Model on empirically calibrating stochastic traffic flow fundamental diagram

Communications in Transportation Research 1 (2021) 100015

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Model on empirically calibrating stochastic traffic flow fundamental diagram

Shuaian Wang [a,**], Xinyuan Chen [b,***], Xiaobo Qu [c,*]

[a] Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hung Hom, Hong Kong, SAR, China
[b] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China
[c] Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden

ARTICLE INFO

Keywords:
Stochastic fundamental diagram
Speed distributions
Traffic control

ABSTRACT

This paper addresses two shortcomings of the data-driven stochastic fundamental diagram for freeway traffic. The first shortcoming is related to the least-squares methods which have been widely used in establishing traffic flow fundamental diagrams. We argue that these methods are not suitable to generate the percentile-based stochastic fundamental diagrams, because the results generated by least-squares methods represent weighted sample mean, rather than percentile. The second shortcoming is widespread use of independent modeling methodology for a family of percentile-based fundamental diagrams. Existing methods are inadequate to coordinate the fundamental diagrams in the same family, and consequently, are not in alignment with the basic rules in probability theory and statistics. To address these issues, this paper proposes a holistic modeling framework based on the concept of mean absolute error minimization. The established model is convex, but non-differentiable. To efficiently implement the proposed methodology, we further reformulate this model as a linear programming problem which could be solved by the state-of-the-art solvers. Experimental results using real-world traffic flow data validate the proposed method.

1. Introduction

Traffic fundamental diagram has been considered as the basis of traffic flow theory (Edie, 1961; Greenberg, 1959; Greenshields, 1935; Underwood, 1961). It describes the relationship between two of the three basic traffic flow parameters: traffic flow (veh/h), speed (km/h), and traffic density (veh/km), and the third one is directly obtained from the fundamental equation of traffic. The seminal work in this field was proposed by Greenshields (1935), who first established a linear model to describe the macroscopic relationship between speed and density. Since then, numerous extensions have been made to better reflect the traffic flow features, e.g., Greenberg (1959), Underwood (1961), Newell (1961), Kerner and Konhäuser (1994), Castillo and Benítez (1995a), Castillo and Benítez (1995b), Li and Zhang (2011), Wu et al. (2011), Keyvan-Ekbatani et al. (2012), Keyvan-Ekbatani et al. (2013), Wang et al. (2011), Wang et al. (2013), Qu et al. (2015), and Ni et al. (2016). The main purpose of these studies is to develop an empirically accurate single-regime model with a small number of meaningful variables for the sake of mathematical elegance and analytical simplicity.

Modeling the stochastic fundamental diagram is an active area of research. Stochastic fundamental diagrams can better depict the traffic flow relations which are essential to develop advanced traffic flow models and traffic management strategies (Zhou and Zhu, 2020). In existing studies, the stochasticity is mainly modeled by adding noise to existing deterministic relations (Wang et al., 2013; Muralidharan et al., 2011) or deriving the probabilistic macroscopic relations from microscopic car-following models (Jabari et al., 2014). Recently, data-driven approaches have attracted attention (Fan and Seibold, 2013; Nikolić et al., 2016, 2019; Qu et al., 2017). Compared to other approaches, data-driven approaches can better exploit the features of traffic flow from real-world data which is obtainable through a variety of facilities, e.g., sensors, and probe vehicles. Nikolić et al. (2016) and Nikolić et al. (2019) study the data-driven probabilistic modeling approach for pedestrian speed-density relations. Fan and Seibold (2013) and Qu et al. (2017) are only published research works that use data-driven approaches to generate percentile-based speed-density relations for freeway traffic. Fan and Seibold (2013) propose two data fitted models based on a specific traffic flow model. Qu et al. (2017) propose a general percentile-based

*Corresponding author. Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden.
** Corresponding author.
*** Corresponding author.
E-mail addresses: wangshuaian@gmail.com (S. Wang), xinyuan.chen@nuaa.edu.cn (X. Chen), xiaobo@chalmers.se (X. Qu).

https://doi.org/10.1016/j.commtr.2021.100015
Received 18 November 2021; Received in revised form 20 November 2021; Accepted 20 November 2021
Available online 9 December 2021

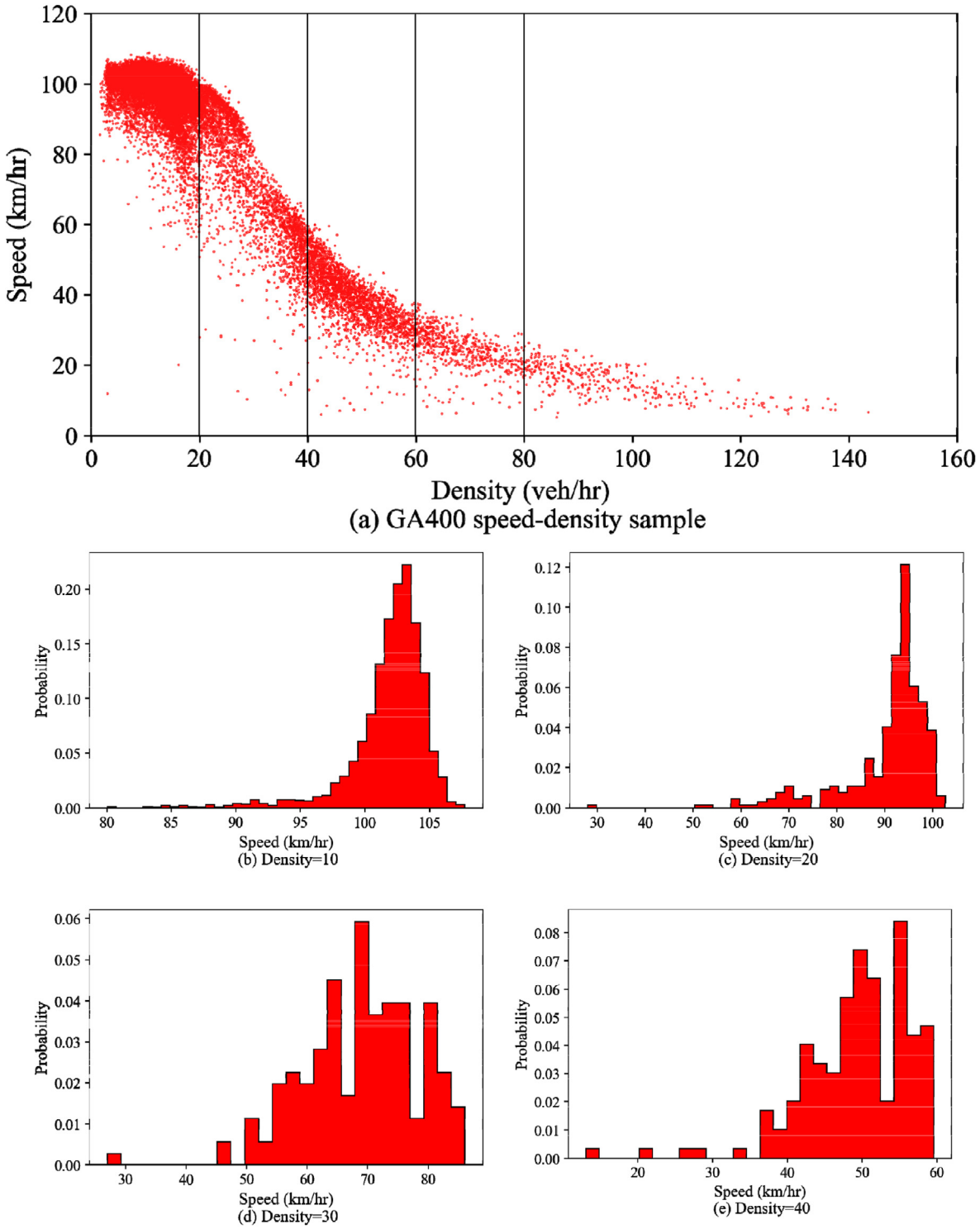

Fig. 1. GA400 speed-density sample.

data-driven approach to establish stochastic speed-density relations that applies to a class of traffic flow models. The percentile-based fundamental diagrams are generated through a weighted least-squares programming problem by assigning different weights to the squared errors in the objective function. Then, for any given density, the corresponding percentile-based speeds can be calculated through the calibrated percentile-based fundamental diagrams. Based on the percentile-based speeds, the cumulative distribution function (CDF) and probability density function (PDF) of speeds at any given density are obtainable.

In this study, we identify two major limitations in existing percentile-based approaches for freeway traffic. First, using least-squares technique leads to less accurate percentile-based fundamental diagrams. The least-squares technique is widely used in transportation studies to reflect a central tendency of a probability distribution. However, the results estimated by the least-squares technique represent the weighted sample mean, rather than sample percentile (details will be described in Section 2.1), and therefore, conflict with the pursuit of the percentile-based fundamental diagrams. Moreover, although sample mean could reflect

the central tendency, its performance is not good under a skewed distribution. The value of sample mean can be easily skewed by the outliers, which results in an unrealistic estimation of central tendency. We note that the distributions of speed are generally skewed. Fig. 1(a) shows the speed-density sample collected by loop detectors from 76 stations on Georgia State Route 400 for continuous observation of one year (referred to as GA400 dataset hereafter). This dataset has been widely used in calibrating and validating traffic flow fundamental diagrams (Wang et al., 2011, 2013; Qu et al., 2015, 2017; Zhang et al., 2018). Given a density, the empirical distribution of speed is obtainable. Fig. 1(b)–(e) show four specific empirical distributions of speed when density equals 10, 20, 30, and 40 (veh/km), which demonstrate the distributions of speed are typically left-skewed. Under this situation, median, i.e., 50th percentile, is a better measurement of the central tendency, because the value of percentile does not depend on all the values in the sample and is unlikely to be skewed by the outliers. When some of the values are more extreme, the effect on the median is smaller. Therefore, to generate the percentile-based fundamental diagrams, a new mathematical modeling approach is needed to replace the conventionally adopted least-squares methods.

Second, the proposed percentile-based calibration approach may generate unrealistic probability estimations. The purpose of percentile-based calibration is to generate the distribution of the random variable of speed. A natural way is to set percentile as 5th, 10th, 15th, … 95th, and for each specific percentile, to solve one optimization model to generate a speed-density curve. However, due to scattering features of speed-density plots, the generated speed-density curves can intersect with each other, e.g., the generated 5th percentile of speed is larger than the generated 10th percentile. This phenomenon violates the basic rule defined in probability theory and statistics, i.e., the generated 5th percentile speed should be no larger than the 10th percentile speed.

1.1. Objectives and contributions

The objective of this study is to propose a new methodology to address these two shortcomings that exist in the aforementioned data-driven stochastic fundamental diagram modeling approaches. In particular, to overcome the first issue, we apply the percentile regression technique in statistics to generate sample percentiles from a given empirical distribution. Then, we proceed to propose a basic model to generate the percentile-based speed-density relations. To address the second issue, the basic model is further extended to a holistic model that coordinates a family of speed-density curves. A transformation is proposed to transform the proposed holistic model into the linear program so that it can be addressed efficiently by the state-of-the-art solvers. Numerical experiments based on GA400 data validate the proposed methodology.

The rest of the paper is organized as follows. In Section 2, we present a new data-driven methodology to establish the percentile-based stochastic fundamental diagrams. A new general modeling framework is presented which is generalizable to deal with the stochasticity of traffic flow using a variety of traffic flow models. This is followed by a case study in Section 3 to demonstrate the applicability and validity of the proposed methodology. Section 4 concludes this study.

2. Methodology

2.1. Percentile via weighted mean absolute error minimization

The percentile-based speed-density model has been proposed to offer a comprehensive picture of the speed-density relationship and reflect the stochasticity of fundamental diagrams. Conventionally, the percentile-based speed-density curves are calculated based on the concept of least-squares. In this study, we note that the least-squares method could result in biased estimations for percentile-based speed-density models. This is because in least-squares methods, the expected loss function is defined by squared errors. Then, the solution of these methods is the weighted sample mean, rather than the sample percentile. We take the sample mean and median as an example to illustrate this issue. It has been shown in statistics (Davino et al., 2013) that the sample mean is generated from the mean squared error (MSE) minimization problem, while the sample median is calculated from the mean absolute error (MAE) minimization problem. To make the paper self-contained, we briefly introduce the relevant theories below.

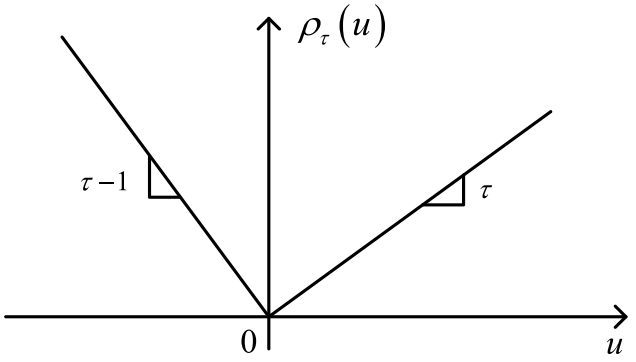

Fig. 2. Loss function.$\rho_\tau(u)$

Consider a random speed V with cumulative distribution function $F_V(v) = P(V \le v)$. The mean $\bar{v}$ of random speed V can be calculated by solving the following MSE minimization problem.

$$\min_{\bar{v}} E(\bar{v}) = \min_{\bar{v}} \int_{\mathbb{R}} (v-\bar{v})^2 dF_V(v). \tag{1}$$

The objective function $E(\bar{v})$ is defined as a Lebesgue integral of squared errors with respect to (w.r.t.) the cumulative distribution function $F_V(v)$. Differentiating problem (1) w.r.t. $\bar{v}$, we have:

$$0 = \int_{\mathbb{R}} (v-\bar{v}) dF_V(v) = \int_{\mathbb{R}} v dF_V(v) - \bar{v}\int_{\mathbb{R}} dF_V(v) = \int_{\mathbb{R}} v dF_V(v) - \bar{v}. \tag{2}$$

Re-arranging Eq. (2) gives:

$$\bar{v} = \int_{\mathbb{R}} v dF_V(v). \tag{3}$$

Eq. (3) is exactly the definition of the mean of random variable V. Let $(V_1 \ldots V_n)$ be independent, identically distributed real random variables with the common cumulative distribution function $F_V(v)$, and 1_A be the indicator of event A. Let the cumulative distribution function $F_V(v)$ be replaced by the empirical distribution function $\widehat{F}_n(v)$:

$$\widehat{F}_n(v) = \frac{1}{n}\sum_{i=1}^{n} 1_{V_i \le v} \tag{4}$$

where n denotes the total number of observations in the sample. The corresponding MSE minimization problem (1) becomes:

$$\min_{\bar{v}} \widehat{E}(\bar{v}) = \min_{\bar{v}} \frac{1}{n}\sum_{i=1}^{n} (v_i - \bar{v})^2 \tag{5}$$

where v_i is the observation of random variable V_i. The objective function of (5), denoted as $\widehat{E}(\bar{v})$, becomes the sum of squared errors divided by the size of sample. The optimization problem (5) meets the definition of ordinary least squares problem. The optimal solution of (5) is also equal to the definition of the sample mean $\frac{1}{n}\sum_{i=1}^{n} v_i$. Therefore, the least-squares methods are suitable to estimate the mean speed.

To estimate the percentiles, the MAE minimization is needed. Let $\tilde{v}$ be the 100τth percentile of the random speed V, $\tilde{v} = F_V^{-1}(\tau) = \inf\{v : F_V(v) \ge \tau\}$, $\tau \in (0,1)$. Define the loss function as $\rho_\tau(u) = u \cdot (\tau - \mathbf{1}_{(u<0)})$, where u is any real number and $\mathbf{1}_{(u<0)}$ is an indicator:

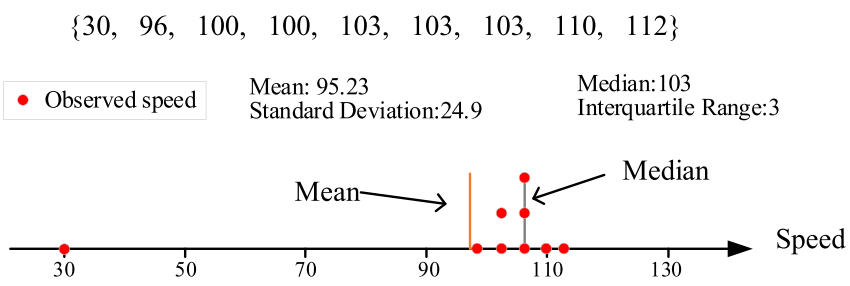

Fig. 3. An illustrative example (unit of speed: km/h).

$$\mathbf{1}_{(u<0)} := \begin{cases} 1, & if\ u < 0 \\ 0, & if\ u \geq 0. \end{cases} \tag{6}$$

As shown in Fig. 2, $\rho_\tau(u)$ is a piecewise linear function. In this study, u is the residual between v and $\tilde{v}$, $u = v - \tilde{v}$. Given any $\tau \in (0,1)$, the positive orthant part of $\rho_\tau(u)$ is described by a linear function with the slope of τ and intercept of 0, and the negative orthant part of $\rho_\tau(u)$ is described by another linear function with the slope of $\tau - 1$ and intercept of 0.

A specific 100τth percentile $\tilde{v}$ can be obtained by minimizing the MAE of u.

$$\begin{aligned} \min_u E_\tau(\rho_\tau(u)) &= \min_{\tilde{v}} \int_{\mathbb{R}} \rho_\tau(v-\tilde{v}) dF_V(v) \\ &= \min_{\tilde{v}} \left\{ (\tau-1) \int_{-\infty}^{\tilde{v}} (v-\tilde{v}) dF_V(v) + \tau \int_{\tilde{v}}^{+\infty} (v-\tilde{v}) dF_V(v) \right\} \end{aligned} \tag{7}$$

where $E_\tau(\rho_\tau(u))$ is a Lebesgue integral of $\rho_\tau(u)$ w.r.t. the cumulative distribution function $F_V(v)$. Differentiating model (7) w.r.t. $\tilde{v}$, we have:

$$0 = (1-\tau) \int_{-\infty}^{\tilde{v}} dF_V(v) - \tau \int_{\tilde{v}}^{+\infty} dF_V(v) = F_V(\tilde{v}) - \tau. \tag{8}$$

Re-arranging Eq. (8) gives:

$$\tau = F_V(\tilde{v}). \tag{9}$$

Hence, $\tilde{v} = F_V^{-1}(\tau)$ which meets the definition of 100τth percentile of random speed V. When the $F_V(v)$ is replaced by the empirical distribution function (4), The MAE minimization problem (7) becomes:

$$\min_u \widehat{E}_\tau(\rho_\tau(u)) = \min_u \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(u) = \min_{\tilde{v}} \frac{1}{n} \left[(\tau-1) \sum_{v_i < \tilde{v}} (v_i - \tilde{v}) + \tau \sum_{v_i \geq \tilde{v}} (v_i - \tilde{v}) \right]. \tag{10}$$

The objective function $\widehat{E}_\tau(\rho_\tau(u))$ becomes the sum of $\rho_\tau(u)$ divided by the size of sample. The optimal solution $\tilde{v}$ of the MAE minimization problem (10) is the sample 100τth percentile. The intuition is the same as for the population quantile. Given that τ is taken as weight w.r.t. the positive part of errors and $\tau - 1$ is the weight w.r.t. the negative part of errors, the optimization problem (10) is actually a weighted mean absolute error, i.e. WMAE, minimization problem. In this regard, the notion of an ordering of the sample observations or the problem of finding the sample 100τth percentile is equivalent to the WMAE minimization problem. Therefore, we can conclude that the WMAE is suitable to estimate the percentiles.

2.1.1. An illustrative example

We use a small example in Fig. 3 to illustrate the difference between the mean squares and absolute error minimizations. Suppose $\mathbf{x}$ is a set of speed (km/h) observed from the same density level, that is $\mathbf{x} = \{30, 96, 100, 100, 103, 103, 103, 110, 112\}$. We take the sample mean and median as an example to obtain some insights. The corresponding mean squares minimization problem can be written as:

$$\min_{\bar{x}} \frac{1}{9} \left[(\bar{x}-30)^2 + (\bar{x}-96)^2 + 2(\bar{x}-100)^2 + 3(\bar{x}-103)^2 + (\bar{x}-110)^2 + (\bar{x}-112)^2 \right]. \tag{11}$$

The optimal solution of (11) is $\bar{x} = 95.2$, which equals the sample mean of $\mathbf{x}$. The mean absolute error minimization problem is written as:

$$\min_{\tilde{x}} \frac{1}{9} \left[|\tilde{x}-30| + |\tilde{x}-96| + 2|\tilde{x}-100| + 3|\tilde{x}-103| + |\tilde{x}-110| + |\tilde{x}-112| \right]. \tag{12}$$

The optimal solution of (12) is $\tilde{x} = 103$, which equals the sample median of $\mathbf{x}$. The results demonstrate the least-squares method reports the mean, rather than the percentile of a given distribution. More importantly, Fig. 3 reveals that the mean could be skewed significantly by the outliers.[1] The mean $\bar{x}$ is lower than all of the speed points except for one since the outlier is so far from the rest of the distribution that has skewed the mean. We note that this phenomenon is common in practice, as shown in Fig. 1. In this situation, median is more robust to estimate the central tendency with a skewed distribution. The outliers would not change the median, since the median is not related to the specific value of outliers. Therefore, a new percentile-based stochastic fundamental diagram should be established based on the WMAE.

2.2. Percentile-based stochastic speed-density models

2.2.1. Basic percentile-based stochastic speed-density model

Section 2.1 shows that the percentiles can be expressed as the solution of a WMAE minimization problem. In this section, we apply this idea to calibrate a family of percentile-based speed-density curves. We use Greenshield's linear model as an example to explain this new model. Suppose the observed data is $(\mathbf{k}_{data}, \mathbf{v}_{data}) = \{(k_i, v_i), i = 1, \ldots, n\}$ and $\tilde{v}_\tau(k)$ is the calibrated 100τth percentile of speed when density equals k. The free-flow speed v_f^τ and the jam density k_{jam}^τ are two type parameters to be calibrated. The objective of the new model is to calibrate the 100τth percentile-based speed-density curve. Mathematically, τ is defined as:

$$\tau = \frac{\sum_{i=1}^{n} g_\tau(k_i, v_i)}{\sum_{i=1}^{n} g_\tau(k_i, v_i) + \sum_{i=1}^{n} h_\tau(k_i, v_i)} \tag{13}$$

and

$$g_\tau(k_i, v_i) = \begin{cases} \tilde{v}_\tau(k_i) - v_i, \tilde{v}_\tau(k_i) > v_i \\ 0, \text{otherwise} \end{cases} \tag{14}$$

$$h_\tau(k_i, v_i) = \begin{cases} v_i - \tilde{v}_\tau(k_i), \tilde{v}_i \geq v_\tau(k_i) \\ 0, \text{otherwise.} \end{cases} \tag{15}$$

[1] Note that although we call them "outliers", they are not wrong records and should not be simply deleted. For instance, in transportation engineering, these "outliers" may represent extremely congested situations and should not be removed.

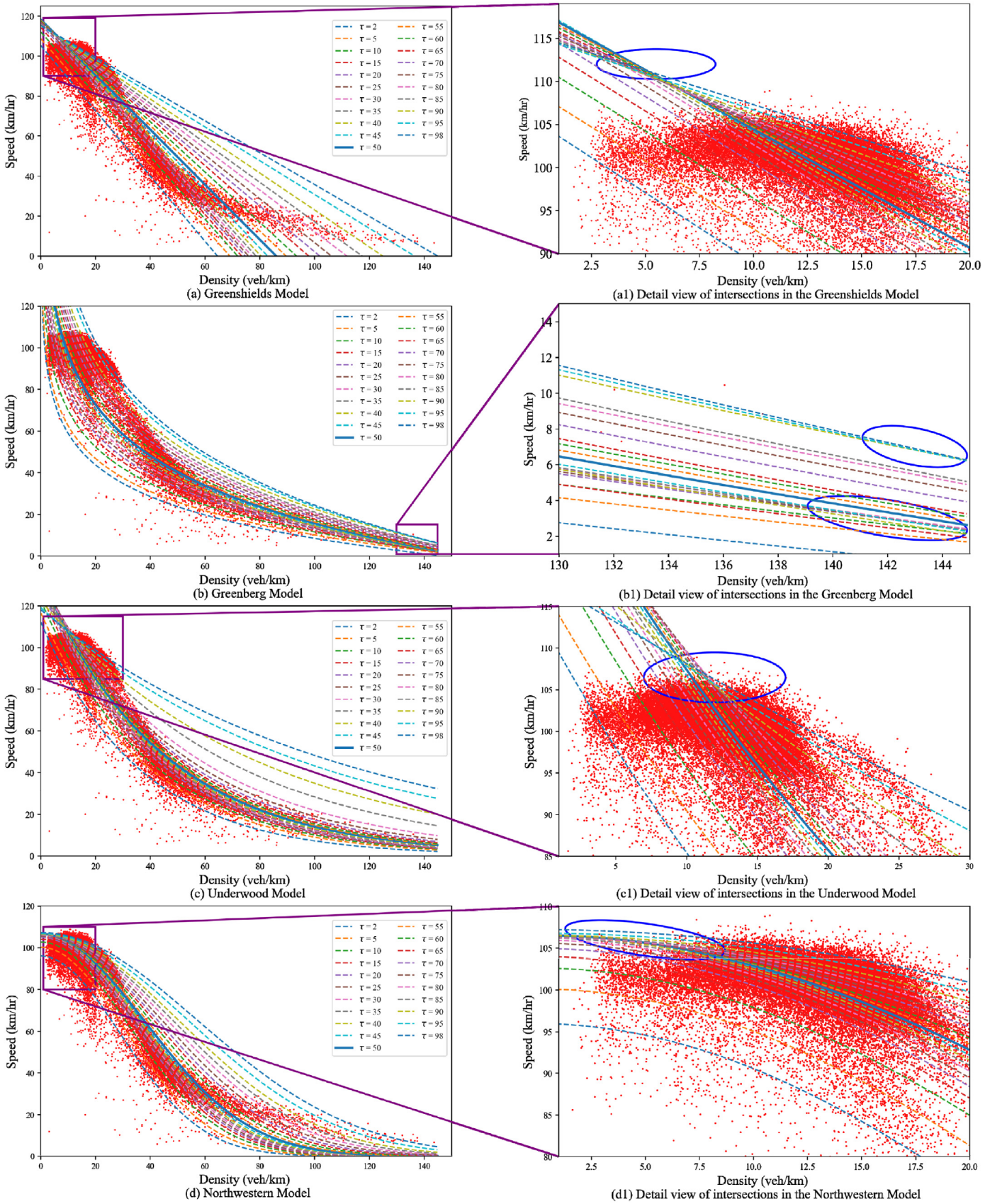

Fig. 4. Family of speed-density curves.

The 100τth percentile-based speed-density calibration model can be formulated as follows:

[M1]

$$\min_{k_{jam},v_f} \widehat{E}\left(\rho_\tau\left(k_{jam},v_f\right)\right)=\min_{k_{jam},v_f}\frac{1}{n}\left[(1-\tau)\sum_{i=1}^{n}\varpi_i g_\tau(k_i,v_i)+\tau\sum_{i=1}^{n}\varpi_i h_\tau(k_i,v_i)\right] \tag{16}$$

subject to

$$\tilde{v}_\tau(k_i)=v_f^\tau\left(1-\frac{k_i}{k_{jam}^\tau}\right), i=1,\ldots,n \tag{17}$$

$$v_f^\tau,\ k_{jam}^\tau>0 \tag{18}$$

and constraints (14), (15), where ϖ_i is the weight for observation (v_i,k_i) which can handle the sample selection bias and improve the calibration accuracy for some models (see Qu et al., 2015 for a method of determining ϖ_i). The objective (16) minimizes the WMAE of $v_i-\tilde{v}_\tau(k_i)$. Constraints (17) define the Greenshield's model. Constraints (14) and (15) represent the absolute error between v_i and $\tilde{v}_\tau(k_i)$. Constraints (18) are non-negative constraints.

It can be shown according to quantile regression theory (Davino et al., 2013) that Eq. (13) is satisfied at the optimal solution to the model (14)–(18) and therefore we can solve the model (14)–(18) to calibrate the 100τth percentile-based speed-density curve. This means that the fundamental diagram calibrated by model (16)–(18) is the 100τth percentile-based speed-density curve based on Greenshield's model. By changing τ between 0 and 1, the stochastic fundamental diagram is obtainable. Similarly, if we replace Eq. (17) by other functions, the corresponding optimization models can be established to generate the 100τth percentile-based speed-density curve w.r.t. other speed-density models.

2.2.2. Holistic percentile-based stochastic speed-density model

The basic percentile-based stochastic speed-density model can obtain the distribution of speed. However, different curves may intersect at some points (unless all the lines are parallel), which means for a particular density k, the predicted lower, e.g., the 5th percentile of $\tilde{v}$, is larger than the predicted higher, e.g. 10th percentile, which is logically wrong. To overcome this problem, we set a domain of interest $[k_{\min},k_{\max}]$ a priori and propose a new holistic percentile-based stochastic speed-density model. Within the domain of interest, a set of constraints is imposed on the percentile-based curves to ensure the predicted lower percentile value is no larger than that of the predicted higher percentile. For the freeway traffic, the traffic states we aim to study range from the free flow to the traffic jam. Therefore, the lower bound of the domain equals zero, $k_{\min}=0$ and the upper bound of the domain is the jam density, $k_{\max}=k_{jam}$, which can be determined empirically from the speed-density plots. Let m be the number of percentiles to be calibrated, $\boldsymbol{\tau}=(\tau_1,\ldots\tau_m)$ be the vector of τ_i, and $\tilde{\mathbf{v}}=(\tilde{v}_{\tau_1},\ldots,\tilde{v}_{\tau_m})$ be the vector of calibrated percentile-based speed-density models. For arbitrary $i\in\{1,\ldots,m\}$, $\tilde{v}_{\tau_i}$ represents 100τ_ith percentile-based speed-density model. The vectors of jam density $\mathbf{k}_{jam}=(k_{jam}^1,\ldots,k_{jam}^m)$ and free-flow speed $\mathbf{v}_f=(v_f^1,\ldots,v_f^m)$ are the parameters to be calibrated. Then, the holistic percentile-based stochastic speed-density model can be formulated as follows:

[M2]

$$\min_{\mathbf{k}_{jam},\mathbf{v}_f} Z\left(\mathbf{k}_{jam},\mathbf{v}_f\right)=\min_{\mathbf{k}_{jam},\mathbf{v}_f}\left\{\frac{1}{n}\left[\sum_{j=1}^{m}\left((1-\tau_j)\sum_{i=1}^{n}\varpi_i g_{\tau_j}(k_i,v_i)\right)+\sum_{j=1}^{m}\left(\tau_j\sum_{i=1}^{n}\varpi_i h_{\tau_j}(k_i,v_i)\right)\right]\right\} \tag{19}$$

subject to

$$\tilde{v}_{\tau_j}(k_{\min})\le\tilde{v}_{\tau_{j+1}}(k_{\min}),\forall j=1,\ldots m-1 \tag{20}$$

$$\tilde{v}_{\tau_j}(k_{\max})\le\tilde{v}_{\tau_{j+1}}(k_{\max}),\forall j=1,\ldots m-1 \tag{21}$$

$$\tilde{v}_{\tau_j}(k_i)=v_f^j\left(1-\frac{k_i}{k_{jam}^j}\right),\forall i=1,\ldots,n,j=1,\ldots,m \tag{22}$$

$$g_{\tau_j}(k_i,v_i)=\begin{cases}\tilde{v}_{\tau_j}(k_i)-v_i,\tilde{v}_{\tau_j}(k_i)>v_i\\0,\text{otherwise}\end{cases},\forall i=1,\ldots,n,j=1,\ldots,m \tag{23}$$

$$h_{\tau_j}(k_i,v_i)=\begin{cases}v_i-\tilde{v}_{\tau_j}(k_i),v_i\ge\tilde{v}_{\tau_j}(k_i)\\0,\text{otherwise}\end{cases},\forall i=1,\ldots,n,j=1,\ldots,m \tag{24}$$

$$\mathbf{v}_f,\mathbf{k}_{jam}>0. \tag{25}$$

The objective (19) integrates a family of speed-density models $\tilde{v}_{\tau_j},\forall j=1,\ldots,m$. Constraints (20) and (21) ensure the calibrated speed-density model v_{τ_j} is smaller than $v_{\tau_{j+1}}$ on the interval $[k_{\min},k_{\max}]$. Constraints (22) represent a family of speed-density model with unknown parameters. Constraints (23) and (24) define the absolute error function. Constraints (25) are non-negative constraints.

2.3. Model transformation

The proposed [M2] cannot be directly addressed. Therefore, we proceed to reformulate it as a linear programming problem so that it can be efficiently solved. We introduce $2nm$ artificial variables $\mathbf{p}=(p_{ij},i=1,\ldots,n;j=1,\ldots,m)$ and $\mathbf{q}=(q_{ij},i=1,\ldots,n;j=1,\ldots,m)$ to represent the positive and negative parts of the vector of errors. Then, [M2] is equivalent to:

[M3]

$$\min_{\mathbf{k}_{jam},\mathbf{v}_f,\mathbf{p},\mathbf{q}}=\left\{\frac{1}{n}\left[\sum_{j=1}^{m}\left((1-\tau_j)\sum_{i=1}^{n}\varpi_i p_{ij}\right)+\sum_{j=1}^{m}\left(\tau_j\sum_{i=1}^{n}\varpi_i q_{ij}\right)\right]\right\} \tag{26}$$

subject to

$$\tilde{v}_{\tau_j}(k_i)-p_{ij}+q_{ij}=v_i,\forall i=1,\ldots,n,j=1,\ldots,m \tag{27}$$

$$p_{ij},q_{ij}\ge0,\forall i=1,\ldots,n,j=1,\ldots,m \tag{28}$$

and constraints (20), (21), (22), and (25). Constraints (27) are the complementary slackness condition. Constraints (28) are non-negative constraints. [M3] is a linear programming problem, and therefore, can be solved by state-of-the-art solvers.

We note that [M3] is a general modeling framework that can be extended to various cases, such as the nonlinear models by some linearization methods (Qu et al., 2017).

3. Case study

The GA400 data has been widely used in the research of traffic flow fundamental diagram, e.g., Wang et al. (2011), Wang et al. (2013), Qu et al. (2015), Qu et al. (2017), and Zhang et al. (2018). As a follow up research, we use the same data set as these studies. By selecting different percentile values α, a family of percentile-based speed-density curves can be generated through [M1]. Fig. 4 shows the curves w.r.t. the Greenshields model (1935), Greenberg model (1959), Underwood model (1961), and Northwestern model (1967). Weights are considered during the calibration of Greenberg model to improve the performance (Qu et al., 2015). The red dots are the GA400 data. The thick solid curve represents the models which percentile value α is equal to 50. The other dash curves represent speed-density models with respect to $\alpha=2$, 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 98.

Based on these percentile-based curves, the stochastic fundamental

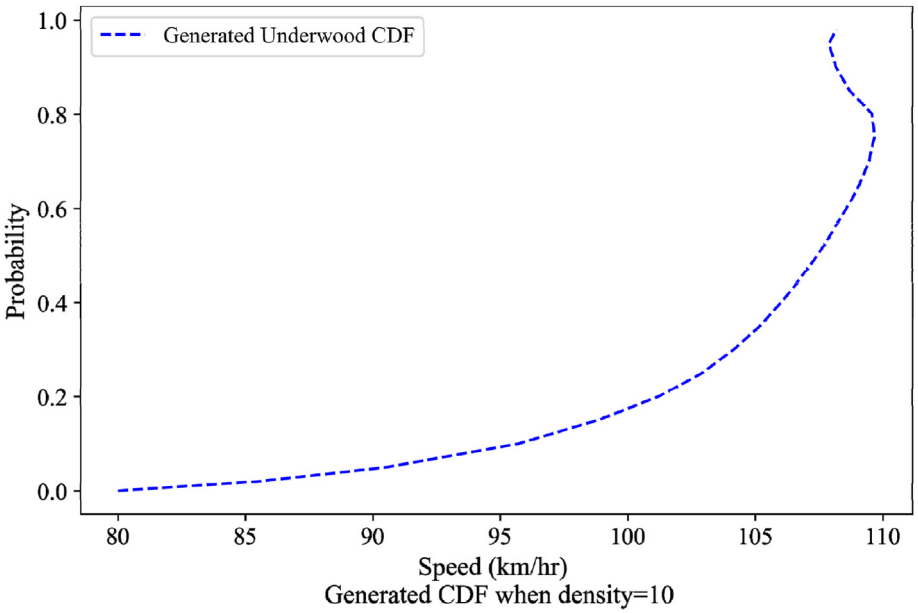

Fig. 5. Generated CDF of Underwood model.

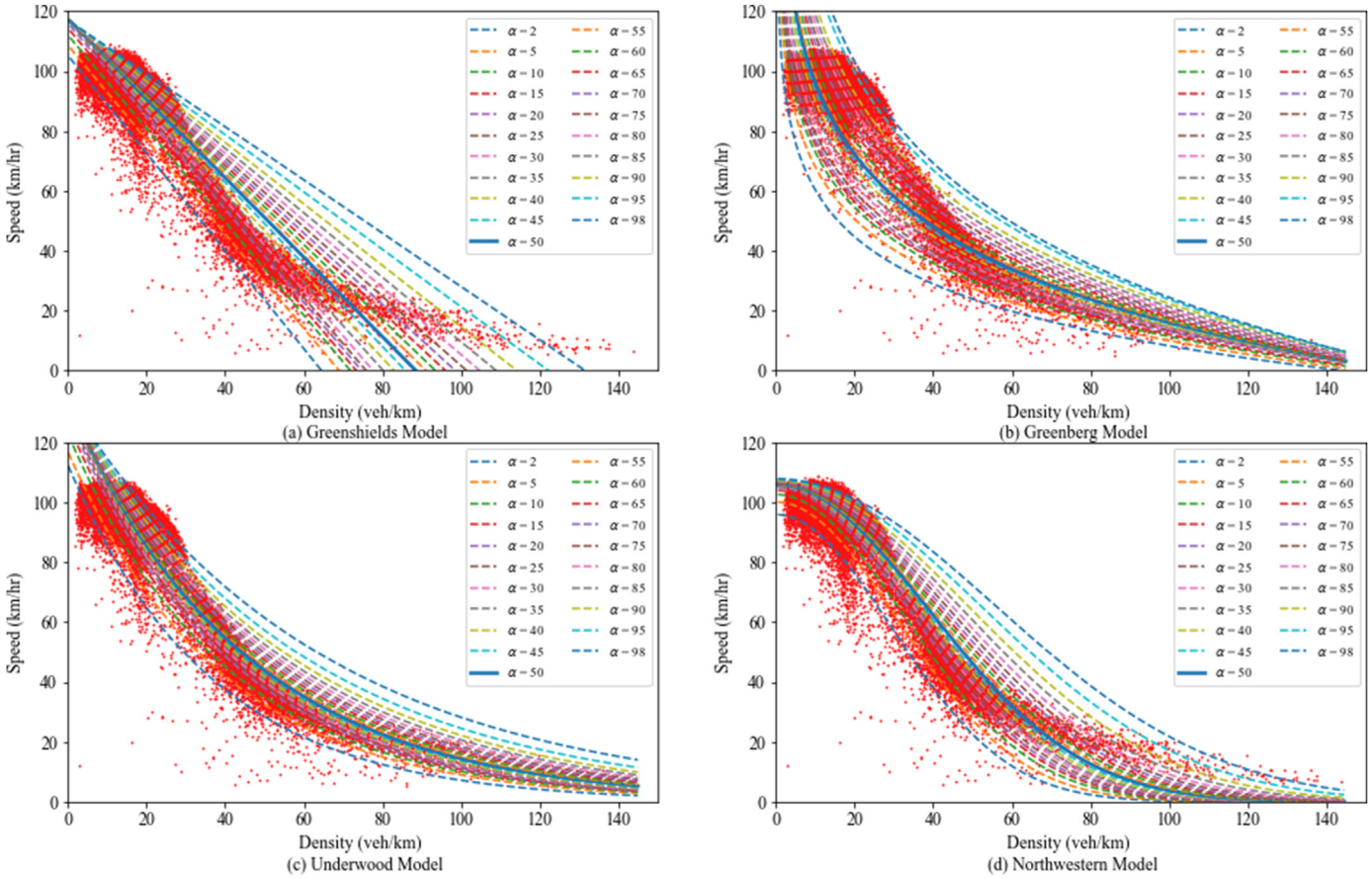

Fig. 6. Family of speed-density curves.

diagrams can be established accordingly for the entire range of traffic conditions. For any given density k, we can obtain the corresponding percentile-based speeds. Then, the CDF and PDF of speeds at a given density k can be derived. However, as mentioned above, the drawback of the basic percentile-based stochastic speed-density model is also obvious. Fig. 4 shows that the predicted speed-density curves of Greenshields, Greenberg, Underwood, and Northwestern model intersect at some points on the density interval [0, 145]. These intersections are highlighted in the right-side figures. For a particular density k within this interval, the predicted lower percentile of speed is larger than the

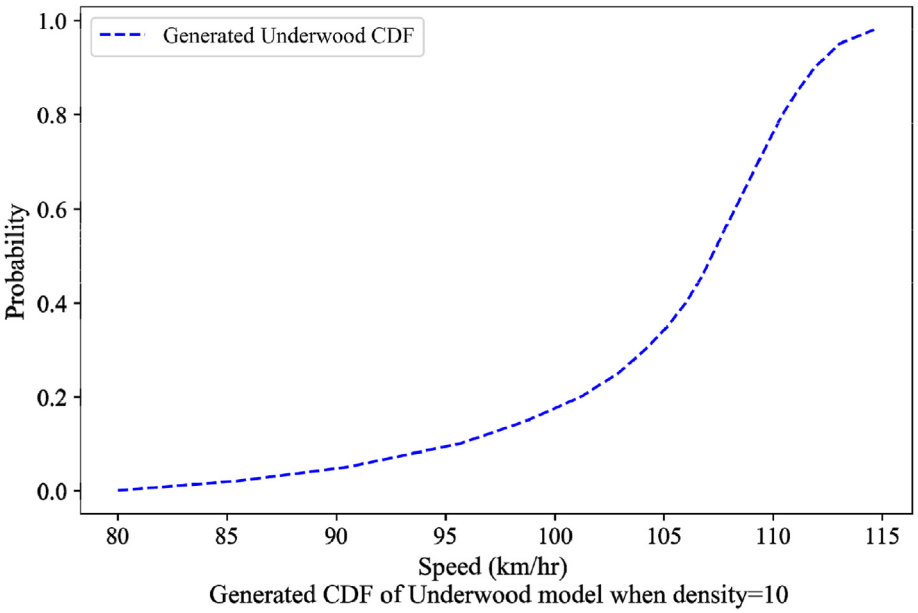

Fig. 7. Generated CDF of integrated Underwood model.

predicted higher percentile, which is logically wrong. Let us take the Underwood model as an example. Fig. 5 shows the generated CDF graph turns around as speed increases. This phenomenon implies that the generated PDF graph is negative (less than zero) within this range.

To address this issue, we proceed to apply the holistic percentile-based stochastic speed-density model to generate the speed-density curves. The domain of interest is set on the interval [0, 145] to ensure the generated CDF is monotonically increasing at any given density. Fig. 6 presents the generated curves w.r.t. the Greenshields, Greenberg, Underwood, and Northwestern model. The results show that the integrated model can reasonably solve this issue, the generated curves no longer interact with each other. Fig. 7 further shows that the generated CDF of Underwood curves becomes monotonically increasing, which implies this issue has been solved.

4. Conclusions

In this paper, we propose a new methodology to calibrate stochastic traffic flow fundamental diagrams. We first develop a basic model to generate the percentile-based speed-density relations, and then propose a holistic modeling framework to overcome unrealistic quantities generated by the basic model. A set of constraints are imposed on the percentile-based fundamental diagrams to ensure the generated CDF at any given density is monotonically increasing. The proposed model is then transformed into a linear programming problem by introducing the artificial variables. By selecting different percentile values, the speed distribution at any given density can be generated. The proposed methodology is validated in the numerical experiment. Compared with previous methods, the proposed modeling approach is more accurate and reasonable. The family of percentile-based fundamental diagrams is able to depict a variety of traffic phenomena, such as the diversity among drivers with different perceptions, responses, and driving habits, and transportation problems (Wang and Meng, 2012; Qi et al., 2021).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Castillo, J.M.D., Benítez, F.G., 1995a. On the functional form of the speed-density relationship-I: general theory. Transport. Res. Part B. 29. https://doi.org/10.1016/0191-2615(95)00008-2.

Castillo, J.M.D., Benítez, F.G., 1995b. On the functional form of the speed-density relationship-II: empirical investigation. Transport. Res. Part B. 29. https://doi.org/10.1016/0191-2615(95)00009-3.

Davino, C., Furno, M., Vistocco, D., 2013. Quantile Regression: Theory and Applications. Wiley. http://www.wiley.com/go/quantile_regression.

Edie, L.C., 1961. Car-following and steady-state theory for noncongested traffic. Oper. Res. 9. https://doi.org/10.1287/opre.9.1.66.

Fan, S., Seibold, B., 2013. Data-fitted first-order traffic models and their second-order generalizations. Transport. Res. Rec. 32–43. https://doi.org/10.3141/2391-04.

Greenberg, H., 1959. An analysis of traffic flow. Oper. Res. 7. https://doi.org/10.1287/opre.7.1.79.

Greenshields, B.D., 1935. A study of traffic capacity. In: 14 Annual Meeting of the Highway Research Board Proceedings.

Jabari, S.E., Zheng, J., Liu, H.X., 2014. A probabilistic stationary speed-density relation based on Newell's simplified car-following model. Transp. Res. Part B Methodol. 68. https://doi.org/10.1016/j.trb.2014.06.006.

Kerner, B.S., Konhäuser, P., 1994. Structure and parameters of clusters in traffic flow. Phys. Rev. E. 50. https://doi.org/10.1103/PhysRevE.50.54.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., et al., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. Transp. Res. Part B Methodol. 46. https://doi.org/10.1016/j.trb.2012.06.008.

Keyvan-Ekbatani, M., Papageorgiou, M., Papamichail, I., 2013. Urban congestion gating control based on reduced operational network fundamental diagrams. Transport. Res. C Emerg. Technol. 33. https://doi.org/10.1016/j.trc.2013.04.010.

Li, J., Zhang, H.M., 2011. Fundamental diagram of traffic flow: new identification scheme and further evidence from empirical data. Transport. Res. Rec. https://doi.org/10.3141/2260-06.

Muralidharan, A., Dervisoglu, G., Horowitz, R., 2011. Probabilistic graphical models of fundamental diagram parameters for simulations of freeway traffic. Transport. Res. Rec. https://doi.org/10.3141/2249-10.

Newell, G.F., 1961. Nonlinear effects in the dynamics of car following. Oper. Res. 9. https://doi.org/10.1287/opre.9.2.209.

Ni, D., Leonard, J.D., Jia, C., et al., 2016. Vehicle longitudinal control and traffic stream modeling. Transport. Sci. 50. https://doi.org/10.1287/trsc.2015.0614.

Nikolić, M., Bierlaire, M., De Lapparent, M., et al., 2019. Multiclass speed-density relationship for pedestrian traffic. Transport. Sci. 53. https://doi.org/10.1287/trsc.2018.0849.

Nikolić, M., Bierlaire, M., Farooq, B., et al., 2016. Probabilistic speed-density relationship for pedestrian traffic. Transp. Res. Part B Methodol. 89. https://doi.org/10.1016/j.trb.2016.04.002.

Qi, W., Shen, B., Yang, Y., 2021. Modeling drivers' scrambling behavior in China: an application of theory of planned behavior. Travel Behaviour and Society 24, 164–171. https://doi.org/10.1016/j.tbs.2021.03.008.

Qu, X., Wang, S., Zhang, J., 2015. On the fundamental diagram for freeway traffic: a novel calibration approach for single-regime models. Transp. Res. Part B Methodol. 73, 91–102. https://doi.org/10.1016/j.trb.2015.01.001.

Qu, X., Zhang, J., Wang, S., 2017. On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. Transp. Res. Part B Methodol. 104, 256–271. https://doi.org/10.1016/j.trb.2017.07.003.

Underwood, R.T., 1961. In: Speed, Volume, and Density Relationships: Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic.

Wang, H., Li, J., Chen, Q.Y., et al., 2011. Logistic modeling of the equilibrium speed-density relationship. Transport. Res. Part A Policy Pract. 45. https://doi.org/10.1016/j.tra.2011.03.010.

Wang, H., Ni, D., Chen, Q.Y., et al., 2013. Stochastic modeling of the equilibrium speed-density relationship. J. Adv. Transport. 47. https://doi.org/10.1002/atr.172.

Wang, S., Meng, Q., 2012. Sailing speed optimization for container ships in a liner shipping network. Transport. Res. E Log. 48 (3), 701–714. https://doi.org/10.1016/j.tre.2011.12.003.

Wu, X., Liu, H.X., Geroliminis, N., 2011. An empirical analysis on the arterial fundamental diagram. Transp. Res. Part B Methodol. 45. https://doi.org/10.1016/j.trb.2010.06.003.

Zhang, J., Qu, X., Wang, S., 2018. Reproducible generation of experimental data sample for calibrating traffic flow fundamental diagram. Transport. Res. Part A Policy Pract 111, 41–52. https://doi.org/10.1016/j.tra.2018.03.006.

Zhou, J., Zhu, F., 2020. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. Transport. Res. C Emerg. Technol. 115. https://doi.org/10.1016/j.trc.2020.102614.

Shuaian (Hans) Wang is a Professor at The Hong Kong Polytechnic University (PolyU). Prior to joining PolyU, he worked as a faculty member at Old Dominion University, USA, and University of Wollongong, Australia. Prof. Wang's research interests include shipping operations management, green shipping, big data in shipping, port planning and operations, urban transport network modeling, and logistics and supply chain management. Prof. Wang has published over 150 journal papers. He dedicates to rethinking and proposing innovative solutions to improve the efficiency of maritime and urban transportation systems, to promote environmental friendly and sustainable practices, and to transform business and engineering education.

Xinyuan Chen is a Lecturer in the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics (NUAA), in China. Prior to joining NUAA, he received his PhD degree from Monash University in Australia and worked as a postdoctoral fellow at The Hong Kong Polytechnic University. His research interests include urban transport network modeling, parallel computing in transport system analysis, multimodal transport system modeling and optimization, stochastic fundamental diagram, and maritime search and rescue.

Xiaobo Qu is a Chair Professor with the Department of Architecture and Civil Engineering, Chalmers University of Technology in Sweden, and is also affiliated with the School of Vehicle and Mobility, Tsinghua University, China. His research is focused on improving large, complex and interrelated urban mobility systems by integrating with emerging technologies. To date, Prof. Qu has secured research funding well above 9 million Euros from the Australian Research Council, Swedish Innovation Agency Vinnova, STINT, and European Union. He has published over 120 journal articles published at top tier journals in the area of transportation. He is an elected member in Academia Europaea-The Academy of Europe.