Interpolation Estimation Method of Tropospheric Delay for Long Baseline Network RTK Based on Support Vector Machine

Jian Deng[1,2], Miaoqiang Xu[3,*], Xuexiang Yu[3] and Aiguo Zhang [1,2]

[1]School of Computer and Information Engineering, Xia men University of Technology, Xia men 361024, China

[2]Big Data Institute of Digital Natural Disaster Monitoring in Fujian, Xia men University of Technology, Xia men 361024, China

[3] Surveying and Mapping Institute, Anhui University of Science and Technology, Huai nan 232000, China

*Correspondence: xmq2618@163.com

Abstract: In order to solve the problem of low precision of tropospheric delay interpolation under the long-distance sparse reference station, a method of tropospheric delay interpolation estimation based on support vector machine (SVM) theory was proposed. Firstly, the troposphere was assumed to be an infinitely thin single-layer membrane, and with the known information of reference stations, the puncture point coordinates and zenith tropospheric delay of each reference station on each visible satellite in the single-layer membrane were obtained. Then, the puncture point coordinates and zenith tropospheric delay were taken as training samples to optimize the appropriate core parameters of SVM and establish the SVM model of tropospheric delay. Finally, two sets of network RTK data with different lengths were selected to compare and analyze the effect of the interpolated tropospheric delay. The results showed that under the long-distance reference station, with the SVM tropospheric delay model established in this paper, the tropospheric delay accuracy of the interpolation estimation is better than 2cm, and the estimation error is generally stable, and the interpolation effect of satellite in different systems is basically the same.

1. Introduction

Network RTK is a high precision and real-time relative positioning technology based on global navigation satellite system (GNSS). This technology refers to establishing a plurality of GNSS reference stations more uniformly in an area, forming mesh coverage for the area, so as to establish a spatial error model based on the known coordinates of the reference stations and the satellite information observed in real time. With this model, GNSS users in the region can obtain error correction information such as tropospheric and ionospheric delay [1], thereby improving the real-time and accuracy of positioning. It can provide users with real-time centimetre positioning accuracy, which has been widely used in various fields such as power energy infrastructure construction, informatization, energy and environmental engineering.

At present, the distance between reference stations in network RTK is generally 30~70km. It is assumed that if the distance between reference stations can be extended, the number of reference stations will be reduced within the same area, which will greatly reduce the costs of system construction and maintenance. This is of great strategic significance for the establishment of a network RTK service system under the sparse reference stations for the underdeveloped western regions, poor

environmental conditions, and even the whole country. Therefore, the research on the key technology of network RTK under long-distance reference stations environment has important theoretical significance and practical value. For the network RTK technology, the ultimate goal is to use the spatial correlation errors (mainly ionospheric and tropospheric error) that the reference stations have solved, and combine the spatial position relationship between the mobile station and the surrounding reference stations to interpolate the differential correction information of mobile users in real time, so as to achieve centimetre-level high-precision positioning. Therefore, the accuracy of interpolation correction in tropospheric spatial error model is directly related to the validity and reliability of RTK positioning, which is one of the key technologies of network RTK.

The research on interpolation model for network RTK was relatively mature at home and abroad, and many improved theories and methods were put forward to solve the deficiency of traditional model, and good results had been achieved [2]-[5]. Li [6] established the multivariate linear regression model, combined estimation model, grey model and BP neural network model of tropospheric delay in short baseline network RKT, and proposed a weight coefficient determination method based on reliability of the combined model to eliminate the influence of tropospheric delay. With genetic algorithm and BP neural network technology, Chen [7] established a high-precision regional fusion model (GA-BPEGNOS model) based on EGNOS model, which improved the accuracy of tropospheric delay correction for short baseline network RTK. However, most of the existing researches were aimed at the study of spatial error correction within the network RTK coverage under medium and long baseline (<100km). When the distance between reference stations increases (150-200km), the spatial error correlation decreases, and the accuracy of tropospheric delay estimated by the conventional model also decreases. Therefore, this paper introduced the theory of support vector machine (SVM). The troposphere was assumed to be an infinite thin single-layer film, and the SVM model of regional tropospheric delay was established to realize the real-time estimation of tropospheric delay of user terminal.

2. Support vector machine

SVM [8], [9] is a two-class classification model, and also is a machine learning method based on the principle of structural risk minimization. Its essence is to transform the input space into a high-dimensional space by using a nonlinear transformation defined by the inner product function, and seek a nonlinear relationship between the input variable and the input variable in this high-dimensional space. Because the SVM algorithm has a rigorous mathematical theory foundation and good function approximation ability, it can solve practical problems such as finite samples, nonlinearity and high dimension, which makes it widely used in the field of regression.

The specific algorithm of the SVM method for nonlinear regression estimation is as follows [10]. Assuming that the data sample is $(x_i, y_i), i = 1, 2, \ldots, n; x_i \in R^m$, which are input parameters; $y_i \in R$, which are the output parameters. n is the sample number. For linear regression, let the linear regression function $f(x) = <w, x> + b$.w is the weight vector; b is the offset. If all the training data are fitted with linear function under the precisionε, then

$$\begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i, \ \xi_i \geq 0 \\ f(x_i) - y_i \leq \varepsilon + \xi_i^*, \ \xi_i^* \geq 0 \end{cases} \qquad i = 1, 2, \ldots, n \qquad (1)$$

Among them, ξ_i and ξ_i^* are relaxation factors. Based on the principle of structural risk minimization, the problem can be transformed into the minimization optimization problem with constraint condition (1) seeking (2)

$$\frac{1}{2} ww + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \qquad (2)$$

Where, C is the penalty factor.

According to the duality theory of nonlinear programming, the Lagrangian equation is established to convert the minimum value problem into the maximum value problem of the dual problem

$$f(x) = \sum_{i=1}^{n} (a_i + a_i^*) <x, x_i> + b \qquad (3)$$

Where, x_i is the support vector, a_i，a_i^* and b are the parameters obtained by the regression.

For the nonlinear regression estimation, a Kernel function $K(x_i, x_j)$ satisfying Mercer can be introduced instead of the inner product $\Phi(x_i)\Phi(x_j)$ of the mapping function, and the nonlinear regression function is obtained.

$$f(x) = \sum_{i=1}^{n}(a_i + a_i^*)K(x_i, x) + b \qquad (4)$$

There are three main types of kernel functions in use:

Linear kernel $K(x_i, x_j) = <x_i, x_j>$;
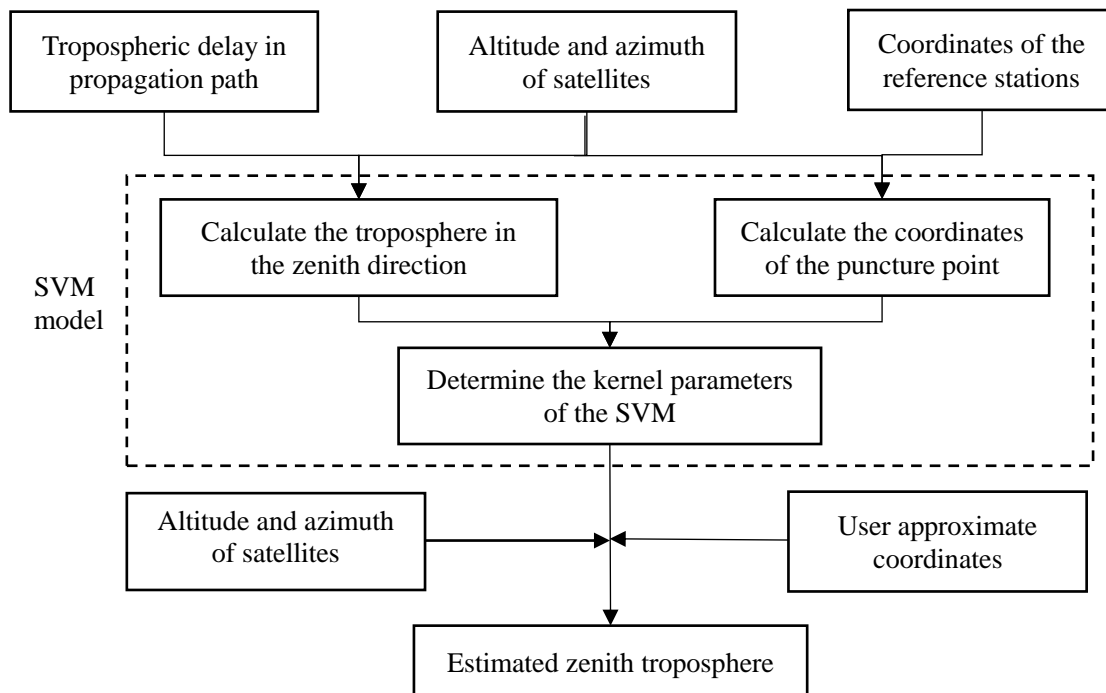
Polynomial kernel $K(x_i, x_j) = (<x_i, x_j> + 1)^d$;

Gaussian radial basis kernel $K(x_i, x_j) = exp(-\|x_i - x_j\|^2 / 2\sigma^2)$

Where, d is the degree of polynomials, and σ is the width of the Gaussian distribution.

3. Establishment of tropospheric model based on SVM

3.1 Overall process of model establishment

Figure 1 showed the process of SVM model establishment. The regional spatial troposphere was assumed to be an infinitely thin single-layer membrane [11]. Firstly, with the known information of reference stations, the puncture point coordinates of each reference station on each visible satellite in the single-layer membrane were obtained. Secondly, the tropospheric delay of each reference station corresponding to each satellite in the propagation path was transformed into the tropospheric delay in the zenith direction at the puncture point. Then, the puncture point coordinates and zenith tropospheric delay were taken as training samples to optimize the appropriate core parameters of SVM and establish the SVM model of tropospheric delay. Finally, for any receiver within the coverage of the network RTK, using the constructed model and taking the coordinates of the puncture point corresponding to the mobile user as input, the satellite tropospheric delay of the receiver can be estimated. In this way, the real-time and precision of RTK positioning were improved.

Figure 1. Process of SVM model establishment

3.2 Determination of the coordinates of the puncture point

First of all, the angle named α between the receiver station to the center of the earth and the satellite to the center of the earth is calculated [10].

$$\alpha = \frac{\pi}{2} - E - \arcsin(\frac{R}{R+H}cosE) \qquad (5)$$

Where E is the satellite elevation angle corresponding to the receiver; R is the radius of the earth, with a value of 6378.1363km; H is the height of the central troposphere, and the average height of the troposphere is different at different latitudes. The average troposphere height is 17-18 km in low latitudes, 10-12 km in the middle latitudes, and 8-9 km in the polar troposphere. Then the longitude λ_p and the latitude φ_p of the puncture point are:

$$\begin{cases} \lambda_p = \lambda_U + \arcsin(\frac{sin\alpha sinA}{cos\varphi_p}) \\ \varphi_p = \arcsin(sin\varphi_U cos\alpha + cos\varphi_U sin\alpha cosA) \end{cases} \qquad (6)$$

In the formula, (λ_U, φ_U) is the latitude and longitude of the receiver station; A is the satellite azimuth corresponding to the receiver station.

3.3 Calculation of the zenith troposphere

At different times, the elevation angle of the satellite to the monitoring station is constantly changing, and the value of the troposphere is related to the length of the propagation path of the radio wave. Therefore, the larger the elevation angle of the satellite to the monitoring station, the smaller the tropospheric delay. Conversely, the greater the tropospheric delay. In order to calculate the tropospheric delay at different times more accurately, the tropospheric delay is converted from the propagation path direction to the zenith direction of the puncture point.

Combined with the propagation path of tropospheric delay DDT and satellite altitude angle E, the conversion formula for tropospheric delay calculation is as follows [12]:

$$VDDT = DDT \cdot cos\beta \qquad (7)$$
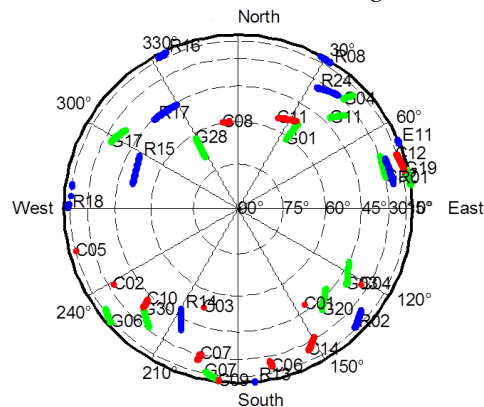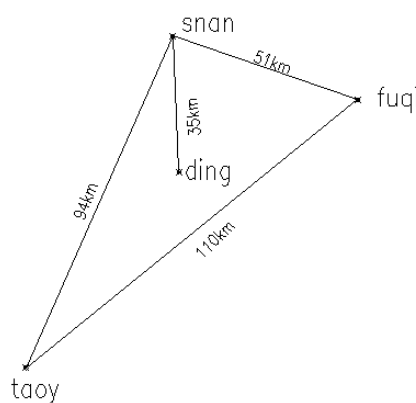
$$\beta = \arcsin(\frac{R}{R+H}cosE) \qquad (8)$$

Where β is the zenith distance of the satellite at the puncture point.

3.4 Model establishment and interpolation estimation

Based on the puncture point coordinates and zenith tropospheric delay, the Gaussian radial basis kernel function was selected from the Libsvm toolbox in matlab platform. Firstly, with a part of the sample data, the approximate values of C and σ were determined through the SVM based on the particle swarm optimization algorithm. Then, all the sample data were used to conduct multiple experiments in the rough range, and finally the values of kernel function parameters and penalty factors were determined, so as to establish the tropospheric delay model of SVM. For the RTK users in the area, as long as the current approximate location was uploaded to the data service centre, the established interpolation model can be used to estimate the tropospheric delay of each satellite corresponding to the station, in this way, the precision of RTK positioning was improved.

4. Experiments and Discussion

4.1 Case1: Estimation of tropospheric delay in network RTK under medium and long distance
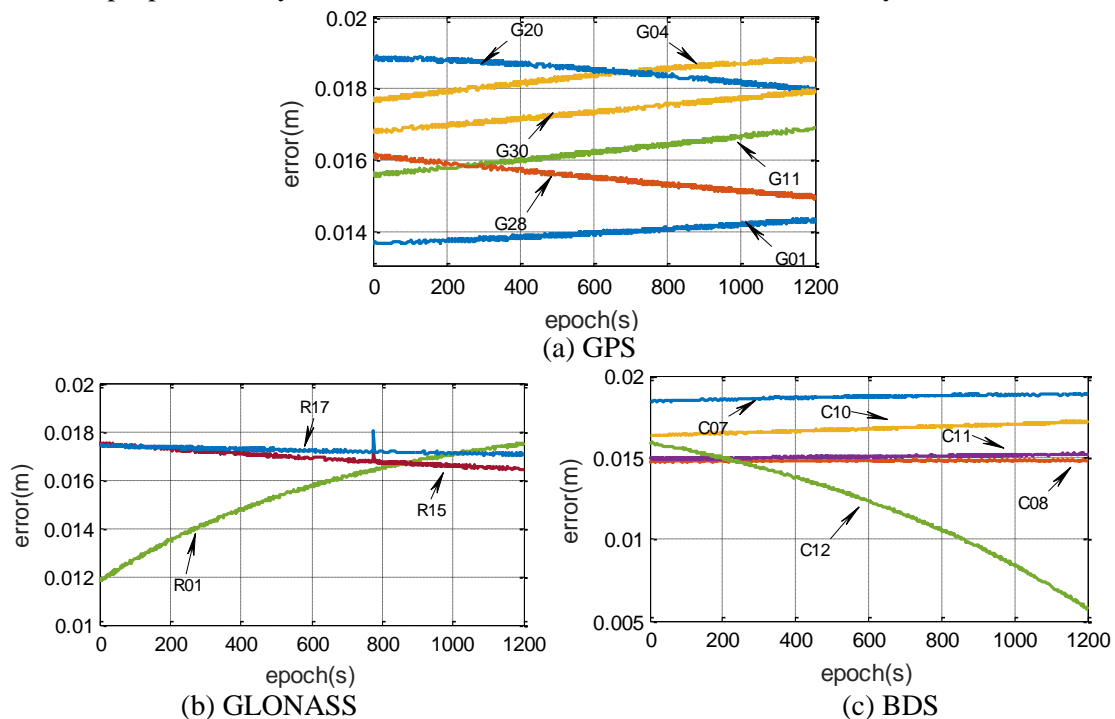

Figure 2. Distribution map of stations in suzhou network RTK   **Figure 3.** Distribution map of satellites

Four reference stations were selected from Suzhou network RTK in China to verify the interpolating effect of the tropospheric model based on SVM in the medium and long distance. Figure 2 shows the approximate location distribution of the reference stations. Among them, snan, fuqi and taoy stations were set as reference stations, and the distance between each reference station is 51km, 94km and 110km, respectively, ding station as the user receiver station, about 35km away from the main reference station. The data of each station was selected from 21:20:01 to 21:40:00 on December 18, 2014 for 1200 seconds. Each second of the selected data includes GPS, GLONASS, and BDS satellite systems, figure 3 shows the distribution of all visible satellites during the observation period. After data pre-processing, the number of GPS, GLONASS and BDS satellites in the available satellites with elevation angles above 20 degrees is 6, 3 and 5 respectively. In practical applications, it is necessary to construct a SVM model and perform tropospheric delay estimation based on the reference stations and the observed satellite information every second.

In each second, firstly, according to the known coordinates of the three reference stations and the current altitude angle and azimuth of satellites, the coordinates of puncture point for each satellite corresponding to each station were calculated by formula (6). Then, using equation (7), the tropospheric delay of the three reference stations corresponding to each satellite were converted from the propagation path to the zenith direction at the puncture site. Due to the Suzhou network RTK belongs to the mid-latitude area, the average height H of the troposphere was determined to be 11 km; Finally, the coordinates of each puncture point were taken as input values, corresponding to the zenith tropospheric delay as output values, so as to determine the values of the kernel function parameter and the penalty factor. Thereby the tropospheric delay SVM model was established at the moment.

On the basis of the above, the latitude and longitude of puncture point for station ding were taken as the input values, and the zenith tropospheric delay at the puncture point of each satellite was estimated by using the SVM model established at the current time, and then the delay was converted from the zenith direction of the puncture point to the propagation path direction. Finally, compared with the actual value, the estimated effect of the model is checked. Figure 4 shows the difference between the estimated tropospheric delay and the actual value of each satellite in the three systems.

(a) GPS

(b) GLONASS

(c) BDS

Figure 4. Comparison between tropospheric delay estimated by SVM and actual value
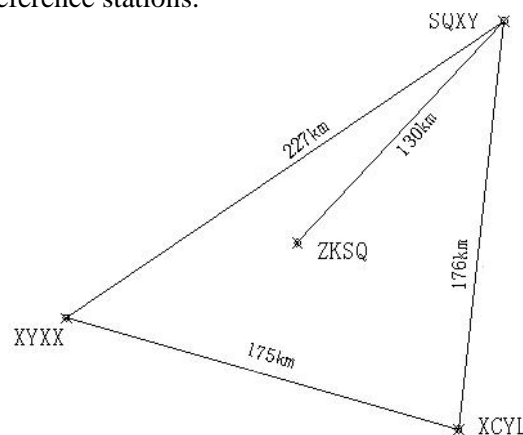
As can be seen from Fig.4, in this case, tropospheric delay estimation of GPS, GLONASS and BDS all achieve a good estimation effect. The estimated value is relatively close to the actual value, and the error fluctuation is small. Table 1 shows the maximum, median and average errors of each system in tropospheric delay estimation. Under the medium and long distance network RTK environment, the maximum deviation of tropospheric delay estimated for GPS, GLONASS and BDS are all less than 2cm, and the median error is between 0.01cm and 0.02 cm, with high estimation accuracy. Among them, the mean error of the tropospheric delay for BDS satellite is the smallest, followed by GPS system and GLONASS system, but in general, the mean errors of the three systems are relatively close, all of which are around 1.6cm.

Table 1 comparison between tropospheric delay estimated and actual value.

Serial number	Satellite number	Maximum error / m	Medium error / m	Serial number	Satellite number	Maximum error / m	Medium error / m
1	G01	0.0144	0.0140	8	R15	0.0175	0.0169
2	G04	0.0189	0.0183	9	R17	0.0181	0.0172
3	G11	0.0169	0.0162	10	C07	0.0189	0.0187
4	G20	0.0189	0.0185	11	C08	0.0149	0.0148
5	G28	0.0162	0.0155	12	C10	0.0172	0.0168
6	G30	0.0180	0.0174	13	C11	0.0152	0.0151
7	R01	0.0175	0.0155	14	C12	0.0159	0.0122
GPS Average error / m		0.0166	**GLONASS Average error / m**		0.0165	**BDS Average error /m**	0.0154

4.2 Case2: Estimation of tropospheric delay in network RTK under long distance
Four reference stations were selected from Henan network RTK in China to verify the interpolating effect of the tropospheric model based on SVM in long distance. Figure 5 shows the approximate location distribution of the reference stations.
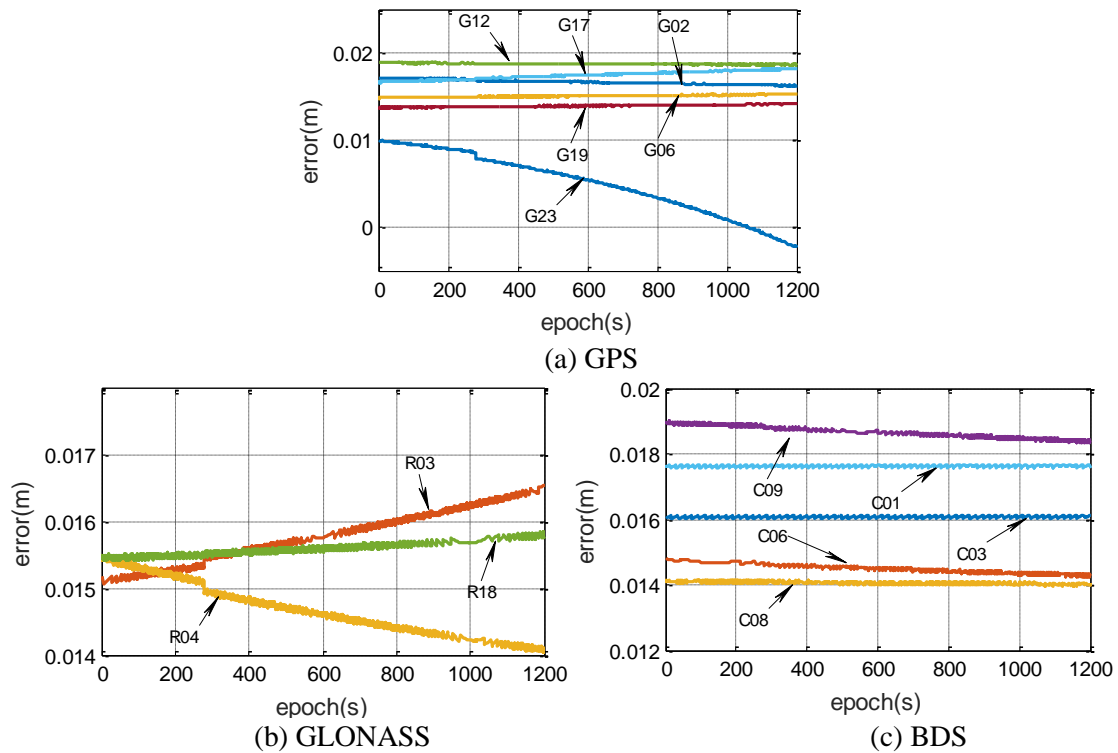

Figure 5. Distribution map of stations in henan network RTK

Among them, SQXY, XYXX and XCYL stations were set as reference stations, and the distance between each reference station is 175km, 227km and 176km, respectively, ZKSQ station as the user

receiver station, about 130km away from the main reference station. The data of each station was selected from 19:30:00 to 19:49:59 on March 1, 2016 for 1200 seconds. Each second of the selected data includes GPS, GLONASS, and BDS satellite systems. After data pre-processing, the number of GPS, GLONASS and BDS satellites in the available satellites with elevation angles above 20 degrees is 6, 3 and 5 respectively.

(a) GPS

(b) GLONASS

(c) BDS

Figure 6. Comparison between tropospheric delay estimated by SVM and actual value

Table 2 comparison between tropospheric delay estimated and actual value.

Serial number	Satellite number	Maximum error/ m	Medium error / m	Serial number	Satellite number	Maximum error / m	Medium error / m
1	G01	0.0144	0.0140	8	R15	0.0175	0.0169
2	G04	0.0189	0.0183	9	R17	0.0181	0.0172
3	G11	0.0169	0.0162	10	C07	0.0189	0.0187
4	G20	0.0189	0.0185	11	C08	0.0149	0.0148
5	G28	0.0162	0.0155	12	C10	0.0172	0.0168
6	G30	0.0180	0.0174	13	C11	0.0152	0.0151
7	R01	0.0175	0.0155	14	C12	0.0159	0.0122
GPS Average error / m		0.0166	**GLONASS Average error / m**		0.0165	**BDS Average error /m**	0.0154

The method flow of establishing regional SVM tropospheric model and estimating tropospheric delay of user station in real time is the same as case 1. At last, the estimated tropospheric delay is compared with the actual value so as to check the estimated effect of the model. Figure 6 shows the difference between the estimated tropospheric delay and the actual value of each satellite in the three systems.

As can be seen from figure 6, tropospheric delay estimation of GPS, GLONASS and BDS also achieve a good estimation effect in this case. The estimated value is relatively close to the actual value, and the error fluctuation is small. Similarly, table 2 shows the maximum, median and average errors of each system in tropospheric delay estimation. Under the long distance network RTK environment, the maximum deviation of tropospheric delay estimated for GPS, GLONASS and BDS are all less than 2cm, and the median error is between 0.01cm and 0.02 cm, with high estimation accuracy. Among them, the mean error of the tropospheric delay for GPS satellite is the smallest, followed by GLONASS system and BDS system, but in general, the mean errors of the three systems are relatively close, all of which are around 1.5cm.

5. Conclusions

The validity and reliability of the spatial error model of regional tropospheric delay is one of the key technologies in network RTK. In this paper, the spatial troposphere was assumed to be an infinitely thin single-layer membrane, and the SVM theory was introduced to realize the real-time construction of regional tropospheric delay model. The comparative analysis of the above cases can lead to the following conclusions:

(1) From single satellite, more than 90% of the satellites have a stable variation in the estimation error of the tropospheric delay, with fluctuations ranging from -3 mm to 3mm. The estimation error of individual satellites fluctuates greatly. The preliminary analysis is due to the low and high variation of satellite height angle during the observation process.

(2) The estimation results of different system satellites are basically the same, that is, the SVM model established in this paper is independent of the type of satellite system.

(3) Under different distances of network RTK, the tropospheric delay accuracy estimated by the SVM tropospheric delay model has little difference, and the accuracy is better than 2cm.

6. Acknowledgments

7. References

[1]   Snay R A and Soler T 2008 *Journal of Surveying Engineering* **134** 95-104
[2]   Huang D F, Li C G, Wu Y Q, Lv G P and Zhou L T 2007 *Chinese Journal of Surveying and Mapping* **36** 256-261
[3]   Tang W M, Liu J N, Liu H and Bai Q B 2007 *Journal of Wuhan University: Information Science Edition* **32** 1156-59
[4]   Wei E, Chai H, An Z and Liu J 2006 *Positioning* **5** 76-81
[5]   Gao X W, Guo J W, Mi J Z, Zhao C M and Zhu H Z 2009 *Journal of Surveying and Mapping* **34** 52-54
[6]   Li J F, Wang Y Q and Guo J Y 2015 *Journal of Surveying and Mapping Science and Technology* **32** 450-454
[7]   Chen Y and Hu W S 2018 *Surveying and Mapping Engineering* **27** 46-52
[8]   Li H 2012 *Statistical Learning Method* (Beijing: Tsinghua University Press)
[9]   Deng N Y and Tian Y J 2004 *A New Method in Data Mining* (Beijing: Science Press)
[10]   Li M L, Hu Y Z, Zhou C, Zhao Z Y and Zhang Y N 2015 *Journal of Xidian University* **42** 147-153
[11]   Schaer S 1999 *Mapping and Predicting the Earth's ionosphere Using the Global Positioning System* (Bern: The University of Bern)
[12]   Zhou P 2014 *Research on ionospheric delay model based on CORS data* (Xu Zhou: China University of Mining and Technology)