Fusion of multi-resolution data for estimating speed-density relationships

Lu Bai[a,b,*] PhD, Associate Professor xinyuesther@126.com, Wai Wong[c,*] PhD, Lecturer wai.wong@canterbury.ac.nz, Pengpeng Xu[d] PhD, Associate Professor pengpengxu@yeah.net, Pan Liu[a,b] PhD Professor pan_liu@hotmail.com, Andy H.F. Chow[e] PhD Associate Professor andychow@cityu.edu.hk, William H.K. Lam[f] PhD Professor Emeritus william.lam@polyu.edu.hk, Wei Ma[f] PhD, Assistant Professor wei.w.ma@polyu.edu.hk, Yu Han[a,b] PhD, Associate Professor, yuhan@seu.edu.cn, S.C. Wong[g] PhD, Chair Professor hhecwsc@hku.hk.

[a] Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing, China

[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou #2, Nanjing, China

[c] Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch 8041, New Zealand

[d] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China

[e] Department of Advanced Design and Systems Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China

[f] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Yuk Choi Road, Hong Kong, China

[g] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

*Corresponding author

ABSTRACT

Estimating traffic flow models, such as speed-density relationships, using data from multiple sources with different temporal resolutions is a prevalent challenge encountered in real-world scenarios. The resolution incompatibility is often intuitively addressed by averaging the high-resolution (HR) data to synchronize with the low-resolution (LR) data. This paper shows that ignoring the variability of HR data within the LR interval during the averaging process could lead to systematic data point distortions, resulting in biased model estimations. The average absolute biases of models estimated from the average data increase with the lost variability of HR data within the LR intervals. Subsequently, it proves that for any given complete average data dataset, there must exist an optimal dataset that minimizes the average absolute bias in model estimations introduced by the averaging process. A novel procedure for determining the practical optimal dataset is proposed. To test the proposed method, real-world HR data from four sites in Hong Kong and Nanjing, China were collected to mimic situations with multi-resolution data. Results demonstrated that the proposed method can significantly reduce the average absolute biases of models estimated from the determined practical optimal dataset, as compared to models estimated from the complete average dataset.

***Keywords*:** speed-density relationship; variability; resolution incompatibility; multi-resolution data; data fusion

1. Introduction

The speed-density relationship offers a profound understanding of traffic dynamics, which is fundamental for traffic modeling, congestion assessment, capacity estimation, and incident detection and management (Bai et al., 2021; Cheng et al., 2021; Dabiri and Kulcsár, 2022; Kodupuganti and Pulugurtha, 2023; Mohammadian et al., 2021; Nigam and Srivastava, 2023; Simon et al, 2022; Xu et al, 2023; Wang et al., 2021; Wang et al., 2022; Yin et al., 2022; Wong and Wong, 2016; Wong et. al., 2019). The accurate estimation of the speed-density relationship necessitates comprehensive and high-quality traffic data. Typically, stationary sensors, such as video camera and loop detector, are used as primary data sources for such estimation (Ambühl and Menendez, 2016; Bramich et al. 2021; Qian et al., 2017; Qu et al., 2017; Saffari et al., 2020; 2022; Wong et al., 2021; Zockaie et al., 2018). These sensors capture traffic data including flow, speed, and density by aggregating individual vehicle information over a consistent time interval. For instance, the widely used NGSIM I-80 dataset in traffic flow research (Coifman, 2015; Jabari and Liu, 2012; 2013; Jabari et al., 2014; Siqueira et al., 2016) was initially captured by seven synchronized digital video cameras and transcribed into vehicle trajectories, providing precise vehicle locations within the study area every one-tenth of a second. The trajectory data was aggregated over 2-min intervals to obtain the traffic flow and concentration (Qian et al., 2017). Qu et al. (2017) utilized loop detector data from 76 stations along Georgia State Route 400 to calibrate the speed distribution and establish the stochastic relationship between traffic speed and density within a link. The raw data was aggregated to calculate average speed, flow, and occupancy over a 20-second period, and then further aggregated over 5-minute intervals. The aggregated data was also used by Wang et al. (2011) and Qu et al. (2015). Recently, Bai et al. (2021) used the data from the Journey Time Indication System (JTIS) in Hong Kong to investigate the influences of speed heterogeneity and rainfall intensity on the link-based speed-density relation. The individual vehicle information recorded by Autoscope video detectors was aggregated over a 2-min period to obtain the average speed, traffic count, and speed variance. Bramich et al. (2022) assessed the effectiveness of 50 empirical traffic flow models using loop detector data collected from 25 cities, which were also typically aggregated over 3- or 5-min intervals. However, in a large urban network, collecting high-quality traffic data for every road link is often impractical. Stationary sensors, due to their costly installation and maintenance, are usually used to collect high-precision traffic data on a limited number of strategic links. For instance, the Kowloon Peninsula region in Hong Kong comprises 3,321 road links, but only 14 of them are equipped permanently with video detectors for capturing traffic data over the year. The remaining non-strategic links lack such high-precision traffic data.

Advancements in urban intelligent transportation systems have expanded the sources from which traffic information can be obtained (Ali-Eldin and Elmroth, 2021; Han et al., 2023; Ikonomakis et al., 2022; Liu et al., 2022; Zhu et al, 2022). For those non-strategic links without high-precision traffic data, a cost-effective approach to obtain traffic information involves integrating data from multiple sources. For instance, in Hong Kong, three primary transport monitoring systems are deployed: the JTIS, Traffic Speed Map (TSM), and Annual Traffic Census (ATC). The JTIS utilizes Autoscope video traffic detectors at major roads across Hong Kong to collect real-time traffic data, such as the space mean speed, its variance, and the traffic count. These data enable the JTIS to provide average journey time estimates for several major routes in Hong Kong, with an update interval of 2 min. However, due to its high cost, the JTIS covers only a limited number of prominent road links. In comparison, the TSM and ATC provide more widespread coverage. The TSM, an advanced real-time traffic speed system, provides speed information at 2-min intervals for 518 major roads in Hong Kong, derived from the automatic vehicle identification systems for commercial vehicles. However, the TSM does not provide traffic flow data for the roads. The ATC is a continual program that regularly monitors road traffic conditions using pneumatic air-tubes and inductive loop detectors from 1,662 detector stations. It implements a sampling strategy for selecting the location and time to measure traffic flow conditions within acceptable precision levels at a reasonable cost (Faghri and Chakroborty, 1994; Lam et al, 2003; Sharma et al., 1996; Wang and Yan, 2022). The ATC offers comprehensive insights into the annual average daily traffic of 88.5% of trafficable roads in Hong Kong and the hourly, daily, and monthly variabilities in traffic flow patterns (Transport Department, 2017). Based on ATC data, traffic flow information for these roads in Hong Kong can be obtained in 60-min intervals. TSM and ATC data can be combined to generate a comprehensive dataset comprising traffic speed and flow information for a wide coverage of roads in Hong Kong.

While combining data from multiple sources to obtain traffic information has clear cost benefits, using the combined data to estimate the speed-density relationship can be challenging. Data obtained from various sources often have different temporal resolutions. For example, the traffic speed data provided by the TSM over 2-min intervals represents high-resolution (HR) data recorded over short time periods. In contrast, the traffic flow data provided by the ATC over 60-min intervals represents low-resolution (LR) data recorded over longer time periods. A conventional and straightforward approach to align the temporal resolutions is to average the HR data over the LR interval. However, averaging HR data over the LR interval would lose valuable information on traffic variability in the HR data. Specifically, when the HR data exhibits significant variability, the averaged data over the LR interval may significantly deviate from the actual HR data, resulting in biased estimations of the model parameters (Wong and Wong, 2015, 2015, 2019; Wong et al., 2019; Xu et al., 2023). Adaptive Kalman filtering can be considered an alternative approach for handling multi-resolution data. This data-driven method involves an iterative mathematical process using a set of equations and successive data inputs to estimate system states when HR data cannot be directly measured (Chui and Chen, 1991; West and Harrison, 1997). However, the transferability of adaptive Kalman filtering may be constrained by significant geographical disparities, primarily due to its high dependency on data. This limitation could potentially affect the accuracy of model estimations.

This study delves into the complexities of speed-density relationship estimation on links with multi-resolution data. First, it uncovers a systematic distortion of data points caused by the averaging process where the variability of HR data in an LR interval is disregarded. Model estimations based on average data with systematic distortions could lead to biased model parameters. Second, an average absolute bias is proposed to objectively quantify the embedded bias. The average absolute bias increases in proportion to the lost variability. Third, it proves that for any given complete average data dataset, there must exist an optimal dataset that minimizes the average absolute bias in model estimations. Fourth, a practical optimal dataset determination procedure is proposed. To verify the applicability and transferability of the proposed method, four sites in Hong Kong and Nanjing, China with HR data are employed to mimic the situation with multi-resolution data. A comprehensive analysis, considering five traffic flow models from different model families and two LR intervals, is conducted for the selected sites. Results from the case study further demonstrates that average-data-based models estimated from the identified practical optimal datasets consistently outperforms those estimated from the complete datasets. This work contributes to the field by uncovering the commonly overlooked issue of biased model estimations arising from average data, and providing a practical, robust, and transferable method for estimating traffic flow models in situations with multi-resolution data.

The remaining sections of this paper are structured as follows: Section 2 delves into the causes for biased traffic flow model estimations arising from average data. Section 3 proves the existence of an optimal dataset. Section 4 proposes a novel method for practical optimal dataset determination. Section 5 presents the case study demonstrating the applicability and transferability of the proposed method. Section 6 concludes the study.

2. Biased speed-density relationship arising from average data

Consider any set $\boldsymbol{R}$ containing $|\boldsymbol{R}|$ sets of HR data, where $|\boldsymbol{R}|$ is the total number of LR intervals. $\forall r \in [1,\ |\boldsymbol{R}|]$, define u_{rm} and k_{rm} respectively to be the mth observation of the HR speed and HR density within the rth LR interval, where $m \in [1,\ M]$ and M is the total number of HR data point within rth LR interval. Based on these HR data, the speed-density relationship can be modeled by

$$u_{rm} = F(k_{rm}; \boldsymbol{\omega_R}) + \varepsilon_{rm}, \tag{1}$$

where $F(.)$ is a highly differentiable nonlinear speed-density function, $\boldsymbol{\omega_R} = \{\omega_{\boldsymbol{R}1}, \omega_{\boldsymbol{R}2}, \ldots, \omega_{\boldsymbol{R}n}\}$ is the vector of model parameters estimated using HR data from set $\boldsymbol{R}$, and ε_{rm} is the random error.

Nevertheless, in most real-world scenarios, these HR data are unavailable for non-strategic links. Most often, only data from multiple sources with different temporal resolutions are accessible. Consider situations where HR speed and LR density are available. To estimate the speed-density relationship based on these data, a common approach to address the resolution incompatibility involving averaging the HR speed data to match with the resolution of the LR density data. Denote $\overline{u}_r$ and σ_{ur}^2 respectively to be the average speed and speed variance of the HR speed data within the rth LR interval, and $\overline{k}_r$ and σ_{kr}^2 respectively to be the average density and density variance of the HR density data within the rth LR interval. The averaging process yields average data points $(\overline{k}_r, \overline{u}_r)$ with a compatible resolution. **Proposition 1** asserts that such approach could lead to systematic vertical data point shifting of HR data to the average data, which is denoted by D_r.

Proposition 1. *Given that the HR data within the LR interval is subject to variability, averaging HR data to align with the resolution of the LR data results in systematic vertical data point shifting by* D_r, $\forall r \in [1, |\boldsymbol{R}|]$, *where* $D_r = \frac{1}{2!}\frac{\partial^2 F(\overline{k}_r;\boldsymbol{\omega_R})}{\partial k_{rm}^2}\sigma_{kr}^2$.

Proof. Approximate u_{rm} by a Taylor series expansion with the center at $k_{rm} = \overline{k}_r$, $\forall m \in [1, M]$,

$$\begin{aligned} u_{rm} = F\left(\overline{k}_r;\boldsymbol{\omega_R}\right) + \frac{\partial F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}}\left(k_{rm}-\overline{k}_r\right) + \frac{1}{2!}\frac{\partial^2 F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}^2}\left(k_{rm}-\overline{k}_r\right)^2 \\ + \cdots + \frac{1}{n!}\frac{\partial^n F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}^n}\left(k_{rm}-\overline{k}_r\right)^n + \varepsilon_{rm}. \end{aligned} \tag{2}$$

A commonly adopted approach to address the resolution incompatibility is averaging the HR speed data to match the resolution of the LR density data. By averaging all the HR speed data over the rth LR interval, the relationship between the average speed, average density, and HR density can be expressed as follows:

$$\begin{aligned} E(u_{rm}) = \overline{u}_r = \frac{u_{r1}+\cdots+u_{rM}}{M} = F\left(\overline{k}_r;\boldsymbol{\omega_R}\right) + \frac{\partial F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}}\frac{\sum_{m=1}^{M}\left(k_{rm}-\overline{k}_r\right)}{M} \\ + \frac{1}{2!}\frac{\partial^2 F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}^2}\frac{\sum_{m=1}^{M}\left(k_{rm}-\overline{k}_r\right)^2}{M} + \cdots + \frac{1}{n!}\frac{\partial^n F\left(\overline{k}_r;\boldsymbol{\omega_R}\right)}{\partial k_{rm}^n}\frac{\sum_{m=1}^{M}\left(k_{rm}-\overline{k}_r\right)^n}{M} + \frac{\sum_{m=1}^{M}\varepsilon_{rm}}{M}. \end{aligned} \tag{3}$$

Eq. (3) shows that in addition to the zeroth-order moment, $F(\overline{k}_r; \boldsymbol{\omega_R})$, the higher order moments also contribute to the average speed, $\overline{u}_r$, or the expectation function, $E(u_{rm})$. It is important to note that although k_{rm} is unavailable and replaced by $\overline{k}_r$, $F(\overline{k}_r; \boldsymbol{\omega_R})$ is identical to the true model $F(k_{rm}; \boldsymbol{\omega_R})$ because their shapes are determined by the same model form along with the same set of model parameters $\boldsymbol{\omega_R}$. In other words, the average speed $\overline{u}_r$ differs from the true model $F(k_{rm}; \boldsymbol{\omega_R})$ if the sum of all the terms, except for $F(\overline{k}_r; \boldsymbol{\omega_R})$, on the right-hand side of Eq. (3) is non-zero, or the HR speed data within the LR interval is subject to variability. In general, the contribution of each term on the right-hand side of Eq. (3) usually decreases with its term order. Therefore, for simplicity, $\overline{u}_r$ can be approximated by its quadratic approximation of the expectation function, $E_2(u_{rm})$, which is obtained by truncating all the terms behind the second-order term in Eq (3):

$$E_2(u_{rm}) = F(\overline{k}_r; \boldsymbol{\omega_R}) + \frac{\partial F(\overline{k}_r; \boldsymbol{\omega_R})}{\partial k_{rm}} \frac{\sum_{m=1}^{M}(k_{rm} - \overline{k}_r)}{M} + \frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega_R})}{\partial k_{rm}^2} \frac{\sum_{m=1}^{M}(k_{rm} - \overline{k}_r)^2}{M} \quad (4)$$

As $\frac{\sum_{m=1}^{M}(k_{rm} - \overline{k}_r)}{M} = 0$ and $\frac{\sum_{m=1}^{M}(k_{rm} - \overline{k}_r)^2}{M} = \sigma_{kr}^2$, it follows

$$E_2(u_{rm}) = F(\overline{k}_r; \boldsymbol{\omega_R}) + \frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega_R})}{\partial k_{rm}^2}\sigma_{kr}^2. \quad (5)$$

Define $D_r = E_2(u_{rm}) - F(\overline{k}_r; \boldsymbol{\omega_R})$, $\forall r \in [1, |\boldsymbol{R}|]$. Thus, the vertical difference between the average data points and the true model can be approximated by

$$D_r = \frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega_R})}{\partial k_{rm}^2}\sigma_{kr}^2. \quad (6)$$

When the HR speed data are substituted by the average speed, all of the data points, (k_{rm}, u_{rm}), $\forall m \in [1, M]$, within the rth LR interval shift to the average data point, $(\overline{k}_r, \overline{u}_r)$. This induces a systematic vertical data point shifting by D_r. The direction of the systematic vertical data point shifting mainly dependent on the convexity of F. Since $F(.)$ is a highly differentiable nonlinear speed-density function, $\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega_R})}{\partial k_{rm}^2} \neq 0$ in general. Thus, $D_r = 0 \Leftrightarrow \sigma_{kr}^2 = 0$.

■

When the speed-density relationship is estimated directly using average speed $\overline{u}_r$ and average density $\overline{k}_r$, $\forall r \in [1, |\boldsymbol{R}|]$, the corresponding least squares function, S, can be expressed as

$$\min S = \sum_{r=1}^{|\boldsymbol{R}|} \left[\overline{u}_r - F\left(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}\right)\right]^2, \tag{7}$$

where $\widehat{\boldsymbol{\omega}}_{\boldsymbol{R}} = \{\widehat{\omega}_{\boldsymbol{R}1}, \widehat{\omega}_{\boldsymbol{R}2}, \dots, \widehat{\omega}_{\boldsymbol{R}n}\}$ is the vector of model parameters estimated based on the average data from set $\boldsymbol{R}$. Upon minimization, $F\left(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}\right) \cong \overline{u}_r$. If any $\overline{u}_r$ exhibits a non-zero systematic vertical data point distortion D_r, the $\widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}$ is biased. Figure 1 illustrates the discrepancy between the HR-data-based and average-data-based speed-density relationships. The arrows illustrate the directions and magnitudes of the systematic vertical data point shifting from the HR data points to the average data point. For details on the systematic data point distortion mechanism, please refer to Wong and Wong (2019).

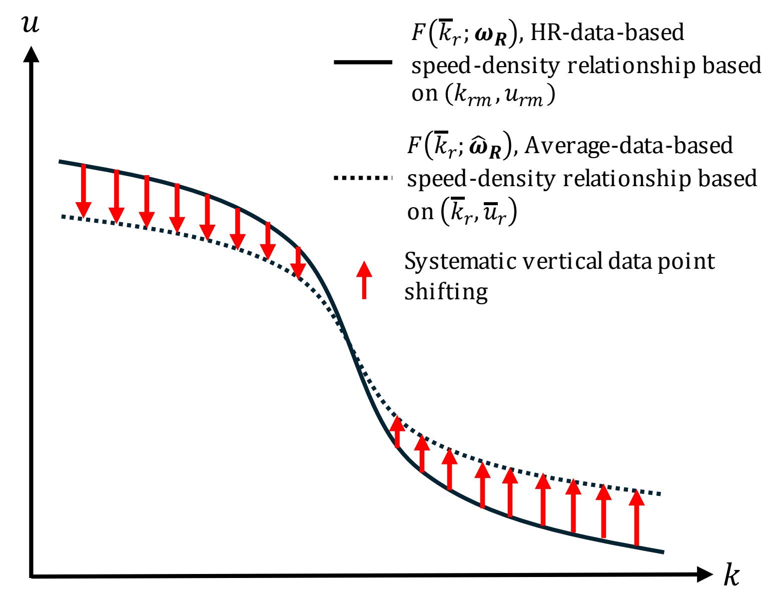

Figure 1. Illustration of the discrepancy between the HR-data-based speed-density relationship and the average-data-based speed-density relationship.

To quantify the bias embedded in the estimated average-data-based speed-density model, only the magnitude of the difference between the two models is considered. Define the absolute difference between the two models at a point associated with the rth LR interval, $|F(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}) - F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})|$, as the absolute bias at that point, $|\varepsilon|_r$. **Proposition 2** states that the average absolute bias of the average-data-based model is dependent on the variability of the HR density within each LR interval.

Proposition 2. *The average absolute bias of the average-data-based speed-density relationship estimated from the average data of set* $\boldsymbol{R}$ *is given by*

$$\overline{|\varepsilon|}_{\boldsymbol{R}} \cong \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}\left|\frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})}{\partial k_{rm}^2}\sigma_{k_r}^2\right| = \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}|D_r|, \tag{8}$$

where $\overline{|\varepsilon|}_{\boldsymbol{R}} \cong 0 \Leftrightarrow \sigma_{k1}^2 = \ ... = \sigma_{kr}^2 = \cdots = \sigma_{k|\boldsymbol{R}|}^2 = 0.$

Proof. Considering the rth LR interval, $\forall r \in [1, |\boldsymbol{R}|]$, as $F(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}) \cong \overline{u}_r \cong E_2(\overline{u}_r)$,

$$F(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}) - F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}}) \cong \frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})}{\partial k_{rm}^2}\sigma_{kr}^2. \tag{9}$$

To quantify the bias embedded in the average-data-based speed-density relationship estimated from the average data of set $\boldsymbol{R}$, the average absolute bias $\overline{|\varepsilon|}_{\boldsymbol{R}}$ is obtained by taking average of $|F(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}) - F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})|$, as shown in Eq. (10):

$$\overline{|\varepsilon|}_{\boldsymbol{R}} = \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}\left|F(\overline{k}_r; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}) - F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})\right|. \tag{10}$$

Substituting Eq. (9) into Eq. (10), $\overline{|\varepsilon|}_{\boldsymbol{R}}$ can be expressed as

$$\overline{|\varepsilon|}_{\boldsymbol{R}} \cong \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}\left|\frac{1}{2!}\frac{\partial^2 F(\overline{k}_r; \boldsymbol{\omega}_{\boldsymbol{R}})}{\partial k_{rm}^2}\sigma_{k_r}^2\right| = \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}|D_r|. \tag{11}$$

Thus, $\overline{|\varepsilon|}_{\boldsymbol{R}} \cong 0 \Leftrightarrow \sigma_{k1}^2 = \ ... = \sigma_{kr}^2 = \cdots = \sigma_{k|\boldsymbol{R}|}^2 = 0.$

■

If $\overline{|\varepsilon|}_{\boldsymbol{R}} \cong 0$, it implies that the discrepancy between the HR-data-based and average-data-based speed-density relationships is minimal, and thus the total biases embedded in the estimated parameters $\widehat{\boldsymbol{\omega}}_{\boldsymbol{R}}$ are also minimal.

3. Existence of an optimal dataset

The process of averaging the HR data to match with the LR data yields average data points ($\overline{k}_r$, $\overline{u}_r$), $\forall r \in [1,\ |\boldsymbol{R}|]$, with a compatible resolution. **Proposition 1** asserts that such process could result in systematic vertical data point shifting. Thus, average data points ($\overline{k}_r$, $\overline{u}_r$) comprise both the information carried over from the HR data and the systematic data point distortion produced by the averaging process. **Proposition 2** states that the average absolute bias, $\overline{|\varepsilon|}_{\boldsymbol{R}}$, is given by the average of $|D_r|$, $\forall r \in [1,\ |\boldsymbol{R}|]$. If the average data point with the highest value of $|D_r|$ is discarded from average data of set $\boldsymbol{R}$, it is anticipated that the average absolute bias, $\overline{|\varepsilon|}_{\boldsymbol{R}}$, will decrease. This is because removing the data point with the greatest distortion has a more significant impact on bias reduction compared to the loss of information from the removal due to the initial large size of the dataset. However, as more data points are removed, the loss of information could become the dominant effect due to the diminished size of the remaining dataset. This could result in an increase in the average absolute bias, $\overline{|\varepsilon|}_{\boldsymbol{R}}$, due to the substantial information depletion. **Proposition 3** states that an optimal dataset with the least average absolute bias exists.

Proportion 3. *Given any set $\boldsymbol{R}$, $\exists$ an optimal set $\boldsymbol{C}\ s.t.\ \boldsymbol{C} \subseteq \boldsymbol{R} \wedge$ the average absolute bias of the average-data-based speed-density relationship estimated from the average data of set $\boldsymbol{C}$, $\overline{|\varepsilon|}_{\boldsymbol{C}}$, is minimized.* $\boldsymbol{C} = \boldsymbol{R} \Leftrightarrow \sigma_{k1}^2 = \ \ldots = \sigma_{kr}^2 = \cdots = \sigma_{k|\boldsymbol{R}|}^2 = 0$.

Proof. Given any set $\boldsymbol{R}$, it can be decomposed into two subsets $\boldsymbol{R_1}$ and $\boldsymbol{R_2}\ s.t.\ \boldsymbol{R_1} \subseteq \boldsymbol{R}$, $\boldsymbol{R_2} \subseteq \boldsymbol{R}$, $\boldsymbol{R_1} \cap \boldsymbol{R_2} = \emptyset$, and $\boldsymbol{R_1} + \boldsymbol{R_2} = \boldsymbol{R}$. $\boldsymbol{R_2}$ comprises a set of discarded HR data with relatively large values of $|D_{r_2}|$ and $\boldsymbol{R_1}$ contains a set of remaining HR data with relatively small values of $|D_{r_1}|$

$s.t. \, |D_{r_2}| \geq |D_{r_1}|, \forall r_2 \in [1, \; |\boldsymbol{R_2}|]$ and $r_1 \in [1, |\boldsymbol{R_1}|]$. The average absolute bias of the average-data-based speed-density relationship estimated from the average data of set $\boldsymbol{R_1}$ is given by

$$\overline{|\varepsilon|}_{\boldsymbol{R_1}} \cong \frac{1}{|\boldsymbol{R_1}|} \sum_{r_1=1}^{|\boldsymbol{R_1}|} \left| F\left(\bar{k}_{r_1}; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega_R}\right) \right|, \tag{12}$$

where $\widehat{\boldsymbol{\omega}}_{\boldsymbol{R_1}} = \left\{\widehat{\omega}_{\boldsymbol{R_1}1}, \widehat{\omega}_{\boldsymbol{R_1}2}, \ldots, \widehat{\omega}_{\boldsymbol{R_1}n}\right\}$ is the vector of model parameters estimated based on the average data from set $\boldsymbol{R_1}$. It follows

$$\overline{|\varepsilon|}_{\boldsymbol{R_1}} \cong \frac{1}{|\boldsymbol{R_1}|} \sum_{r_1=1}^{|\boldsymbol{R_1}|} \left| \left[F\left(\bar{k}_{r_1}; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right)\right] + \left[F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega_R}\right)\right] \right|. \tag{13}$$

Eq. (13) decomposes the average absolute bias, $\overline{|\varepsilon|}_{\boldsymbol{R_1}}$, into $\left[F\left(\bar{k}_{r_1}; \widehat{\boldsymbol{\omega}}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right)\right]$, representing the discrepancy between the average-data-based and HR-data-based models estimated from set $\boldsymbol{R_1}$, and $\left[F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega_R}\right)\right]$, representing the discrepancy between the HR-data-based models estimated from set $\boldsymbol{R_1}$ and the HR-data-based models estimated from set $\boldsymbol{R}$. Using Eq. (9),

$$\overline{|\varepsilon|}_{\boldsymbol{R_1}} \cong \frac{1}{|\boldsymbol{R_1}|} \sum_{r_1=1}^{|\boldsymbol{R_1}|} \left| \frac{1}{2!} \frac{\partial^2 F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right)}{\partial k_{r_1 m}^2} \sigma_{kr_1}^2 + \left[F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right) - F\left(\bar{k}_{r_1}; \boldsymbol{\omega_R}\right)\right] \right|. \tag{14}$$

Considering the initial stage of data point removal where $\boldsymbol{R_1} = \boldsymbol{R}$ and $\boldsymbol{R_2} = \emptyset$, as $F\left(\bar{k}_r; \boldsymbol{\omega_R}\right) - F\left(\bar{k}_r; \boldsymbol{\omega_R}\right) = 0$,

$$\overline{|\varepsilon|}_{\boldsymbol{R}} \cong \frac{1}{|\boldsymbol{R}|} \sum_{r=1}^{|\boldsymbol{R}|} \left| \frac{1}{2!} \frac{\partial^2 F\left(\bar{k}_r; \boldsymbol{\omega_R}\right)}{\partial k_{rm}^2} \sigma_{kr}^2 \right|. \tag{15}$$

When the first average data point with the highest value of $|D_r|$ is discarded from average data of set $\boldsymbol{R}$, $|\boldsymbol{R_1}| = |\boldsymbol{R}| - 1$ and $|\boldsymbol{R_2}| = 1$. Since $|\boldsymbol{R}| \cong |\boldsymbol{R_1}| \gg |\boldsymbol{R_2}|$, the loss of information is minimal and $F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right) \cong F\left(\bar{k}_{r_1}; \boldsymbol{\omega_R}\right)$. Thus,

$$\overline{|\varepsilon|}_{\boldsymbol{R_1}} \cong \frac{1}{|\boldsymbol{R_1}|} \sum_{r_1=1}^{|\boldsymbol{R_1}|} \left| \frac{1}{2!} \frac{\partial^2 F\left(\bar{k}_{r_1}; \boldsymbol{\omega}_{\boldsymbol{R_1}}\right)}{\partial k_{r_1 m}^2} \sigma_{kr_1}^2 \right|. \tag{16}$$

Since $|D_{r_2}| \geq |D_{r_1}|, \forall r_2 \in [1, \; |\boldsymbol{R_2}|]$ and $r_1 \in [1, |\boldsymbol{R_1}|]$,

$$\frac{1}{|\boldsymbol{R_1}|}\sum_{r_1=1}^{|\boldsymbol{R_1}|}\left|\frac{1}{2!}\frac{\partial^2 F\left(\overline{k}_{r_1};\boldsymbol{\omega}_{\boldsymbol{R_1}}\right)}{\partial k_{r_1 m}^2}\sigma_{kr_1}^2\right| \leq \frac{1}{|\boldsymbol{R}|}\sum_{r=1}^{|\boldsymbol{R}|}\left|\frac{1}{2!}\frac{\partial^2 F\left(\overline{k}_r;\boldsymbol{\omega}_{\boldsymbol{R}}\right)}{\partial k_{rm}^2}\sigma_{k_r}^2\right|. \tag{17}$$

Eq. (17) shows that $\overline{|\varepsilon|}_{\boldsymbol{R_1}} \leq \overline{|\varepsilon|}_{\boldsymbol{R}}$. This implies that as $|\boldsymbol{R_2}|$ initially increases from zero to one, $\overline{|\varepsilon|}_{\boldsymbol{R_1}}$ gradually decreases.

Considering the later stage of data point removal where $|\boldsymbol{R_2}|$ is sufficiently large $s.t.$ $|D_1| \cong \cdots \cong \left|D_{r_1}\right| \cong \cdots \cong \left|D_{|\boldsymbol{R_1}|}\right| \cong 0, \forall r_1 \in [1, |\boldsymbol{R_1}|]$,

$$\overline{|\varepsilon|}_{\boldsymbol{R_1}} \cong \frac{1}{|\boldsymbol{R_1}|}\sum_{r_1=1}^{|\boldsymbol{R_1}|}\left|F\left(\overline{k}_{r_1};\boldsymbol{\omega}_{\boldsymbol{R_1}}\right) - F\left(\overline{k}_{r_1};\boldsymbol{\omega}_{\boldsymbol{R}}\right)\right|. \tag{18}$$

When an additional average data point is removed from set $\boldsymbol{R_1}$, define $\boldsymbol{R}_{\boldsymbol{1}}^{'}$ to be the new set containing the remaining HR data and $\boldsymbol{R}_{\boldsymbol{2}}^{'}$ to be the new set comprising the discarded HR data $s.t.$ $|\boldsymbol{R}'_{\boldsymbol{1}}| = |\boldsymbol{R_1}| - 1$ and $|\boldsymbol{R}'_{\boldsymbol{2}}| = |\boldsymbol{R_2}| + 1$. Since $|D_1| \cong \cdots \cong \left|D_{r'_1}\right| \cong \cdots \cong \left|D_{|\boldsymbol{R}'_{\boldsymbol{1}}|}\right| \cong 0$, $\forall r'_1 \in [1, |\boldsymbol{R}'_{\boldsymbol{1}}|]$,

$$\overline{|\varepsilon|}_{\boldsymbol{R}'_{\boldsymbol{1}}} \cong \frac{1}{|\boldsymbol{R}'_{\boldsymbol{1}}|}\sum_{r_1{}'=1}^{|\boldsymbol{R}'_{\boldsymbol{1}}|}\left|F\left(\overline{k}_{r'_1};\boldsymbol{\omega}_{\boldsymbol{R}'_{\boldsymbol{1}}}\right) - F\left(\overline{k}_{r'_1};\boldsymbol{\omega}_{\boldsymbol{R}}\right)\right|. \tag{19}$$

As $|\boldsymbol{R_2}|$ is sufficiently large, the loss of information due to this additional removal is non-negligible. Thus,

$$\frac{1}{|\boldsymbol{R_1}|}\sum_{r_1=1}^{|\boldsymbol{R_1}|}\left|F\left(\overline{k}_{r_1};\boldsymbol{\omega}_{\boldsymbol{R_1}}\right) - F\left(\overline{k}_{r_1};\boldsymbol{\omega}_{\boldsymbol{R}}\right)\right| \leq \frac{1}{|\boldsymbol{R}'_{\boldsymbol{1}}|}\sum_{r_1{}'=1}^{|\boldsymbol{R}'_{\boldsymbol{1}}|}\left|F\left(\overline{k}_{r'_1};\boldsymbol{\omega}_{\boldsymbol{R}'_{\boldsymbol{1}}}\right) - F\left(\overline{k}_{r'_1};\boldsymbol{\omega}_{\boldsymbol{R}}\right)\right|. \tag{20}$$

Eq. (20) shows that $\overline{|\varepsilon|}_{\boldsymbol{R_1}} \leq \overline{|\varepsilon|}_{\boldsymbol{R}'_{\boldsymbol{1}}}$. This implies that when $|\boldsymbol{R_2}|$ is sufficiently large, as $|\boldsymbol{R_2}|$ further increases, $\overline{|\varepsilon|}_{\boldsymbol{R_1}}$ increases.

As $\overline{|\varepsilon|}_{\boldsymbol{R_1}}$ decreases during the initial stage of data point removal and increases during the later stage of data point removal, $\exists$ an optimal set $\boldsymbol{C}$ $s.t.$ $\boldsymbol{C} \subseteq \boldsymbol{R}$ and the average absolute bias of the average-data-based speed-density relationship estimated from the average data of set $\boldsymbol{C}$, $\overline{|\varepsilon|}_{\boldsymbol{C}}$, is minimized. It is trivial to prove that $\boldsymbol{C} = \boldsymbol{R} \Leftrightarrow \sigma_{k1}^2 = \ldots = \sigma_{kr}^2 = \cdots = \sigma_{k|\boldsymbol{R}|}^2 = 0$.

■

4. Practical optimal dataset determination

To minimize the average absolute bias, it is essential to determine the optimal dataset. The metric $|D_r|, \forall r \in [1, |\boldsymbol{R}|]$, measures the variability of HR density data within the rth LR interval. Define $|D^c|$ to be the critical value of $|D_r|$. $\forall r \in [1, |\boldsymbol{R}|]$, if $|D_r| > |D^c|$, then the associated data are discarded. The remaining data form the optimal set $\boldsymbol{C}$. However, evaluating $|D_r|$ presents challenge as it necessitates the traffic flow model, $F(.)$, and density variance of the HR density data within the rth LR interval, $\sigma_{k_r}^2$, which both are unavailable. This practical challenge hinders the determination of exact optimal dataset $\boldsymbol{C}$.

Given the inherent correlation between speed and density in the same physical transportation system, it follows that the variability of HR density should also be positively correlated with the variability of HR speed. In essence, as HR speed data are available, it becomes feasible to indirectly assess the variability of HR density via a metric quantifying the variability of HR speed. In this study, the coefficient of variation of HR speed is chosen as a proxy measure for quantifying the variability of HR density. The coefficient of variation of HR speed data within the rth LR interval, denoted as CV_{ur}, is given by

$$CV_{ur} = \frac{\sigma_{ur}}{\overline{u}_r}, \tag{21}$$

where $\overline{u}_r$ is the mean of the HR speed data within the rth LR interval, σ_{ur} is the standard deviation of the HR speed data within the rth LR interval and $\sigma_{ur} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(u_{rm} - \overline{u}_r)^2}, \forall r \in [1, |\boldsymbol{R}|]$. Define CV_u^c to be the critical value of CV_{ur} that corresponds to $|D^c|$ in the dimension of $|D_r|$. $\forall r \in [1, |\boldsymbol{R}|]$, if $CV_{ur} > CV_u^c$, then the associated data are discarded. The remaining data constitute a dataset $\boldsymbol{C}'$, which differs from the exact optimal dataset $\boldsymbol{C}$ and is termed as a practical optimal dataset for application. As the value of $|D^c|$ is unknown in practice, obtaining the exact optimal dataset $\boldsymbol{C}$ is not feasible. Nevertheless, due to the inherent correlation between speed and density in the same physical transportation system, the practical optimal dataset $\boldsymbol{C'}$ serves as a suitable substitute for the exact optimal dataset $\boldsymbol{C}$. Therefore, identifying CV_u^c is crucial for determining the practical optimal dataset.

The geographical proximity often leads to a correlation between the traffic dynamics of a non-strategic link and a nearby strategic link. Hence, it is not unreasonable to anticipate that the CV_u^c of the non-strategic link should be similar to that of the strategic link. By utilizing the available HR data from the strategic link, the CV_u^c corresponding to the $|D^c|$ in the dimension of $|D_r|$ can be identified. This CV_u^c can be used for practical optimal dataset determination for the non-strategic link. Detailed procedures of practical optimal dataset determination are outlined as follows:

(1) Estimate the HR-data-based traffic flow model for the strategic link.

(2) Enumerate a set of candidate $|D^c|$ values for a given LR interval.

(3) For each candidate $|D^c|$, construct the corresponding candidate optimal dataset, set $\boldsymbol{C}$, by removing data points with $|D_r| > |D^c|$, estimate the average-data-based traffic flow model, and evaluate the average absolute bias, $\overline{|\varepsilon|}_{\boldsymbol{C}}$, using **Proposition 2**.

(4) Identify the candidate $|D^c|$ with the least value of $\overline{|\varepsilon|}_{\boldsymbol{C}}$ as the $|D^c|$.

(5) Establish the $|D_r| - CV_{ur}$ relationship based on the HR data of the strategic link and identify the CV_u^c corresponding to the $|D^c|$.

(6) Use the identified CV_u^c to determine the practical optimal dataset, set $\boldsymbol{C'}$, for the non-strategic link by removing data points with $CV_{ur} > CV_u^c$.

5. Case study

To validate and demonstrate the applicability and performance of the proposed method, real-world HR traffic data from four sites in Hong Kong and Nanjing, China were employed to simulate scenarios where HR data is available for a strategic link and multi-resolution data is accessible for non-strategic links.

5.1. Data collection and processing

The four sites included a major urban three-lane expressway in Hong Kong Island (Site 1), an urban two-lane road in Hong Kong Island (Site 2), an urban two-lane road of in Kowloon Peninsula (Site 3), and an urban four-lane road in Nanjing (Site 4). The HR traffic data for Sites 1 and 2 was collected between January 1 and December 31, 2017. For Site 3, data was collected from January 1 and December 31, 2018, and for Site 4, data was collected between September 1 to November 30, 2023.

For each of these sites, the space mean speed u_m and flow rate f_m were recorded at 2-min intervals, with m representing the index of the mth observation. The space mean speed recorded at the 2-min interval was used as HR speed. HR density k_m is calculated using the formula f_m/u_m. The raw traffic data from these four sites was cleaned to ensure their validity. Firstly, outlier observations caused by malfunctioning traffic detectors were removed. Then, all observations with traffic counts less than five were excluded due to their unreliability. The resulting dataset consisted of 251,721 observations from Site 1, 242,369 observations from Site 2, 256,073 observations from Site 3 and 45,536 observations from Site 4. The availability of the HR data of the four sites enables the estimations of HR-data-based models and the evaluations of average absolute biases.

To replicate situations where multi-resolution data is available for non-strategic links, average data has to be constituted. The average speed $\overline{u}$ and average flow rate $\overline{f}$ was obtained by taking average of the speed and flow data recorded at 2-min intervals, respectively. The LR interval was chosen to be either 30-min or 60-min interval. M is the total number of HR data point within a LR interval. The average density $\overline{k}$ over the LR interval can be approximated as $\overline{f}/\overline{u}$.

Site 1 was chosen to mimic a strategic link with HR data for the determination of CV_u^c. Sites 2, 3 and 4 were selected to simulate non-strategic roads with multi-resolution data. The CV_u^c identified based on HR data from Site 1 was then applied to Sites 2, 3 and 4 for determining practical optimal datasets for traffic flow model estimations. Note that both Sites 1 and 2 were located in Hong Kong Island, making Site 1 a reasonable proxy for a nearby strategic link in relation to the non-strategic link of Site 2. However, Site 3 was situated in Kowloon Peninsula, which is geographically detached from Hong Kong Island, and Site 4 was located in another city. Therefore, applying the CV_u^c identified from Site 1 to Sites 3 and 4 tested its transferability across different networks.

5.2. Model selection

Since the pioneering work by Greenshields et al. (1935), the understanding of speed-density relationships has evolved significantly with the advent of analytical and experimental models (Cheng et al., 2021; Mohammadian et al., 2021; Wang et al., 2022; Yin et al., 2022). These traffic flow models have exhibited diverse model formulations and parameters. For instance, Greenshields et al (1935) proposed a basic linear model with parameters of the free-flow speed u_f and jam density k_j to depict the decreasing relationship between speed and density, which laid the foundation for subsequent developments. The overall performance of this model was enhanced through several modifications by Gazis et al. (1961), Pipes (1967), Drew (1964), and Jayakrishnan et al. (1995). Newell (1961) and Franklin (1961) proposed a nonlinear traffic flow model that uses the free-flow speed u_f, jam density k_j, and kinematic wave speed C_j at jam density as model parameters. Del Castillo and Benitez (1995) then further refined Newell's model, derived through dimensional analysis of a general car-following model. The refined model incorporates essential properties that speed-density relationships must satisfy. Underwood (1961) proposed an exponential functional form of the speed-density model that incorporates the free-flow speed u_f and optimal density k_0. This model was further extended by Drake et al. (1967). More recently, Wang et al. (2011) recently introduced a family of logistic speed-density models, namely 3PL, 4PL, and 5PL, which include varying numbers of parameters. These models incorporate the free-flow speed u_f, the optimal density k_0, and one to three additional parameters. Furthermore, Cheng et al. (2021) proposed a novel S-shaped three-parameter (S3) traffic flow model to depict the relationships among flow, speed, and density. This model incorporates the free-flow speed u_f, optimal density k_0, and maximum flow inertia coefficient m. To ensure that a diverse representation of different functional forms is considered, five traffic flow models were carefully selected to test the practicality and effectiveness of the proposed method. Each selected model represents one of the abovementioned model families. Table 1 summarizes the name, formulation, and parameters of the selected traffic flow models, along with their respective model families.

Table 1. Selected traffic flow models

Model	Functional form	Parameters	Model family
S3 model	$u = \dfrac{u_f}{[1+(k/k_0)^m]^{\frac{2}{m}}}$	u_f, k_0, m	A new family of s-shaped three-parameter traffic flow model
4PL model	$u = u_b + \dfrac{u_f - u_b}{1+\exp\left(\frac{k-k_0}{\theta}\right)}$	u_f, k_0, u_b, θ	The family of logistic speed-density models
Underwood-class model	$u = u_f \exp\left[-\frac{1}{n}\left(\frac{k}{k_0}\right)^n\right]$	u_f, k_0	The family of Underwood-type models
NF model	$u = u_f\left\{1-\exp\left[\frac{C_j}{u_f}\left(1-\frac{k_j}{k}\right)\right]\right\}$	u_f, k_j, C_j	The family of NF-type models
Pipe's model	$u = u_f\left(1-\frac{k}{k_j}\right)^2$	u_f, k_j	The family of Greenshields-type models

Note: u represents the space mean speed; k represents the density; u_f represents the free-flow speed; k_0 represents the optimal density; k_j represents the jam density; C_j represents the absolute value of the kinematic wave speed at jam density; m and θ are the parameters in different models.

5.3. Determination of CV_u^c

Site 1 was selected to simulate a strategic link with HR data for the determination of CV_u^c. As five traffic flow models (shown in Table 1) and two LR intervals (30-min and 60-min) were considered, there were a total of ten cases and ten CV_u^c to be determined. For each case, the HR-data-based traffic flow model was first estimated using the HR data from Site 1. In addition to the complete dataset, candidate optimal datasets were constituted based on a set of selected candidate $|D^c|$ ranging from 30 km/h to 1 km/h with a step of 1 km/h. Based on the candidate optimal datasets, the average-data-based traffic flow model was estimated and the average absolute biases were evaluated.

The results for $|D^c|$ determinations for the S3 models are presented in Tables 2 and 3 for the 30-min and 60-min LR intervals, respectively. For ease of presentation, the results for the complete dataset and the selected candidate $|D^c|$ ranging from 12 km/h to 8 km/h with a step of 1 km/h are presented. For both the 30-min and 60-min LR intervals, clear convex relationships in the average absolute bias were observed during the data point removal processes. For the S3 model with a 30-min LR interval, the average absolute bias dropped from 2.5627 km/h to 0.1026 km/h as data points were removed from the complete dataset using a candidate $|D^c|$ value of 12 km/h. Subsequently, as the data point removal process continued, the average absolute bias reached its lowest value of 0.0835 km/h at a candidate $|D^c|$ value of 10 km/h. However, any further removal of data points with a lower candidate $|D^c|$ value resulted in an increase of the average absolute bias. Similarly, for the 60-min LR interval, the average absolute bias decreased from 2.5627 km/h to 0.3186 km/h as data points were removed from the complete dataset using a candidate $|D^c|$ value of 12 km/h. As the data point removal process further proceeded, the average absolute bias reached its lowest value of 0.2772 km/h at a candidate $|D^c|$ value of 10 km/h. However, any further removal of data points with a lower candidate $|D^c|$ value resulted in an increase of the average absolute bias. The results of determining $|D^c|$ for the other four traffic flow models for the 30-min and 60-min LR intervals are presented in Table A1-A8 in **Appendix A**. Similar clear convex relationships in the average absolute bias were observed during the data point removal processes. These results empirically validated **Proposition 3** that for any given set $\boldsymbol{R}$, $\exists$ an optimal set $\boldsymbol{C}$ $s.t.$ $\boldsymbol{C} \subseteq \boldsymbol{R}$ and the average absolute bias of the average-data-based model estimated from the average data of set $\boldsymbol{C}$, $\overline{|\varepsilon|}_{\boldsymbol{C}}$, is minimized.

Table 2. Determination of $|D^c|$ for the S3 model with a 30-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.42	0.0114	2.5627
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	27.81	0.0181	
	$m/\hat{m}$	2.573	0.0033	2.698	0.0023	
12	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.12	0.0114	0.1026
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.15	0.0203	
	$m/\hat{m}$	2.573	0.0033	2.560	0.0024	
11	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.12	0.0112	0.0896
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.17	0.0207	

	$m/\hat{m}$	2.573	0.0033	2.560	0.0024	
10	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.12	0.0111	**0.0835**
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.21	0.0212	
	$m/\hat{m}$	2.573	0.0033	2.559	0.0023	
9	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.13	0.0111	0.1479
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.29	0.0219	
	$m/\hat{m}$	2.573	0.0033	2.553	0.0023	
8	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.17	0.0108	0.3090
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.39	0.0178	
	$m/\hat{m}$	2.573	0.0033	2.537	0.0023	

Table 3. Determination of $|D^c|$ for the S3 model with a 60-min LR interval

Candidate $|D^c|$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.42	0.0114	2.5627
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	27.81	0.0181	
	$m/\hat{m}$	2.573	0.0033	2.698	0.0023	
12	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.17	0.0114	0.3186
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	30.96	0.0204	
	$m/\hat{m}$	2.573	0.0033	2.542	0.0024	
11	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.16	0.0113	0.2932
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.01	0.0207	
	$m/\hat{m}$	2.57	0.0033	2.542	0.0024	
10	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.16	0.0111	**0.2772**
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.03	0.0212	
	$m/\hat{m}$	2.573	0.0033	2.542	0.0023	
9	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.18	0.0111	0.3036
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.07	0.0219	
	$m/\hat{m}$	2.573	0.0033	2.537	0.0023	
8	$u_f/\hat{u}_f$(km/h)	82.10	0.0165	82.20	0.0108	0.3394
	$k_0/\hat{k}_0$ (veh/km/lane)	31.22	0.0252	31.18	0.0178	
	$m/\hat{m}$	2.573	0.0033	2.524	0.0023	

Figures 2a and 2b illustrate the established linear relationships between $|D_r|$ and CV_{ur}, with R^2 values of 0.8626 and 0.8998 for the 30-min and 60-min LR intervals, respectively. These high R^2 values indicated strong positive correlations exist between $|D_r|$ and CV_{ur}, providing empirical support for the hypothesized inherent correlation between speed and density within the same physical transportation system. Using the established $|D_r| - CV_{ur}$ relationships, the CV_u^c values corresponding to the $|D^c|$ value of 10 km/h for the 30-min LR interval and 60-min LR interval were determined to be 0.3535 and 0.3483, respectively. In each case, the complete dataset was divided into two groups based on the identified CV_u^c value. Subsequently, a Kolmogorov-Smirnova test was utilized to compare the $|D_r|$ values of members from the two groups. The maximum differences in the cumulative probability functions were found to be 0.821, and 0.794 for the cases with 30-min LR interval and 60-min LR interval, respectively. Consequently, the null hypothesis, which assumes that the $|D_r|$ values of the two groups were drawn from the same distribution, was rejected. This implies that the distributions of the $|D_r|$ values in the two groups were statistically significantly different. The results of determining CV_u^c for the other four traffic flow models for the 30-min and 60-min LR intervals are illustrated in Figure B1-B4 in **Appendix B**. Similar linear relationships between $|D_r|$ and CV_{ur} with high R^2 values were observed. Table 4 summarizes the determined CV_u^c for the ten cases with different traffic flow models and LR interval combinations. It is evident that the identified ten CV_u^c only varied within a small range from 0.3483 to 0.4959.

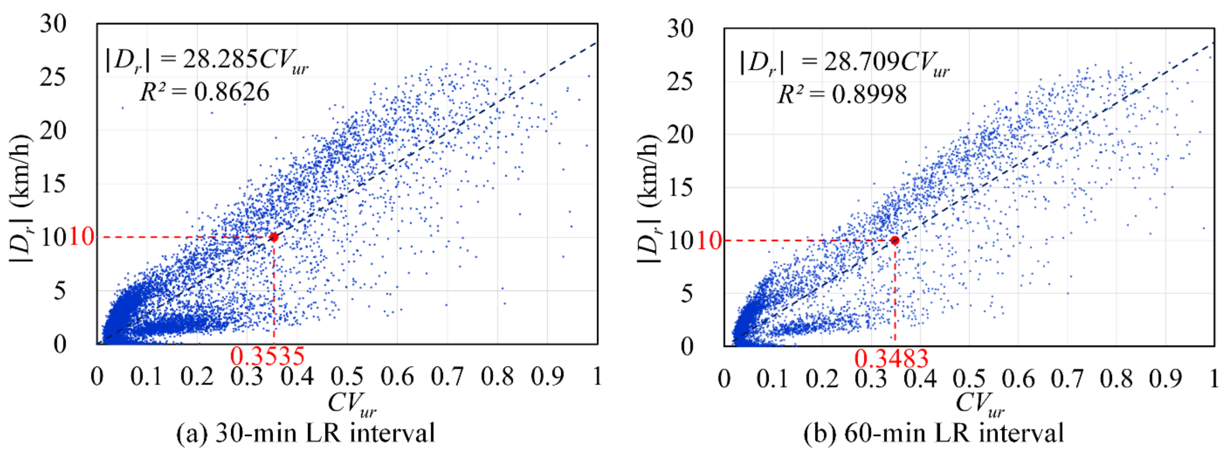

Figure 2. Established linear relationships between $|D_r|$ and CV_{ur} for the S3 model with (a) 30-min LR interval and (b) 60-min LR interval.

Table 4. Summary of CV_u^c for practical optimal dataset determinations

Model	Functional form	Parameter	LR interval	$\lvert D^c \rvert$ (km/h)	CV_u^c
S3 model	$u = \dfrac{u_f}{[1+(k/k_0)^m]^{\frac{2}{m}}}$	u_f, k_0, m	30	10	0.3535
			60	10	0.3483
4PL model	$u = u_b + \dfrac{u_f - u_b}{1+\exp\left(\dfrac{k-k_0}{\theta}\right)}$	u_f, k_0, u_b θ	30	10	0.3960
			60	9	0.3603

Underwood-class model	$u = u_f \exp\left[-\frac{1}{m}\left(\frac{k}{k_0}\right)^m\right]$	u_f, k_0, m	30	15	0.4768
			60	15	0.4639
NF model	$u = u_f\left\{1 - \exp\left[\frac{C_j}{u_f}\left(1 - \frac{k_j}{k}\right)\right]\right\}$	u_f, k_j, C_j	30	12	0.4056
			60	12	0.3966
Pipe's model	$u = u_f\left(1 - \frac{k}{k_j}\right)^2$	u_f, k_j	30	14	0.4959
			60	14	0.4836

5.4. Applicability and transferability of CV_u^c

Sites 2, 3 and 4 were chosen to mimic non-strategic links with multi-resolution data. For both the 30-min and 60-min LR intervals, the CV_{ur} values of the HR speed data from Sites 2, 3 and 4 were evaluated. The CV_u^c that was identified based on HR data from Site 1 was then applied to Sites 2, 3 and 4 to determine practical optimal datasets for traffic flow model estimations. Due to the narrow range of CV_u^c identified for the ten cases in the previous subsection, a CV_u^c value of 0.4 was chosen for ease of application. The practical optimal datasets for both the 30-min and 60-min LR intervals at the three sites comprised data points with a CV_{ur} less than or equal to 0.4. As five traffic flow models and two LR intervals were considered, a total of ten traffic flow models were estimated based on the constituted practical optimal datasets for each of these sites. For evaluation purposes, the HR-data-based models using HR data and the average-data-based models using the complete datasets were also estimated for the ten cases at each of these sites.

Table 5 presents the model estimation results of the S3 model for both the 30-min and 60-min LR intervals at the three sites. Results indicate that the average-data-based models estimated from the constituted practical optimal datasets consistently outperformed the models estimated from the complete datasets due to the reduced average absolute biases. For instance, in the case of Site 3 with the 30-min and 60-min LR intervals, the average absolute biases of the average-data-based model based on the complete datasets at 4.377 km/h and 6.503 km/h were significantly reduced to 1.982 km/h and 2.229 km/h, respectively, when the models were estimated from the practical optimal datasets, leading to notable decreases of 54.7% and 65.7% in the average absolute biases. In general, the variability of HR data within the LR interval increases with the length of the LR interval. The lost variability of HR data within the LR intervals during the averaging process also increases with the length of the LR interval. Consequently, the average absolute bias of average-data-based models for the 30-min LR interval was generally smaller than that of models for the 60-min LR interval. The model estimation results of the other four traffic flow models for both the 30-min and 60-min LR intervals at the three sites are presented in Tables C1 to C4 in **Appendix C**. Similarly, results demonstrate that the average-data-based models estimated from the practical optimal datasets consistently outperformed the models estimated from the complete datasets owing to the substantial reduction in the average absolute biases.

The average absolute bias of an average-data-based model can be minimized if the practical optimal dataset can be determined. In theory, the accuracy of the proposed procedures for practical optimal dataset determination is mainly governed by three key factors: (1) the granularity of the enumeration of candidate $|D^c|$ in Step 2 of the proposed procedures, (2) the strength of the correlation between speed and density in the same physical transportation system, and (3) the strength of the correlation between the traffic dynamics of a non-strategic link and that of a nearby strategic link. However, in this case study, a single CV_u^c value of 0.4 was applied to all cases. Moreover, while it was still reasonable to use Site 1 as a proxy for a nearby strategic link to the non-strategic link of Site 2 due to their geographical closeness in Hong Kong Island, Site 3 was located in Kowloon Peninsula that is geographically detected from Hong Kong Island and Site 4 was even situated in another city. The first and third key governing factors could barely be satisfied. Nevertheless, results of all the cases still show that the average-data-based models estimated from the practical optimal datasets consistently outperformed the models estimated from the complete datasets, suggesting the robustness to the choice of CV_u^c and the transferability to different networks of the proposed method. These favorable properties were guaranteed by **Proposition 3**, which states that for any given set $\boldsymbol{R}$, $\exists$ an optimal set $\boldsymbol{C}\ s.t.\ \boldsymbol{C} \subseteq \boldsymbol{R}$ and $\overline{|\varepsilon|}_{\boldsymbol{C}}$ is minimized, and $\boldsymbol{C} = \boldsymbol{R} \Leftrightarrow \sigma_{k1}^2 = \ ... = \sigma_{kr}^2 = \cdots = \sigma_{k|\boldsymbol{R}|}^2$. In most real-world situations, it is nearly impossible to have zero variability in HR data within the LR interval across the entire observation period. Therefore, removing an adequate amount of data with high CV_{ur} values should generally result in more accurate model estimations. From a practical standpoint, if HR data from a nearby strategic link is available, it is recommended to update the CV_u^c for the non-strategic link. Nonetheless, in cases where HR data is unavailable, the reported case study provides empirical support for considering a CV_u^c value of 0.4 as a viable alternative option.

Table 5. Model estimations of the S3 model for the 30-min and 60-min LR intervals at Sites 2, 3 and 4

Site	LR interval (min)	Parameter	HR-data-based model		Average-data-based model based on complete dataset		Average-data-based model based on practical optimal dataset	
			Mean	SD	Mean	Average absolute bias (km/h)	Mean	Average absolute bias (km/h) [% change]
2	30	$u_f/\hat{u}_f$(km/h)	67.38	0.018	67.37	0.826	67.58	**0.305 [-63.1%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	44.69	0.075	42.16		44.24	
		$m/\hat{m}$	2.34	0.005	2.40		2.31	
	60	$u_f/\hat{u}_f$(km/h)	67.38	0.018	67.40	1.223	67.66	**0.453 [-63.0%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	44.69	0.075	40.83		44.40	
		$m/\hat{m}$	2.34	0.005	2.42		2.28	
3	30	$u_f/\hat{u}_f$(km/h)	66.10	0.021	66.74	4.377	66.63	**1.982 [-54.7%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	31.02	0.039	29.64		30.41	
		$m/\hat{m}$	6.64	0.039	5.81		6.30	
	60	$u_f/\hat{u}_f$(km/h)	66.10	0.021	67.00	6.503	66.82	**2.229 [-65.7%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	31.02	0.039	28.33		30.18	
		$m/\hat{m}$	6.64	0.039	6.84		6.74	
4	30	$u_f/\hat{u}_f$(km/h)	65.76	0.072	65.73	0.583	65.66	**0.136 [-76.7%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	32.41	0.449	31.71		32.50	
		$m/\hat{m}$	1.99	0.025	2.02		1.99	
	60	$u_f/\hat{u}_f$(km/h)	65.76	0.072	65.66	0.857	65.86	**0.294 [-65.7%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	32.41	0.449	30.33		31.99	
		$m/\hat{m}$	1.99	0.025	2.03		1.98	

6. Conclusions

Estimating traffic flow models based on multi-resolution data is a common occurrence in real-world scenarios. A straightforward approach to address this resolution incompatibility is to average the HR data to align with the LR data. However, this study has demonstrated the importance of considering the variability of HR data within the LR interval in the process of estimating traffic flow models. It has been proven that neglecting this variability could lead to systematic distortions in the data and, consequently, biased model estimations. To quantify the bias introduced into average-data-based models due to the lost variability, the average absolute bias was proposed. Most importantly, this study proved that for any given complete average data dataset, there must exist an optimal dataset that minimizes the average absolute bias in model estimations introduced by the averaging process. Subsequently, the novel procedure for determining the practical optimal dataset was proposed.

To verify the applicability of the proposed method, real-world HR traffic data were collected from four sites in Hong Kong and Nanjing to simulate the scenario where only multi-resolution data was available. The results have consistently demonstrated that the average-data-based models estimated from the determined practical optimal datasets outperformed the models estimated from the complete datasets. This case study provides empirical support for the robustness and transferability of the proposed method, offering a solution to the challenges associated with collecting complete HR traffic data and providing a reliable method for traffic flow model estimation in situations involving multi-resolution data. While this study focuses on estimating the speed-density relationship based on multi-resolution data and reducing biases in the estimated model, exploring the complex system transition dynamics within the speed-density relationship is also important. Future research will aim to enhance the understanding of these dynamics by assessing the HR data and average data using metrics such as the signal-to-noise ratio.

Acknowledgments

The work described in this paper was supported by funding from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. 17202223, R5029-18, and R7027-18), the Strategic Public Policy Research Funding Scheme (Project No. S2019.A8.013.19S) and the National Natural Science Foundation of China (Project No. 51925801, 52302376 and 52302433). The last author was also supported by the Francis S Y Bong Professorship in Engineering.

References

Ambühl, L., and Menendez, M. (2016). Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C: Emerging Technologies*, 71, 184–197.

Ali-Eldin, A., Elmroth, E. (2021). Data management, communication systems and the edge: Challenges for the future of transportation. *Communications in Transportation Research*, 1, 100024.

Bai, L., Wong, S.C., Xu, P., Chow, A.H.F., Lam, W.H.K. (2021). Calibration of stochastic link-based fundamental diagram with explicit consideration of speed heterogeneity. *Transportation Research Part B: Methodological*, 150, 524–539.

Bramich, D.M., Menéndez, M., Ambühl, L. (2022). Fitting empirical fundamental diagrams of road traffic: a comprehensive review and comparison of models using an extensive data set. *IEEE Transactions on Intelligent Transportation Systems*, 23(9), 14104-14127.

Cheng, Q., Liu, Z., Lin, Y., Zhou, X. (2021). An s-shaped three-parameter (S3) traffic stream model with consistent car following relationship. *Transportation Research Part B: Methodological*, 153, 246–271.

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

Chui, C., Chen, G. (1991). Kalman filtering with real-time applications. *Springer-Verlag.*

Dabiri, A., Kulcsár, B. (2022). Incident indicators for freeway traffic flow models. *Communication in Transportation Research,* 2, 100060.

Del Castillo, J., Benitez, F. (1995). On the functional form of the speed-density relationship. I: general theory, II: empirical investigation. *Transportation Research Part B: Methodological*, 29(5), 373–406.

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

Drew, R. (1964). Theoretical approaches to the study and control of freeway congestion. T*exas Transportation Institute, Texas A&M University*. Research Report 24-11964.

Faghri A, Chakroborty P. (1994). Development and evaluation of a statistically reliable traffic counting program. *Transportation Planning and Technology*, 18(3): 223-237.

Franklin R. E. (1961) The Structure of a Traffic Shock Wave. *Civil Engineering Pulb. Wks. Rev*., 56, 1186-l 188.

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

Gazis, C., Herman, R., Rothery, W. (1961). Nonlinear follow-the-leader models of traffic flow. *Operation. Research*. 9, 545–567.

Han, Y., Wang, M., Leclercq, L. (2023). Leveraging reinforcement learning for dynamic traffic control: A survey and challenges for field implementation. *Communication in Transportation Research,* 3, 100104.

Ikonomakis, A., Nielsen, D., Holst, K., Dietz, J., Galeazzi, R. (2022). Validation and correction of auto-logged position measurements. *Communication in Transportation Research*, 2, 100051.

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

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

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

Kodupuganti, S.R., Pulugurtha, S.S. (2023). Are facilities to support alternative modes effective in reducing congestion?: modeling the effect of heterogeneous traffic conditions on vehicle delay at intersections. *Multimodal Transportation*, 2(1), 100050.

Lam, W.H.K., Hung, W.T., Lo, H.K, Lo, H.P., Tong, C.O., Wong, S.C., Yang, H. (2003). Advancement of the annual traffic census in Hong Kong. *Transport*, 156(2), 103–115.

Liu, B., Wang, J., Xie, J., Chen, J., Duan, G., Ye, H., Peng, B. (2022). Microscopic trajectory data-driven probability distribution model for weaving area of channel change. *Journal of Automotive Safety and Energy*, 13 (2): 333-340.

Mohammadian, S., Zheng, Z., Haque, M., Bhaskar, A. (2021). Performance of continuum models for realworld traffic flows: Comprehensive benchmarking. *Transportation Research Part C: Emerging Technologies*,147, 132–167.

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

Nigam, A., Srivastava, S. (2023). Hybrid deep learning models for traffic stream variables prediction during rainfall. *Multimodal Transportation*, 2(1), 100052.

Pipes, A. (1967). Car following models and the fundamental diagram of road traffic. *Transportation Research*, 1, 21–29.

Qian, W.L., Siqueira, A.F., Machado, R.F., Lin, K., Grant, T.W. (2017). Dynamical capacity drop in a nonlinear stochastic traffic model. *Transportation Research Part B: Methodological*, 105, 328-339.

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

Qu, X., Zhang, J., Wang, S. (2017). On the stochastic fundamental diagram for freeway traffic: model development, analytical properties, validation, and extensive applications. *Transportation Research Part B: Methodological*, 104, 256–271.

Saffari, E., Yildirimoglu, M., Hickman, M. (2020). A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. *Transportation Research Part C: Emerging Technologies*, 119, 102743.

Saffari E, Yildirimoglu M, Hickman M. (2022). Data fusion for estimating Macroscopic Fundamental Diagram in large-scale urban networks. *Transportation Research Part C: Emerging Technologies*, 137: 103555.

Sharma, S.C., Gulati, B.M., Rizak, S.N. (1996). Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, 122(6), 430-439.

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

Transport Department. (2017). *Transport Planning and Design Manual, Volume 8, Chapter 2: The Annual Traffic Census*, Transport Department, Government of the Hong Kong Special Administrative Region.

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

Xu, M., Di, Y., Ding, H., Zhu, Z., Chen, X., Yang, H. (2023). AGNP: Network-wide short-term probabilistic traffic speed prediction and imputation. *Communications in Transportation Research*, 3, 100099.

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

Wang, S., Chen, X., Qu, X. (2021). Model on empirically calibrating stochastic traffic flow fundamental diagram. *Communication in Transportation Research*, 1, 100015.

Wang, S, Yan, R. (2022). "Predict, then optimize" with quantile regression: A global method from predictive to prescriptive analytics and applications to multimodal transportation. *Multimodal Transportation*, 1(4), 100035.

Wang, Y., Yu, X., Guo, J., Papamichail, I., Papageorgiou, M., Zhang, L., Hu, S., Li., Y., Sun, J., (2022). Macroscopic traffic flow modelling of large-scale freeway networks with field data verification: State-of-the-art review, benchmarking framework, and case studies using METANET, *Transportation Research Part C: Emerging Technologies*, 145, 103904.

West, M., Harrison, B.J. (1997). Bayesian forecasting and dynamic models, second edition, *Springer-Verlag*.

Wong, W., Wong, S.C. (2015). Systematic bias in transport model calibration arising from the variability of linear data projection. *Transportation Research Part B: Methodological*, 75, 1–18.

Wong, W., Wong, S.C. (2016). Biased standard error estimations in transport model calibration due to heteroscedasticity arising from the variability of linear data projection. *Transportation Research Part B: Methodological*, 88, 72–92.

Wong, W., Wong, S.C. (2016). Evaluation of the impact of traffic incidents using GPS data. *Proceedings of the Institution of Civil Engineers – Transport* 169 (3), 148-162.

Wong, W., Wong, S.C. (2019). Unbiased estimation methods of nonlinear transport models based on linearly projected data. *Transportation Science*, 53(3), 665–682.

Wong, W., Wong, S.C., Liu H.X. (2019). Bootstrap standard error estimations of nonlinear transport models based on linearly projected data. *Transportmetrica A: Transport Science*, 15(2), 602–630.

Wong, W., Shen, S., Zhao, Y. and Liu, H.X. (2019) On the estimation of connected vehicle penetration rate based on single-source connected vehicle data. *Transportation Research Part B: Methodological*, 126, 169-191.

Wong, W., Wong, S.C., Liu, H.X. (2021). Network topological effects on the macroscopic fundamental diagram. *Transportmetrica B: Transport Dynamics*, 9(1), 376-398.

Yin, R., Zheng, N., Liu, Z., (2022). Estimating fundamental diagram for multi-modal signalized urban links with limited probe data. *Physica A: Statistical Mechanics and its Applications*, 606: 128091.

Zhou, S.X., Cheng, Q., Wu, X., Li P., Belezamo B., Lu, J., Abbasi, M. (2022). A meso-to-macro cross-resolution performance approach for connecting polynomial arrival queue model to volume-delay function with inflow demand-to-capacity ratio. *Multimodal Transportation*, 1(2), 100017.

Zhu, J., Tasic, I., Qu, X. (2022). Flow-level coordination of connected and autonomous vehicles in multilane freeway ramp merging areas. *Multimodal Transportation*, 1(1), 100005.

Zockaie, A., Saberi, M., and Saedi, R. (2018). A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transportation Research Part C: Emerging Technologies*, 86, 245–262.

Appendix A.

Table A1. Determination of $|D^c|$ for the 4PL model with a 30-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	86.84	0.0316	7.8698
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	27.51	0.01290	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	15.59	0.02495	
	$\theta/\hat{\theta}$	10.60	0.0273	8.89	0.01526	
12	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.20	0.0320	2.1927
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	28.76	0.0138	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	14.22	0.0257	
	$\theta/\hat{\theta}$	10.60	0.0273	10.29	0.0169	
11	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.41	0.0323	1.4937
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.00	0.0140	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.95	0.0259	
	$\theta/\hat{\theta}$	10.60	0.0273	10.52	0.0172	
10	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.55	0.0330	**1.4149**
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.21	0.0143	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.74	0.0266	
	$\theta/\hat{\theta}$	10.60	0.0273	10.70	0.0177	
9	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.72	0.0339	1.7166
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.45	0.0146	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.48	0.0274	
	$\theta/\hat{\theta}$	10.60	0.0273	10.92	0.0184	
8	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.81	0.0350	1.8330
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.67	0.0151	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.28	0.0284	
	$\theta/\hat{\theta}$	10.60	0.0273	11.07	0.0192	

Table A2. Determination of $|D^c|$ for the 4PL model with a 60-min LR interval

Candidate $\lvert D^c\rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	86.79	0.0311	10.5739
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	26.47	0.0124	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	16.43	0.0277	
	$\theta/\hat{\theta}$	10.60	0.0273	8.43	0.0138	
11	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	87.86	0.0293	2.7057
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	28.68	0.0131	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	14.26	0.0264	
	$\theta/\hat{\theta}$	10.60	0.0273	10.07	0.0150	
10	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.19	0.0299	1.4650
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.04	0.0135	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.89	0.0270	
	$\theta/\hat{\theta}$	10.60	0.0273	10.43	0.0155	
9	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.44	0.0308	**1.2235**
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.38	0.0140	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.54	0.0279	
	$\theta/\hat{\theta}$	10.60	0.0273	10.73	0.0163	
8	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.74	0.0321	1.7712
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.73	0.0146	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.14	0.0289	
	$\theta/\hat{\theta}$	10.60	0.0273	11.07	0.0173	
7	$u_f/\hat{u}_f$(km/h)	88.00	0.0476	88.72	0.0352	2.0092
	$k_0/\hat{k}_0$ (veh/km/lane)	29.72	0.0215	29.84	0.0155	
	$u_b/\hat{u}_b$ (km/h)	14.00	0.0365	13.07	0.0309	
	$\theta/\hat{\theta}$	10.60	0.0273	11.11	0.0190	

Table A3. Determination of $|D^c|$ for the Underwood-class model with a 30-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	84.59	0.0185	2.5558
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	33.59	0.0168	
	$m/\hat{m}$	1.649	0.0015	1.615	0.0013	
17	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	84.03	0.0138	0.7489
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.63	0.0142	
	$m/\hat{m}$	1.649	0.0015	1.644	0.0011	
16	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.96	0.0137	0.519
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.75	0.0142	
	$m/\hat{m}$	1.649	0.0015	1.648	0.0011	
15	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.89	0.0135	**0.5084**
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.86	0.0143	
	$m/\hat{m}$	1.649	0.0015	1.653	0.0011	
14	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.83	0.0134	0.6170
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.98	0.0144	
	$m/\hat{m}$	1.649	0.0015	1.657	0.0010	
13	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.77	0.0132	0.7188
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	35.08	0.0144	
	$m/\hat{m}$	1.649	0.0015	1.661	0.0011	

Table A4. Determination of $|D^c|$ for the Underwood-class model with a 60-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	84.88	0.0200	3.4246
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	31.93	0.0176	
	$m/\hat{m}$	1.649	0.0015	1.624	0.0015	
17	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	84.13	0.0132	1.2420
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	33.90	0.0155	
	$m/\hat{m}$	1.649	0.0015	1.643	0.0011	
16	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	84.04	0.0130	0.9879
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.11	0.0156	
	$m/\hat{m}$	1.649	0.0015	1.646	0.0011	
15	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.97	0.0129	**0.7625**
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.30	0.0159	
	$m/\hat{m}$	1.649	0.0015	1.648	0.0011	
14	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.87	0.0126	0.7981
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.51	0.0162	
	$m/\hat{m}$	1.649	0.0015	1.655	0.0011	
13	$u_f/\hat{u}_f$(km/h)	83.91	0.0197	83.77	0.0124	0.9282
	$k_0/\hat{k}_0$ (veh/km/lane)	35.54	0.0180	34.70	0.0166	
	$m/\hat{m}$	1.649	0.0015	1.661	0.0011	

Table A5. Determination of $|D^c|$ for the NF model with a 30-min LR interval

Candidate $|D^c|$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	82.11	0.0132	1.3624
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	102.27	0.1092	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.23	0.0489	
14	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.80	0.0107	0.4443
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	105.21	0.1001	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.48	0.0436	
13	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.79	0.0106	0.3945
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	105.42	0.1002	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.48	0.0436	
12	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.79	0.0105	**0.3619**
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	105.71	0.1002	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.44	0.0433	
11	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.78	0.0104	0.4263
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	106.11	0.1010	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.38	0.0433	
10	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.79	0.0103	0.5492
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	106.56	0.1025	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.29	0.0434	

Table A6. Determination of $|D^c|$ for the NF model with a 60-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	82.32	0.0140	2.0599
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	98.96	0.1190	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.20	0.0538	
14	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.859	0.0104	0.7610
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	104.18	0.1082	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.270	0.0458	
13	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.84	0.0102	0.6922
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	104.75	0.1098	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.160	0.0458	
12	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.83	0.0101	**0.6255**
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	105.19	0.1108	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	30.10	0.0458	
11	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.82	0.0099	0.6289
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	106.01	0.1129	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	29.92	0.0457	
10	$u_f/\hat{u}_f$(km/h)	81.66	0.0151	81.81	0.0098	0.8370
	$k_j/\hat{k}_j$ (veh/km/lane)	105.7	0.1226	106.77	0.1163	
	$C_j/\hat{C}_j$ (km/h)	31.01	0.0572	29.76	0.0462	

Table A7. Determination of $|D^c|$ for the Pipe's model with a 30-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	91.14	0.0178	1.2040
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	116.77	0.0629	
16	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.77	0.0156	0.2432
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	121.59	0.0619	
15	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.74	0.0155	0.2062
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	121.89	0.0621	
14	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.72	0.0154	**0.1536**
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	122.27	0.0621	
13	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.69	0.0152	0.2626
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	122.74	0.0625	
12	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.65	0.0151	0.3807
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	123.23	0.0629	

Table A8. Determination of $|D^c|$ for the Pipe's model with a 60-min LR interval

Candidate $\lvert D^c \rvert$ (km/h)	Parameter	HR-data-based model		Average-data-based model		Average absolute bias (km/h)
		Mean	SD	Mean	SD	
∞ (Complete dataset)	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	91.36	0.0192	2.1952
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	112.38	0.0642	
16	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.62	0.0157	0.4679
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	121.01	0.0653	
15	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.56	0.0156	0.4090
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	121.54	0.0658	
14	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.51	0.0154	**0.3189**
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	122.23	0.0662	
13	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.46	0.0153	0.4517
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	122.83	0.0670	
12	$u_f/\hat{u}_f$(km/h)	90.92	0.0184	90.39	0.0152	0.6448
	$k_j/\hat{k}_j$ (veh/km/lane)	122.2	0.0695	123.57	0.0679	

Appendix B.

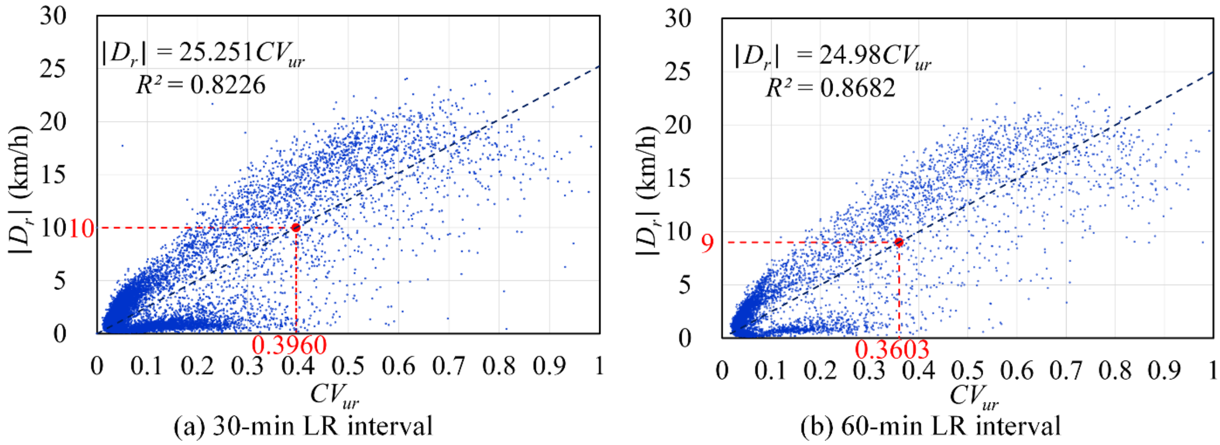

Figure B1. Established linear relationships between $|D_r|$ and CV_{ur} for the 4PL model with (a) 30-min LR interval and (b) 60-min LR interval.

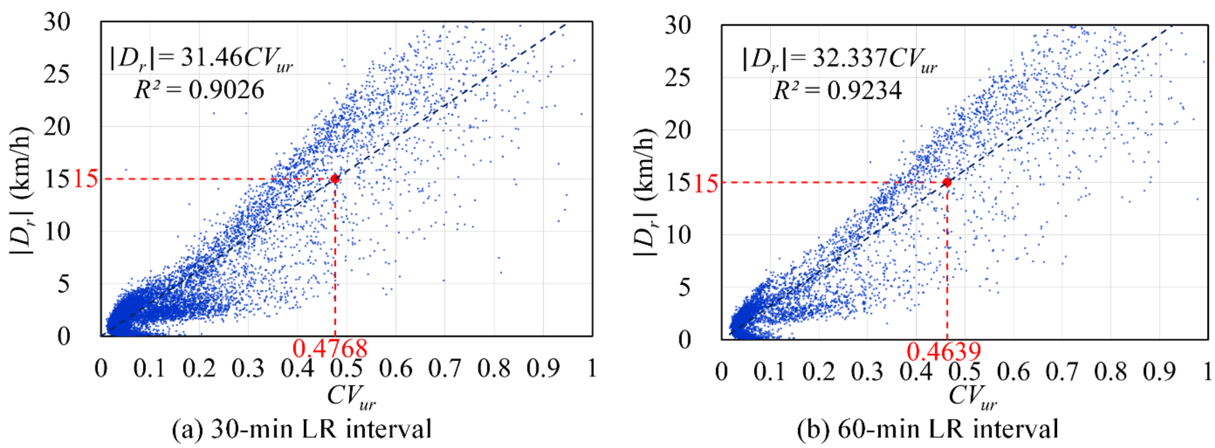

Figure B2. Established linear relationships between $|D_r|$ and CV_{ur} for the Underwood-class model with (a) 30-min LR interval and (b) 60-min LR interval.

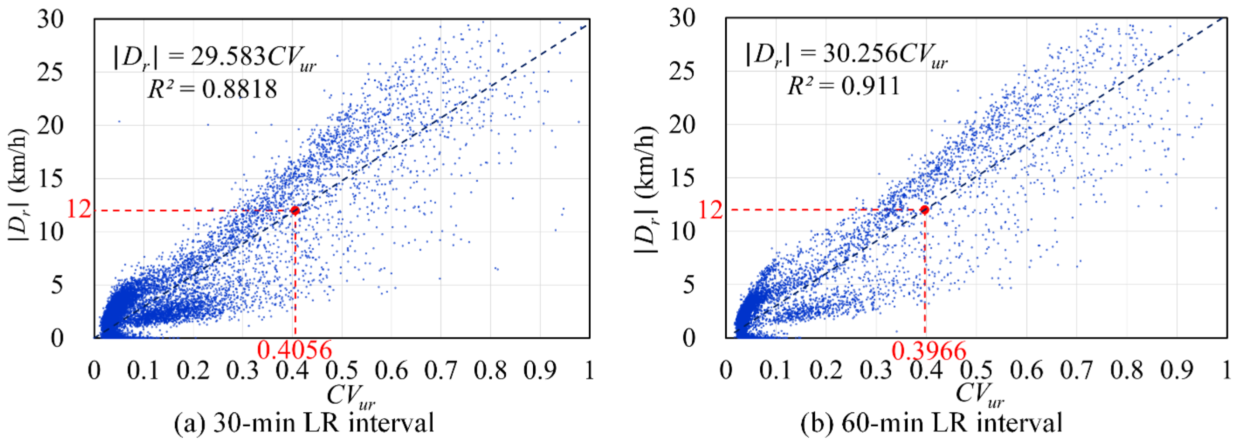

Figure B3. Established linear relationships between $|D_r|$ and CV_{ur} for the NF model with (a) 30-min LR interval and (b) 60-min LR interval

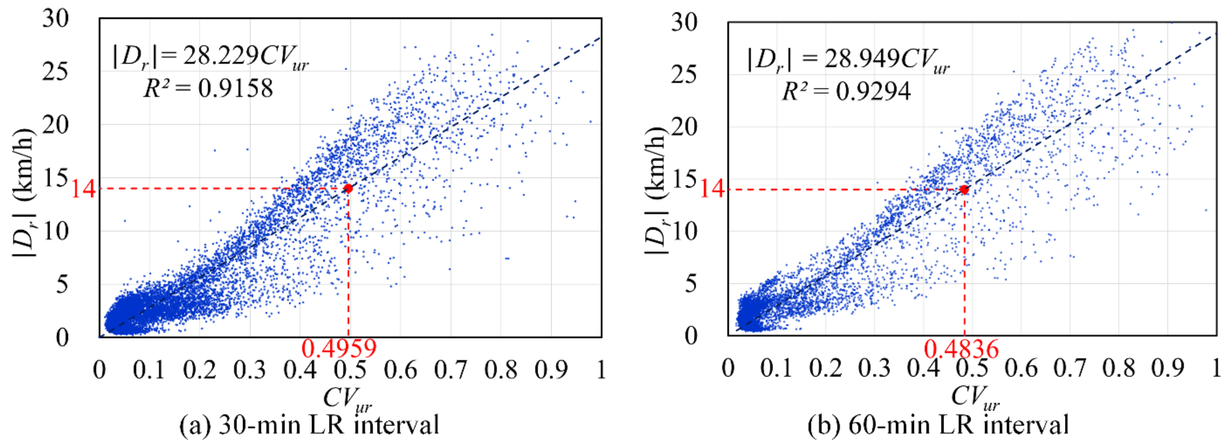

Figure B4. Established linear relationships between $|D_r|$ and CV_{ur} for the Pipe's model with (a) 30-min LR interval and (b) 60-min LR interval

Appendix C.

Table C1. Model estimations of the 4PL model for the 30-min and 60-min LR intervals at Sites 2, 3 and 4

Site	LR interval (min)	Parameter	HR-data-based model		Average-data-based model based on complete dataset		Average-data-based model based on practical optimal dataset	
			Mean	SD	Mean	Average absolute bias (km/h)	Mean	Average absolute bias (km/h) [% change]
2	30	$u_f/\hat{u}_f$(km/h)	75.40	0.089	73.39	8.8416	78.48	**5.4969 [-37.8%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	41.05	0.068	37.63		39.39	
		$u_b/\hat{u}_b$ (km/h)	8.00	0.122	11.92		8.13	
		$\theta/\hat{\theta}$	18.32	0.088	15.02		20.02	
	60	$u_f/\hat{u}_f$(km/h)	75.40	0.089	71.79	15.2712	74.78	**6.1918 [-59.5%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	41.05	0.068	35.34		37.63	
		$u_b/\hat{u}_b$ (km/h)	8.00	0.122	15.15		11.30	
		$\theta/\hat{\theta}$	18.32	0.088	12.65		16.44	
3	30	$u_f/\hat{u}_f$(km/h)	67.30	0.027	68.21	7.8752	67.84	**3.9545 [-49.8%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	40.61	0.036	38.38		39.73	
		$u_b/\hat{u}_b$ (km/h)	7.04	0.050	7.54		7.04	
		$\theta/\hat{\theta}$	6.90	0.0260	7.16		7.01	
	60	$u_f/\hat{u}_f$(km/h)	67.30	0.027	68.46	9.9986	68.09	**3.7317 [-62.7%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	40.61	0.0360	36.24		40.11	
		$u_b/\hat{u}_b$ (km/h)	7.04	0.050	8.45		7.16	
		$\theta/\hat{\theta}$	6.90	0.026	6.70		6.80	
4	30	$u_f/\hat{u}_f$(km/h)	73.98	0.449	73.31	5.3568	73.52	**2.1457 [-59.9%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	26.33	0.229	25.91		26.45	
		$u_b/\hat{u}_b$ (km/h)	7.22	0.415	7.75		7.49	
		$\theta/\hat{\theta}$	12.55	0.291	11.98		12.36	
	60	$u_f/\hat{u}_f$(km/h)	73.98	0.449	72.68	9.2471	73.55	**3.5679 [-61.4%]**
		$k_0/\hat{k}_0$ (veh/km/lane)	26.33	0.229	24.71		25.43	
		$u_b/\hat{u}_b$ (km/h)	7.22	0.415	8.91		7.61	
		$\theta/\hat{\theta}$	12.55	0.291	11.03		11.74	

Table C2. Model estimations of the Underwood-class model for the 30-min and 60-min LR intervals at Sites 2, 3 and 4

Site	LR interval (min)	Parameter	HR-data-based model		Average-data-based model based on complete dataset		Average-data-based model based on practical optimal dataset	
			Mean	SD	Mean	Average absolute bias (km/h)	Mean	Average absolute bias (km/h) [% change]
2	30	$u_f/\hat{u}_f$(km/h)	68.24	0.0194	68.30		68.23	
		$k_0/\hat{k}_0$ (veh/km/lane)	45.95	0.0478	43.84	1.0304	45.14	**0.5319 [-48.4%]**
		$m/\hat{m}$	1.69	0.0025	1.70		1.68	
	60	$u_f/\hat{u}_f$(km/h)	68.24	0.0194	68.36		68.25	
		$k_0/\hat{k}_0$ (veh/km/lane)	45.95	0.0478	42.35	1.6911	45.82	**0.9507 [-43.8%]**
		$m/\hat{m}$	1.69	0.0025	1.71		1.65	
3	30	$u_f/\hat{u}_f$(km/h)	67.13	0.0232	68.14		67.17	
		$k_0/\hat{k}_0$ (veh/km/lane)	33.94	0.0285	33.52	5.2672	34.54	**2.1405 [-59.4%]**
		$m/\hat{m}$	3.55	0.0099	3.01		3.32	
	60	$u_f/\hat{u}_f$(km/h)	67.13	0.0232	68.54		68.45	
		$k_0/\hat{k}_0$ (veh/km/lane)	33.94	0.0285	31.87	6.1056	33.99	**2.2462 [-63.2%]**
		$m/\hat{m}$	3.55	0.0099	2.99		3.20	
4	30	$u_f/\hat{u}_f$(km/h)	60.76	0.0700	61.24		60.16	
		$k_0/\hat{k}_0$ (veh/km/lane)	35.29	0.3496	32.23	2.0457	35.87	**0.6283 [-69.3%]**
		$m/\hat{m}$	1.53	0.0146	1.54		1.54	
	60	$u_f/\hat{u}_f$(km/h)	60.76	0.0700	60.60		60.64	
		$k_0/\hat{k}_0$ (veh/km/lane)	35.29	0.3496	32.29	2.4540	35.51	**0.8186 [-66.6%]**
		$m/\hat{m}$	1.53	0.0146	1.59		1.54	

Table C3. Model estimations of the NF model for the 30-min and 60-min LR intervals at Sites 2, 3 and 4

Site	LR interval (min)	Parameter	HR-data-based model		Average-data-based model based on complete dataset		Average-data-based model based on practical optimal dataset	
			Mean	SD	Mean	Average absolute bias (km/h)	Mean	Average absolute bias (km/h) [% change]
2	30	$u_f/\hat{u}_f$(km/h)	66.60	0.0159	66.67	1.9870	66.61	**0.6438 [-67.6%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	135.98	0.3707	129.10		134.49	
		$C_j/\hat{C}_j$ (km/h)	26.78	0.1053	27.34		26.44	
	60	$u_f/\hat{u}_f$(km/h)	66.60	0.0159	66.73	2.5870	66.64	**0.8624 [-66.7%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	135.98	0.3707	125.22		135.70	
		$C_j/\hat{C}_j$ (km/h)	26.78	0.1053	27.59		25.59	
3	30	$u_f/\hat{u}_f$(km/h)	68.85	0.0274	69.09	1.4717	68.29	**0.4076 [-72.3%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	141.07	0.3405	139.61		140.44	
		$C_j/\hat{C}_j$ (km/h)	27.80	0.0977	29.37		26.70	
	60	$u_f/\hat{u}_f$(km/h)	68.85	0.0274	69.30	1.9718	68.86	**0.5324 [-73.0%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	141.07	0.3405	136.90		140.05	
		$C_j/\hat{C}_j$ (km/h)	27.80	0.0977	29.37		28.32	
4	30	$u_f/\hat{u}_f$(km/h)	64.97	0.0577	64.89	1.2569	64.89	**0.6578 [-47.7%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	144.44	1.5347	146.86		144.73	
		$C_j/\hat{C}_j$ (km/h)	13.03	0.4848	12.82		12.80	
	60	$u_f/\hat{u}_f$(km/h)	64.97	0.0577	64.96	1.8324	64.95	**0.9230 [-49.6%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	144.44	1.5347	143.14		144.43	
		$C_j/\hat{C}_j$ (km/h)	13.03	0.4848	12.51		12.43	

Table C4. Model estimations of the Pipe's model for both the 30-min and 60-min LR intervals at Sites 2, 3 and 4

Site	LR interval (min)	Parameter	HR-data-based model		Average-data-based model based on complete dataset		Average-data-based model based on practical optimal dataset	
			Mean	SD	Mean	Average absolute bias (km/h)	Mean	Average absolute bias (km/h) [% change]
2	30	$u_f/\hat{u}_f$(km/h)	71.20	0.0157	71.42	1.0782	71.10	**0.3252 [-69.8%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	104.90	0.0737	99.82		106.29	
	60	$u_f/\hat{u}_f$(km/h)	71.20	0.0157	71.58	1.8222	70.78	**1.1538 [-36.7%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	104.90	0.0737	96.33		109.55	
3	30	$u_f/\hat{u}_f$(km/h)	69.38	0.0288	72.46	2.9440	70.77	**1.1214 [-61.9%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	165.72	0.2375	161.58		165.68	
	60	$u_f/\hat{u}_f$(km/h)	69.38	0.0288	74.17	4.3132	71.65	**1.9953 [-53.7%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	165.72	0.2375	161.55		164.22	
4	30	$u_f/\hat{u}_f$(km/h)	66.60	0.0563	66.63	1.5762	66.67	**0.5213 [-66.9%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	92.94	0.4555	89.87		91.07	
	60	$u_f/\hat{u}_f$(km/h)	66.60	0.0563	66.81	2.6984	66.87	**1.2146 [-55.0%]**
		$k_j/\hat{k}_j$ (veh/km/lane)	92.94	0.4555	84.07		88.92	