JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2015; **49**:279–296
Published online 16 December 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.1258

Parametric modeling of the heteroscedastic traffic speed variance from loop detector data

Haizhong Wang[1]*, Zhixia Li[2], David Hurwitz[1] and Jianjun Shi[3]

[1]*School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, U.S.A.*
[2]*Department of Civil and Environmental Engineering, University of Wisconsin Madison, Madison, WI 53706, U.S.A.*
[3]*Beijing Key Laboratory of Transportation Engineering, Beijing University of Technology, Beijing, 100124, China*

SUMMARY

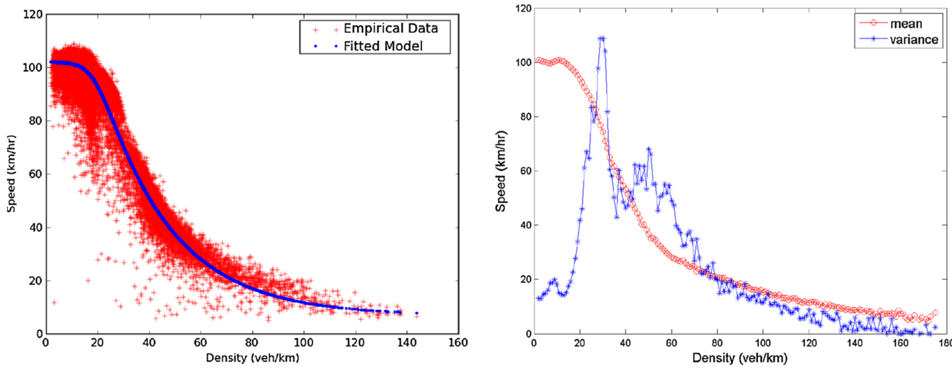
Traffic speed variance is defined as a measure of the dispersion of space mean speeds among drivers. Empirical speed–density observations exhibit a structured traffic speed variance, which has been found to be associated to the roadway crash rate, the fatality rate, and travel time variability. The objective of this paper is to propose a generalized traffic speed variance function to describe this structured variance. The proposed speed variance function is a response of the speed–density curve with two additional parameters. The estimation of the model parameters in the proposed traffic speed variance function can be carried out through an iterative nonlinear least-square algorithm (i.e., LevenbergMarquardt). A series of logistic speed–density curve with varying parameters are used in the proposed traffic speed variance function with different levels of performance. The proposed traffic speed variance model can potentially help to unveil the underlying mechanism of empirical traffic phenomenon such as spontaneous congestion or capacity reduction.

KEY WORDS: traffic speed variance model; speed–density relationship; virtual loop detector data

1. INTRODUCTION

Nonconstant variability appears in numerous fields of scientific inquiry such as chemical and bioassay [1]; traffic flow is no exception. The wide-scattering effects (Figure 1(a)) in the equilibrium speed–density relationship are well-understood and accepted by transportation researchers and professionals. The mean curve of the wide-scattering equilibrium speed–density relationship (Referring to Figure 1(b)) has provoked sufficient modeling efforts using both deterministic and stochastic modeling techniques [2, 3]. However, the traffic speed variance (or equivalently the speed variability), which has been determined to be associated with roadway crash frequency, travel time, and its variability on both highway and major arterial roads [4], has not been not sufficiently addressed in literature. Lave [5] noted in his paper "Speeding, Coordination, and the 55 MPH Limit" that traffic speed variance contributes to fatal rate, not speed. On the basis of his state cross-section data analysis of 1981 and 1982, it was found that there is hardly any statistically discernible relationship between fatality rate and average speed, whereas there is a strong relationship with traffic speed variance. Guo and Smith [6] used a linear stochastic univariate traffic speed series to forecast short-term traffic speed variance; however, this research focused on short-term forecasting instead of traffic speed variance models. Therefore, qualitative and quantitative description of traffic speed variance is significant to both researchers and transportation professionals. With this information, the impact of traffic speed variation on traffic flow operation and management can be better understood. Instead of exploring the potential applications of traffic speed variances (such as road crash rate, travel time, and its variability estimation), this paper

*Correspondence to: Haizhong Wang, School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, U.S.A. E-mail: Haizhong.Wang@oregonstate.edu



(a) Scattered Plot (b) Mean and Variance

Figure 1. The scattered plot of an empirical speed–density relationship and its mean, variance curves.

focuses on the mathematical modeling of traffic speed variances regarding the variance curve's nonlinearity and heteroscedasticity exhibited from the empirical traffic speed variances. The impetus to model speed variation arose from the need to empirically account for observed traffic dynamics such as wide-scattering plots of the empirical fundamental diagram, and the onset of congestion as traffic density varies from low to high.

1.1. Objectives of research

The purpose of this paper is to provide a framework for modeling traffic speed variance. This problem is an intricate one as traffic speed variability depends on diverse dynamic factors, such as driver behavior heterogeneity and consistency, vehicle operating factors and types; and static factors, including roadway characteristics such as horizontal and vertical curvatures, the posted speed limit and design speed. Accurate modeling of the speed variability phenomenon has to take these numerous factors into consideration, which may require substantial data preparation and computation to perform the analysis. To address these issues, a method for modeling structured traffic speed variance exhibited in empirical traffic observations as an unsymmetrical parabolic shape curve has been developed. The proposed modeling methodology is tailored for traffic speed variance on uninterrupted traffic flow facilities. Therefore, this work contributes to the body of knowledge for advanced traffic operations and management. The authors illustrate the suggested approach through a case study and experimental results.

1.2. Outline of paper

The remainder of this paper is organized as follows. The background of traffic speed variance implications is briefly reviewed in Section 2, and the research problem is defined in this section using empirical observations. Section 3 of the paper provides a novel modeling of the traffic speed variance with a focus on the nonlinearity and heterogeneity of the empirically observed traffic speed variance. In this context, a generalized traffic speed variance function is proposed in Section 3.2 in which the speed variance is a response of the speed–density curve with two additional parameters based on the smoothed shape of the speed variance curve (ignoring the local spikes). The choice of the speed–density curves is provided in Section 3.3 in which a series of speed–density curves with varying model parameters are presented. In Section 4.1, the estimation of model parameters of the traffic speed variance function and the speed–density function are provided. In Section 5, we briefly summarize our findings.

DOI: 10.1002/atr

2. PROBLEM FORMULATION

2.1. Background

The knowledge of traffic speed variance (or variation of traffic speeds) has prevalent applications in numerous areas, for example, speed estimation using single/double-loop detector data [7,8], reduction of crash rates on roadways [9,10], and the design of roadway characteristics [11]. Previous research efforts approached this problem from the following directions: (1) From a policy and safety perspective, speed variance and speed variance reduction are serious considerations for the setting of different speed limits for passenger cars, and heavy trucks [12], and highway work zone safety control [13]. Recently, Lu and Chen [14] analyzed the impacts of speed dispersion influence on traffic safety using empirical data from China and the Netherlands. Similiar to Lu and Chen's study [14], Vadeby and Forsman [15] studied the state of knowledge regarding speed distribution and traffic safety. To be more specific, how changes in real speed distribution will impact the accident risk using three different traffic safety measures. Aarts and Schagen [16] identified quantitative relationships between individual driving speed and the risk of road crash rates; the authors concluded that larger speed differentials between vehicles are related to higher crash rates. Additionally, Graves *et al.* [[17]] proposed a model of the optimal speed limit that explicitly recognized the roles of average speed, speed variance, and the enforcement level. (2) From a traffic operation and management standpoint, Saifallah [18] showed that the density of maximum throughput is near the density of maximum speed variance, which agrees with the observation that maximum speed variance occurs at the critical density where capacity is usually obtained. Collins *et al.* [[19]] conducted research about traffic speed variability on rural two-lane highways and the relationship between the increase of speed variance and the increase of crash potential. Research efforts about traffic speed variances on multilane highways were performed to quantify the influence of specific access design factors on speed variance using statistical techniques for improving safety performance [20]. Recently, Rakha [21] proposed a relationship between time mean speed and space mean speed variances, as well as space mean speed and travel time variance. This review of what has been carried out is not intended to be complete but to emphasize that a better understanding of traffic speed variance and its contributing factors is extremely important to the transportation profession. A variety of contributing factors (i.e., driver's lane changing behavior, number of lanes, driver/vehicle heterogeneity [4]) affect traffic speed variance, but the most significant factor was identified by Garber as the difference between the design speed and the posted speed limit [22]. From a safety perspective, traffic speed variance directly relates to crash frequency and could help contribute to the identification of crash-prone locations on highways and arterials [4,22,23,19].

2.2. Notation

The following notation is utilized to formulate the problem.

Parameter	Description
σ_i^2	traffic speed variance
s_i^2	the empirical variance
i	index of traffic density
j	index of jam density
k	traffic density
k_j	jam density
k_c	critical traffic density
k_t	transition density
$V(k)$	traffic speed as a function of density k
$V(k,\theta)$	traffic speed as a function of traffic density k and a set of parameters θ
ε	a random error, which is the discrepancy between the empirical traffic speed and the model speed
τ	a parameter in the generalized traffic speed variance function
δ^2	a parameter in the generalized traffic speed variance function, the physical meaning of this parameter is the traffic speed variance when traffic is at free-flow condition

DOI: 10.1002/atr

v_f	free-flow speed
v_b	the finite traffic speed at stop-and-go conditions
v_d	road design speed
θ_1	a scale parameter that describes how the curve is stretched out over the whole density range
θ_2	a parameter that controls the lopsidedness of the curve
5PL	the five-parameter logistic speed–density model
4PL	the four-parameter logistic speed–density model
3PL	the three-parameter logistic speed–density model

2.3. Problem definition

The problem can be viewed as finding a mathematical function to describe change in traffic speed variance as traffic density varies. The authors start with the empirical traffic speed variances observations collected from Georgia State Route 400 as can be seen from Figure 2 (which were collected by virtual loop detectors on freeway GA400). From the structured empirical traffic speed variances (Figures 3 and 4), one can see that the variance of traffic speed is quite comparable with the mean speed in a wide range [24]. Apparently, an ignorable existence of a structured variance has been verified from the empirical traffic observations. One key question is how can we represent this variance curve mathematically. Before that, the authors first start the formulation of traffic speed variance function with its nonlinearity and heterogeneity.

It is shown that in the paper, the structure of the empirical traffic speed variance is different from weekdays to weekends. Because of the decrease of the number of vehicles on Saturday and Sundays on GA400, the traffic flow on both directions of GA400 is free flow most of the time. Therefore, the structure of empirical traffic speed variance over the weekends only contains the free-flow phase but not the congested phase because there is hardly any congestion during weekends on this highway. The magnitude of the traffic speed variance varies larger during weekends than weekdays; that is, the

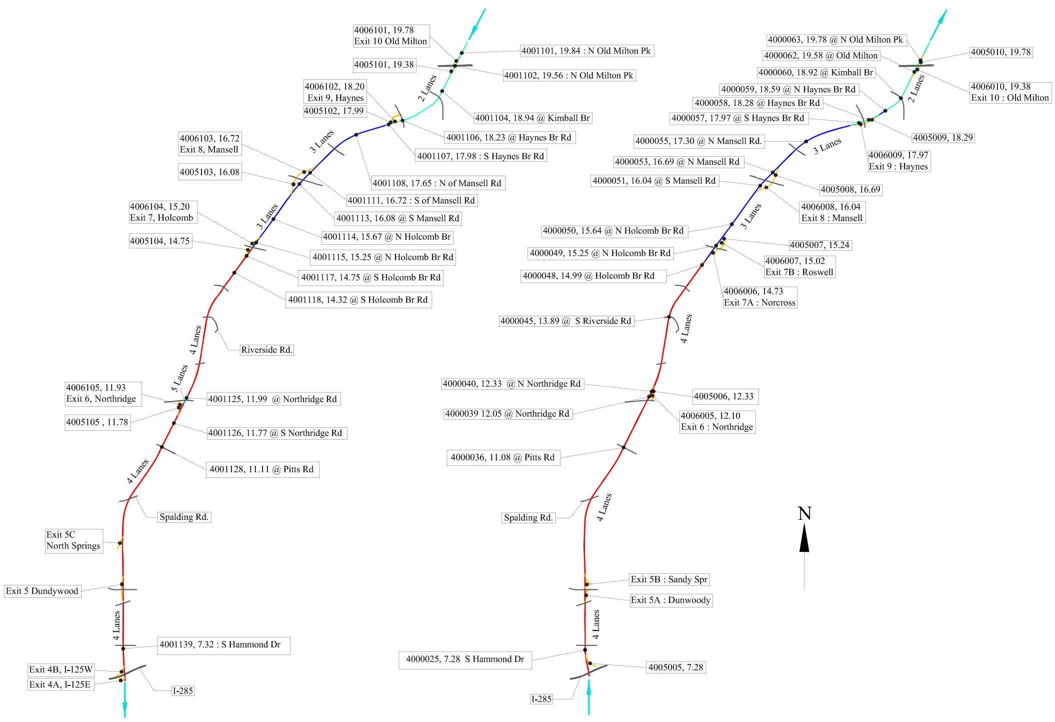
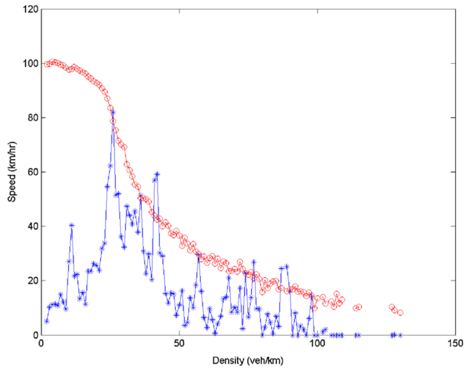
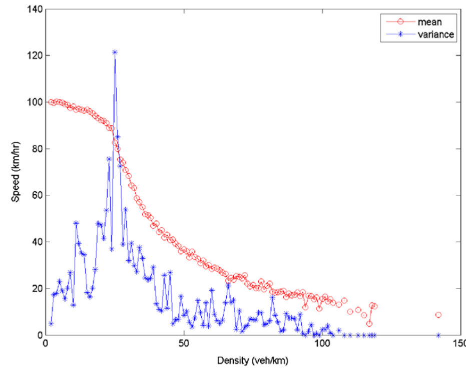


Figure 2. Study site: GA400 southbound and northbound with 100 stations.

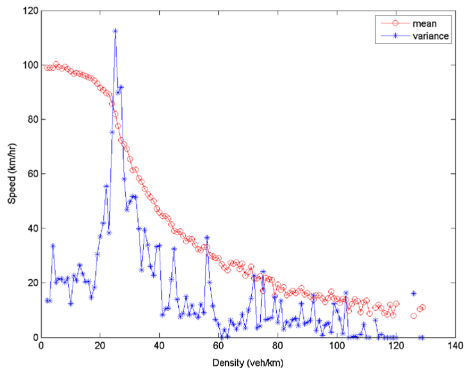
DOI: 10.1002/atr



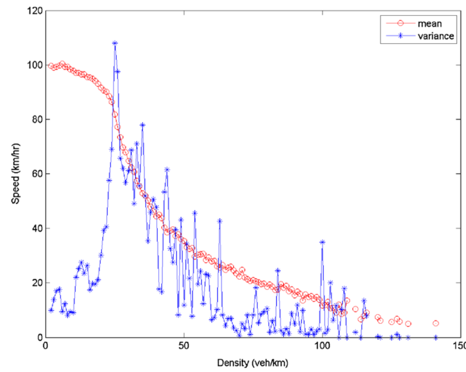
(a) Monday



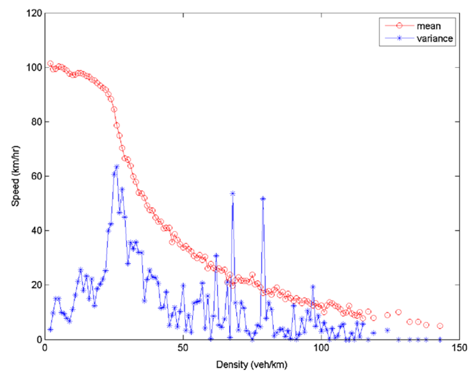
(b) Tuesday



(c) Wednesday



(d) Thursday



(e) Friday

Figure 3. Weekday change of traffic speed variance from 1-year observations at station 4000026 with time aggregation level 5 minutes. The units on the *y*-axis is km/hour and vehs/km on the *x*-axis, the same on the following figures.

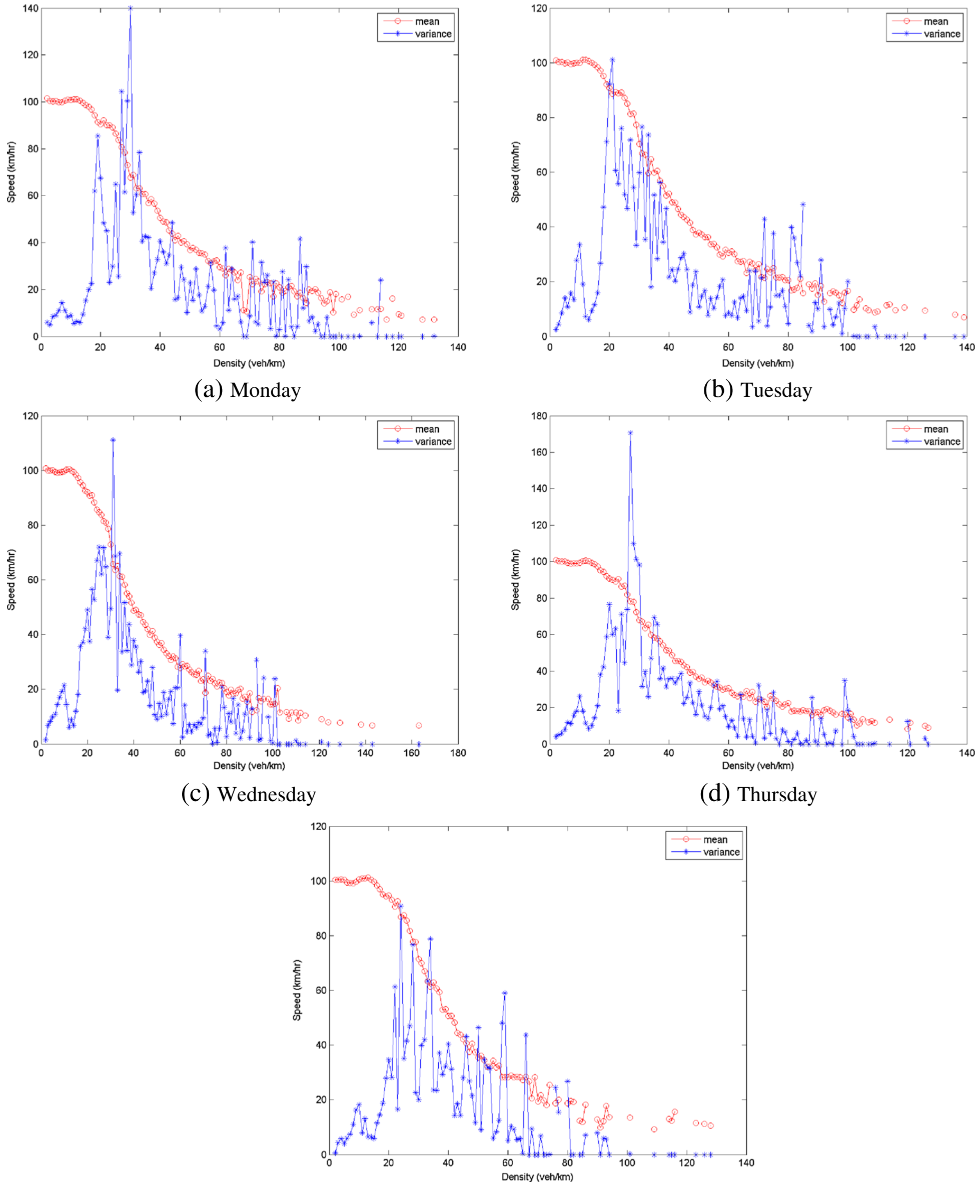
spikes of the maximum speed variance are typically two or three times the free-flow speed, which is significantly larger than the normal weekdays. The model can be used to identify a density value for flow breakdown of the critical density 40 vehs/km during weekdays or weekends.

3. THE MODELING OF TRAFFIC SPEED VARIANCE

3.1. The nonlinearity and heterogeneity of speed variance

Nonconstant variability is prevalent in numerous of scientific fields. In general, there are two cases in the difference of variances [1]: (1) the variance $\sigma_i - \sigma_{i+1}$ is small, and (2) the variance $\sigma_i - \sigma_{i+1}$ is large. If the speed variance $\sigma_i - \sigma_{i+1}$ is small, we can be confident in the approximation of the variances with homogeneity by assuming $Var(\varepsilon_{ij}) = \sigma^2$ [1]. In the case of a large $\sigma_i - \sigma_{i+1}$, the physical

DOI: 10.1002/atr



(a) Monday (b) Tuesday (c) Wednesday (d) Thursday (e) Friday

Figure 4. Weekday change of traffic speed variance from 1-year observations at station 4001118 with time aggregation level 5 minutes.

interpretation is that various driver groups behave differently under changing traffic densities. The empirical evidence is against the assumption of homogeneous variances. In this case, the real heterogeneous variation of σ^2 is approximated by a function f called the variance function such that $Var(\varepsilon_{ij}) = v(k_i,\delta^2,\theta,\tau)$. In most situations, f is assumed to depend on $v(k,\theta)$. For example, $v(k_i,\delta^2,\theta,\tau) = \sigma_2 v(k_i,\theta)^\tau$, where τ is a set of parameters that have to be estimated or assumed to be known already. Modelers usually can simplify the necessary assumptions by assuming that the vector θ_p varies in the interior of an interval. The function $v(k_i,\theta)$ is assumed to be twice continuously differentiable with respect to the parameters θ [1,2].

To model the nonlinearity and heterogeneity, the authors considered the errors in the model, which can be described by

$$\varepsilon = V(k) - v(k,\theta) \tag{1}$$

DOI: 10.1002/atr

in which $V(k)$ is the empirical traffic speed and $v(k,\theta)$ is the speed value given by a speed–density model. For each value of density k, the model can be rewritten as

$$V_{il}(k) = v(k_i, \theta) + \varepsilon_{il} \tag{2}$$

with l varying from 1 to n_i (n_i is the number of speed observations over a long period under density k_i) and i from 1 to j (k_j is jam density). For each value of i, the empirical variance of speed can be calculated by

$$s_i^2 = \frac{1}{n_i}\sum_{l=1}^{n_i}(V_{il} - V_{i*})^2 \tag{3}$$

with $V_{i*} = \frac{1}{n_i}\sum_{l=1}^{n_i} V_{ij}$. From the empirical data, the variance of traffic speeds is heterogeneous, so we assume that $Var(\varepsilon_{il}) = \sigma_i^2$. The model is given by the following

$$V_{il}(k) = v(k_i, \theta) + \sigma_i^2 \tag{4}$$

with $Var(\varepsilon_{il}) = \sigma_i^2$ and $E(\varepsilon_{il}) = 0$, where $l = 1, \ldots, n_i$; $i = 1, \ldots, j$; and total number of observations equal $n = \sum_{i=1}^{j} n_i$. $v(k_i,\theta)$ is given by equations in Section 3.3. ε_{il} are independent Gaussian random variables. By construction, ε is a random error that is equal to the discrepancy between empirical traffic speed and the model speed $v(k_i,\theta)$. θ is a vector of p parameters $\theta_1, \theta_2, \ldots, \theta_p$.

3.2. Parametric modeling of the variance

A variance function needs to be determined to estimate the heterogeneous traffic speed variance. To make the choice of variance function, some qualitative or quantitative indications of empirical traffic speed variance are needed [1]. Figures 3–5 plot the empirical mean of speed–density observations over 1 year and its corresponding speed variance on both weekdays and weekends. From the figures, the authors observe that the variance of the empirical observations first grows as traffic density increases and then decreases with the maximum achieved at an intermediate density around $k_c(35 \rightarrow 45)$ (veh/km) typically called the critical density. The exhibited empirical traffic speed variance can be depicted by a parabola that is smaller at the two ends: free flow and congested. The authors also notice that the general shape of the empirical traffic speed variance is similar to each other from Monday to Friday (weekdays); however, the empirical variance shape is different on Saturday and Sunday from weekdays. This difference can be explained by the traffic pattern change from weekdays to weekends. During weekends, the number of commuters drop significantly so that the vehicles travel at higher speeds on GA400. The proposed traffic speed variance functions in this research can be applied to both weekdays and weekends, but obviously, the traffic speed variance curve on weekends only shows the free-flow conditions.

For an increasing variance, there are essentially two scenarios [1]. The first scenario is that the variance varies as a power of the response

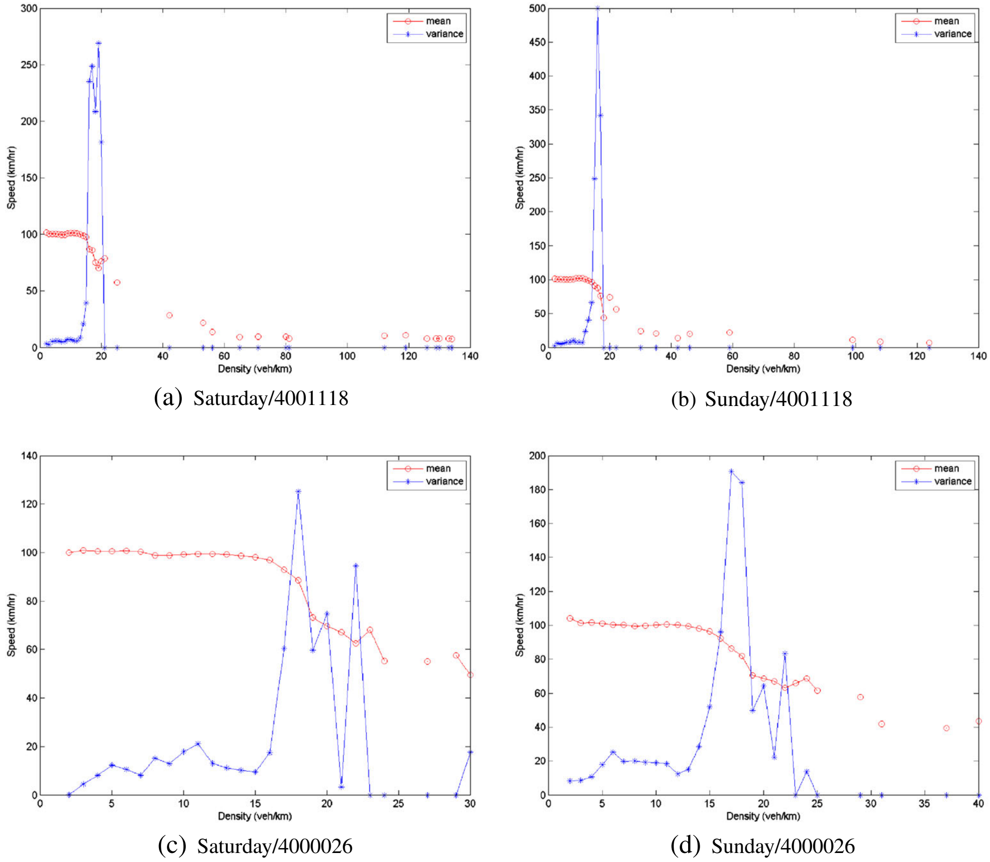
$$\sigma_i^2 = \delta^2\phi(k_i, \theta, \tau) = \delta^2 v(k_i, \theta)^\tau \tag{5}$$

The other one is that the variance varies as a linear function of the response

$$\sigma_i^2 = \delta^2\phi(k_i, \theta, \tau) = \delta^2(1 + \tau v(k_i, \theta)) \tag{6}$$

Judging from the empirical observations of traffic speed variance in the previous section, we found that these two variance functions are not appropriate to model a parabola-shaped variance function.

DOI: 10.1002/atr



(a) Saturday/4001118 (b) Sunday/4001118

(c) Saturday/4000026 (d) Sunday/4000026

Figure 5. Weekend change of traffic speed variance from 1-year observations with time aggregation level 5 minutes.

For a variance function varying like a parabola, the most generalized model is given by

$$\sigma_i^2 = \delta^2 + \delta^2\tau_1(v_{\max} + \tau_2 - v(k_i,\theta))(v(k_i,\theta) - v_{\min}) \tag{7}$$

in which $v_{\max}$ is the maximum value and $v_{\min}$ is the smallest value of $v(k_i,\theta)$. For the empirical traffic speed variance, the variance function that we adopt is given by

$$\sigma_i^2 = \delta^2\left(1.0 + \tau v(k_i,\theta)\left(v_f - v(k_i,\theta)\right)\right) \tag{8}$$

in which δ and τ are parameters; $v(k,\theta)$ adopts the deterministic five-parameter logistic speed–density models, but it opens the existing single-regime speed–density models listed in [2]. A slight change to this model can be made by replacing the free-flow speed term v_f with a highway design speed; that is, a higher v_f, here called v_d, will yield the following model:

$$\sigma_i^2 = \delta^2(1.0 + \tau v(k_i,\theta)(v_d - v(k_i,\theta))) \tag{9}$$

The speed–density relationship embedded in the variance function is open to the existing speed–density models, for example, a Greenshield model. If a Greenshield speed–density model is utilized in the generalized traffic speed variance function with additional parameters, a specific traffic speed variance function based on it will be in the form of

DOI: 10.1002/atr

$$\sigma_i^2 = \delta^2\left(1.0 + \tau v_f\left(1 - k_i/k_j\right)\left(v_d - \left(v_f - v_f k_i/k_j\right)\right)\right) \quad (10)$$

A plot of the traffic variance corresponding to the Greenshield model can be seen from Figure 6. A linear speed–density model leads to a parabolic traffic speed variance function, but we observe that the variance function based on the Greenshield model differs from the empirical traffic speed variance in terms of the magnitude and symmetry.

The motivation of using Greenshield model to test the proposed traffic speed variance function is twofold. Firstly, Greenshield model is a classic speed–density model that is a simple linear function with only two parameters. Second, the Greenshield model is easy to understand and apply in different scenarios. The major purpose of this paper is to show the newly proposed logistic speed–density models and its corresponding traffic speed variance models; therefore, the authors decided not to compare the newly developed model against other existing nonlinear models such as the Greenberg model and the Underwood model to divert the focus. There are varying parameters used in different nonlinear speed–density models, which made the comparison to be a challenging task. This is because it is not convincing to make conclusive remarks to compare models with totally different model parameters. However, this paper will compare the three newly proposed logistic speed–density models with similar sets of parameters.

3.3. *Choice of speed–density curves and corresponding variance functions*

The empirical speed–density observations exhibit a reversed “S” shape, which makes logistic modeling a natural candidate [1,2]. The authors extended the series of logistic speed–density models to the generalized variance function to obtain a series of traffic speed variance function based on the five-parameter (5PL), four-parameter (4PL), and three-parameter (3PL) logistic speed–density models developed in [2]. The performance of the family of logistic speed–density model to the empirical data can be viewed from Figure 7.

The authors found that the logistic speed–density function describes the empirical data [2]. Equation 11 is the most general five-parameter logistic speed–density model (5PL) in sigmoidal shape. In which, v_f and v_b are the upper and lower asymptotes, respectively. Specific to our case, v_f represents free-flow speed. v_b is the average travel speed under stop-and-go conditions. This parameter assumes that traffic has finite movements even in congested situations [25]. θ_1 is a scale parameter that describes how the curve is stretched out over the whole density range, and θ_2 is a parameter that controls the lopsidedness of the curve. The parameter k_t is the turning point at which the speed–density curve makes the transition from free flow to congested flow. Therefore, a sigmoidal shape five-parameter logistic speed–density model and the corresponding traffic speed variance function take a functional form of

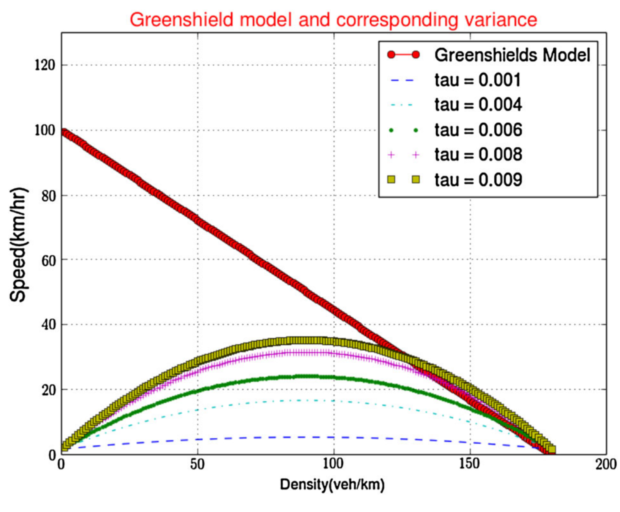
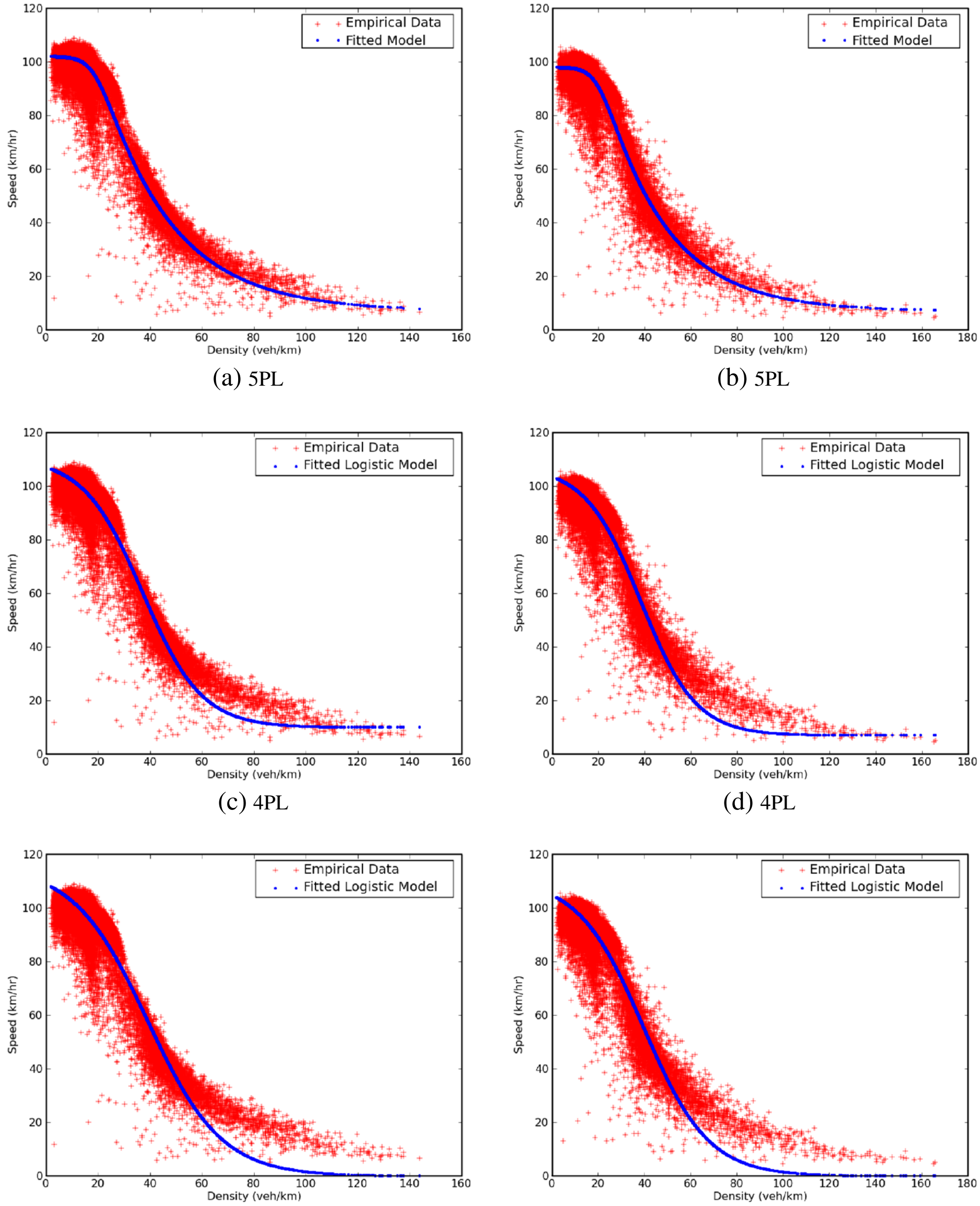


Figure 6. The variance function based on a Greenshield model.

DOI: 10.1002/atr



(a) 5PL (b) 5PL

(c) 4PL (d) 4PL

(e) 3PL (f) 3PL

Figure 7. Performance of five-parameter (top), four-parameter (middle), and three-parameter (bottom) logistic speed–density model fitting to the same set of empirical data.

$$\begin{cases} v(k,\theta) = v_b + \dfrac{v_f - v_b}{\left(1 + \exp\left(\dfrac{k - k_t}{\theta_1}\right)\right)^{\theta_2}} \\ \sigma_i^2 = \delta^2\left(1.0 + \tau v(k,\theta)\left(v_f - v(k,\theta)\right)\right) \end{cases} \tag{11}$$

The performance of the traffic variance function based on a five-parameter logistic speed–density model under varying parameters of δ^2 and τ is demonstrated in Figure 8. A four-parameter logistic speed–density model (4PL) is obtained by reducing the parameter θ_2, and its corresponding variance function is given by

$$\begin{cases} v(k,\theta) = v_b + \dfrac{v_f - v_b}{1 + \exp\left(\dfrac{k - k_c}{\theta_1}\right)} \\ \sigma_i^2 = \delta^2\left(1.0 + \tau v(k,\theta)\left(v_f - v(k,\theta)\right)\right) \end{cases} \tag{12}$$

DOI: 10.1002/atr

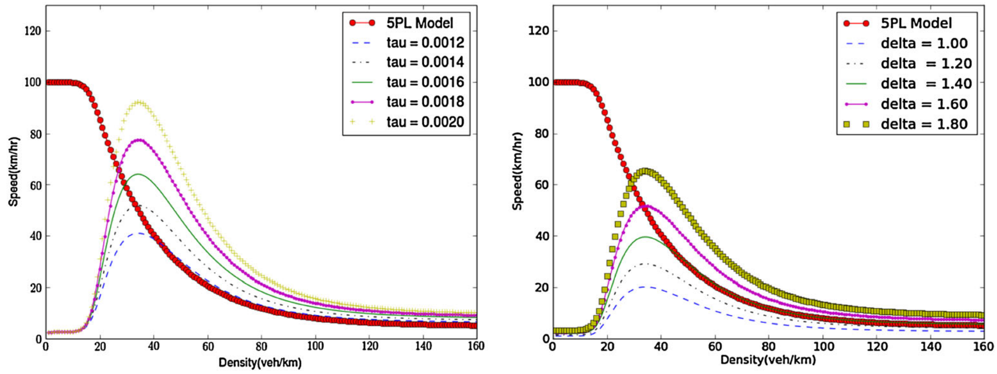


Figure 8. The performance of the traffic variance function based on a five-parameter logistic speed–density model under varying parameters of δ^2 and τ.

The performance of a four-parameter logistic speed–density model can be viewed from Figure 9 when the magnitude of the parameters is varied. The traffic variance function based on this four-parameter logistic speed–density model is plotted in Figure 9(e) and (f) when the four parameters are fixed, whereas the two additional parameters in the variance function is varied.

The physical meaning of the other parameters remains unchanged. Different from the 5PL, the 4PL captures the critical traffic density k_c instead of k_t. The three-parameter logistic speed–density model (3PL) can be obtained by removing the user-specified average travel speed v_b at stop-and-go traffic conditions, and the variance function based on this 3PL curve is given by

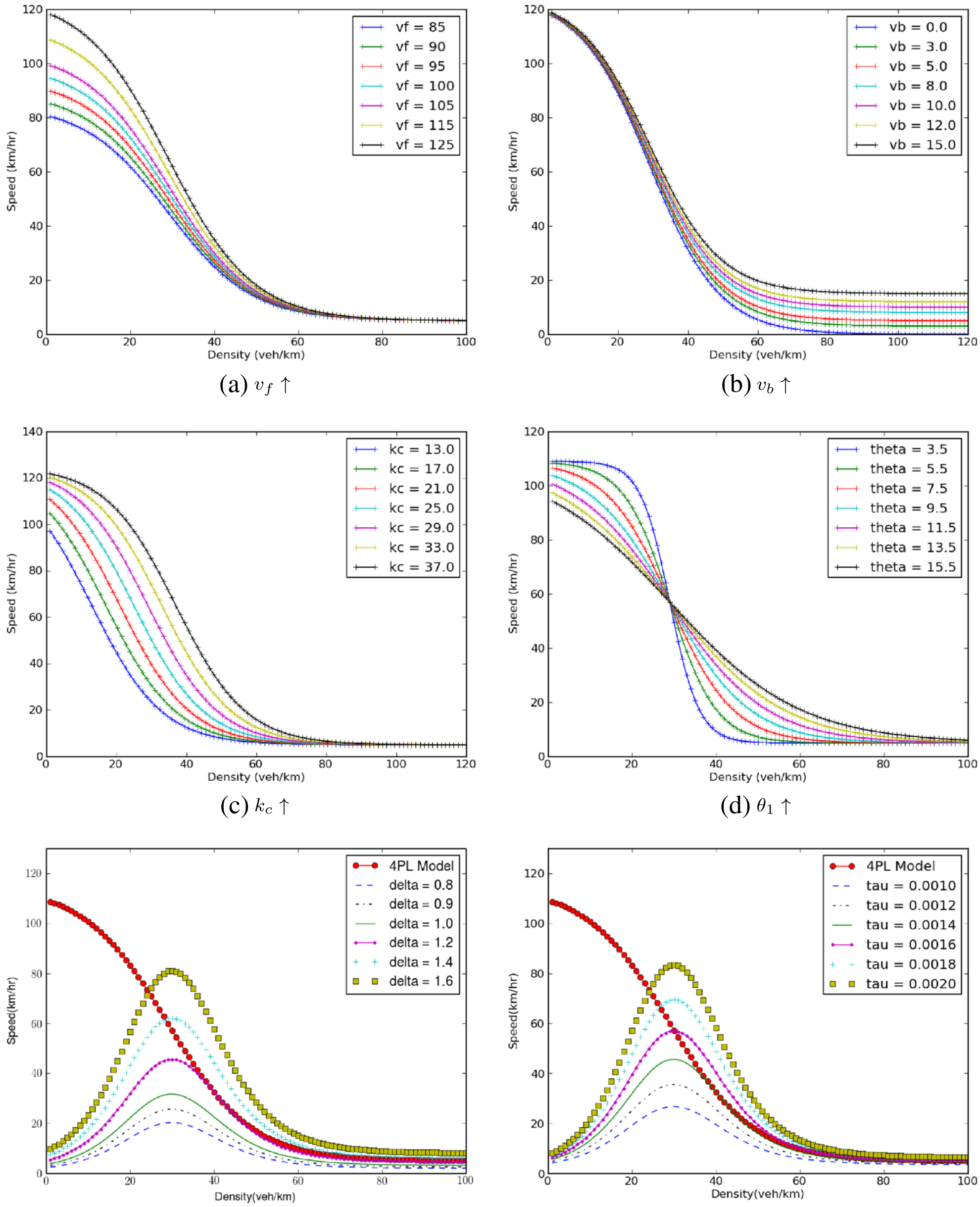
$$\begin{cases} v(k,\theta) = \dfrac{v_f}{1+\exp\left(\dfrac{k-k_c}{\theta_1}\right)} \\ \sigma_i^2 = \delta^2\left(1.0+\tau v(k,\theta)\left(v_f - v(k,\theta)\right)\right) \end{cases} \tag{13}$$

The effects of varying the three parameters $v_f, v_b,$ and θ_1 of the three-parameter logistic speed–density model can be viewed from Figure 10(a)–(c). The performance of the corresponding variance function based the 3PL model is demonstrated in Figure 10(d) and (e).

4. EXPERIMENTAL RESULTS

4.1. Estimation of model parameters

As aforementioned, the modeling of traffic speed variance is dependent on the speed–density relationship. Figures 11 and 12 show the performance of the proposed variance functions with 3PL, 4PL, and 5PL logistic speed–density models when compared with empirical mean and speed variance. It is evident that the selection of parameters in the speed–density model and the traffic speed variance function is critical to the performance when fitting the models to empirical data. Thus, the identification of the optimal parameters in the series of models is essential to reduce the residuals between the model estimation and the empirical data. In this paper, the estimation of the model parameters such as $v_f, v_b, k_t,$ $k_c, \theta_1, \theta_2, \tau,$ and δ^2 was performed through a least-square algorithm [26], which has been implemented in Scipy.Optimize [27]. The estimation of traffic speed variance from empirical data can be carried out through obtaining the parameters from the speed data observation point. The parameters are typically location-based, but they will vary in a certain range that can be obtained from empirical observations. The two parameters in the traffic speed variance functions are more difficult to obtain than the parameters in the speed–density curve because most of the parameters in the speed–density curve has physical meaning, whereas the physical meaning of the two parameters in the traffic speed variance function is unclear. Luckily, the range of two parameters in the traffic speed variance function can be specified; for simpler applications, the two parameters can be assumed constant, but in reality, the two parameters are also location based.



(a) $v_f \uparrow$ (b) $v_b \uparrow$

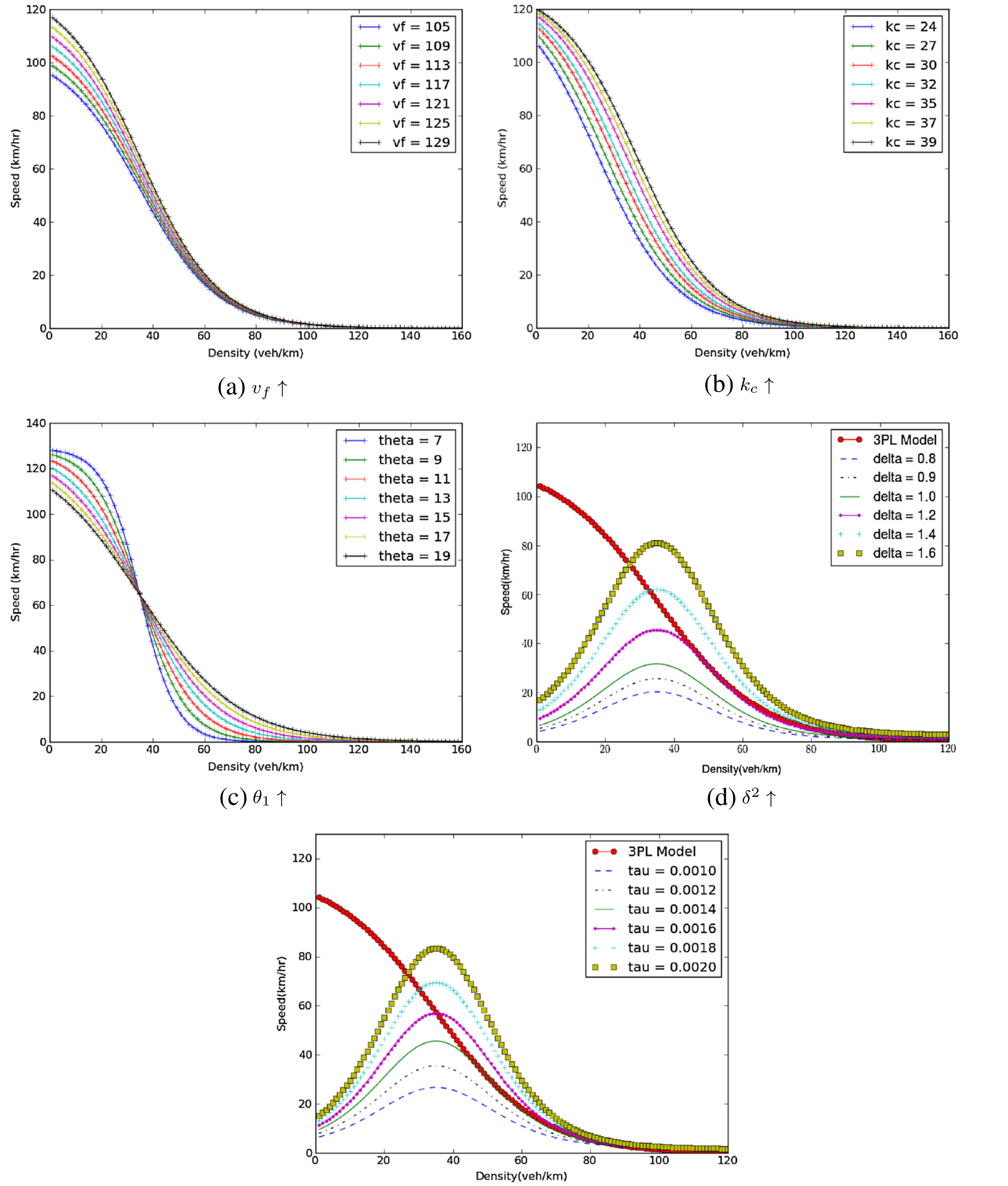
(c) $k_c \uparrow$ (d) $\theta_1 \uparrow$

(e) $\delta^2 \uparrow$ (f) $\tau \uparrow$

Figure 9. Effects of varying parameters v_f, v_b, k_c, and θ_1 on the four-parameter logistic speed–density model and its corresponding variance functions with varying parameters of τ and δ^2.

Table 1 listed the optimal model parameters for the proposed variance functions in Section 3.3, which are dependent on the series of logistic speed–density models [2]. Column S of Table 1 represents the station ID (i.e., detector stations). The model parameters in the speed–density model are obtained through an iterative least-squares procedure, whereas the two additional parameters in the variance function are obtained by a maximum likelihood estimation method because we assume that the error term ε_i is a Gaussian variable. For more details about the statistical estimation techniques, interested readers are referred to [1]. From the magnitude of the estimated parameters particularly δ^2 and τ, the authors observe that δ^2 is relatively stable, whereas τ suffers a large variation. The existence of a constant term in the variance model, in this case δ^2, can be explained in both practical and theoretical ways. The physical meaning of δ^2 is the maximum possible variance when traffic density is nearly 0 (corresponds to free-flow condition). This implies the fact that drivers from different driver groups

DOI: 10.1002/atr



(a) $v_f \uparrow$ (b) $k_c \uparrow$ (c) $\theta_1 \uparrow$ (d) $\delta^2 \uparrow$ (e) $\tau \uparrow$

Figure 10. Effects of varying parameters v_f, k_c, θ_1 on the three-parameter logistic speed–density (3PL) model and the corresponding variance function based on 3PL model with varying parameters: δ^2 and τ.

(aggressive or intimidate, old or young) have their own preferred free-flow speed, and this location-specific parameter is dependent on empirical data. To frame it in a more theoretical sense, the existence of this parameter can be verified by a likelihood ration test by expressing the hypothesis as $\{\delta^2 = 0\}$; the results signify a better fit when $\{\delta^2 \neq 0\}$ [1].

4.2. Analysis of results

In this section of the paper, the authors selected four data sets collected from detectors 4001119, 4001120, 4001122, and 4001123, respectively. Technically, all the 78 observations from the basic segments of GA400 in Figure 2 can be used in the experiments by varying the parameters in the speed–density relationship and the corresponding variance functions based on it. The optimal parameters

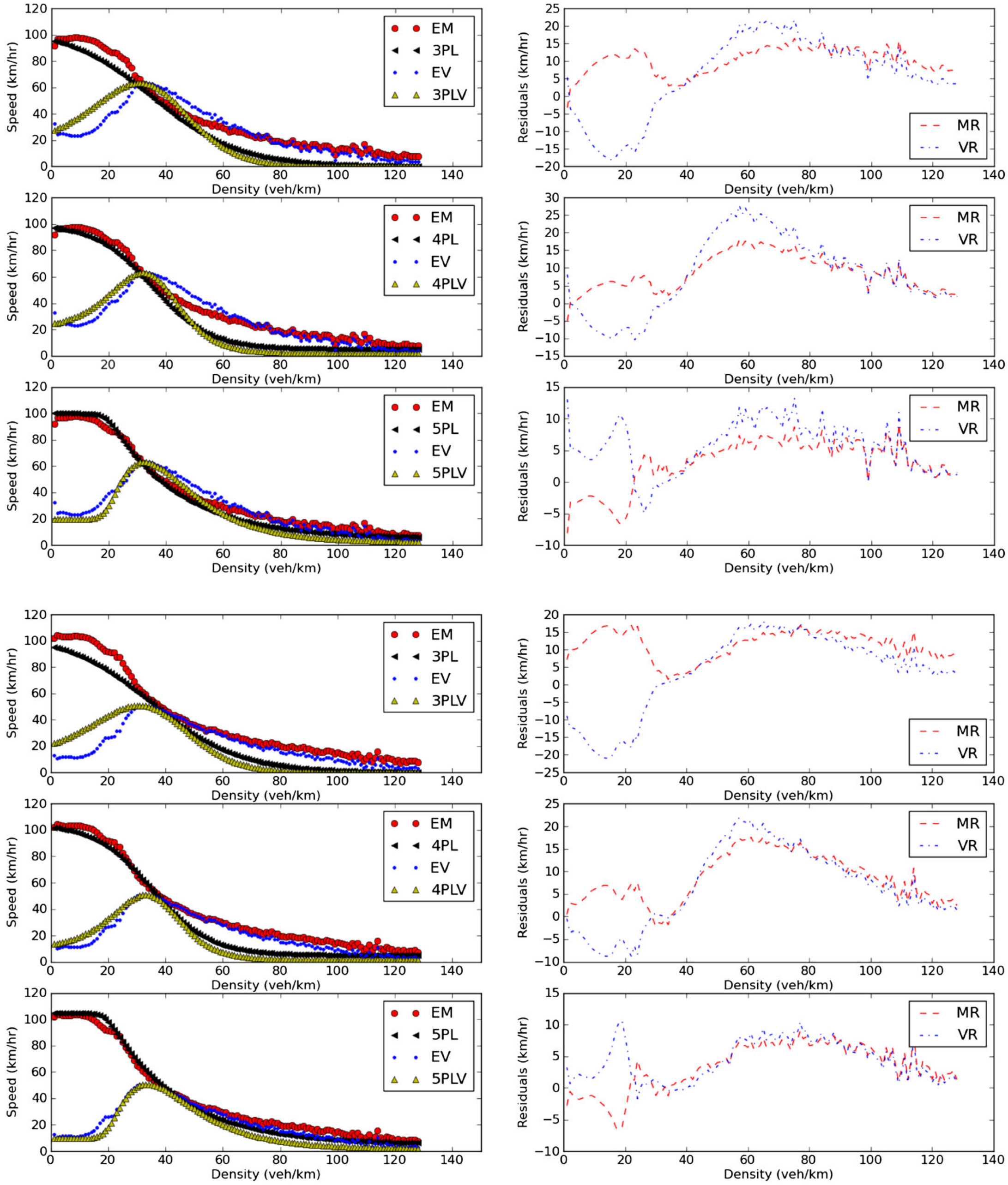


Figure 11. Performance of the three-parameter, four-parameter, and five-parameter logistic speed–density models against the empirical mean and residuals between their corresponding variance models and empirical variance at station 4001119 and 4001120.

in the speed–density curve and its corresponding variance function can be estimated by a least-square iterative procedure (this paper used the leastsq in scipy.optimize [27]). To evaluate the performance of the proposed traffic speed variance models, the authors will compare the 3PL, 4PL, and 5PL speed–density models to the empirical mean of a speed–density observation; simultaneously, the authors also compare the traffic speed variance functions based on the 3PL, 4PL, and 5PL models to the empirical traffic speed variances as can be seen from the left half in Figures 11 and 12. The EM represents the empirical mean of the scattered speed–density observation, whereas EV is the empirical traffic speed variance at this location. Similarly, the *n*PL ($n = 3, 4, 5$) represents the mean curve estimated from the *n*-parameter logistic speed–density models. On the right-hand side of Figures 11 and 12, the MR indicates the mean residual, which is the difference between the empirical mean and the mean

DOI: 10.1002/atr

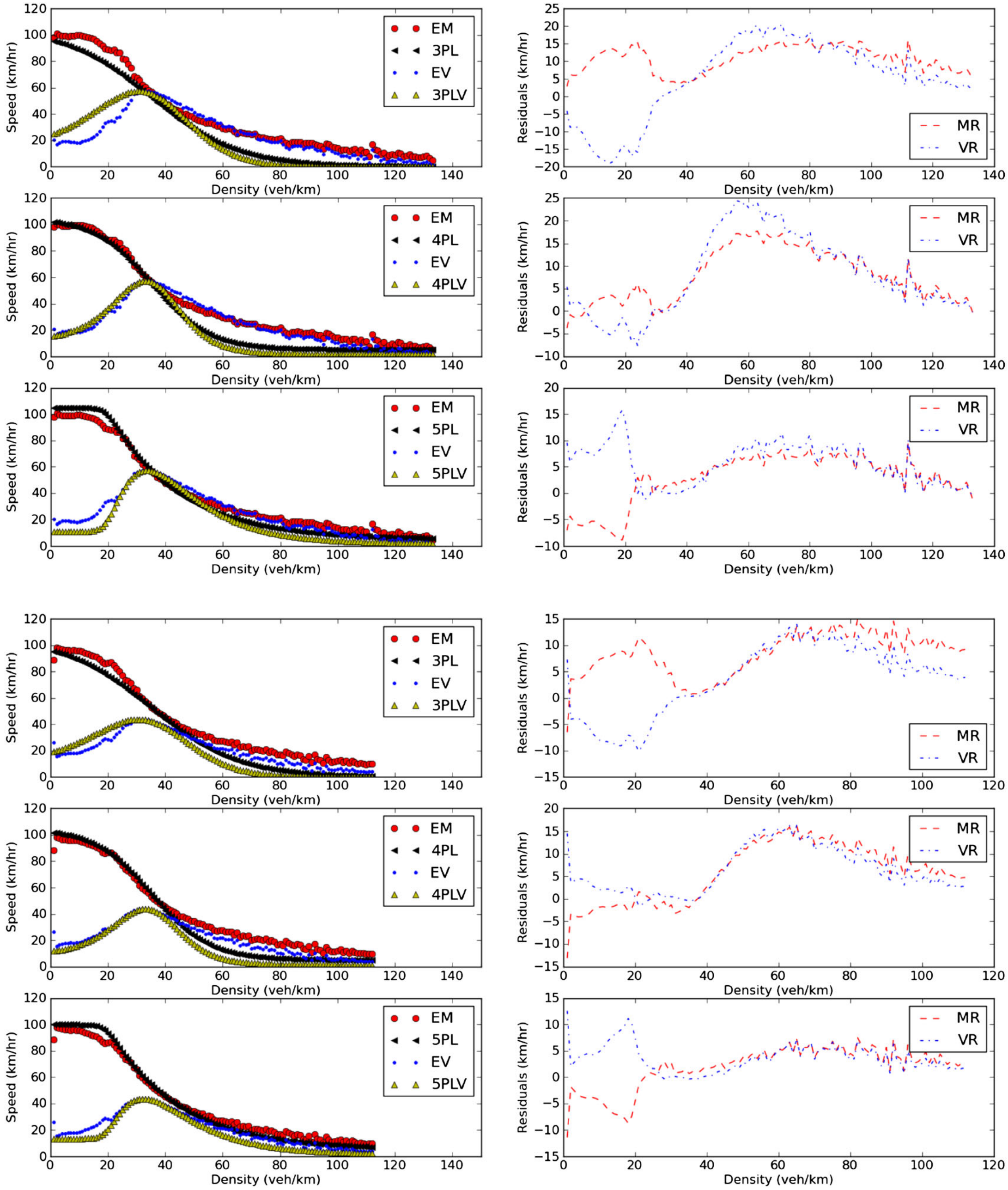


Figure 12. Performance of the three-parameter, four-parameter, and five-parameter logistic speed–density models against the empirical mean and residuals between their corresponding variance models and empirical variance at station 4001122 and 4001123.

estimated from the n-parameter logistic speed–density model; the VR is essentially the variance residuals, which is the difference between the empirical variance and the variance values estimated from the corresponding traffic speed variance functions proposed in Section 3.3. It is obvious from the results shown in Figures 11 and 12 that the 5PL model and its corresponding variance function performs better than other two models in terms of both residuals from the mean and variance. The performance of the logistic speed–density models fitting to the empirical mean has been provided in [2]. Therefore, the authors will focus on the comparison of the proposed traffic speed variance functions to the empirical traffic speed variance. Readers may have noticed that the empirical traffic speed variance we are comparing against with is different from the structured empirical traffic speed variance, which shows many kinks as can

DOI: 10.1002/atr

Table I. Estimated parameters for the variance function and the five-parameter logistic speed–density model (speed v_f in km/hour, density k_t in veh/km).

S	δ^2	τ	v_f	k_t	θ_1	θ_2	S	δ^2	τ	v_f	k_t	θ_1	θ_2
01	1.2	0.0014	107.44	17.53	1.8768	0.0871	25	1.2	0.005	96.09	20.04	3.6202	0.1323
02	1.2	0.004	99.92	16.12	2.1098	0.0947	26	1.3	0.003	99.93	20.34	3.0470	0.1269
03	1.4	0.010	106.89	14.40	1.7388	0.0714	27	1.3	0.04	96.14	22.89	4.5292	0.1941
04	2.4	0.008	47.52	32.61	0.1188	0.1835	28	1.3	0.006	99.88	14.30	0.2418	0.0106
05	1.3	0.004	86.57	24.39	1.0094	0.0401	29	1.1	0.003	106.80	16.95	2.4325	0.1059
06	1.6	0.003	92.71	21.72	3.9212	0.1835	30	1.6	0.005	91.04	22.21	3.2091	0.1138
07	1.4	0.002	99.39	21.26	3.8762	0.1928	31	1.8	0.002	88.99	28.77	5.1484	0.1499
08	1.5	0.013	95.06	20.33	3.15	0.1628	32	1.3	0.003	97.05	19.61	2.2104	0.0746
09	1.1	0.002	111.06	17.01	2.5501	0.1074	33	1.4	0.002	95.69	22.43	3.0685	0.1249
10	1.3	0.012	96.16	19.03	2.0220	0.0938	34	1.3	0.007	98.05	22.24	4.4141	0.1688
11	1.3	0.009	97.64	17.52	2.2787	0.0899	35	1.1	0.005	107.96	21.24	4.1009	0.1736
12	1.2	0.008	100.67	12.63	2.0386	0.0899	36	1.4	0.009	101.92	21.67	3.7766	0.1478
13	1.3	0.006	103.02	15.52	2.0674	0.0857	37	1.3	0.008	98.47	21.07	3.7207	0.1326
14	1.3	0.010	98.97	20.20	3.1219	0.1179	38	1.2	0.005	106.31	19.42	4.6129	0.1802
15	1.3	0.009	98.60	17.69	3.0240	0.1202	39	1.1	0.006	110.36	17.44	3.8358	0.2096
16	1.2	0.004	105.51	16.48	3.3903	0.1404	40	1.2	0.008	108.57	16.86	2.6591	0.1181
17	1.3	0.006	99.35	13.25	1.8755	0.0926	41	1.3	0.004	105.64	19.89	3.3156	0.1181
18	1.3	0.004	102.12	18.99	3.34	0.1231	42	1.3	0.008	105.40	18.67	3.8394	0.1387
19	1.4	0.007	98.08	19.97	3.53	0.1300	43	1.0	0.009	109.58	18.19	2.7535	0.1140
20	1.2	0.005	104.22	18.06	3.3054	0.1110	44	1.4	0.015	99.68	19.64	2.7885	0.0995
21	1.4	0.008	100.06	19.22	3.3051	0.1189	45	1.3	0.009	101.67	18.83	2.7745	0.0985
22	1.3	0.007	100.06	19.22	3.3051	0.1189	46	1.1	0.005	108.45	17.83	2.6356	0.1199
23	1.3	0.010	97.45	20.98	4.9820	0.1901	47	1.2	0.008	108.24	15.25	2.4505	0.1259
24	1.1	0.003	114.36	17.55	4.7015	0.1901	48	1.3	0.006	101.68	21.37	4.5757	0.1254
25	1.5	0.013	89.34	17.42	5.3515	0.2271	49	1.4	0.014	96.72	17.79	5.2587	0.1696
26	1.2	0.010	110.55	12.29	2.0450	0.0714	50	1.7	0.016	87.05	21.47	3.5007	0.1060
27	1.4	0.012	99.11	22.67	5.3573	0.1994	51	1.5	0.008	95.37	16.70	2.4823	0.1341
28	1.4	0.009	98.08	28.67	6.61	0.4005	52	1.3	0.012	102.50	15.05	1.7244	0.0810
29	1.3	0.008	104.20	22.27	4.82	0.1787	53	1.3	0.008	99.22	17.10	1.7835	0.0716
30	1.3	0.008	105.09	24.04	5.5045	0.2693	54	1.5	0.012	89.83	17.75	1.8373	0.0623
31	1.5	0.007	97.72	23.90	5.1731	0.2009	55	0.9	0.005	137.75	15.72	1.2974	0.0512
32	1.6	0.012	95.57	22.53	4.4708	0.1535	56	1.3	0.006	97.12	15.72	1.2974	0.0512
33	1.9	0.015	72.32	21.28	1.2734	0.0341	57	1.6	0.008	87.67	16.91	1.2206	0.0512
34	1.5	0.018	92.66	21.17	5.0568	0.1681	58	1.4	0.009	94.76	16.73	1.8321	0.0618
35	1.2	0.016	103.94	11.24	2.7827	0.0653	59	1.2	0.014	102.27	13.37	1.5524	0.0650
36	1.8	0.008	88.81	19.97	3.8557	0.1687	60	1.4	0.010	91.08	19.45	2.0860	0.0592
37	1.4	0.013	101.40	14.28	3.7376	0.1380	61	1.3	0.012	91.40	20.18	2.6633	0.0911
38	1.3	0.009	99.64	18.22	3.67	0.1413	62	1.3	0.008	94.35	15.14	1.8434	0.0604
39	1.4	0.011	102.99	17.87	3.45	0.1410	63	1.2	0.006	108.43	13.53	1.1753	0.0450

be seen from Figures 3–5. The proposed traffic speed variance function can be potentially applied to model travel time variability and relationship between speed variance and accident rates. The exploration of these applications of the proposed traffic speed variance models is actively pursued by the authors.

5. SUMMARY AND CONCLUSIONS

The primary motivation for this research originated from an empirical need to account for the traffic speed variances for the heterogeneous traffic flow systems. The second motivation stems from the fact that transportation professionals in general have a better understanding of the equilibrium speed–density relationship, but a limited knowledge of how traffic speed variance varies as traffic density increases from 0 to jam density because there is evidence that speed variance is somehow associated with road crash rates and so on. The authors have proposed a generalized variance function to model the structured empirical traffic speed variance. The variance function captures the nonlinear and heterogeneous nature of a parabola-shaped empirical variance. The proposed variance function has two features: it is dependent on the speed–density curve, and it contains two additional parameters that have to be set either as constants for simplification or to be estimated from empirical data.

DOI: 10.1002/atr

The major findings of this research are as follows:

(1) The structured traffic speed variance is the results of naturally occurring macroscopic traffic conditions. The empirical variance takes a parabolic shape that first increases to a local maximum and then decreases as traffic density increases.
(2) The pattern of structured traffic speed variance is different between weekdays and weekends. It is found that this pattern is consistent on either weekdays (from Monday to Friday) or weekends (Saturday and Sunday), but this proposed speed variance model works better for weekdays than weekends data.
(3) A parametric traffic speed variance function is used to model traffic speed variance in terms of its nonlinearity and heterogeneity. The model parameters are calibrated through empirical data.
(4) The proposed variance function matches the empirically observed traffic speed variances. In particular, the five-parameter logistic speed–density model and its corresponding speed variance function describe the empirical variance better than the three-parameter and four-parameter logistic speed–density models and their corresponding variance functions.

For future directions, the authors will address some of the limitations of the proposed traffic speed variance modeling framework, for example, removing the dependency on the speed–density relationship and reducing the number of model parameters in the variance function. One feature of the proposed variance functions is its continuity over the whole range of traffic density. Multiregime traffic speed variance functions or turning point models could be explored and compared with single-regime models. Obviously, a multiregime traffic speed variance model sacrifices the mathematical elegance (i.e., differentiability) for limited improvements on empirical accuracy. With the developed logistic speed–density models and their corresponding traffic speed variance functions, it is likely to develop a stochastic logistic speed–density model that incorporates both mean and variance into considerations.

As a logical next step, the authors will explore various potential applications of the proposed traffic speed variance models. One potential application is to establish the connection between traffic speed variance and the crash rates from the empirical data. However, the challenge for this research is the unavailability of the proper dataset. Another potential application is to use the proposed traffic speed variance model to evaluate travel time reliability either on a corridor or in a network.

ACKNOWLEDGEMENTS

The authors would like to thank the three anonymous reviewers for their thoughtful comments that certainly improved the readability and quality of the paper.

REFERENCES

1. Huet S, Bouvier A, Poursat M-A, Jolivet E. *Statistical Tools for Nonlinear Regression: a Pracital Guide with S-PLUS and R Examples*, 2nd edn. Springer: Springer-Verlag New York, Inc, 2004.
2. Wang H, Li J, Chen Q-Y, Ni D. Logistic modeling of equilibrium speed–density relationship, *Transportation Research Part A: Policy and Practice* 2011; **45**:554–566.
3. Wang H, Ni D, Chen Q-Y, Li J. Stochastic modeling of equilibrium speed–density relationship. *Journal of Advanced Transportation*. DOI:10.1002/atr.172
4. Park S, Ritchie SG. Exploring the relationship between freeway speed variance, lane changing, and vehicle heterogeneity. In *The 83rd Annual Meeting of Transportation Research Board*, Washington, DC, 2004.
5. Lave CA. Speeding, coordination, and the 55 mph limit. *American Economic Review* 1985; **75**(5):1159–1164.
6. Guo J, Smith BL. Short term speed variance forecasting using linear stochastic modeling of univariate traffic speed series. *Tech. Rep. UVACTS-15-17-10*, Center for Transportation Studies, University of Virginia, July 2007.
7. Dailey DJ. A statistical algorithm for estimating speed from single loop volume and occupancy measurements. *Transportation Research Part B: Methodological* 1999; **33**:313–322.
8. Cofiman B. Improved velocity estimation using single loop detectors. *Transportation Research Part A: Policy and Practice* 2001; **35**:863–880.
9. Varhelyi A, Makinen T. The effects of in-car speed limiters: field studies. *Transportation Research Part C: Emerging Technologies* 2001; **9**:191–211.
10. Son HD, Kweon Y-J, Park BB. Development of crash prediction models with individual vehicular data. *Transportation Research Part C: Emerging Technologies* 2011; **19**:1353–1363.
11. Ko J, Guensler RL, Hunter MP. An exploration of the relationship between roadway characteristics and speed variation. *Transportation Research Record: Journal of Transportation Research Board* 2009; **2092**:1–10.

12. Federal Highway Administration. *Speed Management Work Plan*. FHWA-National Highway Traffic Safety Administration: Washington, DC, April 1997.
13. Kang K-P, Chang G-L. A robust model for optimal time-of-day speed control at highway work zones. *IEEE Transactions on Intelligent Transportation Systems* 2006; **7**(1):115–123.
14. Lu F, Chen X. Analyzing the speed dispersion influence on traffic safety. *International Conference on Measuring Technology and Mechatronics Automation*, Zhangjiajie, China, 2009.
15. Vadeby A, Forsman A. Traffic safety and speed dispersion. *Tech. Rep.*, Swedish National Road and Transport Research Institute, 2012.
16. Aarts L, van Schagen I. Driving speed and the risk of road crashes: a review. *Accident Analysis & Prevention* 2006; **38**:215–224.
17. Graves PE, Lee DR, Sexton RL. Speed variance, enforcement, and the optimal speed limit. *Economics Letters* 1993; **42**(2–3):237–243.
18. Benjaafar S, Dooley K, Setyawan W. Cellular automata for traffic flow modeling, Technical Report CTS 97-09, *Intelligent Transportation Systems Institute* 1997.
19. Collins J, Fitzpatrick K, Bauer KM, Harwood DW. Speed variability on rural two-lane highways. *Journal of Transportation Research Record* 1999; **1658**:60–69.
20. Lu JJ. *Guideline Development for Minimizing Operating Speed Variance of Multilane Highways by Controlling Access Design*. University of Tennessee: Knoxville.
21. Rakha H, Zhang W. Estimating traffic stream space-mean speed and reliability from dual and single loop detectors. *Transportation Research Record* 2005; **1925**:38–47. DOI:10.3141/1925-05
22. Garber NJ, Gadiraju R. Factors affecting speed variance and its influence on accidents. *Journal of Transportation Research Record* 1989; **1213**:64–71.
23. Garber NJ, Gadiraju R. Impact of differential speed limits on the speed of traffic and the rate of accidents. *Journal of Transportation Research Record* 1992; **1375**:44–52.
24. Li J, Wang H, Chen Q-Y, Ni D. Analysis of LWR model with fundamental diagram sugject to uncertainties. *Transportmetrica* 2012; **8**(6):1–19.
25. Drake JS, Schofer JL, May AD. A statistical analysis of speed–density hypotheses. *Highway Research Record* 1967; **156**:53–87.
26. Ledvij M. Curve fitting made easy. *The Industrial Physicist* 2003. http://www.aip.org/tip/INPHFA/vol-9/iss-2/p24.pdf
27. Jones E, Oliphant T, Peterson P, *et al*. SciPy: open source scientific tools for Python 2001. http://www.scipy.org/ (Accessed on April 2010).