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
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

$$X(n) = \sum_{j=1}^{d} \text{imf}_j(n) + r_e,$$  
(1)

where $\text{imf}_j(n)$ denotes the $j^{th}$ IMF components constrained to be zero-mean, $r_e$ 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 (minima 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, $\text{imf}_j(n)$.
6) Compute the residual with: $X(n)$ = $X(n)$ - $\text{imf}_j(n)$.
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 a white noise $w(n)$ to the time series: $\tilde{X}(n) = X(n) + w(n)$.
2) Decompose $\tilde{X}(n)$ 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 $\lambda$ of the added white noises should decrease with the following statistical rule

$$\lambda = \frac{\lambda}{\sqrt{N}},$$  
(2)

where $N$ is the number of ensemble members, $\lambda$ 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 $w^j(n)$ ($j = 1, 2, \ldots, T$) to the time series:

$$X^j(n) = X(n) + w^j(n).$$

2) Decompose $X^j(n)$ ($j = 1, 2, \ldots, T$) and compute:

$$\overline{\text{imf}}_j(n) = \frac{1}{T} \sum_{j=1}^{T} \text{imf}_j(n).$$

3) Calculate the residue: $r_j(n) = X(n) - \overline{\text{imf}}_j(n)$.

4) Decompose $r_j(n) + \lambda \cdot E_j(w^j(n))$ ($j = 1, 2, \ldots, T$) and compute:

$$\overline{\text{imf}}_j(n) = \frac{1}{T} \sum_{j=1}^{T} E_j(r_j(n) + \lambda \cdot E_j(w^j(n))).$$

5) For $s = 2, 3, \ldots, S$, compute the $s^{th}$ residual:

$$r_{s}(n) = r_{s-1}(n) - \overline{\text{imf}}_j(n).$$

6) Decompose $r_{s}(n) + \lambda \cdot E_s(w^s(n))$ ($j = 1, 2, \ldots, T$) and compute:

$$\overline{\text{imf}}_j(n) = \frac{1}{T} \sum_{j=1}^{T} E_s(r_s(n) + \lambda \cdot E_s(w^s(n))).$$

7) Iterate steps 5) and 6) for next $s$.

Observe that the $\lambda$ ($s = 1, 2, \ldots, T$) 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.

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>
<thead>
<tr>
<th>IMFs</th>
<th>Hurst</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMF_1</td>
<td>0.11</td>
<td>-3.70e-05</td>
<td>0.03</td>
</tr>
<tr>
<td>IMF_2</td>
<td>0.37</td>
<td>-5.06e-05</td>
<td>0.01</td>
</tr>
<tr>
<td>IMF_3</td>
<td>0.14</td>
<td>-4.97e-05</td>
<td>0.02</td>
</tr>
<tr>
<td>IMF_5</td>
<td>0.41</td>
<td>-3.55e-05</td>
<td>0.02</td>
</tr>
<tr>
<td>Total</td>
<td>0.30</td>
<td>-7.33e-05</td>
<td>-4.45e-02</td>
</tr>
</tbody>
</table>

Fig. 2. CEEMD decomposition of x-coordinate time series.
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

\[
\tilde{X} = G(-7.33e - 5.4.45e - 2) + AR_4 + \sum_{i=6}^{15} AR_i + ARMA_{16})
\]

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 epochs. 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.

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.
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