Research Article

An Improved Car-Following Speed Model considering Speed of the Lead Vehicle, Vehicle Spacing, and Driver’s Sensitivity to Them

Shuaiyang Jiao,1 Shengrui Zhang,1 Zongzhi Li,2 Bei Zhou,1 and Dan Zhao1

1School of Highway, Chang’an University, Xi’an 710064, China
2Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA

Correspondence should be addressed to Shengrui Zhang; zhangsr@chd.edu.cn

Received 23 April 2019; Accepted 24 December 2019; Published 14 January 2020

This paper introduces an improved car-following speed (CFS) model that simultaneously considers speed of the lead vehicle, vehicle spacing, and driver’s sensitivity to them. Specifically, the proposed model extends the Helbing-Tilch model and Yang et al. model developed based on the principle of grey relational analysis where vehicle spacing is considered as the primary factor contributing to car-following speed choices. A computational experiment is conducted for model calibration using vehicle spacing, speed, and acceleration data derived from vehicle trajectory data of the Next Generation Simulation (NGSIM) project sponsored by the Federal Highway Administration (FHWA). It shows that speed of the lead vehicle and vehicle spacing significantly affect speed of the lag vehicle. Further, model validation is carried out using an independent NGSIM dataset by comparing vehicle speed predictions made by the calibrated CFS model with Helbing-Tilch model and Yang et al. model as benchmarks. Compared with speed prediction results of the benchmark models, mean relative errors, root mean square errors, and equal coefficient of speed predictions of the CFS model have reduced by 72.41% and 61.85%, 70.14% and 57.99%, and 33.15% and 14.48%, respectively. The findings of model validation reveal that the CFS model could improve the accuracy of speed predictions in the car-following process.

1. Introduction

In the last half-century, extensive research has been conducted worldwide to develop traffic flow models aimed to explain the nature of traffic flow characteristics and reveal the internal mechanism of congestion in normal and incident-affected traffic conditions, among which car-following models have been introduced as an effective means to describe traffic flow dynamics at the microscopic level.

Car-following models were first introduced in the 1950s to analyze the kinematic relationship between consecutive vehicles along one travel lane without overtaking maneuvers [1, 2]. These models could largely be grouped into two classifications with modeling concepts stemming from engineering and driver behavior perspectives [3, 4]. The optimal velocity (OV) model that was initiated by Bando et al. [5] is a notable example of the engineering-based car-following model [6]. It assumes that each vehicle has an optimal car-following speed (CFS) dependent on spacing between the lead and lag vehicles. The CFS function is further presumed as monotonically increasing with an upper bound. Based on this work, Helbing and Tilch [7] calibrated the CFS model that has been widely cited worldwide [8–18].

With continuing advancements in this area, new CFS models have emerged. For instance, Davis [19] proposed a refined CFS model by simultaneously considering factors of vehicle spacing and driver’s reaction time. In the same period, Hasebe et al. [20] proposed a modified CFS model by considering multiple vehicle spacing measurements associated with several lead vehicles. Extending from the work of Bando et al. [5], Nagai et al. [21] proposed a multiphase CFS model to capture the impacts of different traffic states. Considering the gradual acceleration of vehicles, Li et al. [22] incorporated a two-stage acceleration process into the original CFS model. Batista and Twrdy [23] reviewed some of the existing CFS models and conducted model calibration...
using field vehicle trajectory data. Speed predictions from the calibrated models were found to be similar to the results obtained from models developed by Bando et al. [5]. Further, some researchers have considered additional factors other than vehicle spacing. For instance, Tian et al. [24] proposed a CFS model containing factors of spacing and speed difference of two consecutive vehicles. Tang et al. [25] argued that the optimal speed of the lag vehicle is correlated not only with actual vehicle spacing but also with the vehicle spacing perceived by driver of the lag vehicle. Wang et al. [26] developed a CFS model attributable to vehicle spacing with both lower and upper bounds. Moreover, Wang et al. [27] proposed another CFS model incorporating factors of driver’s reaction time and maximum acceleration. Different from the above models, Yang et al. [28] proposed a new CFS model based on field data. They argued that the functional relationship between vehicle spacing and speed in the CFS model is logarithmic rather than hyperbolic tangent. Based on the findings of model validation, the proposed model was found to be superior to the Helbing-Tilch model in the accuracy of speed predictions. Further, some researchers have proposed more realistic car-following models based on real-world traffic conditions. For example, the effect of honk and leading vehicle’s taillight has been considered in several car-following models [29, 30]. It has been revealed that the proposed novel models could improve traffic flow stability and safety without compromising efficient capacity utilization, leading to a new direction to analyze car-following behavior.

The review of existing CFS models indicates that no research has been conducted to consider speed of the lead vehicle as a factor contributing to decisions of the lag vehicle in choosing the optimal car-following speed. With expanded field deployments of Intelligent Transportation System (ITS) installations and automated data collection techniques, real-time data on vehicle trajectories, pavement surface conditions, and climatic and weather features in the car-following process become more readily available [31–34]. The emerging vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication technologies render operations of vehicles in a cooperative manner, which could potentially improve traffic mobility and safety [35, 36]. As such, data on speed of the lead vehicle could be extracted and utilized to assess its impacts on the optimal speed of the lag vehicle in the car-following process. For this reason, the current study proposes an improved CFS model stemming from notable Helbing-Tilch model and Yang et al. model, but it goes one step beyond them by simultaneously considering vehicle spacing and speed of the lead vehicle, as well as driver’s sensitivity to them in modeling car-following speeds. Since the driver’s reaction time (or delay time) is integrated into the proposed model, it could verify stability of a cooperative car-following platoon after enabling V2V and V2I communication technologies. Specifically, reaction time can be used to characterize packet loss and transmission delay time.

The remainder of the paper is organized as follows: Section 2 begins with an introduction to the basic principle of grey relational analysis and elaborates on the proposed CFS model. Section 3 is concerned with model calibration using vehicle spacing and speed data derived from vehicle trajectory data of the Next Generation Simulation (NGSIM) project sponsored by the U.S. Federal Highway Administration (FHWA) [37]. Section 4 conducts model validation that utilizes an independent set of data on vehicle spacing and speeds generated from the NGSIM project dataset for cross comparisons of car-following speed predictions made by the calibrated CFS model, along with the Helbing-Tilch model and Yang et al. model. Multiple performance measures are employed to evaluate model predictability. Finally, Section 5 draws a summary and provides the study conclusion.

2. Methodology

2.1. Basic Principle of Grey Relational Analysis. The car-following process refers to the progression of dynamic changes in trajectories of two consecutive vehicles in a travel lane without overtaking maneuvers within a certain time interval. The grey relational analysis (GRA) is well suited to model the evolving process of a dynamic system by assessing the relationship of variables utilized to describe it [38]. In this analysis, a grey relational grade is computed by comparing the degree of geometric similarity of a column of reference data with several columns of comparison data to establish relational coefficients between the two sets of data. The higher the value of the grey relational grade, the higher the extent of relevance between the comparison and reference datasets.

Following the basic principle of grey relational analysis, impacts of factors concerning speed of the lead vehicle and vehicle spacing on speed of the lag vehicle in the car-following process could be appraised. Without loss of generality, we treat speed of the lag vehicle \( v_n \) as the column of reference data, which is denoted by the set \( Y = \{ Y(k) | k = 1, 2, \ldots, t \} \), and consider four columns of comparison data, speed of the lead vehicle \( v_{\text{ref}} \), vehicle spacing \( \Delta x \), difference in vehicle speeds \( \Delta v \), and acceleration rate of the lead vehicle \( a_{\text{ref}} \), characterized by the set \( X_i = \{ X_i(k) | k = 1, 2, \ldots, t \} \), \( i = 1, 2, 3, 4 \). Owing to non-commensurable units among vehicle spacing, speed, and acceleration rate, values of reference and comparison data need to be converted to dimensionless forms to facilitate subsequent analysis. Data conversion can be made as follows:

\[
y(k) = \frac{Y(k)}{Y(1)}, \quad k = 1, 2, \ldots, t, \tag{1}
\]

\[
x_i(k) = \frac{X_i(k)}{X_i(1)}, \quad k = 1, 2, \ldots, t; \quad i = 1, 2, 3, 4. \tag{2}
\]

The grey relational coefficient \( \xi_i(k) \) can be calculated by

\[
\xi_i(k) = \frac{\min, \min |y(k) - x_i(k)| + \rho \min, \min |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \min, \min |y(k) - x_i(k)|}. \tag{3}
\]

Setting \( \Delta_i(k) = |y(k) - x_i(k)| \), we get
\[ \xi_i(k) = \frac{\min, \min \Delta_i(k) + \rho \min, \min \Delta_i(k)}{\Delta_i(k) + \rho \min, \min \Delta_i(k)}, \quad (4) \]

where \( \rho \in (0, 1) \) is the distinguishing coefficient. A lower value of \( \rho \) represents a high level of significance in the difference between grey relational coefficients. Typically, the value of \( \rho \) is set as 0.5 [38].

The grey relational grade \( r_i \) is calculated by
\[ r_i = \frac{1}{t} \sum_{k=1}^{t} \xi_i(k), \quad k = 1, 2, \ldots, t. \quad (5) \]

With grey relational grades computed between speed of the lag vehicle and factors concerning speed of the lead vehicle, vehicle spacing, speed differences, and acceleration rate of the lead vehicle, the significance of their impacts on speed of the lag vehicle could be determined.

### 2.2. Two Benchmark Models

To demonstrate the necessity of introducing the proposed CFS model for improved accuracy of car-following speed prognosis, two classical models are selected as benchmarks for cross comparisons. The first one is the Helbing-Tilch model [7], which is of the following specification:
\[ V(\Delta x(t)) = V_1 + V_2 \cdot \tanh[C_1(\Delta x(t) - l_i) - C_2], \quad (6) \]

where \( V(\Delta x(t)) \) is car-following speed with a vehicle spacing \( \Delta x \) at time \( t \); \( V_1, V_2, C_1, C_2 \) are model parameters; and \( l_i \) is length of the lead vehicle. An exemplary set of model coefficients is as \( V_1 = 6.75 \text{ m/s}, \quad V_2 = 7.91 \text{ m/s}, \quad C_1 = 0.13 \text{ m}^{-1}, \quad \text{and} \quad C_2 = 1.57. \)

The second benchmark model is the Yang et al. model purposely developed by Yang et al. [28] to modify the Helbing-Tilch model for enhanced predictability of speed for the lag vehicle in the dynamically evolving car-following process. The model is as below:
\[ V(\Delta x(t)) = m \cdot \ln\left(\frac{\Delta x(t)}{n}\right), \quad (7) \]

where \( V(\Delta x(t)) \) is car-following speed with a vehicle spacing \( \Delta x \) at time \( t \); \( m \) is a nonzero constant; and \( n \) is the equivalent vehicle length as the total of vehicle length and safe clearance of consecutive vehicles. An exemplary set of model coefficients is as \( m = 8.83 \) and \( n = 5.5 \text{ m} \).

For Helbing-Tilch model, as shown in equation (6), when the vehicle spacing at time \( t \), \( \Delta x(t) \), is \( l_i \), \( V(l_i) \) can be calculated as \( V_1 + V_2 \cdot \tanh(-C_2) \). Applying this value to the calibrated Helbing-Tilch model, we obtain \( V(l_i) = -0.9978 \text{ m/s} \), indicating that the lag vehicle performs reversing operations. This is not safe nor does it match real-world situations in that the speed of the lag vehicle would drop to zero at the incomplete stoppage.

For the Yang et al. model depicted by equation (7), when vehicle spacing at time \( t \), \( \Delta x(t) \), is \( n \), the speed of the lag vehicle is reduced to zero, which is also inconsistent with the real-world situation. During peak hours, when vehicle spacing equals the equivalent vehicle length termed as the total of vehicle length and clearance, the speed of the lag vehicle might just be reduced and does not necessarily fall to zero. Moreover, neither of the above models has considered the driver’s sensitivity to changes in vehicle spacing and speed of the lead vehicle in modeling of the car-following process. Once vehicle spacing and dynamics of the lead vehicle got changed, it would take a short interval for the driver of lag vehicle to perceive the change and react to it accordingly. Also, the relative importance of changes in spacing and speed of the lead vehicle in influencing driver’s response decision-making tends to vary according to driver attributes.

#### 2.3. The Proposed CFS Model

To overcome unrealistic predictions of car-following speeds made by the above benchmark models under dense traffic flow circumstances and incorporate driver’s sensitivity to stimuli into the analysis, an improved CFS model is proposed as shown in the following equation:
\[ V_n(t) = \lambda \cdot \ln\left(\frac{\Delta x(t) - \Delta t}{s_{\text{min}}}\right) + k \cdot V_{n-1}(t - \Delta t), \quad (8) \]

where \( V_n(t) \) is speed of the lag vehicle at time \( t \); \( V_{n-1}(t - \Delta t) \) is speed of the lead vehicle at time \( t - \Delta t \); \( \Delta t \) is the driver’s reaction time for the lag vehicle; \( \Delta x(t - \Delta t) \) is the vehicle spacing at time \( t - \Delta t \) and \( \Delta x(t - \Delta t) \geq s_{\text{min}} \); \( s_{\text{min}} \) is the minimum vehicle spacing to ensure safe operations; \( \lambda \) is the sensitivity coefficient of driver of the lag vehicle to vehicle spacing; and \( k \) is the response coefficient of driver of the lag vehicle to speed of the lead vehicle.

The proposed model possesses the following essential properties:

1. When \( \Delta x(t - \Delta t) = s_{\text{min}} \), the spacing between the lead and lag vehicles equals the minimum vehicle spacing at time \( t - \Delta t \), which leads to \( V_n(t) = k \cdot V_{n-1}(t - \Delta t) \). This shows that when the minimum vehicle spacing is reached, the lag vehicle becomes insensitive to the spacing relative to the lead vehicle and its speed is only correlated with the speed of the lead vehicle, of which the condition of \( 0 \leq k \leq 1 \) is satisfied so that the speed of the lag vehicle would be lower or equivalent to that of the lead vehicle to ensure the safety of car-following operations

2. When \( V_{n-1}(t - \Delta t) = 0 \), \( V_n(t) = \lambda \cdot \ln[\Delta x(t - \Delta t)/s_{\text{min}}] \). This indicates that when the speed of the lead vehicle reduces to zero at time \( t - \Delta t \), the speed of the lag vehicle only depends on vehicle spacing \( \Delta x(t - \Delta t) \). When vehicle spacing is large enough and speed of the lead vehicle reaches zero, the speed of the lag vehicle will gradually reduce to zero, instead of decreasing to zero abruptly

### 3. Proposed Model Calibration

#### 3.1. Data Description

For the proposed CFS model, efforts are made to calibrate model coefficients using data on vehicle spacing, speed, and acceleration derived from vehicle
trajectory data collected in the field as part of the FHWA’s NGSIM project [37]. Specifically, the vehicle trajectory data contain operational details of vehicles traveling in one direction of a 640 m long roadway segment of U.S. Road 101 that comprises five through movement lanes and one auxiliary lane. Data collection is for the period of 7:50–8:35 am of a typical weekday initiated on June 15, 2005, and updated on April 4, 2018. The roadway segment was divided into eight sections, each section was equipped with a video camera to record video images of vehicle movements within the section, and the cameras were synchronized to facilitate tracking of vehicle movements along all sections of the roadway segment. The video images of vehicle movements were recorded sequentially according to the order in which each vehicle entered the road segment and were updated in each 0.1 second of time. A customized software called NGVIDEO was employed to extract trajectory details of individual vehicles traversing through the roadway segment. In total, 3,849,725 data entries were recorded. Table 1 presents summary statistics of the vehicle trajectory data.

3.2. Data Processing. The raw data on vehicle trajectories converted from video images are not naturally organized into car-following groups that could be immediately used for model calibration. When a vehicle is traveling along a short roadway section, video images of its movements are taken by a camera. Based on those images, the absence or presence of a car-following group could be identified. In general, no car-following group is formed if no lead vehicle is present in the same travel lane of the vehicle. When the vehicle encounters a lead vehicle in the same travel lane, no car-following group exists if the vehicle spacing is too large. A car-following group is created if there exists a closely spaced lead vehicle in the same travel lane of the lag vehicle.

When a vehicle is traveling along multiple sections of the roadway segment, video images of its movements are captured by multiple synchronized cameras installed at the beginning points of different sections. It might involve no car-following, a single car-following group, or multiple car-following groups. If no lead vehicle is present in the same travel lane of the lag vehicle across all roadway segments, no car-following group exists. If the lead vehicle closely spaced with the lag vehicle in the same travel lane remains unchanged across all roadway segments, one car-following group is developed. If different lead vehicles are closely spaced with the lag vehicle in the same travel lane across all roadway segments, multiple car-following groups are composed. Therefore, screening of vehicle trajectory data is necessary to confirm absence or presence of car-following actions and, if so, sort out the car-following groups in conjunction with relevant trajectory details for each vehicle traversing through multiple sections of the roadway segment.

Figure 1 depicts a data screening procedure to help identify the subset of vehicle trajectory data involving car-following maneuvers. In order to confirm two consecutive vehicles that might belong to one car-following group, Lane_ID is used as a primary key to ensure that they progress along the same travel lane. Further, Frame_ID and Total_Frames are utilized as additional primary keys to make sure that spacing and headway between them meet car-following boundary conditions.

For execution of the proposed data screening procedure, a Python program is developed to extract vehicle trajectory data involving car-following maneuvers. Overall, 3,319,685 data entries (out of 3,849,725 total records) associated with 4,790 car-following groups are extracted. More specifically, the screened dataset contains 710,390, 728,350, 655,672, 655,162, and 570,111 data entries corresponding to 821,967, 950, 1,012, and 1,040 car-following groups distributed from travel lane 1 to lane 5, respectively.

To ensure data validity for subsequent analysis, the screened dataset is further examined to remove potential outliers. In particular, the dataset contains high fidelity field measured values of vehicle spacing, speeds of the lead and lag vehicles, and acceleration rates of the lead and lag vehicles, which are sufficient to analyze microscopic traffic characteristics. With detailed data on the above five parameters available, the Mahalanobis distance method for multivariate analysis can be readily adapted to identify potential outliers [39]. The Mahalanobis distance of two groups of car-following data can be calculated by equation (9):

$$D_M(k, k') = \sqrt{(X_k - X_{k'})^T \cdot S^{-1} \cdot (X_k - X_{k'})},$$

where $D_M(k, k')$ is the Mahalanobis distance computed using field measured data values of all variables associated with comparison car-following groups $k$ and $k'$; $X_k$ and $X_{k'}$ are vectors consisting of average data values of variable $j$ or $j'$ associated with car-following groups $k$ and $k'$, $X_k = (x_{k1}, x_{k2}, \ldots, x_{kj}, x_{kj}', \ldots, x_{k5})^T$, $X_{k'} = (x_{k'1}, x_{k'2}, \ldots, x_{k'j'}, x_{k'j'}, \ldots, x_{k'5})^T$; $x_{kj}$ and $x_{k'j'}$ are field measured values of all variables associated with car-following groups $k$ and $k'$, $\Sigma_{j=1}^{5} x_{kj}/n_k$, $\Sigma_{j=1}^{5} x_{k'j'}/n_{k'}$, $\Sigma_{j=1}^{5} x_{kj}/n_k$, $\Sigma_{j=1}^{5} x_{k'j'}/n_{k'}$, $S^{-1}$ is the inverse matrix of the covariance matrix $S$, consisting of elements of covariance $\text{Cov}(x_{j}, x_{j'})$ of variables $j$ and $j'$ that can be computed by $\text{Cov}(x_{j}, x_{j'}) = \Sigma_{k=1}^{5} n_k \cdot \text{Cov}(x_{kj}, x_{k'j'})/\Sigma_{k=1}^{5} n_k$ and $\text{Cov}(x_{kj}, x_{k'j'}) = \Sigma_{k=1}^{5} \{[(x_{kj} - \bar{x}_k) \cdot (x_{k'j'} - \bar{x}_{k'})]/n_k\}$, $k = 1, 2, \ldots, 5$; $j, j' = 1, 2, \ldots, 5$, where $j=5$ indicating that observed data details of five variables including spacing, speeds, and acceleration rates of consecutive vehicles in different car-following groups are used for analysis; $n_k$ = sample size of variable $j$ or $j'$ associated with car-following group $k$; and $n_{k'}$ = sample size of variable $j$ or $j'$ associated with car-following group $k'$.

With Mahalanobis distance values calculated using data from comparison car-following groups, the Chi-square statistical test can be carried out to sort out data outliers. The use of five variables for the current analysis yields 4 degrees of freedom. At a significance level of 0.005, the critical value of Chi-square statistic becomes $X_{0.005, 4}^2 = 14.86$. Therefore, any data entry of comparison car-following groups with the calculated Mahalanobis distance value exceeding the critical value is an outlier for removal. After data validity checks, the number of effective car-following groups and the lane-based distribution of those groups remain unchanged. However, data entries have reduced slightly by...
1.6% from 3,319,685 to 3,265,261, which are distributed as 699,092,715,795,644,020,644,848, and 561,506 from travel lane 1 to lane 5, respectively. In this study, data entries for travel lanes 1 to 4 are used for preliminary data analysis and proposed model calibration, while remaining data entries for travel lane 5 are used for model validation. Table 2 lists exemplary data entries associated with one car-following group of vehicles using travel lane 3 with a 1.5-second refreshing rate.

3.3. Preliminary Data Analysis. Having prepared valid data entries associated with car-following groups of vehicles using travel lanes 1 and 4, equations (1)–(5) are executed to compute grey relational grades for speed of the lag vehicle in relation to vehicle spacing, speed of the lead vehicle, speed difference between the lead and lag vehicles, and acceleration rate of the lead vehicle, respectively. Table 3 exhibits the grey relational analysis results.

It can be seen from Table 3 that grey relational grades of speed of the lag vehicle in relation to speed of the lead vehicle and vehicle spacing are relatively high. This indicates that speed of the lag vehicle is highly affected by speed of the lead vehicle and vehicle spacing in the car-following process. In contrast, lower values of grey relational grades are obtained between speed of the lag vehicle and speed difference, as well as acceleration rate of the lead vehicle. This reveals that their impacts on speed of the lag vehicle are relatively less significant.

To intuitively demonstrate the findings of grey relational analysis, Figures 2 and 3 illustrate examples of time-varying correlations of speed of the lag vehicle with car-following spacing and speed of the lead vehicle and also with speed difference and acceleration rate of the lead vehicle using data associated with car-following groups 778–783 in travel lane 1, groups 1,241–1,246 in lane 2, groups 1,233–1,240 in lane 3, and groups 934–938 in lane 4, respectively. Consistent with grey analysis results, the graphic illustrations indicate that

---

**Table 1: Summary statistics of vehicle trajectory data for calibrating the proposed model.**

<table>
<thead>
<tr>
<th>Data items</th>
<th>Min</th>
<th>1st quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd quartile</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle_ID</td>
<td>1</td>
<td>506</td>
<td>934</td>
<td>1,024</td>
<td>1,426</td>
<td>2,783</td>
</tr>
<tr>
<td>Vehicle_Class</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Frame_ID</td>
<td>5</td>
<td>2,474</td>
<td>4,021</td>
<td>4,100</td>
<td>5,746</td>
<td>8,906</td>
</tr>
<tr>
<td>Total_Frames</td>
<td>160</td>
<td>531</td>
<td>718</td>
<td>728</td>
<td>911</td>
<td>1,328</td>
</tr>
<tr>
<td>Lane_ID</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Leading_ID</td>
<td>0</td>
<td>450</td>
<td>882</td>
<td>976</td>
<td>1,390</td>
<td>3,109</td>
</tr>
<tr>
<td>Lagging_ID</td>
<td>0</td>
<td>465</td>
<td>901</td>
<td>991</td>
<td>1,408</td>
<td>3,109</td>
</tr>
<tr>
<td>Spacing (m)</td>
<td>0.00</td>
<td>13.29</td>
<td>18.43</td>
<td>20.78</td>
<td>25.64</td>
<td>237.05</td>
</tr>
<tr>
<td>Speed (m/s)</td>
<td>0.00</td>
<td>6.11</td>
<td>9.75</td>
<td>9.43</td>
<td>12.32</td>
<td>29.05</td>
</tr>
<tr>
<td>Acceleration (m/s²)</td>
<td>−3.41</td>
<td>−0.26</td>
<td>0.00</td>
<td>0.05</td>
<td>0.44</td>
<td>3.41</td>
</tr>
</tbody>
</table>

---

**Figure 1: Car-following data screening procedure.**
Table 2: Exemplary data entries associated with one car-following group of vehicles using travel lane 3 (1.5-second refreshing rate).

<table>
<thead>
<tr>
<th>Frame_ID</th>
<th>Spacing Δx (m)</th>
<th>Lag vehicle</th>
<th>Lead vehicle</th>
<th>Mahalanobis distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speed v_n (m/s)</td>
<td>Acceleration a_n (m/s²)</td>
<td>Speed v_{n-1} (m/s)</td>
<td>Acceleration a_{n-1} (m/s²)</td>
</tr>
<tr>
<td>211</td>
<td>23.13</td>
<td>7.01</td>
<td>-0.05</td>
<td>7.70</td>
</tr>
<tr>
<td>212</td>
<td>23.20</td>
<td>7.00</td>
<td>-0.04</td>
<td>7.70</td>
</tr>
<tr>
<td>213</td>
<td>23.27</td>
<td>7.02</td>
<td>0.29</td>
<td>7.70</td>
</tr>
<tr>
<td>214</td>
<td>23.34</td>
<td>7.09</td>
<td>1.02</td>
<td>7.66</td>
</tr>
<tr>
<td>215</td>
<td>23.40</td>
<td>7.23</td>
<td>1.82</td>
<td>7.57</td>
</tr>
<tr>
<td>216</td>
<td>23.43</td>
<td>7.42</td>
<td>2.19</td>
<td>7.45</td>
</tr>
<tr>
<td>217</td>
<td>23.41</td>
<td>7.62</td>
<td>1.82</td>
<td>7.33</td>
</tr>
<tr>
<td>218</td>
<td>23.36</td>
<td>7.76</td>
<td>1.01</td>
<td>7.24</td>
</tr>
<tr>
<td>219</td>
<td>23.29</td>
<td>7.83</td>
<td>0.29</td>
<td>7.21</td>
</tr>
<tr>
<td>220</td>
<td>23.22</td>
<td>7.84</td>
<td>-0.04</td>
<td>7.20</td>
</tr>
<tr>
<td>221</td>
<td>23.16</td>
<td>7.84</td>
<td>-0.04</td>
<td>7.20</td>
</tr>
<tr>
<td>222</td>
<td>23.09</td>
<td>7.84</td>
<td>0.00</td>
<td>7.21</td>
</tr>
<tr>
<td>223</td>
<td>23.03</td>
<td>7.84</td>
<td>0.00</td>
<td>7.21</td>
</tr>
<tr>
<td>224</td>
<td>22.97</td>
<td>7.84</td>
<td>0.00</td>
<td>7.19</td>
</tr>
<tr>
<td>225</td>
<td>22.90</td>
<td>7.76</td>
<td>-1.27</td>
<td>7.09</td>
</tr>
</tbody>
</table>

Table 3: Summary of grey relational grades using data associated with car-following groups.

<table>
<thead>
<tr>
<th>Comparison variable</th>
<th>Vehicle spacing Δx</th>
<th>Speed of the lead vehicle v_{n-1}</th>
<th>Speed difference Δv</th>
<th>Acceleration of the lead vehicle a_{n-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey relational grade</td>
<td>0.9391</td>
<td>0.9849</td>
<td>0.8942</td>
<td>0.8657</td>
</tr>
</tbody>
</table>

Figure 2: An example of time-varying correlations of speed of the lag vehicle with car-following spacing and speed of the lead vehicle. (a) Line 1, (b) line 2, (c) line 3, and (d) line 4.
speed of the lag vehicle is more closely related to vehicle spacing and speed of the lead vehicle than to speed difference and acceleration rate of the lead vehicle.

Figure 4 provides disaggregated information on vehicle maneuvers by car-following group in different travel lanes. In general, the same vehicle spacing or speed of the lead vehicle could correspond to different speeds of the lag vehicle. This is largely attributable to varying sensitivity levels of the driver of the lag vehicle to vehicle spacing and speed of the lead vehicle in different traffic conditions. This observation confirms that speed of the lead vehicle is not the only factor influencing speed of the lag vehicle in the car-following process. Although similar correlation patterns are discovered, the ranges of varying speeds of the lag vehicle correlated to the same vehicle spacing and speed of the lead vehicle appear to be larger for vehicles involving car-following operations in travel lanes 2 and 3. This may be explained by the fact that outcomes of decision-making by the driver of the lag vehicle to maintain desired car-following speeds tend to vary more greatly to keep abreast of more complicated traffic conditions in the middle lanes. Moreover, a logarithmic correlation trend is detected between speed of the lag vehicle and vehicle spacing, whereas a linear trend is identified between speeds of the lag and lead vehicles.

3.4. Proposed Model Calibration. Further to using the dataset comprising 2,703,755 data entries associated with 3,750 car-following groups of vehicles using travel lanes 1 and 4 for preliminary data analysis, the same dataset is utilized for proposed model calibration. For equation (8), we denote the following notations: $Y_i = V_n(t)$, $X_{1i} = \ln(\Delta x(t - \Delta t)/s_{\min})$, $X_{2i} = V_{n-1}(t - \Delta t)$, $\alpha_1$, and $k = \alpha_2 (i = 1, 2, \ldots, n)$. Thus, a multiple linear statistical model without an intercept is formulated as

$$Y_i = \alpha_1 \cdot X_{1i} + \alpha_2 \cdot X_{2i},$$

where $Y_i$ is variable related to speed of the lag vehicle; $X_{1i}$ is the transformed variable of vehicle spacing; $X_{2i}$ is the variable concerning speed of the lead vehicle; $\alpha_1$ is sensitivity coefficient of the lag vehicle to vehicle spacing; $\alpha_2$ is response coefficient of the lag vehicle to the speed of the lead vehicle.
As a practical matter, the minimum vehicle spacing is set as equivalent to the 99th percentile of all vehicle spacing in the dataset sorted in descending order, which is 6.67 m. In addition, the optimal speed of the lag vehicle must be kept as zero to avoid crash occurrences when vehicle spacing equals the minimum vehicle spacing and speed of the lead vehicle becomes zero.

Figure 4: Illustration of correlations of speed of the lag vehicle with vehicle spacing and speed of the lead vehicle. (a) Vehicle No. 783 following vehicle No. 778 in lane 1, (b) vehicle No. 1,246 following vehicle No. 1,241 in lane 2, (c) vehicle No. 1,240 following vehicle No. 1,233 in lane 3, and (d) vehicle No. 938 following vehicle No. 934 in lane 4.
The NLOGIT6 software package is executed for model calibration. Table 4 summarizes calibrated coefficients of the proposed CFS model. As seen in the table, both variables concerning vehicle spacing and speed of the lead vehicle are found to be statistically significant at a significance level of 1%. The adjusted $R^2$ is 0.97. The coefficients of the two variables are different, which indicates that vehicle spacing plays a different level of relative importance from speed of the lead vehicle on the driver’s choice of a desired car-following speed. Also, the coefficient corresponding to the variable of speed of the lead vehicle is 0.8653. This shows that the contribution of speed of the lead vehicle to the optimal speed of the lag vehicle in the car-following process is at most 86.53%, which is at the critical point of reaching the minimum vehicle spacing, namely, $\Delta x(t - \Delta t) = s_{\text{min}}$ and $X_{1,\text{critical}} = \ln(\Delta x(t - \Delta t)/s_{\text{min}}) = 0$.

### 4. Model Validation

In order to examine the ability to which the proposed model is used for car-following speed predictions, model validation is carried out using a separate dataset associated with the highway segment of the NGSIM project involved with car-following data collection. The model validation dataset consists of 561,506 data entries associated with 1,040 car-following groups of vehicles using travel lane 5 of the NGSIM segment.

Model validation is accomplished in three steps. First, the same dataset used for developing coefficients of the proposed CFS model that contains 2,703,755 data entries associated with 3,750 car-following groups of vehicles using travel lanes 1 and 4 of the NGSIM segment is further utilized to calibrate coefficients of the Helbing-Tilch model and Yang et al. model. Next, the calibrated CFS model, Helbing-Tilch model, and Yang et al. model are separately applied to the model validation dataset for car-following speed predictions. Finally, the three sets of data on car-following speed predictions are compared with the real-world car-following speed measurements contained in the model validation dataset and assessed by different evaluation criteria to confirm the superiority of the proposed CFS model.

#### 4.1. Benchmark Model Calibration

To facilitate the Helbing-Tilch model calibration, Taylor expansion is performed to expand the key component $\tanh[C_1(\Delta x(t) - 6.67) - C_2]$ as $8.3725, 27.6471, 0.0127,$ and 0.0003. Without loss of generality, only the first three terms of Taylor expansion expression are retained. Setting $Y_i = V_n(\Delta x(t))$, we obtain the approximate expression of equation (6) in the following:

$$ Y_i = \alpha_0 + \alpha_1 \cdot \Delta x(t)_1 + \alpha_2 \cdot (\Delta x(t)_2)^3 + \alpha_3 \cdot (\Delta x(t)_3)^5. \quad (11) $$

Using the model calibration dataset, the coefficients in equation (11) can be estimated using multiple linear regression analysis as summarized in Table 5. Substituting the Taylor expansion of $\tanh[C_1(\Delta x(t) - l_c) - C_2]$ to the Helbing-Tilch model as equation (6) and setting vehicle length $l_c$ equivalent to the minimum vehicle spacing of 6.67 m, a system of equations is created between $\alpha_0$, $\alpha_1$, $\alpha_2$, and $\alpha_3$ and coefficients $V_1$, $V_2$, $C_1$, and $C_2$ of Helbing-Tilch model. Solving for the simultaneous equations would yield values of $V_1$, $V_2$, $C_1$, and $C_2$ as 8.3725, 27.6471, 0.0127, and 0.0003.

To facilitate the Yang et al. model calibration, we denote $Y_1 = V_n(\Delta x(t))$ and $Y_2 = \ln(\Delta x(t)/n)$. This yields the zero-intercept linear form of the Yang et al. model as equation (7) as

$$ Y_i = m \cdot X_i \quad (12) $$

In the model calibration, the minimum vehicle spacing of 6.67 m established from the model calibration data is adopted as the safe vehicle spacing, namely, $n = 6.67$ m in the Yang et al. model. Using the model calibration dataset, the coefficients in equation (12) can be estimated using simple linear regression analysis as summarized in Table 6.

#### 4.2. Model Evaluation Criteria

Based on the calibrated CFS model in conjunction with the Helbing-Tilch model and Yang et al. model, data values of relevant variables in the model validation dataset are applied to individual models for car-following speed predictions. The three sets of predicted car-following speeds are then compared with the field speed measurements using different model evaluation criteria to assess the model’s prediction power. Table 7 summarizes model evaluation criteria, including mean relative errors (MRE), root mean square errors (RMSE), and equal coefficient (EC) [40, 41].

#### 4.3. Model Evaluation Results

Table 8 presents the results of model performance evaluation. As seen in Table 8, the average MRE values of the proposed model, Helbing-Tilch model, and Yang et al. model are 10.23%, 45.24%, and 29.31%; the average RMSE values of the corresponding models are 4.7625, 16.8212, and 10.8330; and the average EC values of related models are 0.9330, 0.7107, and 0.8353, respectively. Compared with those of the Helbing-Tilch model and Yang et al. model, the average MRE, RMSE, and EC values of the proposed model have reduced by 72.41% and 61.85%, 70.14% and 57.99%, and 33.15% and 14.48%, respectively. Moreover, the minimum and maximum values as well as the range of those extreme values of MRE, RMSE, and EC criteria computed by the proposed model are significantly lower than those of the Helbing-Tilch model and Yang et al. model.

Figure 5 further illustrates distributions of MRE, RMSE, and EC values concerning the three models. Particular to the proposed model, most of the MRE values are below 0.3, RMSE values are below 10.0, and the EC values are over 0.8.
These observations further reveal that the proposed model outperforms the Helbing-Tilch model and Yang et al. model.

5. Summary and Conclusion

This study has introduced an improved model for car-following speed predictions stemming from the widely cited Helbing-Tilch model and Yang et al. model developed based on the principle of grey relational analysis where vehicle spacing is considered as the primary factor in determining car-following speeds. The proposed model goes one step beyond the above models by simultaneously considering vehicle spacing, speed of the lead vehicle, and driver’s sensitivity to them for car-following speed predictions. Unlike the Helbing-Tilch model and Yang et al. model, the improved model could readily describe some critical car-following conditions that are consistent with real-world situations. In particular, the lag vehicle would choose to slow down or come to a full stop rather than always abruptly reducing the speed to zero when the spacing of two consecutive vehicles reaches the minimum spacing level. In addition, the time lag effect between vehicle spacing and speed of the lead vehicle perceived by driver of the lag vehicle and driver’s response to choose an optimal car-following speed is incorporated into the improved model. The relative importance of vehicle spacing and speed of the lead vehicle in contributing to determination of speed of the lag vehicle might not be equivalent to each other.

A computational experiment is conducted for model calibration and validation using vehicle spacing, speed, and

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model coefficient</th>
<th>t-statistic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant term</td>
<td>-9.3173</td>
<td>-258.90</td>
</tr>
<tr>
<td>Vehicle spacing $\Delta x(t)_i$</td>
<td>7.6629</td>
<td>873.80</td>
</tr>
<tr>
<td>Third powered vehicle spacing $(\Delta x(t)_i)^3$</td>
<td>-3.3027</td>
<td>-452.79</td>
</tr>
<tr>
<td>Fifth powered vehicle spacing $(\Delta x(t)_i)^5$</td>
<td>6.1323</td>
<td>316.08</td>
</tr>
</tbody>
</table>

Adjusted $R^2 = 0.61$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model coefficient</th>
<th>t-statistic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle spacing $X_{ii}$</td>
<td>27.7230</td>
<td>4,585.64</td>
</tr>
</tbody>
</table>

Adjusted $R^2 = 0.89$

**Table 7: Summary of model evaluation criteria.**

<table>
<thead>
<tr>
<th>Model evaluation criterion</th>
<th>Description</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean square errors</td>
<td>$\text{MRE} = (1/N)\left(\sum_{i=1}^{N}</td>
<td>\hat{Y}_i - O_i</td>
</tr>
<tr>
<td>Root mean square errors</td>
<td>$\text{RMSE} = \sqrt{\sum_{i=1}^{N} (\hat{Y}_i - O_i)^2/N}$</td>
<td>The smaller the RMSE, the better the model predictivity</td>
</tr>
<tr>
<td>Equal coefficient</td>
<td>$\text{EC} = 1 - (\sqrt{\sum_{i=1}^{N} (\hat{Y}<em>i - O_i)^2} / (\sqrt{\sum</em>{i=1}^{N} \hat{Y}<em>i^2} + \sqrt{\sum</em>{i=1}^{N} O_i^2})$</td>
<td>EC ranges between 0 and 1, and an EC value of 0.9 or above generally indicates a good fit</td>
</tr>
</tbody>
</table>

Note. $\hat{Y}_i$ is the $i^{th}$ predicted car-following speed, $O_i$ is the $i^{th}$ field speed measurement, and $N$ is the total number of data points ($i = 1, 2, \ldots, N)$.

**Table 8: Summary of model validation results.**

<table>
<thead>
<tr>
<th>Performance criterion</th>
<th>Proposed model</th>
<th>Helbing-Tilch model</th>
<th>Yang et al. model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>3.30%</td>
<td>6.12%</td>
<td>5.93%</td>
</tr>
<tr>
<td>1st quartile</td>
<td>8.21%</td>
<td>35.61%</td>
<td>18.79%</td>
</tr>
<tr>
<td>Mean</td>
<td>10.23%</td>
<td>45.24%</td>
<td>29.31%</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>12.62%</td>
<td>45.76%</td>
<td>33.09%</td>
</tr>
<tr>
<td>Max</td>
<td>13.77%</td>
<td>55.57%</td>
<td>41.75%</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>2.1133</td>
<td>2.7332</td>
<td>2.5181</td>
</tr>
<tr>
<td>1st quartile</td>
<td>3.9662</td>
<td>12.2857</td>
<td>8.1394</td>
</tr>
<tr>
<td>Mean</td>
<td>4.7625</td>
<td>16.8212</td>
<td>10.8330</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>5.2657</td>
<td>17.6351</td>
<td>12.5353</td>
</tr>
<tr>
<td>Max</td>
<td>5.9414</td>
<td>22.2920</td>
<td>15.4097</td>
</tr>
<tr>
<td>EC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>0.6300</td>
<td>0.3087</td>
<td>0.3054</td>
</tr>
<tr>
<td>1st quartile</td>
<td>0.9135</td>
<td>0.6231</td>
<td>0.7617</td>
</tr>
<tr>
<td>Mean</td>
<td>0.9330</td>
<td>0.7107</td>
<td>0.8353</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>0.9255</td>
<td>0.6951</td>
<td>0.8084</td>
</tr>
<tr>
<td>Max</td>
<td>0.9473</td>
<td>0.7753</td>
<td>0.8816</td>
</tr>
</tbody>
</table>
acceleration data derived from vehicle trajectory data of the FHWA’s NGSIM project. Preliminary data analysis and calibration of the proposed model indicate that vehicle spacing and speed of the lead vehicle significantly affect speed of the lag vehicle in the car-following process, and their relative importance in contributing to car-following speed determination appears to be unequal.

For the purpose of model validation, the Helbing-Tilch model and Yang et al. model are employed as benchmarks. The same set of data for calibrating the proposed model is utilized to estimate the Helbing-Tilch model and Yang et al. model coefficients. Subsequently, an independent set of NGSIM data is used for car-following speed predictions separately made by the proposed model, Helbing-Tilch model, and Yang et al. model. The three sets of car-following speed predictions are compared with field speed measurements and assessed by multiple model performance criteria. The relative errors, root mean square errors, and equal coefficient values computed in accordance with car-following speed predictions by the proposed model have reduced by 72.41% and 61.85%, 70.14% and 57.99%, and 33.15% and 14.48%, respectively.

The findings of model validation reveal that the proposed model could improve the accuracy of car-following speed predictions. With multiple factors including vehicle spacing, speed of the lead vehicle, and sensitivity of the lag vehicle’s driver to them simultaneously considered in the improved car-following speed model, a new venue is offered to develop effective countermeasures centered on multiple speed contributing factors for highway mobility and safety management. It should be noted that the datasets used for preliminary data analysis, model calibration, and validation are collected from a freeway segment during peak periods of a typical weekday. Expanded analyses are desirable to generalize findings of the current study by considering the following aspects and conducting cross comparisons: (1) interstate, multilane, and two-lane highways; (2) weekdays versus weekends; (3) AM and PM peak, adjacent-to-peak, and off-peak periods of a typical day; (4) undersaturated, oversaturated, and mixed traffic conditions; and (5) travel-
lane specific analysis. It is expected that the generalized models could further enhance the accuracy and precision of car-following speed predictions.

Data Availability

The data for the numerical analysis used in the present study are derived from the NGSIM program; the NGSIM Program US Route 101 datasets are freely available for download at https://www.fhwa.dot.gov/publications/research/operations/its/06135/index.cfm.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge publication support by National Natural Science Foundation of China (no. 71871029), China Postdoctoral Science Foundation (no. 2015M582593), and the Fundamental Research Funds for the Central Universities, CHD (no. 300102218401, no. 300102218404, and no. 300102219306).

References


