Transportation Research Part C 21 (2012) 148–162

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks

Yu Wei, Mu-Chen Chen *

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

ARTICLE INFO

Article history:
Received 16 August 2010
Received in revised form 27 June 2011
Accepted 29 June 2011

Keywords:
Forecasting
Short-term metro passenger flow
Empirical mode decomposition
Neural networks

ABSTRACT

Short-term passenger flow forecasting is a vital component of transportation systems. The forecasting results can be applied to support transportation system management such as operation planning, and station passenger crowd regulation planning. In this paper, a hybrid EMD–BPN forecasting approach which combines empirical mode decomposition (EMD) and back-propagation neural networks (BPN) is developed to predict the short-term passenger flow in metro systems. There are three stages in the EMD–BPN forecasting approach. The first stage (EMD Stage) decomposes the short-term passenger flow series data into a number of intrinsic mode function (IMF) components. The second stage (Component Identification Stage) identifies the meaningful IMFs as inputs for BPN. The third stage (BPN Stage) applies BPN to perform the passenger flow forecasting. The historical passenger flow data, the extracted EMD components and temporal factors (i.e., the day of the week, the time period of the day, and weekday or weekend) are taken as inputs in the third stage. The experimental results indicate that the proposed hybrid EMD–BPN approach performs well and stably in forecasting the short-term metro passenger flow.

1. Introduction

Short-term passenger flow forecasting is a vital component of transportation systems which can be used to fine-tune travel behaviors, reduce passenger congestion, and enhance service quality of transportation systems. The forecasting results of short-term passenger flow can be applied to support transportation system management such as operation planning, and station passenger crowd regulation planning.

Short-term traffic forecasting (Smith et al., 2002; Vlahogianni et al., 2004) and short-term passenger demand forecasting (Tsai et al., 2009) are successful applications of short-term transportation forecasting in the literature. The short-term transportation forecasting approaches can be generally divided into two categories: parametric and non-parametric techniques (Smith et al., 2002; Vlahogianni et al., 2004). Parametric techniques and non-parametric techniques refer to the functional dependency assumed between independent variables and the dependent variable.

In the traditional parametric techniques, historical average (Smith and Demetsky, 1997), smoothing techniques (Williams et al., 1998), and autoregressive integrated moving average (ARIMA) (Hansen et al., 1999; Lee and Fambro, 1999) have been applied to forecast transportation demand. Particularly, ARIMA has become one of the common parametric forecasting approaches since the 1970s. The ARIMA model is a linear combination of time-lagged variables and error terms. The ARIMA model has been widely applied in forecasting short-term traffic data such as traffic flow, travel time, speed, and occupancy (e.g., Ahmed and Cook, 1979; Hamed et al., 1995; Lee and Fambro, 1999). In addition, with the characteristics of seasonality and trends in traffic data, some researchers have applied seasonal ARIMA to predict traffic flow (Williams and Hoel, 2003;

* Corresponding author.
E-mail address: ittchen@mail.nctu.edu.tw (M.-C. Chen).

0968-090X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trc.2011.06.009

Tan et al., 2009) and international air passenger flow (Faraway and Chatfield, 1998; Lim and McAleer, 2002; Chen et al., 2009). As stated in Brooks (2008), ARIMA performs well and robustly in modeling linear and stationary time series. However, the applications of ARIMA or seasonal ARIMA models are limited because they assume linear relationships among time-lagged variables so that they may not capture the structure of non-linear relationships (Zhang et al., 1998).

For the non-parametric techniques, several methods have been used to forecast the transportation demand such as neural networks (Dougherty, 1995; Vlahogianni et al., 2004), non-parametric regression (Smith et al., 2002; Clark, 2003), Kalman filtering models (Wang and Papageorgiou, 2007), and Gaussian maximum likelihood (Tang et al., 2003). Among these non-parametric techniques, neural networks have been frequently adopted as the modeling approach because they possess the characteristics of adaptability, nonlinearity and arbitrary function mapping capability (Zhang et al., 1998). Essentially, neural networks can deal with complex non-linear problems without *a priori* knowledge regarding the relationships between input and output variables (Zhang et al., 1998). Recently, several previous studies have developed neural network based models for traffic and transportation forecasting which include multilayer perceptron neural networks (Smith and Demetsky, 1997; Van Arem et al., 1997; Lee et al., 2006), Kalman filter based multilayer perceptron (Vythoulkas, 1993), time-delay neural networks (Zhang, 2000; Dia, 2001), radial basis function neural networks (Zheng et al., 2006), dynamic neural networks (Chen and Grant-Muller, 2001), state-space neural networks (Van Lint et al., 2005), and the support vector machine for regression (Vanajaksi and Rilett, 2007; Castro-Neto et al., 2009), etc. Multilayer perceptron (MLP) neural networks have been frequently adopted in the literature since they have been demonstrated to outperform ARIMA in solving non-linear problems.

In order to construct more efficient neural networks, several previous studies have proposed hybrid MLP models by incorporating the ARIMA model (Zhang, 2003; Zeng et al., 2008), seasonal ARIMA model (Tseng et al., 2002; Zhang and Qi, 2005), genetic algorithms (Vlahogianni et al., 2005), fuzzy logic rules (Rojas et al., 2008; Zhang and Ye, 2008), and so on. The above hybrid models have been demonstrated to possess better performance than pure MLP neural networks in short-term traffic forecasting.

In the last two decades, traditional travel demand forecasting models including trip generation, trip distribution, mode choice, and assignment steps have been applied to forecast the passenger demand in the literature (e.g., Alfa, 1986; Wirasinghe and Kumarage, 1998; Jovicic and Hansen, 2003; Bar-Gera and Boyce, 2003). However, the applications of traditional travel demand models are limited because a linear relationship among variables is assumed, and a large set of explanatory variables needs to be calibrated. Recently, some researchers have applied neural networks and hybrid models to improve travel demand forecasting. Zhou et al. (2007) developed a hybrid approach, which combines a back-propagation neural network and three traditional travel demand forecasting models (trip generation, trip distribution, and mode choice) to simulate the travelers' inter-city behaviors. In addition, Tsai et al. (2009) developed two neural network based approaches depending on distinctive railway data for short-term railway passenger demand forecasting, in which some temporal features were adopted to improve the forecasting performance. Extracting meaningful patterns or features embedded implicitly in data by data pre-processing may enhance the capability of forecasting models, but relatively few studies have focused on developing data pre-processing techniques in neural network based forecasting models.

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; Huang et al., 2003a). EMD is an empirical, intuitive, direct and adaptive data processing method developed especially for dealing with nonlinear 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 (Huang et al., 1998). Therefore, according to the scale, the physical feature of each mode can be identified. EMD has been successfully applied in several fields such as ocean waves (Veltcheva and Soares, 2004), biomedical engineering (Liang et al., 2005; Jiang and Yan, 2008), signal processing (Li and Meng, 2006; Blanco-Velasco et al., 2008), wind engineering (Li and Wu, 2007), earthquake engineering (Dong et al., 2008), and financial time series (Zhang et al., 2008).

As EMD can decompose data into a small number of independent components, some researchers developed hybrid forecasting approaches by combining EMD with forecasting models to obtain better performance in various fields of signal processing (Shen et al., 2008), short-term electronic loading (Zhu et al., 2007), and traffic engineering (Hamad et al., 2009). Shen et al. (2008) used EMD to extract the signal feature for improving multi-class support vector machines in signal processing. Zhu et al. (2007) presented a hybrid method based on EMD and support vector machines for short-term electronic load forecasting. Additionally, Hamad et al. (2009) applied a hybrid approach of EMD and back-propagation neural networks to predict short-term travel speed by using a set of real-life loop detector data. The experimental results in Hamad et al. (2009) revealed that the hybrid approach, which takes advantage of extracting components by EMD, outperforms the traditional traffic volume prediction model. However, Hamad et al. (2009) only focused on short-term travel speed forecasting for freeway systems. To date, the short-term passenger flow forecasting still remains insufficiently researched in the literature.

The research motivation of this paper is that short-term passenger flow forecasting is an important element in metro systems; however, relatively few studies have focused on this issue. In addition, for the forecasting model of short-term passenger flow, it is essential to capture the characteristics of short-term passenger flow and to embed these characteristics into the forecasting model to enhance forecasting performance. Hence, this paper attempts to develop a hybrid approach which combines EMD and neural networks for short-term passenger flow forecasting.

There are some distinctions between the hybrid approach developed in this paper and the existing approach developed by Hamad et al. (2009). From a methodology perspective, Hamad et al. (2009) did not discuss how to identify the suitable EMD

components for constructing forecasting models. This paper uses statistical analysis to identify the suitable EMD components. In addition, this paper proposes a scheme to systematically identify or combine the extracted EMD components as input variables for the forecasting model. Meanwhile, temporal factors are considered because these factors can represent the features of the passenger flow. Hence, the proposed hybrid forecasting approach, which considers the extracted EMD components and temporal factors, is developed to effectively forecast the short-term passenger flow. As for the perspective of forecasting tasks, Hamad et al. (2009) focused on short-term travel speed forecasting of freeway systems, and the forecasting results can be applied to forecast other traffic parameters (e.g., travel time) for an Intelligent Transportation System. This paper focuses on short-term passenger flow forecasting for metro systems, and the forecasting results can be used as the basis for operations management in metro systems. Furthermore, Hamad et al. (2009) considered EMD essential components in the forecasting model. In this paper, different types of input variables and variable transformations are systematically investigated to obtain the appropriate input variables for the forecasting model.

The remainder of this paper is organized as follows. Section 2 describes EMD, neural networks, and the hybrid forecasting approach. Section 3 illustrates the component extraction by using EMD, and presents the forecasting model's design. To investigate the performance of the proposed hybrid approach, Section 4 compares the results obtained by the proposed hybrid approach and those by various back-propagation neural network based models and Seasonal Autoregressive Integrated Moving Average. Section 5 finally concludes this paper.

2. Methodology

2.1. Empirical mode decomposition

Empirical mode decomposition (EMD) can deal with non-linear and non-stationary data. The main idea of EMD is to decompose 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 data set 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.

The essence of EMD is the sifting process which extracts IMFs from the original data. The algorithm of EMD is described as follows (Huang et al., 1998):

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 envelope as

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

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 designed as a proto-intrinsic mode function,

$$h_1(t) = x(t) - m_1(t). \tag{2}$$

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 $x(t)$,

$$r_1(t) = x(t) - h_1(t). \tag{3}$$

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 when 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:

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

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. 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. Neural networks

Multilayer perceptron (MLP) is one of the most prevalent neural networks. It has the capability of complex mapping between inputs and outputs that makes it possible to approximate nonlinear functions (Zhang et al., 1998; Balestrassi et al., 2009). The architecture of MLP consists of multiple layers, which include an input layer, one or more hidden layers, and an output layer. Each layer comprises several neurons connected to the neurons in neighboring layers. Since MLP contains many interacting nonlinear neurons in multiple layers, it can capture complex phenomena.

MLP is trained by using the historical data to capture the characteristics of data. The parameters of MLP, including weights and biases, are adjusted iteratively by a process of minimizing the global error or the total square error. Consider a 3-layer MLP network consisting of n input neurons, a hidden layer of s hidden neurons and a layer of m output neurons. MLP operates as follows (Kumar, 2005; Ince and Trafalis, 2006). Input neurons receive a pattern vector $x = (x_1, x_2, \ldots, x_n)$ from the dataset and propagate it to all neurons in hidden layers. Each hidden neuron h firstly computes the net input (y_h) and generates an output (Y_h) as shown in following equations.

$$y_h = \sum_{i=1}^{n} w_{ih} x_i, \tag{5}$$

$$Y_h = f(y_h) = f\left(\sum_{i=1}^{n} w_{ih} x_i\right), \tag{6}$$

where w_{ih} means the weight between the ith neuron of the input layer and the hth neuron of the hidden layer, $f(\cdot)$ is the activation function. The activation function, also named the transfer function, determines the relationship between inputs and outputs of a node and a network. The logistic function as shown in Eq. (7) is often used as the activation function in hidden layers (Zhang, 2003),

$$f(x) = \frac{1}{1 + \exp(-x)}. \tag{7}$$

Each output neuron j receives outputs of hidden layer as inputs, and continuously repeats all the operations described above until the termination condition is met. The iterative process of output O_j is expressed by Eq. (8) (Kumar, 2005; Ince and Trafalis, 2006).

$$O_j = f\left(\sum_{h=1}^{s} Y_h w_{hj}\right) = f\left(\sum_{h=1}^{s} f\left(\sum_{i=1}^{n} x_i w_{ih}\right) w_{hj}\right), \tag{8}$$

where n is the number of input neurons, and s is the number of hidden neurons.

This paper applies MLP with a back-propagation algorithm to develop a short-term passenger flow forecasting model. The back-propagation algorithm is one of the most popular learning algorithms adopted in MLP; this is a supervised learning algorithm which minimizes the global error by using the generalized delta rule and the gradient steepest descent method. The learning procedure of a back-propagation neural network (BPN) is described as follows (Kumar, 2005). Readers are referred to Kumar (2005) for further details.

Step 1: Perform input data normalization, and present input data to the input layer.
Step 2: Determine the network architecture and parameters including learning rate, momentum, activation function, and initial weights.
Step 3: Compute the error over output neurons by comparing the generated outputs with the desired outputs.
Step 4: Compute the weight changes. Apply the error to compute the weight change between the hidden layer and output layer, and the weight change between the input layer and hidden layer.
Step 5: Update weights.

Hidden layer to output layer : $w_{hj}^{k+1} = w_{hj}^{k} + \Delta w_{hj}^{k}$,

Input layer to hidden layer : $w_{ih}^{k+1} = w_{ih}^{k} + \Delta w_{ih}^{k}$,

where i, h, and j represent the input layer, hidden layer, and output layer, respectively. w_{hj}^{k+1} is the weight change between the hidden layer and output layer in $(k+1)$th iteration, w_{ih}^{k+1} is the weight change between the input layer and hidden layer in $(k+1)$th iteration, and Δw_{hj}^{k} and Δw_{ih}^{k} represent the net weight changes.

Step 6: Repeat Steps 1–5 until the global error satisfies a predefined threshold.

The challenging task of using BPN is to determine the number of hidden layers, the number of neurons in each hidden layer, learning rate, momentum, and activation function. These parameters and network architecture may be set by using pilot tests. However, some previous studies (e.g., Zhang et al., 1998; Balestrassi et al., 2009) have provided some guidelines or methods to determine them.

2.3. The hybrid EMD–BPN model

In this paper, the proposed hybrid approach for the short-term passenger flow forecasting, namely EMD–BPN, combines EMD with BPN. In EMD–BPN, there are three stages including data decomposition by EMD, EMD component identification and forecasting by BPN. Fig. 1 illustrates the hybrid EMD–BPN approach. These three stages are described as follows.

2.3.1. Stage 1: EMD Stage

The first stage (EMD Stage) decomposes the original passenger flow series data into a number of IMFs, and it also plays the role of data filtering. The extracted IMF components can represent a range of frequencies, from high to low, and reveal various periodic patterns of passenger flow. Each IMF component can represent the local characteristic time scale by itself. The computational requirements of BPN increase to an extreme degree with the number of input nodes. Therefore, identification of input data for BPN is essential to reduce the computational requirements while the forecasting accuracy is maintained (Zheng et al., 2004).

2.3.2. Stage 2: Component Identification Stage

Stage 2 focuses on identifying the meaningful IMFs as inputs for BPN. The extracted IMFs represent a range of frequencies, from high to low. The IMF with higher frequency represents the pattern of a shorter period, whereas the IMF with lower frequency represents the pattern of a longer period. Although state-of-the-art MLP solvers and training algorithms are well prepared to cope with a large number of inputs, the lower correlated inputs (e.g., IMFs with a longer period) can increase dimensionality and computation time. Therefore, in order to build an appropriate short-term forecasting model, it is necessary to identify the meaningful patterns of IMFs.

The meaningful patterns could be measured by the correlation between each IMF component and the original time series data by using statistical methods such as Pearson product moment correlation and Kendal rank correlation. If a Pearson product moment correlation coefficient is positive, it implies that the IMF component change is consistent with the original data change. The higher the correlation coefficient is, the stronger the correlation of the component and original data is. The

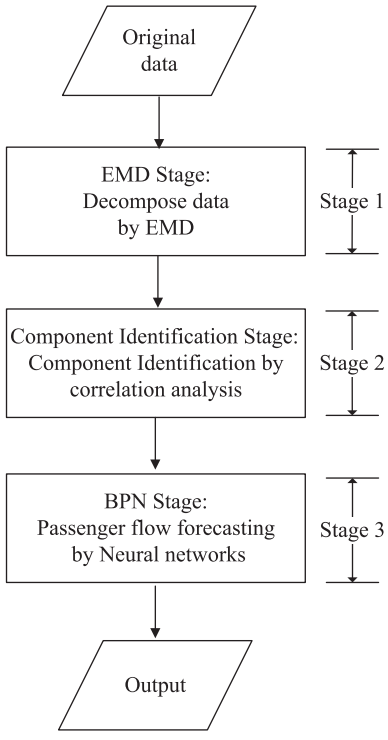

Fig. 1. The hybrid EMD–BPN approach.

other IMF components with lower correlation are classified into the rest of components which can be retained and aggregated as a part of input nodes in the hybrid model. The local characteristic of IMF admits well-behaved Hilbert transforms, from which the instantaneous frequencies can be calculated (Huang et al., 1998).

2.3.3. Stage 3: BPN Stage

For a forecasting model, optimal input feature (attribute, or variable) selection is one of the most important issues. Many researchers have investigated feature selection, and various methods have been proposed. Generally, the feature selection approaches can be divided into two major categories: the filter approach (Blum and Langley, 1997) and the wrapper approach (Kohavi and John, 1997). The filter approach selects a feature subset using a pre-processing step. This approach divides the data set into relevant and irrelevant features by learning algorithms and takes into account the relevant features. The wrapper approach selects an optimal feature subset from the space of possible subsets of features using the induction algorithm itself as a part of the evaluation function (Kohavi and John, 1997). This second approach can obtain the optimal feature subsets, yet it is not easy to use because of the complexity of computation time and space. However, this paper focuses on developing a systematic design guideline for the hybrid EMD–BPN approach, which considers both higher correlated IMFs and lower correlated IMFs (i.e., relevant IMFs and less relevant IMFs on the basis of correlation with the original data). Hence, the EMD–BPN hybrid approach is neither the filter approach nor the wrapper approach.

In order to make the unique pattern of the extracted IMFs to reflect the forecasting model and to enhance the forecasting performance, this paper provides a systematic design guideline on the hybrid EMD–BPN forecasting. The systematic combinations of IMFs are as follows. Firstly, an individual IMF which reveals the respective pattern for original data is taken as an input. Secondly, by considering the dimensionality of the input vector and the computational requirements, the extracted IMFs can be taken as inputs in individual or aggregated form. Thirdly, IMFs are classified into higher correlated IMFs and lower correlated IMFs with the original data in Stage 2. In Stage 3, the higher correlated IMFs are taken as inputs in individual or aggregated form. The lower correlated IMFs are considered as inputs in aggregated form.

The third stage (BPN Stage) in the hybrid model applies BPN to perform the passenger flow forecasting. Several parameters should be determined in building the BPN-based forecasting model. These parameters include the number of input nodes, number of hidden layers, number of hidden nodes, activation function, learning rate, and momentum. The number of input nodes (i.e., input variables) should be determined carefully because it can affect either the learning or prediction capability. The number of input nodes is probably the most critical parameter for a time series forecasting model, since it contains important information embedded in the data (Zhang et al., 1998). Input variables have to capture the passenger flow dynamics and the associated patterns directly affecting the accuracy of forecasting. In the hybrid model, IMFs can be regarded as multiple series data and input variables. The number of time-lags of input variables is based on the rolling step and the rolling horizon. The rolling step represents the resolution of IMFs series used for prediction. The rolling horizon is equal to the number of time-lags of input variables in the input layer. If the rolling horizon is too large, the forecasting model responds slowly to the fluctuation of passenger flow. On the contrary, if the rolling horizon is too small, the forecasting model overreacts to the fluctuation of passenger flow.

In the output layer, output variables include the forecasting step and the forecasting horizon. The forecasting step, also called the step ahead, represents the time interval at which prediction is executed or updated. The forecasting horizon indicates the extent of time ahead to which the forecast is referring. Generally, a lengthier forecasting horizon leads to less accuracy of the forecasting model. However, in the relevant forecasting literature, a variety of views exist on determination of the rolling step, the rolling horizon, the forecasting step and the forecasting horizon. For the forecasting step, a 5-min interval or more (Park et al., 1998), 10-min interval (Vythoulkas, 1993), and 15-min interval for traffic forecasting in the Highway Capacity Manual (2000) and Smith and Demetsky (1997) have been suggested. It has been suggested that the rolling step should equal the forecasting step (Hamad et al., 2009). Obviously, this can be understood intuitively by considering the fact that determination of these variables depends on the application areas of forecasting models.

In this paper, the short-term passenger flow forecasting can provide information for operations management. The appropriate forecasting horizon and accuracy can be the base for train service planning. Considering these essential parameters mentioned above, the rolling step or forecasting step is identified as a 15-min interval based on the collected time intervals. The rolling horizon is identified as six rolling steps. The forecasting horizon is identified from 15 min to 60 min (i.e., four 15-min intervals). The arrangement is similar to a sliding window. This paper uses the sliding window concept to build the data set of short-term passenger flow forecasting. The sliding window size (i.e., the rolling horizon) in the input layer is equal to six steps, and the sliding window size (i.e., the forecasting horizon) in the output layer is up to four steps. In addition, in the proposed EMD–BPN approach, the temporal factors of the passenger flow can be additionally considered. These temporal factors include the day of the week, the time period of the day, and weekday or weekend.

From a theoretical viewpoint, the benefits of the hybrid model can be discussed with the particular characteristics of short-term passenger flow forecasting, and these characteristics are appropriately considered in the designed methodology. From a forecasting perspective, the specific characteristics of short-term passenger flow reveal the various patterns such as peak period pattern, daily pattern, weekly pattern, and monthly pattern. A better forecasting model can be built with these characteristics of short-term passenger flow. The IMFs extracted from the original data by using EMD may be able to reveal the abovementioned patterns. In EMD, no calibration is required in extracting IMFs. In addition, unlike other pre-processing methods, EMD does not change the properties of the original data and adapts to be a data pre-processing technique. There-

fore, EMD can completely reveal the properties of the original data because each IMF component represents a unique characteristic of the original data. Hence, these characteristics can appropriately be taken into account by the proposed EMD–BPN approach.

3. Experiment setup

3.1. Data

A passenger flow dataset of rapid transit is collected to investigate the viability of the proposed EMD–BPN approach for forecasting the short-term passenger flow. The dataset was collected from the Muzha Line of Taipei Rapid Transit Corporation (TRTC) during the period from May 1 to May 31, 2008. Muzha Line is a medium capacity transit system. There was one public holiday (i.e., May 1, the International Labor Day) in the 4-week period of data collection. The forecasting results may not be affected by this public holiday because most people besides laborers are still on duty. There are 2325 observations collected by an automatic fare collection system with a 15-min interval. Taking the operation time of TRTC into consideration, the time period of data collection for each day was from 6:00 AM to 00:45 AM the next day.

3.2. IMF component extraction

In the first stage (EMD Stage), the time series data of passenger flow are decomposed into nine IMF components (IMF C1, IMF C2, . . . , and IMF C9) and one residue (IMF C10) by using EMD, as shown in Fig. 2 (Chen and Wei, 2011). All the extracted IMF components are graphically illustrated in the order in which they are extracted, to indicate the order of frequency (or

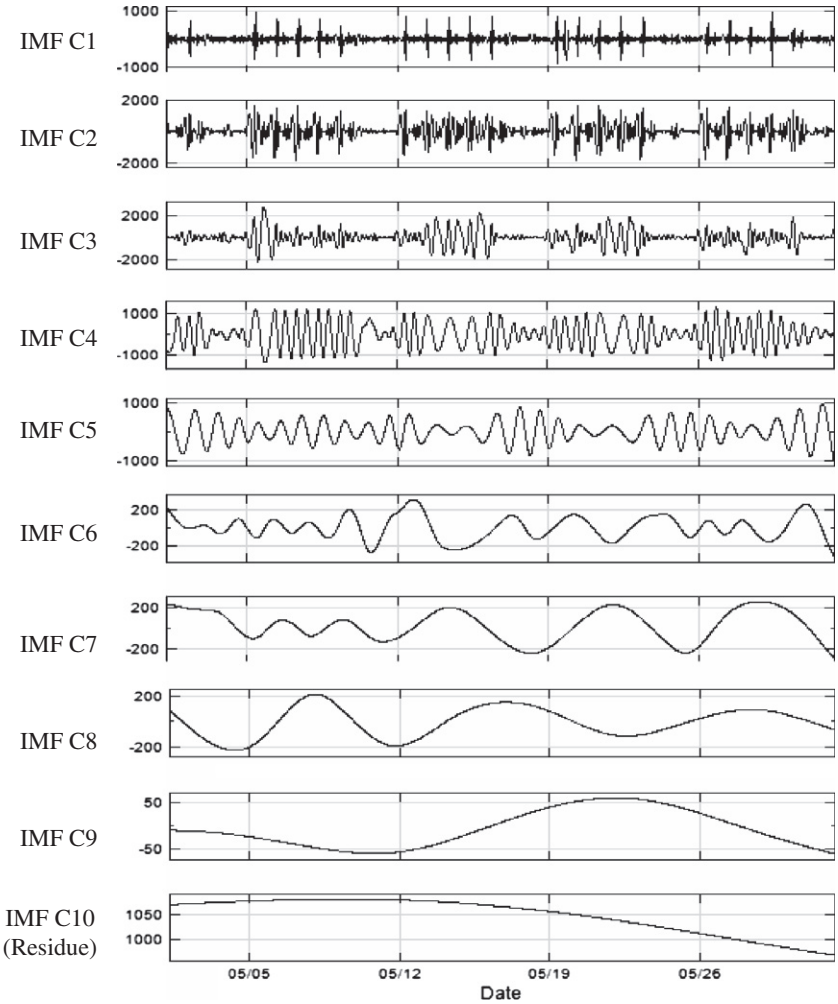

Fig. 2. The extracted IMF components (Chen and Wei, 2011).

period) from the highest frequency to the lowest one. 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 high time variant or noise in the original passenger flow data, while the last few components represent the long period components. The last component is the residue of sifting, which generally represents the trend of the time series. In this paper, EMD components are obtained by using Visual Signal 1.2.

3.3. Identification of IMF components

In the second stage (Component Identification Stage), in order to explore the physical phenomenon of extracted IMF components, Pearson product moment correlation and Kendal rank correlation are applied to measure the correlation between each IMF component and the original time series data. As summarized in Table 1, IMFs C1–C5, C7 and C8 are significantly different at a 99% confidence level, and Pearson correlation coefficients show they are positively correlated to the original time series. Pearson correlation coefficients of IMFs C3–C5 are respectively 0.545, 0.535 and 0.406, which indicate a stronger positive correlation. Observing the Kendall correlation coefficients, IMFs C3–C5 have higher ranks; the coefficients of these components are 0.376, 0.374 and 0.345, respectively. Observing Table 1, the results of Pearson correlation coefficients and Kendall correlation coefficients are consistent. From the percentage power of each IMF, IMFs C2–C4 have higher percentage power values, and IMF C3 has the highest percentage power (34.4%). From Table 1, the percentage power values of IMFs C1, C6–C9 and residue are not significant. To summarize, IMFs C2–C5 are identified as the meaningful components for short-term passenger flow series data. By calculating the instantaneous frequencies, IMFs C2–C5 represent the peak period pattern, semi-service period pattern, semi-daily pattern and daily pattern of short-term passenger flow. Readers are referred to Chen and Wei (2011) for further details of extracted IMF components.

3.4. Forecasting model design

In the third stage (BPN Stage), the collected data are divided into two independent data sets, a training data set and test data set, before performing the forecasting model. The training data set contains 1800 passenger flow data points from May 1 to May 24, constituting 77% of the original data, and the test data set contains 525 passenger flow data points from May 25 to May 31, or 23% of the original data. The training data set is fed into the BPN-based forecasting model to train the weights and biases of neurons. The test data set is used to validate the effectiveness of forecasting models. With these two data sets, the number of input variables is based on the rolling horizon, and the number of output variables is based on the forecasting horizon. The data are normalized before feeding to BPN. In this paper, BPN is run by using NeuralShell 2.

Since the 3-layer network architecture has been frequently used and suggested in the literature (e.g., Balestrassi et al., 2009), the hybrid EMD–BPN model consists of an input layer, a hidden layer and an output layer. These three layers are designed as follows.

3.4.1. Input layer

The elements of the input layer include passenger flow data, passenger patterns (i.e., IMFs components) and temporal factors. Passenger patterns involve the meaningful components (IMFs C2–C5) and the other components (IMFs C1, C6–C10) with respect to input feature identification. The meaningful components could be taken as inputs in individual or aggregated form because they are significantly correlated to the original data. The other components which possibly are the noise in the original data could also be considered as inputs in individual or aggregated form by considering the dimensionality of the

Table 1
The correlation coefficients and percentage power.

IMF	Mean[a]	Standard deviation[b]	Pearson correlation coefficient	Kendall correlation coefficient	Percentage power (%)
C1	1.025	163.495	.070[c]	−.001	1.76
C2	3.237	465.833	.165[c]	−.030[d]	20.50
C3	6.411	622.810	.545[c]	.376[c]	34.40
C4	13.532	327.334	.535[c]	.374[c]	28.30
C5	25.641	213.940	.406[c]	.345[c]	10.90
C6	57.143	83.391	.027	.027[d]	1.33
C7	124.070	71.116	.083[c]	.032[d]	1.70
C8	185.874	115.573	.058[c]	.024	1.02
C9	495.050	13.780	.007	.011	0.12
C10	1488.096	34.399	.013	.004	–
Residue					

[a] Mean of amplitude for each component.
[b] Standard deviation of amplitude for each component.
[c] Correlation is significant at 0.01 level (2-tailed).
[d] Correlation is significant at 0.05 level (2-tailed).

input vector and the computational requirement. The number of input nodes is based on the rolling step and the rolling horizon in the input layer. The rolling step is determined as a 15-min interval, and the rolling horizon is set to six rolling steps. Taking six rolling steps and one forecasting horizon as an example, the first six data points (passenger flow from 6:00 AM to 7:30 AM) of each day are used as input data. The 7th data point of each day is set to the target. Sequentially, the six data points of passenger flow from 6:15 AM to 7:45 AM are used as the next input data. The process continues until the last observation (passenger flow at 00:45 AM next day) becomes the target. Additionally, the temporal factors including the day of the week, the time period of the day, and weekday or weekend are taken into account. According to the above description, the number of input neurons ranges from 9 to 33 nodes in the input layer.

The BPN specific parameters include the activation function, the learning rate, and the momentum. In the previous studies, the logistic function is frequently used as the activation function (Zhang et al., 1998; Maier and Dandy, 2000). Therefore, the logistic function is used in this paper. For the learning rate and momentum, the setting is also chosen with reference to the previous studies (Zhang et al., 1998; Maier and Dandy, 2000; Tseng et al., 2002). The learning rate is set to 0.2, the momentum is set to 0.8, and initial weights are set to 0.3.

3.4.2. Hidden layer

The activation function, learning rate, momentum, and initial weights are set the same as those of the input layer. The number of hidden neurons is referred to as the default value in NeuroShell 2.

3.4.3. Output layer

The number of output neurons depends on the forecasting step and the forecasting horizon. As discussed above, prediction can be performed for one or more steps ahead. In this paper, forecasts up to four forecasting steps ahead are used, i.e., the prediction horizon extends to 60 min at most. Therefore, the number of output neurons is from one to four.

Four hybrid EMD–BPN models are investigated in which different IMFs are used as inputs; these hybrid models are EMD–BPN1, EMD–BPN2, EMD–BPN3, and EMD–BPN4. In EMD–BPN1, each extracted IMF of C1–C10 is individually used to build its own forecast model, and the predicted passenger flow is then obtained by summing all forecasts of C1–C10. In EMD–BPN2, only four meaningful IMFs, C2–C5, are individually used as inputs for building the forecasting model. In EMD–BPN3, besides IMFs C2–C5 being individually used as inputs, the other IMFs (C1, C6–C10) are also aggregated as an input variable. In EMD–BPN4, IMFs C2–C5 components are aggregated as an input variable and the other IMFs (C1, C6–C10) are also aggregated as another input variable for building the forecasting model. The number of output neurons for these hybrid models depends on the forecasting steps ahead. Taking the architecture of model EMD–BPN3 for one forecasting step ahead as an example, the three layers of EMD–BPN3 are designed as follows. The input layer includes the day of the week, the time period of the day, weekday or weekend, six-step rolling horizon of each one of IMFs C2–C5 (i.e., 24 nodes), and six-step rolling horizon of aggregation of IMFs C1, C6–C10 (i.e., six nodes). In such a design, thus, there are 33 input nodes. There is one output node because of one forecasting step ahead.

3.5. Performance measures

Several performance measures have been applied in the literature to investigate the viability of time series forecasting models. Two widely used performance indexes, Mean Absolute Percent Error (MAPE) and Variance of Absolute Percentage Error (VAPE), are applied as the performance measures for comparisons in the literature (e.g., Balestrassi et al., 2009; Zhang and Ye, 2008). MAPE measures the mean prediction accuracy and VAPE reflects the prediction stability. MAPE and VAPE are respectively defined in following equations:

$$MAPE(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100, \tag{9}$$

$$VAPE(\%) = Var\left(\frac{|y_i - \hat{y}_i|}{y_i}\right) \times 100, \tag{10}$$

where y_i is the observation value at the ith time interval, $\hat{y}_i$ is the predicted value at the ith time interval, and N is number of observations. Eqs. (9) and (10) respectively calculate the average and the variance of relative error between observation value and prediction value in all intervals. It is obvious that a prediction model with a larger MAPE is not as accurate as one with a smaller MAPE. Similarly, a prediction model with a larger VAPE is not as stable as one with a smaller VAPE.

4. Results and analysis

4.1. Significance of temporal factors

Four BPN-based models, namely BPN-TF0, BPN-TF1, BPN-TF2, and BPN-TF3, are investigated to verify the significance of temporal factors (i.e., the day of the week, the time period of the day, and weekday or weekend) in forecasting the passenger flow. These four BPN-based models primarily differ in the temporal factors considered as inputs in addition to the passenger

flow. BPN-TF0 considers only the passenger flow as inputs, and this model does not consider any temporal factor. In addition to the passenger flow, BPN-TF1 takes the day of week as inputs. Besides the inputs considered in BPN-TF1, BPN-TF2 further takes the time period of day as inputs. BPN-TF3 considers the three abovementioned temporal factors as inputs. The types of inputs considered in these four BPN-based models are summarized in Table 2.

The MAPE values of four BPN-based models for the test data set up to four steps ahead are shown in Table 3. As can be seen in Table 3, BPN-TF3 performs best because its average MAPE (6.90%) is lowest, whereas BPN-TF0 performs worst due to having the highest average MAPE (10.84%). Obviously, the forecasting accuracy of BPN-TF3, which considers all temporal factors, outperforms the other three BPN-based models. Additionally, from the VAPE values of four BPN-based models for the test data set as shown in Table 4, BPN-TF3 reveals the best forecasting stability because the average VAPE value of BPN-TF3 is the smallest (i.e., 0.37%). The results indicate that temporal factors influence the accuracy and stability of BPN-based forecasting models in predicting the passenger flow.

4.2. Experimental results of hybrid approach

Four hybrid forecasting models, namely EMD–BPN1, EMD–BPN2, EMD–BPN3, and EMD–BPN4, designed above are used to investigate the effectiveness of the hybrid EMD–BPN approach. These four hybrid models primarily differ in the extracted IMFs, which are used as inputs, and these models are summarized in Table 2.

Table 3 summarizes the results of MAPE values obtained by the four hybrid EMD–BPN models (i.e., all the three temporal factors are considered as inputs) from one step ahead to four steps ahead in the test data set. The average MAPE values of EMD–BPN3 (6.89%) and EMD–BPN4 (6.24%) for the test data set are lower. Comparing the results of EMD–BPN4 and BPN-TF3, the average accuracy of EMD–BPN4 improves 11% compared to that of BPN-TF3 for the test data set. Additionally, the multiple comparisons tests have been used to validate the forecasting results. For multiple comparisons tests, Demšar (2006) pointed out that the Wilcoxon signed rank test and Friedman test are robust non-parametric tests for comparing two or more classifiers. The Wilcoxon signed rank test is suitable for comparing two classifiers, and the Friedman test is suitable for comparing multiple classifiers (Demšar, 2006). In this study, the Wilcoxon signed rank test has been used to test whether the forecasting results of BPN-TF3 and EMD–BPN3 have a significant difference or not. The test results ($Z = -5.774$, $p = 7.73\text{E}{-9}$) demonstrate that the forecasting results of BPN-TF3 and EMD–BPN3 are significantly different at a 95% confidence level. Similarly, by using the Wilcoxon signed rank test, the forecasting results ($Z = -11.078$, $p = 1.58\text{E}{-28}$) of BPN-TF3 and EMD–BPN4 are also significantly different at a 95% confidence level. The forecasting results ($Z = -5.714$, $p = 1.106\text{E}{-8}$) of EMD–BPN3 and EMD–BPN4 are also significantly different at a 95% confidence level. The respective p-values remain below 0.05 for these comparing combinations. Further, the Friedman test has been used to validate whether the forecasting results of BPN-TF3, EMD–BPN3 and EMD–BPN4 have a significant difference or not. The test results ($\chi^2(2) = 81.619$, $p = 1.89\text{E}{-18}$) indicate that the forecasting results of these models have a significant difference at a 95% confidence level. This can confirm that EMD–BPN3 and EMD–BPN4 outperform BPN-TF3. Obviously, EMD–BPN3 and EMD–BPN4 can improve the accuracy of passenger flow prediction. This implies that individual or aggregated IMFs C2–C5 embedded into inputs can effectively enhance the forecasting performance. Furthermore, as shown in Table 3, the average MAPE value of EMD–BPN2, which considers only IMFs C2–C5 for the test data set, is 9.75%. IMFs C2–C5 are a subset of components of the original data, yet the EMD–BPN2 forecasting model based on these meaningful IMFs C2–C5 components possesses prediction ability quite near that of the other forecasting models based on all IMF components (e.g., EMD–BPN3 and EMD–BPN4). These results indicate that IMFs C2–C5 could be regarded as highly influential inputs in the hybrid forecasting models.

The MAPE value increases gradually from one step ahead to four steps ahead in these models. The results reveal that forecasting more steps ahead could produce higher MAPE. Notice that the higher MAPE implies lower forecasting accuracy. Furthermore, the daily prediction profile could be clearly illustrated with the daily average relative error (i.e., the average MAPE value for the same time period). Taking EMD–BPN4 as an example, as shown in Fig. 3, the daily average relative errors of EMD–BPN4 in morning peak periods (7:30–9:30) and afternoon peak periods (17:30–19:30) are lower than that of BPN-

Table 2
The types of inputs considered in BPN-based models and the hybrid EMD–BPN models.

Model	Types of inputs			
BPN-based models				
BPN-TF0	Passenger flow			
BPN-TF1	Passenger flow	The day of the week		
BPN-TF2	Passenger flow	The day of the week	The time period of the day	
BPN-TF3	Passenger flow	The day of the week	The time period of the day	Weekday or weekend
The hybrid EMD–BPN models				
EMD–BPN1	Temporal factors[a]	Individual IMFs C1–C10		
EMD–BPN2	Temporal factors[a]	Individual IMFs C2–C5		
EMD–BPN3	Temporal factors[a]	Individual IMFs C2–C5	Aggregated IMFs (C1, C6–C10)	
EMD–BPN4	Temporal factors[a]	Aggregated IMFs C2–C5	Aggregated IMFs (C1, C6–C10)	

[a] Temporal factors include the day of the week, the time period of the day, and weekday or weekend.

Table 3
The MAPE values of BPN-based models and the hybrid EMD–BPN models for test data set up to four steps ahead.

MAPE (%)	Number of forecasting steps ahead				
	1	2	3	4	Average
BPN-based models					
BPN-TF0	7.88	9.58	11.88	14.00	10.84
BPN-TF1	8.42	9.40	10.50	13.92	10.56
BPN-TF2	6.42	6.30	7.81	7.65	7.05
BPN-TF3	6.50	7.10	6.79	7.22	6.90
The hybrid EMD–BPN models					
EMD–BPN1	5.04	6.39	7.82	8.85	7.03
EMD–BPN2	9.15	10.01	9.40	10.46	9.75
EMD–BPN3	6.09	6.56	7.04	7.88	6.89
EMD–BPN4	5.26	5.82	6.70	7.18	6.24

Table 4
The VAPE values of BPN-based models and the hybrid EMD–BPN models for test data set up to four steps ahead.

VAPE (%)	Number of forecasting steps ahead				
	1	2	3	4	Average
BPN-based models					
BPN-TF0	0.48	1.34	2.84	5.85	2.63
BPN-TF1	0.61	0.84	1.57	4.57	1.90
BPN-TF2	0.31	0.26	0.49	0.46	0.38
BPN-TF3	0.30	0.39	0.35	0.45	0.37
The hybrid EMD–BPN models					
EMD–BPN1	0.21	0.45	0.71	0.90	0.57
EMD–BPN2	0.64	0.82	0.68	1.01	0.79
EMD–BPN3	0.28	0.31	0.38	0.49	0.37
EMD–BPN4	0.18	0.23	0.35	0.43	0.30

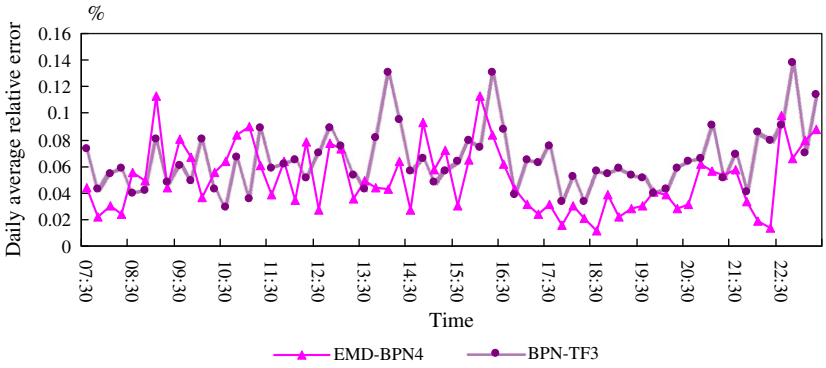

Fig. 3. Daily average relative error of EMD–BPN4 and BPN-TF3.

TF3 for one forecasting step ahead in the test data set. The results reveal that the prediction accuracies of morning and afternoon peak periods are higher for EMD–BPN4. Finally, from Fig. 4, the observation values and prediction values of the proposed EMD–BPN4 in one step forecasting ahead are quite close.

Table 4 summarizes the VAPE values of the four hybrid EMD–BPN models for the test data set. The average VAPE values of EMD–BPN4 and EMD–BPN3 are 0.30% and 0.37%, respectively. The results reveal that EMD–BPN4 and EMD–BPN3 are more stable than BPN-TF3 in the test data set. The results of VAPE and MAPE are consistent for the four hybrid models and BPN-TF3. EMD–BPN3 and EMD–BPN4 produce lower average MAPE and VAPE values. That is to say, the proposed hybrid EMD–BPN models (i.e., EMD–BPN4 and EMD–BPN3) have the higher prediction accuracy and stability among these hybrid models and BPN-TF3.

Regarding the training time of these forecasting models, taking one step forecasting ahead as an example, the average training time of BPN-TF3 is about 33 min, whereas the average training times of EMD–BPN1, EMD–BPN2, EMD–BPN3, and EMD–BPN4 are 33, 57, 61, and 38 min, respectively. That is to say, the training times of BPN-TF3 and EMD–BPN1 are shorter. The training time of EMD–BPN4 is relatively short because IMFs C2–C5 and IMFs (C1, C6–C10) are aggregated respectively as

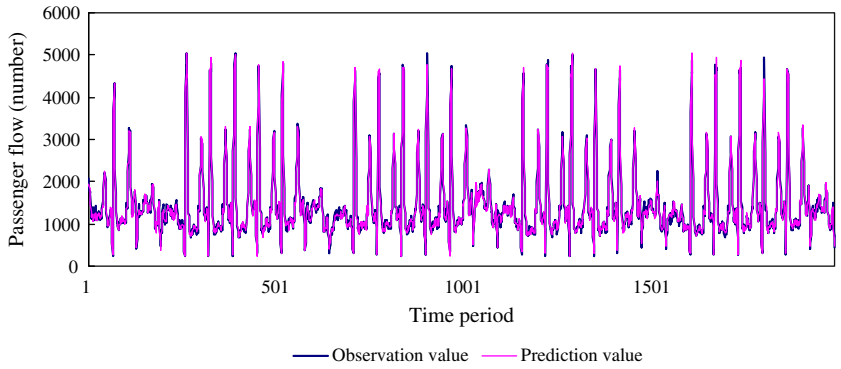

Fig. 4. The comparisons of observation value and prediction value of EMD–BPN4.

Table 5
Parameter estimation of SARIMA model.

Variables	Coefficient	Std. error	*t*-Statistic	Prob.
AR(1)	1.753096	0.045536	38.49898	0
AR(2)	−0.795622	0.038226	−20.81363	0
SAR(75)	−0.453026	0.011042	−41.02617	0
SAR(150)	−0.847745	0.010425	−81.31598	0
MA(1)	−0.745287	0.052987	−14.06539	0
MA(2)	0.213115	0.031823	6.696886	0
MA(3)	−0.167249	0.034936	−4.787233	0
SMA(75)	0.335899	0.019267	17.43421	0
SMA(150)	0.781867	0.018903	41.36144	0
R-squared	0.919795	Mean dependent var		−1.206612
Adjusted *R*-squared	0.919384	S.D. dependent var		570.4633
S.E. of regression	161.9711	Akaike info criterion		13.01842
Sum squared resid	41030968	Schwarz criterion		13.04909
Log likelihood	−10229.99	Hannan–Quinn criter.		13.02982
Durbin-Watson stat	1.984416			

inputs in the short-term forecasting model. The aggregation of IMFs reduces the dimensionality of input variables of EMD–BPN4. Additionally, the computation time of EMD is relatively short (within 10 s). The CPU time is based on an IBM compatible PC with an Intel Core 2 Duo CPU running at 1.6 GHz with 1.5 GB RAM. Observing the forecasting accuracy, stability and computation time of these models, EMD–BPN4 can be an appropriate model for forecasting the passenger flow. Although the computation times of these models are relatively long, a more powerful computer can be used to reduce the CPU time.

4.3. Comparisons with SARIMA model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model developed by Box and Jenkins (1976) is a conventional time series forecasting approach for analyzing time series with characteristics of seasonality and tendency. A number of studies have compared the performance of SARIMA with that of other methods such as exponential smoothing, and neural network (De Gooijer and Hyndman, 2006). In order to evaluate the forecasting accuracy and stability, this study compares EMD–BPN with SARIMA.

In SARIMA$(p, d, q)(P, D, Q)_S$, S is the seasonal length; p and P represent the orders of non-seasonal and seasonal autoregressive parameters; q and Q represent the orders of non-seasonal and seasonal moving average parameters; d and D represent the numbers of the regular and seasonal differences required. The SARIMA model uses three iterative steps, model identification process, parameter estimation and model diagnosis, to obtain the prediction results. The orders of the optimal SARIMA model are determined with validity information criteria such as Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and Hannan–Quinn information criterion (HQIC).

The passenger flow data set is divided into in-sample (i.e., training data set) and out-of-sample (i.e., test data set) in the SARIMA model. Similar to the training and test of EMD–BPN models, the in-sample data set contains 1800 (77%) passenger flow data points, and the out-of-sample data set contains 525 (23%) passenger flow data points. After performing the SARIMA process iteratively, the optimal SARIMA$(2, 0, 3)(2, 1, 2)_{75}$ model is obtained based on the previously mentioned validity information criteria as listed in Table 5. The model is presented as:

$$(1 - 1.753L + 0.796L^2)(1 + 0.453L^{75} + 0.848L^{150})(1 - L^{75})y_t = (1 - 0.745L + 0.213L^2 - 0.167L^3)(1 + 0.336L^{75} + 0.782L^{150})\varepsilon_t. \quad (11)$$

That is,

$$\begin{aligned} y_t = {} & 1.753y_{t-1} - 0.796y_{t-2} + 0.547y_{t-75} - 0.959y_{t-76} + 0.435y_{t-77} - 0.395y_{t-150} + 0.692y_{t-151} - 0.314y_{t-152} \\ & + 0.848y_{t-225} - 1.486y_{t-226} + 0.675y_{t-227} + \varepsilon_t - 0.745\varepsilon_{t-1} + 0.213\varepsilon_{t-2} - 0.167\varepsilon_{t-3} + 0.336\varepsilon_{t-75} - 0.25\varepsilon_{t-76} \\ & + 0.072\varepsilon_{t-77} - 0.056\varepsilon_{t-78} + 0.782\varepsilon_{t-150} - 0.582\varepsilon_{t-151} + 0.167\varepsilon_{t-152} - 0.13\varepsilon_{t-153}, \end{aligned}$$

where y_t denotes the observed passenger flow at time t, $t = 1, 2, \ldots, 2325$, L is the backward shift operator ($L^n y_t = y_{t-n}$), and ε_t is the estimated residual at time t. The residual term ε_t could be independently and identically distributed as normal random variables with a mean of zero and a constant variance.

The MAPE value of SARIMA for test data set is 30.39%, and the VAPE value of SARIMA is 7.58%. Obviously, the SARIMA model does not perform well. The forecasting accuracy and stability of the SARIMA model are the worst among all models investigated in this paper. The results indicate that the SARIMA model is not suitable for constructing the forecasting model for non-linear passenger flow due to the assumption of linearity.

5. Discussion

The results of short-term passenger flow forecasting can provide useful information for decision makers of metro systems. With the results of short-term passenger flow forecasting, decision makers can appropriately adjust the operation plans (e.g., headway and train dispatching), and activate the station passenger crowd regulation plan and emergency response plan provided that the predicted passenger flow is higher than the pre-determined threshold. The operation plans could be slightly modified based on the fluctuation of passenger flow to ensure that the required service level of the metro systems can be met. The minor modifications can be made in the operation plans such as train service plan, train schedule, timetable and crew schedule. For example, if decision makers identify passenger congestion in a specific period by observing the forecasts, they can arrange extra trains and adjust the train headway to relieve the passenger congestion and ensure the level of service in metro systems. In addition, decisions makers can assign the appropriate number of staff with a reasonable operation cost or add volunteers to deal with the passenger congestion.

In this paper, the hybrid EMD–BPN approach is developed to forecast the short-term passenger flow in metro systems. The meaningful IMFs C2–C5 components are extracted from passenger flow data by applying EMD, and they are taken as inputs for the forecasting model. Comparing the results of four BPN-based models, temporal factors (i.e., the day of the week, the time period of the day, and weekday or weekend) influence the performance of BPN-based forecasting models. In addition, by comparing the performance of four hybrid EMD–BPN models and the BPN-TF3 model, the proposed hybrid EMD–BPN models, in which IMFs C2–C5 are embedded individually or in aggregated form into inputs (i.e., EMD–BPN3 and EMD–BPN4), can generate better forecasting accuracy and stability in terms of MAPE and VAPE for one to four steps ahead. Furthermore, comparing the performance of various models, the proposed EMD–BPN approach outperforms BPN-based models and SARIMA.

Based on the above results, four points are worth noting. First, the EMD method adapts to be a data pre-processing technique. The proposed EMD–BPN model appropriately considers the characteristics of short-term passenger flow and enhances forecasting performance. Second, the extracted IMFs C2–C5 components represent the periodic patterns of passenger flow. The proposed EMD–BPN models, which consider the individual or aggregated IMFs C2–C5 components, can enhance forecasting accuracy and stability since the smoothed and regularized IMFs components could lead to improvement in passenger flow forecasting. Third, the proposed EMD–BPN model provides a systematic design guideline to include the unique patterns of the extracted IMFs as inputs and to enhance forecasting performance. By the systematic combinations of IMFs, the empirical results support the effectiveness and stability of the hybrid EMD–BPN model. For example, EMD–BPN4, in which IMFs C2–C5 are taken as an input variable in aggregated form, not only enhances the forecasting performance of the prediction model but also reduces dimensionality of the input vector. Considering both the forecasting performance and the dimensionality reduction of the input layer, EMD–BPN4 may be more suitable than the other hybrid EMD–BPN models. Finally, according to the forecasting performance of various forecasting models (BPN-based models, hybrid EMD–BPN models and SARIMA), BPN adapts to solve the non-linear complexity of passenger flow problem. Traditional time series methods such as the SARIMA model cannot deal with non-linearity well due to the assumption of linearity.

6. Conclusions

An accurate and stable passenger flow forecasting model can be applied to support transportation system management such as operation planning, revenue planning, and facility improvement. This paper proposes a hybrid EMD–BPN approach for short-term passenger flow forecasting with three stages. In the first stage (EMD Stage), by applying EMD, the IMF components are extracted from the short-term passenger flow data. The extracted components possess the unique characteristics of the passenger flow. In the second stage (Component Identification Stage), the meaningful components are identified by traditional statistical methods. In the third stage (BPN Stage), BPN is applied to perform the passenger flow forecasting. MAPE and VAPE are used to measure the forecasting accuracy and stability of forecasting models. The experimental results of four BPN-based models reveal that temporal factors (i.e., the day of the week, the time period of the day, and weekday or

weekend) can influence the accuracy and stability of forecasting models. In addition, the meaningful components (i.e., IMFs C2–C5 in this study) can also affect the performance of the hybrid EMD–BPN approach. Comparing various forecasting models (i.e., the BPN-based models, the hybrid EMD–BPN models, and SARIMA) with the passenger flow of a metro system, the proposed hybrid EMD–BPN approach outperforms the other models.

EMD can effectively reflect the characteristics of passenger flow and enhance forecasting accuracy, yet the problem of mode mixing may appear, since the time series data have intermittency. Mode mixing is defined as the fact that one IMF component may include both high frequency and low frequency, and a frequency may exist in different IMFs. In this paper, the passenger flow of the metro system may show mode mixing because the time series data may reveal intermittency. Mode mixing not only reduces the capability of EMD in identifying the true timescales of IMF components but also affects the forecasting ability of the hybrid EMD–BPN approach. To overcome such a problem, the ensemble empirical mode decomposition (EEMD) method developed by Wu and Huang (2004, 2008) can be used to extract IMF components. Developing a hybrid forecasting model by incorporating EEMD may lead to more accurate and stable forecasting results, and this would be valuable future work.

Acknowledgements

The authors are thankful to Taipei Rapid Transit Corporation for providing the necessary data. This work is partially supported by the National Science Council, Taiwan, ROC under Grant NSC 95-2221-E-009-361-MY3. The authors also wish to thank the editor and reviewers for their helpful comments, which have led the authors to consider more deeply the subject of short-term passenger flow forecasting.

References

Ahmed, M.S., Cook, A.R., 1979. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. Transportation Research Board Record 722, 1–9.
Alfa, A.S., 1986. A review of models for the temporal distribution of peak traffic demand. Transportation Research Part B 20 (6), 491–499.
Balestrassi, P.P., Popova, E., Paiva, A.P., Marangon Lima, J.W., 2009. Design of experiments on neural network's training for nonlinear time series forecasting. Neurocomputing 72 (4–6), 1160–1178.
Bar-Gera, H., Boyce, D., 2003. Origin-based algorithms for combined travel forecasting models. Transportation Research Part B 37 (5), 405–422.
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.
Brooks, C., 2008. Introductory Econometrics for Finance. Cambridge University Press Inc..
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial Intelligence 97 (1–2), 245–271.
Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., Lee, D.H., 2009. AADT prediction using support vector regression with data-dependent parameters. Expert Systems with Applications 36 (2), 2979–2986.
Chen, C.-F., Chang, Y.-H., Chang, Y.-W., 2009. Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. Transportmetrica 5 (2), 125–140.
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, M.-C., Wei, Y., 2011. Exploring time variants for short-term passenger flow. Journal of Transport Geography 19 (4), 488–498.
Clark, S., 2003. Traffic prediction using multivariate nonparametric regression. Journal of Transportation Engineering 129 (2), 161–168.
Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 2, 1–30.
De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. International Journal of Forecasting 22 (3), 443–473.
Dia, H., 2001. An object-oriented neural network approach to short-term traffic forecasting. European Journal of Operational Research 131, 253–261.
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.
Dougherty, M., 1995. A review of neural networks applied to transport. Transport Research Part C 3 (4), 247–260.
Faraway, J., Chatfield, C., 1998. Time series forecasting with neural networks: a comparative study using the airline data. Applied Statistics 47 (2), 231–250.
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.
Hamed, M.M., Al-Masaeid, H.R., Bani Said, Z.M., 1995. Short-term prediction of traffic volume in urban arterials. Journal of Transportation Engineering 121 (3), 249–254.
Hansen, J.V., McDoald, J.B., Nelson, R.D., 1999. Time series prediction with genetic-algorithms designed neural networks: an empirical comparison with modern statistical models. Journal of Computational Intelligence 15 (3), 171–183.
Highway Capacity Manual, 2000. Transportation Research Board, National Research Council, Washington, DC.
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 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.
Ince, H., Trafalis, T.B., 2006. A hybrid model for exchange rate prediction. Decision Support Systems 42 (2), 1054–1062.
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.
Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artificial Intelligence 97 (1–2), 273–324.
Kumar, S., 2005. Neural Networks: A Classroom Approach. McGraw-Hill Education.
Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transportation Research Board 1678, 179–188.
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.
Liang, H.L., Bressler, S.L., Desimone, R., Fries, P., 2005. Empirical mode decomposition: a method for analyzing neural data. Neurocomputing (65–66), 801–807.
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.
Lim, C., McAleer, M., 2002. Time series forecasts of international travel demand for Australia. Tourism Management 23 (4), 389–396.
Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environmental Modelling and Software 15 (1), 101–124.
Park, B., Messer, C.J., Urbanik, T., 1998. Short-term freeway traffic volume forecasting using radial basis function neural network. Annual Meeting of the Transportation Research Record 1651, 39–47.
Rojas, I., Valenzuela, O., Rojas, F., Guillen, A., Herrera, L.J., Pomares, H., Marquez, L., Pasadas, M., 2008. Soft-computing techniques and ARMA model for time series prediction. Neurocomputing 71, 519–537.
Shen, Z.-X., Huang, X.-Y., Ma, X.-X., 2008. An intelligent fault diagnosis method based on empirical mode decomposition and support vector machine. Proceedings of the Third International Conference on Convergence and Hybrid Information Technology 1 (11–13), 865–869.
Smith, B.L., Demetsky, M.J., 1997. Traffic flow forecasting: comparison of modeling approaches. Journal of Transportation Engineering 123 (4), 261–266.
Smith, B.L., Williams, B.M., Keith Oswald, R., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C 10 (4), 303–321.
Tan, M.-C., Wong, S.C., Xu, J.-M., Guan, Z.-R., Zhang, P., 2009. An aggregation approach to short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 10 (1), 60–69.
Tang, Y.F., Lam, William H.K., Ng, Pan L.P., 2003. Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong. Journal of Transportation Engineering 129 (3), 271–277.
Tsai, T.-H., Lee, C.-K., Wei, C.-H., 2009. Neural network based temporal feature models for short-term railway passenger demand forecasting. Expert Systems with Applications 36 (2), 3728–3736.
Tseng, F.-M., Yu, H.-C., Tzeng, G.-H., 2002. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change 69 (1), 71–87.
Van Arem, B., Kirby, H.R., Der, Van., Vlist, M.J.M., Whittaker, J.C., 1997. Recent advances and applications in the field of short-term traffic forecasting. International Journal of Forecasting 13 (1), 1–12.
Vanajaksi, L., Rilett, L.R., 2007. Support vector machine technique for the short term prediction of travel time. IEEE Intelligent Vehicles Symposium, pp. 600–605.
Van Lint, J.W.C., Hoogendoorn, S.P., Van Zuylen, H.J., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Research Part C 13 (5–6), 347–369.
Veltcheva, A.D., Soares, C.G., 2004. Identification of the components of wave spectra by the Hilbert Huang transform method. Applied Ocean Research 26 (1–2), 1–12.
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.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transportation Research Part C 13 (3), 211–234.
Vythoulkas, P.C., 1993. Alternative approaches to short term forecasting for use in driver information systems. Proceedings of the 12th International Symposium on Transportation and Traffic Theory, Berkeley, CA, pp. 485–506.
Wang, Y., Papageorgiou, M., 2007. Real-time freeway traffic state estimation based on extend Kalman filter: a case study. Transportation Science 42 (2), 167–181.
Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. Journal of Transportation Engineering 129 (6), 664–672.
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, Z., Huang, N.E., 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proceedings of Royal Society of London. A 460, 1597–1611.
Wu, Z., Huang, N.E., 2008. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis 1, 1–41.
Zeng, D., Xu, J., Gu, J., Liu, L., Xu, G., 2008. Short term traffic flow prediction using hybrid ARIMA and ANN models. Proceedings of Workshop on Power Electronics and Intelligent Transportation System, pp. 621–625.
Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14 (1), 35–62.
Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.
Zhang, H.M., 2000. Recursive prediction of traffic conditions with neural network models. Journal of Transportation Engineering 126 (6), 472–481.
Zhang, G.P., Qi, M., 2005. Neural network forecasting for seasonal and trend time series. European Journal of Operational Research 160 (2), 501–514.
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.
Zhang, Y., Ye, Z., 2008. Short-term traffic flow forecasting using fuzzy logic system methods. Journal of Intelligent Transportation Systems 12 (3), 102–112.
Zheng, S., Tian, J.W., Liu, J., Xiong, C.Y., 2004. Novel algorithm for image interpolation. Optical Engineering 43 (4), 856–865.
Zheng, W.Z., Lee, D.H., Shi, Q.X., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. Journal of Transportation Engineering 132 (2), 114–121.
Zhou, Q., Lu, H.-P., Wei, X., 2007. New travel demand models with back-propagation network, natural computation. Proceedings of the Third International Conference on Natural Computation, Haikou, China, pp. 311–317.
Zhu, Z., Sun, Y., Li, H., 2007. Hybrid of EMD and SVMs for short-term load forecasting. Proceedings of IEEE International Conference on Control and Automation, pp. 1044–1047.