens to be similar to the commonly used one, but some distinct differences shown in Equations (15), (16), (17), and (18) should be noted in the software implementation.

#### 2.4 Estimating VED with the TEC model and LEO satellite RO data (the shape function model)

According to the slab thickness theory, VTEC is distributed to different altitude, varying with certain proportionality in the ionosphere. Therefore, the electron density is able to be defined as below (Davies 1990; Hernandez-Pajares et al. 2000; Garcia-Fernandez 2004).

$$N(\lambda, \theta, h) = \text{VTEC}(\lambda, \theta) \cdot F(h) \quad (25)$$

where  $\lambda$  is longitude,  $\theta$  is latitude,  $h$  is altitude,  $N(\lambda, \theta, h)$  is the electron density at point  $(\lambda, \theta, h)$ ,  $\text{VTEC}(\lambda, \theta)$  is the vertical TEC at point  $(\lambda, \theta)$ ,  $F(h)$  is the shape function with regard to altitude  $h$ .

$F(h)$  is firstly solved using the RO data. Then 3D electron density or the VED profile in a given location can be retrieved by Equation (25) (Hajj et al. 2002; 2004; Garcia-Fernandez 2004).

#### 2.5 Separating the outlier effect and the Shape Function model error

The VED profiles, which are retrieved by the Shape Function model as Equation (25), include two parts of errors from the following two steps:

- (1) Generating the VTEC; and
- (2) Calculating the VED profile by the Shape Function model.

The above errors are classified into two groups: (a) from the ground based GPS measurement outlier(s) and, (b) from all other errors except for that from (a).

The errors from group (a) are the outlier effect, and the errors from group (b) are the Shape Function model error which includes all the errors (except for ground based GPS outliers). In this paper, the outlier effect is from the ground based GPS measurement outliers, whilst the RO GPS measurement outlier influence is treated as part of the Shape Function model error.

The biases from above two groups of errors are estimated as follows:

Based on Section 2.3, we assume that the outlier effect can be removed by the  $W$ -test. Therefore, the VED bias from the Shape Function model error can be expressed as:

$$Bias_{exclu-outlier}(i) = N_{with-w-test}(i) - N_{ionosonde}(i) \quad i = 1, \dots, n \quad (26)$$

where  $i$  is the  $i^{th}$  layer ionosphere;  $n$  is the total layer number of the ionosphere;  $Bias_{exclu-outlier}$  is the bias from Step (2) or the Shape Function model error;  $N_{with-w-test}$  is the electron density generated by Equation (25), in which the  $W$ -test is used to detect outliers in the ground-based GPS data;  $N_{ionosonde}$  is the “real” electron density from the ionosonde profile.

The VED bias from both the outlier and the Shape Function model error is expressed as:

$$Bias_{inclu-all-biases}(i) = N_{without-w-test}(i) - N_{ionosonde}(i) \quad i = 1, \dots, n \quad (27)$$

where  $Bias_{inclu-all-biases}$  is the bias from both Step (1) (or the outlier) and Step (2) (or the Shape Function model error);  $N_{without-w-test}$  is the electron density generated by Equation (25), in which the  $W$ -test is not used when the  $VTEC$  is generated (the  $w$ -test can't calculate  $VTEC$ . It is just an outlier detection tool).

The VED bias only from the outlier (or outlier effect) is calculated by combining Equations (26) and (27) as:

$$Bias_{exclu-Shape-Function}(i) = Bias_{inclu-all-biases}(i) - Bias_{exclu-outlier}(i) \quad i = 1, \dots, n \quad (28)$$

where  $Bias_{exclu-Shape-Function}$  is the bias from the outlier and  $Bias_{exclu-outlier}$  is the bias from the Shape Function model error.

The VED profiles, which are related to one of the 3 biases shown in Equations (26)-(28), are computed as:

$$N_{exclu-outlier}(i) = N_{with-w-test}(i) \quad i = 1, \dots, n \quad (29)$$

$$N_{exclu-Shape-Function}(i) = N_{ionosonde}(i) + Bias_{exclu-Shape-Function}(i) \quad i = 1, \dots, n \quad (30)$$

$$N_{inclu-all-biases}(i) = N_{without-w-test}(i) \quad i = 1, \dots, n \quad (31)$$

where  $N_{exclu-outlier}$  is the retrieved electron density without the outlier bias;  $N_{exclu-Shape-Function}$  is the electron density without the Shape Function model bias; and  $N_{inclu-all-biases}$  is the electron density including both the outlier and the Shape Function model biases.

The numerical results from Equations (29)-(31) will be presented in the following Section.

### 3. Testing results and analysis

In this study, firstly, the TEC distributions are generated using the TSVD with/without the  $W$ -test. Secondly, based on the retrieved TEC, their related VED profiles are generated. Thirdly, the biases from the outlier and the Shape Function model are separated and analysed.

In the testing, the assumed ionospheric shell height was set as 300 km and the satellite cut-off angle was set as  $25^{\circ}$ . And one group of real data set A (local daytime), and a simulated data set B were used.

(1) Data set A includes:

(a) the ground based data from the eight Australian permanent GPS stations including ade1, alice, karr, mobs, str1, tid1, tow2 and yar2 in UTC 02:34:30 --- 02:35:00 (or local time 12:34:30 --- 12:35:00) on 16/10/2006.

(b) the RO data from the RO event occurred between COSMIC L001 and GPS31 at UTC 02:34:15 to 02:37:10 on 16/10/2006;

(c) 3 ionosonde profiles at Townsville, Canberra and Camden (at UTC 02:35, 02:33 and 02:35 respectively). They will be compared with the 3 retrieved VED profiles in the above 3 locations.

(2) Data set B (simulated data):

Using data set A, the STECs are generated. If one of the STEC is manually increased and its  $W$ -value reaches more than the threshold value, the observation related to this STEC is an outlier. In the simulation test, total 24 STECs (related to 3 different locations) are manually changed and 24 outliers are simulated. In this time, data set A is defined as data set B.

### 3.1 Test one --- using data set A

Fig. 1 shows the  $W$ -value variations with regards to the observations before (the red ‘\*\*\*’ line) and after (the ‘\_\_\_’ blue line) the  $W$ -test, indicating that there are observations with the  $W$ -value of bigger than the chosen threshold 2.58, which are also associated with the big residuals shown in Fig. 2 (the red ‘\*\*\*’ line and the ‘\_\_\_’ line related to before and after  $W$ -test): When the TSVD and the  $W$ -test algorithm were used, the 4 outliers were identified and removed. Consequently, as shown in Fig. 1, all the  $W$ -values for the remaining observations are smaller than the threshold; the residuals shown in Fig. 2 become smaller, resulting in more reliable estimates for the TEC parameters.

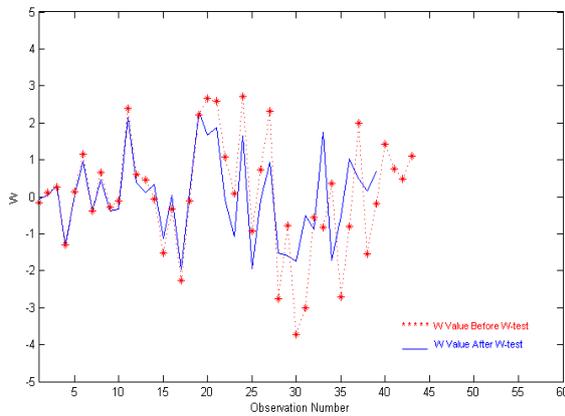


Figure 1:  $W$ -values for all the observations before and after outlier detection based on data set A

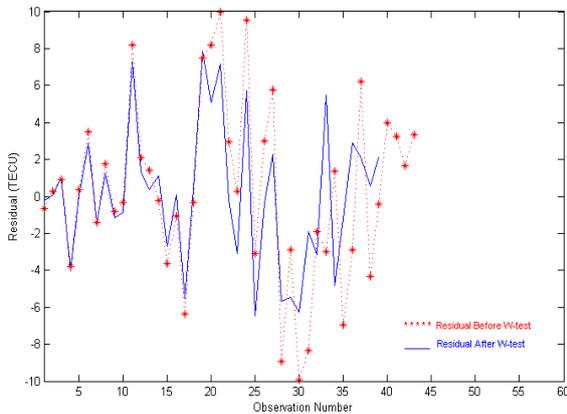


Figure 2: Residuals for all the observations before and after outlier detection based on data set A

Figs. 3-5 show the effects from the outlier and the Shape Function model error based on the retrieved VED profiles generated by Equations (29)-(31) at 3 locations including Townsville, Canberra and Camden. In Fig. 9, the red ‘\*\*\*’ line is the ‘real’ VED profile from the ionosonde. The blue ‘. . .’ line is the VED profile in which the outlier effect has been removed. The cyan ‘\_\_\_’ line shows the VED profile without the Shape

Function model error effect. The green ‘+++’ line is the retrieved VED profile which combines the effects from the outlier and the Shape Function model error.

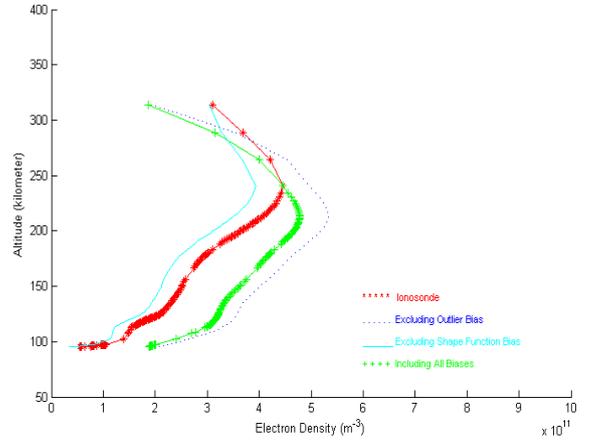


Figure 3: Retrieved VED profiles with different biases at Townsville based on data set A

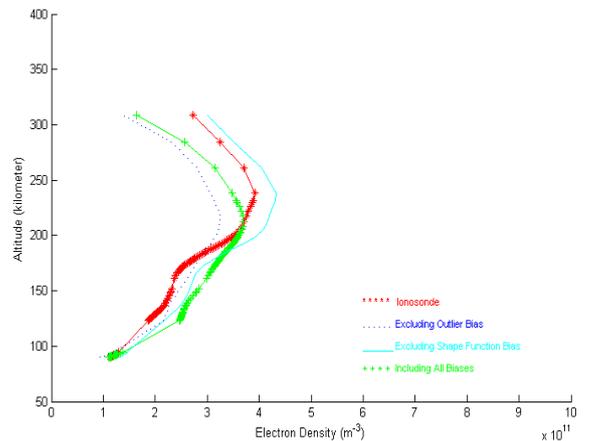


Figure 4: Retrieved VED profiles with different biases at Canberra based on data set A

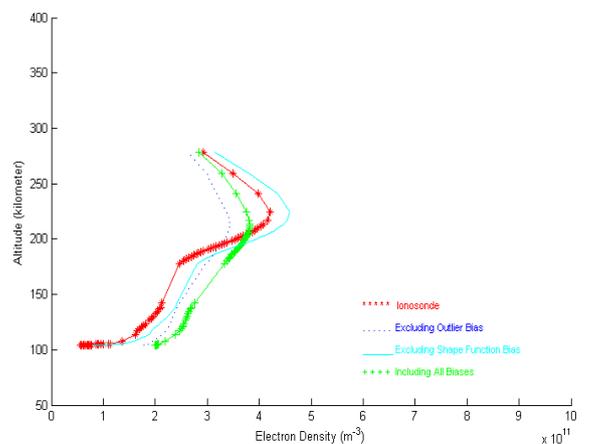


Figure 5: Retrieved VED profiles with different biases at Camden based on data set A

Based on the figure, the Shape Function model error makes the electron density become bigger (as the blue “ . . . ” line) in most of the layers of the ionosphere. The outlier impact is reverse (as the cyan “\_” line). And also this outlier effect is smaller than that from the Shape Function model error in most of the ionospheric layers because the cyan line has smaller deviation to the red line than the blue line. Similar to Fig. 3, Figs. 4-5 are the results at Canberra and Camden.

At these 2 locations, the influences from the outlier are all smaller than that from the Shape Function model error in most of the ionospheric layers. More detail comparisons about the VED profiles are showed in Table 1.

Table 1 has the following items:

(1)  $RMS_{relative}$  : quantitatively compare different VED profile accuracies as defined below:

$$RMS_{relative} = \frac{1}{\text{mean value}} \sqrt{\frac{\sum_{i=1}^n [N_{model}(i) - N_{ionosonde}(i)]^2}{n}} \quad (32)$$

where  $n$  is the total sample number in the VED profile, the  $N_{model}(i)$  is the  $i^{\text{th}}$  sample item from the VED profiles generated by the model, the  $N_{ionosonde}(i)$  is the  $i^{\text{th}}$  corresponding sample item in ionosonde VED profile and the  $\text{mean value}$  is the average value of the VED profile;

(2) “TSVD\_incl\_all\_bias”: related to the VED profile

which includes the both biases from the outlier and the Shape Function model, generated by Equation (31);

(3) “Deviation\_incl\_all\_bias”: the item (2) (including  $N_m F_2$  and  $h_m F_2$ ) minus the ionosonde value, which can be positive (retrieved  $N_m F_2 / h_m F_2$  more than that from the ionosonde) or negative (retrieved  $N_m F_2 / h_m F_2$  less than that from the ionosonde);

(4) “TSVD\_excl\_outlier”: related to the VED profile which only includes the Shape Function model error impact, generated by Equation (29);

(5) “Deviation\_excl\_outlier”: the item (4) (including  $N_m F_2$  and  $h_m F_2$ ) minus the ionosonde value, which can be positive or negative;

(6) “TSVD\_excl\_Shape\_Funct”: related to the VED profile which only includes the outlier impact, generated by Equation (30);

(7) “Deviation\_excl\_Shape\_Funct”: the item (6) (including  $N_m F_2$  and  $h_m F_2$ ) minus the ionosonde value;

(8) “Outlier Effect (mean)”: describe the outlier impact and as defined below:

$$\text{Outlier Effect (mean)} = \frac{1}{n} \left( \sum_{i=1}^n \frac{\text{Bias}_{\text{exclu-Shape-Function}}(i)}{N_{ionosonde}(i)} \right) \quad (33)$$

where  $\text{Bias}_{\text{exclu-Shape-Function}}$  is the outlier effect and is calculated by Equation (31).

Table 1: Comparisons among the VED profiles from different modelling procedures on 16/10/2006

	Townsville (UTC 02:35)			Canberra (UTC 02:33)			Camden (UTC 02:35)		
	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)
Ionosonde	-	4.5	242	-	4.0	238	-	4.21	228
TSVD_incl_all_bias	36.49	4.78	212	15.62	3.69	213	29.57	3.82	214
Deviation_incl_all_bias	-	0.28	-30	-	-0.31	-25	-	-0.42	-14
TSVD_excl_outlier	45.95	5.33	215	13.31	3.25	213	25.22	3.43	214
Deviation_excl_outlier	-	0.83	-27	-	-0.75	-25	-	-0.78	-14
TSVD_excl_Shape Funct	17.92	3.94	241	13.13	4.33	238	12.63	4.59	224
Deviation_excl_Shape Funct	-	-0.55	-1	-	0.33	0	-	0.38	-4
Outlier Effect (mean)	17.63	-	-	13.79	-	-	16.46	-	-

As shown in Table 1, at Townsville, the  $RMS_{relative}$  of the VED including both the outlier and the Shape Function model error effects is 36.49%. It is smaller than the  $RMS_{relative}$  excluding the outlier impact (45.95%). The reason is that the outlier makes the retrieved VED

smaller than the “real” one and the Shape Function model error makes the retrieved VED bigger than the “real” one in most of the layers of the ionosphere as Fig. 3 displays. Therefore, these two impacts compensate each other and generate the above results.

Also, the  $RMS_{relative}$  of the VED excluding Shape Function model error influence (17.92%) is smaller than the  $RMS_{relative}$  of the VED excluding the outlier influence (45.95%).

In  $N_m F_2$  and  $h_m F_2$ , the deviations excluding the outlier impact are  $0.83 \times 10^{11}/m^3$  and  $-27km$ , and the deviations excluding the Shape Function model error are  $-0.55 \times 10^{11}/m^3$  and  $-1km$  which, in the absolute value, are smaller than the first pair. The combined deviations including both the outlier and the Shape Function impacts are  $0.28 \times 10^{11}/m^3$  and  $-30km$  respectively. The Outlier Effect (mean) is 17.63% which is close to the  $RMS_{relative}$  of the VED excluding the Shape Function model error impact (17.92%).

At Canberra, The  $RMS_{relative}$  of VEDs excluding the outlier effect and the  $RMS_{relative}$  of VED excluding the Shape Function error are 13.31% and 13.13% respectively. They are close each other. In  $N_m F_2$  and  $h_m F_2$ , the deviations excluding the outlier effect are  $-0.75 \times 10^{11}/m^3$  and  $-25km$ , and the deviations excluding the Shape Function model error are  $0.33 \times 10^{11}/m^3$  and  $0km$  which, in the absolute value, are still smaller than the first pair. The combined deviations including both the outlier and the Shape Function impacts are  $0.31 \times 10^{11}/m^3$  and  $-25km$ . The Outlier Effect (mean) is 13.79%.

At Camden, The  $RMS_{relative}$  of VEDs excluding the outlier influence and the  $RMS_{relative}$  of VED excluding the Shape Function model error are 25.22% and 12.63% respectively. When the outlier impact and the Shape Function model error act simultaneously, a bigger  $RMS_{relative}$  (29.57%) including both the outlier and the Shape Function model error impacts is generated. In  $N_m F_2$  and  $h_m F_2$ , the deviations excluding the outlier impact are  $-0.78 \times 10^{11}/m^3$  and  $-14km$ , and the deviations excluding the Shape Function model error are  $0.38 \times 10^{11}/m^3$  and  $-4km$  which, based on the absolute value, are still smaller than the first pair. The combined deviations including both the outlier and the Shape Function effects are  $0.42 \times 10^{11}/m^3$  and  $-14m$ . The Outlier Effect (mean) is 16.46%.

In Summary, the results have shown that, together with the ionosonde data, the  $W$ -test can numerically separate the two biases from the outlier impact and the Shape Function model error. Generally speaking, the Shape Function model error has a bigger effect on the  $RMS_{relative}$  of the retrieved VED profile, the  $N_m F_2$  and the  $h_m F_2$  than that from the outlier impact.

In this data set, four outliers were detected and their effects on  $RMS_{relative}$  are 17.63%, 13.79% and 16.46% in Townsville, Canberra and Camden, respectively. These outliers have a big negative influence on the accuracy of the estimated TEC model. More numerical results generated by simulated data are shown in the following Section.

### 3.2 Test two --- using data set B

This section further demonstrates the impact of outlier magnitude on the retrieved VED profile through using the  $W$ -test method within the traditional TSVD by the simulated outliers. Using data set A, the calculated STEC from the 5<sup>th</sup> observation was 9.94TECU and was not an outlier (with a  $W$ -value of 0.14, which is smaller than the threshold of 2.58). But, we intended to manually increase that value about 3 times and it reaches to the value of 30TECU for this observation, which is obviously an outlier, then the daytime data has 5 outliers. Fig. 6 shows that the absolute  $W$ -value related to the 5<sup>th</sup> observation data reached to 6.65 ( $>2.58$ ), and the absolute residual related to the 5<sup>th</sup> observation was increased to 17.79 TECU as shown in Fig. 7.

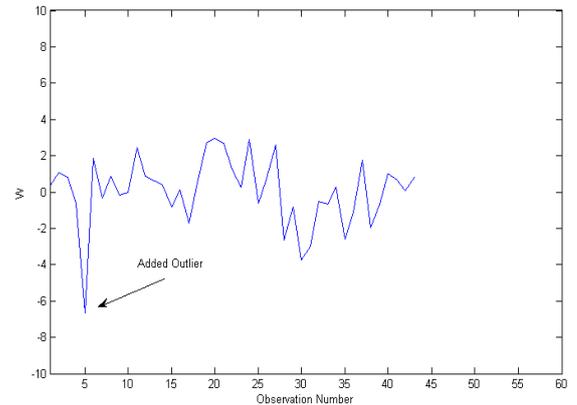


Figure 6:  $W$ -values for all the observations before the outlier detection based on data set B

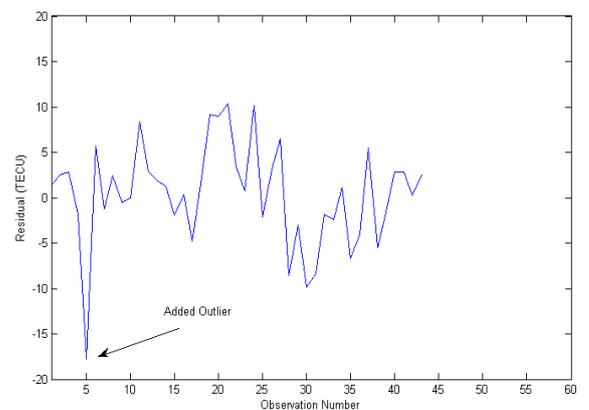


Figure 7: Residuals for all the observations before the outlier detection based on data set B

Table 2 shows the effects from the outlier and the Shape Function model error when the STEC from the 5<sup>th</sup> observation data is simulated as 30TECU. At Townsville, Canberra and Camden, the  $RMS_{relative}$ s of the VED profiles excluding the Shape Function model error (or only including the outlier influence) are increased to 27.59%, 15.55% and 13.4% from 17.92%, 13.13% and 12.63% (as in Table 1) respectively. Also, In  $N_m F_2$ , the deviations from the outlier impact reach  $-0.77 \times 10^{11}/m^3$ ,  $0.44 \times 10^{11}/m^3$  and  $0.43 \times 10^{11}/m^3$  from  $-0.55 \times 10^{11}/m^3$ ,  $0.33 \times 10^{11}/m^3$  and  $0.38 \times 10^{11}/m^3$  (as in Table 1) at the above 3 locations respectively. In  $h_m F_2$ , when the outlier is added, the deviations from the outlier impact are still as -1km, 0km and -4km and there are no variation at the above 3 locations. Finally, the Outlier

Effect (mean) at these locations are increased to 26.76%, 16.27% and 16.71% from 17.63%, 13.79% and 16.46% (as in Table 1) respectively, which are consistent with the  $RMS_{relative}$  of the VED profiles excluding the Shape Function model error.

To further demonstrate the impact of the outlier magnitude on the accuracy of the VED, which is measured by two indexes: the  $RMS_{relative}$  excluding the Shape Function model error and the Outlier Effect (mean), the 5<sup>th</sup> observation STEC was changed to 30, 40, ..., 100 TECU (as the simulated outliers), and the variations of the above two indexes are shown in Table 3.

Table 2: Comparisons among the VEDs from different modelling procedures for the simulated outlier data set on 16/10/2006 5<sup>th</sup> STEC as 30 TECU (original 9.94TECU)

	Townsville (UTC 02:35)			Canberra (UTC 02:33)			Camden (UTC 02:35)		
	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)	$RMS_{relative}$ (%)	$N_m F_2$ ( $10^{11}$ )	$h_m F_2$ (km)
Ionosonde	-	4.5	242	-	4.0	238	-	4.21	228
TSVD_incl_all_bias	31.52	4.55	213	16.92	3.8	213	29.51	3.86	214
Deviation_incl_all_bias	-	0.05	-29	-	-0.2	-25	-	-0.35	-14
TSVD_excl_outlier	45.94	5.33	215	13.31	3.25	213	25.22	3.43	214
Deviation_excl_outlier	-	0.83	-27	-	-0.75	-25	-	-0.78	-14
TSVD_excl_Shape_Funct	27.59	3.73	241	15.55	4.44	238	13.4	4.64	224
Deviation_excl_Shape_Funct	-	-0.77	-1	-	0.44	0	-	0.43	-4
Outlier Effect (mean)	26.76	-	-	16.27	-	-	16.71	-	-

Table 3: Outlier Magnitude Variation and Its Effect

16/10/2006										
TEC related to the 5 <sup>th</sup> observation (TECU)		Original	Manually Increased to (outlier simulation)							
		9.94	30	40	50	60	70	80	90	100
Townsville UTC 02:35	$RMS_{relative}$ (%)	17.92	27.59	32.66	38.04	43.73	49.77	56.22	63.07	70.37
	TSVD_excl_Shape_Funct									
	Outlier Effect (mean)	17.63	26.76	31.23	35.73	40.28	44.86	49.49	54.15	58.85
Canberra UTC 02:33	$RMS_{relative}$ (%)	13.13	15.55	16.76	17.97	19.19	20.41	21.61	22.83	24.05
	TSVD_excl_Shape_Funct									
	Outlier Effect (mean)	13.79	16.27	17.48	18.71	19.95	21.2	22.43	23.7	24.97
Camden UTC 02:35	$RMS_{relative}$ (%)	12.63	13.4	13.83	14.27	14.72	15.18	15.64	16.12	16.62
	TSVD_excl_Shape_Funct									
	Outlier Effect (mean)	14.46	16.71	16.86	16.99	17.13	17.25	17.36	17.47	17.57

As shown in Table 3, the ‘‘Original’’ column shows the situation when no outliers are added and there are only 4 original outliers as discussed in Section 3.1. Based on the results in Table 3, at Townsville, the  $RMS_{relative}$  excluding the Shape Function model error increases significantly from 17.92% to 70.37% when the STEC reaches 100TECU from 9.94TECU. Also, the Outlier

Effect (mean) has a big jump too (from 17.63% to 58.85%). At Canberra, the increase is not very big. The  $RMS_{relative}$  excluding the Shape Function model error increases from 13.13% to 24.05% and Outlier Effect (mean) is from 13.79% to 24.97%. At Camden, above two indexes change much small, they are only from

12.63% to 16.62% (the first index), and from 14.46% to 17.57% (the second index).

In summary, the magnitude of the outlier is proportional to its impact on the estimated VED, and also this effect is location dependent.

#### 4. Concluding Remarks

The initial testing results have demonstrated that the proposed quality control procedure with the outlier detection and exclusion can have significant impacts on the TEC modelling results based on the spherical harmonic function and the Shape Function models. It is expected that such a quality control procedure could also be implemented in other TEC modelling processes where the least-squares or Kalman filtering method is used.

Furthermore, the proposed method can also increase the accuracy of the Shape Function model. Because this model can be influenced by two groups of error sources: the outlier (from the TEC model) and the Shape Function model itself. The Shape Function model error is systematic in nature and can be analysed, while the outliers can be eliminated using the proposed quality control procedure. Based on the biases separation method, after the Shape Function model biases at some locations for the target region are calculated, the distribution map of such Shape Function model biases in this target region at a given time, can be generated by an interpolation technology. Consequently, if we want to calculate the VED profile at a given location within the target region, the Shape Function model bias at this location can be first acquired by the retrieved distribution map of this Shape Function model biases, and then is removed, resulting in more accurate VED profiles.

#### Acknowledgments

This research is sponsored by an Australian Research Council Linkage APAI Project with the industry partner, IPS Radio and Space Services, Australian Bureau of Meteorology. The ionosonde data used in this paper were provided by Dr Mike Terkildsen from the IPS.

#### References

- Baarda, W. (1968) *A testing procedure for use in geodetic networks*. Netherlands Geod.Comm., New Series, 2(5), Delft.
- Davies K (1990) *Ionospheric Radio, Institution of Electrical Engineers*. Peter Peregrinus Ltd Press, United Kingdom.

- Fjeldbo G, Eshleman VR (1968) *The Atmosphere of Mars Analyzed by Integral Inversion of the Mariner IV Occultation Data*. Planet. Space Sci., 16, 1035-1059.
- Garcia-Fernandez M (2004) *Contribute to the 3D Ionospheric Sounding with GPS data*, PhD thesis, Universitat Politecnica de Catalunya (UPC), Spain.
- Hajj GA, Romans LJ (1998) *Ionospheric electron density profiles obtained with the Global Positioning System: Results from the GPS/MET experiment*. Radio Science, 33, 175-190.
- Hajj GA, Wilson BD, Wang C, Pi X, Rosen IG (2004) *Data Assimilation of Ground GPS Total Electron Content into a Physics-Based Ionospheric Model by Use of the Kalman Filter*. Radio Science, 39, RSIS05,doi:10.1029/2002RS002859.
- Hajj GA, Kursinski ER, Romans LJ, Bertiger WI, Leroy SS (2002) *A Technical Description of Atmospheric Sounding by GPS Occultation*. Journal of Atmospheric and Solar-Terrestrial Physics, 64: 451-469.
- Hartl DA (2007) *Tomographic Reconstruction of 2-D Atmospheric Trace Gas Distribution from Active DOAS Measurements*, PhD thesis, University of Heidelberg, Germany.
- Hernandez-Pajares M, Juan M, Sanz J (2000) *Improving the Abel inversion by adding ground GPS data to LEO radio occultation in ionospheric sounding*. Geophysical Research Letter, 27(16): 2473-2476
- Kaplan ED (1996) *Understanding GPS: Principles and Applications*. Artech House Publishers, Boston.
- Komjathy A (1997) *Global Ionospheric Total Electron Content Mapping Using the Global Positioning System*, PhD thesis, University of Brunswick, Canada
- Liao X (2000) *Carrier Phase Based Ionosphere Recovery Over A Regional Area GPS Network*, PhD thesis, University of Calgary, Canada.
- Liu Z (2004) *Ionosphere Tomographic Modeling and Applications Using Global Positioning System (GPS) Measurements*, PhD thesis, University of Calgary, Canada.
- Liu Z, Gao Y, Skone S (2005) *A Study of Smoothed TEC Precision Inferred from GPS Measurement*, Earth Planets Space, 57: 999-1007.

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, 54:13-16.

Strang G, Borre K (1997) ***Linear Algebra, Geodesy and GPS***. Wellesley-Cambridge Press, America.

Wang J, Chen Y (1999) ***Outlier detection and reliability measures for singular adjustment models***, Geomatics Research Australia, 71: 57-72.

Xue XJ, Kozaczek KJ, Kurtz SK, Kurtz DS (2000) ***A Direct Algorithm for Solving-conditioned Linear Algebraic Systems***, Advances in X-ray Analysis, 42: 629-632

### **Biography**

Dr Jinling Wang is an Associate Professor in the School of Surveying & Spatial Information Systems, University of New South Wales (UNSW). His major research interest is in the areas of geo-positioning, mapping and navigation using multi-sensors. He is a Fellow of the Royal Institute of Navigation, UK, and a Fellow of the International Association of Geodesy (IAG). Dr Wang is a member of the Editorial Board for the international

journal GPS Solutions, and was Chairman of the study group (2003-2007) on pseudolite applications in positioning and navigation within the IAG's Commission 4. He was elected 2004 President of the International Association of Chinese Professionals in Global Positioning Systems (CPGPS), and the Founding Editor-in-Chief for the Journal of Global Positioning Systems (2002-2007), Conference Convener of the 2004 International Symposium on GPS/GNSS in Sydney, Australia; Technical Chair for the world's premier GPS/GNSS conference (ION GNSS2008) in the USA. He holds a PhD in GPS/Geodesy from the Curtin University of Technology, Perth, Australia. He has published over 200 papers in journals and conference proceedings as well as two commercial software packages. to find more details about the research activities within his research group at UNSW: Visit website: [www.gmat.unsw.edu.au/wang](http://www.gmat.unsw.edu.au/wang).