Estimation of vehicular journey time variability by Bayesian data fusion with general mixture model

Xinyue Wu, Andy H.F. Chow *Member, IEEE*, Li Zhuang, Wei Ma *Member, IEEE*, William H.K. Lam, and S.C. Wong

***Abstract*—This paper presents a Bayesian data fusion framework for estimating journey time variability that uses a mixture distribution model to classify feeding data into different traffic states. Different from most studies, the proposed framework offers a generalized statistical foundation for making full use of multiple traffic data sources to estimate the vehicular journey time variability. Feeding data collected from multiple data sources are classified based on the associated traffic conditions, and the corresponding estimation biases of the individual data sources are determined by arbitrary distributions. The proposed framework is implemented and tested on a Hong Kong corridor with actual data collected from the field. Different statistical distributions of prior and likelihood knowledge are applied and compared. The findings of the case study show significant improvement in the journey time estimations of the proposed method compared with the individual measurements. The results also highlight the benefit of incorporating a traffic state classifier and prior knowledge in the fusion framework. This study contributes to the development of reliability-based intelligent transportation systems based on advanced traffic data analytics.**

***Index Terms*—Journey time variability, Bayesian data fusion, general mixture model, traffic state classification, automatic vehicle identification.**

I. INTRODUCTION

ESTIMATION of vehicular journey times and their associated variability in congested road networks is crucial for the development and operation of reliable intelligent transportation systems [1], [2]. There have been a large number of studies estimating and modeling journey times with different data sources [3]. Tam and Lam [4] propose an estimation model based on Automatic Vehicle Identification (AVI) system. Celikoglu [5] presents a fundamental diagram based approach utilizing data collected from microwave sensors. Jenelius and Koutsopoulos [6] develop a statistical model using low-frequency GPS data. Qiu *et al.* [7] and Han *et al.* [8] use deep learning methods to estimate journey time with taxi data. Reviews on point travel time estimation using different types of data can be referred to [3]. However, the aforementioned methods only operate with the single data source. With the increasing availability of traffic data from multiple sources, recent transportation studies have begun to explore the use of data fusion techniques to estimate journey time [9]–[11]. Existing data fusion techniques can be generally divided into three groups: statistical, probabilistic-based and artificial intelligence-based techniques [12], [13]. Early data fusion models apply simple statistical methods, such as weighted average and convex combination, to fuse the different measurements [14]–[16]. However, these methods lack flexibility and fail to provide accurate and reliable combinations. To overcome the drawbacks, Kalman Filter (KF) is arguably one of the most widely applied statistical data fusion techniques in the traffic domain [17], [18]. Chu *et al.* [19] apply a standard linear KF to combine traffic data collected from loop detectors and probe vehicles to estimate freeway journey times. Han [20] compares the journey time estimation performance of the extended KF and unscented KF for multi-sensor data sources. Trinh *et al.* [21] propose the incremental KF to estimate the traffic state using loop detectors and GPS data. Nevertheless, the estimation performance of the KF is closely related to the explicitness and complexity of the adopted traffic model, which limits the estimation precision to some extent. As it is desirable to combine different data sources without complicated traffic models, studies on traffic data fusion have been further explored via probabilistic-based and artificial intelligence-based algorithms [22], [23]. El Faouzi *et al.* [24] apply Dempster-Shafer evidence theory to estimate the travel time by fusing conventional traffic detector and probe vehicle data. Guo and Yang [25] propose a support degree algorithm based on the credibility and similarity among license plate recognition data, loop detector data and probe vehicle data to estimate the travel time. Zhu *et al.* [26] estimate link travel time using artificial neural networks with three different sources. Sun *et al.* [27] propose a relation learning framework to estimate travel time with heterogeneous data.

It can be found that most journey time estimation algorithms presented in the literature have largely focused on deterministic estimation with no consideration of the associated variability. Therefore, recent studies have investigated the journey time variability in order to better evaluate the performance of transport networks using a single data source. A large number of empirical studies estimate the variability

We would like to thank all anonymous referees for their constructive comments. This study was supported by a Research Impact Fund (RIF) sponsored by the Hong Kong Research Grants Council (Project No.: R5029-18). The study was supported by the Strategic Public Policy Research Funding Scheme (Project No.: S2019.A8.013.19S). The authors would also like to thank Hong Kong Transport Department and Autotoll Limited for providing the traffic data. *(Corresponding author: Andy H. F. Chow.)*

Xinyue Wu and Andy H.F. Chow are with the Department of Systems Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China (e-mail:andychow@cityu.edu.hk).

Li Zhuang is with the School of Cyber Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China.

Wei Ma and William H.K. Lam are with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

S.C. Wong is with the Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China.

by fitting the travel time distribution into a specific assumed distribution [28]–[30]. Uno *et al*. [31] analyze the journey time reliability with the assumption of lognormal distribution using bus probe data. Sumalee *et al*. [32] propose an algorithm to estimate the link travel time distributions based on the cell transmission model, and the probability mass function of the journey time is fitted to a skew-normal distribution using loop detector data. Rahmani *et al*. [33] present a non-parametric model to reduce the bias from probe vehicle data in order to estimate the travel time distribution. Ma *et al*. [34] proposed a generalized Markov chain approach to estimate trip travel time distribution with automatic vehicle location data. Chen *et al*. [35] develop a copula-based model for arterial travel time distribution estimation using AVI and trajectory data separately. Nevertheless, most journey time variability estimation algorithms have not effectively leveraged the benefit of multi-source data. In a recent study, to improve the estimation performance with data fusion techniques, Shi *et al*. [36] propose an improved Dempster–Shafer evidence theory to estimate the travel time distribution using multi-source data with consideration of spatial correlations. Mil and Piantanakulchai [37] propose a Bayesian based model using maximum likelihood to estimate travel time. Based on this study, Gemma *et al*. [38] incorporate the Gaussian mixture model to represent the ground truth in the data fusion process. Saffari *et al*. [39] also apply Bayesian model to estimate macroscopic fundamental diagrams with multiple data sources. However, all these fusion models rely on strong Gaussian assumptions for the consideration of inferential tractability but overlook the skewness associated with journey time.

Despite these research efforts, we have not seen a unified framework with general formulations for estimating journey time variability with the use of different data sources under different traffic conditions. As far as our knowledge, most studies on Bayesian data fusion use non-informative prior probability due to its simplicity, and there is no existing work that investigates the impact of prior information on journey time variability. In practical scenarios, traffic data collected from different detectors might suffer from the problem of non-Gaussian measurement errors caused by traffic control, sensor failures, or impulsive data processing [40]. The common Gaussian assumptions might lead to biased or unreliable fusion results. Moreover, existing traffic state classification models based on Gaussian mixtures, *e.g.* Wang *et al*. [41] and Liu *et al*. [42], ignore the heavy-tailed characteristics of journey time distribution under different traffic states.

Therefore, this paper presents a generalized and transferable statistical model for estimating journey time variability with heterogeneous data sources taking into account of the prevailing traffic conditions via a mixture distribution model based state classification algorithm. The prior knowledge and the corresponding systematic errors associated with different sources are incorporated into the underlying posterior distribution models following the classification and identification of traffic states in the data fusion framework. Different from [37]–[39], the data fusion model proposed herein tests the benefit of introducing different informative prior knowledge and allows arbitrary distributions to estimate measurement errors. This work contributes to the state-of-the-art in three aspects: first, our work offers a generalized and transferable statistical model based on the Bayesian theory for making full use of multiple data sources, going beyond the conventional Gaussian error assumptions. Then, the analysis on the influence of prior knowledge advances our understanding on the interplay between the different types of shared priors and observation error likelihoods. Finally, the application of the general mixture model makes the data classification more flexible in capturing the effect of skewness and other distributional characteristics in journey time with associated traffic states.

The rest of this paper is organized as follows: Section 2 presents the methodology including the problem settings and use of notations. Section 3 presents a case study of the selected Hong Kong corridor with journey time data inferred from different sensors. Section 4 provides some concluding remarks.

II. Methodology

The proposed vehicular journey time estimation framework consists of two components: a data classification procedure based on a mixture distribution model and a Bayesian data fusion framework. The input data are first fed through the mixture distribution model and classified according to the traffic states that the data are associated with. The classified traffic data are then processed and integrated by the Bayesian fusion algorithm for deriving a corresponding journey time estimation. The entire framework consists of various parameters which need to be determined through a training process. The accuracy of the trained data classification and fusion framework is then evaluated through a testing process with use of a set of performance metrics. Fig. 1 summarises the proposed journey time distribution estimation framework. Details of each component, including the training and testing procedures, are presented in the following sections.

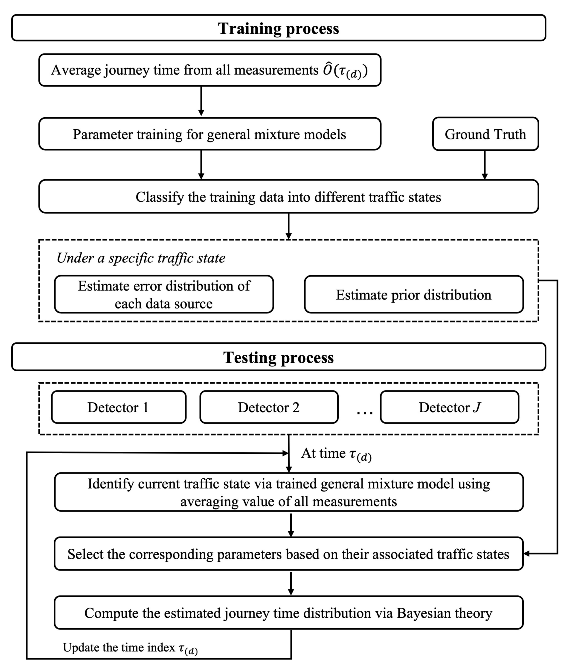


Fig. 1: Flowchart of the proposed methodology

A. Traffic state classification

For each day d considered in the analysis, we define $\tau_{(d)} = \{1, 2, ..., \mathcal{T}_{(d)}\}$ to be a series of time intervals spanning the day d. We also define $S(\tau_{(d)}) = s$ to be a discrete variable representing the prevailing traffic condition in state s, where $s \in \mathcal{S}$, during a specific time interval $\tau_{(d)}$ on day d. The notation $\mathcal{S}$ denotes the set of all traffic states considered in the analysis. A simple example can be a binary case $s = \{0, 1\}$, where $S(\tau_{(d)}) = 0$ means the traffic is in 'free-flow' state during time $\tau_{(d)}$, and $S(\tau_{(d)}) = 1$ means the traffic is in 'congested' state at that time $\tau_{(d)}$.

The traffic state $S(\tau_{(d)})$ is considered to be stochastic with the probability $P(S(\tau_{(d)} = s)$ being in a specific state s given by π_s for all $s \in \mathcal{S}$. We consider all states s in $\mathcal{S}$ are mutually exclusive and comprehensively exhaustive, and this gives $\sum_{\forall s \in \mathcal{S}} \pi_s = 1$.

We further define $\Psi_{\boldsymbol{\theta}_s}(\hat{O})$ to be the probability distribution model of observed journey time $\hat{O}$, in which the distribution model parameters $\boldsymbol{\theta}_s$ are dependent on the prevailing traffic state $S(\tau_{(d)}) = s$ at the time of interest $\tau_{(d)}$.

The objective of the data classification algorithm is to determine from which traffic state $S(\tau_{(d)})$ the data are collected based on the average journey times $\hat{O}(\tau_{(d)})$ inferred from the measurements. Given the inferred journey times $\hat{O}(\tau_{(d)}) = \hat{o}$ from all measurements, the state classification is based on the probability $P\left((S(\tau_{(d)}) = s)|(\hat{O}(\tau_{(d)}) = \hat{o})\right)$ which is determined from the inferred journey time $\hat{O}(\tau_{(d)}) = \hat{o}$ as stated in the proposition below:

Proposition II.1. *Given the observed journey time $\hat{O}(\tau_{(d)})$ at time $\tau_{(d)}$ being $\hat{o}$, the probability of the corresponding traffic state $S(\tau_{(d)})$ being in s at the time can be determined as*

$$P\left((S(\tau_{(d)}) = s)|(\hat{O}(\tau_{(d)}) = \hat{o})\right) = \frac{\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O}(\tau_{(d)}) = \hat{o})}{\sum_{\forall s \in \mathcal{S}}[\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O}(\tau_{(d)}) = \hat{o})]}. \tag{1}$$

Proof. The probability $P\left((S(\tau_{(d)}) = s)|(\hat{O}(\tau_{(d)}) = \hat{o})\right)$ can be expressed by Bayes' theorem over all possible traffic states s as:

$$P\left((S(\tau_{(d)}) = s)|(\hat{O}(\tau_{(d)}) = \hat{o})\right) = \frac{P(S(\tau_{(d)}) = s)P\left((\hat{O}(\tau_{(d)}) = \hat{o})|(S(\tau_{(d)}) = s)\right)}{P\left(\hat{O}(\tau_{(d)}) = \hat{o}\right)}. \tag{2}$$

Moreover, the probability $P\left(\hat{O}(\tau_{(d)}) = \hat{o}\right)$ in (2) can be established by the theorem of total probability with use of the inferred journey time distribution model $\Psi_{\boldsymbol{\theta}_s}$ over possible states s as

$$P\left(\hat{O}(\tau_{(d)}) = \hat{o}\right) = \sum_{\forall s \in \mathcal{S}}[P(S(\tau_{(d)}) = s)\Psi_{\boldsymbol{\theta}_s}(\hat{O}(\tau_{(d)}) = \hat{o})]. \tag{3}$$

Substituting (3) into (2), and the definitional "$P(S(\tau_{(d)}) = s) = \pi_s$" for all s, gives (1).

□

Given the classification probability function (1) and the inferred journey time $\hat{O}(\tau_{(d)}) = \hat{o}$ at the time of interest $\tau_{(d)}$, a corresponding traffic state $s^* \in \mathcal{S}$ is assigned such that the associated $P\left((S(\tau_{(d)}) = s^*)|(\hat{O}(\tau_{(d)}) = \hat{o})\right)$ is maximized, i.e.

$$\begin{aligned} s^* &= \arg\max_s P\left((S(\tau_{(d)}) = s)|(\hat{O}(\tau_{(d)}) = \hat{o})\right) \\ &= \arg\max_s \left\{\frac{\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O}(\tau_{(d)}) = \hat{o})}{\sum_{\forall s \in \mathcal{S}}[\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O}(\tau_{(d)}) = \hat{o})]}\right\} \end{aligned} \tag{4}$$

The classification rule (4) requires determination of the parameters $\boldsymbol{\Theta} = [\boldsymbol{\theta}_s, \boldsymbol{\pi}_s]$, for all $s \in \mathcal{S}$. That is, we would need to determine the model parameters $\boldsymbol{\theta}$ in the inferred journey time distribution model $\Psi_{\boldsymbol{\theta}_s}$, and the probabilities $P(S(\tau_{(d)}) = s) = \pi_s$ of the occurrences of each traffic state s in $\mathcal{S}$.

1) Determination of parameters Θ: To determine $\boldsymbol{\Theta}$, we first establish the following general mixture model (GMM, [43]), parameterized by $\boldsymbol{\Theta}$, for the distribution of inferred journey times:

$$\Psi_{\boldsymbol{\Theta}}(\hat{O}) = \sum_{\forall s \in \mathcal{S}} \pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O}) \tag{5}$$

with $\Psi_{\boldsymbol{\theta}_s}(\hat{O})$ being the conditional distribution function of $\hat{O}$ on traffic state s as in (1) and (4), for all $s \in \mathcal{S}$. The parameters $\boldsymbol{\Theta}$ in the mixture distribution model (5) are to be determined by using the maximum likelihood estimation (MLE) via an expectation-maximization (EM) algorithm with consideration of the latent variable [43].

Given a set of N independent journey time measurements: $\{\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N\}$ inferred at different times for determining the parameters $\boldsymbol{\Theta}$ in the mixture model, we can formulate a corresponding likelihood function with respect to these measurements over all possible states $s \in \mathcal{S}$ as:

$$\mathcal{L}(\boldsymbol{\Theta}|\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N) = \prod_{n=1}^{N} \sum_{\forall s \in \mathcal{S}} [\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O} = \hat{o}_n)]. \tag{6}$$

It is noted that the likelihood function is a function of the parameter set $\boldsymbol{\Theta} = [\boldsymbol{\theta}_s, \boldsymbol{\pi}_s]$, for all $s \in \mathcal{S}$, given a specific set of inferred observations $\hat{O} = \{\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N\}$.

The likelihood function in (6) can further be converted into the logarithmic form as:

$$\tilde{\mathcal{L}}(\boldsymbol{\Theta}|\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N) = \sum_{n=1}^{N} log\left(\sum_{\forall s \in \mathcal{S}} [\pi_s \Psi_{\boldsymbol{\theta}_s}(\hat{O} = \hat{o}_n)]\right). \tag{7}$$

The objective herein is to seek a set of parameters $\boldsymbol{\Theta}^*$, where

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} \tilde{\mathcal{L}}(\boldsymbol{\Theta}|\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N) \tag{8}$$

The solution $\mathbf{\Theta}^*$ in (8) is to be solved by the the expectation–maximization (EM) algorithm [43] as follows:

1) Initialize $\mathbf{\Theta}^{(0)}$ at iteration $\kappa = 0$.
2) **E-step**: Calculate the expected value $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(\kappa)})$ of the log-likelihood function $\tilde{\mathcal{L}}$ with the parameters set $\mathbf{\Theta}^{(\kappa)}$ at the current iteration κ:
$$Q(\mathbf{\Theta}, \mathbf{\Theta}^{(\kappa)}) = \mathbb{E}\left[\tilde{\mathcal{L}}(\mathbf{\Theta}|\hat{o}_1, \hat{o}_2, \cdots, \hat{o}_N)|\mathbf{\Theta}^{(\kappa)}\right], \quad (9)$$
in which the expectation is taken over all component distributions $\Psi_{\boldsymbol{\theta}_s}$, for all $s \in \mathcal{S}$, in the mixture model.
3) **M-step**: Update $\mathbf{\Theta}$ by maximizing the log-likelihood function (9):
$$\Theta^{(\kappa+1)} = \arg\max_{\mathbf{\Theta}} Q(\mathbf{\Theta}, \mathbf{\Theta}^{(\kappa)}) \quad (10)$$
4) Compute the log-likelihood value with the updated parameters. If the change in the log-likelihood function values is less than a predefined threshold, or a predefined maximum number of iterations has been reached, the algorithm will stop. Otherwise, return to Step 2 (the E-step).

B. Data fusion

Following the classification of vehicular journey time data, we then have a Bayesian data fusion framework which aims to estimate the posterior distribution for the actual journey times $T(\tau_{(d)})$ at times $\tau_{(d)}$ after incorporating all associated observations $O_j(\tau_{(d)})$ collected from J multiple sources, where $J \geq 1$. It is also considered that measurements taken from each source j, where $j = 1, 2, ..., J$, are associated with an error e_j whose statistical properties are assumed to be known from historical records. The statistical properties can be recognized as the traffic state-dependent systematic error and random error associated with the measurements taken from source j respectively [44].

Given the true but unknown journey time $T(\tau_{(d)})$ at time $\tau_{(d)}$ is t, where $t > 0$, it can be deduced that the corresponding measurement $O_j(\tau_{(d)})$ of the journey time taken by source j will be $O_j(\tau_{(d)}) = o_j = t + e_j$. Here, the state-dependent error e_j follows an arbitrary probability distribution. Specifically, the error e_j from sufficiently large datasets can be assumed to follow Gaussian distribution $N(\mu_j, \sigma_j)$ for all $j = 1, 2, ..., J$. Hence the distribution of measurement $O_j(\tau_{(d)})$ can be expressed by
$$P\Big(O_j(\tau_{(d)}) = o_j|T(\tau_{(d)}) = t\Big) \sim N(t + \mu_j, \sigma_j), \quad (11)$$
in which μ_j and σ_j are estimated based on the prevailing traffic states $S(\tau_{(d)})$. The parameters in the distribution are considered to be dependent on the prevailing traffic states $S(\tau_{(d)})$ that is to be identified in the classification stage presented in the previous section.

We now define $\mathbb{O}_{\mathcal{J}}(\tau_{(d)}) = [O_j(\tau_{(d)})] = [o_1, o_2, ..., o_J]$ to be the collection of measured journey times from all available sources j at time $\tau_{(d)}$. The posterior probability of the actual journey times $T(\tau_{(d)})$ inferred from these measurements $\mathbb{O}_{\mathcal{J}}(\tau_{(d)})$ is established by using Bayes' theorem as:
$$P\Big(T(\tau_{(d)}) = t|\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big) = \frac{P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)}) = t\Big) \cdot P\Big(T(\tau_{(d)}) = t\Big)}{P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big)} \quad (12)$$
in which t denotes the true (but unknown) value of $T(\tau_{(d)})$, and $P\Big(T(\tau_{(d)}) = t\Big)$ is the prior distribution of $T(\tau_{(d)})$ before taking into account the measurements of $\mathbb{O}_{\mathcal{J}}(\tau_{(d)})$.

We then define the following function of time t
$$g(t) = P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)}) = t\Big) \cdot P\Big(T(\tau_{(d)}) = t\Big) \quad (13)$$
which is recognized as the numerator of (12).

As different sensors operate on the basis of different physical principles, we further suppose measurements made from different sources are independent from each other, which gives
$$\begin{aligned} &P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)}) = t\Big) \\ &= P\left(O_1(\tau_{(d)}) = o_1, \cdots, O_J(\tau_{(d)}) = o_J|T(\tau_{(d)}) = t\right) \\ &= \prod_{j=1}^{J} P\Big(O_j(\tau_{(d)}) = o_j|T(\tau_{(d)}) = t\Big), \end{aligned} \quad (14)$$
and hence the function $g(t)$ can be further expressed as:
$$\begin{aligned} g(t) &= P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)}) = t\Big) \cdot P\Big(T(\tau_{(d)}) = t\Big) \\ &= \left[\prod_{j=1}^{J} P\left(O_j(\tau_{(d)}) = o_j|T(\tau_{(d)}) = t\right) \cdot P\big(T(\tau_{(d)}) = t\big)\right]. \end{aligned} \quad (15)$$

Estimates for the true journey time t can be derived from either an interval estimation or point estimation approach. From an interval estimation perspective, the statistics of the true journey time t are given by (16) and (17) in the following proposition in terms of the function $g(t)$:

Proposition II.2. *The mean $\hat{t}^*$ and standard deviation $\hat{\sigma}^2_{t^*}$ of the posterior estimate, $\hat{t}$, can be determined in terms of $g(t)$ respectively as:*
$$\hat{t}^* = \frac{\int_0^\infty [t \cdot g(t)]\,\mathrm{d}t}{\int_0^\infty g(t)\,\mathrm{d}t} \quad (16)$$
$$\hat{\sigma}^2_{t^*} = \frac{\int_0^\infty \left[\left(t - \hat{t}^*\right)^2 g(t)\right]\mathrm{d}t}{\int_0^\infty g(t)\,\mathrm{d}t} \quad (17)$$

Proof. The mean $\hat{t}^*$ of the posterior estimate can be expressed from the first principle as:
$$\hat{t}^* = \int_0^\infty \left[t \cdot P\Big(T(\tau_{(d)}) = t|\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big)\right]\mathrm{d}t \quad (18)$$

Using Bayes' theorem to revise the right-hand-side expression of (18), we have

$$\hat{t}^* = \frac{\int_0^\infty \Big[t \cdot P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)})=t\Big)P\Big(T(\tau_{(d)})=t\Big)\Big]\,\mathrm{d}t}{P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big)}, \tag{19}$$

in which we note using theorem of total probability that

$$\begin{aligned} &P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big) \\ &= \int_0^\infty \Big[P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)})=t\Big)\cdot P\Big(T(\tau_{(d)})=t\Big)\Big]\,\mathrm{d}t. \end{aligned} \tag{20}$$

Given the expression of $g(t)$ in (15), from (19) and (20) we can have

$$\begin{aligned} \hat{t}^* &= \frac{\int_0^\infty \Big[t \cdot P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)})=t\Big)P\Big(T(\tau_{(d)})=t\Big)\Big]\,\mathrm{d}t}{\int_0^\infty \Big[P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)})=t\Big)P\Big(T(\tau_{(d)})=t\Big)\Big]\,\mathrm{d}t} \\ &= \frac{\int_0^\infty [t\cdot g(t)]\,\mathrm{d}t}{\int_0^\infty g(t)\,\mathrm{d}t}. \end{aligned} \tag{21}$$

Likewise, we can also have

$$\begin{aligned} \hat{\sigma}^2_{t^*} &= \int_0^\infty \Big[\Big(t-\hat{t}^*\Big)^2 P\Big(T(\tau_{(d)})=t|\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big)\Big]\,\mathrm{d}t \\ &= \frac{\int_0^\infty \Big[\Big(t-\hat{t}^*\Big)^2 g(t)\Big]\,\mathrm{d}t}{\int_0^\infty g(t)\,\mathrm{d}t} \end{aligned} \tag{22}$$

following similar working procedure. □

For point estimation, the unknown true journey time t is to be estimated as the most probable value that maximizes the posterior probability in (12), where

$$\hat{t}^*_{MAP}(\tau_{(d)}) = \arg\max_t P\Big(T(\tau_{(d)})=t|\mathbb{O}_{\mathcal{J}}(\tau_{(d)})\Big), \tag{23}$$

in which $\hat{t}^*_{MAP}(\tau_{(d)})$ is regarded as the maximum a posterior (MAP) estimation.

It is noted that the denominator of the posterior probability in (12) is merely the distribution of measurements $\mathbb{O}_{\mathcal{J}}(\tau_{(d)})$ which is independent of the choice of value for $\hat{t}^*_{MAP}(\tau_{(d)})$, Expression (23) for $\hat{t}^*_{MAP}(\tau_{(d)})$ can hence be simplified to

$$\begin{aligned} &\hat{t}^*_{MAP}(\tau_{(d)}) \\ &= \arg\max_t P\Big(\mathbb{O}_{\mathcal{J}}(\tau_{(d)})|T(\tau_{(d)})=t\Big)P\Big(T(\tau_{(d)})=t\Big) \\ &= \arg\max_t [g(t)]. \end{aligned} \tag{24}$$

The journey time estimates above can be solved via various numerical methods depending on the prior distribution model adopted [43]. Closed form solutions for $\hat{t}^*$ and $\hat{\sigma}^2_{t^*}$ may be available given the specific prior and error distribution as presented below (see also [37]).

Case A: Uniform prior distribution with Gaussian error distribution

Should the prior distribution be uniform where the probability $P(T(\tau_{(d)}) = t)$ is a constant, the point estimate $\hat{t}^*_{MAP}(\tau_{(d)})$ will coincide with the mean value of the interval estimate $\hat{t}^*$ with the symmetricity of the uniform prior distribution [37], [43]. Moreover, expression (24) can be reduced to:

$$\begin{aligned} &\hat{t}^*_{MAP}(\tau_{(d)}) = \hat{t}^* \\ &= \arg\max_t \left[\prod_{j=1}^{J} P\left(O_j(\tau_{(d)})=o_j|T(\tau_{(d)})=t\right)\right] \end{aligned} \tag{25}$$

in which the constant term $P(T(\tau_{(d)}) = t)$ in $g(t)$ in (24) can be removed from the maximization.

Given the J data sources with the respective measurements o_j, and errors $e_j \sim N(\mu_j, \sigma_j)$, at prevailing time $\tau_{(d)}$, maximizing the right-hand-side of expression (25) gives the standard deviation ($\hat{\sigma}_{t^*}$) and expected value of $\hat{t}^*_{MAP}$ at time $\tau_{(d)}$ as [37], [43]:

$$\hat{\sigma}^2_{t^*} = \Big[\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\Big]^{-1}, \tag{26}$$

$$\hat{t}^* = \hat{t}^*_{MAP} = \Big[\sum_{j=1}^{J}\frac{o_j-\mu_j}{\sigma_j^2}\Big]\hat{\sigma}^2_{t^*}. \tag{27}$$

Case B: Gaussian prior distribution with Gaussian error distribution

Should the prior distribution of the journey time be Gaussian, say $N(\lambda, \delta)$, i.e.:

$$P\Big(T(\tau_{(d)})=t\Big) = \frac{1}{\sqrt{2\pi}\delta}\cdot exp\Big[-\frac{(t-\lambda)^2}{2\delta^2}\Big], \tag{28}$$

we also have the point estimate $\hat{t}^*_{MAP}(\tau_{(d)})$ coinciding with the mean value of the interval estimate $\hat{t}^*$ due to the symmetricity of the Gaussian prior distribution [37], [43]. Expression (24) also becomes:

$$\begin{aligned} &\hat{t}^*_{MAP}(\tau_{(d)}) = \hat{t}^* \\ &= \arg\max_t \Big[\prod_{j=1}^{J} P\left(O_j(\tau_{(d)})=o_j|T(\tau_{(d)})=t\right)P\Big(T(\tau_{(d)})=t\Big)\Big] \\ &= \arg\max_t \Bigg\{\frac{1}{\sqrt{2\pi}\delta}\Big[\prod_{j=1}^{J}\frac{1}{\sqrt{2\pi}\sigma_j}\Big]\exp\Big[-\Big(\frac{(t-\lambda)^2}{2\delta^2} \\ &+ \sum_{j=1}^{j}\frac{(o_j-t-\mu_j)^2}{2\sigma_j^2}\Big)\Big]\Bigg\} \end{aligned} \tag{29}$$

Maximizing the right-hand-side of (29) gives the following MAP estimated statistics of the true journey time t:

$$\hat{\sigma}^2_{t^*} = \left[\frac{1}{\delta^2}+\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\right]^{-1} \tag{30}$$

$$\hat{t}^* = \hat{t}^*_{MAP} = \Big[\frac{\lambda}{\delta^2}+\sum_{j=1}^{J}\frac{o_j-\mu_j}{\sigma_j^2}\Big]\hat{\sigma}^2_{t^*} \tag{31}$$

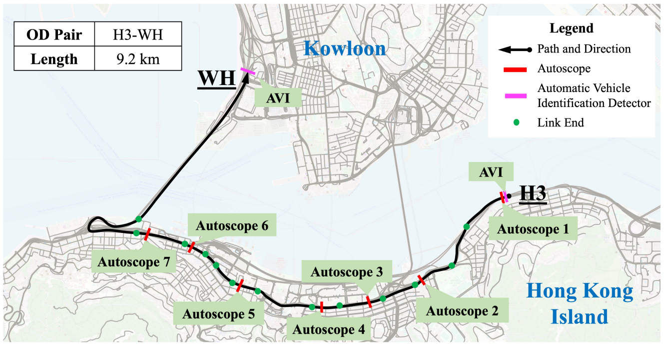
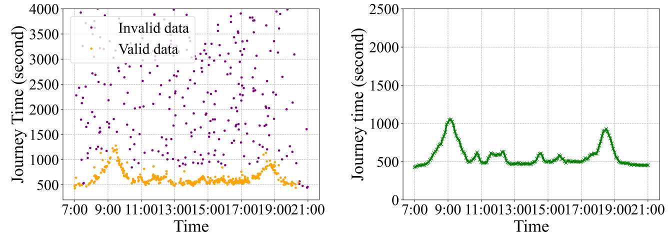


Fig. 2: Case study corridor



(a) AVI (b) Autoscope

Fig. 3: Journey times inferred from AVI system and Autoscope system - 24 January 2018

III. Case Study

A. Data collection

The fusion algorithm is now implemented and tested with a case study using real-world traffic data collected from a selected Hong Kong corridor as shown in Fig. 2. The selected corridor connects the Island Eastern Corridor in Hong Kong Island with the Western Harbour Crossing in Kowloon. The total length of the study route is 9.2 kilometers and the journey time of the study path under free-flow conditions is 7.6 minutes.

Data are collected from two different sources from 7:00 to 21:00 on all weekdays during the study period: 1 November 2017 to 31 March 2018. Regarding the data sources, we first have a pair of RFID (radio frequency identification)-based AVI sensors deployed at both ends of the corridor with which the journey times of detected vehicles can be derived by matching their identifications at the two ends. It is known that journey times inferred from AVI could contain various disturbances due to mis-identification of vehicles, detouring, or stopping of vehicles en-route [1], [45]. A filtering algorithm is adopted herein based upon the ST-DBSCAN (Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise) method [46], [47] to remove the invalid journey time observations. The filtered journey time data are aggregated and interpolated based on the median value of the valid measurements every 5-min period. Fig. 3(a) shows an example of filtered and aggregated journey time observations on 24 January 2018.

We further have cameras deployed under the Autoscope system at seven locations along the corridor with which sectional journey times of the vehicles can be derived through image processing. Given the location-wide measurements, the spatio-temporal speed field of the study corridor is constructed by using the established kernel-based Adaptive Smoothing Method (ASM) [48], with which the journey time profiles can be constructed by using a frozen field method [45] with a temporal resolution of 5-minute. Fig. 3(b) is an example of journey time records converted from the sectional cameras along the corridor on 24 January 2018.

The traffic data are divided into two sets in the experiments: data collected during 1 November 2017 and 28 February 2018 are regarded as the training dataset; data collected during 1 March 2018 and 31 March 2018 are regarded as the testing dataset. For evaluation, validated ground truths of the journey times are also provided by the Hong Kong Transport Department.

B. State classification

We start with exploring the performance of the state classification method with the feeding data. The parameters of the mixture distribution model are first determined with the training dataset for traffic state classification. Three of the most commonly used probability distributions in the literature are chosen for constructing the mixture model: Gaussian, lognormal and Gamma distribution models. We will also study the effect of having different number of traffic states on the performance of the classification algorithm.

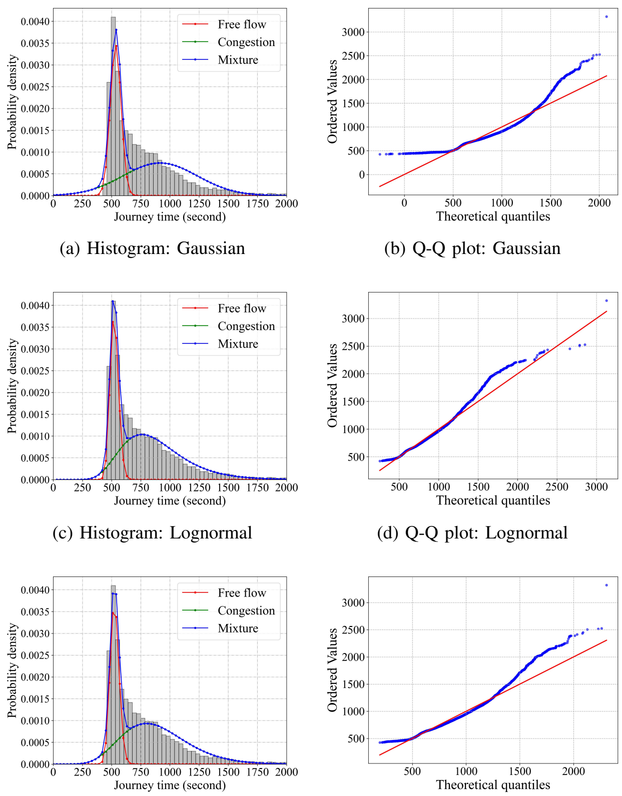
Fig. 4 first shows the empirical distribution (see the histogram in the figure) of all journey times inferred from the data sources during the training period, and their comparison against the probability density function generated by the general mixture model. We start with considering two states (free-flow and congested) in the general mixture model. It is noted that the lognormal mixture model has the best fit among the three distribution model settings considered herein as shown in Fig. 4(b), Fig. 4(d) and Fig. 4(f). However, the deviation from the diagonal line suggests the presence of an additional distinct traffic state cluster in the journey time data. The speed-flow data inferred from the Autoscope camera at location ‘Autoscope 3’ (see Fig. 2) are used to further illustrate the classification performance, with the results obtained from lognormal mixture model shown in Fig. 6(a). The blue dots in the figures are the measurements classified to be free flow (i.e., upper portion of the speed-flow scatter plot) while the red dots represent data classified as congested (i.e., lower portion of the speed-flow scatter plot). The figure reveals that the traffic data within the transition areas between congested and free-flow conditions cannot be fully distinguished by the general mixture model with only two components.

To improve the classification results, we further add a third traffic state (the transition state) into the general mixture model setting and the corresponding results are shown in Fig. 5 and Fig. 6(b). Compared to Fig. 4 and Fig. 6(a), our results generally reveal that the general mixture model with incorporation of three traffic states could identify the observations within the transition states and hence deliver an improved classification performance (see also [49]). Table

TABLE I: Statistics of journey time estimates with different general mixture model settings

Unit: seconds	Free Flow			Congested		
	Mean	STD	Weight(%)	Mean	STD	Weight(%)
Gaussian	534.84	46.53	40.3	915.84	317.97	59.7
Lognormal	523.86	36.49	34.2	883.60	287.67	65.8
Gamma	526.32	38.41	35.5	892.24	288.79	64.5

	Free Flow			Transition			Congestion		
	Mean	STD	Weight(%)	Mean	STD	Weight(%)	Mean	STD	Weight (%)
Gaussian	522.25	35.69	35.3	771.77	161.15	48.6	1237.65	375.40	16.1
Lognormal	519.30	33.70	33.5	746.69	156.56	45.0	1175.11	358.05	21.5
Gamma	520.26	34.16	34.0	756.95	158.02	46.7	1202.47	363.56	19.3

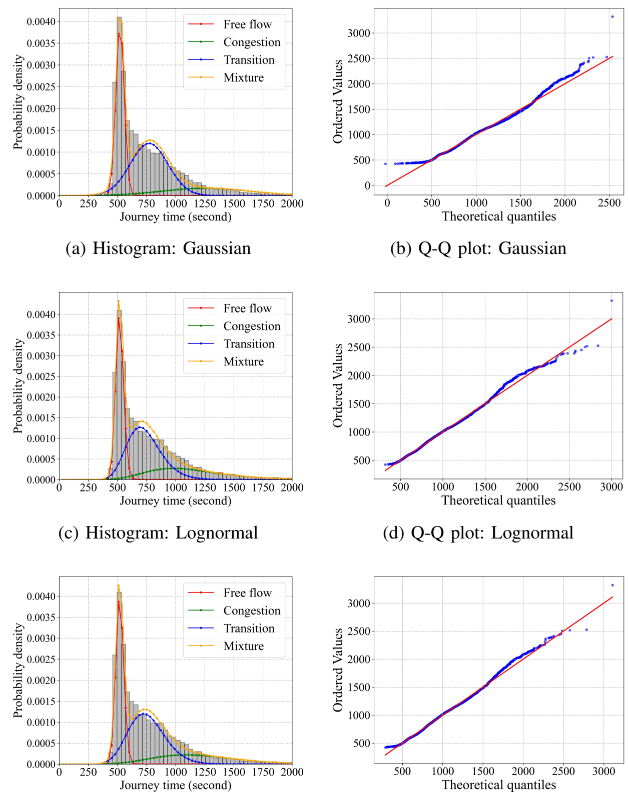


(a) Histogram: Gaussian (b) Q-Q plot: Gaussian

(c) Histogram: Lognormal (d) Q-Q plot: Lognormal

(e) Histogram: Gamma (f) Q-Q plot: Gamma

Fig. 4: General mixture model with two traffic states



(a) Histogram: Gaussian (b) Q-Q plot: Gaussian

(c) Histogram: Lognormal (d) Q-Q plot: Lognormal

(e) Histogram: Gamma (f) Q-Q plot: Gamma

Fig. 5: General mixture model with three traffic states

I further summarizes the statistics of the state classification generated by the three different mixture models (Gaussian, lognormal, and Gamma) with two and three state components. The results show that the lognormal and Gamma mixture models perform similarly when incorporating either two or three state components into the mixture model.

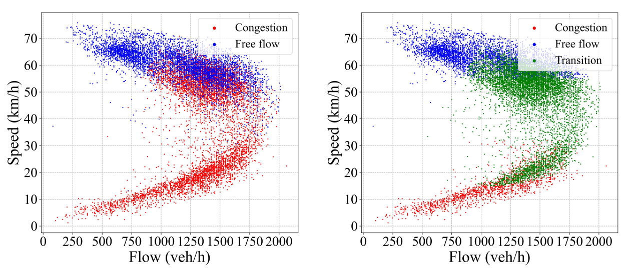
With the prevailing traffic state classification, Table II shows the statistics of journey time measurement errors associated with different mixture distribution model settings. In general, the measurement errors under congested conditions have larger mean and standard deviation values than those under free-flow or transition conditions. This is consistent with our understanding of journey time characteristics. We attribute the higher variability of travel time errors to the different congestion phases along the complex study corridor that contains signalized intersections, frontage access with entries and exits, and a cross-harbour tunnel [50]. The inherent nature of the study corridor and the varying level of traffic flow makes the lower variability almost unlikely to be observed during congestion period. Moreover, it can be observed that the measurement errors classified by lognormal mixture model generally have the smallest mean and standard deviation. This suggests that the lognormal mixture could better describe the journey time distribution features and achieve more accurate classification results.

To validate the error distribution, we compare the histogram of errors with several different types of distributions, including Gaussian, skew-normal, Gaussian mixture, generalized normal and logistic distributions, as shown in Fig. 7. It can be found that the errors associated with different data sources and traffic states do not exhibit clear Gaussian distributions. Fig. 8 further presents the Q-Q plots of Autoscope errors under free

TABLE II: Statistics of journey time estimate errors with different mixture model settings (unit: [seconds])

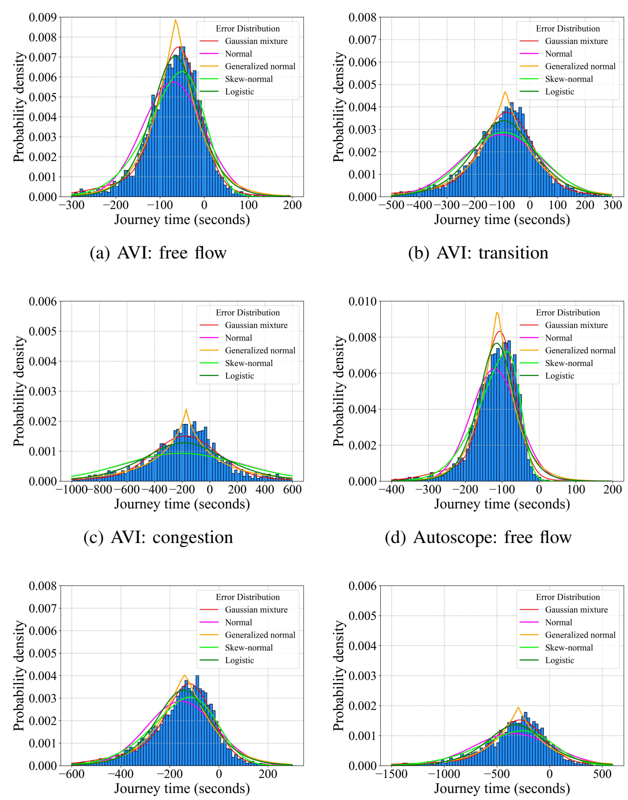
Mixture	Sources	Free Flow		Congested	
		Mean	STD	Mean	STD
Gaussian	AVI	-72.24	75.06	-134.09	275.55
	Autoscope	-120.92	69.56	-219.01	255.42
Lognormal	AVI	-71.38	71.65	-128.58	263.88
	Autoscope	-121.48	65.79	-209.04	246.09
Gamma	AVI	-71.36	72.07	-129.83	266.30
	Autoscope	-121.14	66.34	-211.16	248.03

Mixture	Sources	Free Flow		Transition		Congestion	
		Mean	STD	Mean	STD	Mean	STD
Gaussian	AVI	-71.24	71.58	-108.74	155.67	-208.66	491.85
	Autoscope	-121.21	65.82	-165.12	152.75	-385.05	411.83
Lognormal	AVI	-70.99	69.38	-101.62	144.29	-195.22	432.36
	Autoscope	-120.79	64.01	-154.58	139.34	-346.01	371.80
Gamma	AVI	-71.19	70.61	-103.92	149.02	-204.18	458.29
	Autoscope	-121.11	65.07	-158.66	144.30	-365.37	389.93



(a) Two traffic states (b) Three traffic states

Fig. 6: Clustering of speed-flow data with different numbers of traffic states classified by Lognormal mixture model



(a) AVI: free flow (b) AVI: transition

(c) AVI: congestion (d) Autoscope: free flow

(e) Autoscope: transition (f) Autoscope: congestion

Fig. 7: Error distributions of two data sources under different traffic states classified by Lognormal mixture model

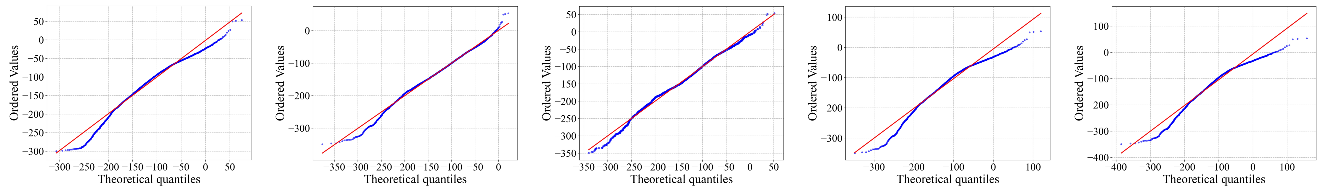
flow conditions with different error distributions. Both figures indicate that the skew-normal and Gaussian mixtures show better fitting performance for both data sources.

On the characteristics of the data sources, we find that both AVI and Autoscope tend to underestimate the path journey times with the mean errors over all traffic states being negative. Therefore, it is necessary to use non-zero mean error distributions to account for such systematic bias. Meanwhile, the AVI data produce a smaller absolute mean error but a larger standard deviation compared with the Autoscope system. Therefore, it is necessary to fuse the data collected from different data sources for more reliable estimation on journey time distribution.

C. Journey time estimation with data fusion

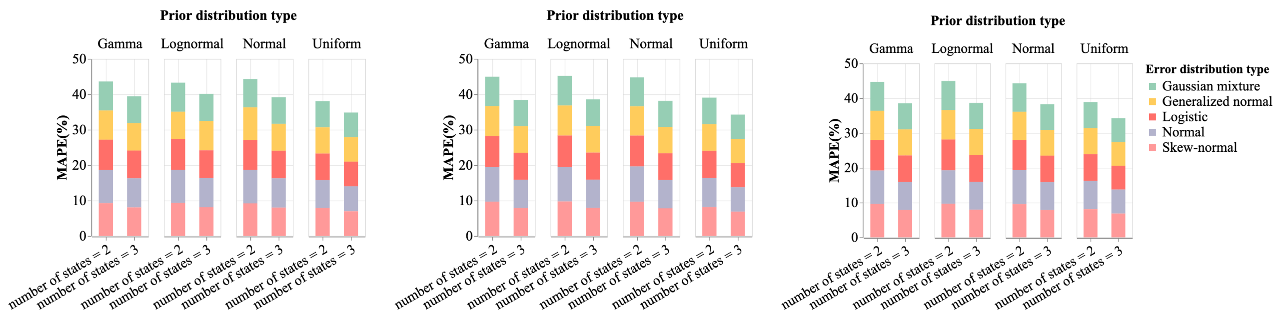
The trained mixture model based classification is now incorporated into the Bayesian data fusion framework. The metrics shown herein include both the point estimation (*i.e.*, mean absolute error (MAE) and mean absolute percentage error (MAPE)) and interval estimation (*i.e.*, average coverage error (ACE), captured mean width interval (CMWI) and index F [51]), in which the confidence level (*i.e.* α) for the interval estimation is set to be 90%.

Fig. 9 compares the stacked MAPE results produced by different classification mixture models and fusion model settings. We first note the effect of different settings of the mixture distribution models for traffic state classification incorporated in the data fusion framework. Similar to the results shown in the training process, it is found that all mixture models with the incorporation of three traffic states (free-flow, transition, congested) could lead to higher eventual estimation accuracies compared to those with only two states (free-flow and congested). It is also interesting to note that the best mixture model identified from the training process (*i.e.*, lognormal mixture model) performs similarly to the Gaussian mixture model in the overall fusion framework. In fact, the benefit of adopting a more computationally demanding lognormal mix-



(a) Gaussian (b) Skew-normal (c) Gaussian mixture (d) Generalized normal (e) Logistic

Fig. 8: Distribution analysis of Autoscope errors under free flow conditions using Q-Q Plot



(a) Classified by Gaussian mixture (b) Classified by Lognormal mixture (c) Classified by Gamma mixture

Fig. 9: Comparison of MAPE with different prior and error distribution settings with data classified by different general mixture models

TABLE III: Accuracies of journey time estimates under different model settings ($\alpha = 90\%$)

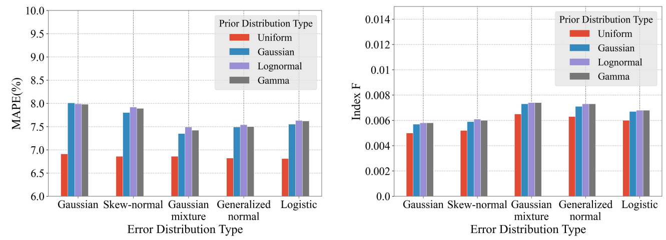
Prior	Error	MAE (unit: [s])	MAPE(%)	ACE(%)	CMWI (unit: [s])	Index F
Uniform	Gaussian	72.49	6.91	3.55	402.63	0.0050
	Skew-normal	72.89	6.86	3.21	388.52	0.0052
	Gaussian mixture	73.89	6.86	-0.71	308.46	0.0065
	Generalized normal	73.85	6.82	-0.12	313.92	0.0063
	Logistic	72.73	6.81	1.09	334.74	0.0060
Gaussian	Gaussian	86.14	8.01	-1.41	355.04	0.0057
	Skew-normal	83.34	7.81	-1.95	344.57	0.0059
	Gaussian mixture	79.26	7.35	-7.01	277.73	0.0073
	Generalized normal	80.38	7.43	-6.54	283.30	0.0071
	Logistic	80.32	7.54	-4.40	302.89	0.0067
Lognormal	Gaussian	86.72	7.99	-2.26	342.88	0.0058
	Skew-normal	86.12	7.92	-3.81	331.78	0.0061
	Gaussian mixture	81.74	7.49	-8.51	273.26	0.0074
	Generalized normal	82.79	7.54	-8.01	272.3	0.0073
	Logistic	82.60	7.63	-5.50	294.43	0.0068
Gamma	Gaussian	85.68	7.98	-2.06	346.43	0.0058
	Skew-normal	85.22	7.89	-3.41	336.24	0.0060
	Gaussian mixture	80.65	7.42	-8.06	272.42	0.0074
	Generalized normal	81.83	7.50	-7.41	274.45	0.0073
	Logistic	81.97	7.62	-5.10	297.17	0.0068

ture distribution model for state classification is not significant when compared with other mixture distribution models in the overall fusion framework. The observations herein suggest that the selection of the number of traffic states to be incorporated in the classification is more important than the choice of distribution models in traffic state classification.

Table III summaries all statistics of the journey time estimates obtained by different model settings given the best classification model (lognormal mixture) with respect to given validated ground truths and Fig. 10 presents a comparative analysis of point estimation metric MAPE and interval estimation index F with cluster columns. On the choice of prior distribution, it is found that the best prior information identified from the testing results is the uniform distribution with consideration point estimation metrics (*i.e.*, MAE and MAPE). This suggests that the benefit of incorporating a prior assumption on journey time characteristics does not help to improve the point estimates of journey times in the present case study. However, the benefits of incorporating prior knowledge of journey times are realized in interval estimations as the index F values delivered by the prediction with the incorporation of Gaussian, lognormal and Gamma prior distributions are higher than that with uniform distribution in all settings. This implies that prior assumptions with information inferred from historical data could reduce the estimated interval width with the same confidence ($\alpha = 90\%$) and hence higher reliability of journey

TABLE IV: Accuracies of point journey time estimates under different classified traffic states

	Free Flow			Transition			Congestion		
	AVI	Autoscope	Fused	AVI	Autoscope	Fused	AVI	Autoscope	Fused
MAE(s)	75.76	120.66	37.64	119.41	161.24	71.52	280.55	428.66	179.91
MAPE	11.77%	19.00%	5.95%	13.12%	17.60%	7.99%	16.13%	24.59%	10.00%



(a) MAPE

(b) Index F

Fig. 10: Comparison of MAPE and index F metrics with different prior and error distribution settings

time estimates. Among the informative prior distributions, it is noted that the journey times estimated by Gaussian prior distribution deliver a similar or even better performance to those estimated by lognormal and Gamma prior on some occasions in terms of MAPE and index F. This may suggest that the skewness captured by both lognormal and Gamma distribution does not significantly impact the journey time distribution modeling, as the symmetric Gaussian performs similarly in this case. It can be understood for the fact that the classification of data based on traffic states may lead to a reduction in skewness within each individual subset. The finding echoes some of the previous studies on journey time distribution modeling [28], [52]–[54].

We now investigate the impact of the error distribution on the estimation results. It can be observed from Fig. 10(a) that the best fitting model identified from the Q-Q plots in the training process (*i.e.*, Gaussian mixture model) could achieve the best point estimation performance. However, it also results in the underestimation of variability as shown in the value of ACE. We attribute this to the fact that the greater informativeness derived from individual data sources gives stronger constraints to the parameters of interest, which leads to a smaller value of uncertainty.

It can be seen that there is a trade-off between estimation accuracy and corresponding reliability due to the interactions between prior and likelihood functions. To further analyze this effect, Fig. 11 compares the shapes of posterior distribution under different combinations of prior and likelihood distributions. As shown in Fig. 11(a) and Fig. 11(c), the narrower error distribution makes the likelihood dominate the updating process and leads to a smaller variance in the posterior. Compared to the uncertainty estimation, it is found that the choice of error distribution has a limited impact on the MAP estimates. In contrast, the choice of prior contributes less to the location and shape of the posterior distribution in this scenario since they are less informative compared to the likelihood as presented in Fig. 11(a) and Fig. 11(b). However, we can still see the benefits of incorporating prior as they could help to avoid the posterior probabilities being concentrated in a neighborhood of possibly inaccurate information sources and hence reduce the accumulation of errors in the fusion process.

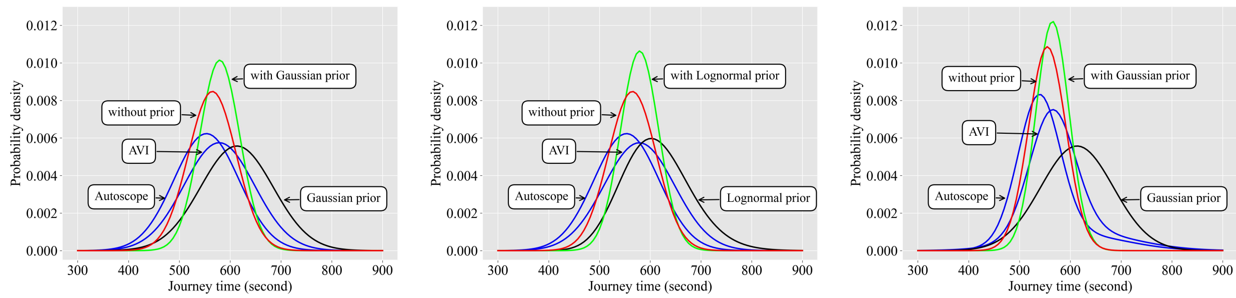
Considering the influence of prior and error distribution, the combination of Gaussian prior and skew-normal error distributions is regarded as the optimal model setting scenario as it strikes a balance between the point and interval estimation. Fig. 12 further compares the cumulative distribution of the estimation errors under the best scenario with the uniform prior setting. The Hong Kong Transport Department requires that fusion results over 95% of journey time estimates need to have absolute errors of less than 20%. It can be found that the fusion model incorporated with Gaussian prior distributions could reach the requirement although the uniform prior distribution can deliver even better performance. Moreover, the journey time intervals estimated with the uniform prior are much larger than those with the Gaussian prior distribution given the same confidence level $\alpha = 90\%$.

Table IV further presents the point estimation metrics under different classified traffic states achieved by the best model setting (combination of lognormal mixture classification, Gaussian prior and skew-normal error distributions). In general, the results show that the journey times processed by the fusion algorithm can outperform those obtained directly from individual sources via the point and interval detectors deployed on-site in terms of accuracy. It is also found that the estimation error is the smallest under free-flow condition, with larger improvements gained in MAPE reduction under the congested condition.

Fig. 13 further visualizes the corresponding results of journey time interval estimations on selected weekdays. The red dots in the figure represent the validated measurements from the field, and the shaded area represents the interval estimation of journey time with a confidence α of 90%. It is observed that the validated measurements are well captured within the estimated journey time interval with surges of journey time during congested periods recognized associated with wider prediction intervals on all selected days.

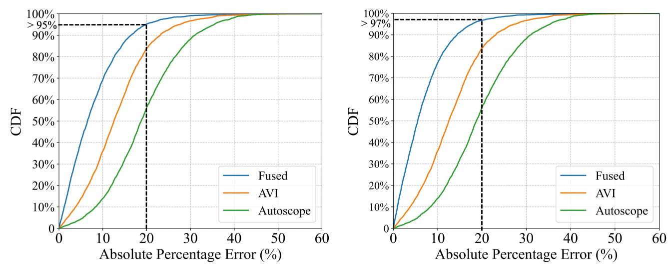
IV. Conclusion

This paper presents a Bayesian data fusion framework with a general mixture distribution traffic state classification model for the estimation of vehicular journey time distributions using heterogeneous data. In contrast to most studies, the proposed data fusion framework provides associated variability estimation with a generalized and transferable statistical model. The framework incorporates traffic state classification based on a general mixture distribution model, with the journey time variability then being estimated based on the road traffic conditions. Feeding data collected from multiple data sources are classified based on their associated traffic conditions, and



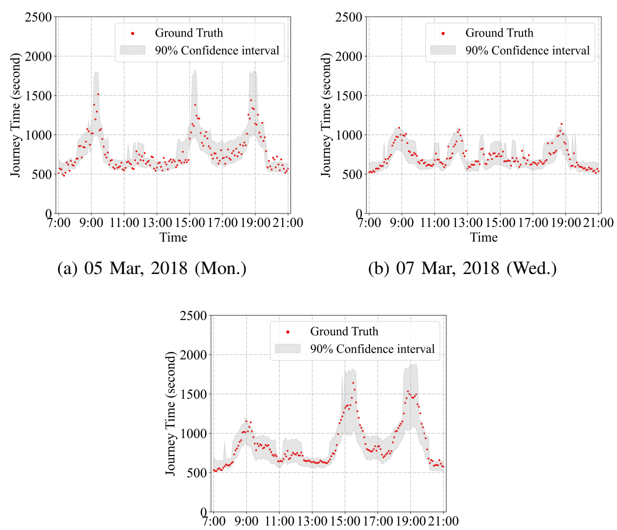
(a) Gaussian prior with Gaussian error (b) Lognormal prior with Gaussian error (c) Gaussian prior with Gaussian mixture error

Fig. 11: Effect of prior and individual likelihood distribution on the estimated results



(a) Gaussian prior (b) Uniform prior

Fig. 12: Cumulative distribution of journey time estimation errors with different prior distributions (with skew-normal error)



(a) 05 Mar, 2018 (Mon.) (b) 07 Mar, 2018 (Wed.)

(c) 09 Mar, 2018 (Fri.)

Fig. 13: Estimated journey time variability on different weekdays (confidence level: $\alpha = 90\%$)

the corresponding biases of the individual data sources are investigated by using arbitrary distributions. Lastly, the vehicular journey times and their associated variability are estimated by Bayesian theory and different statistical distributions of prior and error likelihood are applied and compared.

The proposed framework is implemented and tested using actual data collected in a real-world setting on a Hong Kong highway corridor. The experimental results reveal significant improvements in the estimations of journey times and the associated variability compared with measurements obtained from individual sources. We then use the underlying general mixture distribution model to further analyze the effect of incorporating traffic state classification. The results show that preclassifying the feeding data according to their associated traffic conditions can improve the estimation accuracy. We also find that the number of traffic conditions included in the model influences the fusion performance, and that classifying data into three clusters can better describe the traffic conditions. However, the improvement is more significant in the free-flow condition than the congested condition. This finding concurs with the difficulties that are frequently encountered when modeling the characteristics of congested traffic given the complicated dynamics involved [1]. The classification results of the different mixture models also indicate that the lognormal mixture can better describe the journey time characteristics than the Gaussian and Gamma mixture models.

Our test results using different algorithmic settings show that the uniform prior distribution provides the best deterministic estimation. This suggests that incorporating prior knowledge did not improve the deterministic journey time estimation. We believe this could be because the correlation among the day-to-day journey time data is not significant as the journey time patterns are vulnerable to multiple exogenous factors. Nevertheless, it is observed that the inclusion of informative prior distributions can avoid error accumulation from observations in the fusion process. Moreover, our results also find that the choice of error distribution contributes more significantly to the interval estimation compared with point estimation. Considering the balance between ACE and CMWI, Gaussian prior with skew-normal error distributions can significantly reduce the width of the estimated intervals while maintaining the coverage of the journey time records at an acceptable level, in contrast to the non-informative distributions.

The proposed data fusion framework contributes to the development of reliability-based intelligent transportation systems through the improved estimation of the journey time and the associated variability. Our framework also offers a generalized statistical foundation for making full use of multiple traffic data sources. Currently the Bayesian fusion framework is limited to the genuine likelihood without incorporating weights to multiple data sources in order to account for their variations in accuracy or confidence levels. Future work we have been working on is to construct a real-time learning algorithm that can refine the weights assigned to different data sources with respect to prevailing traffic conditions via a weighted likelihood function in the Bayesian inference [55].

References

[1] A. H. F. Chow, A. Santacreu, I. Tsapakis, G. Tanaksaranond, and T. Cheng, "Empirical assessment of urban traffic congestion," *Journal of Advanced Transportation*, vol. 48(8), pp. 1000–1016, 2014.

[2] R. Zhong, J. Luo, H. Cai, A. Sumalee, F. Yuan, and A. H. F. Chow, "Forecasting journey time distribution with consideration to abnormal traffic conditions," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 292–311, 2017.

[3] U. Mori, A. Mendiburu, M. Álvarez, and J. A. Lozano, "A review of travel time estimation and forecasting for advanced traveller information systems," *Transportmetrica A: Transport Science*, vol. 11, no. 2, pp. 119–157, 2015.

[4] M. L. Tam and W. H. K. Lam, "Application of automatic vehicle identification technology for real-time journey time estimation," *Information Fusion*, vol. 12, no. 1, pp. 11–19, 2011.

[5] H. B. Celikoglu, "Flow-based freeway travel-time estimation: A comparative evaluation within dynamic path loading," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 772–781, 2013.

[6] E. Jenelius and H. N. Koutsopoulos, "Travel time estimation for urban road networks using low frequency probe vehicle data," *Transportation Research Part B: Methodological*, vol. 53, pp. 64–81, 2013.

[7] J. Qiu, L. Du, D. Zhang, S. Su, and Z. Tian, "Nei-tte: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2659–2666, 2020.

[8] L. Han, B. Du, J. Lin, L. Sun, X. Li, and Y. Peng, "Multi-semantic path representation learning for travel time estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13 108–13 117, 2021.

[9] F. Guo, J. W. Polak, R. Krishnan *et al.*, "Predictor fusion for short-term traffic forecasting," *Transportation research part C: emerging technologies*, vol. 92, pp. 90–100, 2018.

[10] K. Tang, S. Chen, and A. J. Khattak, "Personalized travel time estimation for urban road networks: A tensor-based context-aware approach," *Expert Systems with Applications*, vol. 103, pp. 118–132, 2018.

[11] K. Tang, S. Chen, and Z. Liu, "Citywide spatial-temporal travel time estimation using big and sparse trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 4023–4034, 2018.

[12] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges–a survey," *Information Fusion*, vol. 12, no. 1, pp. 4–10, 2011.

[13] G. Guido, S. S. Haghshenas, A. Vitale, and V. Astarita, "Challenges and opportunities of using data fusion methods for travel time estimation," in *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1. IEEE, 2022, pp. 587–592.

[14] K. Choi and Y. Chung, "A data fusion algorithm for estimating link travel time," *ITS journal*, vol. 7, no. 3-4, pp. 235–260, 2002.

[15] C. Bachmann, B. Abdulhai, M. J. Roorda, and B. Moshiri, "A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling," *Transportation research part C: emerging technologies*, vol. 26, pp. 33–48, 2013.

[16] N.-E. El Faouzi and L. A. Klein, "Data fusion for its: techniques and research needs," *Transportation Research Procedia*, vol. 15, pp. 495–512, 2016.

[17] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.

[18] Y. Yuan, J. Van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, "Real-time lagrangian traffic state estimator for freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 59–70, 2012.

[19] L. Chu, S. Oh, and W. Recker, "Adaptive kalman filter based freeway travel time estimation," in *84th TRB Annual Meeting, Washington DC*. Citeseer, 2005.

[20] J. Han, "Multi-sensor data fusion for travel time estimation," Ph.D. dissertation, Imperial College London, 2012.

[21] X.-S. Trinh, D. Ngoduy, M. Keyvan-Ekbatani, and B. Robertson, "Incremental unscented kalman filter for real-time traffic estimation on motorways using multi-source data," *Transportmetrica A: Transport Science*, vol. 0, no. 0, pp. 1–27, 2021.

[22] E. F. Nour-Eddin, "Bayesian and evidential approaches for traffic data fusion: methodological issues and case study," Transportation Research Board 85th Annual Meeting, Tech. Rep., 2006.

[23] J. Zhao, Y. Gao, Y. Qu, H. Yin, Y. Liu, and H. Sun, "Travel time prediction: Based on gated recurrent unit method and data fusion," *IEEE Access*, vol. 6, pp. 70 463–70 472, 2018.

[24] N.-E. El Faouzi, L. A. Klein, and O. De Mouzon, "Improving travel time estimates from inductive loop and toll collection data with dempster–shafer data fusion," *Transportation research record*, vol. 2129, no. 1, pp. 73–80, 2009.

[25] Y. Guo and L. Yang, "Reliable estimation of urban link travel time using multi-sensor data fusion," *Information*, vol. 11, no. 5, p. 267, 2020.

[26] L. Zhu, F. Guo, J. W. Polak, and R. Krishnan, "Urban link travel time estimation using traffic states-based data fusion," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 651–663, 2018.

[27] F. Sun, R. Gao, W. Xing, Y. Zhang, W. Lu, J. Fang, and S. Liu, "Deep fusion for travel time estimation based on road network topology," *IEEE Intelligent Systems*, vol. 37, no. 3, pp. 98–107, 2022.

[28] S. Clark and D. Watling, "Modelling network travel time reliability under stochastic demand," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 119–140, 2005.

[29] M. Fosgerau and D. Fukuda, "Valuing travel time variability: Characteristics of the travel time distribution on an urban road," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 83–101, 2012.

[30] M. M. Rahman, S. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 453–466, 2018.

[31] N. Uno, F. Kurauchi, H. Tamura, and Y. Iida, "Using bus probe data for analysis of travel time variability," *Journal of Intelligent Transportation Systems*, vol. 13, no. 1, pp. 2–15, 2009.

[32] A. Sumalee, T. Pan, R. Zhong, N. Uno, and N. Indra-Payoong, "Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: Algorithm and case studies," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 263–285, 2013.

[33] M. Rahmani, E. Jenelius, and H. N. Koutsopoulos, "Non-parametric estimation of route travel time distributions from low-frequency floating car data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 343–362, 2015.

[34] Z. Ma, H. N. Koutsopoulos, L. Ferreira, and M. Mesbah, "Estimation of trip travel time distribution using a generalized markov chain approach," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 1–21, 2017.

[35] M. Chen, G. Yu, P. Chen, and Y. Wang, "A copula-based approach for estimating the travel time reliability of urban arterial," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 1–23, 2017.

[36] C. Shi, B. Y. Chen, W. H. Lam, and Q. Li, "Heterogeneous data fusion method to estimate travel time distributions in congested road networks," *Sensors*, vol. 17, no. 12, p. 2822, 2017.

[37] S. Mil and M. Piantanakulchai, "Modified bayesian data fusion model for travel time estimation considering spurious data and traffic conditions," *Applied Soft Computing*, vol. 72, pp. 65–78, 2018.

[38] A. Gemma, L. Mannini, S. Carrese, E. Cipriani, and U. Crisalli, "A gaussian mixture model and data fusion approach for urban travel time forecast," in *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2021, pp. 1–6.

[39] E. Saffari, M. Yildirimoglu, and M. Hickman, "Data fusion for estimating macroscopic fundamental diagram in large-scale urban networks," *Transportation Research Part C: Emerging Technologies*, vol. 137, p. 103555, 2022.

[40] W. Fang, W. Zhuo, Y. Song, J. Yan, T. Zhou, and J. Qin, "∆free-lstm: An error distribution free deep learning for short-term traffic flow forecasting," *Neurocomputing*, vol. 526, pp. 180–190, 2023.

[41] C. Wang, Q. Zhu, Z. Shan, Y. Xia, and Y. Liu, "Fusing heterogeneous traffic data by kalman filters and gaussian mixture models," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 276–281.

[42] X. Liu, L. Pan, and X. Sun, "Real-time traffic status classification based on gaussian mixture model," in *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2016, pp. 573–578.
[43] K. P. Murphy, "Probabilistic Machine Learning: An Introduction," *The MIT Press, Cambridge, MA*, 2022.
[44] A. Ang and W. Tang, "Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering (2nd Edition)," *John Wiley & Sons, Hoboken, NJ*, 2007.
[45] A. H. F. Chow, "Heterogeneous urban traffic data and their integration through kernel-based interpolation," *Journal of Facilities Management*, vol. 14(2), p. 165–178, 2016.
[46] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.
[47] D. Birant and A. Kut, "St-dbscan: An algorithm for clustering spatial–temporal data," *Data & knowledge engineering*, vol. 60, no. 1, pp. 208–221, 2007.
[48] M. Treiber and D. Helbing, "Reconstructing the spatio-temporal traffic dynamics from stationary detector data," *Cooperative Transportation Dynamics*, vol. 1, no. 3, pp. 3–1, 2002.
[49] B. Heydecker and J. Addison, "Analysis and modelling of traffic flow under variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 206–217, 2011.
[50] J. Van Lint, H. J. Van Zuylen, and H. Tu, "Travel time unreliability on freeways: Why measures based on variance tell only half the story," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 258–277, 2008.
[51] X. Yuan, C. Chen, M. Jiang, and Y. Yuan, "Prediction interval of wind power using parameter optimized beta distribution based lstm model," *Applied Soft Computing*, vol. 82, p. 105550, 2019.
[52] Y. Hollander and R. Liu, "Estimation of the distribution of travel times by repeated simulation," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 2, pp. 212–231, 2008.
[53] J. Kim and H. S. Mahmassani, "Compound gamma representation for modeling travel time variability in a traffic network," *Transportation Research Part B: Methodological*, vol. 80, pp. 40–63, 2015.
[54] F. Zheng, J. Li, H. Van Zuylen, X. Liu, and H. Yang, "Urban travel time reliability at different traffic conditions," *Journal of Intelligent Transportation Systems*, vol. 22, no. 2, pp. 106–120, 2018.
[55] E. Pompe, "Introducing prior information in weighted likelihood bootstrap with applications to model misspecification," *arXiv preprint arXiv:2103.14445*, 2021.

Xinyue Wu received the B.S. degree in traffic management from Dalian Maritime University, Dalian, China, in 2017, and the M.S. degree in transport from Imperial College London and University College London, U.K., in 2018. She is currently pursuing her Ph.D. degree at the Department of Systems Engineering, City University of Hong Kong. Her research interests include data fusion and analytics, deep learning, and their applications in intelligent transport systems.

Andy H. F. Chow (Member, IEEE) is currently an Associate Professor in Systems Engineering at the City University of Hong Kong. His research lies in developing tools for analyzing and managing transport systems. His doctoral dissertation on optimal control of dynamic transport networks completed in London received a Gordon Newell Memorial Dissertation Prize in 2008. Dr. Chow is a Board Member of the Hong Kong Society for Transportation Studies (HKSTS), Chartered Member of the Chartered Institute of Logistics and Transport (CILTHK), Member of the Institute of Electrical and Electronics Engineers (IEEE).

Li Zhuang received the B.Eng. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, in 2016, and the Ph.D. degree in data science from the City University of Hong Kong, Hong Kong. She is currently an Assistant Professor at the School of Cyber Science and Engineering, Southeast University, China. Her research interests focus on computer vision, machine learning, and deep learning with applications on intelligent transportation and renewable energy.

Wei Ma (Member, IEEE) received bachelor's degrees in Civil Engineering and Mathematics from Tsinghua University, China, master degrees in Machine Learning and Civil and Environmental Engineering, and PhD degree in Civil and Environmental Engineering from Carnegie Mellon University, USA. He is currently an assistant professor with the Department of Civil and Environmental Engineering at the Hong Kong Polytechnic University (PolyU). His research focuses on intersection of machine learning, data mining, and transportation network modeling, with applications for smart and sustainable mobility systems.

William H.K. Lam received the B.S. and M.S. degrees from the University of Calgary, Calgary, AB, Canada, in 1978 and 1980, respectively, and the Ph.D. degree in transportation engineering from the University of Newcastle upon Tyne, Newcastle, U.K., in 1992. He is currently an Emeritus Professor of Civil and Transportation Engineering with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong. Prof. Lam is a member of the International Advisory Committee of the International Symposium on Transportation and Traffic Theory (ISTTT). He is the Founding Editor-in-Chief of Transportmetrica and is now one of the Co-Editors-in-Chief of Transportmetrica A: Transport Science. His research interests include transport planning and traffic forecasting, ITS technology and development, smart surveillance and traffic simulation, public transport, and pedestrian studies.

S.C. Wong received the B.Sc. (Eng.) and M.Phil. degrees from The University of Hong Kong (HKU), and the Ph.D. degree in transport studies from University College London (UCL), U.K. He is currently a Chair Professor with the Department of Civil Engineering, HKU. His research interests include optimization of traffic signal settings, continuum modeling for traffic equilibrium problems, land use and transportation problems, dynamic highway and transit assignment problems, urban taxi services, and road safety.