https://doi.org/10.1016/j.trb.2017.07.003

On the Stochastic Fundamental Diagram for Freeway Traffic: Model Development, Analytical Properties, Validation, and Extensive Applications

Xiaobo Qu[a], Jin Zhang[b], Shuaian Wang[c]

[a] School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia. Email: drxiaoboqu@gmail.com (Corresponding Author)

[b] Griffith School of Engineering, Griffith University, Gold Coast 4222 NSW, Australia.

[c] Department of Logistics & Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. Email: wangshuaian@gmail.com

Abstract

In this research, we apply a new calibration approach to generate stochastic traffic flow fundamental diagrams. We first prove that the percentile based fundamental diagrams are obtainable based on the proposed model. We further prove the proposed model has continuity, differentiability and convexity properties so that it can be easily solved by Gauss-Newton method. By selecting different percentile values from 0 to 1, the speed distributions at a given density can be derived. The model has been validated based on the GA400 data and the calibrated speed distributions perfectly fit the speed-density data. This proposed methodology has wide applications. First, new approaches can be proposed to evaluate the performance of calibrated fundamental diagrams by taking into account not only the residual but also ability to reflect the stochasticity of samples. Secondly, stochastic fundamental diagrams can be used to develop and evaluate traffic control strategies. In particular, the proposed stochastic fundamental diagram is applicable to model and optimize the connected and automated vehicles at the macroscopic level with an objective to reduce the stochasticity of traffic flow. Last but not the least, this proposed methodology can be applied to generate the stochastic models for most regression models with scattered samples.

Keywords: Stochastic Fundamental Diagram; Speed Distributions; Traffic Control.

1. INTRODUCTION

The traffic flow fundamental diagram has been considered as the foundation of traffic flow theory. It addresses the relationship among three fundamental parameters of traffic flow: traffic flow (vehs/hour), speed (km/hour), and traffic density (vehs/km). As flow is the product of speed and density, this relationship usually refers to flow – density or speed –density relationship. Since the seminal Greenshields model (Greenshields et al., 1935) was proposed, numerous studies have been done to improve this over-simplified relationship empirically and/or analytically (Greenberg, 1959; Newell, 1961; Underwood, 1961; Edie, 1961; Kerner and Konhäuser, 1994; Del Castillo and Benítez, 1995a&b; Li and Zhang, 2001; Wu, 2002; MacNicolas, 2008, Ji et al., 2010; Wang et al., 2011; Wu et al., 2011; Dervisoglu, 2012; Keyvan-Ekbatani et al., 2012&2013). The main focus of these studies is to develop accurate deterministic speed-density models with two or three practically meaningful parameters[1].

1.1 Six prominent speed – density models

In this section, we introduce a few prominent speed – density models. Greenberg (1959) propose a logarithmic function to represent this relationship. The main drawback of this model is that speed tends to infinity when density tends to zero. In order to overcome this limitation, Underwood (1961) put forward an exponential model. However, this model is not able to predict speeds at high densities. Newell (1961), Drake et al. (1967), and Wang et al. (2011) also propose their speed – density models in order to better represent this fundamental relationship. Table 1 lists six prominent speed – density models. These models can be used to determine the road capacities (Wu and Rakha, 2009), developing macroscopic traffic flow models (Phegley, 2013), and anticipate the traffic downstream (Kühne, 1984), and model traffic control strategies (Wang et al., 2014).

Table 1: Six well known speed-density models

Models	Function

[1] Note that the fundamental diagram has recently been extended to network level (i.e. macroscopic fundamental diagram), which deals with interrupted flow (e.g. Daganzo and Geroliminis, 2008; Geroliminis and Daganzo, 2008; Chiu et al., 2010; Leclercq, 2014; Keyvan-Ekbatani et al., 2015).

Greenshields et al. (1935)	$v = v_f\left(1 - \frac{k}{k_j}\right)$	v_f, k_j
Greenberg (1959)	$v = v_o \ln\left(\frac{k_j}{k}\right)$	$v_o,\ k_j$
Underwood (1961)	$v = v_f \exp\left(-\frac{k}{k_o}\right)$	v_f, k_o
Northwestern (Drake et al., 1967)	$v = v_f \exp\left[-\frac{1}{2}\left(\frac{k}{k_o}\right)^2\right]$	v_f, k_o
Newell (1961)	$v = v_f\left\{1 - \exp\left[-\frac{\lambda}{v_f}\left(\frac{1}{k} - \frac{1}{k_j}\right)\right]\right\}$	v_f, k_j, λ
Wang et al. (2011)	$v = \frac{v_f}{1 + \exp\left(\frac{k - k_c}{\theta}\right)}$	$v_f,\ k_c, \theta$

1.2 Stochasticity of speed – density samples

Although frequently being called a traffic "flow", freeway traffic is in fact far more complex than deterministic, predictable, and homogeneous fluids governed by physical laws. Indeed, freeway traffic flow possesses inherent random characteristics as it is composed by a variety of heterogeneous vehicles with distinct mechanical and electronic features, which are driven by a group of diversified drivers with different perceptions, responses, and driving habits on freeways with varied geometric features. In fact, there is a consensus in the literature that microscopic variables of traffic flow should be modeled as random variables (e.g. Breiman, 1963; Haight, 1963; Cowan, 1971&1975; Branston, 1976; Hoogendoorn and Bovy, 1998; Mahnke and Kaupužs, 1999; Jabari and Liu, 2012; Jabari and Liu, 2014). However, the traffic flow fundamental diagram, which refers to the relationship between speed, density, and flow, is predominantly treated as deterministic (e.g. Lighthill and Whitham, 1955; Zhang, 1998; Aw and Rascle, 2000; Wang et al., 2011; Coifman, 2014). Figure 1 shows the speed – density sample collected by loop detectors from 76 stations on the Georgia State Route 400 (referred to as GA400 dataset hereafter). This dataset has been widely used in calibrating and validating traffic flow fundamental diagrams (e.g. Wang et al., 2011; Qu et al., 2015). The raw GA400 data at each station are aggregated to average speed, flow and occupancy over a 20s sampling period. The raw data is further aggregated every 5 minutes. It is believed that the time interval is long enough (i.e. 15 times of the sampling period) for describing equilibrium fundamental diagrams (Wang et al., 2011; Coifman, 2014&2015; Ponnu and Coifman, 2015; Coifman et al., 2016). As can be seen in the figure, these samples are rather scattered throughout the entire range of traffic flow. In this regard, without taking into account the heterogeneity in vehicles, road geometry, and drivers, deterministic speed-density relationships limit their capacity to practically represent traffic flow and may result in inaccurate or misleading results in modelling traffic control strategies.

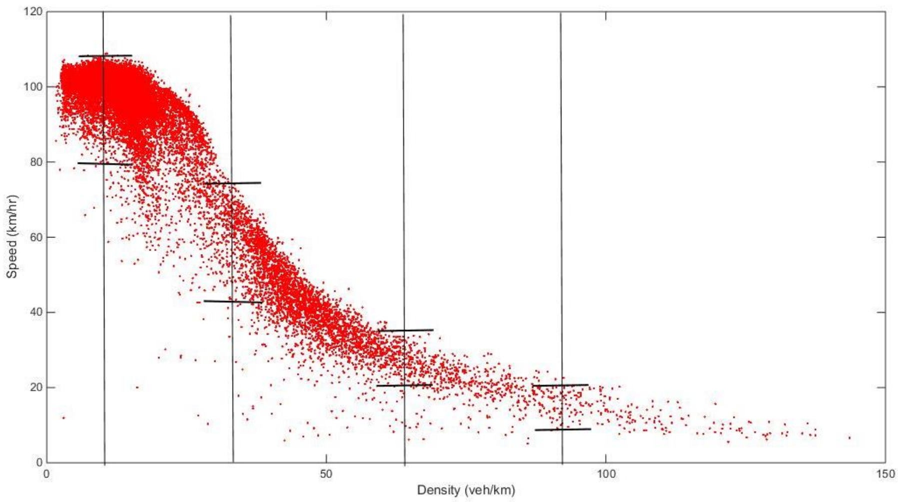


Figure 1: GA400 speed - density sample

A few pioneering attempts have been made to model the traffic flow fundamental diagram in a stochastic manner. Soyster and Wilson (1973) propose a simple stochastic flow–concentration model for traffic on hills using a Poisson process. Kerner (1998) gives the range of speed at a density based on three phase traffic flow models. Muralidharan et al. (2011) propose a probabilistic graphical traffic fundamental diagram based on the triangular deterministic model. Wang et al. (2013) propose a macroscopic stochastic approach to model the equilibrium speed-flow relationship. Jabari and Liu (2014) develop a probabilistic stationary speed-density relation based on Newell's simplified microscopic car-following model. Fan and Seibold (2013) introduce a new varying parameter, the empty-road velocity, to reflect the randomness in the Aw-Rascle-Zhang (ARZ) model. These pioneering studies have laid a solid foundation for stochastic traffic flow modelling. However, a few important points are yet to be resolved. The underlying assumptions for Soyster and Wilson (1973) are very specific and ideal and thus limit its applicability. In contrast to a speed distribution, only a range of speeds are given in Kerner (1998). The latter four studies are essentially on the basis of macroscopic or microscopic traffic models with a few analytical properties. Wang et al. (2013) approximate the variance and mean by using Karhunen-Loève expansion and explicit stochastic speed-density relationships are not obtainable in this research. To the best of our knowledge, Muralidharan et al. (2011), Jabari and Liu (2014) and Fan and Seibold (2013) are the only three published research works that are able to generate the distributions of speed or flow as a function of density and percentile based flow/speed-density relations are obtainable. However, they are all on the basis of a specific traffic flow model and thus are not generalizable to deal with the stochasticity of traffic flow using other traffic flow models such as the ones in Table 1.

1.3 Contributions and organization

In this paper, we develop a generic approach to generate a stochastic fundamental diagram. An optimization model based on the theorem of total probability is employed to calibrate the speed distributions as a function of density throughout the entire range of traffic states. We further prove 1) the optimal solution with respect a new parameter α is an unbiased estimator for the $100\alpha^{\text{th}}$ percentile based speed-density curve; 2) the proposed optimization model is convex for the four two-parameter models listed in Table 1 so the Gauss-Newton method can be applied to solve it; and 3) we design an approach that is able to efficiently calibrated the stochastic fundamental diagrams for the two three-parameter models in Table 1. We further apply the proposed methodology to calibrate the stochastic fundamental diagrams based on GA400 data and the resulting speed distributions perfectly match the observed data. It is believed that the proposed methodology has wide applicability to all speed-density models.

The rest paper is organized as follows. In Section 2, we present the optimization model and prove its convexity for Greenshields model. Its extensions to other models are illustrated in Section 3. This is followed by a case study in Section 4 to demonstrate the applicability and validity of the proposed methodology. Section 5 concludes.

2. METHODOLOGY

In this section, we first briefly introduce the deterministic speed – density models proposed in Qu et al. (2015). This is followed by a concept of percentile based speed – density curve and its mathematical representation. We further transform the mathematical representation to an optimization model in Section 2.2. The analytical properties of the proposed model's objective function based on a linear speed – density relationship, including continuity, differentiability, and convexity, have been rigorously proved in Section 2.3. Due to these analytical properties, the optimization model can be easily solved by Gauss-Newton method.

2.1 Deterministic Speed-Density Models

In order to establish a generalized stochastic speed-density diagram, we firstly need to select a deterministic speed-density model. A stochastic model can be developed on the basis of the selected deterministic model. Suppose the observed data is $(k^{data}, v^{data}) \in \{(k_i, v_i), i = 1, 2, ..., m\}$ and $v(k^{data})$ is the calibrated speed value by using the selected model when density equals k^{data}. Let us use the Greenshield's model as an example. The weighted least square method (WLSM) proposed by Qu et al. (2015) is used to calibrate the two parameters, which are free flow speed v_f and the jam density k_{jam} (we use k_{jam} instead of k_j to avoid confusion):

[M]

$$\min C(v_f, k_{jam}) = \sum_{i=1}^{m} \varpi_i \left(v_i - v(k_i)\right)^2 \tag{1}$$

subject to:

$$v(k_i) = v_f\left(1 - \frac{k_i}{k_{jam}}\right), i = 1, 2, ..., m \tag{2}$$

$$v_f > 0 \tag{3}$$

$$k_{jam} > 0 \tag{4}$$

where ϖ_i is the weight for observation (k_i, v_i) which accounts for the sample selection bias. The general weight determination method in Qu et al. (2015) is applied for this research, which is illustrated as below. In reality, the weight for an observation is essentially the distance between this observation with its next one. In doing so, if a particular traffic state is overrepresented, lower weights (i.e smaller distance among two adjacent samples) will be given to the corresponding observations to guarantee that this state would not dominate the calibration process; in contrast, if a particular traffic state is underrepresented, higher weights will be given. In this way, the sample selection bias can be largely eliminated.

Step 1: Rank the observations in terms of their densities. We thus have

$$\left(v_{(1)}, k_{(1)}\right), \left(v_{(2)}, k_{(2)}\right), \cdots, \left(v_{(i)}, k_{(i)}\right), \cdots, \left(v_{(m)}, k_{(m)}\right) \tag{5}$$

where $k_{(1)} \le k_{(2)} \le \cdots \le k_{(i)} \le \cdots \le k_{(m)}$ and $v_{(i)}$ is the corresponding speed in observation (i).

Step 2: Define $(\hat{u})$ as the largest index (i) that corresponds to the same density as $k_{(1)}$, that is,

$$\hat{u} := \arg\max\{i = 1, 2, \cdots m \mid k_{(i)} = k_{(1)}\} \tag{6}$$

Then,

$$\varpi_{(i)} = \frac{k_{(\hat{u}+1)} - k_{(1)}}{\hat{u}}, i = 1, 2, \cdots, \hat{u} \tag{7}$$

Step 3: Define $u = \hat{u} + 1$. Redefine $(\hat{u})$ as the largest index (i) that corresponds to the same density as $k_{(u)}$, that is,

$$\hat{u} := \arg\max\{i = u, u+1, u+2, \cdots, m \mid k_{(i)} = k_{(u)}\} \tag{8}$$

If $\hat{u} < m$, set

$$\varpi_{(i)} = \frac{k_{(\hat{u}+1)} - k_{(u-1)}}{2(\hat{u} - u + 1)}, i = u, u+1, u+2, \cdots, \hat{u} \tag{9}$$

and repeat Step 3. Else,

$$\varpi_{(i)} = \frac{k_{(m)} - k_{(u-1)}}{m - u + 1}, i = u, u+1, u+2, \cdots, m \tag{10}$$

and stop. □

Note that the only difference in calibrating different non-linear speed-density models is equation (2). If we select other models in Table 1, the calibration results will reflect the respective deterministic speed-density models. In other words, if we replace eq. (2) by other functions in Table 1, the corresponding deterministic speed-density functions can be calibrated by following the same procedure.

2.2 Stochastic Speed-Density Models

Having obtained the calibrated deterministic speed-density models, we apply an optimization model to calibrate a family of percentile-based speed-density curves by introducing another parameter α in the model [M'] to be presented later. We again use Greenshield's linear model as an example to explain the new model. The objective of the new model [M'] is to calibrate a $100\alpha^{\text{th}}$ percentile based speed-density curve such that the ratio between weighted residual of observations below the calibrated curve and the total residual is α. In other words, α is defined as the ratio between weighted residual of observations below the calibrated percentile based curve and the total residual. The calibrated curve, denoted as $v_\alpha\left(k_i\right)$, is actually the $100\alpha^{\text{th}}$ percentile based speed-density curve. The mathematical representation of α is,

$$\frac{\sum_{i=1}^{m}\varpi_i\left|g_\alpha(k_i,v_i)\right|}{\sum_{i=1}^{m}\varpi_i\left|v_i-v_\alpha\left(k_i\right)\right|}=\alpha \tag{11}$$

and

$$g_\alpha(k_i,v_i)=\begin{cases} v_\alpha\left(k_i\right)-v_i, \text{ if } v_i-v_\alpha\left(k_i\right)<0 \\ \quad 0, \text{ otherwise} \end{cases} \tag{12}$$

As $g_\alpha(k_i,v_i)$ is non-negative, eq. (11) is simplified as

$$\frac{\sum_{i=1}^{m}\varpi_i g_\alpha(k_i,v_i)}{\sum_{i=1}^{m}\varpi_i\left|v_i-v_\alpha\left(k_i\right)\right|}=\alpha \tag{13}$$

We further construct an example to better explain eqs. (11)-(13). Please refer to Appendix I.

Our key finding for the stochastic speed-density model is:

Theorem 1: Eq. (13) is satisfied at the optimal solution to the following optimization model [M'] and therefore we can solve [M'] to calibrate the $100\alpha^{\text{th}}$ percentile based speed-density line:

[M']
$$\min C\left(k_{jam}, v_f\right) = \sum_{i=1}^{m} (1-2\alpha)\varpi_i \left(g_\alpha(k_i, v_i)\right)^2 + \sum_{i=1}^{m} \alpha\varpi_i \left(v_i - v_\alpha\left(k_i\right)\right)^2 \tag{14}$$

subject to:

$$v_\alpha\left(k_i\right) = v_f\left(1 - \frac{k_i}{k_{jam}}\right), i = 1, 2, \dots, m \tag{15}$$

$$g_\alpha(k_i, v_i) = \begin{cases} v_\alpha\left(k_i\right) - v_i, \text{ if } v_i - v_\alpha\left(k_i\right) < 0 \\ 0, \text{ otherwise} \end{cases} \tag{16}$$

$$v_f > 0 \tag{17}$$

$$k_{jam} > 0 \tag{18}$$

Theorem 1 is implied by Lemma 1 and Corollary 2 that will be shown in the next sub-section. Theorem 1 means that, the fundamental diagram calibrated by Model [M'] is the $100\alpha^{\text{th}}$ percentile based speed-density curve based on Greenshield's model. By changing α between 0 and 1, the stochastic fundamental diagram is obtainable. Again, here if we replace eq. (16) by other functions in Table 1, the corresponding optimization models can be established to generate the $100\alpha^{\text{th}}$ percentile based speed-density curve with respect to other speed-density models.

2.3 Analytical properties: continuity, differentiability, and convexity

This section aims to examine the relations between Eq. (13) and model [M'] in order to gain insights into the stochasticity and deepen our understanding of the problem.

It should be noted that when $\alpha = 0$ or when $\alpha = 1$, the optimal objective value of model [M'] is 0 and there are an infinite number of optimal solutions of v_f and k_{jam}. As the two extreme cases with $\alpha = 0$ and $\alpha = 1$ are not of much help in practice, we assume in the sequel that $0 < \alpha < 1$.

2.3.1 A transformed model

We first develop an equivalent model to [M']:

Lemma 1: Define new decision variables

$$x = v_f, \quad y = v_f / k_{jam} \tag{19}$$

Then model [M'] is equivalent to the following model:

[M'']
$$\min_{x>0,y>0} C(x,y) := \sum_{i=1}^{m} (1-2\alpha)\varpi_i \left(g_\alpha\left(k_i, v_i\right)\right)^2 + \sum_{i=1}^{m} \alpha\varpi_i \left(v_i - v_\alpha\left(k_i\right)\right)^2 \tag{20}$$

$$v_\alpha\left(k_i\right) = x - yk_i \tag{21}$$

$$g_\alpha(k_i, v_i) = \begin{cases} v_\alpha\left(k_i\right) - v_i, \text{ if } v_i - v_\alpha\left(k_i\right) < 0 \\ 0, \text{ otherwise} \end{cases} \tag{22}$$

In order to simplify the notation, we hereby introduce

$$f_\alpha(k_i, v_i) = \begin{cases} v_i - v_\alpha\left(k_i\right), \text{ if } v_i - v_\alpha\left(k_i\right) > 0 \\ 0, \text{ otherwise} \end{cases} \tag{23}$$

The objective function of [M''] will be simplified as

$$\min_{x>0,y>0} C(x,y) := \sum_{i=1}^{m} \alpha\varpi_i \left(f_\alpha\left(k_i, v_i\right)\right)^2 + \sum_{i=1}^{m} (1-\alpha)\varpi_i \left(g_\alpha\left(k_i, v_i\right)\right)^2 \tag{24}$$

2.3.2 Continuity, differentiability, and convexity of $C(x, y)$

In model [M''], $f_\alpha\left(k_i, v_i\right)$ and $g_\alpha(k_i, v_i)$ are actually functions of x and y. Therefore, model [M''] actually minimizes the bi-variate function $C(x, y)$ over $x > 0$ and $y > 0$. It is easy to see that $C(x, y)$ is a continuous function of x and y.

Lemma 2: $C(x, y)$ is differentiable over $x > 0$ and $y > 0$.

Proof: Given a particular $\bar{x} > 0$ and a particular $\bar{y} > 0$, we classify the observations into three sets I_1, I_2, and I_3 in this way: all observations in sets I_1 are above the calibrated line, all observations in sets I_2 are on the calibrated line, and all observations in sets I_3 are below the calibrated line.

Then, the right partial derivative of $C(x,y)$ over x at $(\bar{x},\bar{y})$ is

$$\left.\frac{\partial C}{\partial x}\right|_{x\to\bar{x}^+,y=\bar{y}} = \sum_{i\in I_1}(-2)\alpha\varpi_i(v_i-\bar{x}+\bar{y}k_i)+\sum_{i\in I_2}2(1-\alpha)\varpi_i(\bar{x}-\bar{y}k_i-v_i)+\sum_{i\in I_3}2(1-\alpha)\varpi_i(\bar{x}-\bar{y}k_i-v_i) \tag{25}$$

The left partial derivative of $C(x,y)$ over x at $(\bar{x},\bar{y})$ is

$$\begin{aligned}\left.\frac{\partial C}{\partial x}\right|_{x\to\bar{x}^-,y=\bar{y}} &= \sum_{i\in I_1}(-2)\alpha\varpi_i(v_i-\bar{x}+\bar{y}k_i)+\sum_{i\in I_2}(-2)\alpha\varpi_i(v_i-\bar{x}+\bar{y}k_i) \\ &+\sum_{i\in I_3}2(1-\alpha)\varpi_i(\bar{x}-\bar{y}k_i-v_i)\end{aligned} \tag{26}$$

The definition of set I_2 implies

$$\bar{x}-\bar{y}k_i=v_i,\quad i\in I_2 \tag{27}$$

The above three equations mean that

$$\left.\frac{\partial C}{\partial x}\right|_{x\to\bar{x}^+,y=\bar{y}} = \left.\frac{\partial C}{\partial x}\right|_{x\to\bar{x}^-,y=\bar{y}} \tag{28}$$

Therefore, $C(x,y)$ is differentiable over x. Similarly, we can prove that $C(x,y)$ is differentiable over y. □

Lemma 3: $C(x,y)$ is not necessarily twice differentiable. However, it is twice differentiable when $\alpha=0.5$.

Proof: Given a particular $\bar{x}>0$ and a particular $\bar{y}>0$, similar to the above proof, the right twice partial derivative of $C(x,y)$ over x at $(\bar{x},\bar{y})$ is

$$\left.\frac{\partial^2 C}{\partial x^2}\right|_{x\to\bar{x}^+,y=\bar{y}} = \sum_{i\in I_1}2\alpha\varpi_i+\sum_{i\in I_2}2(1-\alpha)\varpi_i+\sum_{i\in I_3}2(1-\alpha)\varpi_i \tag{29}$$

The left twice partial derivative of $C(x,y)$ over x at $(\bar{x},\bar{y})$ is

$$\left.\frac{\partial^2 C}{\partial x^2}\right|_{x\to\bar{x}^-,y=\bar{y}} = \sum_{i\in I_1}2\alpha\varpi_i+\sum_{i\in I_2}2\alpha\varpi_i+\sum_{i\in I_3}2(1-\alpha)\varpi_i \tag{30}$$

Evidently, the above two equations are the same if $\alpha = 0.5$ or if set I_2 is empty. Otherwise the above two equations are different. □

Lemma 4: $C(x, y)$ is strictly convex.

Proof: Rearranging terms, we have

$$C(x,y)=\sum_{i=1}^{m}\left[\alpha\varpi_i\left(\max(v_i-v_\alpha(k_i),0)\right)^2+(1-\alpha)\varpi_i\left(\max(v_\alpha(k_i)-v_i,0)\right)^2\right] \tag{31}$$

It is easy to see that $\alpha\varpi_i\left(\max(v_i-v_\alpha(k_i),0)\right)^2+(1-\alpha)\varpi_i\left(\max(v_\alpha(k_i)-v_i,0)\right)^2$ is a strictly convex function of $v_\alpha(k_i)$. As $v_\alpha(k_i)=x-yk_i$ is a linear function of x and y, $C(x, y)$ is strictly convex over x and y. □

Lemma 4 implies that

Corollary 1: The optimal solution to model [M''] is unique and any local minimum to $C(x, y)$ is global minimum. □

As a result, model [M''] can be easily solved by Gauss-Newton algorithm.

2.3.3 Relation between the definition of percentile based speed-density line and optimization model [M'']

The optimal solution to model [M''], denoted by (x_α, y_α), must satisfy the first-order optimality condition that $\partial C/\partial x = 0$, $\partial C/\partial y = 0$ at (x_α, y_α). That is,

Theorem 2: The optimal solution to model [M''] (x_α, y_α) satisfies

$$\left.\frac{\partial C}{\partial x}\right|_{x=x_\alpha, y=y_\alpha} = \sum_{i\in I_1}(-2)\alpha\varpi_i(v_i - x_\alpha + y_\alpha k_i) + \sum_{i\in I_3} 2(1-\alpha)\varpi_i(x_\alpha - y_\alpha k_i - v_i) = \sum_{i=1}^{m}(-2)\alpha\varpi_i f_\alpha(k_i, v_i) + \sum_{i=1}^{m} 2(1-\alpha)\varpi_i g_\alpha(k_i, v_i) = 0 \tag{32}$$

That is,

$$\frac{\sum_{i=1}^{m}\varpi_i \bullet g_\alpha(k_i, v_i)}{\sum_{i=1}^{m}\varpi_i \bullet f_\alpha(k_i, v_i) + \sum_{i=1}^{m}\varpi_i \bullet g_\alpha(k_i, v_i)} = \alpha \tag{33}$$

$$\left.\frac{\partial C}{\partial y}\right|_{x=x_\alpha, y=y_\alpha} = \sum_{i\in I_1} 2\alpha\varpi_i(v_i - x_\alpha + y_\alpha k_i)k_i + \sum_{i\in I_3}(-2)(1-\alpha)\varpi_i(x_\alpha - y_\alpha k_i - v_i)k_i = \sum_{i=1}^{m} 2\alpha\varpi_i f_\alpha(k_i, v_i)k_i + \sum_{i=1}^{m}(-2)(1-\alpha)\varpi_i g_\alpha(k_i, v_i)k_i = 0 \tag{34}$$

That is,

$$\frac{\sum_{i=1}^{m}\varpi_i \bullet g_\alpha(k_i, v_i)\bullet k_i}{\sum_{i=1}^{m}\varpi_i \bullet f_\alpha(k_i, v_i)\bullet k_i + \sum_{i=1}^{m}\varpi_i \bullet g_\alpha(k_i, v_i)\bullet k_i} = \alpha \tag{35}$$

Eq. (33) in Theorem 2 implies that

Corollary 2: Eq. (13) is satisfied at the optimal solution to model [M'']. That is, the portion of lower weighted error out of the total weighted error is equal to α. □

Theorem 1, which is the most important finding, is now implied by Lemma 1 and Corollary 2.

Note that eq. (35) in Theorem 2 further implies that the portion of lower weighted error that is further weighted by the density out of the total weighted error that is further weighted by the density is still equal to α. This means that the percentile 100α is balanced between low density areas and high density areas.

We define the weighted average of estimated speeds:

$$\begin{aligned}\bar{v}_\alpha &:= \sum_{i=1}^{m} \varpi_i v_\alpha(k_i) = \sum_{i=1}^{m} \varpi_i \left[v_i - f_\alpha(k_i, v_i) + g_\alpha(k_i, v_i) \right] \\ &= \sum_{i=1}^{m} \varpi_i v_i - \sum_{i=1}^{m} \varpi_i f_\alpha(k_i, v_i) + \sum_{i=1}^{m} \varpi_i g_\alpha(k_i, v_i)\end{aligned} \tag{36}$$

Proposition 1: $\bar{v}_\alpha > \bar{v}_{0.5}$ when $\alpha > 0.5$, and $\bar{v}_\alpha < \bar{v}_{0.5}$ when $\alpha < 0.5$.

Proof: Eq. (32) in Theorem 2 implies

$$\alpha \sum_{i=1}^{m} \varpi_i f_\alpha(k_i, v_i) - (1-\alpha) \sum_{i=1}^{m} \varpi_i g_\alpha(k_i, v_i) = 0 \tag{37}$$

Eq. (37) shows that

$$\sum_{i=1}^{m} \varpi_i f_\alpha(k_i, v_i) = \sum_{i=1}^{m} \varpi_i g_\alpha(k_i, v_i) \quad \text{when } \alpha = 0.5 \tag{38}$$

$$\sum_{i=1}^{m} \varpi_i f_\alpha(k_i, v_i) < \sum_{i=1}^{m} \varpi_i g_\alpha(k_i, v_i) \quad \text{when } \alpha > 0.5 \tag{39}$$

$$\sum_{i=1}^{m} \varpi_i f_\alpha(k_i, v_i) > \sum_{i=1}^{m} \varpi_i g_\alpha(k_i, v_i) \quad \text{when } \alpha < 0.5 \tag{40}$$

Combined with eq. (36), we have

$$\bar{v}_{\alpha=0.5} = \sum_{i=1}^{m} \varpi_i v_i \;,\; \bar{v}_{\alpha>0.5} > \sum_{i=1}^{m} \varpi_i v_i \text{ , and } \bar{v}_{\alpha<0.5} < \sum_{i=1}^{m} \varpi_i v_i \tag{41}$$

3. EXTENSIONS TO NON-LINEAR SPEED-DENSITY MODELS

According to Qu et al. (2015), the linear fundamental diagram does not perform very well compared to other models. As a result, we also need to develop probabilistic non-linear fundamental diagrams. Fortunately, the three two-parameter nonlinear models can all be easily linearized in the form of eq. (21). We use Greenberg model as an example, which can be transformed as

$$v = v_0 \left(\ln(k_j) - k^{(g)} \right), \text{where } k^{(g)} = \ln(k) \tag{42}$$

By defining new decision variables

$$x = v_0 \ln\left(k_j\right), \quad y = v_0 \tag{43}$$

We have

$$v = x - yk^{(g)}, \text{where } k^{(g)} = \ln\left(k\right) \tag{44}$$

Similarly, we can linearize the other two two-parameter nonlinear models. Table 2 summarizes the linearization of the three two-parameter nonlinear models.

Table 2: Linearization of the three two-parameter non-linear models

Models	Original form	Linearized form
Greenberg (1959)	$v = v_o \ln\left(\frac{k_j}{k}\right)$	$v = x - yk^{(g)}$, where $k^{(g)} = \ln\left(k\right), x = v_0 \ln\left(k_j\right), y = v_0$
Underwood (1961)	$v = v_f \exp\left(-\frac{k}{k_o}\right)$	$v^{(u)} = x - yk$, where $v^{(u)} = \ln\left(v\right), x = \ln\left(v_f\right), y = \frac{1}{k_0}$
Northwestern (Drake et al., 1967)	$v = v_f \exp\left(-\frac{1}{2}\left(\frac{k}{k_o}\right)^2\right)$	$v^{(n)} = x - yk^{(n)}$, where $v^{(u)} = \ln\left(v\right), k^{(n)} = k^2, x = \ln\left(v_f\right), y = \frac{1}{2k_0{}^2}$

As can be seen in Table 2, all of the three models can be linearized to the form of eq. (21). We can easily prove that all the properties and algorithms of Model [M’] are also applicable to the corresponding optimization models based on the linearized Greenberg, Underwood and Northwestern models. The differentiability and strict convexity properties are both guaranteed so that Gauss-Newton method can be used accordingly. It should be pointed out that the methodology is only applicable when using speed-density relations that can be linearized.

For three-parameter models, we propose a numerical approach to obtain satisfactory calibration results. We first assume that one parameter is known and examine whether we can linearize the models in the form of eq. (21). For Newell model, we have

$$v = v_f \left[1 - \exp\left\{-\frac{\lambda}{v_f}\left\{\frac{1}{k} - \frac{1}{k_j}\right\}\right\}\right] \tag{45}$$

Eq. (50) is equivalent to

$$v_f \ln\left(\frac{v_f}{v_f - v}\right)\frac{1}{\lambda} = \frac{1}{k} - \frac{1}{k_j} \tag{46}$$

If we assume v_f is known, we can calibrate $1/\lambda$ and $1/k$ using the known v_f and the speed-density data. Note that the relationship between $1/\lambda$ and $1/k$ is linear. In other words, eq. (46) can be linearized as

$$k^{(e)} = x - y\lambda^{(e)} \tag{47}$$

where

$$\lambda^{(e)} = \frac{1}{\lambda} \tag{48}$$

$$k^{(e)} = \frac{1}{k} \tag{49}$$

$$x = \frac{1}{k_j} \tag{50}$$

$$y = v_f \ln\left(\frac{v_f - v}{v_f}\right) \tag{51}$$

Similarly, we can linearize the 3PL model (Wang et al., 2011). Table 3 summarizes the linearization of these two three-parameter non-linear models.

Table 3: Linearization of the three-parameter non-linear models

Models	Original form	Linearized form
Newell (1961)	$v = v_f\left[1 - \exp\left\{-\frac{\lambda}{v_f}\left\{\frac{1}{k} - \frac{1}{k_j}\right\}\right\}\right]$	$k^{(e)} = x - y\lambda^{(e)}$, where $\lambda^{(e)} = \frac{1}{\lambda}$, $k^{(e)} = \frac{1}{k}$ $x = \frac{1}{k_j}$, and $y = v_f \ln\left(\frac{v_f - v}{v_f}\right)$
Wang et al. (2011)	$v = \frac{v_f}{1 + \exp\left(\frac{k - k_c}{\theta}\right)}$	$k = x - y\theta$, where $x = k_c$, and $y = \ln\left(\frac{v}{v_f - v}\right)$

As both models can be linearized in the form of eq. (21), the continuity, differentiability, and convexity all hold for their corresponding optimization models if v_f is known.

Accordingly, if v_f is known, the Gauss-Newton method is applicable to generate the stochastic speed-density diagrams based on the two three-parameter models. As v_f represents the free flow speed, it has a very compact domain, say from 90 km/hour to 120 km/hour. In this regard, we can simply enumerate all possible values of v_f (with the precision of e.g. 1 km/hour) to obtain the global minimum for Model [M''] with respect to the two three-parameter models.

In sum, as the above transformation simply assumes that one of the three parameters is known, the exact solution is not obtainable. However, due to the compact domain of the calibration parameters, a satisfactory solution can be numerically estimated based on enumeration. As such, the three-parameter cases are not analytically handled in this research and only a numerical approach is provided for obtaining satisfactory solutions.

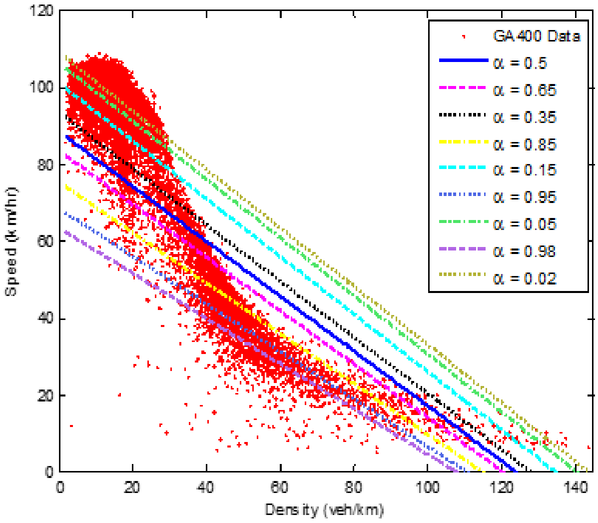
4. CASE STUDY

4.1 Data description

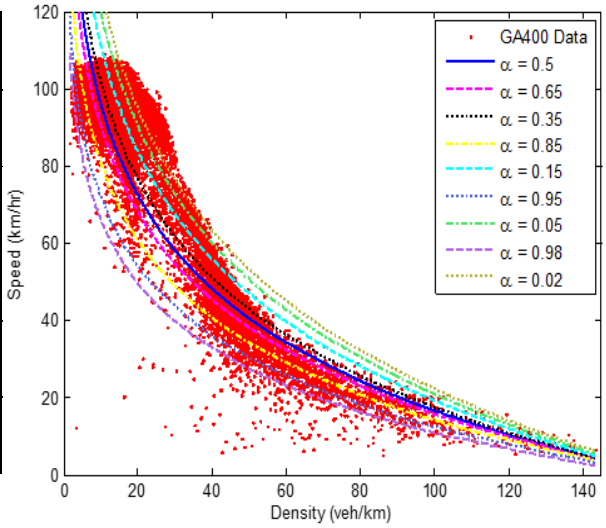
In this case study, we use the data collected by loop detectors from 76 stations on the Georgia State Route 400 for continuous observation of one year. The raw data at each station was originally aggregated to average speed, flow and occupancy over a 20s sampling period. It is further aggregated to 5 minutes to establish equilibrium traffic flow fundamental diagram. This aggregated data has been widely used in research community for fundamental diagram research (e.g. Wang et al., 2011&2014; Qu et al., 2015). As a follow up research, we use exactly the same aggregated data with the above three studies.

4.2 Results

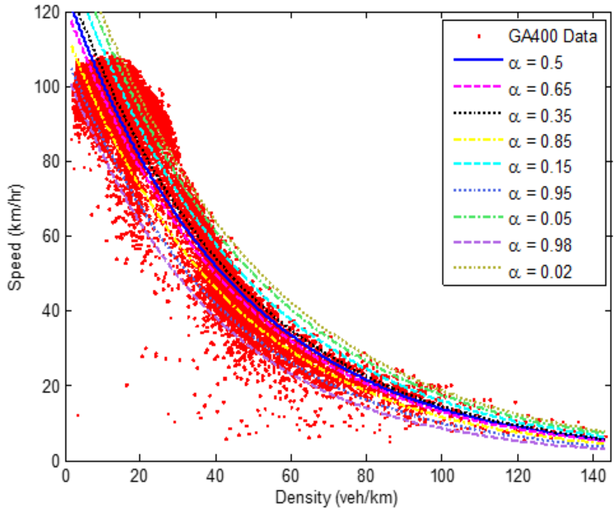
As per discussed in the previous sections, the optimization models all have sound analytical properties such as continuity, differentiability and convexity. By selecting different α values, a family of percentile based speed-density curves are obtainable. Figure 2 presents the curves with respect to the Greenshield's model (1935), Greenberg model (1959), Underwood model (1961), Northwestern model (1967), Newell model (1961), and 3PL model (2011). The red dots are the GA400 data. The thick solid line represents the models calibrated by WLSM (i.e. $\alpha = 0.5$). The other solid lines represent the flow-density curves generated from the optimization models with respect to $\alpha = 0.98,\ 0.95, 0.85, 0.65, 0.35, 0.15,\ 0.05,$ and 0.02 . As the percentile-based curves are generated, the stochastic fundamental diagram can be established accordingly for the entire range of traffic conditions. In other words, given a density value, the speed distributions can be obtained accordingly. We compare the generated speed distributions and the empirical speed distributions in Section 4.3.



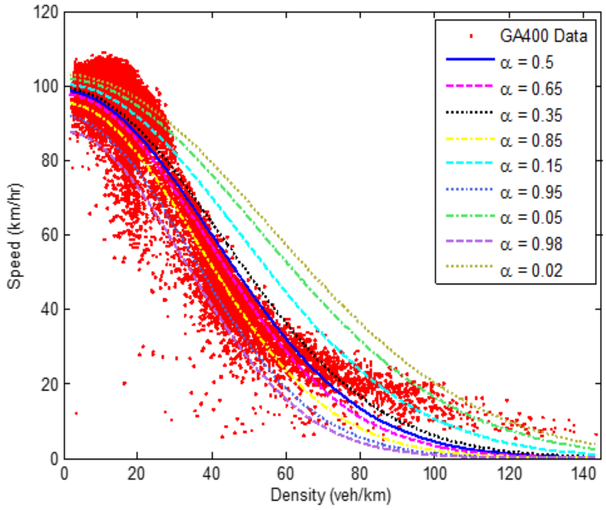
(a) Greenshields Model



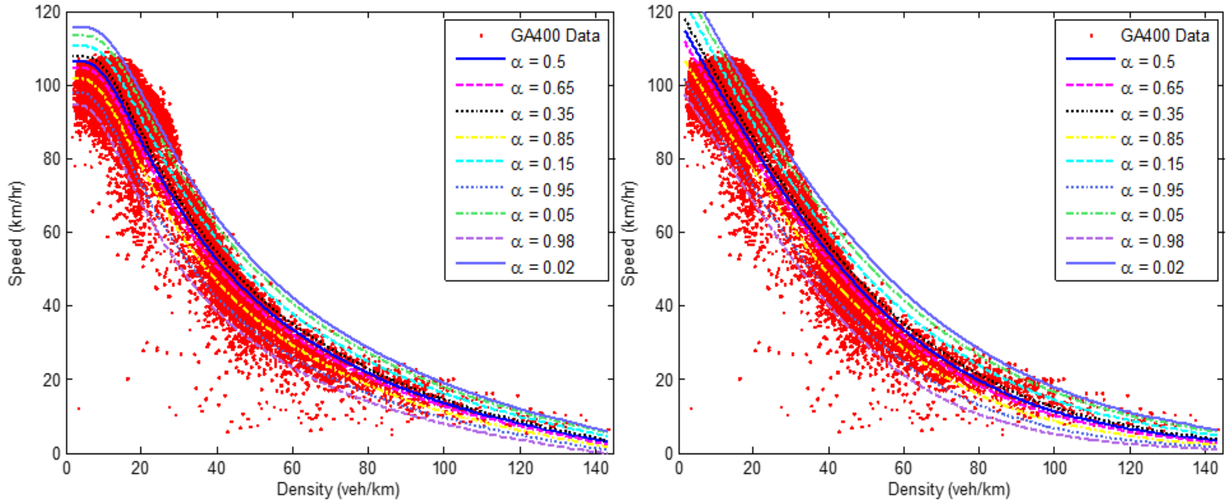
(b) Greenberg Model



(c) Underwood Model



(d) Northwestern Model

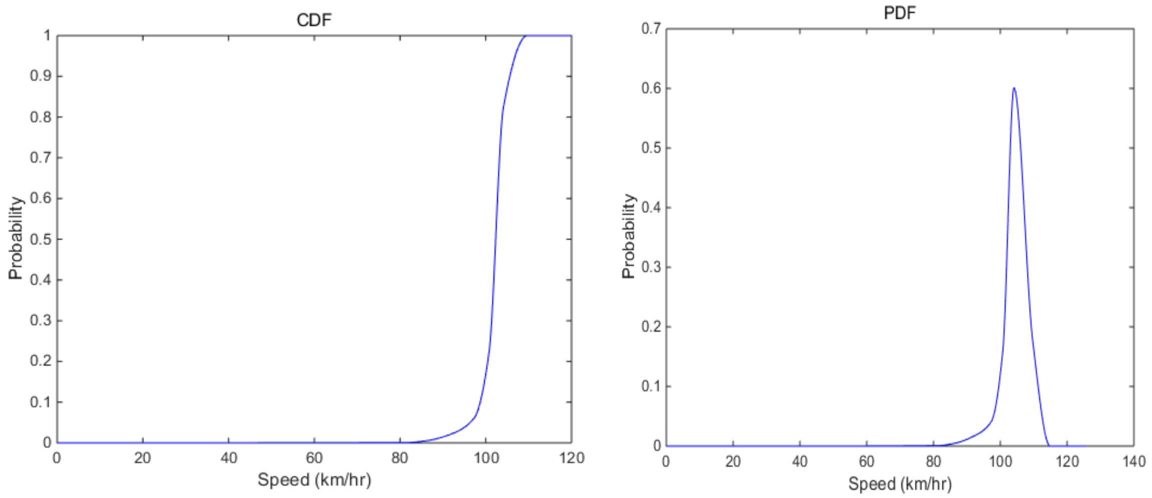


(e) Newell Model (f) 3PL Model

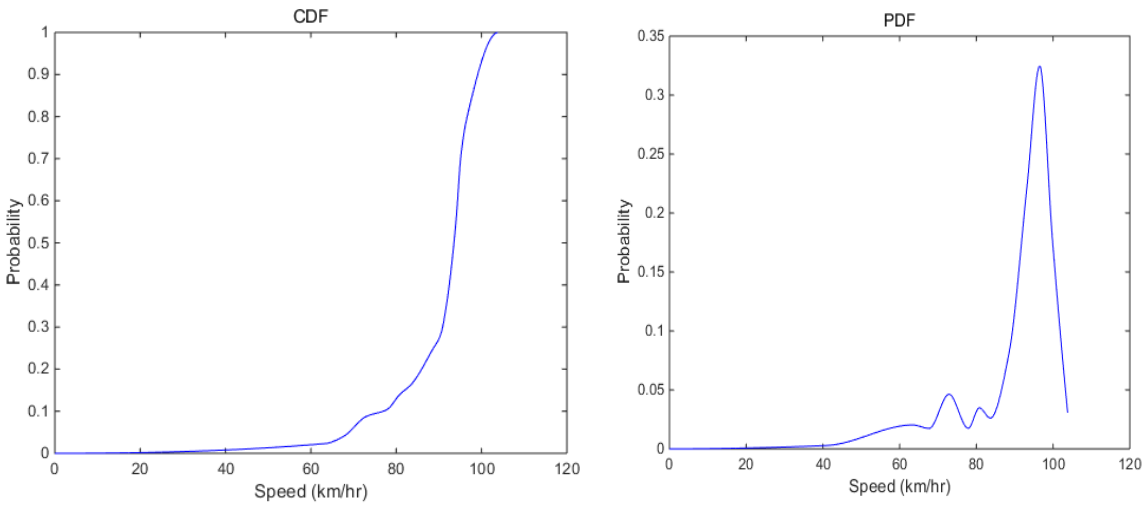
Figure 2: Family of flow-density curves

4.3 Validation

According to the Figure 2, for any given density, we can have all the corresponding percentile based speeds. In other words, the cumulative distribution function (CDF) and probability density function (PDF) of speeds at any given density are obtainable. Let us use the Underwood model as an example. When the density equals 10 veh/km and 20 veh/km, the generated cumulative distribution and probability density graphs are shown in Figure 3.



CDF when k = 10 *PDF when k = 10*

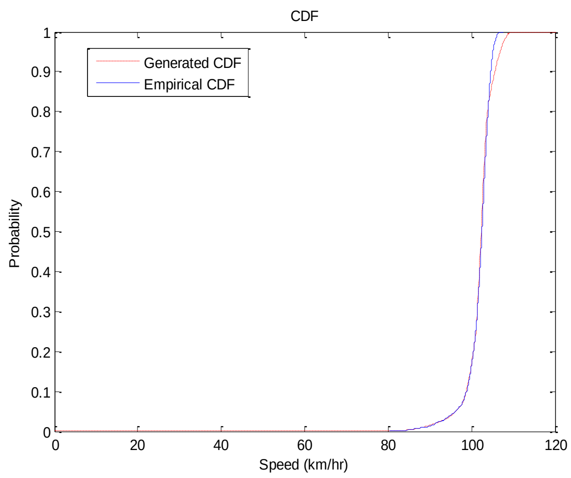


CDF when k = 20 *PDF when k = 20*

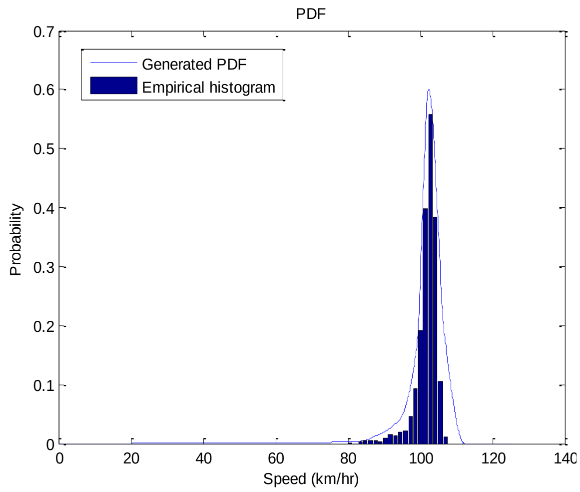
Figure 3: CDF and PDF of speeds

In order to validate the performance of the generated CDFs and PDFs, we further use the GA400 data to generate the empirical CDFs and PDFs with respect to the two densities. As can be seen in Figure 4, the generated CDFs and PDFs perfectly re-establish the empirical CDFs and PDFs. Hypothesis tests also suggest that the generated distributions fit the data very well. It should be noted that the empirical speed PDF when density equals 20 veh/km has a zigzag section from 60 km/hour to 85 km/hour. Surprisingly, the proposed model based approach can still capture this zigzag pattern and practically re-establish the pattern of our speed-density samples. We further compare the empirical results with generated results for other intervals when density is less than 40 veh/km and observe similar patterns. Therefore, this model can indeed generate the speed distributions at any given densities and the stochastic fundamental diagram is thus established.

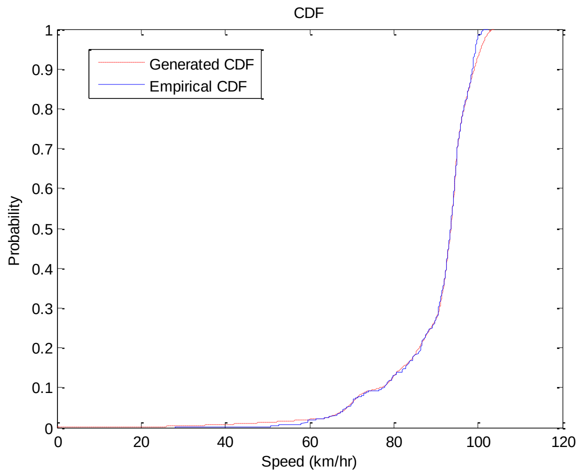
Note that the data points at higher densities are not enough to generate complete histograms for validation. The comparison results when density equals to 40 vehs/km, 60 vehs/km, and 80 vehs/km are presented in Appendix II. Although the histograms are incomplete, the generated PDFs and CDFs still reasonably re-establish the empirical ones. This further demonstrates the effectiveness and robustness of the proposed methodology.



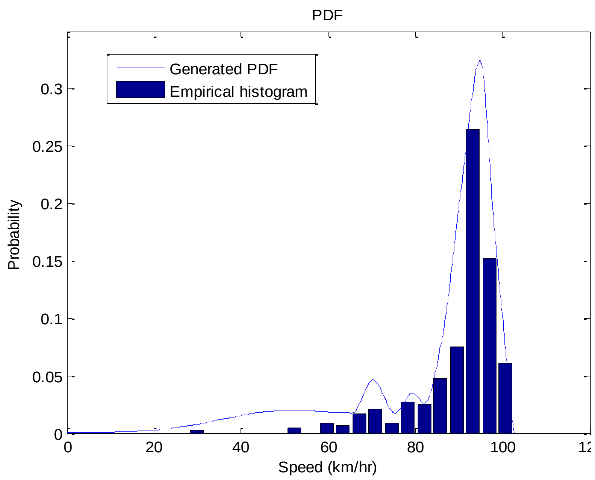
Empirical vs generated CDFs when k =10



Empirical vs generated PDFs when k =10



Empirical vs generated CDFs when k =20



Empirical vs generated CDFs when k =20

Figure 4: Empirical vs generated CDFs/PDFs

5. CONCLUSIONS

In this paper, we apply an optimization model based on the theorem of total probability to generate stochastic fundamental diagrams. In the proposed model, we introduce a new parameter α. We first prove that the solution of the proposed model with respect to any given α from 0 to 1 is actually the α^{th} percentile based fundamental diagram. Then we prove the proposed optimization model has continuity, differentiability and convexity properties so that it can be easily solved by Gauss-Newton method. By selecting different α values from 0 to 1, the speed distribution at a given density is obtainable. We further validate that the calibrated speed distributions perfectly fit our collected data.

In the past, as the fundamental diagram is a deterministic curve, researchers use the residual to assess the performance of a calibrated fundamental diagram. In doing so the stochastic nature of traffic flow is totally ignored. With this proposed stochastic fundamental diagram, new approaches can be proposed to evaluate the performance of calibrated fundamental diagrams by taking into account not only the residual but also stochasticity. This can be a follow-up study for this research.

Stochastic fundamental diagrams are also very useful in developing and evaluating traffic control strategies (Jabari and Liu 2013; Siqueira et al., 2016). First, as a deterministic model represents 50th percentile speed-density curve, it is incapable of handling 50% scenarios of traffic (underestimate 50% scenarios of flow or overestimate 50% scenarios of speed), which is not robust enough. In fact, 15th percentile is usually used for traffic engineering (e.g. to determine the operating speed). With this stochastic fundamental diagram, we can also use 15th percentile based speed-density curve to model and assess traffic control strategies with an attempt to deal with 85% scenarios. More importantly, it has been well recognized that the performance of our freeway systems can be substantially improved if the heterogeneity of traffic flow dynamics and stochasticity of fundamental diagrams can be controlled (Keyvan-Ekbatani et al., 2012; Punzo and Montanino, 2016). Unfortunately, existing traffic control strategies, which are based on deterministic models, are incapable of controlling the stochasticity of fundamental diagrams. With the proposed stochastic fundamental diagrams, new traffic control strategies can be developed with an objective to minimize the gap between upper and lower limit of our fundamental diagrams (e.g. 15th percentile based and 85th percentile based speed-density curves)

As discussed in the introductory section, the heterogeneities in drivers, vehicles, and road geometries result in stochasticity of fundamental diagrams. The recent research and development of the connected and automated vehicles provide a possible solution to reduce or even eliminate the heterogeneities (Du et al., 2015; Zhou et al., 2016a). A few pioneering studies have been done in optimising the trajectories of connected and automated vehicles to improve the safety, efficiency, and sustainability at the microscopic level (Ma et al., 2016; Zhou et al., 2016b). With this proposed stochastic fundamental diagram, it can also be modelled at the macroscopic level with an objective to reduce the stochasticity of fundamental diagrams.

This proposed methodology can be applied to generate the stochastic models for any linear or non-linear regression models with scattered sample. For example, it has been well recognized that vehicle fuel consumption is highly correlated to the vehicle speed. By using the proposed methodology, the stochastic speed-fuel functions can be developed.

It should be pointed out that appropriate data aggregation plays a key role in modelling traffic flow fundamental diagram. In this research, the speed and density data is aggregated to 5 minutes for equilibrium model development. As a future work, we will examine the stochasticity of samples and states by applying the proposed methodology to the aggregation level of 30 seconds.

Acknowledgement

This research is supported by UTS bluesky and ECR grants. The authors are indebted to the Editor in Chief and three anonymous reviewers for their insightful comments which indeed significantly improve the quality of this research. Further, special thanks are given to Dr. Daiheng Ni from UMass Amherst and Dr. Haizhong Wang from Oregon State University for their data support.

REFFERENCE

Aw, A. and Rascle, M. (2000). Resurrection of second-order models of traffic flow. Society for Industrial and Applied Mathematics, 60(3), 916–938.

Chiu, Y. C., Zhou, L. and Song, H. (2010). Development and calibration of the Anisotropic Mesoscopic Simulation model for uninterrupted flow facilities. Transportation Research Part B, 44(1), 152–174.

Coifman, B., (2014). Revisiting the empirical fundamental relationship. Transportation Research Part B, 68, 173–184.

Coifman, B., (2015). Empirical flow-density and speed-spacing relationships: evidence of vehicle length dependency. Transportation Research Part B, 78, 54–65.

Coifman, B., Wu, M., Redmill, K., and Thornton, D.A., (2016). Collecting ambient vehicle trajectories from an instrumented probe vehicle high quality data for microscopic traffic flow studies. Transportation Research Part C, 72, 254-271.

Cowan, R. J. (1971). A road with no overtaking. Australian Journal of Statistics, 13(2), 94–116.

Cowan, R. J. (1975). Useful headway models. Transportation Research, 9(6), 371–375.

Daganzo, C.F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B, 42(9), 771–781.

Del Castillo, J. M. and Benítez, F. G. (1995a). On the functional form of the speed-density relationship – I: General theory. Transportation Research Part B, 29(5), 373–389.

Del Castillo, J. M. and Benítez, F. G. (1995b). On the functional form of the speed-density relationship – II: Empirical investigation. Transportation Research Part B, 29(5), 391–406.

Dervisoglu, G. (2012). Automatic calibration of freeway models with model-based sensor fault detection. PhD Dissertation. University of California, Berkeley.

Drake, J. S., Schofer, J. L. and May, A. D. (1967). A statistical analysis of speed–density hypotheses. Highway Research Record, 154, 112–117.

Du, L., Han, L. and Chen, S. (2015). Coordinated online in-vehicle routing balancing user optimality and system optimality through information perturbation. Transportation Research Part B, 79, 121–133.

Edie, L. C. (1961). Car-following and steady state theory for non-congested traffic. Operations Research, 9(1), 66–76.

Fan. S. and Seibold, B. (2013). Data-fitted first-order traffic models and their second-order generalisations: Comparison by trajectory and sensor data. Transportation Research Record, 2391, 32–43.

Geroliminis, N. and Daganzo, C.F. (2008). Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transportation Research Part B, 42(9), 759–770.

Greenberg, H. (1959). An analysis of traffic flow. Operations Research, 7, 255–275.

Greenshields, B. D., Bibbins, J. R., Channing, W. S. and Miller, H. H. (1935). A study of traffic capacity. Highway Research Board Proceedings, 14, 448–477.

Haight, F. A. (1963). Mathematical theories of traffic flow. Academic Press, London, UK.

Hoogendoorn, S.P. and Bovy, P.H.L. (1998). A new estimation technique for vehicle-type specific headway distributions. Transportation Research Record, 1646, 18–28.

Jabari, S.E. and Liu, H.X. (2012). A stochastic model of traffic flow: theoretical foundations. Transportation Research Part B, 46, 156–174.

Jabari, S.E. and Liu, H.X. (2013). A stochastic model of traffic flow: Gaussian approximation and estimation. Transportation Research Part B, 47, 15–41.

Jabari, S.E. and Liu, H.X. (2014). A probabilistic stationary speed–density relation based on Newell's simplified car-following model. Transportation Research Part B, 68, 205–223.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S. and Qian, X. (2010). Investigating the shape of the macroscopic fundamental diagram using simulation data. Transportation Research Record, 2161, 40–48.

Kerner, B. S. (1998). Experimental features of self-organization in traffic flow. Physical Review Letters, 81(17), 3797.

Kerner, B.S. and Konhäuser, P. (1994). Structure and parameters of clusters in traffic flow. Physical Review E, 50(1), 54–83.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I. and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. Transportation Research Part B, 46(10), 1393–1403.

Keyvan-Ekbatani, M., Papageorgiou, M. and Papamichail, I. (2013). Urban congestion gating control based on reduced operational network fundamental diagrams. Transportation Research Part C, 33, 74–87.

Keyvan-Ekbatani, M., Papageorgiou, M., and Knoop, V.L. (2015). Controller design for gating traffic control in presence of time-delay in urban road networks. Transportation Research Part C, 59, 308-322.

Kühne, R.D., (1984). Macroscopic freeway model for dense traffic-stop-start waves and incident detection. In: Proceedings of the 9th International Symposium on Transportation and Traffic Theory, pp. 21–42.

Leclercq, L., Chiabaut, N. and Trinquier, B. (2014). Macroscopic fundamental diagram: a cross-comparison of estimation method. Transportation Research Part B, 62, 1–12.

Lehmann, E. L. and George, C. (1998). Theory of Point Estimation (2nd ed.). New York: Springer.

Li, J. and Zhang, H.M. (2001). Fundamental diagram of traffic flow: new identification scheme and further evidence from empirical data. Transportation Research Record, 2011, 50–59.

Lighthill, M.J. and Whitham, G.B. (1955). On kinematic waves II: A theory of traffic flow on long crowded roads. Proceedings of Royal Society A, 229, 281–345.

Ma, J., Li, X., Zhou, F., Hu, J. and Park, B.B. (2016). Parsimonious shooting heuristic for trajectory design of connected automated traffic part II: Computational issues and optimization. Transportation Research Part B, 95, 421-441.

MacNicholas, M. J. (2008). A simple and pragmatic representation of traffic flow. In: Symposium on The Fundamental Diagram: 75 years, Transportation Introduction Research Board, Woods Hole, MA.

Mahnke, R., and Kaupužs, J. (1999). Stochastic theory of freeway traffic. Physical Review E, 59(1), 117.

Muralidharan, A., Dervisoglu, G., and Horowitz, R. (2011). Probabilistic graphical models of fundamental diagram parameters for simulations of freeway traffic. Transportation Research Record, 2249: 78-85.

Newell, G. F. (1961). Nonlinear effects in the dynamics of car following. Operations Research, 9(2), 209–229.

Phegley, R.T., Horowitz, R., and Gomes, G. (2013). Fundamental diagram calibration: A stochastic approach to linear fitting. Transportation Research Board (TRB) 93rd Annual Meeting, Paper Number 14-4928, Session 2013.

Ponnu, B., Coifman, B., (2015). Speed-spacing dependency on relative speed from the adjacent lane: new insights for car following models. Transportation Research Part B, 82, 74–90

Punzo, V. and Montanino, M. (2016). Speed or spacing? Cumulative variables, and convolution of model errors and time in traffic flow models validation and calibration. Transportation Research Part B, 91, 21–33.

Qu, X., Wang, S. and Zhang, J. (2015). On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. Transportation Research Part B, 73, 91–102.

Siqueira, A. F., Peixoto, C.J.T., Wu, C. and Qian, W.L. (2016). Effect of stochastic transition in the fundamental diagram of traffic flow. Transportation Research Part B, 87, 1–13.

Soyster, A.L. and Wilson, G. R. (1973). A stochastic model of flow versus concentration applied to traffic on hills. Highway Research Board, 456, 28–39.

Underwood, R. T. (1961). Speed, volume, and density relationship: quality and theory of traffic flow. Yale Bureau of Highway Traffic, 141–188.

Wang, H., Li, H., Chen, Q. and Ni, D. (2011). Logistic modeling of the equilibrium speed–density relationship. Transportation Research Part A, 45, 554–566.

Wang, H., Ni, D., Chen, Q.Y. and Li, J. (2013). Stochastic modelling of the equilibrium speed-density relationship. Journal of Advanced Transportation, 47 (1), 126–150.

Wang, Y., Kosmatopoulous, E.B., Papageorgiou, M., Papamichail, I., (2014). Local ramp metering in the presence of a distant downstream bottleneck: theoretical stability analysis. IEEE Transactions on Intelligent Transportation Systems, 15(5), 2024-2039.

Wu, N., (2002). A new approach for modelling of fundamental diagrams. Transportation Research Part A, 36(10): 867–884.

Wu, X., Liu, H.X. and Geroliminis, N. (2011). An empirical analysis on the arterial fundamental diagram. Transportation Research Part B, 45(1), 255–266.

Zhang, H.M. and Kim, T. (2005). A car-following theory for multiphase vehicular traffic flow. Transportation Research Part B, 39, 385–399.

Zhang, H.M., (1998). A theory of non-equilibrium traffic flow. Transportation Research Part B, 32, 485–498.

Zhou, F., Li, X. and Ma, J. (2016a). Parsimonious shooting heuristic for trajectory design of connected automated traffic part I: Theoretical analysis with generalized time geography. Transportation Research Part B, 95, 394-420.

Zhou, M., Qu, X. and Jin, S. (2016b). On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model based approach. IEEE Transactions on Intelligent Transportation Systems. Article in Press.

Appendix I

Example A1: We construct an example to show eqs. (11)-(13). In this example, we assume the weight for each data point is 0.125. The solid line is the calibrated $100\alpha^{\text{th}}$ percentile based speed-density curve based on a linear function form (i.e. $v_\alpha(k)$). As can be seen in Figure A1, there are four data points above the solid line and the rest four are below it. Evidently, eq. (11) represents the ratio between absolute residual of samples below the solid line and the total residual and this curve is actually the 77th percentile speed-density curve according to eq. (13). In other words, α here is 0.77.

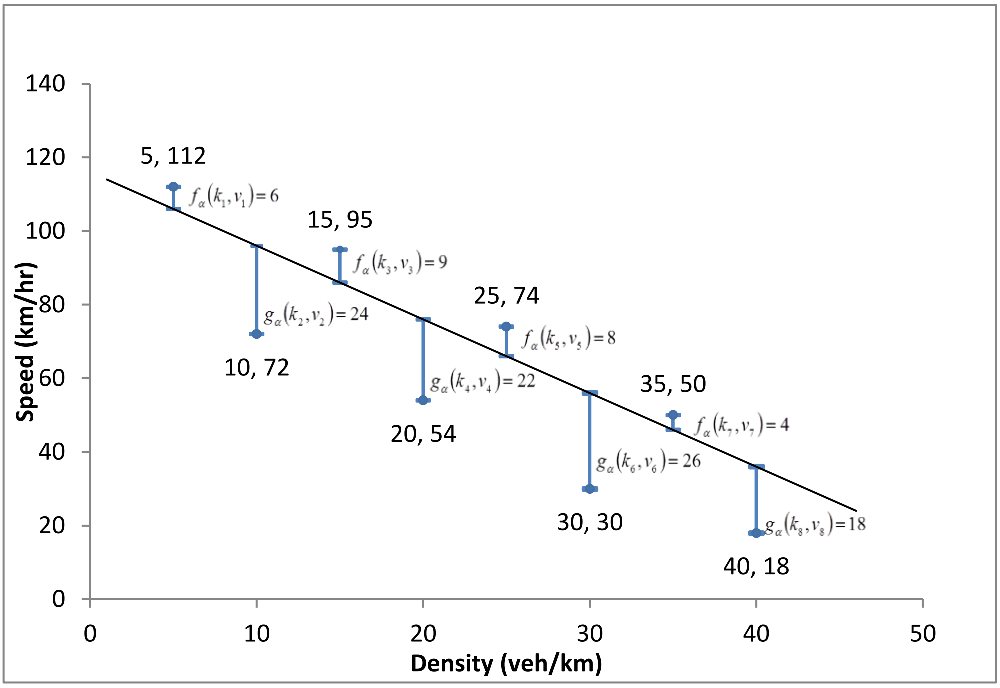
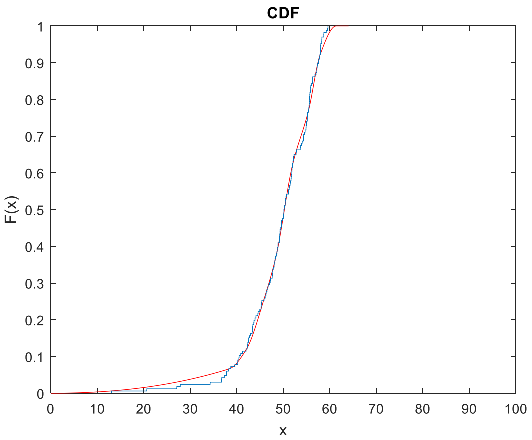


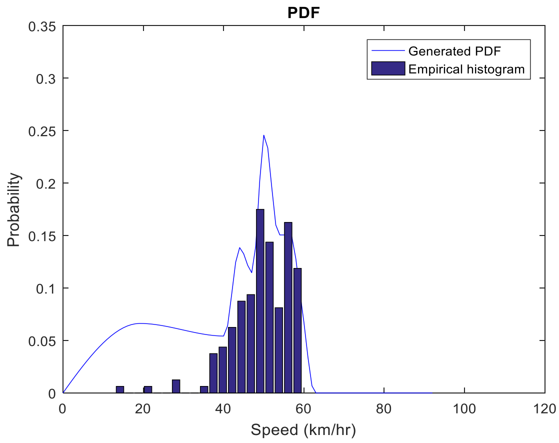
Figure A1: An illustrative example

Appendix II

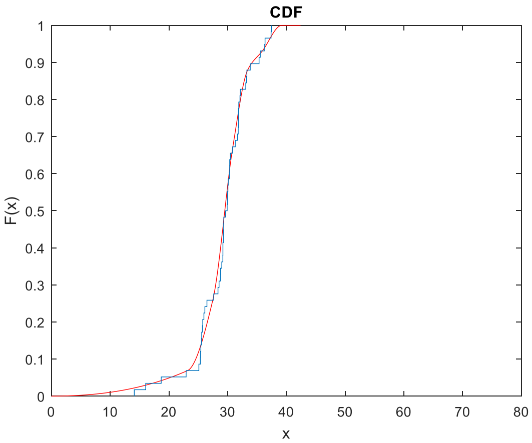
This appendix lists the empirical probability density function (PDF) and cumulative distribution function (CDF) against generated ones when density equals 40 vehs/km, 60 vehs/km, and 80 vehs/km. Note that the number of points at higher densities is significantly less than that at low densities. Consequently, the data points at these densities are not enough to obtain a complete histogram for comparison. For example, when density equals 40 vehs/km, we have no data when speed is less than 40 km/hour; when densities equal 60 vehs/km and 80 vehs/km, some parts of the histograms are apparently missing and lack of statistical power. Nevertheless, similar to Figure 4, the generated PDFs/CDFs still reasonably re-establish the empirical distributions. This further demonstrates the effectiveness and robustness of the proposed methodology.



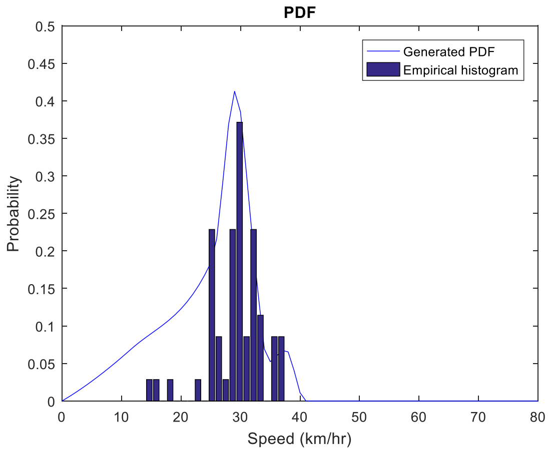
Empirical vs generated CDFs when k = 40



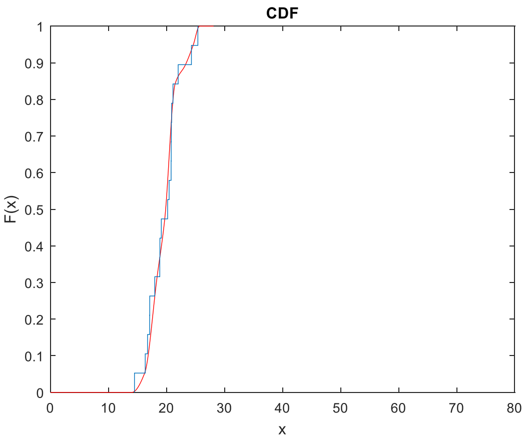
Empirical vs generated PDFs when k = 40



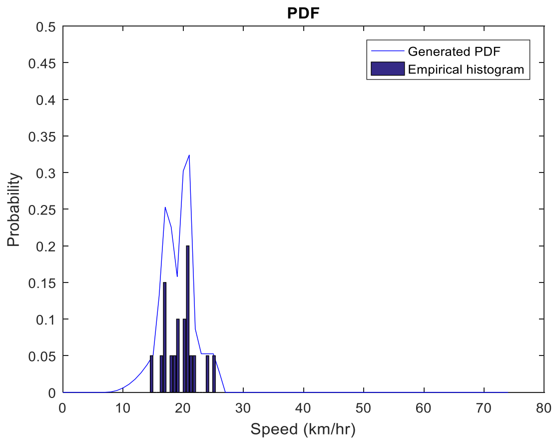
Empirical vs generated CDFs when k = 60



Empirical vs generated PDFs when k = 60



Empirical vs generated CDFs when k = 80



Empirical vs generated PDFs when k = 80

Figure A2: Empirical vs generated PDFs/CDFs (when k = 40, 60, and 80 vehs/km)