A Hybrid Method for Short-term Freeway Travel Time Prediction Based on Wavelet Neural Network and Markov Chain

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Civil Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjce-2017-0231.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>15-Jul-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Yang, Hang ; Tongji University</td>
</tr>
<tr>
<td></td>
<td>Zou, Yajie; Tongji University</td>
</tr>
<tr>
<td></td>
<td>Wang, Zhongyu ; Shanghai Maritime University</td>
</tr>
<tr>
<td></td>
<td>Wu, Bing; Tongji University, Key Laboratory of Road and Traffic Engineering of Ministry of Education</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?:</td>
<td>N/A</td>
</tr>
<tr>
<td>Keyword:</td>
<td></td>
</tr>
</tbody>
</table>
A Hybrid Method for Short-term Freeway Travel Time Prediction Based on Wavelet Neural Network and Markov Chain

Hang Yang
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University
No. 4800 Cao’an Road, Shanghai 201804, China
Phone: +86 21 69588152
Email: 1991hanghang@tongji.edu.cn

Yajie Zou, Ph. D.*
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University
No. 4800 Cao’an Road, Shanghai 201804, China
Phone: +86 21 69588152
Email: zouyajie@gmail.com

Zhongyu Wang, Ph. D.
College of Transport and Communications, Shanghai Maritime University
No. 1550 Haigang Avenue, Shanghai 201306, China
Phone: +86 21 38282300
Email: wangzy@shmtu.edu.cn

Bing Wu, Ph. D.
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University
No. 4800 Cao’an Road, Shanghai 201804, China
Phone: +86 21 69588152
Email: wubing@tongji.edu.cn

(* Corresponding Author)
ABSTRACT

Short-term travel time prediction is an essential input to Intelligent Transportation Systems (ITS). Timely and accurate traffic forecasting is necessary for Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS). Despite several short-term travel time prediction approaches have been proposed in the past decade, especially for hybrid models which consist of machine learning models and statistical models, few studies focus on the over-fitting problem brought by hybrid models. The over-fitting problem deteriorates the prediction accuracy especially during peak hours. This paper proposes a hybrid model which embraces Wavelet Neural Network, Markov Chain and the volatility model (WNN-MAR-VOA) for short-term travel time prediction in a freeway system. The purpose of this paper is to provide deeper insights into underlining dynamic traffic patterns and to improve the prediction accuracy and robustness. This method takes periodical analysis, error correction and noise extraction into consideration and improve the forecasting performance in peak hours. The proposed methodology predicts travel time by decomposing travel time data into three components: a periodic trend presented by a modified Wavelet Neural Network (WNN), a residual part modeled by Markov Chain, and the volatility part estimated by the modified generalized autoregressive conditional heteroscedasticity (GJR-GARCH) model. Forecasting performance is investigated with freeway travel time data from Houston, Texas and examined by three measures: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The results show that the travel times predicted by the WNN-MAR-VOA method are robust and accurate. Meanwhile, the proposed method is able to capture the underlying periodic characteristics and volatility nature of travel time data.
1 **KEY WORDS:** Travel time, Short-term prediction, Wavelet Neural Network, Multi-regime based model, Markov Chain, Volatility analysis
1. INTRODUCTION

Short-term travel time prediction is a fundamental input for Intelligent Transportation Systems (ITS). Traffic congestion is a thorny problem in many cities, especially in those megacities because of the sophisticated traffic network. Travel time is an important measure of roadway traffic condition and thus usually regarded as the optimization objective when addressing the roadway network congestion, which highlights the importance of short-term travel time prediction.

In this study, we introduced a WNN-MAR-VOA method, which embraces Wavelet Neural Network, Markov Chain and GJR-GARCH model to obtain accurate short-term travel time prediction of freeway. Wavelet Neural Network is used to present the periodic pattern, Markov Chain is utilized to correct the residual part of initial prediction, and GJR-GARCH model is employed to predict the volatility part. The proposed method is aimed at providing some insights into the basic structures of the raw data and accurately forecast the highway travel time quantitatively in the near future. The main innovations of the proposed hybrid method list as follows:

- Different from the most existing hybrid models which consist of periodical analysis and the residual part, the proposed hybrid method add residual-correction and volatility-analysis to the hybrid model.
- Modifying the shift factor and the constriction factor of wavelet transform to accelerate the convergence rate of learning procedure and avoid local optimal solution as much as possible.
- Extracting white noise, which is brought by over-fitting problem after the combination of statistical model and machine learning model. As a consequence, the accuracy and robustness of forecasting performance are improved, especially in peak hours.
The error-based Markov Chain method is relatively convenient to classify the Markov states and more flexible in dealing with other travel time data on short-term traffic forecasting in other hybrid model.

2. LITERATURE REVIEW

A number of short-term travel time prediction approaches have been developed in the past decade, including: regression methods (Sun et al. 2003), time series methods (Ishak and Al-Deek 2002), Kalman filtering methods (Nanthawichit et al. 2003), supporting vector machine models (Cheng et al. 2017; Yu et al. 2011), neural networks (Yildirimoglu and Geroliminis 2013), fuzzy logic system (D’Angelo et al. 1999; Tang et al. 2017), and other machine learning models (Hofleitner et al. 2012; Polson and Sokolov 2017; Van Lint et al. 2005; Zhang et al. 2017b). More details about travel time prediction models were summarized in some recent review papers, such as (Vlahogianni et al. 2014) and (Van Lint and Van Hinsbergen 2012).

Short-term travel time prediction is a comprehensive issue. It not only exhibits periodic variations, but also presents residual patterns and noise. Vlahogianni et al. (2006) concluded that exogenous factors such as traffic incidents, weather, and roadway conditions result in these uncertainties. Detecting the regular characteristics in travel time data, as well as extracting deterministic patterns from raw data is of great necessity. Meanwhile, it is usually difficult for traditional forecasting methodologies to capture the sudden drops or increases of travel time (Kuchipudi and Chien 2003), learning and modeling the stochastic traffic behaviors would potentially improve forecasting accuracy. To statistically and explicitly analyze the nature of travel time data, decomposition of the raw data into different components and analyzing them with suitable methods may prove effective (Dimitriou et al. 2008; Kuchipudi and Chien 2003;
Urban traffic-flow theory considers the periodic traffic pattern, over a 24-h period for weekdays, or over a week for both weekdays and weekends (Zhang et al. 2017a). Periodical pattern were observed in many transport predictions. Zou et al. (2015) used a trigonometric regression function to capture the periodic component. Zhang et al. (2014) utilized spectral analysis to exhibits repetitive behaviors of traffic flow data over time. Linear regression, with the stepwise-variable-selection method and more advanced tree-based methods were proposed in Kwon’s Research (Kwon et al. 2000), and day-to-day trends were observed in numerical results. Yu et al. (2011) applied Fourier transform and concluded that smaller cycles exist during the obvious 24-h cycle. Jiang and Adeli (2005) used wavelet analysis and proposed a time-delay model to reveal possible common cyclical components along a two-way freeway.

However, Vlahogianni et al. (2014) pointed out that how to appropriately define and retrieve temporal patterns of travel time data is important, since utilizing this information may greatly boost prediction performance. Detrending based travel time prediction is frequently used nowadays. The key hypothesis of detrending method is separating a certain trend of time series from the remaining fluctuations (Li et al. 2014). Multi-regime based model is the widely used method in detrending based predictions (Cetin and Comert 2006; Kamarianakis et al. 2012). Markov Chain proves valid and practical in multi-regime based models. D'Angelo et al. (1999) used Markov Chain to predict recurrent congestion on freeway corridors. Compared with the single-variable model, Markov Chain model represented obvious superiority on predicting recurrent congestion with the combination of periodical analysis. Yeon et al. (2008) proposed a hybrid discrete Markov Chain to estimate travel time on freeway on the basis of the cyclicality in the daily and weekly travel time data. However, all the classification of Markov status in these
Yang et al.

researches included three parts: uncongested state, semi-congested state and congested state. The difficulty in determining the semi-congested state may influence the prediction accuracy.

Most traditional hybrid models only concentrate on periodic trend and residual part of travel time data. Recently, some researchers began to make use of statistical volatility models in dealing with the over-fitting problem. The main approach is exploring the uncertainty and variability to alleviate the influence brought by noise. Variance is constant in the assumptions of traditional methods, the volatility models forecast time-dependent variances. Engle (1982) introduced ARCH model, which relaxes former assumption and assumes that the changes in volatility of time series is predictable. Although the structure of the ARCH model is simple and easy to understand, describing the volatility process often requires a number of parameters. Bollerslev (1986) proposed GARCH model, which is the extension form of ARCH model, allowing for a much more flexible structure. Some successful implementation of standard GARCH models is demonstrated in the recent studies. Meanwhile, extensions of the GARCH model have also been utilized in modeling traffic volatilities (Karlaftis and Vlahogianni 2009; Zhang et al. 2013). These studies indicated that the apparent changes in the volatility of travel time data brought by noise are predictable and may result from a specific type of non-linear function.

The remainder of the paper is organized as follows. In section 3, we introduce the methodology, which consists of Wavelet Neural Network, Markov Chain and volatility model. In section 4, we describe data collection site and analyze the periodical pattern. Then three parts of hybrid model is clarified respectively. In section 5, we utilize three measures to evaluate the prediction performance of WNN-MAR-VOA method and compare the proposed model with the other five models. In section 6, we provide conclusions and some suggestions for future works.
3. METHODOLOGY

Traditional travel time forecasting methods take constant variance into consideration and forecast the current value as a function of its past values. Zhang et al. (2014) concluded that accurate and reliable travel time forecasting relies on better understanding of the overall underlining components. Therefore, based on the discussion in the introduction, we assume that travel time consists of three components: a periodic trend, a residual component, and the volatility part. The structure of the proposed model is the sum of the periodic trend, the residual part and the variance:

$$x_t = p_t + r_t + v_t$$  \hspace{1cm} (1)

where $x_t$ is the observed travel time index during time interval $t$, $p_t$ is the periodic trend expressed by Mother Wavelet Function (MWF), which proved effective in capturing the periodical patterns of raw data by Zou et al. (2014):

$$y = \cosine(1.75x)e^{-x^2/2}$$  \hspace{1cm} (2)

MWF describes the periodic patterns of travel time in different time intervals, $r_t$ is the correction of the residual part of the data which is presented by Markov Chain after removing the periodic components, $v_t$ is the variance of $x_t$ calculated by GJR-GARCH model.

Wavelet Neural Network model provides the essential framework to highlight the periodic patterns presented in the travel time data, and learning procedure of MWF on the periodic cosines reproduces the cyclic patterns. By removing these periodicities in the initial data, the residual parts of the traffic counts could be corrected by multi-regime based model. The Markov Chain and volatility models are introduced to correct the residual part of the data.

3.1 Wavelet Neural Network Model
Wavelet Neural Network (WNN) is the modified version of BP Neural Network (BPNN). In this paper, an improved wavelet transform is used to accelerate the convergence rate of learning procedure and avoid local optimal solution as much as possible. The topological structure of WNN is showed in Figure 1. $X_1, X_2, \ldots, X_k$ are the input parameters, $P_1, P_2, \ldots P_m$ are predicted values, $W_{ij}$ and $W_{jk}$ are the weights of Neural Network, $Y(x)$ is the wavelet basis function.

![Topological structure of Wavelet Neural Network.](https://mc06.manuscriptcentral.com/cjce-pubs)

As a machine learning approach to detect regular cyclical patterns or periodic variations, WNN is adopted to investigate the cyclicality in the daily and weekly travel time data (Park 2002). Based on the concept of Wavelet Transform and fourier series, travel time $x_k$ collected during different time interval $t$ forms a time series that often exhibits some repetitive behaviors over time. These repetitive behaviors resulted from the mixtures of periodic series with multiple frequencies and amplitudes. The dynamic system function represented by Eq. (3) is approximated using the wavelet transform function and wavelet coefficient expressed in a general form as Eq. (4). More details about Wavelet Transform could be found in Antonini et al. (1992).
In Eq. (3), where $x^t$ and $y^{t-1}$ represent the current travel time status vector and feedback travel time input vector at time $t$ (representing the discrete time series values up to time $t$), respectively; $f(.)$ and $g(.)$ are scalar nonlinear mapping functions; and $e(t)$ is error between the actual and predicted values of future travel time output $y(t)$. In Eq. (4), $M$ is number of wavelets to be discussed in the next section; $\bar{f}(y)$ represents the approximation of $f(y)$, $w_i$ represents the discrete wavelet transform coefficient; and $\psi_{a,b}(y)$ is two-dimensional wavelet expansion functions obtained from the basic wavelet function $\phi(t)$ by simple scaling and translation which is showed as follows:

$$\psi_{a,b}(y) = \frac{1}{\sqrt{|a|}} \phi \left( \frac{y - b}{a} \right)$$

in which the parameter vectors $a \neq 0$ and $b$ denotes the frequency (or scale) and the time (or space) location vectors corresponding to the multidimensional input vector $y$, respectively, and $\mathbb{R}$ is set of real numbers. The notation $L^2(\mathbb{R})$ represents the square summable constructed state space vectors.

Eq. (3) and Eq. (4) are used to modify the shift factor and the constriction factor of MWF. And MWF is chosen as wavelet basis function in the hidden layer which is listed in Eq. (2). The transfer function between input layer and hidden layer is set as follows which has been verified effective in Jiang’s research (Jiang and Adeli 2005):
in which $h(j)$ is the output of $j^{th}$ node in hidden layer, $h_j$ is MWF, $w_{ij}$ is the connection-weight between input layer and hidden layer, $b_j$ is the shift factor of $h(j)$, and $a_j$ is the constriction factor of $h(j)$. Meanwhile, transfer function between hidden layer and output layer is presented in Eq. (7).

$$p(k) = \sum_{i=1}^{l} w_{ik} p(i)$$

in which $w_{ik}$ is the weight between hidden layer and output layer, $p(k)$ is the output of $k^{th}$ node in output layer, $p(i)$ is the output of $i^{th}$ node in hidden layer, $l$ is the number of nodes in hidden layer, and $m$ is the number of nodes in output layer. Then $p_t$ is obtained by calculating $p(k)$ from WNN in each time interval $t$. The main steps of WNN are listed as follows:

1. Calculate the prediction error of WNN

$$e = \sum_{k=1}^{m} y_n(k) - y(k)$$

$y_n(k)$ is the expected output, $y(k)$ is the prediction outcome of WNN.

2. Correct the weights and WMF based on the prediction error

$$w_{n,k}^{(i+1)} = w_{n,k} + \Delta w_{n,k}^{(i+1)}$$

$$a_k^{(i+1)} = a_k + \Delta a_k^{(i+1)}$$

$$b_k^{(i+1)} = b_k + \Delta b_k^{(i+1)}$$
\[ \Delta w_{n,k}^{(i+1)} = -\eta \frac{\partial e}{\partial \Delta w_{n,k}^{(i)}} \]

\[ \Delta d_k^{(i+1)} = -\eta \frac{\partial e}{\partial \Delta d_k^{(i)}} \]

\[ \Delta b_k^{(i+1)} = -\eta \frac{\partial e}{\partial \Delta b_k^{(i)}} \]  

(9)

\( \eta \) is the learning rate of WNN.

3.2 Markov Chain Model

As mentioned in literature review, Multi-regime based predictions is of great necessity in short-term travel time prediction. Markov Chain model is used to classify the different “regime” of the uncertain part of original travel time data. The main function of Markov Chain model is correcting the residual part of the data after removing the periodic components by WNN, and aiming to make the predicted value more closed to the actual travel time. Zhang (2000) demonstrated that predictive error of neural networks in time \( t \) is highly related to that in time \( t-1 \), while has little correlation with any other former times, especially for traffic flow characteristics in peak hours of the weekdays. In this paper, the percentage error of initial prediction made by WNN is modeled as a Markov Chain and classified into several states. Markov Chain model is utilized to minimize the holistic prediction error, state-classification is simplified and convergence rate of Markov Chain is accelerated because there is no need to identify congested states of traffic condition.

The foundation for the construction of Markov Chain is non-after effect; that is, an object's state at the present phase is only related to the state at the previous phase, while has nothing to do with status at any other former phases. Furthermore, forecast result correction needs to be able to satisfy requirements of non-after effect and in accordance with the steady-
Yang et al. state consistency check (Ramezani and Geroliminis 2012). The method about how to classify the Markov state in a more elaborate way is not included in this paper. A binary-decision-diagram-based method was used to determine the state of percentage error margin in this paper. The detailed information of the binary-decision-diagram-based method can be seen in Xie and Beerel (1998). According to the percentage error margin (the percentage that absolute error takes in actual value) in WNN, Markov’s status area can be classified into several states.

<table>
<thead>
<tr>
<th>State of Percentage Error Margin</th>
<th>Interval Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(−∞,−7%]</td>
</tr>
<tr>
<td>B</td>
<td>(−7%,−5%]</td>
</tr>
<tr>
<td>C</td>
<td>(−5%,−3%]</td>
</tr>
<tr>
<td>D</td>
<td>(−3%,−1%]</td>
</tr>
<tr>
<td>E</td>
<td>(−1%,1%]</td>
</tr>
<tr>
<td>F</td>
<td>(1%,3%]</td>
</tr>
<tr>
<td>G</td>
<td>(3%,5%]</td>
</tr>
<tr>
<td>H</td>
<td>(5%,7%]</td>
</tr>
<tr>
<td>I</td>
<td>(7%,+∞)</td>
</tr>
</tbody>
</table>

During state $E_i$ transferring to $E_j$ through a single step, the probability taking place is called one-step transfer probability. We form one-step transfer probability matrix $p=[p_{ij}]_{N \times N}$ is, $p_{ij} \leq 1$. $p_{ij} = m_{ij}/m_i$, $m_i$ means how many times that state $E_i$ appears, $m_{ij}$ denotes that how many times the State transfers from $E_i$ to $E_j$. The transition is a change of state and the one-step transition matrix (e.g., $n = 1$, where $n$ is the number of steps) shows the changing rate from state $i$ to state $j$ as shown below:
in which \( p_{ij} \) represents the transition rate from state \( i \) to state \( j \), which can be expressed in conditional probability as follows:

\[
p_{ij} = P\{X(t+1) = j | X(t) = i\} \quad \text{for } i, j = 1, 2, ..., m, \text{ and } t = 1, 2, 3, ...
\]

in which \( m \) is the total number of possible states.

If the system’s initial state vector is \( S_0 \), then \( S_0 = (S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9) \), among which, \( S_i \) (\( i = 1, 2, ..., 9 \)) is the probability in the state of \( i \); if the state vector is \( S_k \) after \( k \) steps of transfer, we obtain

\[
S_k = S_{k-1} \times p = S_0 \times p^k
\]

which is based on C-K Equation (Medury and Madanat 2013). In every step, the specific range of percentage error in the state of \( i \) with the largest probability, \( k_i \) is chosen to replace the former one which is calculated by initial WNN. So the expression of \( r_i \) is:

\[
r_i = \hat{r}_i (1 - k_i)
\]

in which \( \hat{r}_i \) is the actual value of observed travel time data, \( r_i \) is the residual part of the initial data, \( k_i \) equals to the value of the upper bound of the selected range.

### 3.3 Volatility Model

As mentioned at the beginning of this section, the content of the proposed hybrid method is the sum of the periodic trend, the residual part and the variance. Most traditional models only concentrate on the first two parts. The variance part is considered to simply satisfy the white
noise properties in the assumptions of those models. Conditional heteroscedastic models relax the assumption and treat the variance part $e_t$ as conditionally heteroscedastic. So the expression of $e_t$ is:

$$e_t = z_t \times \sqrt{h_t}$$

(15)

where $h_t = E(e_t^2 | F_{t-1})$, and $\{z_t\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, which can be assumed to follow a standard normal, standard student-t, or generalized error distribution. Eq. (15) indicates that the conditional distribution of $e_t$ is independent. And it is distributed with zero mean and a new variance of $h_t$.

Volatility model is concerned with time-evolution of the conditional variance of travel time data, so addressing the conditional variance of $e_t$ in different ways leads to various heteroscedastic models.

Bollerslev (1986) proposed a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model that allows for a flexible lag structure. Although GARCH model is a well-specified method, this model ignores the empirically important asymmetric effect in travel time data. It is because that travelers’ response to sudden decrease or increase of travel time could be largely different in transportation system. This volatility may lead to the sudden change of driving behavior. Glosten et al. (1993) proposed the GJR-GARCH model with a conditional variance formulation, which successfully captured asymmetric response. Zhang et al. (2014) pointed out that GJR-GARCH allows the conditional variance to respond differently to the past negative and positive innovations. The GJR-GARCH approach is listed below:

$$h_t = a_0 + \sum_{i=1}^{r} \beta_i h_{t-i} + \sum_{j=1}^{u} (\alpha_j e_{t-j}^2 + \gamma_j e_{t-j}^2 I_{t-j})$$

(16)
in which $a_0$ is a constant, $\beta_j, \alpha_j$ and $\gamma_j$ are non-negative parameters. $I_{t-j}$ is a binary variable for $e_t$.

$$I_{t-j} = \begin{cases} 1 & \text{if } e_{t-j} < 0 \\ 0 & \text{if } e_{t-j} \geq 0 \end{cases}$$  \hspace{1cm} (17)

Eq. (15) highlights the difference of $e_{t-j}$ on the conditional variance according to the sign whether it is positive or not. And variance $I_{t-j}$ is an indicator as shown in Eq. (17). The two equations above present that $h_t$ responds differently to positive and negative $e_{t-j}$. It is clear that if all of $\gamma_j$ is zero, this process converts to a standard GARCH(r, u) process. As a consequence, this volatility model is able to capture both volatility clustering and asymmetry in travel time data.

4. NUMERICAL EXPERIMENTS

4.1 Test Site and Data Collection

The travel time data used in this study were collected on a westbound segment of the US-290 freeway section (an east-west US freeway) between I-610 and FM-1960 in Houston, Texas. This segment has an obvious imbalanced traffic flow pattern. The imbalanced characteristic is used to test the validity of volatility part of the hybrid model, and it is also effective when testing the robustness of multi-regime based model. The westbound traffic during the morning peak is less heavy compared with the eastbound traffic due to the directional traffic pattern. The length of the selected freeway section is approximately 12 miles and the free-flow travel time along this stretch is about 13 minutes. There are 6 AVI readers along this stretch, so the segment is divided into 5 links. The distances between each two detectors are 0.8, 2.6, 3.0, 1.5, 4.1 miles. The AVI readers record vehicles’ information, which contain toll tags, ID and timestamps when those
vehicles passed through. By matching their IDs from a pair of AVI readers, the travel time of the
link enclosed by this pair of AVI readers is the difference in the timestamps (Songchitruxsa et al.
2009). Occasional missing data are replaced by the moving average technique.

Travel time data were collected every 30s in 24 hours per day. The sample data are from
January, 2008 to August, 2008. Individual travel times were aggregated into 5-minute intervals
for each link. So there are 288 data points representing five-minute travel time data for each day.
This study is interested in the prediction of travel time data on weekdays (Tuesday-Thursday)
from January to August, 2008. Figure 2(a) shows the locations of the 6 AVI stations as well as 5
resulting links along the US-290. Since the westbound links experienced heavy congestion
during the afternoon peak hour, only the afternoon peak travel time data are used in the study to
evaluate the prediction performance. In this paper, the prediction performance of new model and
the other models is evaluated using the travel time data on link D. The other four links are given
dynamic weight and used as four row vector of input variables in the later training period of
WNN. Thus, our hybrid model considers the spatial information from neighboring links. Travel
time information from the other four links (A, B, C and E) can be useful indicator of an incident
and thus improve prediction of travel time under non-recurrent congestion condition.
FIGURE 2 Tests site and data patterns: (a) Study links along the US-290; (b) Periodical component of travel time during the weekdays.

4.2 Wavelet Neural Network Analysis

From Figure 2(b) we could see that travel time data have similar patterns over different days and different times of the day. This approximate trend may affect the traffic prediction and should be
removed to improve the forecasting results.

In this study, there is no automatic way of determining the training period. To make the choice and comparison of training period, varying training periods of 1, 2, 3, 4, 5 months is considered to compare the training performance, that is, the travel time data in March (10 days), April (11 days), May (13 days), June (14 days) and July (15 days) are used as the possible training periods. The travel time data of the five links in August (12 days) are used as the test set and this test set is used in evaluating the performance of all the possible training periods.

Although increasing the training period beyond 20 days brings slight improvement to the training performance, considering the sufficient travel time data in this study, we chose a sliding 35-day training period of time for WNN-MAR-VOA model. For example, the WNN-MAR-VOA models for travel time prediction on weekdays (3:30 PM—7:30 PM) on August 28, 2008 were trained on the 35 most recent days (i.e., May (1 days), June (8 days), July (15 days) and August(11 days)). Here, 12 weekdays (Tuesday to Thursday) in August, 2008 is chosen to test the proposed hybrid model and the other traditional models. Moreover, according to training time and approximate error, 20 is selected as the number of nodes for the hidden layer; and the network approximate error is 0.00005. Thus, we obtain the simulation and the predicted values of WNN.

4.3 Correction of the residual part

The residual part of the travel time data still shows various patterns, which may result from propagation of error at different time t. The expression of percentage error is presented in Eq. (18):
in which $P_t$ and $\hat{P}_t$ are the observed and predicted travel times at time $t$. According to the predicted error in WNN model, the Markov status region is divided into 9 states and we achieve Markov state transfer situation and Markov state transfer probability matrix $P$, as shown as follows:

\[
P = \begin{bmatrix}
\frac{3}{7} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{2}{7} \\
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 & \frac{2}{5} & \frac{1}{5} & 0 \\
\frac{1}{7} & \frac{2}{7} & \frac{1}{7} & \frac{1}{7} & 0 & \frac{2}{7} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{8} & \frac{1}{4} & \frac{3}{8} & \frac{1}{4} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{8} & 0 & \frac{3}{8} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\
0 & 0 & 0 & \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\
\frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\
\frac{2}{9} & 0 & 0 & 0 & 0 & \frac{1}{9} & \frac{1}{9} & \frac{2}{9} & \frac{1}{3}
\end{bmatrix}
\]  

(19)

Based on C-K Equation, the new range of percentage error in interval $t$ between 3:30 PM and 7:30 PM on link D, 12 weekdays in August, 2008 is obtained. We calculated the percentage error of every point-in-time between 3:30 PM and 7:30 PM for every 5 minutes. Then $r_t$ is accessible on the basis of Eq. (14).

4.4 Hybrid Model Prediction

As mentioned in the introduction part, existing literature applied the GARCH model in traffic volatility forecasting and show promising results in the study by Yang et al. (2010). One main shortcoming of the GARCH model is treating the positive and negative volatility symmetrically. Dimitriou et al. (2008) suggested that the nature of traffic data is non-symmetric volatility. The GJR-GARCH model proved to be able to capture the non-symmetric effect by Wang (2009). In
this paper, GJR-GARCH model is utilized to forecast volatility nature in travel time data. As the
GARCH-type models with order of (1,1) usually proved effective enough to provide accurate
and reliable forecasting in the study of Hansen and Lunde (2005), the GJR-GARCH(1,1) model
is chosen to determine the volatility part of the hybrid model. The parameter estimation of the
model is the maximum likelihood method. The “garch” toolbox in software Matlab provides a
comprehensive set of methods for univariate GARCH modeling process and therefore is used to
present the GJR-GARCH model.

As is demonstrated before, Markov Chain model is used to correct the percentage error
for the results calculated by WNN, and then residual part is corrected and a new error sequence
is regarded as the input to the GJR-GARCH model. One advantage of Markov Chain model is
non-parameter setting and the error-based model is flexible when dealing with other issues. As a
consequence, the proposed hybrid model should be compared with other single and hybrid
models. The other benchmarks are listed in section 5.

5. RESULTS AND DISCUSSION

5.1 Experiment Preparation

In this section, the prediction performance of the WNN-MAR-VOA model is evaluated using the
travel time data on link D in August. In addition, the prediction performance between the
proposed model and the other two conventional prediction models is compared. One of the
conventional prediction models is Autoregressive Integrated Moving Average, ARIMA,
assuming that the value of the predicted travel time series at time t takes a linear, weighted form
of the observed travel times in the previous several time intervals (Fei et al. 2011). And the other
one is Back Propagation Neural Network (BPNN), which is the simplified form of WNN and is
Yang et al. widely used in travel time prediction (Vlahogianni et al. 2004). The mathematical equation of an ARIMA \((p,d,q)\) is described as follows (Zou et al. 2015):

\[
(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1 - B)^d X_t = (1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_q B^q)Z_t
\]

where \(p\) is the number of autoregressive terms, \(d\) is the number of non-seasonal differences and \(q\) is the number of lagged forecast errors; \(\{X_t\}\) is stationary, \(\{Z_t\}\) is a Gaussian white noise series with mean zero and variance \(B\) is the smooth weighting factor. As for BPNN, readers can find more detailed information in (Kamarianakis et al. 2012; Yu et al. 2011).

Meanwhile, the combinations of BPNN, ARIMA and Markov Chain model, volatility model are also considered in comparison models (e.g. BPNN-MAR-VOA model includes BPNN, Markov Chain and volatility model, whose structure is similar with the proposed hybrid model). Five models are listed as benchmarks: BPNN model, ARIMA model, WNN model, BPNN-MAR-VOA model and ARIMA-MAR-VOA model. Same as the hybrid model, the GARCH(1,1) process is employed in BPNN-MAR-VOA and ARIMA-MAR-VOA. This study takes use of the selected five models as the benchmark model to investigate the effectiveness of the hybrid approach. Then the prediction performance of these six models is compared. Using a sliding 35-day training period for the proposed hybrid model, the other five models are used to predict the 5-minute travel times on link D for the weekdays (3:30 PM – 7:30 PM, Tuesday – Thursday) of August, 2008. The prediction results are compared with the observed travel time data.

The mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) are used to compare the prediction performance between the six models. The performance indexes (MAE, MAPE and RMSE) were calculated as follows:
\[ MAE = \frac{1}{n} \sum_{t=1}^{n} |\hat{P}_t - P_t| \]  

(21)

\[ MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\hat{P}_t - P_t}{P_t} \right| \]  

(22)

\[ RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\hat{P}_t - P_t)^2} \]  

(23)

\( n \) is the number of observations, \( P_t \) and \( \hat{P}_t \) are the observed and predicted travel times at time \( t \).

5.2 Prediction Performance

![Graph showing prediction performance](image-url)
Forecasting performance on August 21, 2008 is presented here as the example of all the 12 weekdays in August, 2008. Figure 3 shows forecasting performance of the selected models in two different forms. Figure 3 (a) shows the prediction performance in view of the comparison between actual values and predicted values of six models in the afternoon peak (3:30 PM-7:30 PM, August 21, 2008). Figure 3 (b) is used to demonstrate whether the specific model is robust and reliable peak hours by viewing percentage errors. Note that the blue line in Figure 3 (a) shows the actual value of travel time data. For qualitative analysis, travel time remained almost constant in the uncongested traffic condition, the predicted travel times from all six models can fit the observed travel times closely. The red line, which represents the hybrid model, fitting the curve-trend of the actual value well and perform better than the other five models. The red line stays “closest” to the actual data line. While ARIMA model performs relatively unsatisfied
Yang et al.

compared with others, especially in congested conditions. The possible reason is that ARIMA is linear-weight based, so it can hardly capture the sophisticated nature of travel time in peak hours. Another finding which should be noted is that the proposed hybrid model predicts the sudden change of travel time around 18:45 timely, while the prediction-delay can be seen clearly around 18:45 in other five models.

The similar result can be seen more clearly in Figure 3 (b). The absolute value of percentage error stays relatively low in uncongested conditions before around 4:45 PM. At 4:50 PM, the actual data of travel time experienced a rapid increase. The prediction of the other five models deviates from the observed value far more than the hybrid model. Meanwhile, a substantial volatility is observed in all models, except for WNN-MAR-VOA. Between 4:50 PM and 6:50 PM, the absolute value of percentage error of ARIMA is higher than others in almost all the period of time. With the lowest value of absolute percentage error, the hybrid model shows great robustness and reliability in congested conditions in these 2 hours. While for uncongested conditions in 3:30 PM to 4:45 PM, BPNN-MAR-VOA model and ARIMA model perform slightly better than the hybrid model. For BPNN-MAR-VOA model, the comprehensive structure and the error correction part may explain its performance in this stage. For ARIMA model, the linear-weight function is suitable for fitting the actual data in steady state. The similar results are observed in the other 11 weekdays. More detailed information can be observed from three performance indexes (MAE, MAPE and RMSE). Bold values indicate the smallest MAE, MAPE and RMSE values.

Table 1 Forecasting results of six models

<table>
<thead>
<tr>
<th></th>
<th>8/05</th>
<th>8/06</th>
<th>8/07</th>
<th>8/12</th>
<th>8/13</th>
<th>8/14</th>
<th>8/19</th>
<th>8/20</th>
<th>8/21</th>
<th>8/26</th>
<th>8/27</th>
<th>8/28</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPNN</td>
<td>4.20</td>
<td>23.12</td>
<td>24.00</td>
<td>21.94</td>
<td>16.02</td>
<td>37.66</td>
<td>17.60</td>
<td>15.36</td>
<td>25.20</td>
<td>22.56</td>
<td>13.08</td>
<td>19.60</td>
</tr>
</tbody>
</table>
As shown in Table 1, almost all goodness-of-fit values indicate that the proposed hybrid model can consistently outperform the other five models. With the combination of Markov Chain and the volatility model, the value of RMSE experiences a slight decrease. It represents an
average improvement of 20% in forecasting accuracy. For August 14th, the traffic incident may occurred cause all models show larger prediction error than usual. Prediction performance of WNN-MAR-VOA model is better than others on that day, so our model has potential to deal with accidents/incidents condition. Noting that combination model sometimes perform better than the single model. As for the single model, WNN outperforms BPNN as well as ARIMA, and BPNN shows higher accuracy than ARIMA. While the value of RMSE of BPNN-MAR-VOA is lower than WNN in some cases, and the similar observation can be seen between ARIMA-MAR-VOA and BPNN. As the combination model usually needs more time than single model to predict, computation cost will be discussed in section 5.3 to further evaluate the forecasting performance of all models.

5.3 Computational Cost

As introduced before, the dataset used in this study covers the whole 12 workdays of August in 2008, and data sampling interval is 5 min. Therefore, the amount of data is 2,808 lines for each automobile passing by. The dataset is divided into two parts, 80% of data are used for training purpose, and the rest is used for validation. The computational costs for each prediction algorithm are list in Table 2. The computation environment used herein applies Matlab 2016 on a laptop server(Surface Book) with one Intel Core i5-6300U @2.4 GHz CPU, 8 GB RAM, and the Windows 10 operating system.

Table 2 Computation cost of algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Dataset</th>
<th>Training time(s)</th>
<th>Prediction time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNN</td>
<td>Total 2,808 lines, 2,232(80%) lines for training, 576(20%) lines for validation</td>
<td>72s</td>
<td>1s</td>
</tr>
<tr>
<td>WNN</td>
<td></td>
<td>90s</td>
<td>1s</td>
</tr>
<tr>
<td>ARIMA</td>
<td></td>
<td>10s</td>
<td>1s</td>
</tr>
<tr>
<td>MAR+VOA</td>
<td></td>
<td></td>
<td>6s</td>
</tr>
</tbody>
</table>
From Table 2, we can know the time cost of prediction of MAR-VOA method in this study is about 6 second (s) whereas the total training time for the other three algorithms are about 72s, 108s and 10s, respectively.

The total computational cost is determined by two main factors. The first one is total training time of different prediction algorithms, which is highly related to their own structures. The second part is prediction time of BPNN, WNN and ARIMA, which is only 1s. Meanwhile, Markov Chain and the volatility model contribute only 6s to the total computation cost. The calculation time of MAR+VOA is explained by Equations (10) and (16), which is determined by the previous prediction value. It is clear that the computational cost of error-correction and noise-extraction steps for the MAR+VOA calculation is typically small, the main computational cost is caused by total training time for the prediction algorithms. More importantly, computation costs of all models are less than 2min, so the prediction results could be delivered to drivers during 5 min, which is the sampling interval. Therefore, the proposed hybrid model does not significantly increase the computational cost at the stage of prediction.

6. CONCLUSIONS AND FUTURE WORK

This paper proposed a new hybrid model for short-term travel time prediction in a freeway system with travel time data. It is based on the concept of decomposing travel time into three separate components: a periodic trend, a residual part, and a volatility part. The significant cyclical components are extracted from travel time data by applying Wavelet Neural Network(WNN) for presenting the periodic pattern, the residual part is corrected by Markov Chain and the volatility part is analyzed by GJR-GARCH model. Freeway travel time data collected from six AVI readers located on US-290 between I-610 and FM-1960 in Houston, Texas were employed to estimate the forecasting performance of the hybrid model and the other
Yang et al.

five models. BPNN, ARIMA and WNN are existing models. Meanwhile, BPNN-MAR-VOA, ARIMA-MAR-VOA are the combination models which have the similar structure with the hybrid model proposed in this paper. Comparisons between the new hybrid model with the other five models demonstrated that decomposing the travel time data into the listed three parts can perform better accuracy, reliability and robustness in terms of MAE, RMSE and MAPE.

This method takes periodical analysis, error correction and noise extraction into consideration and improve the forecasting performance. Then, the proposed hybrid method has its advantage in easing the over-fitting effect brought by the combination of statistical model and machine learning model. In addition, the error-based Markov Chain makes the hybrid model more flexible when analyzing other travel time data on short-term traffic forecasting. This model also has advantages in interpreting the underlying dynamic characteristics of travel time data and capturing the sudden changes of travel time ahead of time.

Limitations of the proposed models and algorithms are also discussed for future research. A more elaborated method of status-classification in Markov Chain process needs to be further studied for practical applications. Also, the deterministic algorithms on determining the number of nodes in the hidden layer of WNN as well as the size of the training samples should also be investigated. In addition, examining the robustness of the hybrid model under interrupted traffic flow conditions or incidents/accidents situations is also meaningful if the data is available. A regime-switching hybrid prediction model, which takes the alternating traffic regimes (i.e., recurrent and non-recurrent congestion conditions) into account may be considered in future research.

ACKNOWLEDGMENTS

This research is sponsored jointly by the National Natural Science Foundation of China (grant
no. 51608386) and Shanghai Sailing Program (grant no. 16YF1411900). The authors would also like to thank Ms Zhen Chen, and Dr Xu Miao for their reviews and constructive comments.
REFERENCE

Yang et al.


