Exploring time variants for short-term passenger flow

Mu-Chen Chen *, Yu Wei

Institute of Traffic and Transportation, National Chiao Tung University, 4F, No. 118, Section 1, Chung-Hsiao W. Road, Taipei 100, Taiwan, ROC

ABSTRACT

Passenger flow is a fundamental element in a transportation system. It is important to explore the time variants of short-term passenger flow for transportation planning and operation. When the data are sufficiently analyzed, transportation planners not only can make better decisions, but also enhance the performance of transportation systems. The data of short-term passenger flow may be difficult to analyze due to its exotic oscillation. Hilbert–Huang transform (HHT) has recently been developed for analyzing non-linear and non-stationary data. In this paper, the proposed time variants exploration method includes two stages: empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA). A real passenger flow dataset is collected from Taipei rapid transit corporation (TRTC) to investigate the viability of the proposed time variants exploration approach. The intrinsic mode functions (IMFs) extracted by EMD can represent the local characteristics of passenger flow and imply its meaningful time variants such as peak period pattern, semi-service period pattern, semi-daily pattern and daily pattern. By comparing the results of HHT with that of fast Fourier transform (FFT), it indicates that HHT can obtain the narrower frequency band, accurately capture time–frequency–energy distribution, and help to enhance the performance of transportation systems. The results show that HHT is an effective approach for exploring the time variants of short-term passenger flow in a metro system.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Passenger flow is one of the essential elements in transportation systems. It is also an essential input for transportation management, including transportation planning, transportation infrastructure construction, facility improvement, operation planning, revenue management, and so on. When the data are sufficiently analyzed, transportation planners are able to make better decisions with more useful information.

During the past two decades, a number of researchers have developed various travel demand forecasting models to predict passenger flow and trends, such as conventional travel demand modeling, and multiple regression (e.g., Alfa, 1986; Wirasinghe and Kumarage, 1998; Kulshreshtha and Nag, 2000; Golias, 2002; Jovicic and Hansen, 2003; Bar-Gera and Boyce, 2003; Varagouli et al., 2005; Wardman, 2006; Tsekeris and Stathopoulos, 2006; Zhou and Kockelman, 2009). The conventional travel demand forecasting model is a sequential demand modeling method considering trip generation, trip distribution, mode choice, and assignment modules. Such an approach has been widely applied in the travel forecasting phase of the urban transportation planning process (Alfa, 1986; Wirasinghe and Kumarage, 1998; Jovicic and Hansen,

2003). Most previous studies have focused on building models that consist of a large set of explanatory variables to predict the long-term demand and modal split. However, the travel demand model needs to aggregate the abovementioned modules, in which each module may include various methods. In addition, Bar-Gera and Boyce (2003) pointed out that the sequential demand modeling method suffers from the inconsistent consideration of travel times and congestion effects of route choice in the modeling procedure. The inconsistent consideration of congestion effects was argued as the shortcoming of sequential demand modeling methods (Bar-Gera and Boyce, 2003). Furthermore, some studies have revealed that the conventional forecasting model of travel demand lacks volume variation when it faces the variants of economic activity (Kulshreshtha and Nag, 2000; Varagouli et al., 2005).

In recent years, the multiple regression method and multivariate dynamic econometric time series have been applied to construct transportation demand models of passenger flow (Kulshreshtha and Nag, 2000; Kuby et al., 2004; Varagouli et al., 2005; Wardman, 2006). These methods considered some external factors such as economic variables to predict passenger flow. In the literature, demand forecasting models have commonly been constructed under the assumption of linear and stationary passenger flow time series, which may not be realistic. In addition, a train traffic model incorporated a dynamic equation based on the evolution of train headways and train passenger loads to estimate the variants of passenger flow on railways (Assis and Milani, 2004).

* Corresponding author. Tel.: +886 2 2349 4967; fax: +886 2 2349 4953.
E-mail addresses: ittchen@mail.nctu.edu.tw, iemcchen@yahoo.com.tw (M.-C. Chen).

Nomenclature

π	a mathematical constant which is the ratio of the circumference of a circle to its diameter	$r_n(t)$	the final residue
P	the Cauchy principal value of the singular integral	$X(t)$	the original time series
$a(t)$	the amplitude function of time t	$Y(t)$	the Hilbert transform of IMF
$c_j(t)$	the jth IMF component	$Z(t)$	a complex conjugate pair combination of $X(t)$ and $Y(t)$
$h(\omega)$	the marginal spectrum	$\theta(t)$	the phase function of time t
$H(\omega, t)$	the Hilbert spectrum	$\omega(t)$	the instantaneous frequency of Hilbert transform

---

Recently, Chen et al. (2009) have studied the diurnal pattern of subway ridership in New York with the socio-demographics of population.

In the last decade, several methods have been developed for the short-term analysis of passenger traffic flow in transportation systems and networks. These methods include principal component analysis (Abdel-Aty and Pemmanaboina, 2006; Paris and Van den Broucke, 2008), wavelet analysis (Huang, 2003; Jiang and Hon, 2005; Xie and Zhang, 2006), neural networks (Vlahogianni et al., 2004; Lee et al., 2006), and support vector regression (Wu et al., 2004; Vanajaksi and Rilett, 2007), etc. Other methods circumventing the assumptions of linearity and stationarity include the vector autoregressive method (Chandra and Al-Deek, 2009) and error components analysis (Frejinger and Bierlaire, 2007), which can address the problem of spurious regression.

Although several forecasting models have been developed to predict passenger flow to date, analysis of the time variants of passenger flow without the assumptions of linearity and stationarity is still needed. Without this assumption, the obtained results will be more realistic for transportation system management and planning. Furthermore, the data collection cost can be extensively reduced, provided that transportation planners only collect the passenger flow for analysis.

A methodology for analyzing non-linear and non-stationary data named Hilbert–Huang transform (HHT) has recently been introduced by Huang et al. (1998). HHT primarily consists of two stages: empirical mode decomposition (EMD) and Hilbert spectra analysis (HSA) (Huang et al., 1998, 2003a). EMD is an empirical, intuitive, direct and adaptive data processing method developed especially for dealing with non-linear and non-stationary data. Basically, EMD applies a sifting process to decompose data into a small number of independent and nearly periodic intrinsic modes based on the local characteristic time scale. Therefore, according to the scale, the physical implications of each mode can be identified. In the second stage, the instantaneous frequency, determined by Hilbert transform, can provide a much sharper identification of imbedded structures in data (Huang et al., 1998, 2003a, 2004).

HHT has been successfully applied in several fields such as ocean waves (Hwang et al., 2003), biomedical engineering (Balocchi et al., 2004; Liang et al., 2005; Jiang and Yan, 2008; Su et al., 2008), signal processing (Tao et al., 2005; Xie and Wang, 2006; Li and Meng, 2006; Rai and Mohanty, 2007; Blanco-Velasco et al., 2008; Guo et al., 2008), wind engineering (Li and Wu, 2007), and earthquake engineering (Dong et al., 2008). Several previous studies have applied EMD to extract intrinsic mode functions (IMFs) (Balocchi et al., 2004; Liang et al., 2005; Li and Meng, 2006; Wu, 2007; Blanco-Velasco et al., 2008; Guhathakurta et al., 2008; Guo et al., 2008; Li et al., 2008; Zhang et al., 2008). However, most applications are primarily limited to the studies of nature science and engineering (Zhang et al., 2008). Relatively few applications in social science can be found. For example, references (Huang et al., 2003a; Wu, 2007; Guhathakurta et al., 2008; Zhang et al., 2008) have utilized EMD to analyze financial time series. Additionally, Hamad et al. (2009) has applied EMD to analyze traffic volume data. Hamad et al. (2009) applied a combined approach of EMD and back-propagation neural networks to predict traffic volume by using a set of real-life loop detector data. From the experimental results, it was revealed that the hybrid approach, which takes advantage of EMD, outperforms the traditional traffic volume prediction model involving a simulation model and time series method. However, Hamad et al. (2009) have only focused on building a forecasting model of short-term traffic volume. The time variants of passenger flow, which plays an essential role in transportation systems, still remains insufficiently investigated. HSA can be incorporated with EMD to further explore the time variants of passenger flow.

Analyzing the time variants of short-term passenger flow is an important issue in building a forecasting model of passenger demand. With the time variants of short-term passenger flow, transportation planners can make effective operation plans such as station passenger crowd regulation planning, transportation resource planning and human resource planning and so onto improve transportation performance. Notice that "short-term" in this paper refers to a forecasting horizon of 15 min, which is common in the analysis of transportation operations (e.g., Williams et al., 1998; Chen and Grant-Muller, 2001; Xie and Zhang, 2006; Zhang and Ye, 2008). This paper aims at analyzing the time variants of short-term passenger flow in a metro system by using HHT. The remainder of this paper is organized as follows. Section 2 gives an introduction to the concept and algorithm of HHT. Section 3 introduces the data of metro passenger flow, and the results obtained by HHT. Section 4 compares the results of HHT with that of fast Fourier transform, and discusses the implications of the results for a metro system. Section 5 describes the applications of results obtained by HHT. Finally, Section 6 concludes this paper.

2. Methodology

In the first stage, the short-term passenger flow is decomposed into a small number of independent and nearly periodic intrinsic modes based on the local characteristic time scale by itself via EMD. Second, the time–frequency–energy distribution is obtained by HSA.

2.1. Empirical mode decomposition

Empirical mode decomposition (EMD) developed by Huang et al. (1998) is a signal analysis method which can deal with non-linear and non-stationary data. The main idea of EMD is to decompose the original time series data into a finite and small number of oscillatory modes based on the local characteristic time scale by itself. Each oscillatory mode, which is similar to a harmonic function, is expressed by an intrinsic mode function (IMF). IMFs have to satisfy the following two conditions (Huang et al., 1998):

1. in the whole dataset of a signal, the number of extrema and the number of zero crossings must either equal or differ at most by one, and
2. at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

As mentioned above, the first condition resembles the traditional requirement of a narrow band for a stationary Gaussian process. The second condition is to modify the global requirement to a local one. Ideally, the local mean of data should be zero. However, it is impossible to define the local mean for non-stationary data by using a local scale. In order to avoid defining the local averaging time scale, Huang et al. (1998) proposed forcing the local symmetry with the local mean of the envelopes defined by the local maxima and the local minima. By using EMD, the well-defined instantaneous frequency can be obtained.

The essence of EMD is the sifting process which extracts IMFs from the original data. The algorithm of EMD is described as follows:

Step 1: identify all the local extrema including minimum values and maximum values in time series data $x(t)$.

Step 2: generate the upper and lower envelopes, $e_{max}(t)$ and $e_{min}(t)$, by a cubic spline line.

Step 3: calculate the mean value $m_1(t)$ from the upper and lower envelopes, then generate the mean envelop as:

$$m_1(t) = (e_{min}(t) + e_{max}(t))/2 \qquad (1)$$

Step 4: calculate the difference between the time series data $x(t)$ and mean value $m_1(t)$. The first difference $h_1(t)$ is designated as proto-intrinsic mode function,

$$h_1(t) = x(t) - m_1(t) \qquad (2)$$

Step 5: check whether the proto-intrinsic mode function $h_1(t)$ satisfies the properties of IMF. Ideally, $h_1(t)$ should be an IMF. However, it may generate a new extremum, and shift or exaggerate the existing extrema in the sifting process.

If properties of $h_1(t)$ satisfy all the requirements of an IMF, $h_1(t)$ is denoted as the ith IMF $c_i(t)$, and substitutes residue $r_1(t)$ for the original time series data $x(t)$.

$$r_1(t) = x(t) - h_1(t) \qquad (3)$$

Otherwise, $h_1(t)$ is not an IMF. Then, it substitutes $h_1(t)$ for the original time series $x(t)$.

Step 6: repeat Steps 1–5. The sifting process stops while the residue satisfies one of the termination criteria. First, the residue or the ith component is smaller than the predetermined threshold or becomes a monotonic function such that no more IMF can be extracted. Second, the number of zero crossings and extrema is the same as that of the successive sifting step (Huang et al., 2003b).

By using the above algorithm, the original time series data $x(t)$ can be decomposed into n modes and a residue as follows:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t) \qquad (4)$$

where n is the number of IMFs, $c_j(t)$ represents IMFs which are nearly orthogonal to each other and periodic, and $r_n(t)$ is the final residue, which is a constant or a trend. By the sifting process, each IMF is independent and specific for expressing the local characteristics of the original time series data. Generally, the first component has the highest frequency, which represents the shortest period variants in the time series data, whereas the residue represents the lowest frequency. Consequently, the set of IMFs is derived from high frequency to low frequency. In addition, EMD can also be taken as a filter of high pass, band pass or low pass.

2.2. Hilbert spectral analysis

The second stage of HHT is Hilbert spectral analysis (HSA), which is performed to obtain the time–frequency–energy distribution. The following is a brief introduction to HAS from Huang et al. (1998). After the decomposition procedure in EMD, the Hilbert transform of IMF $Y(t)$ can be obtained as:

$$Y(t) = \frac{1}{\pi} P \int \frac{X(t')}{t - t'} dt', \qquad (5)$$

where π represents a mathematical constant (i.e., pi) which is the ratio of the circumference of a circle to its diameter, and P indicates the Cauchy principal value of singular integral. In Eq. (5), the Hilbert transform is defined as the convolution of $X(t)$ with $1/t$. It emphasizes the local properties of the time series data $X(t)$. With this definition, an analytic signal $Z(t)$ is a complex conjugate pair combination of $X(t)$ and $Y(t)$, and it takes the form

$$Z(t) = X(t) + iY(t) = a(t)e^{i\theta(t)} \qquad (6)$$

where

$$a(t) = \sqrt{X^2(t) + Y^2(t)} \qquad (7)$$

and

$$\theta(t) = \arctan\left(\frac{X(t)}{Y(t)}\right) \qquad (8)$$

$a(t)$ indicates the amplitude, which is a function of time, and $\theta(t)$ represents the phase, which is also a function of time. In Eq. (6), the polar coordinate expression shows the local nature of representation. It is the best local fit of amplitude and phase-varying trigonometric function to the original time series data $X(t)$. The instantaneous frequency of Hilbert transform is defined as:

$$\omega(t) = \frac{d\theta(t)}{dt} \qquad (9)$$

After performing Hilbert transform on each IMF, the original time series data $X(t)$ can be expressed as the real part (RP) in the following form:

$$X(t) = RP \sum_{j=1}^{n} a_j(t) \exp(i\theta_j(t)) = RP \sum_{j=1}^{n} a_j(t) \exp\left(i \int \omega_j(t) dt\right). \qquad (10)$$

The residue $r_n(t)$ is dropped, because it may be a monotonic function or a smaller value than the predetermined threshold. Eq. (10) represents both the amplitude and the frequency of each component as a function of time. With both the amplitude and frequency being a function of time, the frequency–time distribution of the amplitude is called a Hilbert spectrum $H(\omega, t)$. The square of amplitude $a_j(t)$ can be taken as energy, so the Hilbert spectrum can express the energy distribution in the original time series data on each frequency. According to the definition of Hilbert spectrum, the marginal spectrum $h(\omega)$ can be defined as:

$$h(\omega) = \int_0^T H(\omega, t) dt \qquad (11)$$

where T is the total data length. The marginal spectrum offers a measure for the total amplitude contribution from each frequency. It presents the cumulative amplitude over the entire data span.

The advantages of HHT method can be summarized as follows. First, EMD can decompose the non-stationary and non-linear data into a small number of IMFs by the sifting process. Each component represents the local characteristic time scale of original data by itself. All IMFs are nearly independent of each other. Therefore, each IMF is local, adaptive and orthogonal. Second, each IMF is one

portion of the original data. The original data can also be completely reconstructed by summing up the components and residue. Third, since the instantaneous frequency is obtained by using Hilbert transform, it is possible to observe the changing frequency of each component. Hence, we can analyze the data in the time–frequency–energy space.

3. Data and results

3.1. Data

A dataset of rapid transit passenger flow is used to investigate the viability of HHT for exploring the time variants for short-term passenger flow. The dataset of passenger flow was collected from the Muzha line of Taipei rapid transit corporation (TRTC) during the period from May 1 to May 31, 2008. The selected data, which contained 2976 records as shown in Fig. 1, were collected by automatic fare collection system every 15 min. The service hours of TRTC are from 6:00 AM to 1:00 AM the next day.

3.2. Intrinsic mode decomposition

By using EMD, the time series data of passenger flow are decomposed into nine IMF components and residue as shown in Fig. 2. The time variants characteristics of passenger flow are clearly demonstrated in Fig. 2. All the extracted IMF components are graphically illustrated in the order in which they are extracted, to indicate the change of frequency (or periodic) from the highest frequency to the lowest frequency. To start with, the high frequency (or short period) components are obtained in the first few components (e.g., IMF C1), and the low frequency (or long period) components are given in the last few components (e.g., IMFs C6–C9). The first few components represent the highly time variants or noise in the original passenger flow data, while the last few components represent the long periodic components. The last component is the residue of sifting, which generally represents the trend of the time series. However, this paper focuses on analyzing the time variants of short-term passenger flow, so the trend is not essential and it is not extracted from the one-month data.

In order to demonstrate the intrinsic meaning of IMFs, the study reconstructs these components. Each of these IMFs is superimposed on the original time series data. With step-by-step adding of IMFs, the mean square error (MSE) between the aggregated data of IMFs and the original time series data is 1.04265E-13. Since the data are only kept to a few decimal places, this MSE may be the round-off error in computation. Therefore, the extracted IMFs and residue can appropriately represent the original data.
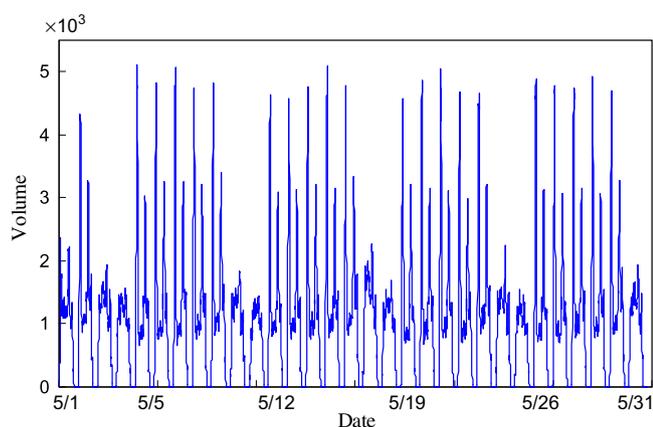
3.3. Statistical analysis

In order to explore the physical phenomenon of the extracted IMF components, Pearson product moment correlation and Kendal rank correlation are applied in this paper to measure the correlation between each IMF component and the original time series data. As summarized in Table 1, IMFs C1–C5 and C7–C8 positively correlated to the original time series in terms of Pearson correlation coefficient. Pearson correlation coefficients of IMFs C3–C5 are respectively 0.545, 0.535 and 0.406, which indicate a stronger positive correlation. Observing from the Kendall correlation coefficients, IMFs C3–C5 have higher ranks, the coefficients of which are 0.376, 0.374 and 0.345, respectively. The results of Pearson correlation coefficient and Kendall correlation coefficient are consistent. With the percentage power of each IMF, IMFs C2–C4 have higher power, and IMF C3 has the highest percentage power (34.4%). From Table 1, the percentage powers of IMFs C1, C6–C9 and residue are not significant. To summarize, IMFs C2–C5 are demonstrated to be meaningful components for original time series data. The physical phenomena of IMFs C2–C5 are going to be discussed in next section.

3.4. The results of Hilbert spectrum

After extracting IMFs by EMD, HHT is then applied to separate the original time series into nearly orthogonal components. In this paper, Visual Signal 1.2 (www.ancad.com.tw) is used to obtain EMD components and HHT. As mentioned above, each component has the well-behaved Hilbert transform property. The instantaneous frequency can present more information about the time variants of passenger flow in the frequency–time domain. The instantaneous frequencies of the first five IMFs are graphically illustrated in Fig. 3. The mean instantaneous frequency of IMF C2 is 0.2445 cycles per hour (i.e. 4.1 h per cycle). In other words, the mean instantaneous frequency of IMF C2 represents the periodic pattern of peak period. The mean instantaneous frequencies of IMFs C3–C5 are 0.1464, 0.0725 and 0.0389 cycles per hour, respectively. After transformation, the cycles of IMFs C3–C5 are about 6.8, 13.8, and near 24 h per cycle. Due to the service hours of the TRTC metro system are from 6:00 AM to 1:00 AM the next day, IMF C3 could be regarded as the semi-service pattern of passenger flow. Similarly, the mean instantaneous frequencies of IMF C4 could be regarded as the semi-daily pattern of passenger flow because of one half daily hours. The mean instantaneous frequencies of IMF C5 could be defined as the daily pattern of passenger flow because of daily hours. To summarize, the instantaneous frequencies of IMFs C2–C5 imply the time variants of passenger flow, and they are peak period pattern, semi-service period pattern, semi-daily pattern and daily pattern.

As discussed above, the Hilbert spectrum represents the energy distribution of the collected time series data in both frequency and time scale. It provides more information about amplitude variants of measured time series data. The time–frequency–energy representation of the original data can thus be expressed by the distribution of the amplitude in both frequency and time scale. By performing Hilbert transform, Hilbert spectrum of all IMFs and the selected meaningful IMFs C2–C5 can be obtained. The Hilbert marginal spectrum of all IMFs and the selected meaningful IMFs C2–C5 are illustrated in Fig. 4a and b. For the Hilbert spectrum of original data, which displays the energy concentration as a whole, the strongest energy is located at around the afternoon peak periods (17:30 PM–19:00 PM) on Monday, May 5 during the whole observation period. Notice that the graphs of Hilbert spectrum are not shown in this paper for simplification, and more detailed analysis is available from the authors. From the theoretical fundamental of HHT mentioned above, the square of amplitude can be

Fig. 1. The time series of passenger flow of the Muzha line.
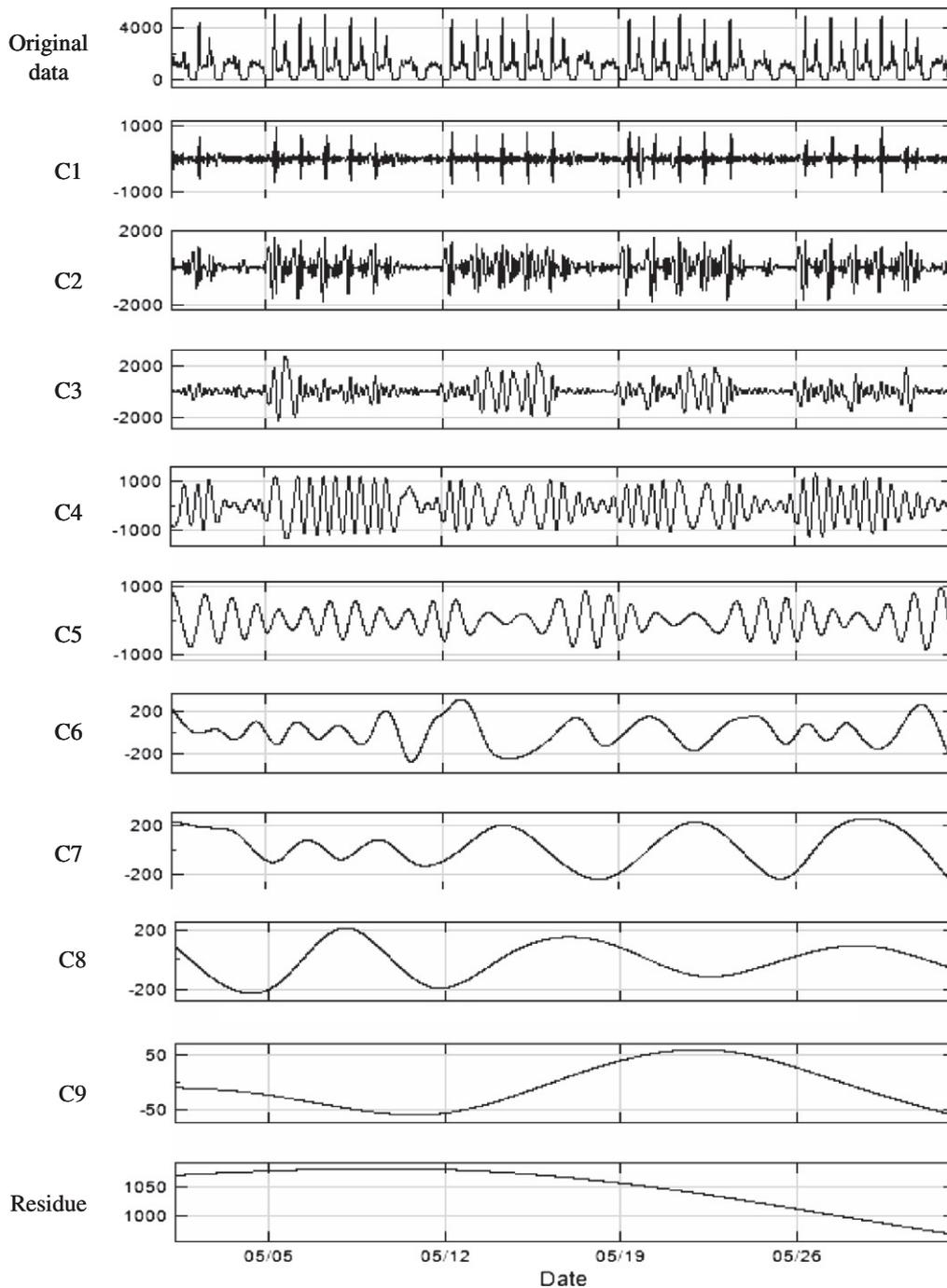
Fig. 2. The original data of passenger flow and its IMFs.

taken as energy. The larger amplitude reveals larger energy. Hence, the energy information reflecting the real fluctuation of passenger flow indicates that the largest fluctuation exists at afternoon peak periods. With the HHT spectrum, metro planners can effectively observe the fluctuations of short-term passenger flow in the metro system, and make decisions based on the observation. After performing Hilbert transform with the selected meaningful IMFs, the Hilbert spectrum clearly shows the energy distribution that concentrates on weekdays (Monday–Friday) and weekends (Saturday–Sunday). The weekdays have a higher frequency variants and a stronger energy. Comparing the results of Hilbert spectra, it demonstrates that the fluctuation patterns of weekdays and weekends which include the selected meaningful IMFs C2–C5 are

more obvious than that of original passenger flow data. Obviously, since the noise and the low frequency are eliminated by EMD, the passenger flow can be more effectively analyzed.

Hilbert marginal spectra of individual IMFs C2–C5 are illustrated in Fig. 5a–d. Clearly, it can be observed from these figures that the energy of each IMF concentrates on narrower frequency bands. According to the Hilbert spectrum of IMF C2, the energy concentrates on Monday–Friday and indicates a higher frequency variant. The highest energy locates around the frequency of 0.1020 cycles per hour for IMF C2, as shown in Fig. 5a. That is to say, the strongest energy distributes at the higher frequency. In addition, the frequency of 0.2510 cycles per hour presents relatively smaller energy. The results may imply that IMF C2 preserves

Table 1
The correlation coefficients and percentage power of components.

IMF	Mean	Standard deviation	Pearson coefficient	Kendall coefficient	Percentage power (%)
C1	1.025	163.49500	.070[b]	−.001	1.76
C2	3.237	465.83342	.165[b]	−.030[a]	20.50
C3	6.411	622.81081	.545[b]	.376[b]	34.40
C4	13.532	327.33391	.535[b]	.374[b]	28.30
C5	25.641	213.94020	.406[b]	.345[b]	10.90
C6	57.143	83.39142	.027	.027[a]	1.33
C7	124.070	71.11571	.083[b]	.032[a]	1.70
C8	185.874	115.57342	.058[b]	.024	1.02
C9	495.050	13.78034	.007	.011	0.12
Residue	1488.096	34.39925	.013	.004	–

[a] Correlation is significant at 0.05 levels (2-tailed).
[b] Correlation is significant at 0.01 levels (2-tailed).

both the frequencies of 0.1020 and 0.2510 cycles per hour. Similarly, the energy concentrates on weekdays and indicates a lower frequency variant for IMFs C3 and C4 as Hilbert spectrum. According to the Hilbert marginal spectrum of IMFs C3–C4, the highest energy in turn locates around the frequencies of 0.0942 and 0.0784 cycles per hour as shown in Fig. 5b and c.

4. Comparisons and discussions

4.1. Comparisons

The fast Fourier transform (FFT) method, which is an efficient method to compute the discrete Fourier transform (DFT) and its inverse, has been widely applied in analyzing random data. The primary property of DFT is that it can convert discrete-time data into its discrete-frequency components, being a harmonic function. FFT can be used to depict the global frequency distribution and to analyze stationary and linear data.

In order to evaluate the performance of HHT in exploring the time variants of passenger flow, three types of comparisons are made as follows: comparing HHT to FFT with the original time series data, comparing HHT to FFT with individual IMFs C2–C5, and comparing HHT to FFT with the superimposed IMFs C2–C5. Visual Signal 1.2 is also used to compute FFT.

4.1.1. Comparison with original time series

For DFT, power spectral density (PSD) can show the strength of the energy with a function of frequency. The probability density function of power spectral density can be estimated by using the parametric and non-parametric estimation techniques. In this paper, the Kernel smoothing density function (Elgammal et al., 2003; Gurwicz and Lerner, 2005), which is a general non-parametric estimation technique, is used to define the probability density function of FFT. A Gaussian function is used as the Kernel function herein. In addition, PSD is equivalent to the square of the mean amplitude of DFT. The total variance of the original time series is recovered upon integrated PSD over the frequency range. A Hanning window is used in the Fourier spectrum.

With the passenger flow, HHT and FFT spectra can be obtained by performing HHT and FFT, respectively. From Fig. 6a, the frequency band varies to a wide extent from the span of original data by using FFT. The results of FFT only demonstrate the global frequency distribution including main frequencies of 0.0417, 0.0833, and 0.2083 cycles per hour. On the other hand, the HHT spectrum indicates the local time scale of each IMF and display the energy concentration. That is to say, HHT spectrum preserves both the frequency and the energy resolution in the time domain. For the frequency resolution of the entire passenger flow data, the frequency band of HHT distributes 0.0392–0.0784 cycles per hour as shown in Fig. 4a. Obviously, the frequency distribution of HHT demonstrates a narrower range than that of FFT for the entire passenger flow data.

4.1.2. Comparison with individual IMFs C2–C5

As the definition of the first condition of IMF defined in Section 2, each IMF should be a stationary and narrow band signal. According to the first condition of IMF, each IMF should be a stationary
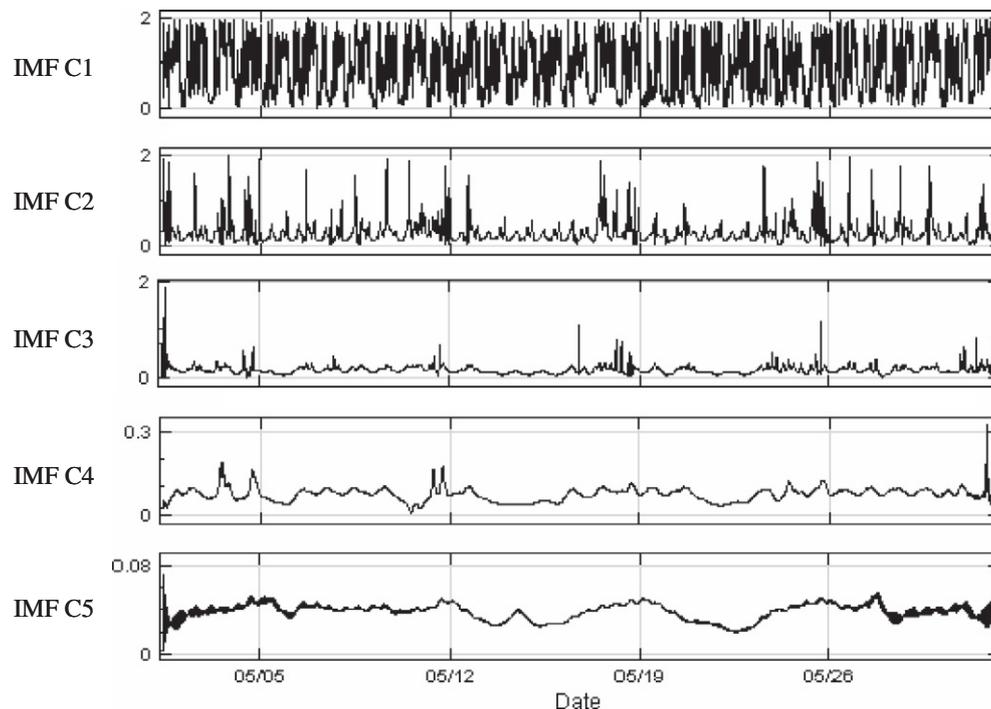
Fig. 3. The instantaneous frequencies for the first five IMFs.

a Hilbert marginal spectrum of original data

b Hilbert marginal spectrum of superimposed IMFs C2-C5
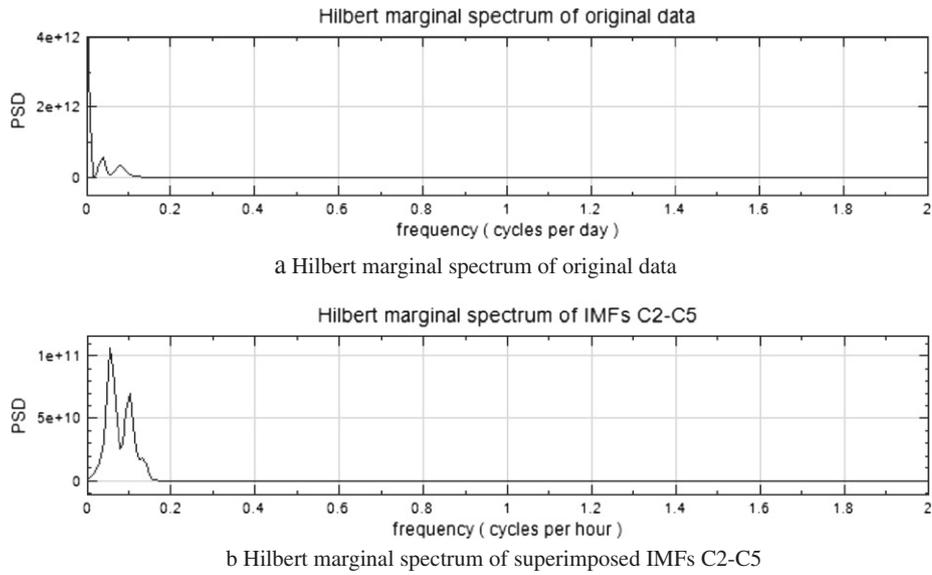
Fig. 4. Hilbert marginal spectra for passenger flow and IMFs C2–C5.

a Hilbert marginal spectrum of IMF C2

b Hilbert marginal spectrum of IMF C3

c Hilbert marginal spectrum of IMF C4

d Hilbert marginal spectrum of IMF C5

Fig. 5. Hilbert marginal spectra of individual IMFs C2–C5.

a Fourier spectrum of original data

b Fourier spectrum IMF C2

c Fourier spectrum of IMF C3

d Fourier spectrum of IMF C4

e Fourier spectrum of IMF C5
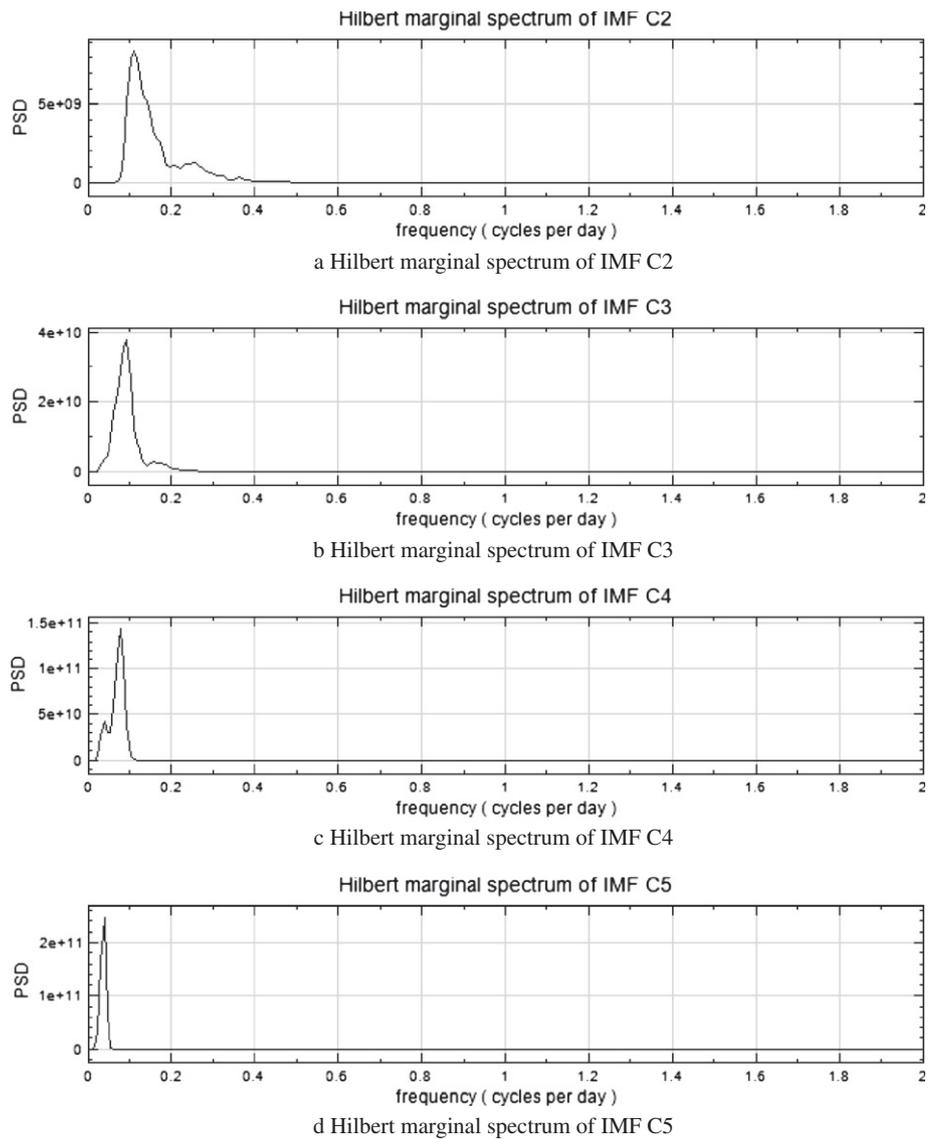
f Fourier spectrum of IMFs C2-C5
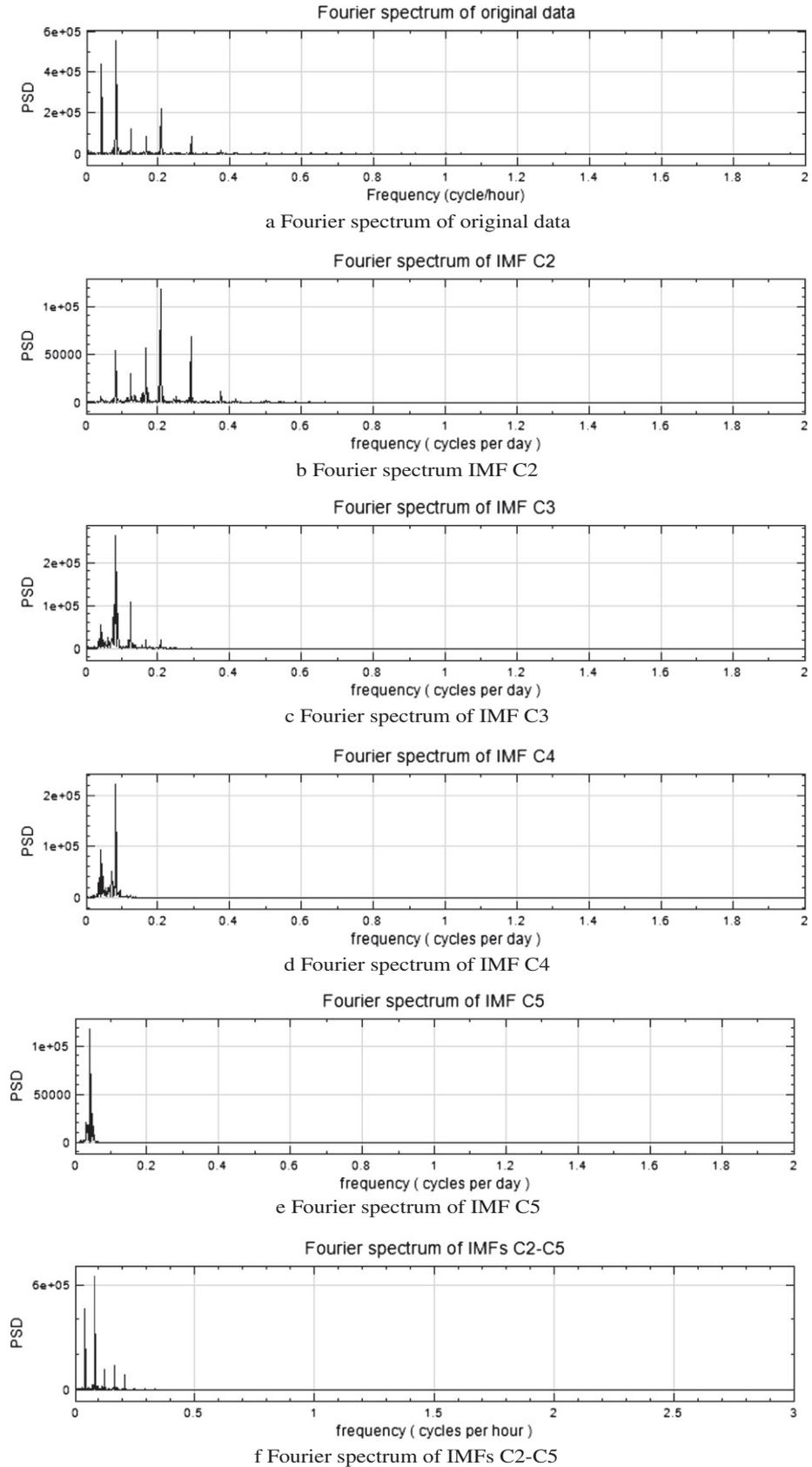
Fig. 6. Fourier spectra of original data and IMFs C2–C5.

and narrow band signal; therefore FFT is also applied to analyze the meaningful IMFs C2–C5 for comparison purpose.

Fig. 6b–e illustrates the spectra of IMF C2–C5 by using FFT. The FFT spectrum of IMF C2 can capture the high frequency, yet the frequency band varies widely. Although the main frequency of IMF C2 is located at 0.2083 cycles per hour (4.8 h per cycle), the other frequencies of 0.2917 and 0.0833 cycles per hour are also significant, shown in Fig. 6b. The main frequency of IMF C3 is obviously detected at 0.0833 cycles per hour (12 h per cycle). Similarly, FFT spectra of IMFs C4 and C5 can be identified clearly as well. The main frequencies of IMF C4 are clearly identified to be 0.0833 cycles per hour (12 h per cycle) and 0.0417 cycles per hour (24 h per cycle). From Fig. 6e, the main frequency of IMF C5 clearly is 0.0417 cycles per hour (24 h per cycle). These results show that FFT method is only capable of capturing the main global frequencies (cycles) of the selected component IMFs C2–C5.

Hilbert marginal spectra of individual IMFs C2–C5 are illustrated in Fig. 5a–d. The Hilbert spectrum and the Hilbert marginal spectrum of individual IMFs C2–C5 can represent the information of time–frequency–energy distribution as discussed in Section 3. Observing both HHT and FFT spectra, there are some distinctions for IMFs C2–C5. Firstly, the results of FFT represent the global frequency for individual component, but those of HHT present the local frequency which can illustrate the time variants of passenger flow. The time variants can reflect the fluctuation of passenger flow. Secondly, HHT generates a narrower frequency band than FFT for individual IMFs C2–C4. Taking IMF C2 as an example, there is obviously a stronger energy distribution at about 0.1020 cycles per hour with HHT (refer to Fig. 5a). However, the frequency of IMF C2 with FFT distributes dispersedly at 0.0833, 0.2083, and 0.2917 cycles per hour (refer to Fig. 6b). Obviously, HHT generates a narrower frequency band than that of FFT. In case of IMF C5, both HHT and FFT generate a similar frequency distribution at around 0.0417 cycles per hour (24 h per cycle). These results indicate that the frequencies captured by HHT and FFT are close at low frequency.

4.1.3. Comparison with superimposed IMFs C2–C5

As mentioned above, IMFs C2–C5 are demonstrated to be meaningful components for original time series data. Therefore, HHT and FFT are additionally adopted to analyze the superimposed IMFs C2–C5. From Fig. 6f, the frequency of FFT distributes obviously on 0.0417 and 0.0833 cycles per hour, and it only captures semi-daily period and daily period patterns including 12 h per cycle and 24 h per cycle since the noise and IMFs of low frequency are removed. However, HHT clearly illustrates the energy distribution that is concentrated on weekdays and weekends. From Hilbert spectra and Hilbert marginal spectra of superimposed IMFs C2–C5 (refer to Fig. 4b), the first two highest energy locates at the frequencies of 0.0549 and 0.1020 cycles per hour. The frequency of 0.0549 cycles per hour means 18.2 h per cycle and represents the approximate service periods of passenger flow in the metro system (from 6:00 AM to 1:00 AM the next day). The frequency of 0.1020 cycles per hour means 9.8 h per cycle and represents semi-service periods of passenger flow (one half of service hours, i.e., 9.5 h). HHT can capture service period and semi-service period patterns since the noise and IMFs of low frequency are removed. That is, for superimposed IMFs C2–C5, FFT only captures semi-daily and daily patterns, but HHT can additionally capture semi-service period and service period patterns. In addition, the Hilbert spectrum reveals that morning peak periods (8:00 AM–9:00 AM) on weekdays hold a stronger energy distribution. It reflects the peak characteristics of the metro system. Notice that the information can only be provided by HHT.

4.2. Discussions

In this paper, the time series data of short-term passenger flow in a metro system are decomposed into nine independent intrinsic modes and a residue with various frequencies. The individual IMFs C2–C5 illustrates the time variants of passenger flow such as peak period pattern, semi-service period pattern, semi-daily pattern and daily pattern.

Exploring the extracted IMFs by HHT brings out some interesting patterns of the time variants in short-term passenger flow. For example, according to our prior knowledge, the passenger flow of weekdays and weekends may behave similarly, since a large number of passengers take the Muzha line to visit Muzha Zoo, which is located near the terminal station. However, from the results, the weekday period has a stronger energy distribution among the discovered patterns of time variants. This indicates that the leisure trips on weekends are less than the work trips on weekdays. Decision makers in a metro system can arrange an appropriate service plan with respect to the patterns of time variants to save operating costs and improve performance.

The distinct features of HHT in the analysis of passenger flow are stated as follows. First, previous studies mainly applied traditional methods such as multiple regression to analyze passenger flow. They focused on studying the influential explanatory variables (e.g., population, employment, land use, social economic, fares, service level), and on discussing the explanation ability, so that the physical entities of passenger flow have been neglected. However, the HHT method can overcome the disadvantages of traditional methods such as the assumptions of linearity and stationarity, and reflect the fluctuations of passenger flow more accurately. By using the sifting process, IMFs can completely reveal the unique pattern by themselves (e.g., IMFs C2–C5). Second, most previous studies focused on studying the short-term traffic flow. However, this study obtains the time variants of passenger flow in the metro line by using HHT. HHT can provide more information (e.g., time–frequency–energy) to help decision makers establish metro operation plans such as equipment maintenance plans and train service plans (e.g., headway), service levels, train schedules, timetables and crew schedules.

By comparing HHT and FFT, two points are worth to be mentioned. From a theoretical point of view, FFT is restricted to analyze linearity and stationary data. It provides the global frequency distribution and generates a wider frequency range to fit harmonic function. However, HHT, which is free of linearity and stationarity assumption, is adaptive to analyze non-linearity and non-stationary data. EMD can serve as a filter to extract the meaningful components from the original time series. Hilbert transform offers the instantaneous frequency to reflect the time variants of time series data and it obtains the time–frequency–energy distribution. Therefore, HHT is more adaptive to analyze real-word complexity data than FFT. Provided that FFT is applied to analyze non-linear and non-stationary data, two issues should be noticed. First, to fit non-stationary time series data with a constant amplitude and frequency, the sinusoidal functions will require a much wider range of frequency. Second, the non-linear harmonic distortion will cause a leakage effect from the low frequency to generate a high frequency range of spectrum. In practice, FFT is incapable of separating high frequency noise, which may present the external environment changes such as weather change (a sunny day/a rainy day), special activities, etc. However, HHT can eliminate the high frequency noise and reveal several passenger patterns. For the case of superimposed IMFs C2–C5, HHT captures semi-service and service patterns, but FFT obtains semi-daily and daily patterns. Semi-service and service patterns imply the characteristic of passenger flow during the service period, yet semi-daily and daily patterns reveal the characteristic of passenger flow during the service

period and the non-service period. However, the characteristic of passenger flow during the non-service period is meaningless because the operation strategy focuses on the passenger flow demand during the service period. Obviously, semi-service and service patterns obtained by HHT can reveal more information of time variants of passenger flow than that of semi-daily and daily patterns obtained by FFT. By considering the characteristics of semi-service and service patterns, planners can generate proper operation strategies and enhance the performance of metro systems. Therefore, HHT is superior to FFT in analyzing non-stationary and non-linear data of passenger flow in a metro system.

5. Applications

Muzha line is a computerized, automated and driverless metro system, and it is controlled by the operation control center. Managers of metro systems can observe the dynamic fluctuation of passenger flow through the time–frequency–energy distribution revealed in the Hilbert spectrum. For example, in the original passenger flow data, the strongest energy locates around afternoon peak periods (17:30 PM–19:00 PM) on Monday, May 5. The time–frequency–energy distribution may reflect the congestion of passenger flow at some specific time periods in the metro system. In such a situation, managers can make operation plans such as arranging additional trains to the main line in advance to relieve the passenger congestion. Mangers can observe whether the similar pattern of congestion appears during the same periods on any weekdays.

By using the time–frequency–energy distribution of passenger flow, train schedules and crew schedules in transportation systems could be planned in advance. The Hilbert spectrum of superimposed IMFs C2–C5 clearly indicates the morning peak period (8:00 AM–9:00 AM) on weekdays. Based on extracted morning peak characteristics of the metro system, planners of metro system can set up various operation plans such as train schedules, crew schedules and train headways in morning peak and off-peak periods on weekdays to ensure the required service level of the metro system. Additionally, managers can thus arrange the appropriate number of staff to enhance the service level with a reasonable operation cost. Furthermore, managers can examine the capacity of station facility and equipment to ensure that it meets safety regulations of transportation systems.

6. Conclusions

In this paper, a signal processing method developed recently, namely HHT, has been adopted to analyze the short-term passenger flow of a metro system. The HHT approach can extract the oscillations embedded in data without setting any subjective preliminary assumption. The IMFs extracted by EMD present some useful time variants patterns of short-term passenger flow including peak period pattern, semi-service period pattern, semi-daily pattern and daily pattern. The results of HSA also indicate the time–frequency–energy distribution, which can be used to generate effective operation plans in a metro system. Additionally, HHT spectra are compared favorably to the traditional FFT spectra for presenting various frequencies. The comparisons indicate that HHT can obtain the narrower frequency band, and capture more information represented by a time–frequency–energy distribution, which can help to enhance the transportation system planning. From the results obtained in this study, HHT has been demonstrated to be an effective approach for discovering the time variant characteristics of short-term passenger flow.

Although HHT can effectively reflect the characteristics of passenger flow, the HHT method has some shortcomings (Huang and Wu, 2008). First, HHT can not directly explain the time vari-

ants of original data. While a finite set of IMFs can be generated by EMD, these IMFs are lack of theoretical fundamentals for explaining the physical meanings. Generally, it is necessary to interpret the meanings of extracted IMFs with domain knowledge. Second, the present EMD method can not analyze the data more than two dimensions (apart from the time dimension). Hence, the time variants of passenger flow in terms of spatial dimension can not be obtained in this study. Recently, some researchers have tried to develop the EMD approach for extracting IMFs from two-dimensional data (Bhuiyan et al., 2008; Shi et al., 2009). However, this paper analyzes the variants of passenger flow in an entire metro line with two dimensions including time and passenger flow by using HHT. In the future, it is worth to develop a two-dimensional EMD approach (i.e., temporal-station pair-passenger flow EMD) to additionally analyze the time variants of passenger flow in terms of stations. Third, one IMF component may include both high frequency and low frequency, and a frequency may exist in different IMFs such that the explanation of IMFs is complicated. The future work can further apply ensemble empirical mode decomposition (EEMD) method (Wu and Huang, 2008), which adds white noise in the sifting process, to deal with the problem of mode mixing. Fourth, although EMD and HHT can generate useful service patterns, they do not have the forecasting capability. However, the meaningful IMFs can be used as the input variables of forecasting models such as neural networks and support vector machines for regression to enhance the forecasting performance. Developing the hybrid forecasting model by incorporating with EMD is also a valuable future work.

Acknowledgements

References

Abdel-Aty, M.A., Pemmanaboina, R., 2006. Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data. IEEE Transactions on Intelligent Transportation Systems 7 (2), 167–174.

Alfa, A.S., 1986. A review of models for the temporal distribution of peak traffic demand. Transportation Research Part B 20 (6), 491–499.

Assis, W.O., Milani, B.E.A., 2004. Generation of optimal schedules for metro lines using model predictive control. Automatica 40 (8), 1397–1404.

Balocchi, R., Menicucci, D., Santarcangelo, E., Sebastiani, L., Gemignani, A., Ghelarducci, B., Varanini, M., 2004. Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition. Chaos Solitons & Fractals 20 (1), 171–177.

Bar-Gera, H., Boyce, D., 2003. Origin-based algorithms for combined travel forecasting models. Transportation Research Part B 37 (5), 405–422.

Bhuiyan, S.M.A., Adhami R.R., Khan J.F., 2008. Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. EURASIP Journal on Advances in Signal Processing, p. 164.

Blanco-Velasco, M., Weng, B., Barner, K.E., 2008. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Computers in Biology and Medicine 38 (1), 1–13.

Chandra, S.R., Al-Deek, H., 2009. Predictions of freeway traffic speeds and volumes using vector autoregressive models. Journal of Intelligent Transportation Systems 13 (2), 53–72.

Chen, H.B., Grant-Muller, S., 2001. Use of sequential learning for short-term traffic flow forecasting. Transportation Research Part C 9 (5), 319–336.

Chen, C., Chen, J., Barry, J., 2009. Diurnal pattern of transit ridership: a case study of the New York city subway system. Journal of Transport Geography 17 (3), 176–186.

Dong, Y., Li, Y., Xiao, M., Lai, M., 2008. Analysis of earthquake ground motions using an improved Hilbert–Huang transform. Soil Dynamics and Earthquake Engineering 28 (1), 7–19.

Elgammal, A., Duraiswami, R., Davis, L.S., 2003. Efficient Kernel density estimation using the fast Gauss transform with applications to color modeling and tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (11), 1499–1504.

Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transportation Research Part B 41 (3), 363–378.

Golias, J.C., 2002. Analysis of traffic corridor impacts from the introduction of the new Athens metro system. Journal of Transport Geography 10 (2), 91–97.

Guhathakurta, K., Mukherjee, I., Chowdhury, A.R., 2008. Empirical mode decomposition analysis of two different financial time series and their comparison. Chaos, Solitons & Fractals 37 (4), 1214–1227.

Guo, K., Zhang, X., Li, H., Meng, G., 2008. Application of EMD method to friction signal processing. Mechanical Systems and Signal Processing 22 (1), 248–259.

Gurwicz, Y., Lerner, B., 2005. Bayesian network classification using spline-approximated kernel density estimation. Pattern Recognition Letters 26 (11), 1761–1771.

Hamad, K., Shourijeh, M.T., Lee, E., Faghri, A., 2009. Near-term travel speed prediction utilizing Hilbert–Huang transform. Computer-Aided Civil and Infrastructure Engineering 24 (8), 551–576.

Huang, D.-W., 2003. Wavelet analysis in a traffic model. Physica A 320 (1–2), 298–308.

Huang, N.E., Wu, Z.H., 2008. A review on Hilbert–Huang transform: method and its applications to geophysical studies. Reviews of Geophysics 46 (2), RG2006.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A 454 (1971), 903–995.

Huang, N.E., Wu, M.-L., Qu, W., Long, S.R., Shen, S.S.P., 2003a. Applications of Hilbert–Huang transform to non-stationary financial time series analysis. Applied Stochastic Models in Business and Industry 19 (3), 245–268.

Huang, N.E., Wu, M.L., Long, S.R., Shen, S.S.P., Qu, W.D., Gloersen, P., Fan, K.L., 2003b. A confidence limit for the empirical mode decomposition and the Hilbert spectral analysis. Proceedings of the Royal Society of London. A 459 (2037), 2317–2345.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 2004. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society London. A 460 (2046), 1597–1611.

Hwang, P.A., Huang, N.E., Wang, D.W., 2003. A note on analyzing nonlinear and nonstationary ocean wave data. Applied Ocean Research 25 (4), 187–193.

Jiang, X., Hon, H.A., 2005. Dynamic wavelet neural network model for traffic flow forecasting. Journal of Transportation Engineering 131 (10), 771–779.

Jiang, R., Yan, H., 2008. Studies of spectral properties of short genes using the wavelet subspace Hilbert–Huang transform (WSHHT). Physica A 387 (16–17), 4223–4247.

Jovicic, G., Hansen, C.O., 2003. A passenger travel demand model for Copenhagen. Transportation Research Part A 37 (4), 333–349.

Kuby, M., Barranda, A., Upchurch, C., 2004. Factors influencing light-rail station boardings in the United States. Transportation Research Part A 38 (3), 223–247.

Kulshreshtha, M., Nag, B., 2000. Structure and dynamics of non-suburban passenger travel demand in Indian railways. Transportation 27 (2), 221–241.

Lee, S., Lee, Y.-I., Cho, B., 2006. Short-term travel speed prediction models in car navigation systems. Journal of Advanced Transportation 40 (2), 123–139.

Li, H.G., Meng, G., 2006. Detection of harmonic signals from chaotic interference by empirical mode decomposition. Chaos, Solitons & Fractals 30 (4), 930–935.

Li, Q.S., Wu, J.R., 2007. Time–frequency analysis of typhoon effects on a 79-storey tall building. Journal of Wind Engineering and Industrial Aerodynamics 95 (12), 1648–1666.

Li, K.P., Gao, Z.Y., Zhao, X.M., 2008. Multiple scale analysis of complex networks using the empirical mode decomposition method. Physica A 387 (12), 2981–2986.

Liang, H., Lin, Q.-H., Chen, J.D.Z., 2005. Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease. IEEE Transactions on Biomedical Engineering 52 (10), 1692–1701.

Paris, H., Van den Broucke, S., 2008. Measuring cognitive determinants of speeding: a application of the theory of planned behavior. Transportation Research Part F 11 (3), 168–180.

Rai, V.K., Mohanty, A.R., 2007. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. Mechanical Systems and Signal Processing 21 (6), 2607–2615.

Shi, W.Z., Tian, Y., Huang, Y., Mao, H.X., Liu, K.F., 2009. A two-dimensional empirical mode decomposition method with application for fusing panchromatic and multispectral satellite images. International Journal of Remote Sensing 30 (10), 2637–2652.

Su, Z.-Y., Wang, C.-C., Wu, T., Wang, Y.-T., Tang, F.-C., 2008. Instantaneous frequency–time analysis of physiology signals: the application of pregnant women's radial artery pulse signals. Physica A 387 (2–3), 485–494.

Tao, Q., Chen, Q., Li, L., 2005. Analytic unit quadrature signals with nonlinear phase. Physica D 203 (1–2), 80–87.

Tsekeris, T., Stathopoulos, A., 2006. Gravity models for dynamic transport planning: development and implementation in urban networks. Journal of Transport Geography 14 (2), 152–160.

Vanajaksi, L., Rilett, L.R., 2007. Support vector machine technique for the short term prediction of travel time. IEEE Intelligent Vehicles Symposium 600, 605.

Varagouli, E.G., Simos, T.E., Xeidarkis, G.S., 2005. Fitting a multiple regression line to travel demand forecasting: the case of the prefecture of Xanthi, Northern Greece. Mathematical and Computer Modeling 42 (7–8), 817–836.

Vlahogianni, E.I., Golia, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. Transport Reviews 24 (5), 533–557.

Wardman, M., 2006. Demand for rail travel and the effects of external factors. Transportation Research Part E 42 (3), 129–148.

Williams, B.M., Durvasula, P.K., Brown, D.E., 1998. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. Transportation Research Record 1644, 132–141.

Wirasinghe, S.C., Kumarage, A.S., 1998. An aggregate demand model for intercity passenger travel in Sri Lanka. Transportation 25 (1), 77–98.

Wu, M.-C., 2007. Phase correlation of foreign exchange time series. Physica A 375 (2), 633–642.

Wu, Z., Huang, N.E., 2008. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis 1 (1), 1–41.

Wu, C.-H., Ho, J.-M., Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE Transaction Intelligent Transportation 5 (4), 276–281.

Xie, H., Wang, Z., 2006. Mean frequency derived via Hilbert–Huang transform with application to fatigue EMG signal analysis. Computer Methods and Programs in Biomedicine 82 (2), 114–120.

Xie, Y., Zhang, Y., 2006. A wavelet network model for short-term traffic volume forecasting. Journal of Intelligent Transportation Systems 10 (3), 141–150.

Zhang, Y., Ye, Z., 2008. Short-term traffic flow forecasting using fuzzy logic system methods. Journal of Intelligent Transportation Systems 12 (3), 102–112.

Zhang, X., Lai, K.K., Wang, S.-Y., 2008. A new approach for crude oil price analysis based on empirical mode decomposition. Energy Economics 30 (3), 905–918.

Zhou, B(Brenda)., Kockelman, K.M., 2009. Predicting the distribution of households and employment: a seemingly unrelated regression model with two spatial processes. Journal of Transport Geography 17 (5), 369–376.