# Chapter 25
# Analysis and Application of Extracting GPS Time Series Common Mode Errors Based on PCA

**Gao Han, Zhang Shuangcheng and Zhang Rui**

**Abstract** Characteristics of daily position time series which dated from 2006 to 2011 in South Central of America GPS fiducial network are researched in this paper. A spatial filtering algorithm based on principal component analysis was employed to extract and remove the common mode errors from the coordinate time series. This method promoted the accuracy and reliability of the sites coordinate. Further, we discussed the differences influence in velocity field of GPS stations by test to wipe off the common-mode error. The result shows that the common mode errors can not be neglected in time series analysis, particularly when we deal with the micro deformation, the PCA method can get rid of the common mode errors effectively and improve the reliability of result.

**Keywords** GPS time series · Common mode error · Principal component analysis · Velocity field

## 25.1 Introduction

Nowadays GPS continuous observation has become one significant method of deformation monitoring. There are hundreds of global monitor stations and thousands of regional monitor stations have been built around the world, they can provide us with a unified reference frame to monitor the surface deformation of

G. Han (✉) · Z. Shuangcheng · Z. Rui
College of Geology Engineering and Geomatics, Chang'an University, Xi'an 710054, China
e-mail: gaohan_cd@163.com

Z. Shuangcheng
e-mail: shuangcheng369@vip.163.com

Z. Rui
e-mail: zhangruiwkk@163.com

J. Sun et al. (eds.), *China Satellite Navigation Conference (CSNC) 2013 Proceedings*, Lecture Notes in Electrical Engineering 243,
DOI: 10.1007/978-3-642-37398-5_25, © Springer-Verlag Berlin Heidelberg 2013

global, regional also even smaller scale range. It not only can monitor surface displacement and deformation which caused by the earthquake, but also can monitor tiny tectonic deformation such as post seismic deformation, tectonic movement deformation and so on [1]. However, through the analysis of stations' coordinate time series of GPS regional network we found that there was a kind of space–time correlation error between different stations which called the common mode error (CME) [2–4]. There is no clear conclusion for the source of CME nowadays. It reflects the change on reference framework and scale. Probably it was unsteadily caused by GPS satellite orbit biases, residual error of ocean tide correction and atmospheric pressure tidal (S1 and S2 tidal wave) in essence [5]. CME is the main error source of GPS daily solution, and has negative influence on the extraction of deformation characteristics. Therefore, how to remove the common mode error from the coordinate time series of GPS stations and improve the precision of sites coordinate are very important to the deformation analysis.

In this paper, we used the data of South Central of America GPS regional network which contains about thirteen stations, and utilize the principal component analysis (PCA) to extract and remove the common mode errors from the coordinate time series, improve the accuracy and reliability of the sites coordinate. Further we got the crustal deformation information of experiment region.

## 25.2 The Rationale of Principal Component Analysis

PCA is a multivariate statistics analysis method of selecting a few important variables from multiple variables by the linear transformation. With orthogonal transforms, it turns original random vectors of dependent fraction into new random vectors of independent fraction. The transpositional aim is turning the multiple indicators into a few comprehensive index which used the thought of dimension reduction.

For a GPS regional network which has n stations and observations are carried out for m days, each component (N, E and U) of the residual coordinate time series was expressed $\boldsymbol{X}(t_i, x_j)$ $(i = 1, 2, \cdots m;\ j = 1, 2, \cdots n)$, each column contains a direction data sequences which removed the trend and average, each line represents a component value of all stations in specified epoch, $\boldsymbol{B}$ is covariance matrix of $\boldsymbol{X}$, $b_{i,j}$ was defined by

$$b_{i,j} = \frac{1}{m-1} \sum_{k=1}^{m} \boldsymbol{X}(t_k, x_i) \boldsymbol{X}(t_k, x_j) \tag{25.1}$$

It is a matrix of $n \times n$, can be broken down into

$$\boldsymbol{B} = \boldsymbol{V} \Lambda \boldsymbol{V}^{\boldsymbol{T}} \tag{25.2}$$

where $\boldsymbol{V^T}$ is a $n \times n$ orthogonal matrix which consist of feature vector, $\Lambda$ is a eigenvalue matrix which consist of $k$ nonzero diagonal elements, $\boldsymbol{B}$ is a non-singular matrix, namely $k = n$, the feature vector of $\boldsymbol{B}$ can be written as$(\lambda_1, \boldsymbol{v_1}), (\lambda_2, \boldsymbol{v_2}), \cdots (\lambda_n, \boldsymbol{v_n})$, where $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n$ is a set of orthogonal basis, we describe $\boldsymbol{X}(t_i, x_j)$ with the set of orthogonal basis

$$\boldsymbol{X}(t_i, x_j) = \sum_{k=1}^{n} \boldsymbol{a}_k(t_i)\boldsymbol{v}_k(x_j) \qquad (j = 1, 2 \cdots n) \tag{25.3}$$

where $\boldsymbol{a}_k$ is the temporal amplitude of the $k$th principal component, as shown

$$\boldsymbol{a}_k(t_i) = \sum_{j=1}^{n} \boldsymbol{X}(t_i, x_j)\ \boldsymbol{v}_k(x_j) \qquad (k = 1, 2 \cdots n) \tag{25.4}$$

where $\boldsymbol{v}_k$ is its corresponding eigenvector. Usually the eigenvectors are arranged in the descending order so that the leading few PCs can account for the common mode signature of the entire network, while the higher-order PCs are related to local environmental effects [6]. Can use the cumulative contribution of characteristic value $m_k$ to represent the contribution of each principal component, namely,

$$m_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \qquad (k = 1, 2, \cdots n) \tag{25.5}$$

Set a threshold value such as 0.85, when the cumulative contribution rate to achieve the threshold value, the first P is the primarily model component.

$$\varepsilon_j(t_i) = \sum_{k=1}^{p} \boldsymbol{a}_k(t_i)\ \boldsymbol{v}_k(x_j) \tag{25.6}$$

Since the first p principal components of the detrended time series explain the common mode variations on our network and their eigenvectors have a rather homogeneous spatial pattern, we treat the first p PC as CME.

## 25.3 Analysis of Extracting GPS Time Series CME Based On PCA

In this paper, the stable area's stations of South Central of America were considered in our experiment: primarily we obtained the residual error time series by getting rid of the mean and the trend from the original series; secondly we


**Fig. 25.1** The station of SEU1 network

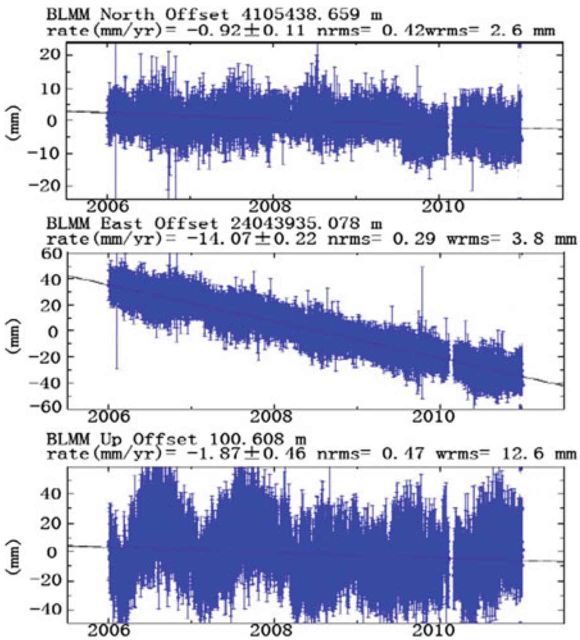



**Fig. 25.2** Raw daily coordinate time series of BLMM

extracted the common-mode error of the experiment area by PCA method; lastly we analyzed the impact of common-mode error on the velocity field of the experiment area.

### 25.3.1 Obtain the GPS Residual Time Series

In the paper we selected thirteen stations of GPS continuous observation which dated from January 1, 2006 to January 1, 2011 in the South Central of America, the distribution of these stations are shown in Fig. 25.1. The daily sites coordinate was estimated from GPS observation by using the GAMIT/GLOBK software packages. In the first step, we obtained the relaxant solution each day with GAMIT software, we used in this solution nearby 4 IGS stations (ALGO, KOUR, GOL2, DRAO) in this solution which position and velocities were well determined in the ITRF2005 to serve as the linking of the local network and global IGS network. In the second step we combined our loosely constrained solution with the SOPAC global solution (igs1 - 1gs6) by using GLOBK softeware. Further, coordinates time series of stations were generated by using GLRED model to identify and remove the surveys or stations which are outliers. Finally, after the clean coordinates time series were got, internal constraints are applied by using a set of globally-distributed

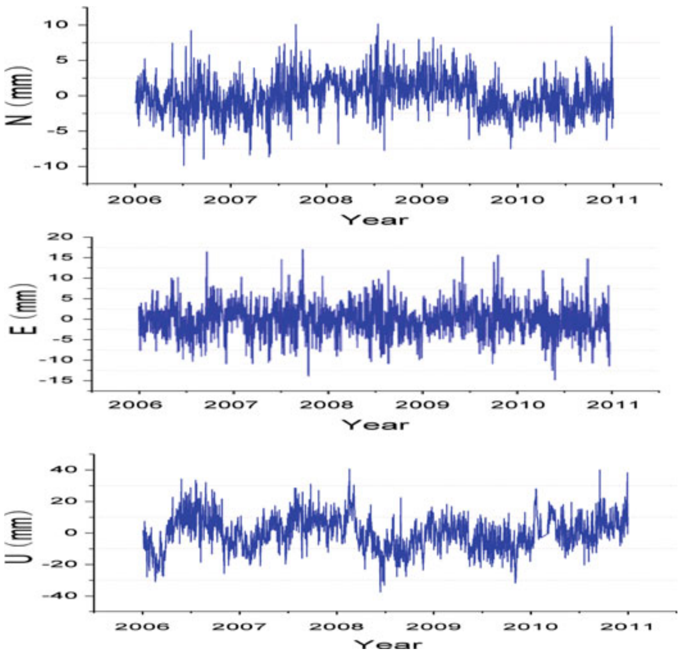



**Fig. 25.3** The residual coordinate time series of BLMM

fiducial station coordinates to define the reference frame for daily station coordinate estimates. Then, we obtained the raw daily coordinate time series in the ITRF2005 (Fig. 25.2, BLMM, for example). Because of the data of some sites is missing seriously, we only process the data from January 1, 2006 to February 25, 2010 by using the PCA method in the subsequent.

When we filter the residual time series with PCA, it requires the time series should be uniformly-space sampling. But it is unlikely to meet such requests in the actual observation. There will always be some days missing. It needs the interpolation management for it to get the uniformly-space sampling time series. Therefore, the paper used cubic spline interpolation. This method is more effective for the continuous data which missing less than 5 days, and the less missing the better [7]. However, when the continuous volume data have bigger loss, the interpolation will appear abnormal after the cubic spline interpolation method was used. In our example, the most continuous lost quantity no more than 5 days in addition to the part of some stations after 2010. Therefore, we use the cubic spline interpolation to filling-in the date which missing below 5 days, and use the Fast Fourier Transform (FFT) interpolated algorithm to filling-in the date which missing more than 5 days. Because of the method of FFT has less damage to the time series and it is better for the situation of whose most continuous lost quantity no more than 50 days.

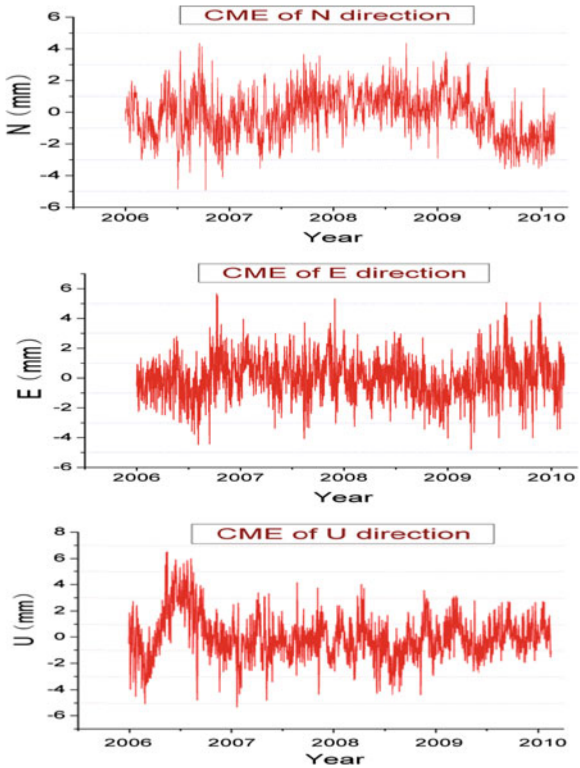
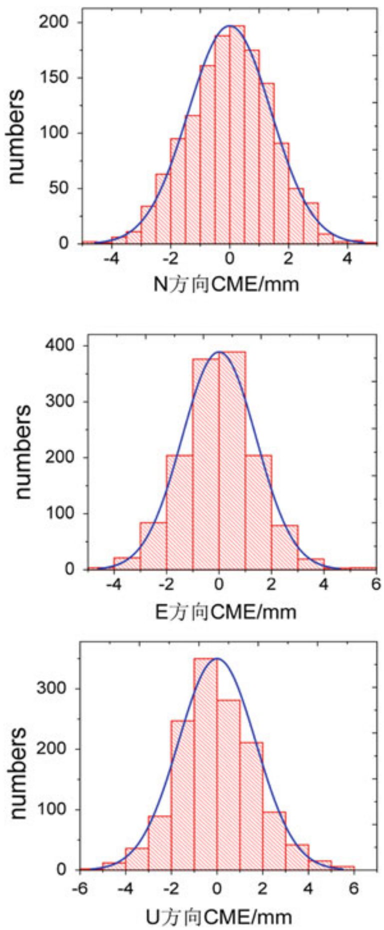



**Fig. 25.4** Time series of common mode errors (N, E, U)

**Fig. 25.5** Statistical graph of CME

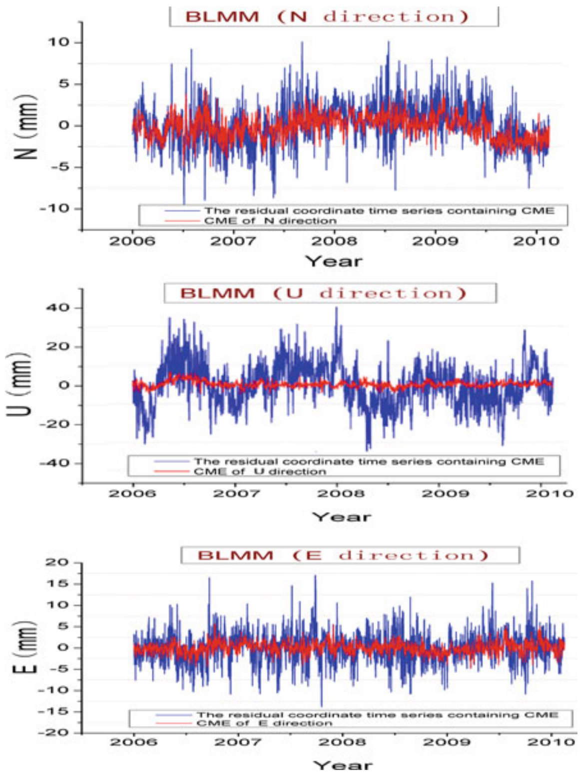



After the interpolating, we used Nikolaidis model (see [8]) to fit the GPS coordinates time series and obtain the GPS residual coordinates time series which removed the mean and the trend. The result will be the input data of PCA in the subsequent. Due to the limit of paper, station of BLMM only be showed at here. Figure 25.3 is the residual coordinate time series of BLMM.

### 25.3.2 Extract of GPS Time Series CME with PCA

Using the above method which called Principal Component Analysis, we processed 13 sites residual time series on SEU1 area-network in the South Central of America. During the PCA analyzing, we extracted and eliminated the components whose accumulative contribution rate reached 85 %. That component represents common mode error. In this experiment, we selected the first three principal components as the common mode error which present the directions of N, E and U. Their accumulative contribution rate was obtained respectively as 86.4, 87.8 and 90.1 % by using the formula (25.5). The common mode errors extracted on the SEU1 area-network which present the directions of N, E and U are shown in the Fig. 25.4. The mean value of N, E and U respectively is 5, 5 and 7 mm. From the following cartogram Fig. 25.5, we can see that common mode error has randomness obviously. Its characteristic is similar to the usual error we see later, CME of all the stations on the regional network were eliminated. The original result of site BLMN was shown in the Fig. 25.6. And by the PCA area filtering, the residual time series of BLMM which has dealed was shown in the Fig. 25.7. From the figure, we found that its horizontal direction is smooth, its height direction has slightly fluctuation. But from the overall view, its amplitude of fluctuation decreased. The method effectively eliminated the common mode errors which improves the precision and stability of the sites coordinate, enhances the robustness of coordinate series.

**Fig. 25.6** The residual coordinate time series of BLMM before removing CME



### 25.3.3 Analysize the Impact of CME on Velocity Field

In order to further discuss the horizontal movement of the SEU1 area-network in the South Central of America, and impacts of the regional common mode error on horizontal velocity. To analyze the impact, in the view of weather or not contain the CME, the velocity field of the area should be achieved separately. Next, comparative analysis was got such as below.

When using GLOBK to obtain the velocity field, we chose the data of 20 days (dated from January 1st to January 20th) every year in 2006–2010. First of all, we combined the everyday relaxation solution of baseline and the global solution which ware calculated from GAMIT. Secondly, we computed the time series by using the GLRED model, and checked the repeatability of baseline and coordinate. Moreover, we eliminated outliers and the CME in the 3.2-part, got the clean time series. Finally, we combined the everyday solution by using the GLOBK modle. And then, the speed of each station in ITRF2005 framework (see Fig. 25.8) was

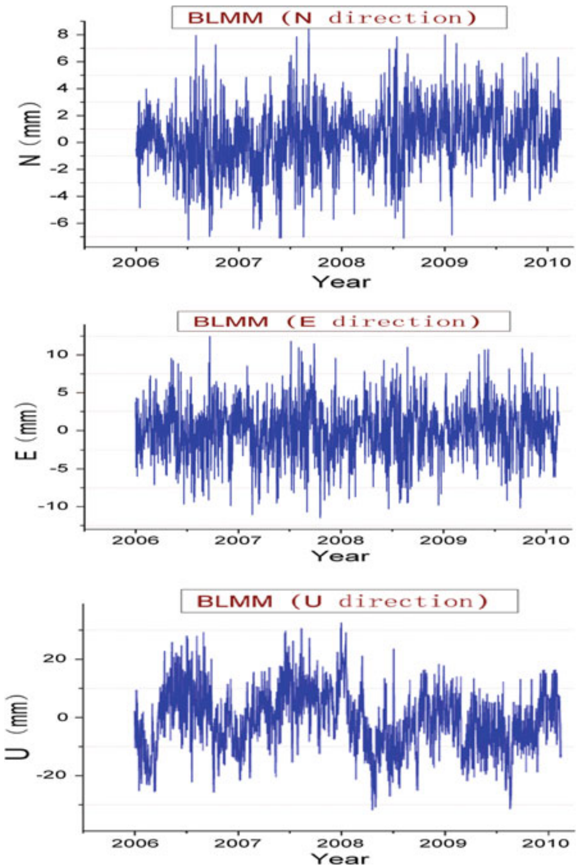



**Fig. 25.7** The residual coordinate time series of BLMM after removing CME

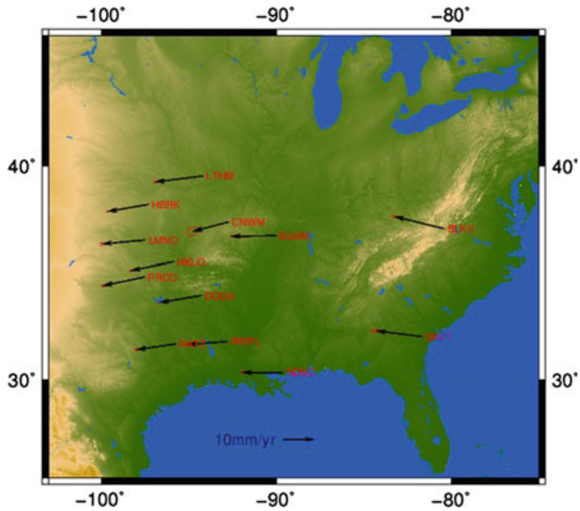



**Fig. 25.8** Station horizontal velocities of North America area (relative ITRF2005) (Error *ellipse* is 95 % confidence level)

estimated. Further, the velocity filed was obtained which relating to the North America plate based on the plate motion model of ITRF2005VEL [9–11]. The horizontal velocity in the east–west direction is between −1.3 mm/a and 3.5 mm/a, and its mean velocity is 1.4 mm/a; the horizontal velocity in the north–south direction is between −3.7 and 2.5 mm/a, and its mean velocity is −2.2 mm/a. Using the same method but the only difference is not eliminating the CME, we can get the velocity filed which relating to the North America plate that including the CME. Comparing it with the velocity filed which eliminating the CME, we can find that the horizontal velocity in the E and direction N are about 1 mm/a. The biggest is 1.2 mm/a. As the Fig. 25.9 shows, blue arrows represent the velocity which

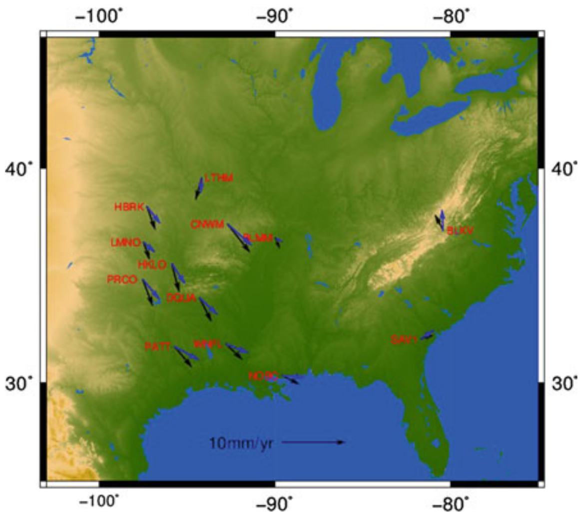



**Fig. 25.9** Station horizontal velocities of SEU1 area (relative to the North America plate) (*blue arrows* represent before removing CME, *black arrows* represent after removing CME)

not removing CME, black arrows represent the velocity which removing CME. From the above results, we can see that the CME has influence on the horizontal velocity in the Central and South America and it can not be ignored when we extract the micro deformation of crustal.

## 25.4 Conclusions

In this paper, we effectively extracted the common mode error of coordinate time series on SEU1 area-network in the South Central of America, it improved the precision and stability of the sites coordinate and enhances the robustness of the coordinate series. Then after we eliminate the CME, the difference of velocity field was further discussed relative to North American Plate. As the result shows that the common-mode error has important influence on the station velocity in some extent, and so it cannot be ignored especially when we extract the micro deformation of crustal.

## References

1. Wu J, SunY, Liu C (2008) Extraction of common mode errors for continuous GPS networks and deformation analysis. J Geodesy Geodyn 28(4):97–101
2. Dong DN, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. J Geophys Res 111:B03405
3. Yunfeng T, Zhengkang S (2009) Progress on reduction of non-tectonic noise in GPS position time series. Acta Seismol Sin 31(1):68–81
4. Zhang J et al (1997) Southern California permanent GPS geodetic array. Error analysis of daily position estimates and site velocities. J Geophys Res 102:18035–18055
5. Yunfeng T, Zhengkang S (2011) Correlation weighted stacking filtering of common-mode component in GPS observation network. Acta Seismol Sin 33(2):198–208
6. Savage J (1995) Principal component analysis of interseismic deformation in southern California. Geophy Res 100:12691–12701
7. Yanqiang W, Liren H (2004) A new interpolation method in time series analyzing. J Geodsy Geodyn 24(4):43–47
8. Nikolaidis R (2002) Observation of geodetic and seismic deformation with the global positioning system. University of California, San Digeo
9. Yian L, Liren H et al (2004) Present-day Crustal movement in China. Seismological Press, Beijing
10. Liu GY (2004) Some key issues relating to high precision GPS positioning and crustal deformation analysis. Institute of Geodesy and Geophysics, CAS, Wuhan
11. Rong M, Zhou W et al (2009) Global plate motion model based on ITRF2005. Hydrogr Surveying charting 29(1):1–3