A GPS Time Series Prediction Model Based on CEEMD

Jia Lu, Xing Chen, and Shuo Feng

***Abstract*—A GPS time series prediction model is presented, based on Complete Ensemble Empirical Mode Decomposition (CEEMD), which has the advantage of improving the prediction accuracy greatly. CEEMD is a new and improved version of Empirical Mode Decomposition, which decomposes a non-linear and non-stationary time series into a finite and often small number of Intrinsic Mode Functions (IMFs) and a residual. For each IMF and the residual, appropriate models are recommended to model them respectively. Due to the reversibility of the decomposition, the final predicted result of the GPS time series is available by summing up predicted results of all IMFs and the residual. Experiment results show that the proposed model behaves much better than the classical time series prediction model.**

***Index Terms*—GPS, time series, complete ensemble empirical mode decomposition, intrinsic mode functions.**

I. Introduction

Recent years have witnessed increasing interest in Global Positioning System (GPS) [1]–[6]. In most cases, GPS can be used for monitoring geophysical phenomena, such as coseismic and tectonic plates movements [7]. Usually, the challenge in above applications is improving the positioning accuracy.

Positioning errors in the GPS, however, are probably inevitable due to ionosphere errors, multipath errors, antenna phase deviation, etc. [8]. Several strategies have already been proposed to eliminate the positioning errors such as Wavelet models [9], FIR filters [10], adaptive filters [11], the Bayesian methods [12], GPS/Pseudolites (PLs) positioning technology [13], the Kalman filter [14], the carrier phase-difference model [15], the fractional Brownian motion (fBm) [16], autoregressive moving average (ARMA) model [17], etc.

ARMA, a time series prediction model, is widely employed in GPS time series analysis [17]–[20], which can develop proper techniques for automatic recognition and calibration of the positioning errors in GPS time series [17], [21]. However, ARMA which is directly adopted to analyse the GPS time series can not always achieve a good prediction accuracy. This is mainly because ARMA aims to analyse stationary time series while the GPS time series is virtually a non-linear [22] and non-stationary [23] process with different components, for example, white noise, pink noise and random work. Therefore, several improved ARMA models were presented to eliminate prediction errors in GPS time series analysis [18], [24].

In fact, the inherent non-linear and non-stationary characteristic in the GPS time series can be described effectively by a fully data adaptive decomposition method, such as Empirical Mode Decomposition (EMD) [1], [25]–[28], Ensemble Empirical Mode Decomposition (EEMD) [27] and Complete Ensemble Empirical Mode Decomposition (CEEMD) [28]. These methods are powerful time series analysis tools showing great promise for non-linear and non-stationary time series analysis.

In this paper, a GPS time series prediction model is proposed, which is based on Complete Ensemble Empirical Mode Decomposition (CEEMD). First, the GPS time series is decomposed into Intrinsic Mode Functions (IMFs) and residual components using CEEMD. Then we calculate autocorrelation function (ACF), partial autocorrelation function (PACF), and the Hurst index for each IMF and the residual. Based on model identification rules, appropriate models, e.g., AR, ARMA or Gaussian processes, are recommended to model each IMF and the residual separately. Finally, based on the reversibility of the decomposition, the final predicted result of the GPS time series can be calculated by summing up predicted results of all IMFs and the residual. Experiment results show that the proposed model behaves much better than the classical time series prediction model.

Empirical Mode Decomposition (EMD) has drawn more and more attention recently. Several improved empirical mode decomposition methods were proposed. Section II provides a detail description. Section III explains the core algorithm of the proposed model step by step based on a practical GPS time series. A convincing comparison between the proposed model and the classical time series prediction model is presented in Section IV. Finally, conclusions with ideas for future work are discussed in Section V.

II. Summary of EMD

Empirical Mode Decomposition (EMD) is a developed technique for the decomposition of non-linear and non-stationary time series. The most distinguished characteristic is that this decomposition is data adaptive and fully reversible. Subsequently, detailed explanations of different empirical mode decomposition methods are provided.

A. Empirical Mode Decomposition

Huang *et al.* [25], [26] first proposed the adaptive time-frequency data analysis technique, named Empirical Mode Decomposition (EMD). It decomposes a time series into several Intrinsic Mode Functions (IMFs) or modes and residual components. Importantly, this decomposition is reversible. That is to say, the original time series can be

Manuscript received June 9, 2015; revised January 24, 2016.

Jia Lu is with China Defense Science and Technology Information Center, Beijing, China, he is also with Xichang Satellite Launch Center, Sichuan, China (e-mail: lujia661@126.com).

Xing Chen and Shuo Feng are with Xichang Satellite Launch Center, Sichuan, China (e-mail: chenxok@gmail.com, fsing1987@163.com).

reconstructed directly by summing up all IMFs and the residual with a negligible error.

Given a time series, $X(n)$, the decomposition of $X(n)$ is operated via the study of consecutive local minima and maxima. $X(n)$ can be expressed as

$$X(n) = \sum_{j=1}^{k} imf_j(n) + r_k, \tag{1}$$

where $imf_j(n)$ denotes the j^{th} IMF components constrained to be zero-mean, r_k stands for a residual trend after k decomposition. Each IMF is defined to have the number of extrema (maxima and minima) and number of zero-crossing equal or differing by one. And the envelope of each IMF is strictly symmetric based on the local maxima and minima respectively. The effective algorithm of EMD can be summarized as follows:

1) Find all extrema (minma and maxima) of $X(n)$.
2) Locate the maximum (minimum) envelope max(n) (min(n)) of $X(n)$ by passing a natural cubic spline through the local maxima (minima).
3) Compute mean values with: $m(n)$=(max(n) +min(n))/2.
4) Remove details with: $d(n)=X(n)-m(n)$.
5) Iterate steps 1) to 4) on $d(n)$ until it is zero-mean according to the stopping criterion; the obtained $d(n)$ is referred to as an IMF, $imf_j(n)$.
6) Compute the residual with: $X(n)=X(n)-imf_j(n)$.
7) Iterate steps 1) to 6) until no more IMFs are available.

EMD has been successfully used in a broad range of research areas by means of extracting useful signals from collected time series generated by noisy non-linear and non-stationary processes. As useful as EMD proved to be, it still suffers from "mode mixing". The mode mixing can be defined as the appearance of oscillations of completely different amplitude in an intrinsic mode function, or the appearance of similar oscillations in different intrinsic mode functions. The undesired consequence is that the individual IMF is far away from physical meaning.

B. Ensemble Empirical Mode Decomposition

To alleviate mode mixing from presence, a new noise-based analysis technique is proposed, Ensemble Empirical Mode Decomposition (EEMD) [27]. In EEMD, the true intrinsic mode function was defined as the amalgamation of the time series and a white noise with a finite amplitude. By means of adding finite noise, EEMD mitigate mode mixing dramatically.

EEMD is developed as follows:

1) Find Add a white noise $w(n)$ to the time series: $\dot{X}(n)=X(n)+w(n)$.
2) Decompose $\dot{X}(n)$ into IMFs.
3) Iterate steps 1) and 2), and use different white noise each time.
4) Calculate means of all IMFs as the final result.

Notably, the amplitude λ of the added white noises should decrease with the following statistical rule

$$\lambda_n = \frac{\lambda}{\sqrt{N}}, \tag{2}$$

where N is the number of ensemble members, λ_n is the difference between the time series and IMFs.

In fact, by increasing the ensemble members and keeping the amplitude of the added white noises, the added white noises is always cancelled each other out (or reduced to a negligibly small level). Several examples have demonstrated that EEMD does significantly alleviate the chance of mode mixing.

C. Complete Ensemble Empirical Mode Decomposition

EEMD solves the mode mixing problem in EMD, however it introduces new ones [28]. In order to conquer these situations, a variation of EEMD, Complete Ensemble Empirical Mode Decomposition (CEEMD) was proposed. CEEMD is a new and improved version with more robust design for "mode mixing". In CEEMD, a particular white noise is added at each step of the decomposition. The CEEMD is described as follows:

1) Add a white noise $w^j(n)(j=1,2,\ldots\mathrm{T})$ to the time series: $X^j(n) = X(n) + w^j(n)$.
2) Decompose $X^j(n)(j=1,2,\ldots\mathrm{T})$ and compute: $\overline{imf_1}(n) = \frac{1}{T}\sum_{j=1}^{T} imf_1^j(n)$.
3) Calculate the residue: $r_1(n) = X(n) - \overline{imf_1}(n)$.
4) Decompose $r_1(n) + \lambda_1 E_1(w^j(n))(j=1,2,\ldots\mathrm{T})$ and compute: $\overline{imf_2}(n) = \frac{1}{T}\sum_{j=1}^{T} E_1(r_1(n) + \lambda_1 E_1(w^j(n)))$. The operator $E_j(\bullet)$ means the j^{th} mode obtained by EMD.
5) For s = 2, 3,…, S, compute the s^{th} residual: $r_s(n) = r_{s-1}(n) - \overline{imf_s}(n)$.
6) Decompose $r_s(n) + \lambda_s E_s(w^j(n))(j=1,2,\ldots\mathrm{T})$ and compute: $\overline{imf_{s+1}}(n) = \frac{1}{T}\sum_{j=1}^{T} E_1(r_s(n) + \lambda_s E_s(w^j(n)))$.
7) Iterate steps 5) and 6) for next s.

Observe that the $\lambda_s(s=1,2,\ldots\mathrm{T})$ can be used to select proper SNR at each step. CEEMD solves the mode mixing problem thoroughly and does better in reducing the sifting iterations. Theoretical analysis and experiments demonstrate that CEEMD provides an exact reconstruction of time series and a better spectral separation of the IMFs than the other two empirical mode decomposition methods, with a numerically negligible error. Therefore, CEEMD is adopted in this paper.

III. The Proposed Model

The GPS time series is in fact a non-linear [22] and non-stationary [23] process. Therefore, the traditional time series prediction models which directly employ the GPS time series may reduce the prediction accuracy. In this paper, we first decompose the GPS time series using CEEMD into a finite set of IMFs and a residual. For each IMF and the residual, then we implement the model verification based on autocorrelation function (ACF), partial autocorrelation function (PACF), and the Hurst index. Next, appropriate models, e.g., AR, ARMA or Gaussian processes, are recommended to model all IMFs and the residual respectively. Based on the reversibility of the decomposition, finally, different models are employed to predict each IMF and the residual separately and by summing up predicted results of all IMFs and the residual, we get the final predicted result of the GPS time series.

To conclude, in this paper a GPS time series prediction model is proposed which is based on CEEMD. The main flowchart of the proposed model is provided in Fig. 1. Subsequently, a practical GPS time series is used as an example to step by step show the core algorithm of the proposed model.

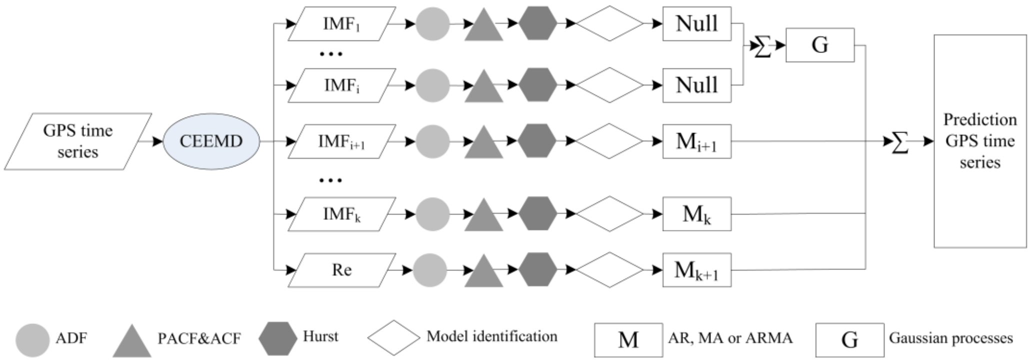


Fig. 1. Flowchart of the propose model.

Step 1: Record the GPS time series using a certain type of GPS receivers. We collect lots of GPS time series from Thursday, March 5, to Sunday, March 8, in 2015 with 100 msec. sampling. Then the GPS time series is preprocessed to get absolute errors for x-coordinate (X), y-coordinate (Y) and z-coordinate (Z) respectively. Next, X is employed to explain the algorithm process. Step 2 to Step 6 can be applied to Y and Z as well.

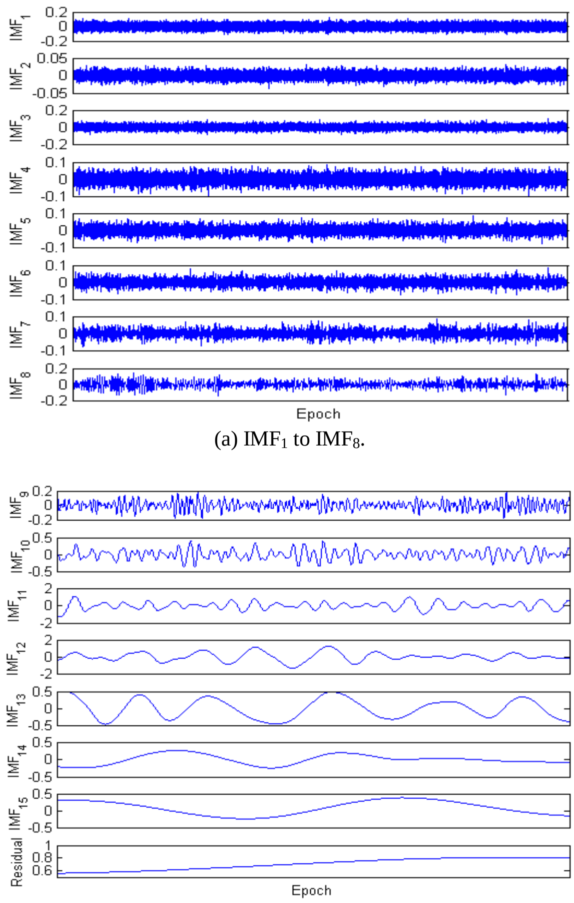
Step 2: Decompose X by CEEMD into a finite set of IMFs and a residual. In this paper, X is decomposed into fifth IMFs and a residual. Fig. 2 gives all IMFs and the residual.

Step 3: Test for the unit root using Augmented Dickey Fuller (ADF) test. For each IMF and the residual, the unit root test is implemented to verify the stationarity. Results show that all IMFs are stationary time series without drift while the residual is stationary time series with drift.

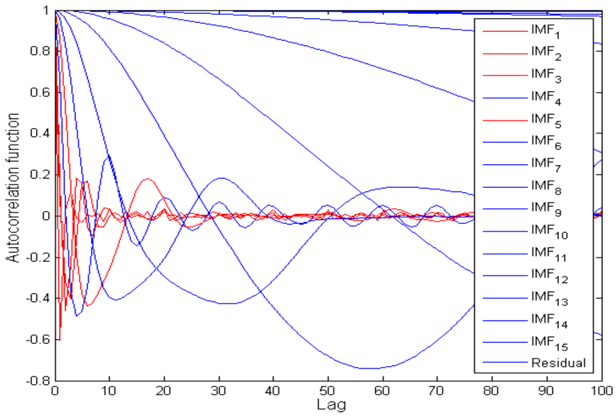
Step 4: Calculate values of autocorrelation function (ACF), partial autocorrelation function (PACF) and the Hurst index. For each IMF and the residual, ACF and PACF are implemented to identify the autoregressive (AR) and moving average (MA) parts of ARMA model. Based on model identification rules, if ACF is trailing and PACF is truncated, the time series can be concluded for the AR model. If ACF is truncated and PACF is trailing, the time series can be described as the MA model. And if both ACF and PACF are trailing, ARMA model is proper for the time series. The calculation results are presented in Fig. 3. Fig. 3 shows that ACF of IMF_4 and IMF_6-IMF_{15} is trailing and PACF of IMF_4 and IMF_6-IMF_{15} is truncated. Therefore, according to model identification rules, the AR model is recommended to model IMF_4 and IMF_6-IMF_{15}. The residual, with a trailing ACF and a trailing PACF, can be modeled by an ARMA model. Since ACF and PACF of IMF_1-IMF_3 and IMF_5 are both truncated, AR, MA, or ARMA are not fit to them. So, the Hurst index is considered. The Hurst index can be used to distinguish the gaussian process from the long-range dependence process. Calculation results of IMF_1-IMF_3 and IMF_5 are presented in Table I. Table I shows that IMF_1-IMF_3 and IMF_5 behave as gaussian processes. Furthermore, Montillet *et al*. [4] showed that summation of the IMFs whose Hurst index is less than 0.5 can be estimated by a gaussian process. Therefore, IMF_1-IMF_3 and IMF_5 are modeled by a gaussian process.

TABLE I: THE HURST INDEXES OF IMF_1-IMF_3 AND IMF_5

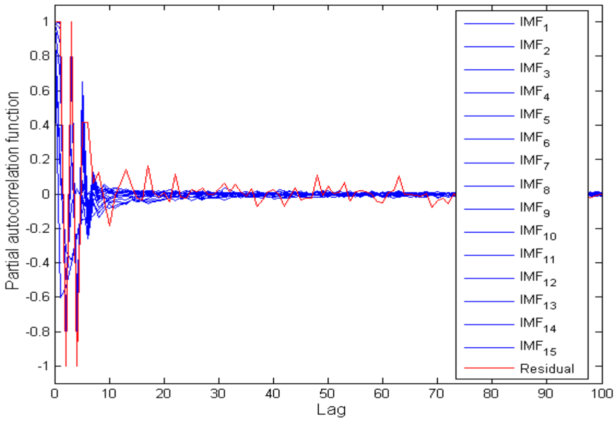
IMFs	Hurst	Mean	Std
IMF_1	0.11	-3.70e-05	0.03
IMF_2	0.37	-5.06e-05	0.01
IMF_3	0.14	-4.97e-05	0.02
IMF_5	0.41	-3.55e-05	0.02
Total	0.30	-7.33e-05	-4.45e-02



(b) IMF_9 to IMF_{15} and the residual.
Fig. 2. CEEMD decomposition of x-coordinate time series.



(a) ACF of IMFs and the residual.



(b) PACF of IMFs and the residual.

Fig. 3. ACF and PACF of IMFs and the residual.

Step 5.1: Parameterize AR models for IMF_4 and IMF_6-IMF_{15} and parameterize an ARMA model for the residual. Denote AR_4, AR_6-AR_{15} and $ARMA_{16}$ as their corresponding models.

Step 5.2: Parameterize a gaussian process for IMF_1-IMF_3 and IMF_5. Denote $G(mu, sigma)$ as the corresponding model, where mu is the mean value and the standard deviation is sigma.

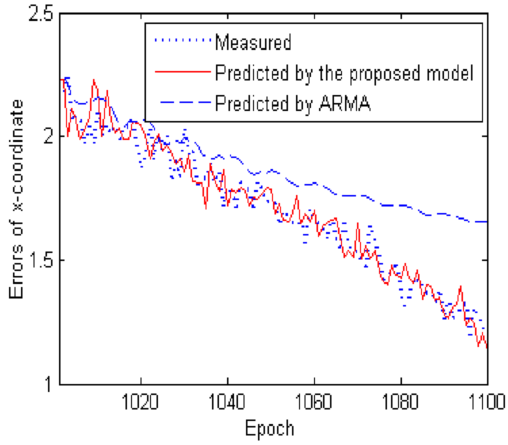
Step 6: Predict X by summing up the predicted result of all IMFs and the residual. This is on account of the fact that CEEMD is fully reversible. Specifically, we have

$$\tilde{X} = G(-7.33e-5, 4.45e-2) + AR_4 + \sum_{i=6}^{15} AR_i + ARMA_{16})) \cdot$$

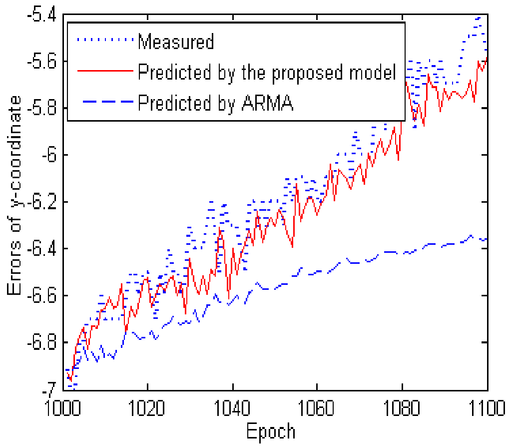
IV. Experimental Results

In this section, ARMA and the proposed model are employed to predict the GPS time series. For each coordinate (x-coordinate, y-coordinate and z-coordinate) of the GPS time series, the first 1000[1] epoches of the coordinate time series are employed to construct the proposed model step by step (Section III). Similarly, ARMA is also parameterized based on the first 1000 epoches. Then we predict the next 100[2] epoches.

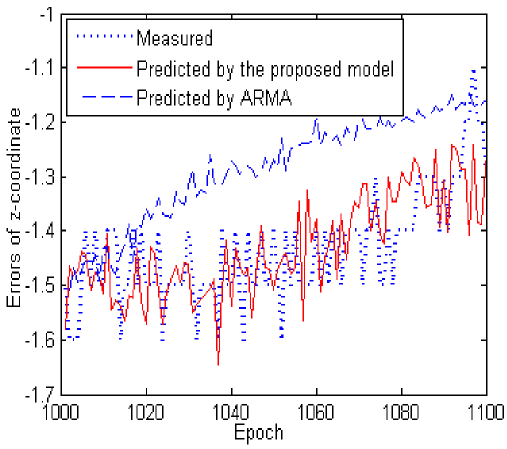
Fig. 4(a) presents the prediction results of the x-coordinate time series. Obviously, the proposed model behaves much better than the classical time series prediction model, ARMA. Similarly, the prediction results for the y-coordinate time series and z-coordinate time series are provided in Fig. 4(b) and Fig. 4(c) separately. It is obvious that the proposed model improves the prediction accuracy greatly.



(a) x-coordinate.



(b) y-coordinate.



(c) z-coordinate

Fig. 4. Comparison between the measured x(y, z)-coordinate time series and predicted results using the proposed model and ARMA model.

V. Conclusions

This paper introduces a GPS time series prediction model based on Complete Ensemble Empirical Mode Decomposition (CEEMD). It has the advantage of improving the prediction accuracy greatly. CEEMD is a robust extension

[1]The value 1000 is modifiable by users on the basis that the recommended AR or ARMA models can be parameterized.

[2]The value 100 is modifiable by users.

of Empirical Mode Decomposition (EMD) method with more robust design for "mode mixing". By decomposing a real GPS time series into Intrinsic Mode Functions (IMFs) and residual components using CEEMD, we find it is more convenient to model the individual IMF and the residual than to directly model the whole GPS time series. Naturally, a much better prediction accuracy is achieved.

Our future work will focus on generalizing the use of the proposed model. For example, the proposed model can be used to study financial time series, fuzzy time series, etc.

References

[1] J. Li, C. Liu, Z. Zeng, and L. Chen, "GPR signal denoising and target extraction with the CEEMD method," *IEEE Geoscience and Remote Sensing Letters*, vol. 99, no. 6, pp. 1–5, 2015.

[2] M. Manataki, A. Vafidis, and A. Sarris, "Evaluation of EMD-based methods on GPR data," in *Proc. Near Surface Geoscience 2014-20th European Meeting of Environmental and Engineering Geophysics*, 2014.

[3] J. Li, L. Chen, S. Xia, P. Xu, and F. Liu, "A complete ensemble empirical mode decomposition for GPR signal time-frequency analysis," *SPIE Defense+ Security. International Society for Optics and Photonics*, 2014, pp. 90770C–90770C.

[4] J.-P. Montillet, P. Tregoning, S. McClusky, and K. Yu, "Extracting white noise statistics in GPS coordinate time series," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 563–567, 2013.

[5] W. Torge and J. Muller, *Walter de Gruyter*, 2012.

[6] Y. M. Zhang, H. S. Shi, Y. Liu, and C. Liu, "Multipath reduction of shaft tower monitoring with GPS in mine," *Advanced Materials Research*, vol. 457, pp. 265–269, 2012.

[7] S. Miyazaki and Y. Hatanaka, "Crustal deformation observed by GSIS new GPS array," *Eos Trans. AGU*, vol. 78, no. 17, pp. 99–134, 1997.

[8] H. Bian, S. Zhang, and Q.-Z. Zhang, "Empirical mode decomposition combined single epoch algorithm for GPS baseline solution," in *Proc. IEEE International Congress on Image and Signal Processing*, 2010, pp. 4056–4058.

[9] L. Ge, S. Han, and C. Rizos, "Multipath mitigation of continuous GPS measurements using an adaptive filter," *GPS Solutions*, vol. 4, no. 2, pp. 19–30, 2000.

[10] C. Satirapod and C. Rizos, "Multipath mitigation by wavelet analysis for gps base station applications," *Survey Review*, vol. 38, no. 295, pp. 2–10, 2005.

[11] J. Wang, J. Gao, and D. Cao, "Dynamic deformation signal extracting model based on a dyadic wavelet transform," *Journal-China University of Mining and Technology (Chinese Edition)*, vol. 36, no. 1, p. 116, 2007.

[12] A. Borghi, L. Cannizzaro, and A. Vitti, "Advanced techniques for discontinuity detection in GNSS coordinate time-series. An Italian case study," *Geodesy for Planet Earth*, 2012, pp. 627–634.

[13] W. Jianpeng, G. Jingxiang, L. Chao, and W. Jian, "High precision slope deformation monitoring model based on the GPS/Pseudolites technology in open-pit mine," *Mining Science and Technology (China)*, vol. 20, no. 1, pp. 126–132, 2010.

[14] D. Yuan, X. Cui, D. Fan, W. Feng, and Y. Yu, "Application of kalman filter method to the date processing of GPS deformation monitoring," in *Proc. 2010 Second International Workshop on Education Technology and Computer Science (ETCS)*, 2010, vol. 2, pp. 269–272.

[15] S. Han, "Carrier phase-based long-range GPS kinematic positioning," University of New South Wales, 1997.

[16] J. Montillet and K. Yu, "Modified leaky LMS algorithms applied to satellite positioning," in *Proc. 2014 IEEE 80th Vehicular Technology Conference (VTC Fall)*, 2014, pp. 1–5.

[17] J. Li, K. Miyashita, T. Kato, and S. Miyazaki, "GPS time series modeling by autoregressive moving average method: Application to the crustal deformation in central Japan," *EPS*, vol. 52, no. 3, pp. 155–162, 2000.

[18] X. Luo, M. Mayer, and B. Heck, "Verification of ARMA identification for modelling temporal correlations of GNSS observations using the ARMASA toolbox," *Studia Geophysica et Geodaetica*, vol. 55, no. 3, pp. 537–556, 2011.

[19] A. Lyubushin and P. Yakovlev, "Properties of GPS noise at Japan islands before and after tohoku mega-earthquake," *Springer Plus*, vol. 3, no. 1, pp. 1–7, 2014.

[20] Z. Qianqian and G. Qingming, "Bayesian methods for outliers detection in GNSS time series," *Journal of Geodesy*, vol. 87, no. 7, pp. 609–627, 2013.

[21] M. R. Islam, M. Rashed-Al-Mahfuz, S. Ahmad, and M. K. I. Molla, "Multiband prediction model for financial time series with multivariate empirical mode decomposition," *Discrete Dynamics in Nature and Society*, 2012.

[22] W. Jian, J.-X. Gao, J.-L. Wang, and C.-H. Xu, "EMD-based GPS baseline solution and validation test," *Journal of China University of Mining and Technology*, vol. 18, no. 2, pp. 283–287, 2008.

[23] S.-S. Jan and C.-C. Sun, "A study on the application of the empirical mode decomposition to GPS signal acquisition process," *Journal of Aeronautics, Astronautics and Aviation, Series A*, vol. 41, no. 2, pp. 83–89, 2009.

[24] J.-P. Montillet and K. Yu, "Modeling geodetic processes with levyn alpha-stable distribution and farima," *Mathematical Geosciences*, pp. 1–20, 2014.

[25] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[26] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: The hilbert spectrum 1," *Annual Review of Fluid Mechanics*, vol. 31, no. 1, pp. 417–457, 1999.

[27] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, 2009, pp. 1–41.

[28] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4144–4147.

Jia Lu received the Ph.D. degree in computer science and technology from National University of Defense Technology, Changsha, China, in 2014.

His research work has been published in *International Conference on Computer Communication and Networks* (ICCCN 2014). His research interests include time series analysis, error estimating coding, wireless channels modeling.

Xing Chen received the master degree in computer application technology from Hefei University of Technology, Hefei, China, in 2012.

His research work has been published in *Journal of Computer Applications*. His research interests include digital image processing, content based image retrieval (CBIR), satellite navigation system.

Shuo Feng received the master degree in communication and information system from Nankai University, China, in 2014

His research interests include error estimating coding, wireless sensor network, Signal and information processing.