Bayesian Traffic State Estimation Using Extended Floating Car Data

Victor Kyriacou, Yiolanda Englezou, *Member, IEEE*, Christos G. Panayiotou, *Senior Member, IEEE*, and Stelios Timotheou, *Senior Member, IEEE*

***Abstract*— Traffic state estimation is a challenging task due to the collection of sparse and noisy measurements from specific points of the traffic network. The emergence of Connected and Automated Vehicles (CAVs) provides new capabilities for traffic state estimation using extended floating car data such as position, speed and spacing information. In this work we propose a Bayesian Traffic State Estimation (BTSE) methodology for estimating the traffic density based on extended floating car data. BTSE utilizes the Bayesian paradigm to express any prior information to derive probability distributions of the traffic density of different road segments of the traffic network. Two variations of the BTSE methodology are developed to handle the offline and online estimation problem. The BTSE methodology is evaluated both using realistic SUMO micro-simulations for M25 Highway, London, U.K., and a real-life vehicle-trajectory dataset from German highways, extracted from videos recorded by drones. The efficiency and accuracy of the BTSE methodology is compared to an existing methodology in the literature. We present results for the estimation performance of the methods showing that the Bayesian methodology consistently results in lower mean absolute percentage error than the compared literature method. The BTSE methodology yields high-quality estimation results even for a low penetration rate of CAVs (e.g. 5%).**

***Index Terms*— Traffic density, traffic monitoring, probe vehicles, connected and automated vehicles, spacing measurements, probabilistic inference.**

I. Introduction

TRAFFIC state estimation (TSE) is a critical process for traffic planning, management and operations. A traffic state is often defined through the main traffic variables, namely density (veh/km), flow (veh/h) and speed (km/h) and is fully described by at most two of the aforementioned variables. TSE refers to the process of inferring traffic variables from a limited amount of noisy observed traffic data on certain road segments. This process depends on the estimation approach utilised for the task, the traffic flow models considered to describe the traffic dynamics, and the available input data.

Various methods have been proposed in the literature for efficient observation of traffic conditions [1]. Generally, estimation approaches can be categorised based on input data and physical or statistical assumptions.

In terms of physical or statistical assumptions, there are four main categories of approaches: (i) model-driven, (ii) data-driven (iii) hybrid, and (iv) streaming-data-driven. *Model-driven* approaches typically rely on traffic flow models, divided in first-order (e.g. the Lighthill-Whitham-Richards (LWR) model [2], [3]) or higher-order extensions (e.g. the Payne-Witham (PW) model [4], [5]). Most model-driven methods make use of the Kalman filter and its variations (see among others [6], [7], [8], [9], [10]) to solve the model equations and assume data obtained from fixed-location sensors. In addition, particle filters (see for example [11], [12], [13]) and the adaptive smoothing filter (e.g. [14], [15]), among many other methods [16], have been proven powerful tools for TSE. Such methods make sure that the estimation process respects basic traffic principles, however they might fail to fully capture the phenomena of real-world traffic when the considered traffic model is not representative of reality. *Data-driven* approaches rely on statistical or machine learning methods and historical data to develop an internal representation between traffic variables, without explicitly considering traffic models. Bayesian inference is an established statistical method to derive accurate TSE results when data are limited [17], [18], [19], [20], while it has been used in combination with model-based approaches for more accurate estimation results [21], [22]. The most widely used machine learning tools include deep neural networks [23], [24], support vector machines [25] and decision trees [26]. The estimation results using data-driven methods are often more accurate, however they cannot guarantee that the results will be physically feasible and they need a large amount of high-quality data to provide accurate results. Also, supervised machine learning methods require data of the true value of the estimation parameter of interest, which often not available. *Hybrid* methods combine desirable features from both model-driven and data-driven approaches in order to ensure accurate results and reduce data requirements [27], [28]. A new framework in the deep learning literature, namely physics-informed deep learning, has gain significant attention the past few years with promising results as these physics-informed regularisers reduce the space of

Manuscript received 8 June 2021; revised 31 January 2022, 12 August 2022, and 20 October 2022; accepted 1 November 2022. Date of publication 16 December 2022; date of current version 8 February 2023. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme through Project KIOS Center of Excellence (CoE) under Grant 739551; in part by the Marie Sklodowska-Curie through Project Bayesian Intelligent Transportation Systems (BITS) under Grant 101003435; and in part by the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development and through the Research Promotion Foundation, under Project CULTURE/BR-NE/0517/14. The Associate Editor for this article was X. Di. *(Victor Kyriacou and Yiolanda Englezou contributed equally to this work.) (Corresponding author: Stelios Timotheou.)*

Victor Kyriacou is with the Graduate School of Informatics, University of Amsterdam, 1012 WX Amsterdam, The Netherlands (e-mail: victorkyriacou@gmail.com).

Yiolanda Englezou, Christos G. Panayiotou, and Stelios Timotheou are with the KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, 1678 Nicosia, Cyprus (e-mail: englezou.yiolanda@ucy.ac.cy; christosp@ucy.ac.cy; timotheou.stelios@ucy.ac.cy).

Digital Object Identifier 10.1109/TITS.2022.3225057

feasible solutions and approximate solutions that are consistent with the chosen models using a limited amount of data (see for example, [29], [30], [31], [32]). Despite the recent success of physics-informed learning some studies have shown that such methods might fail to train [33] or could be computationally expensive [34]. Besides, the available traffic data should match the traffic parameters used in the underlying model. Finally, *streaming-data-driven* approaches use only streaming data (i.e. real-time data) for traffic estimation [35]. Such methods are robust against uncertain phenomena and unpredictable incidents, however they require a large amount of streaming data in order to provide accurate estimations.

In terms of traffic measurements, data is collected using either stationary or fixed-location sensors (e.g., inductive loop detectors and camera-based sensors), termed *stationary data*, or mobile sensors (e.g., GPS, speedometer) mounted on floating vehicles, termed *Floating Car Data (FCD)* [36]. Furthermore, connected and automated vehicles (CAVs) can allow the collection of extended data about the neighbourhood of a vehicle, such as the space and time headway from surrounding vehicles, using sophisticated on-board sensors (e.g., lidars, radars), termed *extended FCD (xFCD)* [37].

Stationary data methods rely on traffic variables, such as flow and occupancy, measured at the fixed locations of sensors. The density and speed can then be estimated by making weak assumptions, e.g., knowledge of the average vehicle length. Stationary sensors have been widely used by road administrators on highways since the early 90's [38]. TSE based on stationary data has been well studied in the literature [39]. However, the capabilities of fixed-location methods are limited due to the sparsity and high installation cost of sensors [40] and the inability of measuring traffic states beyond their fixed locations [41]. Also, such sensors often deliver low-quality and unreliable information, as they are subject to considerable disruptions due to system errors [42].

On the other hand, FCD-based TSE methods provide great opportunities for reducing the dependency on conventional stationary sensors and offer the ability to collect data from a wider spatiotemporal domain. FCD include the position, speed and direction of travel of moving vehicles equipped with appropriate sensors. FCD-based TSE methods allow the continuous observation and analysis of traffic conditions, either *offline* using data from conventional floating vehicles or *online* using data from connected vehicles, to provide wide-range spatiotemporal TSE information [43]. Typically these methods make use of traffic flow models that explicitly manage vehicle trajectory data. A novel model-based approach to integrate GPS data into highway traffic models was proposed in [44] and an extended Kalman filtering approach was developed in [45] to incorporate mobile data using the LWR model. A generalised least squared estimation approach to estimate macroscopic traffic states using multiple data sources was proposed in [46].

Several research works have also developed TSE methods that use both FCD and stationary data. In [17] and [47], data-driven methods were proposed that extract traffic information, e.g. the fundamental diagram, using historical stationary data and use mobile data to estimate the current traffic states. In [48], a Kalman filtering model-based TSE algorithm was developed utilizing only average speed measurements from connected vehicles and flow measurements from fixed-location sensors. The performance of the approach was investigated in a detailed microscopic simulation study under various penetration rates of connected vehicles and traffic conditions [49]. The approach was further extended in [50] and [51] to take into account multi-lane highways.

Recently, TSE methods that utilize xFCD have been proposed. Such data is collected from CAVs that have the capability to measure their *net space headway* or *gap* from the leading vehicle, in addition to their position and speed. In [35] a streaming-data-driven approach was proposed to estimate the traffic density in highways, based on Edie's definitions, using only xFCD with moderate estimation performance; this methodology was also utilised for TSE in an urban area of Barcelona [52]. Furthermore, a TSE method was proposed in [53] that estimates traffic states by differentiating the continuous cumulative counts of vehicles inside a region, obtained through spacing measurements from CAVs and using a conservation law. To reduce microscopic vehicle behaviour fluctuations, an extended approach was developed that estimates the traffic states and fundamental diagram parameters using xFCD, and further updates the traffic states using a model-based data assimilation method [54]. However, the use of deterministic traffic flow models requires offline calibration and may not fully represent reality, introducing extra uncertainty. A recent work proposed two methods based on Bayesian inference and deep learning for the estimation of the traffic flow rate only in free-flow conditions using headway data acquired by CAVs [55]. The proposed methods assume error-free measurements and a large amount of historical data (from multiple days to years) were used to form an initial probability density function of the traffic flow, updated as new measurements were obtained.

In this work we consider the problem of estimating the traffic density in mixed-traffic multi-lane highways with only xFCD without considering any underlying traffic flow model, similar to [35] and [55]. In this context, we propose a data-driven Bayesian methodology and develop algorithms for both offline and online estimation; to the authors knowledge, this is the first work that develops a Bayesian methodology for TSE using only xFCD and does not make use of a traffic model, historical data or data from any other source. The proposed methodology expresses prior beliefs about the traffic states, as a probability distribution, to provide additional information, especially due to the sparsity of xFCD. These prior beliefs combined with a fundamental traffic relationship that links measurements to state variables, yields the posterior probability density that is used to infer the traffic states, i.e. the traffic density in this case. We provide all details of the prior distributions obtained through a proper elicitation technique; other Bayesian methodologies provide general information regarding prior distributions which hinders their practical implementation by interested researchers, e.g. [29]. The way prior distributions are derived in this work eliminates the utilisation of a large amount of historical data needed by other Bayesian methodologies. Furthermore, to overcome the

stationarity nature of probability density functions (PDFs), we propose the estimation of the traffic density for discrete time windows in order to avoid the utilisation of filtering approaches, such as particle filters [56], that are computationally expensive.

The developed Bayesian methodology offers the following main advantages compared to the state-of-the-art: (i) the estimation performance is superior, especially for low CAV penetration rates, due to the additional source of information used along with the measurements, (ii) the inference of the traffic state is obtained in a probability density form, providing more detailed information compared to a point-estimate [57, Chapter 4] and, (iii) the estimation accuracy of the traffic variables does not rely on historical data or traffic models.

The contributions of this work are the following:

- We design a fully Bayesian methodology[1] that uses sparse xFCD collected from CAVs to derive the posterior predictive distribution of the traffic density in different spatiotemporal regions of a specific highway under study.
- We elaborate on the proposed methodology to develop both an offline (data-driven) and an online (streaming-data-driven) Bayesian TSE (BTSE) algorithm. The offline BTSE algorithm uses measurements collected over the entire time-horizon to estimate the traffic density in different time-windows, whereas the online BTSE algorithm uses measurements only from the previous and current time-windows to estimate the traffic density of the current time-window.
- The Bayesian methodology is compared with an existing methodology in the literature proposed by [35]. The superiority of the proposed approach is illustrated using both micro-simulations for a specific section of the M25 Highway in London, U.K., and a real-life dataset collected from German highways [59].

The remainder of the paper is organised as follows. Section II describes the TSE problem when considering mixed-traffic with xFCD. Section III presents an existing solution approach introduced by [35], which will be used to compare results obtained from the proposed method. In Section IV we focus in the solution approach where we develop an offline/online Bayesian approach to obtain the posterior predictive distribution of the traffic density. A simulation study using SUMO (Simulation of Urban MObility) and a real-life dataset are used to validate the proposed methodology in Section V. Section VI discusses the use conditions and limitations of the proposed methodology. Section VII concludes the paper and suggests future research directions.

A. Notation

In the remainder of this paper we use the following notation. All bold letters indicate vectors (lower case) or matrices (upper case), while calligraphic letters denote sets. If $\mathcal{A}$ is a set, we denote as $|\mathcal{A}|$ the cardinality of the specific set. Set $\mathbb{R}_+ = [0, \infty)$ denotes the set of all positive real numbers. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{-1}$, denote the transpose and the matrix inverse respectively. $\mathbf{I}_{n\times n}$ is the $n \times n$ identity matrix and $\mathbf{0}_{n\times n}$ denotes a $n \times n$ matrix that all its elements are zero. The median of a vector $\mathbf{x}$ is defined as the value separating the higher half from the lower half of the vector and is denoted as median($\mathbf{x}$). Furthermore, $x \sim \text{Normal}(\mu, \sigma^2)$, $x \sim \text{IG}(\alpha, \beta)$ and $x \sim \text{Unif}(c, d)$ indicate that x is a random variable drawn from the normal distribution with mean μ, and variance, σ^2, the inverse-gamma distribution with shape and scale parameters $\alpha > 0$ and $\beta > 0$, and the uniform distribution in the range $[c, d]$, respectively.

II. Problem Statement

We consider a traffic network modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the set of vertices $\mathcal{V}$ represents road junctions and the set of edges $\mathcal{E}$ represents road links. An arbitrary road link is considered to be comprised of N^S road segments and N^L lanes. We define the (i, l) space region, $i \in \mathcal{N}^S = \{1, \ldots, N^S\}$, $l \in \mathcal{N}^L = \{1, \ldots, N^L\}$, for the l-th lane of the i-th road segment as

$$\mathcal{R}_{i,l} = \{x | x_{i-1,l} \leq x \leq x_{i,l}\},\ i \in \mathcal{N}^S,\ l \in \mathcal{N}^L,$$

where, $x_{i-1,l}$ and $x_{i,l}$ denote the upstream and downstream boundary in the l-th lane of the i-th road segment.

We consider the estimation of traffic densities of distinct lanes of different road segments over a time-horizon T. The estimation time-window duration considered is T^W, such that $N^W = T/T^W$ different density values are derived over the entire time-horizon. The j-th estimation time-window is defined as $\mathcal{T}_j = \{\tau | \tau_{j-1} \leq \tau \leq \tau_j\}$, $j \in \mathcal{N}^W = \{1, \ldots, N^W\}$, such that $\tau_{N^W} - \tau_0 = T$ and $\tau_j - \tau_{j-1} = T^W$, $\forall j \in \mathcal{N}^W$.

Each space region $\mathcal{R}_{i,l}$, $i \in \mathcal{N}^S$, $l \in \mathcal{N}^L$, and time-window $\mathcal{T}_j$, $j \in \mathcal{N}^W$ define a time-space region, denoted by $\mathcal{A}_{ilj}$ [km$\times$ h], as

$$\mathcal{A}_{ilj} = \{(x, \tau) |\ x_{i-1,l} \leq x \leq x_{i,l}, \tau_{j-1} \leq \tau \leq \tau_j\}, \quad (1)$$

shown in Figure 1. The coordinates of the lower-left and upper-right corner of region $\mathcal{A}_{ilj}$ are $(x_{i-1,l}, \tau_{j-1})$ and $(x_{i,l}, \tau_j)$, respectively. Because the analysis is uniform across all time-space regions, we hereafter drop indices i, l, j, from all measurements, variables and sets and consider that region $\mathcal{A}$ is the same as $\mathcal{A}_{ilj}$, unless otherwise stated.

We consider mixed traffic conditions with two types of vehicles: (i) conventional vehicles (CVs) that do not have measuring capabilities and (ii) CAVs or connected vehicles with the ability of measuring the distance to their leading vehicle; hereafter we use the term 'CAVs' for the second type of vehicles. Let $\mathcal{N}^V$ and $\mathcal{N}^C$ denote the set of all vehicles and the set of CAVs in time-space region $\mathcal{A}$, respectively, where $\mathcal{N}^C \subseteq \mathcal{N}^V$, and $N^V = |\mathcal{N}^V|$, $N^C = |\mathcal{N}^C|$. Each CAV $c \in \mathcal{N}^C$ in $\mathcal{A}$ is equipped with various sensors for measuring:

- The *position*, given as $\mathbf{z}_c = [z_{c,1}, \ldots, z_{c,M_c}]^{\mathrm{T}}$ [km].
- The *net space headway* or *gap*,[2] given as $\mathbf{y}_c = [y_{c,1}, \ldots, y_{c,M_c}]^{\mathrm{T}}$ [km].

[1] A fully Bayesian framework requires prior distributions for all unknown entities in the model, with the posterior distribution effectively capturing all aspects of uncertainties involved. Optimal inference and prediction is achieved through the selection of the most appropriate values for nuisance parameters and hyperparameters using the available data [58].

[2] The *net space headway* or *gap* is defined as the distance between the front-bumper of the ego vehicle and the rear-bumper of its leading vehicle.

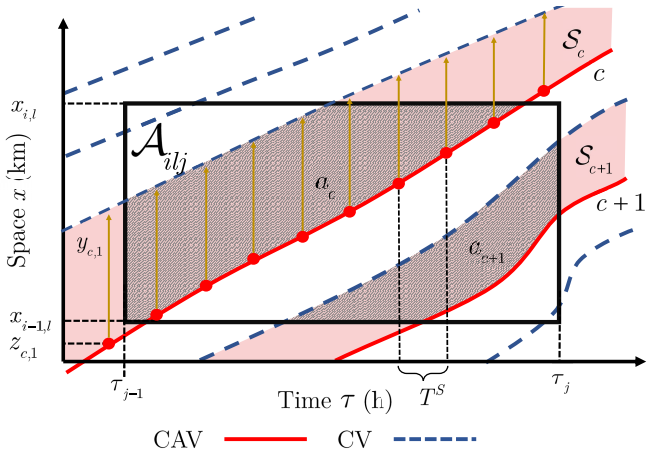

Fig. 1. Time-space region $\mathcal{A}_{ilj}$. Red bullets and yellow arrows indicate the position and spacing measurements recorded from the CAV accordingly. Red solid lines and blue dashed lines indicate the derived trajectories of CAVs and CVs, respectively. The pink shaded area is the time-space region between CAV c and its leading vehicle, $\mathcal{S}_c$, and the hatched area is the time-space region between CAV c and its leading vehicle in $\mathcal{A}_{ilj}$, a_c.

- The *speed*, given as $\mathbf{u}_c = [u_{c,1}, \ldots, u_{c,M_c}]^{\mathrm{T}}$ [km/h].
- The *time-stamp*, given as $\hat{\boldsymbol{\tau}}_c = [\hat{\tau}_{c,1}, \ldots, \hat{\tau}_{c,M_c}]^{\mathrm{T}}$ [h].

Above, M_c is the number of measurements collected by CAV c in the specific time-space region.

CVs do not have the capability to collect any data. The sampling period is set equal to T^S such that each CAV collects a new measurement every T^S time units. Based on these measurements, the following data are calculated from each CAV and communicated to an operation center:

- The *speed* $\mathbf{u}_c$.
- The *time spent* of CAV $c \in \mathcal{N}^C$ in the specific region $\mathcal{A}$, given as $t_c = \hat{\tau}_{c,M_c} - \hat{\tau}_{c,1}$ [h], where $\hat{\tau}_{c,M_c}$ and $\hat{\tau}_{c,1}$ are the time-stamps of the last and first measurements, respectively.
- The *area* of the time-space region between CAV $c \in \mathcal{N}^C$ and its leading vehicle in region $\mathcal{A}$, denoted as $a_c = \mathcal{A} \cap \mathcal{S}_c$ [km×h]. By using measurements $\mathbf{y}_c$ and $\mathbf{z}_c$, we derive $\mathcal{S}_c$ which is the time-space region, between CAV c and its leading vehicle. Finally, we calculate a_c, which is the value of the area between the intersection of $\mathcal{A}$ and $\mathcal{S}_c$, as shown in Figure 1.

The true traffic density, ρ^{true}, in time-space region $\mathcal{A}$ is defined according to [60] as

$$\rho^{true} = \frac{\sum_{n \in \mathcal{N}^V} t_n}{\sum_{n \in \mathcal{N}^V} a_n}. \tag{2}$$

Equation (2) shows that if all vehicles in the specific road segment under study are CAVs then the traffic density is known. Nonetheless, if only a proportion of traffic are CAVs, then the traffic density cannot be estimated exactly.

Given the data provided by CAVs, $\mathbf{u}_c$, t_c and a_c, $\forall c \in \mathcal{N}^C$, the objective of this work is to estimate the traffic density of time-space region $\mathcal{A}$, according to Equation (2).

We consider the estimation of the traffic density both in an *offline* and *online* framework. When offline estimation is performed, xFCD is collected by CAVs for the N^W time-windows, $\mathcal{T}_j$, $\forall j \in \mathcal{N}^W$, and the estimation method uses the total amount of information. For online estimation xFCD is collected for time-windows $\mathcal{T}_{j-1}$ and $\mathcal{T}_j$ and the estimation is performed for time-window $\mathcal{T}_j$, $\forall j \in \mathcal{N}^W$. The online procedure is repeated for the total N^W time-windows.

For simplicity the following assumptions are made. First, we assume that CAVs are randomly selected and have the same driving behaviour as the rest of the vehicles. Furthermore, we consider that measurement errors from different sensors have a cumulative effect resulting in an additive zero mean Gaussian error on density.

III. Existing Solution Approach

An online estimation approach for obtaining traffic variables using only xFCD was proposed in [35] making the assumption that CAVs[3] collect a_c and t_c, $\forall c \in \mathcal{N}_C$. The assumption of random sampling of CAVs defined in the previous section holds for this approach with the additional assumption that the measurements are collected without error. Edie's generalised definitions [60] are utilised to estimate traffic density, flow and speed for individual time-space regions. For example, to estimate traffic density, the set of all vehicles, $\mathcal{N}^V$, in (2) is replaced by $\mathcal{N}^C$, yielding

$$\hat{\rho}^{PTSE} = \frac{\sum_{c \in \mathcal{N}^C} t_c}{\sum_{c \in \mathcal{N}^C} a_c}. \tag{3}$$

Estimator (3) is biased and an inverse correlation exists between the bias and the penetration rate of CAVs, as shown in [35]. The authors concluded that their algorithm provides good estimates for 5-minute and hourly volumes; however, the error was large for low traffic demand, as well as for low CAV penetration rates. The developed Bayesian estimation algorithms are compared to this approach, which we refer to as *point TSE* (PTSE).

IV. Solution Approach

We develop a fully Bayesian approach to derive the probability distribution of the traffic density in a specific time-space region, assuming we have available the measurements acquired by CAVs, as defined in Section II. Bayesian inference is a natural way to update prior beliefs for unknown parameters through the posterior distribution and obtain marginal distributions of interest. The main idea of Bayesian inference is to continually update prior beliefs about events as new evidence, i.e. measurements, is acquired. The proposed methodology is comprised of four phases.

- *Phase A:* Utilise a fundamental definition to link measurements to parameters.
- *Phase B:* Formulate prior information about parameters.
- *Phase C:* Combine the two sources of information from Phase A and Phase B using Bayes' theorem [61] to obtain the posterior distribution.
- *Phase D:* Infer parameters using the resulting posterior distribution.

In the next section we design a Bayesian inference methodology for traffic density estimation both in an offline

[3]The term used by the authors of [35] is *probe vehicles*, instead of CAVs.

(Section IV-B) and online (Section IV-C) framework. Note however that the methodology can be used to estimate other traffic variables such as traffic flow and speed.

A. Bayesian Traffic State Estimation Methodology

1) Phase A: The main objective of Phase A is to utilize a fundamental definition that relates the measured parameters, namely, time spent and area between different CAVs and their leading vehicles of a specific time-space region, with the unknown parameter, the traffic density.

Towards this direction we consider Edie's equation of traffic density, i.e. Equation (2). Initially, we select the statistical formulation

$$\rho = g(\boldsymbol{\alpha}, \mathbf{t}) + \varepsilon, \tag{4}$$

where $\mathbf{t} = [t_1, \ldots, t_c, \ldots, t_{N^C}]^{\mathrm{T}}$ is the N^C-vector of time-spent of all CAVs and $\boldsymbol{\alpha} = [a_1, \ldots, a_c, \ldots, a_{N^C}]^{\mathrm{T}}$ the N^C-vector of each CAV's time-space region area with its leading vehicle. Furthermore, $g(\cdot, \cdot)$ is the model given by (2),

$$g(\boldsymbol{\alpha}, \mathbf{t}) = \frac{\sum_{c \in \mathcal{N}^C} t_c}{\sum_{c \in \mathcal{N}^C} a_c},$$

and ε is the noise which is a Gaussian random variable, $\varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2)$. In contrast with [35] and [55] that make an error-free assumption for the measurements obtained from the CAVs, we assume that the measurements are subject to noise, ε, as shown in (4). Note that if all vehicles in the specific region where CAVs, i.e. $\mathcal{N}^C = \mathcal{N}^V$, then (2) would be exact and known. Here, for a specific region $\mathcal{A}$ we choose to observe a_c for CAV $c \in \mathcal{N}^C$, and assume a prior distribution for the parameters t_c and the model noise σ_ε^2. We utilise the basic concepts of Bayesian inference to derive a probability distribution of the traffic density in region $\mathcal{A}$ by integrating out the unknown parameters.

The likelihood function for Equation (2), $\pi^L(\rho|\mathbf{t}, \boldsymbol{\alpha}, \sigma_\varepsilon^2)$, is

$$\pi^L(\rho|\mathbf{t}, \boldsymbol{\alpha}, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\rho - g(\boldsymbol{\alpha}, \mathbf{t})]^2\right\}. \tag{5}$$

2) Phase B: The objective of Phase B is to formulate any prior beliefs about the parameters of interest, namely, the time spent and the model noise, as a known distribution. Note that a_c is not treated as a random variable because $a_c = \mathcal{A} \cap \mathcal{S}_c$ depends on t_c through $\mathcal{S}_c$. First, we assume that the time spent of each CAV in the specific region follows a prior distribution with a known mean, $\mu_c(u)$, and variance, $\sigma_c^2(u)$, which are functions of the average speed of vehicles, u. Without loss of generality, in order to describe the *a priori* information about $\mathbf{t}$, we choose normal prior distributions, $\pi^B(t_c)$, such that $t_c \sim \text{Normal}(\mu_c(u), \sigma_c^2(u))$, where $\mu_c(u), \sigma_c^2(u) > 0$, $\forall c \in \mathcal{N}^C$, are hyperparameters that can be estimated from the collected data. As $t_1, \ldots, t_{N^C}$, are mutually independent random variables all having a normal distribution, the $N^C \times 1$ random vector $\mathbf{t}$ defined as $\mathbf{t} = [t_1, \ldots, t_{N^C}]^{\mathrm{T}}$ has a multivariate normal distribution with mean the N^C-column vector $\boldsymbol{\mu}(u) = [\mu_1(u), \ldots, \mu_{N^C}(u)]^{\mathrm{T}}$ and covariance matrix the $N^C \times N^C$ diagonal matrix with $\boldsymbol{\Sigma}(u)_{cc'} = \sigma_c^2(u)$ when $c = c'$ and zero otherwise.[4,5] Hence, without loss of generality, we have that

$$\mathbf{t} \sim \text{Normal}(\boldsymbol{\mu}(u), \boldsymbol{\Sigma}(u)). \tag{6}$$

The estimation of the mean and covariance matrix is challenging when the average speed takes continuous values, i.e., $u \in \mathbb{R}_+$. We address this issue separately for the offline and online estimation in Section IV-B and Section IV-C, respectively.

We select an inverse-gamma prior distribution, $\pi^B(\sigma_\varepsilon^2)$, for the error variance, σ_ε^2

$$\sigma_\varepsilon^2 \sim \text{IG}(\beta, \delta),$$

where again $\beta, \delta > 0$ are known hyperparameters. The inverse-gamma prior distribution is a common assumption in Bayesian statistics as it results in an analytical tractable marginal distribution. For $0 < \beta, \delta < 1$, which will be used in this work, the inverse-gamma prior distribution is an *uninformative prior* or *diffuse prior* meaning that it expresses vague or general information, i.e. information that is not subjectively elicited, about the error variance [57].

The joint prior density of $\mathbf{t}$ and σ_ε^2 is given by

$$\begin{aligned}\pi^B(\mathbf{t}, \sigma_\varepsilon^2) &= \pi^B(\mathbf{t})\, \pi^B(\sigma_\varepsilon^2) \\ &= \frac{\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(\mathbf{t} - \boldsymbol{\mu}(u))^{\mathrm{T}} \boldsymbol{\Sigma}(u)^{-1}(\mathbf{t} - \boldsymbol{\mu}(u)) - \frac{\delta}{\sigma_\varepsilon^2}\right\}}{\sigma_\varepsilon^{2\beta + \mathcal{N}^C + 2}\sqrt{|\boldsymbol{\Sigma}(u)|}}.\end{aligned} \tag{7}$$

3) Phase C: In this phase we combine the two sources of information as provided in Phases A and B, i.e. the prior distribution and the chosen definition, to obtain the posterior density, $\pi^A(\mathbf{t}, \sigma_\varepsilon^2|\rho, \boldsymbol{\alpha})$. This is achieved by using Bayes' Theorem to update our prior beliefs for parameters $\mathbf{t}$ and σ_ε^2 by using the likelihood function (5). Note that the values used for the traffic density ρ, are obtained by using Equation (2) and the spacing data collected by CAVs. The unnormalised posterior density satisfies

$$\begin{aligned}&\pi^A(\mathbf{t}, \sigma_\varepsilon^2|\rho, \boldsymbol{\alpha}) \\ &\quad \propto \pi^B(\mathbf{t})\, \pi^B(\sigma_\varepsilon^2) \pi^L(\rho|\mathbf{t}, \boldsymbol{\alpha}, \sigma_\varepsilon^2) \\ &\quad \propto \frac{\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(\mathbf{t} - \boldsymbol{\mu}(u))^{\mathrm{T}} \boldsymbol{\Sigma}(u)^{-1}(\mathbf{t} - \boldsymbol{\mu}(u))\right\}}{\sqrt{|\boldsymbol{\Sigma}(u)|}} \frac{\exp\left\{-\frac{\delta}{\sigma_\varepsilon^2}\right\}}{\sigma_\varepsilon^{2\beta + \mathcal{N}^C + 2}} \\ &\qquad \times \exp\left\{-\frac{1}{2}[\rho - g(\boldsymbol{\alpha}, \mathbf{t})]^2\right\}.\end{aligned} \tag{8}$$

The unknown error variance σ_ε^2 is a nuisance parameter, i.e. is not of immediate interest, but it still must be taken into account when studying parameters which are of interest, i.e. t_c, $c \in \mathcal{N}^C$. Using known results from statistical theory [61],

[4]This assumption is made for simplicity and without loss of generality. Even under this assumption, the results presented in Section V show the superiority of the proposed methodology. Considering the use of a non-diagonal covariance matrix may yield even better results.

[5]This is a special case of a multivariate normal distribution with a diagonal covariance matrix and can be proved by showing that the product of the probability density functions of $t_1, \ldots, t_{N^C}$ is equal to the joint probability density function of $\mathbf{t}$.

we can integrate out σ_ε^2 and obtain the marginal posterior distribution for $\mathbf{t}$:

$$\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha}) = \int_0^\infty \pi^A(\mathbf{t},\sigma_\varepsilon^2|\rho,\boldsymbol{\alpha})d\sigma_\varepsilon^2 \propto \frac{\exp\left\{-\frac{1}{2}(\mathbf{t}-\boldsymbol{\mu}(u))^{\mathrm{T}}\boldsymbol{\Sigma}(u)^{-1}(\mathbf{t}-\boldsymbol{\mu}(u))\right\}}{\sqrt{|\boldsymbol{\Sigma}(u)|}} \times\left[1+\frac{[\rho-g(\boldsymbol{\alpha},\mathbf{t})]^2}{2b}\right]^{-(\beta+\frac{N^C}{2})}, \quad (9)$$

which is basically the prior distribution multiplied by a multivariate t-distribution, with 2β degrees of freedom, mean $g(\boldsymbol{\alpha},\mathbf{t})$ and variance $\frac{\delta}{\beta}\mathbf{I}_{N^C\times N^C}$.

Phase D. The objective of Phase D is to use the posterior distribution for prediction, an important objective of Bayesian inference addressed through the predictive distribution. More specifically, assume that we have ρ, derived through (4) and the collected measurements by a fixed proportion of CAVs, and want to predict $\tilde{\rho}$, hence we want to find the predictive distribution that represents uncertainty in a new observation given previously obtained observations. We denote as $\tilde{\rho}$ the derived traffic density over all vehicles in the specific time-space region, and the posterior predictive density $\pi(\tilde{\rho}|\rho)$ is obtained by integrating out $\mathbf{t}$ with respect to its posterior distribution:

$$\pi(\tilde{\rho}|\rho) = \int_\Omega \pi(\tilde{\rho}|\rho,\mathbf{t},\boldsymbol{\alpha})\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})d\mathbf{t} \quad (10)$$

where $\Omega = [0,\infty)^{N^C}$ is the set of all possible values of time spent for the CAVs deployed in time-window $\mathcal{T}_j$ and $\pi(\tilde{\rho}|\rho,\mathbf{t},\boldsymbol{\alpha})$ is the conditional distribution of $\tilde{\rho}$ given $\mathbf{t}$, $\boldsymbol{\alpha}$ and ρ. The joint prior distribution of ρ and $\tilde{\rho}$, conditional on all unknown parameters $\mathbf{t}$, $\boldsymbol{\alpha}$ and σ_ε^2 is given by

$$\begin{pmatrix}\tilde{\rho}\\ \rho\end{pmatrix}\Bigg|\,\mathbf{t},\boldsymbol{\alpha},\sigma_\varepsilon^2 \sim N\left(\begin{pmatrix}g(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{t}})\\ g(\boldsymbol{\alpha},\mathbf{t})\end{pmatrix},\ \sigma_\varepsilon^2\mathbf{I}_{2\times 2}\right). \quad (11)$$

Standard results for normal distributions can be used to derive the following conditional posterior distribution [62]

$$\tilde{\rho}\mid\rho,\mathbf{t},\boldsymbol{\alpha},\sigma_\varepsilon^2 \sim N\left(g(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{t}}),\sigma_\varepsilon^2\right), \quad (12)$$

which is the likelihood function (5) at new values of the time-spent $\tilde{\mathbf{t}}$ and area of time-space region $\tilde{\boldsymbol{\alpha}}$, obtained from the marginal posterior density $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$. The predictive distribution is given as an average over the marginal posterior density $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$ which contains all the information we know about $\mathbf{t}$ [63]. Hence, to derive the predictive distribution we know that $\pi(\tilde{\rho}|\rho,\mathbf{t},\boldsymbol{\alpha})$ follows the likelihood distribution (5), which is a normal distribution as given in Equation (12). To calculate the posterior predictive distribution we need to marginalise $\pi(\tilde{\rho}|\rho,\mathbf{t},\boldsymbol{\alpha})$ over the posterior of $\mathbf{t}$ given $\rho,\boldsymbol{\alpha}$, i.e. $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$.

The marginal posterior density $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$, however, is not a standard distribution, hence numerical evaluation is required. We employ sampling techniques based on Markov Chain Monte Carlo (MCMC) methods (see [64, Chapter 11]). The idea of MCMC is in a sense to by-pass the mathematical operations rather than to implement them. MCMC methods construct a Markov chain[6] $\tilde{\mathbf{t}}^{(0)},\tilde{\mathbf{t}}^{(1)},\tilde{\mathbf{t}}^{(2)},\ldots,$ with steady state distribution equal to the posterior density, $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$, of interest. The empirical distribution of the first $\tilde{M}$ values, $\boldsymbol{\Xi}_{\mathrm{MCMC}} = [\tilde{\mathbf{t}}^{(1)\mathrm{T}},\ldots,\tilde{\mathbf{t}}^{(r)\mathrm{T}},\ldots,\tilde{\mathbf{t}}^{(\tilde{M})\mathrm{T}}]^{\mathrm{T}}$, will then converge to $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$ as $\tilde{M}\to\infty$ meaning that if we have a sufficiently large sample from any distribution then we effectively have the whole distribution such that we can calculate anything about the distribution (e.g. the mean or the variance) from the sample.

Algorithm 1 Metropolis-Hastings Algorithm

Input: $\tilde{\mathbf{t}}^{(0)}$, $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$
for $r = 1,2,\ldots,\tilde{M}$ **do**
  Propose: $\mathbf{t}^* \sim q(\mathbf{t}^*|\tilde{\mathbf{t}}^{(r-1)})$
  Acceptance probability:
  $P_\alpha(\mathbf{t}^*|\tilde{\mathbf{t}}^{(r-1)}) = \min\left\{1, \frac{q(\tilde{\mathbf{t}}^{(r-1)}|\mathbf{t}^*)\pi^M(\mathbf{t}^*|\rho,\boldsymbol{\alpha})}{q(\mathbf{t}^*|\tilde{\mathbf{t}}^{(r-1)})\pi^M(\tilde{\mathbf{t}}^{(r-1)}|\rho,\boldsymbol{\alpha})}\right\}$
  Sample $\omega \sim$ Unif[0, 1]
  if $\omega < P_\alpha$ **then**
    Accept the proposal: $\tilde{\mathbf{t}}^{(r)} \leftarrow \mathbf{t}^*$
  else
    Reject the proposal: $\tilde{\mathbf{t}}^{(r)} \leftarrow \tilde{\mathbf{t}}^{(r-1)}$
Output: $\boldsymbol{\Xi}_{\mathrm{MCMC}} = [\tilde{\mathbf{t}}^{(1)\mathrm{T}},\ldots,\tilde{\mathbf{t}}^{(\tilde{M})\mathrm{T}}]^{\mathrm{T}}$

A widely used MCMC algorithm that is relative simple is the Metropolis-Hastings algorithm [66], [67], which will be used for the purposes of this work (see Algorithm 1). The first step is to initialise the chain with starting values $\tilde{\mathbf{t}}^{(0)}$ for the random variables. Let the current state of the chain be $\tilde{\mathbf{t}}^{(r)}$. The main loop of the algorithm consists of three components: (i) generate a sample from a proposal density $q(\cdot)$, (ii) compute the acceptance probability, P_α, and (iii) accept or reject the candidate sample with probability P_α, or $1-P_\alpha$, respectively. The Markov property specifies that the distribution of $\tilde{\mathbf{t}}^{(r+1)}$ given all previous draws, $\tilde{\mathbf{t}}^{(r+1)}|\tilde{\mathbf{t}}^{(r)},\tilde{\mathbf{t}}^{(r-1)},\ldots,$ depends only on the most recent value drawn $\tilde{\mathbf{t}}^{(r)}$.

The Metropolis-Hastings algorithm is a general approach for sampling from a target density, in our case $\pi^M(\mathbf{t}|\rho,\boldsymbol{\alpha})$. However, it requires the specification of a proposal density, which must be chosen carefully. Acceptance rates close to $\frac{1}{4}$ where recommended for high dimensional models and close to $\frac{1}{2}$ for models of dimension 1 or 2 [63]. For the purposes of this work, at each MCMC step we propose values for $\mathbf{t}$ from a Normal distribution, $q(\cdot) = \mathrm{Normal}(\cdot,\cdot)$, $\mathbf{t}^* \sim \mathrm{Normal}(\tilde{\mathbf{t}}^{(r)},\nu^2\mathbf{R})$, where $\mathbf{R}$ is the covariance matrix resulting from the nonlinear least squares fit of (2), scaled by a value ν^2.[7] As in many MCMC methods, the draws are regarded as a sample from the

[6]**Markov chain**: A sequence of random variables $\{\boldsymbol{\theta}^{(0)},\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},\ldots\}$, such that at each time $t \geq 0$, the next state $\boldsymbol{\theta}^{(t+1)}$ is sampled from a distribution $f(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)})$ which depends only on the current state of the chain $\boldsymbol{\theta}^{(t)}$. This sequence is called *Markov chain*, and $f(\cdot|\cdot)$ is the *transition kernel* of the chain [65, Chapter 1].

[7]Following [64, Chapter 12], the most efficient proposal distribution has scale $c \approx \frac{2.4}{\sqrt{p}}$, where p the number of unknown parameters. Efficiency is defined in terms of the effective sample size, which gives the number of independent samples from that posterior distribution that would yield the same Monte Carlo error as the autocorrelated Markov chain.

Algorithm 2 An iid Sample From $\pi(\tilde{\rho}|\rho)$

Input: Ξ_{MCMC}, $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$
for $r = 1, 2, \ldots, \tilde{M}$ **do**
  Sample $\tilde{\rho}^{*(r)}$ from $\pi(\tilde{\rho}|\rho, \tilde{\mathbf{t}}^{(r)}, \boldsymbol{\alpha})$
Output: $\tilde{\boldsymbol{\rho}}^* = [\tilde{\rho}^{*(1)}, \ldots, \tilde{\rho}^{*(\tilde{M})}]^{\mathrm{T}}$

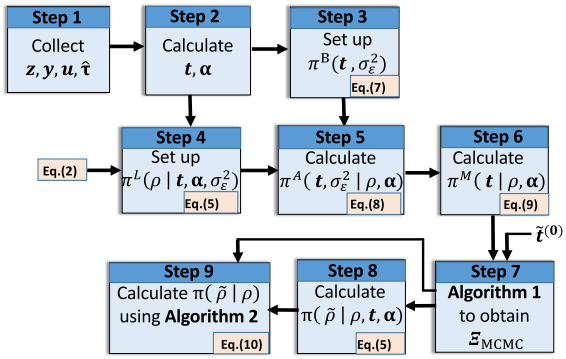

Fig. 2. An illustration of the proposed methodology.

target distribution only after the chain has passed the burn-in time[8] and the effect of the fixed starting value has become so small that it can be ignored.

Once the $\tilde{M}$ samples from the marginal posterior distribution, $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$, are obtained using Algorithm 1, i.e. $\Xi_{\mathrm{MCMC}} = [\tilde{\mathbf{t}}^{(1)\mathrm{T}}, \ldots, \tilde{\mathbf{t}}^{(r)\mathrm{T}}, \ldots, \tilde{\mathbf{t}}^{(\tilde{M})\mathrm{T}}]^{\mathrm{T}}$, we proceed to the calculation of the posterior predictive distribution, $\pi(\tilde{\rho}|\rho)$, by marginalising $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$ over $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$. As $\pi(\tilde{\rho}|\rho)$ might not be available in closed form and it is often easier to sample from this distribution using Monte Carlo methods. Towards this task, each sample in Ξ_{MCMC}, $\tilde{\mathbf{t}}^{(r)}$ is plugged in Equation (12) to sample from $\pi(\tilde{\rho}|\rho, \tilde{\mathbf{t}}^{(r)}, \boldsymbol{\alpha})$ as described in Algorithm 2. Then $\tilde{\rho}^{*(r)}$, $r = 1, \ldots, \tilde{M}$ is an iid sample from the target posterior predictive distribution $\pi(\tilde{\rho}|\rho)$.

Estimating the posterior predictive distribution is very important as this distribution gives the necessary information about unobserved data, in our case the traffic density $\tilde{\rho}$ of a specific time-space region, where there are no CAVs to provide measurements.

Figure 2 summarises the basic idea of the proposed Bayesian TSE methodology, referred to as BTSE in the remainder of this paper.

B. Offline BTSE Algorithm

As mentioned in Section IV-A, the estimation of the mean $\boldsymbol{\mu}(u)$ and covariance matrix $\boldsymbol{\Sigma}(u)$ in Equation (6) is challenging when the average speed u takes continuous values. When the estimation procedure is done offline (Section IV-B) we approximate u by defining N^U discrete speed ranges $[d_{k-1}, d_k)$, $k = 1, \ldots, N^U$, where $d_0 = 0$ and $d_{N^U} = \infty$, associated with different traffic congestion levels. For a specific time-space region the collected data $\mathbf{t}$ and $\boldsymbol{\alpha}$ are included in set $\mathcal{U}_{\tilde{k}}$, where

$$\tilde{k} = \mathrm{arg}_{k=1,\ldots,N^U}\{\tilde{u} \in [d_{k-1}, d_k)\}, \quad \text{and} \tag{13}$$

$$\tilde{u} = \mathrm{median}\left([\mathrm{median}(\mathbf{u}_1), \ldots, \mathrm{median}(\mathbf{u}_{N^C})]^{\mathrm{T}}\right). \tag{14}$$

Parameter $\tilde{k}$ denotes the range index of the median speed $\tilde{u}$ of the median speeds of all CAVs. We use the median, $\tilde{u}$, instead of the average speed, u, as it is a more robust measure against having a small proportion of extremely large or small values. By constructing measurement sets $\mathcal{U}_k$ over all time-windows for a specific space region, we can derive parameters $\boldsymbol{\mu}_k^{off}$ and $\boldsymbol{\Sigma}_k^{off}$ for speed range, $[d_{k-1}, d_k)$, $k = 1, \ldots, N^U$. The proposed algorithm for offline BTSE, referred to as $BTSE_f$ for a specific time-space region is given as follows.

- **Part I:** Initialization.

Step 1: Collect measurements $\mathbf{z} = [\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_c^{\mathrm{T}}, \ldots, \mathbf{z}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_c^{\mathrm{T}}, \ldots, \mathbf{y}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{u} = [\mathbf{u}_1^{\mathrm{T}}, \ldots, \mathbf{u}_c^{\mathrm{T}}, \ldots, \mathbf{u}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$ and $\hat{\boldsymbol{\tau}} = [\hat{\boldsymbol{\tau}}_1^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\tau}}_c^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\tau}}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$.

Step 2: Calculate time spent $\mathbf{t} = [t_1, \ldots, t_{N^C}]^{\mathrm{T}}$ and the area of the time-space region between a CAV and its leading $\boldsymbol{\alpha} = [a_1, \ldots, a_{N^C}]^{\mathrm{T}}$.

Step 3: Set up the prior distribution $\pi^B(\mathbf{t}, \sigma_\varepsilon^2)$ given in (7). Set $\boldsymbol{\mu}(u) = \boldsymbol{\mu}_{\tilde{k}}^{off}$ and $\boldsymbol{\Sigma}(u) = \boldsymbol{\Sigma}_{\tilde{k}}^{off}$ using Equations (13)-(14), and choose hyperparameters β, δ such that the inverse-gamma distribution is an uninformative prior.

- **Part II:** For each time-window $\mathcal{T}_j$, $j \in \mathcal{N}^W$ repeat:

Step 4: Set up the likelihood function $\pi^L(\rho|\mathbf{t}, \boldsymbol{\alpha}, \sigma_\varepsilon^2)$ as shown in Equation (5).

Step 5: Calculate the posterior distribution $\pi^A(\mathbf{t}, \sigma_\varepsilon^2|\rho, \boldsymbol{\alpha})$ using Equation (8).

Step 6: Integrate out the error variance σ_ε^2 to obtain the marginal posterior $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$ given in Equation (9).

Step 7: Take a sample, Ξ_{MCMC}, from the marginal posterior $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$, if not available in closed form, using MCMC based on Algorithm 1.

Step 8: Calculate $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$ using (12), which also follows the likelihood distribution (5).

Step 9: Use sample Ξ_{MCMC} to numerically integrate out parameters $\mathbf{t}$ from $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$ to obtain $\pi(\tilde{\rho}|\rho)$, as shown in Equation (10), using Algorithm 2.

Summarising the above steps, BTSE_f executes Part I once as an initialisation process and iterates over Part II for each individual time-window to result in the estimated traffic densities of a specific time-space region over the entire time-horizon T. This procedure is repeated for all regions of interest. In order to compare estimation results with the PTSE approach [35], we calculate the mean of $\pi(\tilde{\rho}|\rho)$, used as the estimated traffic density $\hat{\rho}^{BTSE_f}$.

C. Online BTSE Algorithm

The proposed algorithm for online BTSE, referred to as $BTSE_o$, differs from BTSE_f by using the ability of CAVs to report data while they travel through a specific time-space region $\mathcal{A}$. The procedure is given as follows.

[8]Burn-in is the procedure of throwing away some iterations at the beginning of an MCMC run.

For each time-window $\mathcal{T}_j$, $j \in \mathcal{N}^W$ repeat:

Step 1: Collect measurements $\mathbf{z} = [\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_c^{\mathrm{T}}, \ldots, \mathbf{z}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_c^{\mathrm{T}}, \ldots, \mathbf{y}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{u} = [\mathbf{u}_1^{\mathrm{T}}, \ldots, \mathbf{u}_c^{\mathrm{T}}, \ldots, \mathbf{u}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$ and $\hat{\boldsymbol{\tau}} = [\hat{\boldsymbol{\tau}}_1^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\tau}}_c^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\tau}}_{N^C}^{\mathrm{T}}]^{\mathrm{T}}$.

Step 2: Calculate time spent $\mathbf{t} = [t_1, \ldots, t_{N^C}]^{\mathrm{T}}$ and the area of the time-space region between a CAV and its leading $\boldsymbol{\alpha} = [a_1, \ldots, a_{N^C}]^{\mathrm{T}}$.

Step 3: Set up the prior distribution $\pi^B(\mathbf{t}, \sigma_\varepsilon^2)$ given in (7). Set $\boldsymbol{\mu}(u) = \boldsymbol{\mu}_{\tilde{k}}^{on}$ and $\boldsymbol{\Sigma}(u) = \boldsymbol{\Sigma}_{\tilde{k}}^{on}$, and choose hyperparameters β, δ such that the inverse-gamma distribution is an uninformative prior.

Step 4: Set up the likelihood function $\pi^L(\rho|\mathbf{t}, \boldsymbol{\alpha}, \sigma_\varepsilon^2)$ as shown in Equation (5).

Step 5: Calculate the posterior distribution $\pi^A(\mathbf{t}, \sigma_\varepsilon^2|\rho, \boldsymbol{\alpha})$ using Equation (8).

Step 6: Integrate out the error variance σ_ε^2 to obtain the marginal posterior $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$ given in Equation (9).

Step 7: Take a sample, $\boldsymbol{\Xi}_{\mathrm{MCMC}}$, from the marginal posterior $\pi^M(\mathbf{t}|\rho, \boldsymbol{\alpha})$, if not available in closed form, using MCMC based on Algorithm 1.

Step 8: Calculate $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$ using (12), which also follows the likelihood distribution (5).

Step 9: Use sample $\boldsymbol{\Xi}_{\mathrm{MCMC}}$ to numerically integrate out parameters $\mathbf{t}$ from $\pi(\tilde{\rho}|\rho, \mathbf{t}, \boldsymbol{\alpha})$ to obtain $\pi(\tilde{\rho}|\rho)$, as shown in Equation (10), using Algorithm 2.

Steps 1 and 2 of Part I are the same with BTSE$_f$. Step 3 is slightly modified in terms of the definition of the mean and variance of $\pi^B(\mathbf{t})$. We consider measurements from the previous and current time-windows, $\mathcal{T}_{j-1}$ and $\mathcal{T}_j$, respectively. Assuming that the congestion level does not change within two time-windows implies that $\boldsymbol{\mu}(u)$ and $\boldsymbol{\Sigma}(u)$ do not depend on u, i.e. $\boldsymbol{\mu}(u) \approx \boldsymbol{\mu}^{on}$ and $\boldsymbol{\Sigma}(u) \approx \boldsymbol{\Sigma}^{on}$. Hence, parameters $\boldsymbol{\mu}^{on}$ and $\boldsymbol{\Sigma}^{on}$ are derived using the collected data $\mathbf{t}$ and $\boldsymbol{\alpha}$ from the previous and current time-windows. The iterative process indicated as Part II in the previous section is no longer needed, hence we proceed to Steps 4-10 as shown in Section IV-B, to calculate the mean of $\pi(\tilde{\rho}|\rho)$, used as the estimated traffic density $\hat{\rho}^{BTSE_o}$. Hence, BTSE$_o$ iterates over Part I for each time-space region. The main difference is that BTSE$_f$ requires the collection of all data for all time-windows to set up the prior distributions and then proceed with the estimation process, whereas BTSE$_o$ utilises only the information provided at the current and previous time-window. As before, in order to compare estimation results with the PTSE approach [35], we calculate the mean of $\pi(\tilde{\rho}|\rho)$, used as the estimated traffic density $\hat{\rho}^{BTSE_o}$.

V. Performance Evaluation

To evaluate the performance of the proposed Bayesian methodology we examine: (i) a simulation study that represents a real network that operates under both free-flow and congested conditions (Section V-A) and (ii) a real-life dataset, the HighD dataset (Section V-B). We utilise the framework for both the offline and online BTSE approaches and compare estimation results with the PTSE approach [35], described in Section III. To obtain the estimation results we consider different CAV penetration rates as well as time-window durations. The metric used to compare the performance of each algorithm is the Mean Absolute Percentage Error (MAPE) between the sth estimated traffic density, $\hat{\rho}_s^{\mathrm{alg}}$, alg $\in$ {PTSE, BTSE$_o$, BTSE$_f$}, and the sth true traffic density ρ_s^{true} given by

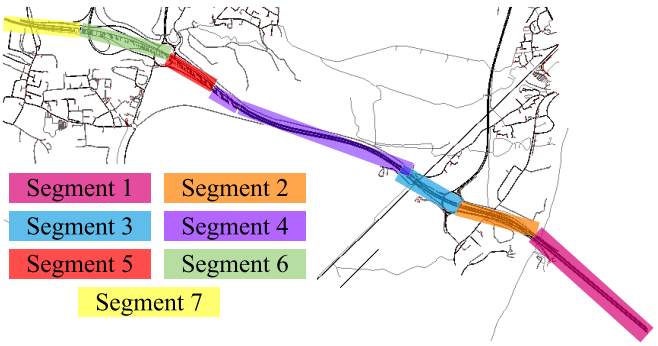

Fig. 3. M25 Highway simulation: SUMO Representation of the M25 highway stretch in London, England.

$$\mathrm{MAPE} = \frac{1}{S}\sum_{s=1}^{S}\frac{|\rho_s^{true} - \hat{\rho}_s^{\mathrm{alg}}|}{\rho_s^{true}}100\%, \tag{15}$$

where S is the total number of estimations. The MAPE is a common measure for evaluating the estimation performance because the error values of the quantity of interest are scaled to percentage units, which makes it easier to interpret.[9] The mean value of the derived probability distribution, $\pi(\tilde{\rho}|\rho)$, is calculated using the iid samples obtained using Algorithm 2 for both BTSE$_f$ and BTSE$_o$, and is used as the estimated traffic density $\hat{\rho}$ in (15). The true value of traffic density is obtained through (2), assuming that all vehicles in the road segment under study are CAVs, hence we have full knowledge of vehicle trajectories.

For both experiments we assume that the model noise ε is a Gaussian random variable with zero mean and unknown variance σ_ε^2, for which we assume $\sigma_\varepsilon^2 \sim \mathrm{IG}(\beta, \delta)$. We set the values of the hyperparameters to $\beta = 0.3$ and $\delta = 0.2$.[10] To eliminate any biases regarding the choice of CAVs we average each result over 10 repetitions and randomly select a different set of CAVs in each experiment to simulate a specific penetration rate. All experiments were executed and algorithms were coded in `Matlab`.

A. Case Study I: M25 Highway Simulation

1) Simulation Setup : For the simulation study we utilise the SUMO microscopic simulator [69]. We simulate traffic of a multi-lane highway road stretch of the M25 motorway in London, England between junctions J_{13} and J_{14}, with the SUMO representation of the network shown in Figure 3.

[9]The mean absolute error (MAE) and root mean squared error (RMSE) were also examined, yielding similar observations to the ones presented in this paper.

[10]The particular values were chosen to ensure an uninformative prior, as well as to eliminate any issues in the posterior inference [68]. A detailed sensitivity analysis indicated that the estimation results are not significantly affected by the values of β and δ.

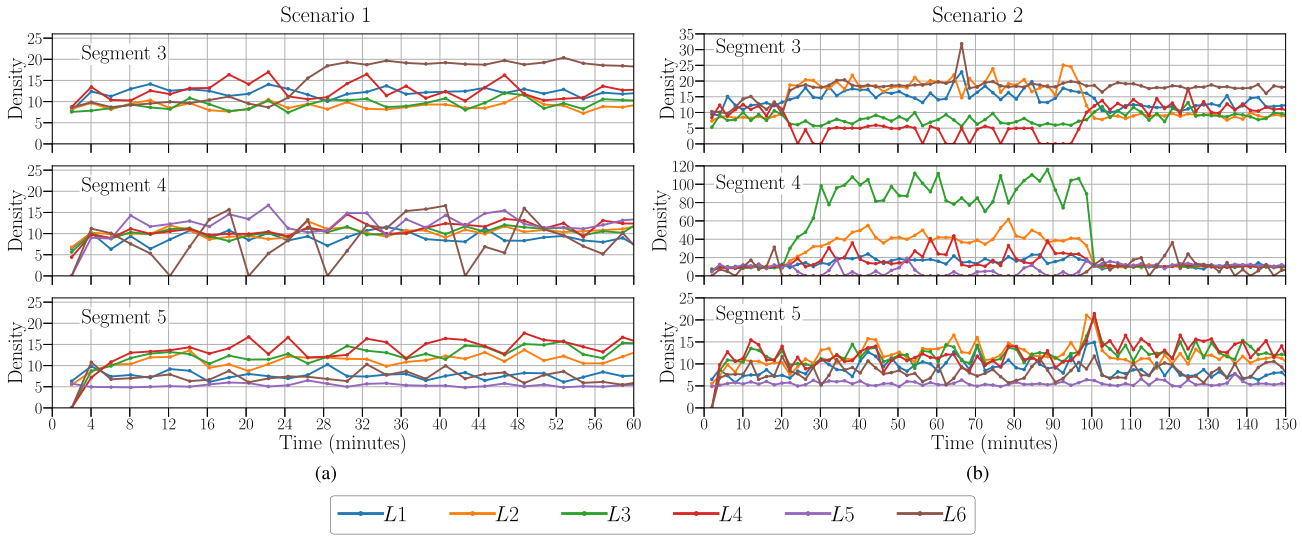

Fig. 4. M25 Highway simulation: Traffic density per lane ($L1 - L6$) for segments 3, 4 and 5 under free-flow and congested conditions, for (a) Scenario 1 and (b) Scenario 2 respectively, when $T^W = 2$ minutes.

TABLE I

M25 HIGHWAY SIMULATION: MAPE (%) CALCULATED FOR BTSE_f, BTSE_o AND PTSE FOR ALL SEGMENTS SHOWN IN FIGURE 3 FOR BOTH SCENARIOS AND PENETRATION RATES (PR) 5%, 10%, 20% AND 30% WHEN $T^W = 2$ MINUTES

Scenario 1

	Segment 1			Segment 2			Segment 3			Segment 4			Segment 5			Segment 6			Segment 7		
PR	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE
5%	11.9	32.6	47.0	10.3	26.9	40.1	14.9	30.8	33.8	13.6	36.6	55.3	11.5	29.1	37.8	15.7	29.7	45.6	14.7	30.1	41.9
10%	11.3	20.7	30.9	10.0	18.3	26.4	14.2	20.4	23.4	12.7	22.4	36.7	10.4	18.5	26.4	14.7	21.6	34.0	14.2	19.9	29.0
20%	11.0	13.4	17.9	9.5	13.1	17.1	14.1	13.8	16.9	12.3	15.3	23.8	10.0	12.2	16.7	12.3	13.3	18.7	14.2	14.4	20.3
30%	10.8	10.3	12.4	9.2	11.2	13.2	14.1	11.3	12.7	11.1	11.9	15.8	9.9	10.1	11.5	12.1	10.0	12.9	14.2	11.1	13.0

Scenario 2

	Segment 1			Segment 2			Segment 3			Segment 4			Segment 5			Segment 6			Segment 7		
PR	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE	BTSE_f	BTSE_o	PTSE
5%	11.2	29.6	42.8	18.9	31.2	41.4	22.1	29.6	40.9	24.4	37.0	50.3	15.7	31.3	43.4	13.6	29.5	42.4	13.6	28.0	39.0
10%	10.1	19.9	29.5	18.1	20.9	31.7	21.6	20.8	28.5	21.8	26.5	36.2	15.0	21.8	30.4	11.8	19.5	28.1	12.7	19.1	27.2
20%	9.4	12.8	16.7	17.5	14.4	18.3	21.3	14.0	17.1	18.7	19.8	24.1	14.3	14.6	18.9	10.9	13.6	17.1	12.6	13.3	16.9
30%	7.5	10.2	12.0	16.9	11.5	13.7	19.1	11.4	12.4	12.2	16.4	17.2	12.3	12.2	13.2	9.6	10.7	11.7	11.3	10.7	11.8

The road stretch is about 8-km long and is separated into seven road segments, with size varying between 0.7-1.6 km, as shown in Figure 3 (segment 1 = 1.6km, segment 2 = 0.6km, segment 3 = 0.75km, segment 4 = 1.5km, segment 5 = 1km, segment 6 = 1.1km, segment 7 = 1km). It consists of two off-ramps (in segments 2 and 5) and two on-ramps (in segments 3 and 7). The number of lanes varies from 4 to 6 and the speed limit is 120 km/h.

Two distinct traffic scenarios are considered:

- **Scenario 1** is an 1-hour traffic simulation representing free-flow traffic conditions. The true traffic density of each lane varies between 0-20 veh/km for all segments.
- **Scenario 2** is a 2.5-hour traffic simulation representing congested traffic conditions. The true traffic density of each lane varies between 0-120 veh/km for all segments. The scenario involves an accident at the downstream boundary of segment 4 which results in closure of lanes L_5, L_6 for 1 hour.

To create traffic within the network, we have considered the Intelligent-Driver Model (IDM) which is one of the build-in car-following models of the SUMO microsimulator. The IDM parameters of individual vehicles are drawn from normal distributions as follows: vehicle length $l_v \sim N(4.9, 0.2)$ [m] conditional on $3.5 \leq l_v \leq 5.5$, maximum speed $u_v^{max} \sim N(28, 3)$ [m/s], acceleration $\theta \sim N(2.5, 0.6)$ [m/s^2], deceleration $\vartheta \sim N(4.5, 0.8)$ [m/s^2], minimum gap distance $d_g \sim N(2.5, 0.4)$ [m], and simulation time-step 0.2 [s]. In the simulation, we consider two types of vehicles: (i) CVs which do not collect any data, and (ii) connected vehicles which measure position $\mathbf{z}_c$, speed $\mathbf{u}_c$ and spacing $\mathbf{y}_c$ every 0.2 seconds (assuming a maximum detection range of 200 m). Based on these measurements, each connected vehicle calculates and communicates to an operation center its speed, $\mathbf{u}_c$, time-spent, t_c, and time-space region, a_c. All results are averaged over ten (10) simulations due to the stochastic nature of SUMO.

The true densities of segments 3, 4 and 5 for Scenarios 1 and 2 are given in Figures 4(a) and 4(b), for $T^W = 2$ minutes. Note that the true density is re-calculated as the time-window changes in order to obtain the MAPE. These represent different segment types found in a highway: on-ramp (segment 3), off-ramp (segment 5) and normal (segment 4). For the BTSE$_f$ we use three speed ranges: [0,40), [40, 80) and [80, ∞) km/h to represent high and moderate congestion and free-flow conditions, respectively. These speed ranges can be subjectively chosen to represent different congestion levels, without affecting the basic steps of the methodology described in Figure 2.

2) Results: We present estimation results for BTSE$_f$, BTSE$_o$ and PTSE algorithms obtained using different penetration rates and time-window durations for the simulation network given in Figure 3 under Scenarios 1 and 2.

Table I illustrates the MAPE (15) of the three estimation approaches (BTSE$_f$, BTSE$_o$ and PTSE) under both scenarios for penetration rates 5%, 10%, 20% and 30%. A different traffic density is estimated for each lane and each segment every $T^W = 2$ minutes. As shown in Table I, the BTSE$_f$ approach results in significantly better MAPE for low penetration rates, i.e. 5% and 10%. For example, the resulting MAPE from BTSE$_f$ is about 2-3 times lower than BTSE$_o$ and 3-4 times lower than PTSE when the penetration rate is 5%. For high penetration rates, the three algorithms yield similar performance. Furthermore, notice that while the performance of BTSE$_o$ and PTSE improves significantly for increasing penetration rate, the corresponding improvement for BTSE$_f$ is negligible. This may occur because the BTSE$_f$ utilises information collected from all time-windows to form the prior in Eq. (7), instead from only $\mathcal{T}_j$ and $\mathcal{T}_{j-1}$ used in the BTSE$_o$. Interestingly, the performance of BTSE$_o$ and PTSE is sometimes marginally better compared to BTSE$_f$ for high penetration rates. This may happen because the information provided from the current and the previous time-windows used in BTSE$_o$ represents better the prior distribution compared to BTSE$_f$ that constructs the prior distribution using data from the entire traffic scenario. In addition, BTSE$_o$ has a better performance compared to PTSE for all segments, especially for low penetration rates. The estimation error of BTSE$_o$ increases from about 11% to 31% as the penetration rate decreases from 30% to 5%, whereas the estimation error of the PTSE increases from about 13% to 45% for the respective rates. For example, let us consider Segment 2. Although the MAPE for BTSE$_o$ and PTSE is relatively close for penetration rate 30%, when the penetration rate decreases to 5%, the MAPE for BTSE$_o$ is about 15% better (26.85% compared to 40.10%).

In addition, the impact of the duration of the time-window T^W is examined in terms of the MAPE for the three algorithms. Results are obtained using varying time-window durations from 2 to 10 minutes, the four different penetration rates of connected vehicles and all segments. Figure 5 depicts the average MAPE for each penetration rate and time-window duration, where the first row of figures shows results for Scenario 1 and the second row for Scenario 2, for the four penetration rates. As shown, there is an overall decrease of the MAPE for all algorithms as the size of the time-window increases, with BTSE$_o$ and PTSE resulting in higher improvement compared to BTSE$_f$. For example, for penetration rate 5%, increasing the time-window duration from 2 to 10 minutes improves the MAPE from about 32% to 15% for BTSE$_o$ and from about 45% to 20% for PTSE, in both scenarios, while the improvement of BTSE$_f$ is around 5%. Additionally, the increase of time-window duration has less impact on the MAPE as the penetration rate increases. For instance, for penetration rate 30% increasing the time-window duration from 2 to 10 minutes yields less than 10% improvement for all approaches and the two scenarios. The impact of the time-window duration is lower for higher penetration rates and the estimation results are more accurate for longer time-windows. These observations are consistent with the fact that more data is collected in longer time-windows, resulting in more accurate estimations. Another reason that the time-window variation might decrease the MAPE is the fact that as the time-windows are increased, the true densities become smoother functions and hence easier to predict. In addition, for higher connected vehicle penetration rates, sufficient data can be provided for smaller time-windows yielding lower MAPE.

The impact of the penetration rate is also examined for all algorithms in both scenarios. In Figure 6 the barplots summarise the average MAPE of each approach using the corresponding penetration rate and all time-window durations from 2 to 10 minutes in each scenario. It is evident that the BTSE$_f$ approach achieves estimation error of about 12% for Scenario 1 and 14% for Scenario 2 using 5% penetration rate. Compared to BTSE$_f$, BTSE$_o$ and PTSE yield about 2 and 3 times higher estimation error for Scenario 1 and about 1.5 and 2 times higher estimation error for Scenario 2, respectively. In order to achieve similar estimation error to BTSE$_f$, BTSE$_o$ requires a 20% and 10% penetration connected vehicle rate for Scenarios 1 and 2, while the corresponding penetration rates for PTSE are 30% and 20%.

In Figure 7 an example of traffic density estimation obtained from each algorithm (PTSE, BTSE$_f$, BTSE$_o$) along with the real traffic density is shown for penetration rates 5% and 30%. These results are obtained for lane 3 of segment 4 under both scenarios, over a 2-minute time-window. As shown, the BTSE methodology provides a PDF whereas the PTSE approach provides a single point estimate of the traffic density. For 30% penetration rate the distributions of BTSE$_f$ and BTSE$_o$ have less variance compared to the ones provided for 5% penetration rate. This means that the traffic density estimation for 5% penetration rate is more 'uncertain' compared to the 30% penetration rate, which is a result of the amount of information utilised in the prior distribution. Both BTSE distributions are 'closer' to the true value of the traffic density compared to the PTSE estimate, resulting in the MAPE difference shown in Table I and Figures 5 and 6.

All the above results were obtained by applying the proposed Bayesian methodology to estimate the traffic density of each lane for a simulation scenario in M25 motorway in both free-flow and congested conditions. The BTSE$_o$ algorithm consistently yields better estimation results compared to PTSE,

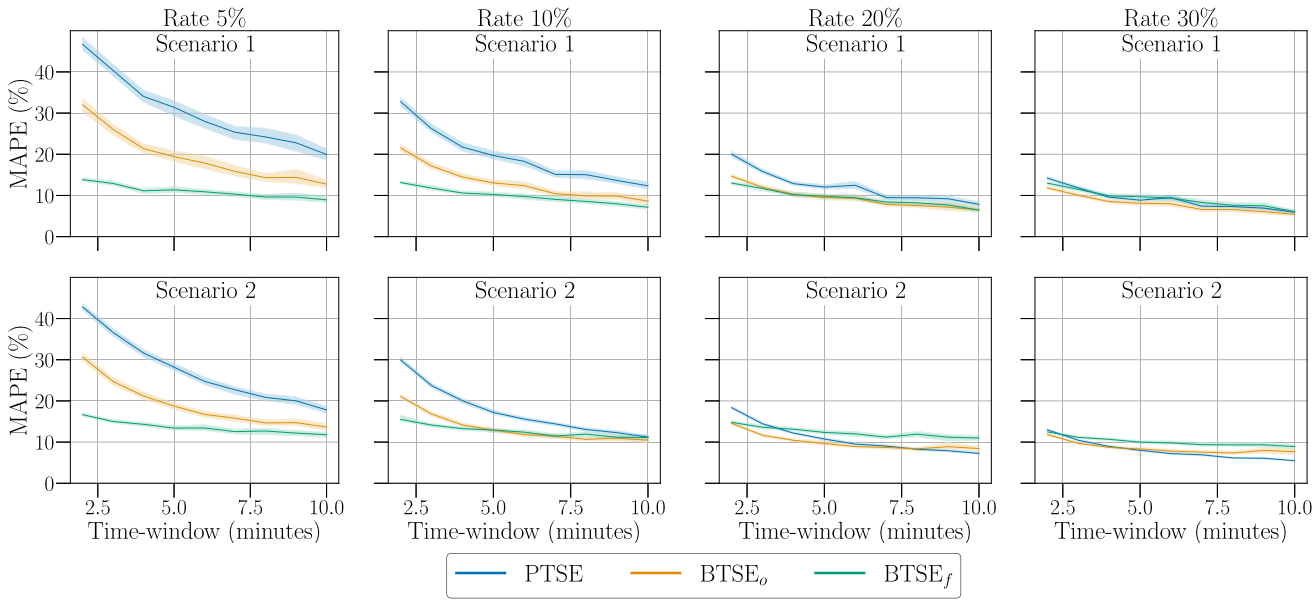

Fig. 5. M25 Highway simulation: Average MAPE for different time-window durations, T^W, for each penetration rate and all segments for Scenario 1 (first row) and Scenario 2 (second row). The three solid lines show the average MAPE for each algorithm and the shaded area around the lines represents the 95% confidence intervals calculated using bootstrapping.

especially for low penetration rates. The BTSE$_f$ algorithm produces significantly better estimations compared to both BTSE$_o$ and PTSE for low penetration rates. The performance of the algorithms is similar for high penetration rates. Next, we examine the performance of BTSE$_f$, BTSE$_o$ and PTSE for a real-life dataset.

B. Case Study II: HighD Real-Life Dataset

1) Dataset Description and Estimation Setup: HighD is an open source dataset consisting of vehicle trajectories recorded at six different locations in highways around Cologne, Germany, using unmanned aerial vehicles (UAV) [59]. The length of each observed road segment varies between 400-420 m. For improved image stability, video recordings were made between the hours 8am and 5pm on days with no wind. The resolution of each video recording is 4K (4096×2160 pixels) at 25 frames per second, with an average video length of 17 minutes. The total recording time is approximately 16.5 hours. The dataset contains information for more than 110,000 vehicles (90,000 passenger vehicles and 20,000 trucks) with a total driven distance of 45,000 km and a total driven time of 447 hours. The dataset consists of information about the frames in which each vehicle appears, as well as its position, velocity, acceleration and the IDs of its surrounding vehicles. Details regarding the time and space headway for each vehicle and each frame are also included. Finally, statistical values such as the minimum, maximum and mean velocity, time and space headway of each vehicle are provided.

We present results for the BTSE$_f$, BTSE$_o$ and PTSE algorithms for traffic density estimation (see Sections IV-A and III). Due to the fact that the length of the videos is around 15 minutes the minimum and maximum time-window duration is set to 1 and 5 minutes, respectively. Four different connected vehicle penetration rates are assumed for the collection of measurements: 5%, 10%, 20% and 30%. Similar to Case Study I, the performance of the proposed methodology is evaluated in terms of the MAPE (15).

2) Results: Figure 8 illustrates the performance of each algorithm evaluated for the four different penetration rates and the results are averaged over all different time-window durations. As a general observation BTSE$_f$ yields the best performance, while BTSE$_o$ is better than PTSE. As the penetration rate of connected vehicles increases the average MAPE decreases with both the BTSE$_f$ and BTSE$_o$ consistently resulting in lower MAPE than the PTSE. Although for a penetration rate of 30% the BTSE$_o$ and BTSE$_f$ yield around 3.5% and 6.5% lower MAPE compared to PTSE, the corresponding performance difference for 5% penetration rate increases to 10.5% and 25.5%, respectively. Interestingly, for 5% and 10% penetration rates, BTSE$_f$ results in about two times lower MAPE compared to PTSE. Moreover, as the penetration rate increases from 5% to 30% the BTSE$_f$ algorithm is improved by around 4%. Note that the estimation performance of BTSE$_o$ and BTSE$_f$ is better than PTSE for 30% penetration rate, in contrast with the corresponding estimation results presented in Case Study I where all algorithms had similar performance.

Figure 9 shows the average MAPE when varying the time-window duration for the four penetration rates. Three main observations can be made. First, all algorithms result in lower error as the time-window duration increases with the BTSE$_f$ and BTSE$_o$ resulting in up to 30% and 10% lower MAPE compared to the PTSE, respectively. Second, it can be seen that BTSE$_f$ provides high-quality estimation results resulting in less than 20% MAPE in almost all cases considered, irrespective of the time-window duration and penetration rate. Third, the rate of improvement of BTSE$_o$ and PTSE is

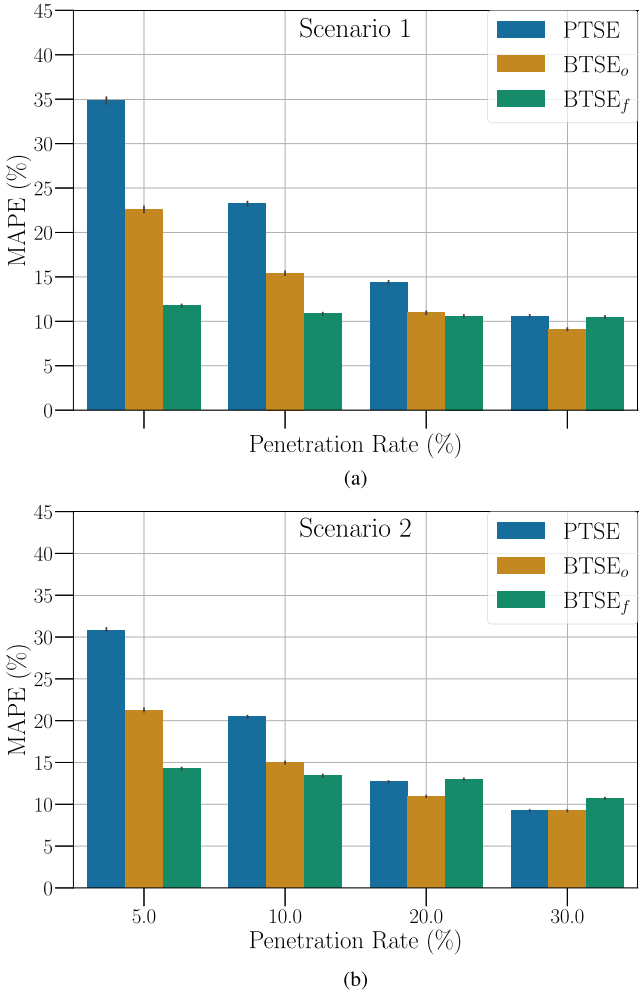

Fig. 6. M25 Highway simulation: Average MAPE for penetration rates 5%, 10%, 20% and 30%. The results are averaged over time-window duration $\{1, 2, \ldots, 9, 10\}$ min. for (a) Scenario 1 and (b) Scenario 2.

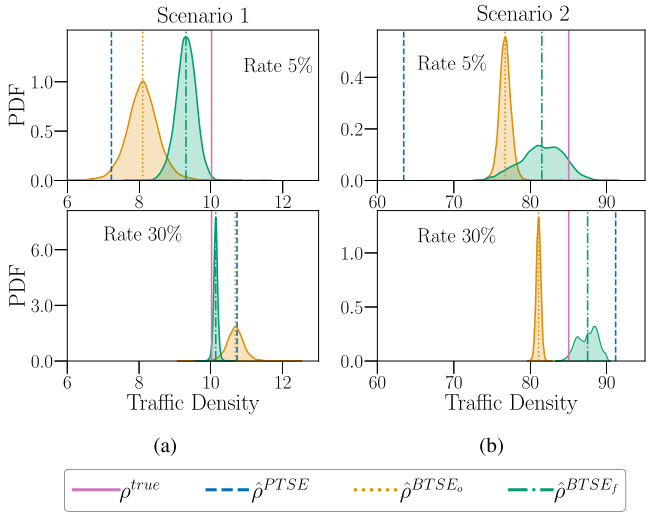

Fig. 7. M25 Highway simulation: Examples of traffic density estimations for a specific time-space region obtained from each methodology using 5% and 30% penetration rates for (a) Scenario 1 and (b) Scenario 2. The probability density function (PDF) of BTSE$_o$ and BTSE$_f$ is depicted with the orange and green shaded areas, respectively.

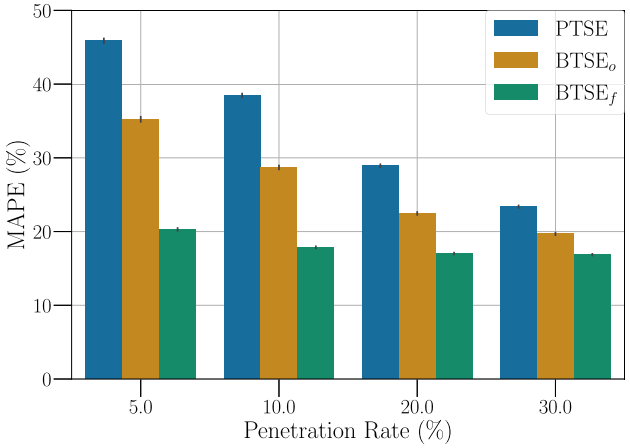

Fig. 8. HighD dataset: Average MAPE for penetration rates 5%, 10%, 20% and 30%. The results are averaged over time-window duration $\{1, 2, 3, 4, 5\}$ min.

similar, with BTSE$_o$ being consistently better by 3% to 10% in terms of the MAPE.

The derived results for Case Study II are in line with the results of Case Study I and support that the proposed Bayesian methodology exhibits more accurate estimation results of the traffic density even for small penetration rates. Finally, note that the estimation results for the PTSE algorithm are inline with the results presented in [35].

VI. Discussion

A. Use Conditions

This work proposes a Bayesian approach that makes use of only xFCD obtained by CAVs or connected vehicles with the capability of measuring the distance to their leading vehicle. The Bayesian framework is a natural way to express uncertainty about an unknown variable, i.e. the traffic density, particularly when the available data is limited and sparse. This method estimates a conditional distribution of the measurements (posterior distribution) by using the available prior information that represents all the available knowledge apart from the data themselves, meaning that no relevant information is omitted from the analysis, and in turn a PDF that represents the traffic density if all vehicles in the highway were CAVs. The most important aspects of this work are that: (i) no traffic model is used, outlasting the need of sufficiently calibrating the model prior to its use as well as avoiding the risk of the model failing to capture real-world traffic due to the ideal assumptions and conditions assumed; (ii) no extensive historical data are required for accurate estimations; (iii) even small penetration rates, i.e. 5% of CAVs deployed in the highway, are sufficient to provide traffic density estimations with small error. Hence, the only requirement for the application of this method in real-life networks is the deployment of a small penetration rate of CAVs in the network of interest that will collect the measurements, which will be transmitted and processed centrally to provide the information needed to follow the steps of the methodology as illustrated in Figure 2 either in an offline or online setup, depending on

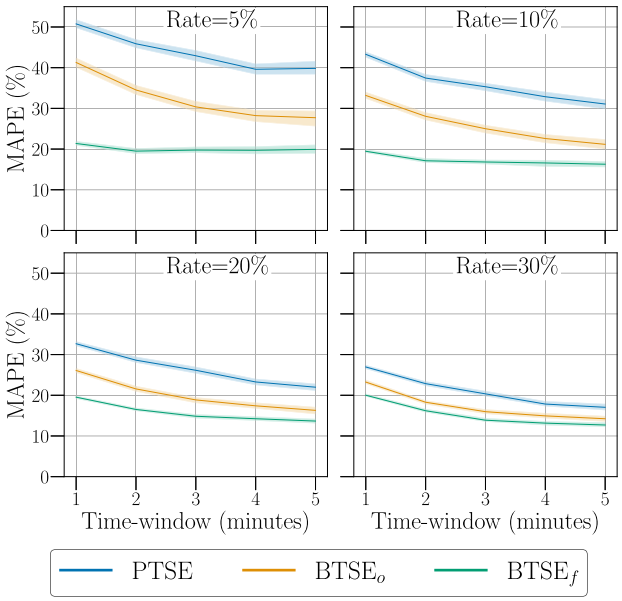

Fig. 9. HighD dataset: Average MAPE for penetration rates 5%, 10%, 20% and 30%. The results are averaged over time-window duration {1, 2, 3, 4, 5} min.

the specifications of the estimation procedure objective of the study. This work has investigated the estimation accuracy for time-windows in the range [2, 10] minutes and road segment lengths in the range [0.4, 1.5] km and penetration rates as small as 5%, showing that the methodology is robust in different time-space combinations.

B. Limitations

As discussed the applicability of this work is straightforward. One issue that might rise is when there are time-space regions with no available measurements, i.e. they are not visited by CAVs and hence no prior information is available. Such an issue might occur under small penetration rates of CAVs, small time-window durations, and short road segments. One approach to resolve it is to consider time-space interpolation techniques for such regions. It is important to note however that zero measurements indicate that the density in such regions is very low, implying that accurate traffic state estimation is not critical.

VII. Conclusion and Future Work

In this paper, we proposed a novel Bayesian methodology, aiming to derive a probability distribution of the traffic density in a specific road segment assuming that we have available measurements acquired by CAVs. We have developed algorithms for both offline and online Bayesian traffic density estimation (BTSE$_f$ and BTSE$_o$, respectively). We compared the BTSE$_f$ and BTSE$_o$ algorithms with an approach previously presented in the literature [35] (PTSE) under two case studies: (a) a simulation scenario in both free-flow and congested conditions and (b) a real-life traffic dataset (highD) [59]. The results of the two case studies are in line and can be summarized as follows. The BTSE$_f$ algorithm is the best among the examined algorithms, yielding up to three and four times better estimation performance compared to BTSE$_o$ and PTSE. The performance of the BTSE$_f$ algorithm is consistent for different penetration rates and time-window durations, yielding excellent results in all examined settings. Comparing the two online algorithms, BTSE$_o$ yields up to 1.5 times lower estimation error compared to PTSE, especially for low penetration rates.

Future plans include the comparison of the proposed Bayesian methodology with other TSE methods in the literature. In terms of the Bayesian methodology we intend to investigate the robustness and dependence of the methodology on the chosen prior distributions, assumed hyperparameters and the effective size of the MCMC algorithm. Furthermore, we intend to use our methodology to estimate traffic flow and speed. Finally, we aim to investigate the correlation between neighbouring time-space regions and integrate macroscopic traffic dynamics to achieve highly accurate network traffic state estimation.

References

[1] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annu. Rev. Control*, vol. 43, pp. 128–151, Jan. 2017.
[2] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 229, no. 1178, pp. 317–345, 1955.
[3] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, 1956.
[4] H. J. Payne, "Model of freeway traffic and control," *Math. Model Public Syst.*, vol. 1, pp. 51–61, Jan. 1971.
[5] G. B. Witham, *Linear Nonlinear Waves*. New York, NY, USA: Wiley, 1974.
[6] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing," *Transp. Res. A, Policy Pract.*, vol. 42, no. 10, pp. 1340–1358, Dec. 2008.
[7] S. Jabari and H. X. Liu, "A stochastic model of traffic flow: Theoretical foundations," *Transp. Res. B, Methodol.*, vol. 46, no. 1, pp. 156–174, Jan. 2012.
[8] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 99–118, May 2016.
[9] Y. Gu, "Bayesian-based traffic state estimation in large-scale networks using big data," Ph.D. thesis, Dept. Civil Environ. Eng., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2017.
[10] F. Zheng, S. E. Jabari, H. X. Liu, and D. C. Lin, "Traffic state estimation using stochastic Lagrangian dynamics," *Transp. Res. B, Methodol.*, vol. 115, pp. 143–165, Sep. 2018.
[11] A. Doucet, N. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York, NY, USA: Springer, 2001.
[12] L. Mihaylova, B. Boel, and A. Hegyi, "Freeway traffic estimation within particle filtering framework," *Automatica*, vol. 43, no. 2, pp. 290–300, Feb. 2007.
[13] M. Hawes, H. M. Amer, and L. Mihaylova, "Traffic state estimation via a particle filter over a reduced measurement space," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–8.
[14] M. Treiber and D. Helbing, "Reconstructing the spatio–temporal traffic dynamics from stationary detector data," *Cooperative Transp. Dyn.*, vol. 1, no. 3, pp. 1–24, 2002.
[15] J. W. C. van Lint and S. P. Hoogendoorn, "A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways," *Comput. Aided Civil Infrastruct. Eng.*, vol. 25, no. 8, pp. 596–612, Nov. 2010.
[16] M. Canaud, L. Mihaylova, N. E. El Faouzi, R. Billot, and J. Sau, "A Bayesian approach to real-time traffic state estimation using a particle PHD filter with appropriate clutter intensity," in *Proc. 92nd Transp. Res. Board Annu. Meeting*, Jan. 2013, p. 13.
[17] T. Neumann, P. L. Bohnke, and L. C. Touko Tcheumadjeu, "Dynamic representation of the fundamental diagram via Bayesian networks for estimating traffic flows from probe vehicle data," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1870–1875.

[18] J. Kim and G. Wang, "Diagnosis and prediction of traffic congestion on urban road networks using Bayesian networks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2595, no. 1, pp. 108–118, Jan. 2016.

[19] D. Ni and J. Leonard, II, "Markov chain Monte Carlo multiple imputation using Bayesian networks for incomplete intelligent transportation systems data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1935, no. 1, pp. 57–67, 2005.

[20] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 89–107, Sep. 2013.

[21] A. Hofleitner, R. Herring, P. Abbeel, and A. Bayen, "Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1679–1693, Dec. 2012.

[22] N. Polson and V. Sokolov, "Bayesian analysis of traffic flow on interstate I-55: The LWR model," *Ann. Appl. Stat.*, vol. 9, no. 4, pp. 1864–1888, 2015.

[23] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.

[24] I. Laña, J. L. Lobo, E. Capecci, J. D. Ser, and N. Kasabov, "Adaptive long-term traffic state estimation with evolving spiking neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 101, pp. 126–144, Apr. 2019.

[25] J. Wang and Q. Shi, "Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory," *Transp. Res. C, Emerg. Technol.*, vol. 27, pp. 219–232, Feb. 2013.

[26] X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, "Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2303–2310, Sep. 2017.

[27] S. E. G. Jabari, D. M. Dilip, D. Lin, and B. Thonnam Thodi, "Learning traffic flow dynamics using random fields," *IEEE Access*, vol. 7, pp. 130566–130577, 2019.

[28] M. Rostami-Shahrbabaki, A. A. Safavi, M. Papageorgiou, P. Setoodeh, and I. Papamichail, "State estimation in urban traffic networks: A two-layer approach," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102616.

[29] A. J. Huang and S. Agarwal, "Physics informed deep learning for traffic state estimation," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[30] R. Shi, Z. Mo, and X. Di, "Physics informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models," in *Proc. 35th AAAI Conf. Artif. Intell.*, vol. 31, 2021, pp. 540–547.

[31] M. Barreau, M. Aguiar, J. Liu, and K. H. Johansson, "Physics-informed learning for identification and state reconstruction of traffic density," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 2653–2658.

[32] R. Shi, Z. Mo, K. Huang, X. Di, and Q. Du, "A physics-informed deep learning paradigm for traffic state and fundamental diagram estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11688–11698, Aug. 2022.

[33] S. Wang, X. Yu, and P. Perdikaris, "When and why PINNs fail to train: A neural tangent kernel perspective," *J. Comput. Phys.*, vol. 449, Jan. 2022, Art. no. 110768.

[34] G. E. Karniadakis et al., "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[35] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transp. Res. C, Emerg. Technol.*, vol. 53, no. 6, pp. 134–150, 2015.

[36] M. Treiber and A. Kesting, "Trajectory and floating-car data," in *Traffic Flow Dynamics*. Cham, Switzerland: Springer, 2013.

[37] W. Huber, M. Ladke, and R. Ogger, "Extended floating-car data for the acquisition of traffic information," in *Proc. 6th World Congr. Intell. Transp. Syst. (ITS)*, 1999, pp. 8–12.

[38] B. D. Greenshields, "A study of traffic capacity," in *Proc. Highway Res. Board*, vol. 14, 1935, pp. 448–477.

[39] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, 2005.

[40] T. Darwish and K. A. Bakar, "Traffic density estimation in vehicular ad hoc networks: A review," *Ad Hoc Netw.*, vol. 24, pp. 337–351, Jan. 2015.

[41] B. Coifman, "Revisiting the empirical fundamental relationship," *Transp. Res. B, Methodol.*, vol. 68, pp. 173–184, Oct. 2014.

[42] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, "Detecting errors and imputing missing data for single-loop surveillance systems," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1855, pp. 160–167, Jan. 2003.

[43] J. C. Herrera, D. B. Work, R. Herring, X. (Jeff) Ban, Q. Jacobsond, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 4, pp. 568–583, Aug. 2010.

[44] J. C. Herrera and A. M. Bayen, "Incorporation of Lagrangian measurements in freeway traffic state estimation," *Transp. Res. B, Methodol.*, vol. 44, no. 4, pp. 460–481, May 2010.

[45] Y. Yuan, J. W. C. van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, "Real-time Lagrangian traffic state estimator for freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 59–70, Mar. 2012.

[46] W. Deng, H. Lei, and X. Zhou, "Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach," *Transp. Res. B, Methodol.*, vol. 57, pp. 132–157, Nov. 2013.

[47] M. R. Wilby, J. J. V. Díaz, A. B. R. Gonzlez, and M. Á. Sotelo, "Lightweight occupancy estimation on freeways using extended floating car data," *J. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 149–163, Apr. 2014.

[48] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3484–3497, Dec. 2016.

[49] M. Fountoulakis, N. Bekiaris-Liberis, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Highway traffic state estimation with mixed connected andconventional vehicles: Microscopic simulation-based testing," *Transp. Res. C, Emerg. Technol.*, vol. 78, pp. 13–33, May 2017.

[50] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation per lane in the presence of connected vehicles," *Transp. Res. B, Methodol.*, vol. 106, pp. 1–28, Dec. 2017.

[51] S. Papadopoulou, C. Roncoli, N. Bekiaris-Liberis, I. Papamichail, and M. Papageorgiou, "Microscopic simulation-based validation of a per-lane traffic state estimation scheme for highways with connected vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 441–452, Jan. 2018.

[52] L. Montero, M. Pacheco, J. Barceló, S. Homoceanu, and J. Casanovas, "Case study on cooperative car data for estimating traffic states in an urban network," *Transp. Res. Rec.*, vol. 2594, no. 1, pp. 127–137, 2016.

[53] T. Seo and T. Kusakabe, "Probe vehicle-based traffic state estimation method with spacing information and conservation law," *Transp. Res. C Emerg. Technol.*, vol. 59, pp. 391–403, Oct. 2015.

[54] T. Seo, T. Kusakabe, and Y. Asakura, "Traffic state estimation with the advanced probe vehicles using data assimilation," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 824–830.

[55] Y. Han and S. Ahn, "Estimation of traffic flow rate with data from connected-automated vehicles using Bayesian inference and deep learning," *Frontiers Future Transp.*, vol. 2, Mar. 2021, Art. no. 644988.

[56] N. Polson and V. Sokolov, "Bayesian particle tracking of traffic flows," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 345–356, Feb. 2018.

[57] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer Series in Statistics). New York, NY, USA: Springer, 1985.

[58] Y. Xu, S. Dass, and T. Maiti, *Bayesian Prediction and Adaptive Sampling Algorithms for Mobile Sensor Networks: Online Environmental Field Reconstruction in Space and Time*. New York, NY, USA: Springer, 2015.

[59] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2118–2125.

[60] L. Edie, "Discussion of traffic stream measurements and definitions," in *Proc. 2nd Int. Symp. Theory Traffic Flow*, 1963, pp. 139–154.

[61] A. O'Hagan and J. Forster, *Bayesian Inference* (Kendall's Advanced Theory of Statistics). London, U.K.: Oxford Univ. Press, 2004.

[62] S. Banerjee, B. Carlin, and A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2004.

[63] A. Gelman, W. R. Gilks, and G. O. Roberts, "Weak convergence and optimal scaling of random walk metropolis algorithms," *Ann. Appl. Probab.*, vol. 7, no. 1, pp. 110–120, Feb. 1997.

[64] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2013.

[65] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. Boca Raton, FL, USA: CRC Press, 1996.

[66] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
[67] W. K. Hastings, "Monte Carlo sampling using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, Dec. 1970.
[68] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian Anal.*, vol. 1, no. 3, pp. 515–534, Sep. 2006.
[69] P. A. Lopez et al., "Microscopic traffic simulation using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2575–2582.

Victor Kyriacou received the B.Sc. degree in computer engineering from the Department of Electrical and Computer Engineering, University of Cyprus, in 2020. He is currently pursuing the master's degree in artificial intelligence with the University of Amsterdam. He was a Research Assistant at the KIOS Research and Innovation Center of Excellence, University of Cyprus, until 2021. His research focused on statistical learning techniques for data analysis and estimation applications on intelligent transportation systems. He received the Best Thesis Award on his final year titled "Probabilistic Traffic State Estimation using Spacing Measurements".

Yiolanda Englezou (Member, IEEE) received the degree in applied mathematics and physical sciences from the National Technical University of Athens in 2014, where she studied mathematics, physics and computer science, specialising in the areas of applied mathematics and statistics, and the Ph.D. degree in statistics from the University of Southampton in 2018. She has been a Post-Doctoral Researcher with the KIOS Research and Innovation Center of Excellence, University of Cyprus, since 2018. The work was funded by the U.K. Engineering and Physical Sciences Research Council and, for a time, the U.K. Atomic Weapon Establishment. Her research interests include Bayesian inference techniques, statistical and machine learning and estimation applications with emphasis on intelligent systems with impact in real life, and design of experiments, focusing on the development of methods for designing experiments for the calibration of physical and computational models. In 2020, she received the Prestigious Marie Sklodowska Curie (MSCA) Widening Fellowship to work on the project "Bayesian Intelligent Transportation Systems (BITS)".

Christos G. Panayiotou (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, in 1994 and 1999, respectively, and the M.B.A. degree from the Isenberg School of Management, Aforementioned University, in 1999. He was a Research Associate at the Center for Information and System Engineering (CISE) and the Manufacturing Engineering Department, Boston University, from 1999 to 2002. In 2002, he joined the University of Cyprus (UCY), where he is a Professor with the Electrical and Computer Engineering (ECE) Department. He is also the Deputy Director of the KIOS Research and Innovation Center of Excellence for which he is also a Founding Member. He has published more than 275 papers in international refereed journals and conferences. His research interests include modeling, control, optimization and performance evaluation of discrete event and hybrid systems, intelligent transportation systems, cyber-physical systems, event detection and localization, fault diagnosis, wireless, ad hoc and sensor networks, smart camera networks, resource allocation, and intelligent buildings. He was a recipient of the 2014 Best Paper Award for the journal *Building and Environment* (Elsevier). He held several positions in organizing committees and technical program committees of numerous international conferences, including the General Chair of the 23rd European Working Group on Transportation (EWGT2020) and the General Co-Chair of the 2018 European Control Conference (ECC2018). He has also served as the Chair of various subcommittees of the Education Committee of the IEEE Computational Intelligence Society. From 2016 to 2020, he was an Associate Editor of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY. He is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the Conference Editorial Board of the IEEE Control Systems Society, the Journal of *Discrete Event Dynamical Systems*, and the *European Journal of Control*.

Stelios Timotheou (Senior Member, IEEE) received the Dipl.-Ing. degree (summa cum laude) in electrical and computer engineering from the National Technical University of Athens in 2005, and the M.Sc. degree (Hons.) in communications and signal processing and the Ph.D. degree in intelligent systems and networks from the Department of Electrical and Electronic Engineering, Imperial College London, in 2006 and 2010, respectively. He was a Research Associate at KIOS, a Visiting Lecturer at the Department of Electrical and Computer Engineering, University of Cyprus, and a Post-Doctoral Researcher at the Computer Laboratory, University of Cambridge. He is an Assistant Professor with the Department of Electrical and Computer Engineering and a Faculty Member with the KIOS Research and Innovation Center of Excellence, University of Cyprus. His research interests include monitoring, control and optimization of critical infrastructure systems, with emphasis on intelligent transportation systems and communication systems. He was a recipient of the 2017 Cyprus Young Researcher in Physical Sciences and Engineering Award by the Cyprus Research Promotion Foundation. He is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.