

# A non-parametric Bayesian framework for traffic-state estimation at signalized intersections

Junchen Jin<sup>a,\*</sup>, Xiaoliang Ma<sup>a</sup>

<sup>a</sup>*System Simulation & Control, Department of Transport Science,  
KTH Royal Institute of Technology, Teknikringen 10, Stockholm 10044, Sweden.*

---

## Abstract

An accurate and practical traffic-state estimation (TSE) method for signalized intersections plays an important role for real-time operations to facilitate efficient traffic management. This paper presents a generalized modeling framework for estimating traffic states at signalized intersections. The framework is non-parametric and data-driven, without the requirement on explicit models of traffic. Additionally, in principle, any type of data source together with any type of signal controller can be incorporated with the proposed framework. The Bayesian filter (BF) approach is the core of the framework and introduces a recursive state estimation process. The required transition and measurement models of the BFs are trained using Gaussian process (GP) regression models with respect to a set of historical data. A Gaussian process model uses kernel functions to describe the proximity among data points, and the hyper-parameters adopted in the GP model are optimized according to the training data. In addition to the detailed derivation of the integration of BFs and GP regression models, an algorithm based on the extended Kalman filter is presented for real-time traffic estimation. The effectiveness of the proposed framework is demonstrated through several numerical experiments using data generated in microscopic traffic simulations. Both fixed-location data (i.e., loop detector) and mobile data (i.e., connected vehicle) are examined with the framework. As a result, the method performs well for the tested traffic conditions. In particular, the estimator provides a competitive estimation accuracy merely using the position information of a small portion of vehicles at the intersection. The approach is suitable for a short-term estimation requirement, which is normally a challenging task in traffic control and operations.

*Key words:* Traffic state estimation, data-driven model, non-parametric framework, Bayesian filters, Gaussian process regression.

---

## 1. Introduction

### 1.1. Background

Traffic-state estimation (TSE) refers to the process of inferring traffic state variables using partially observed and noisy traffic data that are obtained from deployed surveillance technologies. An accurate and practical TSE method plays an important role in both traffic planning and operations of an indispensable component of urban networks, signalized intersections, to mitigate traffic congestion efficiently. In particular, estimating traffic states is of significant importance in measuring the performance of signalized intersections and further optimizing signal timing plans for conventional controllers [16]. Furthermore, one fundamental aspect of most emerging traffic adaptive signal control systems is to understand the evolving process of traffic states [14]. Normally, the more precise and advanced the signal control is, the more accurate and frequent traffic states are required for the system.

Generally speaking, TSE approaches are grouped into two categories, namely those that are model-driven approaches and other that are data-driven approaches, according to the assumptions and input data they each rely upon.

---

\*Corresponding author. Tel: +46 87908499; Email: junchen@kth.se.

In short, a model-driven TSE approach relies on physics models of traffic system, characterized by empirical relationships and necessitating careful model selection and calibration processes. In certain cases, checking the validity of a model or calibrating it requires a large amount of data. In terms of data-driven TSEs, historical data, taking into account various traffic conditions, is essential otherwise the methods may fail if irregular events have occurred. Moreover, the computational cost for training and learning can be rather high. Along with the increasing development of data and sensing technologies, data-driven models appear to have received more attention recently. The following subsection reviews the existing TSE approaches and discusses their various advantages and disadvantages.

### *1.2. Relevant studies*

To date, numerous TSE approaches with various features on highways have been proposed, such as Wang and Papageorgiou [29]. These have all relied heavily upon both the availability of data and highway traffic flow models. For TSE of highway networks, a comprehensive review is presented in Seo et al. [21]. Therein, the authors summarize how the Lighthill-Whitham-Richards (LWR) model is widely employed and Kalman filter methods (or other analogous techniques) are leveraged for model-based TSEs of highways.

For signalized intersections, the TSE problem is more challenging based on the high complexity of the traffic modeling, compared to the situation on highways. In literature, intersection-wide traffic states are estimated based on either fixed-location sensors (e.g., inductive loop detectors and camera-based sensors) or mobile sensors (e.g., global position systems (GPS), cellular phones, bluetooth, connected vehicles, and other tracking devices) [6].

On one hand, when fixed-location sensor data is adopted, TSE methods can generally be categorized by input-output models and shockwave models. Input-output models estimate traffic states according to the conservation relation between the cumulative traffic input and output (i.e., arrival and departure process of vehicles) to a signalized link. These types of models have a simplistic conceptual nature when describing the traffic queuing processes. An example study is proposed by Vigos et al. [27] who employed a Kalman filter algorithm to produce reliable estimates of the number of vehicles. Their approach is enabled with the help of measurements provided by at least two upstream inductive loop detectors, which were typically not in line with practical cases. Later, the method was simplified to allow for TSEs on the basis of one single detector commonly available if a vehicle-actuated signal controller or an adaptive signal controller is deployed [26].

Nevertheless, Skabardonis and Geroliminis [22] and Liu et al. [13] reported that input-output models are limited due to the inability of capturing traffic states when vehicle positions exceed beyond detectors. The shockwave models succeeded in addressing this issue. Specifically, Liu et al. [13] identified a long queue by searching for break points using the detection information from a stop-line detector and an upstream detector. Then, a model was generated to reconstruct the shockwave propagation to recover the evolving process of traffic states beyond the upstream detector based on LWR theory. The shockwave models are not always successful because they require explicit vehicle-arrival rates that cannot be always captured in many cases. For instance, the vehicle arrival is equal to zero under the congested situations when the upstream detector is constantly occupied.

On the other hand, mobile data is complementary to fixed-location sensor data, which is largely involved in TSE, since the era of the so-called “Internet of Things”, and has shown great potential for real-time TSEs at signalized intersections [2, 3, 7, 4, 6, 19, 24]. Specifically, connected vehicles provide superb opportunities for reducing the dependency on conventional vehicle detectors for TSE at signalized intersections.

Among the TSE studies using mobile data, [4] proposed a sampling-based statistical model with a focus on TSE from the location and time information of probe vehicles at an isolated intersection in unsaturated and oversaturated conditions, respectively. The derived models were validated by comparing estimated states with actual states generated from a microscopic simulator, VISSIM. Another characteristic study was conducted by Hofleitner et al. [7], whereby they presented a hybrid modeling framework for estimating and predicting arterial traffic states with streaming GPS probe data. The approach is built on well-established traffic flow models through signalized intersections, and then combined with a machine learning framework to learn the static parameters utilized for the traffic models.

While many studies in literature were devoted to link-based TSE with regards to intersections controlled by fixed-time traffic signal, group-based and lane-based control approaches provided benefits due to more flexible phase structures and sequences for both conventional traffic signal systems [e.g., 9, 15] and adaptive signal systems [8]. Lane-based traffic states are one of the most important inputs for enabling the functionalities of group-based and lane-based controllers. Lee et al. [12] and Zhan et al. [30] pioneered the analysis of traffic states in each lane served at an isolated intersection to support lane-based and group-based signal controllers.

To the best of our knowledge, few of the proposed TSE approaches are designated for intelligent and traffic-adaptive signal controllers. There is still a need to develop robust short-term TSE approaches and enable effective decision-making processes for adaptive signal control systems. Viti and Van Zuylen [28] and Tiaprasert et al. [24] are two rare examples, where they each developed models to estimate number of queued vehicles at an isolated intersection operated by vehicle-actuated signals.

### 1.3. Objectives

Although TSE approaches for signalized intersections have been widely studied in different contexts, most models require prior knowledge of data type, model structure and model parameters. To this end, this study posits existing works in the machine learning field within a generalized framework for real-time TSE at signalized intersections with the following properties:

- the design of a framework that is compatible with both conventional and adaptive signal control systems, and lane-based states can be generated for supporting group- or lane-based signal controllers;
- the developed framework is not restricted by a certain data type, meaning that both fixed-location and mobile data can be adopted;
- the embedded models are non-parametric such that prior knowledge for setting the parameters is not needed.

The rest of the paper is organized as follows. The proposed non-parametric modeling framework is described in section 2. Section 3 focuses on a practical algorithm enabled by extended Kalman filter. Numerical experiments are carried out and the results and analyses using fixed-location data or connected vehicle data are found in section 4. Finally, the paper is concluded and research perspectives are addressed in section 5.

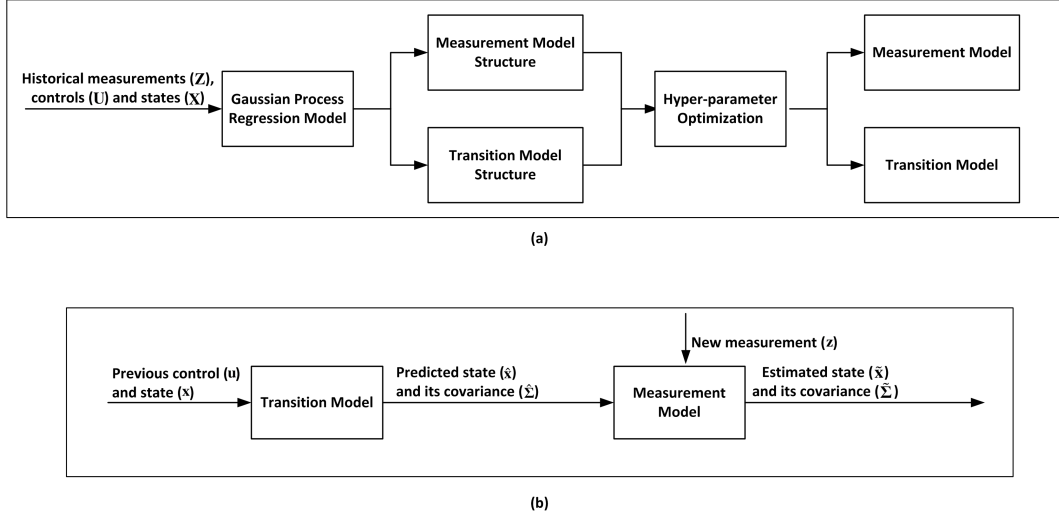
## 2. Non-parametric Bayesian framework for TSE

The Bayesian filter (BF) approach is a commonly used algorithm for estimation when new measurements continuously arrive over time [23]. The algorithm features two operating steps: a prediction step and an update step. During the prediction step, a transition model is utilized to predict the state forward when previous state and control are provided. The predicted state is updated in light of new measurements with a measurement model at the update step. The transition and measurement models are usually parametric models in nature such that the parameters and noise components of them can be optimized or manually tuned according to the provided training data. Accurate parametric models are difficult to obtain, although the basic physics behind to formulate the models is easy to understand. The estimation capabilities may be limited because the parametric models ignore the variations of parameters under different situations if a system having complex relationships between its various elements is incorporated.

A Gaussian Process (GP) is a powerful and non-parametric tool for learning regression functions from sample data [18]. Key advantages of GPs include the ability to provide interval estimates over the state while the size of the interval determines the uncertainty level if the estimated state is equal to the mode value of the obtained distribution [20]. A study by Ko and Fox [10] demonstrated that GP regression models are well-suited for modeling the basic components (i.e., transition and measurement models) of BFs. They provided information regarding how GP regression models are integrated into different versions of BFs.

The overall framework of TSEs using a BF integrated within a GP regression approach, or BFGP for short, is depicted in Fig. 1. In the figure, the framework is partitioned into off-line training (upper block) and on-line estimation (lower block) processes. At the beginning of the off-line training, historical data including measurements ( $Z$ ), controls ( $U$ ) and states ( $X$ ) are inserted as the inputs into the GP models. The measurement and transition models are finally identified by performing a hyper-parameter optimization process for the pre-defined model structures based on the GP regression approach.

During the on-line estimation process, the transition model employs previous control ( $\mathbf{u}$ ) and state ( $\mathbf{x}$ ) to obtain a predicted state ( $\hat{\mathbf{x}}$ ) together with its covariance matrix ( $\hat{\Sigma}$ ) at each estimation step. Using the measurement model, the predicted state and covariance is transferred to the estimated state ( $\tilde{\mathbf{x}}$ ) and covariance ( $\tilde{\Sigma}$ ) by incorporating a newly received measurement ( $\mathbf{z}$ ).



**Fig. 1.** The overall framework of the state estimation using the BFGP with respect to the training process in (a) and on-line estimation process in (b).

### 2.1. Recursive state estimation

The recursive state estimation problem is defined by determining a most likely state at a certain time given all past measurements and control inputs. Hence, the probabilistic law characterizes the state evolution, which might be arrived at by a probability distribution conditioned on the measurements and controls. Such a probability distribution is a posterior probability over the state variables conditioned on the available data, which is abbreviated as a belief distribution. Suppose that the system starts from the initial state,  $\mathbf{x}_0$ , and carries out the initial control,  $\mathbf{u}_0$ , and the first measurement vector is denoted as  $\mathbf{z}_1$ . The belief distribution over a state variable at time  $t$ ,  $\mathbf{x}_t$ , is denoted by  $\text{bel}(\mathbf{x}_t)$ , i.e.,

$$\text{bel}(\mathbf{x}_t) = P(\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{0:t-1}). \quad (1)$$

Then, the estimated state, interpreted as the most likely state, is obtained by

$$\tilde{\mathbf{x}}_t = \arg \max_{\mathbf{x}_t} \text{bel}(\mathbf{x}_t). \quad (2)$$

According to the Markov assumption applied in the BF, the conditional probability distribution of the next state depends only upon the current state and control instead of the sequence of events that preceded such a state-control pair. Such an evolving process for states, controls and measurements is referred as a Markov process model represented by the following equations:

$$P(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1}) = P(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \quad (3)$$

$$P(\mathbf{z}_t | \mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1}) = P(\mathbf{z}_t | \mathbf{x}_t). \quad (4)$$

Here,  $P(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1})$  and  $P(\mathbf{z}_t | \mathbf{x}_t)$ , respectively, represent the state transition probability and measurement probability. Therefore, the belief function can be recursively computed by the following equations (see A for the derivation):

$$\text{bel}(\mathbf{x}_t) = \frac{P(\mathbf{z}_t | \mathbf{x}_t) \overline{\text{bel}}(\mathbf{x}_t)}{\int P(\mathbf{z}_t | \mathbf{x}_t) \overline{\text{bel}}(\mathbf{x}_t) d\mathbf{x}_t}, \quad (5)$$

$$\begin{aligned} \overline{\text{bel}}(\mathbf{x}_t) &= P(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1}) \\ &= \int P(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}. \end{aligned} \quad (6)$$

where  $\overline{\text{bel}}(\mathbf{x}_t)$  represents the initial belief distribution. The initial belief reflects the fact that  $\overline{\text{bel}}(\mathbf{x}_t)$  is a prediction of the current state at time  $t$  based on the posterior distribution of the previous state,  $\text{bel}(\mathbf{x}_{t-1})$ , before incorporating the measurement. Estimating  $\text{bel}(\mathbf{x}_t)$  from  $\overline{\text{bel}}(\mathbf{x}_t)$  is often known as measurement correction or update. Therefore, the concrete implementation of BFs requires three necessary probability distributions. These are the state transition probability,  $P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ , measurement probability,  $P(\mathbf{z}_t|\mathbf{x}_t)$ , and initial belief,  $P(\mathbf{x}_0)$ .

The state transition model is typically described by

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \boldsymbol{\epsilon}, \quad (7)$$

where  $g(\cdot, \cdot)$  denotes a mapping between the previous state-control pair,  $(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ , and the current state,  $\mathbf{x}_t$ .  $\boldsymbol{\epsilon}$  is assumed to follow a white Gaussian process with a covariance matrix  $\boldsymbol{\Sigma}^{tran}$  and a zero mean value, so that the conditional distribution of the current state,  $\mathbf{x}_t$ , is

$$P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) = \mathcal{N}(g(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \boldsymbol{\Sigma}^{tran}). \quad (8)$$

Likewise, assuming that  $\boldsymbol{\zeta}$  also follows a white Gaussian noise process.  $\boldsymbol{\Sigma}^{meas}$  describes the covariance matrix of the measurement model. The measurement model is given by

$$\mathbf{z}_t = h(\mathbf{x}_t) + \boldsymbol{\zeta}. \quad (9)$$

and the measurement distribution is

$$P(\mathbf{z}_t|\mathbf{x}_t) = \mathcal{N}(h(\mathbf{x}_t), \boldsymbol{\Sigma}^{meas}), \quad (10)$$

where  $h(\cdot)$  indicates a mapping between the current state,  $\mathbf{x}_t$ , and the current measurement,  $\mathbf{z}_t$ .

## 2.2. GP regression model

GP has been considered as a powerful framework for regression model in machine learning. This section begins with a general example to introduce the basic principle of GP regression models, followed by a description of predictive modeling. Let us assume a set of training data  $\mathcal{D}$  with the size  $N$ . The data can be represented by

$$\mathcal{D} = \langle \mathbf{X}, \mathbf{y} \rangle, \quad (11)$$

where  $\mathbf{X}$  is an  $N \times M$  matrix, and each row is a  $1 \times M$  vector referring to an instant of input data.  $\mathbf{y}$  represents an  $N \times 1$  matrix containing scalar training outputs. The training data is represented by

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N]^T, \quad (12)$$

$$\mathbf{y} = [y_1, \dots, y_i, \dots, y_N]^T, \quad i = 1, 2, \dots, N. \quad (13)$$

Here,  $\mathbf{x}_i$  and  $y_i$ , respectively, are a column vector and a scalar value of the  $i^{\text{th}}$  training data set.

To apply GP model to the regression problem, each output value is assumed to be drawn from noisy processes, which can be generally represented by

$$y_i = f(\mathbf{x}_i) + \epsilon_i = f_i + \epsilon_i, \quad i = 1, 2, \dots, N, \quad (14)$$

where  $f(\mathbf{x}_i)$  refers to a function mapping between the input,  $\mathbf{x}_i$ , and an instantiation,  $f_i$ .  $\epsilon_i$  obeys a Gaussian distribution with zero-mean and a variance of  $\eta_i$ . Therefore, the probability distribution over an output variable,  $y_i$ , is

$$P(y_i|f_i, \eta_i) = \mathcal{N}(f_i, \eta_i), \quad i = 1, 2, \dots, N. \quad (15)$$

It is assumed that the noise variables, i.e.  $\epsilon_1 \dots \epsilon_N$ , for the corresponding data points are independent of each other such that

$$P(\mathbf{y}|\mathbf{f}, \boldsymbol{\eta}) = \prod_{i=1}^N P(y_i|f_i, \eta_i) = \mathcal{N}(\mathbf{f}, \mathbf{I}^N \boldsymbol{\eta}), \quad (16)$$

where  $\mathbf{I}^N$  denotes an  $N \times N$  unit matrix, and

$$\mathbf{f} = [f_1, \dots, f_i, \dots, f_N]^T, \quad (17)$$

$$\boldsymbol{\eta} = [\eta_1, \dots, \eta_i, \dots, \eta_N]^T, \quad i = 1, 2, \dots, N, \quad (18)$$

In addition, each instantiation,  $f_i, i = 1, 2, \dots, N$ , can be theoretically computed by any functions. An infinite collection of instantiation variables is assumed to follow a GP. According to the definition of GP, any subset variables of the collection (denoted as a vector  $\mathbf{f}$ ) is jointly Gaussian distributed conditioned on a set of input data,  $\mathbf{X}$ , and kernel parameters,  $\boldsymbol{\theta}$ , i.e.,

$$P(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(0, \mathbf{K}), \quad (19)$$

where  $\mathbf{K}$  denotes the covariance matrix determined by a kernel function of the input data, and each of its entries are represented by

$$K_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}), \quad i = 1, 2, \dots, N, j = 1, 2, \dots, N, \quad (20)$$

where  $K(\cdot)$  is a kernel function signifying the level of proximity among data points. Specifically, two instantiation variables (e.g.,  $f_i$  and  $f_j$ ) are more strongly correlated associated with the corresponding data points (e.g.,  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ) that are more similar to each other. The notion of proximity depends on the variations among applications.

Since the marginal distribution of the output variable  $\mathbf{y}$  is conditioned on the input values and hyper-parameters ( $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ ), i.e.,

$$P(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \int P(\mathbf{y}|\mathbf{f}, \boldsymbol{\theta})P(\mathbf{f}|\mathbf{X}, \boldsymbol{\eta})d\mathbf{f}, \quad (21)$$

this can be further represented as a multivariate Gaussian distribution

$$P(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \mathcal{N}(0, \mathbf{C}), \quad (22)$$

where  $\mathbf{C}$  denotes the covariance matrix. Each element of  $\mathbf{C}$  is computed by

$$C_{i,j} = C(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}, \eta_i) = K(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}) + \eta_i \delta_{ij}, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, N, \quad (23)$$

where  $\delta_{ij}$  results from the Kronecker delta function.

In order to make predictions on the output variable for new inputs, let us consider that new input data  $\mathbf{x}_{N+1}$  is inserted into the training data of  $N$  samples, and the output variable of the new data is denoted by  $y_{N+1}$ . According to Equation 22, the joint distribution of the output variables, including the new variable conditioned on the input data points and hyper-parameters is a multivariate Gaussian distribution, given by

$$P(\mathbf{y}^{new}|\mathbf{X}^{new}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \mathcal{N}(0, \mathbf{C}^{new}), \quad (24)$$

where  $\mathbf{X}^{new}$ ,  $\mathbf{y}^{new}$ , and  $\mathbf{C}^{new}$  represent the input variables, output variables, and covariance matrix when including the new input data, i.e.,

$$\mathbf{X}^{new} = \begin{bmatrix} \mathbf{X} \\ \mathbf{x}_{N+1}^T \end{bmatrix}; \mathbf{y}^{new} = \begin{bmatrix} \mathbf{y} \\ y_{N+1} \end{bmatrix}; \mathbf{C}^{new} = \begin{bmatrix} \mathbf{C} & \mathbf{c}_{N+1} \\ \mathbf{c}_{N+1}^T & C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \eta_{N+1}) \end{bmatrix}. \quad (25)$$

$\mathbf{c}_{N+1}$  is an  $N \times 1$  matrix defined by

$$\mathbf{c}_{N+1} = [C(\mathbf{x}_1, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \eta_1), C(\mathbf{x}_2, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \eta_2), \dots, C(\mathbf{x}_N, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \eta_{N+1})]^T. \quad (26)$$

For such a multivariate Gaussian distribution, if two sets of variables are jointly Gaussian, the distribution of one set of variables conditioned on the other fulfils a Gaussian distribution. Therefore, the conditional distribution of the new output variable  $y_{N+1}$  is also Gaussian, i.e.,

$$P(y_{N+1}|\mathbf{X}, \mathbf{x}_{N+1}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \mathcal{N}(v(\mathbf{x}_{N+1}, \mathcal{D}), \Gamma(\mathbf{x}_{N+1}, \mathcal{D})), \quad (27)$$

where  $v(\mathbf{x}_{N+1}, \mathcal{D})$  and  $\Gamma(\mathbf{x}_{N+1}, \mathcal{D})$  refer to mean and variance functions, as derived by

$$v(\mathbf{x}_{N+1}, \mathcal{D}) = \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{y}, \quad (28)$$

$$\Gamma(\mathbf{x}_{N+1}, \mathcal{D}) = C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \eta_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}. \quad (29)$$

For the detailed derivation of the mean and variance functions, readers can refer to B.

**Table 1**  
Summary of the training data

Notation	Description	Composition	Dimensions
$\mathbf{X}^{tran}$	A matrix containing all the input data points for the transition model	$[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N-1}]^T$	$(N-1) \times (M^x + M^u)$
$\mathbf{Y}^{tran}$	A matrix containing all the output data points for the transition model	$[\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{N-1}]^T$	$(N-1) \times M^x$
$\mathbf{X}^{meas}$	A matrix containing all the input data points for the measurement model	$[\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_N]^T$	$(N-1) \times M^x$
$\mathbf{Y}^{meas}$	A matrix containing all the output data points for the measurement model	$[\mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_N]^T$	$(N-1) \times M^z$

### 2.3. The GP-based transition and measurement models

This section illustrates how the GP regression model is integrated into a BF framework. Firstly, let us assume that  $M^x$ ,  $M^u$ , and  $M^z$  represent the number of elements in the state, control, and measurement vectors, and the size of the data points for three vectors are identical, being  $N$ . By relating to the state transition and measurement models in Equation 7 and 9, the training datasets are given by

$$\mathcal{D}^{tran} = \langle \mathbf{X}^{tran}, \mathbf{Y}^{tran} \rangle, \quad (30)$$

$$\mathcal{D}^{meas} = \langle \mathbf{X}^{meas}, \mathbf{Y}^{meas} \rangle, \quad (31)$$

where  $\mathbf{X}^{tran}$  and  $\mathbf{Y}^{tran}$  are the input and output data points for the transition model while the input and output data points for the measurement model are represented by  $\mathbf{X}^{meas}$  and  $\mathbf{Y}^{meas}$ . Notice that each input data point of the transition model is a  $(M^x + M^u) \times 1$  column vector containing both state and control vectors, denoted by

$$\mathbf{s} = \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}. \quad (32)$$

In addition, the definitions of  $\mathbf{X}^{tran}$ ,  $\mathbf{Y}^{tran}$ ,  $\mathbf{X}^{meas}$ , and  $\mathbf{Y}^{meas}$  are summarized in Table 1.

In Equation 27, the predictive distribution of a scalar output is derived. Thus, for an arbitrary element in the state vector at time  $t$ ,  $x_{p,t}$  where  $p = 1, 2, \dots, M^x$ , its distribution given the previous state-control pair,  $(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ , is denoted by

$$P(x_{p,t} | \mathbf{X}^{tran}, \mathbf{s}_{t-1}, \mathbf{y}_p^{tran}, \boldsymbol{\theta}_p^{tran}, \boldsymbol{\eta}_p^{tran}) = \mathcal{N}(v_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}), \Gamma_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran})), \quad (33)$$

where  $\mathbf{y}_p^{tran}$  corresponds to the  $p^{th}$  row vector in the matrix of  $\mathbf{Y}^{tran}$ , and  $\boldsymbol{\theta}_p^{tran}$  and  $\boldsymbol{\eta}_p^{tran}$  are the corresponding hyper-parameters. The  $v_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran})$  and  $\Gamma_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran})$  are the mean and variance functions according to Equations 28 and 29. It is assumed that the dimensions of state variables are independent among each other, and the same applies to the dimensions of the measurement variable. The resulting distribution of the state vector at time  $t$ ,  $\mathbf{x}_t$ ,

$$P(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) = \prod_{p=1}^{M^x} P(x_{p,t} | \mathbf{X}^{tran}, \mathbf{s}_{t-1}, \mathbf{y}_p^{tran}, \boldsymbol{\theta}_p^{tran}, \boldsymbol{\eta}_p^{tran}). \quad (34)$$

Similarly, the probability distribution of the measurement variable at  $t$ ,  $\mathbf{z}_t$ , based on the current state vector,  $\mathbf{x}_t$ , is given by

$$\begin{aligned} P(\mathbf{z}_t | \mathbf{x}_t) &= \prod_{q=1}^{M^z} P(z_{q,t} | \mathbf{X}^{meas}, \mathbf{x}_t, \mathbf{y}_q^{meas}, \boldsymbol{\theta}_q^{meas}, \boldsymbol{\eta}_q^{meas}) \\ &= \prod_{q=1}^{M^z} \mathcal{N}(v_q^{meas}(\mathbf{x}_t, \mathcal{D}^{meas}), \Gamma_q^{meas}(\mathbf{x}_t, \mathcal{D}^{meas})), \end{aligned} \quad (35)$$

where  $z_{q,t}$  represents the  $q^{th}$  element of the measurement vector, and  $\mathbf{y}_q^{meas}$  is the  $q^{th}$  row vector in the matrix of  $\mathbf{Y}^{meas}$ .  $\boldsymbol{\theta}_q^{meas}$  and  $\boldsymbol{\eta}_q^{meas}$  are the hyper-parameters for the  $q^{th}$  measurement model. The mean and variance functions of the  $q^{th}$  measurement model are denoted as  $v_q^{meas}(\mathbf{x}_t, \mathcal{D}^{meas})$  and  $\Gamma_q^{meas}(\mathbf{x}_t, \mathcal{D}^{meas})$ .

### 2.4. Hyper-parameter optimization

The hyper-parameters,  $\boldsymbol{\psi}$ , including the variance of noise variables and kernel parameters, partially govern the accuracy of a GP regression model, where

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix} \quad (36)$$

---

**Algorithm** One-step state estimation

---

**Input:**  $\mathbf{z}_t$ : current measurement;  
 $\mathbf{x}_{t-1}$ : previous estimated state;  
 $\Sigma_{t-1}$ : previous covariance of the estimated state;  
 $\mathbf{u}_{t-1}$ : previous control;  
 $\mathcal{D}^{tran}$ : historical data for the transition model;  
 $\mathcal{D}^{meas}$ : historical data for the measurement model;  
 $\mathcal{M}^{tran}$ : transition model;  
 $\mathcal{M}^{meas}$ : measurement model.

**Output:**  $\mathbf{x}_t$  and  $\Sigma_t$ : current estimated state and covariance;  
 $\mathcal{M}^{tran}$  and  $\mathcal{M}^{meas}$ : transition and measurement models.

- 1: **if**  $\mathcal{D}^{tran} \neq \emptyset$  **and**  $\mathcal{D}^{meas} \neq \emptyset$  **then**
- 2:  $\mathcal{M}^{tran} \leftarrow \text{updateTransitionModel}(\mathcal{D}^{tran});$
- 3:  $\mathcal{M}^{meas} \leftarrow \text{updateMeasurementModel}(\mathcal{D}^{meas});$
- 4: **end if**
- 5:  $\text{updateEKFFModels}(\mathcal{M}^{tran}, \mathcal{M}^{meas});$
- 6:  $\mathbf{x}_t, \Sigma_t \leftarrow \text{GPEKF}(\mathbf{x}_{t-1}, \Sigma_{t-1}, \mathbf{u}_{t-1}, \mathbf{z}_t);$
- 7: **return**  $\mathbf{x}_t, \Sigma_t, \mathcal{M}^{tran}, \mathcal{M}^{meas}$

---

Fig. 2. Pseudo-code of a one-step state estimation using the BFGP modeling framework.

A proper set of hyper-parameters can be inferred by making a point estimate that maximizes a log-likelihood function over the output variable conditioned on the training input variable and hyper-parameters. Using the general representation of a multivariate Gaussian distribution in Equation 22, the likelihood function,  $\mathcal{L}(\boldsymbol{\psi})$ , yields a log-likelihood function as follows:

$$\mathcal{L}(\boldsymbol{\psi}) = \ln P(\mathbf{y}|\mathbf{X}, \boldsymbol{\psi}) = -\frac{1}{2} \ln |\mathbf{C}| - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} - \frac{N}{2} \ln(2\pi). \quad (37)$$

The optimal hyper-parameters are then determined by

$$\boldsymbol{\psi}^* = \arg \max_{\boldsymbol{\psi}} \mathcal{L}(\boldsymbol{\psi}). \quad (38)$$

The maximum likelihood problem can be solved by different optimization methods, such as the scaled-conjugate gradient optimization algorithm [17].

### 3. Computational algorithm

So far, this paper has outlined a general approach for state estimation using the BFGP framework. This section extends the method for real-time state estimation and prediction by integration with Kalman filter.

#### 3.1. One-step estimation process

In our study, a kernel function based on the exponential of a quadratic form is adopted:

$$K(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}) = \theta_1 e^{-\frac{\theta_2}{2} (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)} + \theta_3 + \theta_4 \mathbf{x}_i^T \mathbf{x}_j, \quad i = 1, \dots, N, j = 1, \dots, N, \quad (39)$$

where  $\theta_1$  and  $\theta_2$  are corresponding parameters corresponding to the quadratic form.  $\theta_3$  and  $\theta_4$  correspond to constant and linear terms.

Fig. 2 depicts a one-step estimation process for a newly observed measurement. In addition to the current measurement, the previous estimated state and covariance as well as the previous control are needed to trigger the one-step



---

**Algorithm GPEKF**


---

**Input:**  $\mathbf{x}_{t-1}$ : previous estimated state;  
 $\Sigma_{t-1}$ : previous covariance of the estimated state;  
 $\mathbf{u}_{t-1}$ : previous control;  
 $\mathbf{z}_t$ : new measurement.

**Output:**  $\mathbf{x}_t, \Sigma_t$ .

```

1:  $\hat{\mathbf{x}}_t \leftarrow \text{computePredictedState}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1});$  # according to Equation 40
2:  $\hat{\Sigma}_t \leftarrow \text{computeUncertaintyOfPredictedState}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}, \Sigma_{t-1});$  # according to Equation 42
3:  $\mathbf{K}_t \leftarrow \text{computeKalmanGain}(\hat{\mathbf{x}}_t, \hat{\Sigma}_t);$  # according to Equation 43
4:  $\mathbf{x}_t \leftarrow \text{computeEstimatedState}(\hat{\mathbf{x}}_t, \mathbf{K}_t, \mathbf{z}_t);$  # according to Equation 45
5:  $\Sigma_t \leftarrow \text{computeUncertaintyOfEstimatedState}(\hat{\Sigma}_t, \mathbf{K}_t);$  # according to Equation 46
6: return  $\mathbf{x}_t, \Sigma_t$ ;

```

---

**Fig. 3.** Pseudo-code of the essential steps for GPEKF.

process of BF. If the historical data sets,  $\mathcal{D}^{tran}$  and  $\mathcal{D}^{meas}$ , are available, the transition and measurement models,  $\mathcal{M}^{tran}$  and  $\mathcal{M}^{meas}$ , are updated using the given data sets (see lines 1 – 4, Fig. 2).

Kalman filter is widely used framework to carry out state estimation and prediction for state-space models. However, Equations 34 and 35 show that the transition and measurement models are governed by nonlinear functions. To cope with the nonlinearities, a number of algorithms have been proposed under the BF modeling framework, such as extended Kalman filters (EKF), unscented Kalman filters (UKF), and Particle filters (PF) [23]. EKF is employed in this study, utilizing the first-order Taylor expansion for linearizing the GP-based nonlinear functions (see line 5, Fig. 2). Thereafter, GP-based EKF (GPEKF) utilizes the previous estimated state and covariance,  $\mathbf{x}_{t-1}$  and  $\Sigma_{t-1}$ , respectively, the previous control,  $\mathbf{u}_{t-1}$ , and the current measurement,  $\mathbf{z}_{t-1}$ , to generate the current state,  $\mathbf{x}_t$ , and its covariance matrix,  $\Sigma_t$ , (see line 6, Fig. 2).

### 3.2. GPEKF

Fig. 3 lists the essential steps for GPEKF in pseudo-code, and the mathematical formulations of each step are described subsequently. Lines 1 – 2 represent the prediction step while the update step is found in lines 3 – 5. Assume that the state is dimensionally independent, so the predicted state,  $\mathbf{s}_t$ , can be composed by

$$\hat{\mathbf{x}}_t = \begin{bmatrix} v_1^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \\ \dots \\ v_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \\ \dots \\ v_{M^x}^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \end{bmatrix}, \quad \mathbf{s}_{t-1} = \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{u}_{t-1} \end{bmatrix}, \quad p = 1, 2, \dots, M^x, \quad (40)$$

where the computation of the mean function,  $v_p^{tran}(\cdot, \cdot)$ , can be located in Equation 28.

For linearizing the nonlinear functions, a linear approximation to a function through the function's value and slope is constructed by using the first-order Taylor expansion. Let  $\mathbf{G}_t$  denote the Jacobian of GP mean function with respect to the state transition model, i.e.,

$$\mathbf{G}_t = \begin{bmatrix} \frac{\partial}{\partial \mathbf{s}_{t-1}} v_1^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \\ \dots \\ \frac{\partial}{\partial \mathbf{s}_{t-1}} v_p^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \\ \dots \\ \frac{\partial}{\partial \mathbf{s}_{t-1}} v_{M^x}^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \end{bmatrix}, \quad p = 1, 2, \dots, M^x. \quad (41)$$

Then, the covariance matrix of the predicted state distribution is computed by

$$\hat{\Sigma}_t = \mathbf{G}_t^T \Sigma_{t-1} \mathbf{G}_t + \begin{bmatrix} \Gamma_1^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \Gamma_{M^x}^{tran}(\mathbf{s}_{t-1}, \mathcal{D}^{tran}) \end{bmatrix}, \quad (42)$$

where the variance function for the transition models,  $\Gamma_p^{tran}(\cdot, \cdot)$ ,  $p = 1, 2, \dots, M^x$ , is computed by Equation 29.

Line 3 in the pseudo-code indicates that the Kalman gain is computed by using the predicted state and covariance matrix. Specifically, the formulation is

$$\mathbf{K}_t = \hat{\Sigma}_t \mathbf{H}_t^T (\mathbf{H}_t \hat{\Sigma}_t \mathbf{H}_t^T + \begin{bmatrix} \Gamma_1^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \Gamma_{M^x}^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \end{bmatrix})^{-1}, \quad (43)$$

where the variance functions for the measurement models,  $\Gamma_q^{meas}(\cdot, \cdot)$ ,  $q = 1, 2, \dots, M^x$ , are given according to Equation 29.  $\mathbf{H}_t$  refers to the Jacobian of the GP mean function with respect to the measurement models, i.e.,

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial}{\partial \hat{\mathbf{x}}_t} v_1^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \\ \dots \\ \frac{\partial}{\partial \hat{\mathbf{x}}_t} v_2^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \\ \dots \\ \frac{\partial}{\partial \hat{\mathbf{x}}_t} v_{M^x}^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \end{bmatrix}, \quad q = 1, 2, \dots, M^x, \quad (44)$$

where the mean function for the measurement model,  $v_{M^x}^{meas}(\cdot, \cdot)$ , is resultant from Equation 28.

Finally, the estimated state is calculated by adding the predicted state with the degree of correctness to which the measurement is incorporated into the new state estimate. Such a degree is proportional to the Kalman gain as well as the deviation of the observed current measurement and the predicted measurement. In particular, the current estimated state is

$$\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t (\mathbf{z}_t - \begin{bmatrix} v_1^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \\ \dots \\ v_q^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \\ \dots \\ v_{M^x}^{meas}(\hat{\mathbf{x}}_t, \mathcal{D}^{meas}) \end{bmatrix}), \quad q = 1, 2, \dots, M^x. \quad (45)$$

The updated covariance matrix of the estimated state,  $\Sigma_t$ , is produced through adjustment of the covariance matrix of the predicted state using information gain,  $\mathbf{I}^{M^x} - \mathbf{K}_t \mathbf{H}_t$ , i.e.,

$$\Sigma_t = (\mathbf{I}^{M^x} - \mathbf{K}_t \mathbf{H}_t) \hat{\Sigma}_t. \quad (46)$$

where  $\mathbf{I}^{M^x}$  signifies an  $M^x \times M^x$  unit matrix.

## 4. Numerical experiments

This section demonstrates the presented TSE methodology using different types of traffic data. The traffic data are generated using a microscopic simulator.

### 4.1. Experiment setup

In the experiments, the developed estimation approach was applied to a symbolic and isolated intersection (see Fig. 4a). The deployed signal controller applied the stage-based phasing where traffic lights were altered in a fixed-phase sequence (see Fig. 4b). The detector layout was derived from a typical Swedish detection system. One inductive short detector and one inductive long detector were placed 80 meters and 10 meters away from the stop-line respectively.

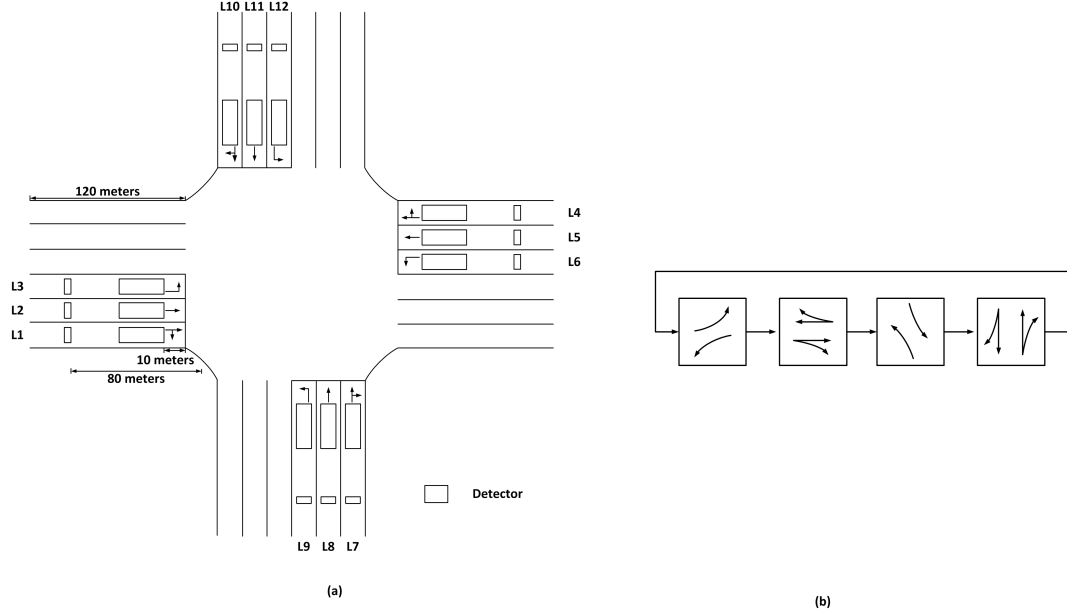


Fig. 4. Lane configuration (a), phase sequence (b) and detected signals from a loop detector (c).

Besides this, a vehicle-actuated timing approach was employed such that the green allocation varied depending on the vehicular presences detected from the loop detectors. In particular, if a loop detector is occupied by vehicle, the connected signal controller would receive a signal of a digital pulse, being 1, and the digital signal returned to 0 as soon as the vehicle has passed through the detector. The timing logic applied is a gap-seeking algorithm for green time extension, as described in Jin et al. [9].

In the present paper, state and control are defined in the same way as for different types of data sources, whereas the definition of measurement may differ according to the deployed monitoring system. The elements of the state vector represent the numbers of vehicles, including those both queued and approaching, on the lanes associated with the intersection. This definition of state has been applied for several adaptive signal control systems [e.g. 1, 5]. The state vector is analytically represented by

$$\mathbf{x}_t = [n_{1,t}, \dots, n_{k,t}, \dots, n_{N^{lane},t}], \quad k = 1, 2, \dots, N^{lane}, \quad (47)$$

where  $n_{k,t}$  denotes the number of vehicles on the  $k^{th}$  lane at  $t$ .  $N^{lane}$  represents the total number of lanes at the intersection, and  $N^{lane} = 12$  for the experiments.

The control data was collected employing general information on traffic light indications (i.e., green, yellow and red), which is accessible by any type of signal controller. For the state transition process, green and non-green light indications make a difference to the change of the number of vehicles on a lane. Next, the control vector is represented by

$$\mathbf{u}_t = [g_{1,t}, \dots, g_{k,t}, \dots, g_{N^{lane},t}], \quad k = 1, 2, \dots, N^{lane}, \quad (48)$$

where  $g_{k,t}$  refers to the percentage of green time within a defined estimation interval extracted from the traffic light that controls the  $k^{th}$  lane at  $t$ .

In the experiments, two types of data sources were incorporated into the estimation framework: loop detector data and connected vehicle data. For loop detector data, the measurements are derived from the received digital signals. Although including both long and short loop detectors would probably increases estimation accuracy, detections from the short detectors are, for generality, utilized in the analysis. A switch of the digital signal refers to “0-1-0”, and each switch implies that a vehicle drives through the loop detector.

The measurement vector for loop detector data is represented as

$$\mathbf{z}_t = [\tilde{n}_{1,t}^{loop}, \dots, \tilde{n}_{k,t}^{loop}, \dots, \tilde{n}_{N^{lane},t}^{loop}]^T, \quad k = 1, 2, \dots, N^{lane}, \quad (49)$$

where  $\tilde{n}_{k,t}^{loop}$  denotes the measurement extracted from the short loop detector associated with the  $k^{th}$  lane. Inspired by the conservation relation of the cumulated vehicles introduced in the input-output models, each element in the measurement vector indicates an approximation of the number of vehicles using such a relation. Then, the measurement element is computed by

$$\tilde{n}_{k,t}^{loop} = \max\{n_{k,t-1} + \frac{Q_{k,t}\Delta t}{300} - \frac{g_{k,t}\Delta t}{t^{pass}}, 0\}, \quad (50)$$

where  $\Delta t$  represents the estimation interval for each update.  $Q_{k,t}$  denotes the number of vehicles passing through a detector placed on the  $k^{th}$  lane over the previous five minutes.  $t^{pass}$  represents the time on average spent by the each vehicle to drive through a lane after it enters the lane. It is assumed that  $t^{pass} = 5$  seconds. Although such an assumption do not consider congested regime where vehicles are queued, the transition model tries to correct this measured error.

In terms of connected vehicle data, vehicle position is accessible for the signal controller when vehicle-to-infrastructure (V2I) communication is enabled. According to the available intersection geometry, the number of connected vehicles on each lane can be extracted instantly. During the experiments, each connected vehicle was registered with an ID when it entered the intersection, and that vehicle ID was recorded by the signal controller. Therefore, the flow of connected vehicles was equal to the number of unique vehicle IDs within a certain time interval.

Let  $Q_{k,t}^{cv}$  represent the number of connected vehicles flowing on lane  $k$  over the previous five minutes. Similar to the measurement definition using information from loop detectors, measurement elements,  $\tilde{n}_{k,t}^{cv}, k = 1, 2, \dots, N^{lane}$ , can be calculated by

$$\tilde{n}_{k,t}^{cv} = \max\{n_{k,t-1} + \frac{Q_{k,t}^{cv}\Delta t}{300r_{k,t}} - 0.2g_{k,t}\Delta t, 0\}, \quad (51)$$

where  $r_{k,t}$  indicates a rough estimation of the penetration rate of connected vehicles at  $t$ , recursively updated by

$$r_{k,t} = \sum_{l=t-1-\tau}^{t-1} \frac{n_{k,l}^{cv}}{n_{k,l}}. \quad (52)$$

Here,  $\tau$  denotes the length of time accounting for updating the penetration rate.  $n_{k,l}^{cv}$  represents the number of connected vehicles on lane  $k$  at time  $l$ . Consequently, the corresponding measurement vector is defined by

$$\mathbf{z}_t = [\tilde{n}_{1,t}^{cv}, \dots, \tilde{n}_{k,t}^{cv}, \dots, \tilde{n}_{N^{lane},t}^{cv}]^T, \quad k = 1, 2, \dots, N^{lane}. \quad (53)$$

According to the difference between the observed and estimated states for a time series, the evaluation of all experiments is based on two criteria: the mean absolute error (MAE) and weighted mean absolute percent error (WAPE). They are analytically computed by

$$MAE = \frac{\sum_{t=1}^T |\mathbf{x}_t^{obs} - \mathbf{x}_t|}{T}, \quad (54)$$

$$WAPE = \frac{\sum_{t=1}^T |\mathbf{x}_t^{obs} - \mathbf{x}_t|}{\sum_{t=1}^T \mathbf{x}_t^{obs}} \quad (55)$$

where  $T$  denotes the total time period of the series, and  $\mathbf{x}_t^{obs}$  is the observed state vector. Literally, MAE indicates the absolute error on average that can be expected from the estimation approach, and WAPE allows for a comparison of estimations under different conditions.

#### 4.2. Data preparation

As mentioned, simulated data is applied to evaluate the effectiveness of the proposed methodology. According to the applied intersection layout, a traffic model was built in an open-source microscopic simulator, SUMO 0.19.0 [11]. A self-developed software program for the signal controller was implemented to connect the SUMO simulator so that the traffic lights in the simulation were governed by the stage-based vehicle-actuated control [9]. To generate the state, control and measurement data, SUMO simultaneously recorded the detection information and the number of vehicles on each lane via a provided application programming interface, called TraCI.

**Table 2**  
Model parameters of the IDM and signal control parameters

Description	Value	Unit
Desired speed	54	km/hour
Standard deviation of desired speed	5	km/hour
Simulation step	0.1	second
Acceleration	1.0	m/s <sup>2</sup>
Comfortable deceleration	-1.5	m/s <sup>2</sup>
Time gap	1.0	seconds
Acceleration exponent	4	-
Minimum gap between vehicles	2	meters
Yellow time	3	seconds
All-red time	2	seconds
Minimum green time	5	seconds
Maximum green time	50	seconds

**Table 3**  
Traffic volume [vehicles/hour] for each turning movement at the studied intersection

Traffic flow scenario	West entry			East entry			South entry			North entry		
	L	T	R	L	T	R	L	T	R	L	T	R
Homogeneous (medium)	125	250	125	125	250	125	125	250	125	125	250	125
Homogeneous (high)	150	300	150	150	300	150	150	300	150	150	300	150
Arterial WE (medium)	125	250	125	125	250	125	70	125	70	70	125	70
Arterial WE (high)	150	300	150	150	300	150	75	150	75	75	150	75
Arterial NS (medium)	70	125	70	70	125	70	125	250	125	125	250	125
Arterial NS (high)	75	150	75	75	150	75	150	300	150	150	300	150
Validation	146	336	124	82	303	123	107	134	126	100	196	79

L, T, and R represent the turning rates of left-turn, through and right-turn movements, respectively.

Even though incorporating multiple vehicle classes (e.g., buses and trucks) into the formulation was possible, this was specifically avoided to simplify the presentation of the main ideas. All vehicles were assumed to be identical passenger cars herein. During the experiments, a commonly-used car-following model, the intelligent driver model (IDM), was adopted and the model parameters were consistent with the calibration results of Treiber and Kesting [25]. The car-following model parameters and signal control parameters are summarized in Table 2.

In this paper, the analysis of the estimation results is divided into two parts based on the applied measurement data types. Various experimental settings (e.g., estimation interval and penetration rate) were investigated for both data sources. For each experiment, three datasets, including the training, validation, and test sets, were generated by simulations. A training dataset refers to a set of data examples used to discover potential relationships among states, controls, and measurements. A validation set is employed to compare the model performance in terms of estimation accuracy based on the aforementioned performance criteria. Finally, a test set is adopted to assess the utility of the proposed estimation approach and offer detailed information on the estimation results.

To generate the simulated data, vehicles were randomly sampled according to a Poisson process with an arrival rate in vehicles per second. Table 3 presents the six applied traffic flow scenarios which produce 500 data points each. The ‘‘Homogeneous’’ scenarios in the table assume that traffic flows are identical for all four approaches of the intersection. Either the north-south or the east-west directions are assumed to be the major roads in the ‘‘Arterial’’ scenarios. The corresponding traffic flow is scaled to the ‘‘medium’’ or ‘‘high’’ level. The ‘‘medium’’ level represents a normal traffic flow condition for the intersection, whereas the ‘‘high’’ level describes a significant increase (approximately 20%) in traffic flow compared to the ‘‘medium’’ scenarios.

**Table 4**

Estimation errors (MAE and MAPE) for the lanes associated with the intersection for four update intervals (i.e., 1, 5, 10, and 20 seconds using a validation set)

Update interval	Estimation error	Lanes											
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
1	MAE[veh]	0.21	0.16	0.13	0.22	0.21	0.09	0.19	0.16	0.12	0.21	0.17	0.08
	WAPE[%]	6.57	7.06	5.57	5.63	7.17	7.16	6.14	7.89	5.38	5.26	7.15	6.33
5	MAE[veh]	0.51	0.46	0.36	0.49	0.47	0.27	0.42	0.38	0.34	0.45	0.43	0.28
	WAPE[%]	7.65	7.89	11.54	8.82	10.34	14.88	13.20	19.14	12.71	10.29	13.67	14.49
10	MAE[veh]	0.69	0.65	0.57	0.66	0.64	0.42	0.59	0.52	0.52	0.62	0.60	0.38
	WAPE[%]	7.15	7.37	15.11	9.62	10.98	21.50	16.14	24.45	16.88	13.69	18.32	21.18
20	MAE[veh]	0.98	0.89	0.86	0.89	0.83	0.63	0.76	0.66	0.75	0.78	0.76	0.65
	WAPE[%]	8.12	7.71	21.12	12.25	13.27	28.86	21.99	24.93	23.19	17.02	22.25	27.03

For the validation data set, the traffic flow rate was randomly generated, and the values opted for are also found in Table 3. Each validation dataset contained 600 simulated data points. In this simulation generating a test set, the traffic flow rate commenced with the “Arterial WE (medium)” scenario. The pattern of traffic flow scenarios was arranged from “Arterial WE (medium)” to “Arterial WE (high)” to “Homogeneous (high)” to “Arterial NS (high)” to “Homogeneous (medium)” to “Arterial NS (medium)”. The simulation lasted for 600 seconds with respect to each traffic flow scenario such that the simulation was carried out for 3,600 seconds in total. Regardless of whether the measurement was derived from loop detector data or connected vehicle data, the same traffic flows were applied when generating the three datasets. Nevertheless, the resulting vehicle generation time were different because of randomness in traffic simulation.

### 4.3. Results and discussions

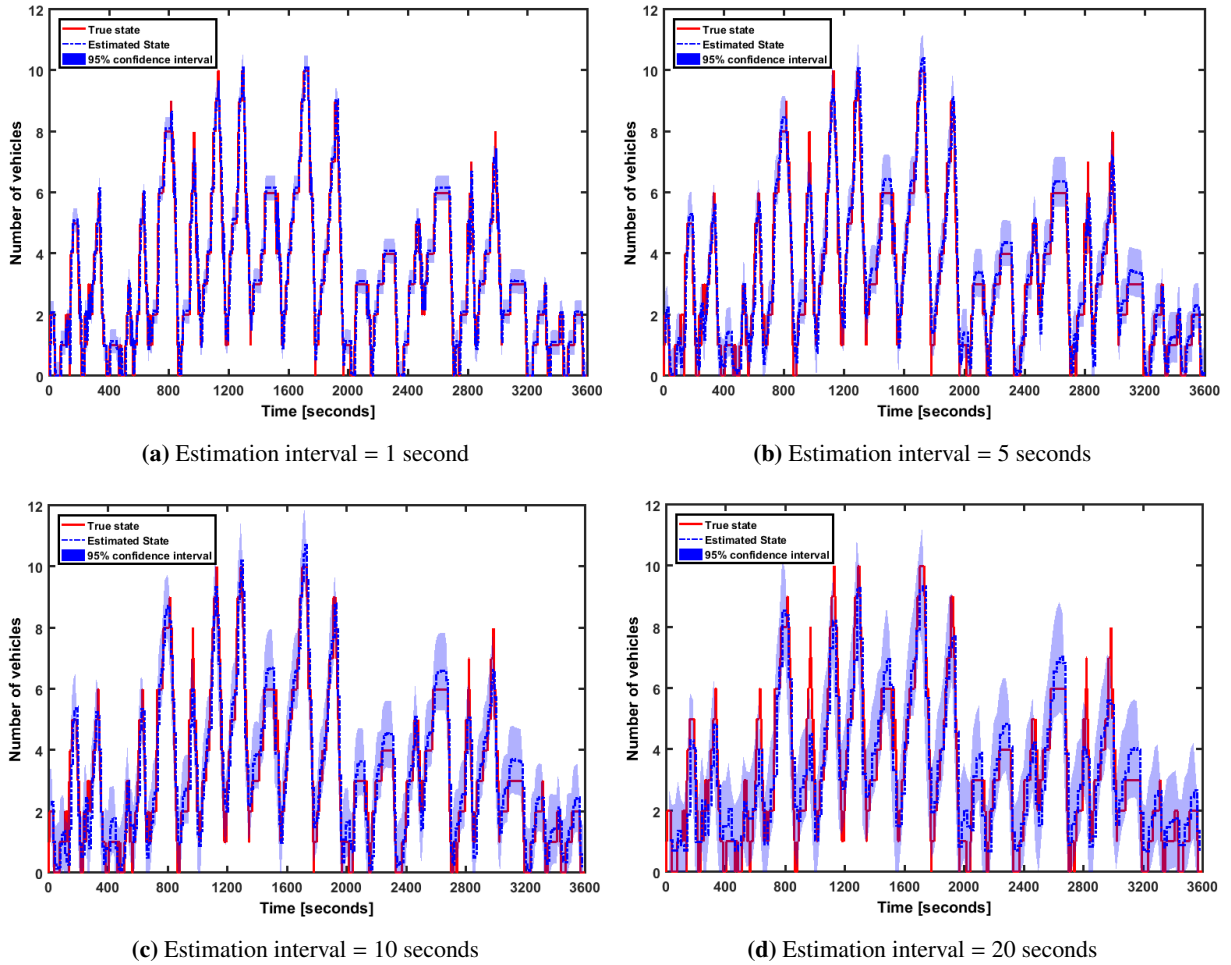
#### 4.3.1. Fixed-location data

For investigating the proposed estimation approach using fixed-location data (loop detector data in this case), four different estimation intervals i.e., 1, 5, 10, and 20 seconds, were applied. The estimated states are compared with the actual states extracted from the SUMO simulations. Table 4 shows the estimation errors (i.e., MAE and MAPE) generated from the validation set. The estimated state follows the ground truth reasonably well for all the tested estimation intervals.

In particular, Table 4 shows that all the WAPes are below 30%, and the corresponding absolute errors on average are below one vehicle in the experiments. The TSE approach is usually more challenging for generating competitive performance when being conducted for estimating states over a shorter time interval. However, the validation results demonstrate the advantage in this regard. Both of the two types of errors are among the smallest for the estimation interval of one second, regardless of lanes. When the estimation interval is one second, all the reported absolute mean errors of the number of vehicles can be as small down to less than 0.3, which is considered negligible.

To provide more insight into the estimation results, Fig. 5 and Fig. 6 display the estimated and actual number of vehicles over time for the simulated test data set with respect to two typical lanes, lanes “L2” and “L8”. Both lanes indicate through directions, and lanes “L2” and “L8” refer to eastbound and northbound directions, respectively. The presences of vehicles on lane “L2” increase gradually and then decrease back to smaller values within the tested simulation period, whereas the counterpart of lane “L8” becomes high after approximately 1,800 seconds because of the saturated traffic condition at the intersection.

In contrast to a point estimator that is proposed, for the most part, in the literature, the estimation approach in this study is with an interval estimator such that a region covering the states with 95% confidence intervals were also drawn in the two figures. Generally speaking, the estimation approach reports satisfactory results. The region of the 95% confidence interval contains almost the actual states, meaning that the coverage probability was close to 100%. Further, when the estimation interval is enlarged, the region of the confidence interval becomes greater irrespective

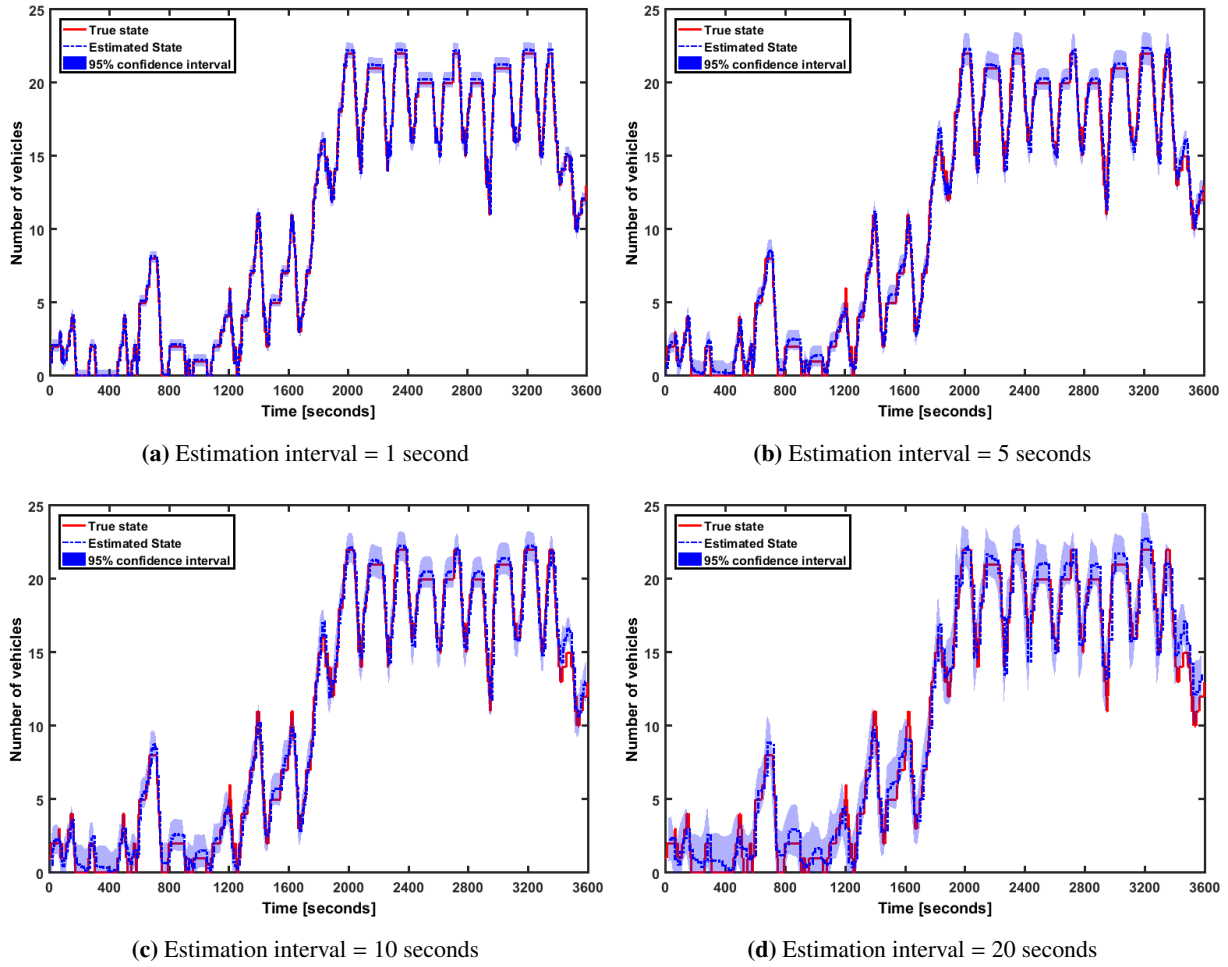


**Fig. 5.** Estimated number of vehicles, the ground truth, and the 95% confidence interval over time for the “L2” lane in the tested scenario with respect to four estimation intervals

of which lane is considered. This is in agreement with the authors’ expectations as additional uncertainties should be introduced by the estimation models if more information between two estimation points is omitted.

If one closely examines Fig. 5a and Fig. 6a, the fluctuation of the estimated number of vehicles on the lanes almost coincides with the actual number when the estimation interval is set to one second. Specifically, relatively high errors are present when the number of vehicles suddenly arrives at an extreme value, but the highest absolute error in the two plots is still below 0.5 vehicles. The absolute errors are kept within 0.2 vehicles over time for the case when the value of the state remains stable (e.g., TSE from 3,000 to 3,600 seconds for lane “L2”).

Nevertheless, from the two figures, a deterioration in estimation accuracy can be apparently seen along with the increase of the estimation interval, especially in the case where the traffic state was altered substantially. For instance, the number of vehicles on lane “L2” switches between approximately ten and zero over a short period from the simulated 800 to 2,000 seconds. Fig. 5d presents that the differences between the actual states and the estimated states are relatively large when the estimation interval is 20 seconds, compared with consequent estimation performance from 2,000 seconds to 3,200 seconds in Fig. 6d when the state is more stabilized with respect to the same estimation interval.



**Fig. 6.** Estimated number of vehicles, the ground truth, and the 95% confidence interval over time for the “L8” lane in the tested scenario with respect to four estimation intervals

**Table 5**

Estimation errors (MAE and MAPE) for the lanes associated with the intersection for four penetration rates (i.e., 25%, 50%, 75% and 100%) when the estimation interval was 1 second

Penetration rate [%]	Estimation error	Lanes											
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
25	MAE[veh]	0.20	0.19	0.22	0.23	0.29	0.09	0.25	0.29	0.22	0.23	0.19	0.21
	WAPE[%]	4.78	6.68	7.03	5.34	10.48	8.79	7.69	13.85	7.70	5.94	8.19	9.39
50	MAE[veh]	0.19	0.16	0.13	0.19	0.21	0.07	0.21	0.16	0.18	0.21	0.17	0.09
	WAPE[%]	4.34	5.86	3.87	4.23	6.96	6.33	6.15	7.67	7.01	5.23	6.61	3.96
75	MAE[veh]	0.18	0.18	0.12	0.21	0.20	0.07	0.20	0.17	0.13	0.21	0.15	0.10
	WAPE[%]	3.86	5.53	3.58	4.06	5.56	6.63	5.75	6.89	4.35	4.95	5.62	5.79
90	MAE[veh]	0.19	0.16	0.13	0.23	0.21	0.08	0.18	0.17	0.12	0.20	0.17	0.08
	WAPE[%]	4.45	4.98	3.75	4.78	5.74	6.58	4.31	5.18	6.37	4.04	5.32	4.24



**Table 6**

Estimation errors (MAE and MAPE) for the lanes associated with the intersection for four penetration rates (i.e., 25%, 50%, 75% and 100%) when the estimation interval was 20 seconds

Penetration rate [%]	Estimation error	Lanes											
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
25	MAE[veh]	0.99	1.16	1.26	1.10	1.22	0.76	0.88	0.83	1.00	0.96	0.98	1.47
	WAPE[%]	7.02	8.65	29.53	9.09	11.00	40.08	25.71	42.38	24.21	20.07	27.74	70.83
50	MAE[veh]	0.61	0.69	1.00	0.64	0.77	0.56	0.73	0.63	0.81	0.81	0.79	0.62
	WAPE[%]	4.27	5.07	22.58	4.54	5.72	26.78	20.86	29.07	17.72	16.38	20.70	26.76
75	MAE[veh]	0.55	0.62	0.89	0.56	0.67	0.56	0.71	0.61	0.70	0.78	0.74	0.58
	WAPE[%]	3.78	4.48	19.65	3.84	4.88	27.27	19.21	27.55	13.00	14.60	17.94	26.57
90	MAE[veh]	0.60	0.64	0.90	0.51	0.64	0.59	0.69	0.64	0.70	0.74	0.71	0.60
	WAPE[%]	4.09	4.54	19.30	3.48	4.64	27.34	17.78	27.61	12.74	13.70	17.01	26.86

#### 4.3.2. Mobile data

Several experiments were further performed to evaluate the effectiveness of the proposed estimation approach using the connected vehicle data. In addition to investigating the effects on estimation accuracy with respect to various estimation intervals, an analysis was carried out on applying different penetration rates of connected vehicles. Four penetration rates, ranging from 0% to 100% (i.e., 25%, 50%, 75%, and 90%) were implemented. Note that neither 0% nor 100% needs to be included in our analysis since TSE cannot be carried out when no information (0% penetration rate) is provided, and meanwhile it is trivial to obtain the state when the information on vehicles at the entire intersection is fully accessible (100% penetration rate).

A similar tendency of the estimation accuracy, compared to the results generated using loop detector-based data, was observed after performing the experiments using a variety of estimation intervals. This means that the estimator also suffers from increase of the estimation interval when connected vehicle data is in use. In addition, sensitivity analysis on the change of the penetration rate was conducted. Table 5 and Table 6, respectively, summarize the estimation results for two typical estimation intervals (i.e., 1 second and 20 seconds) with the validation set in terms of the four penetration rates.

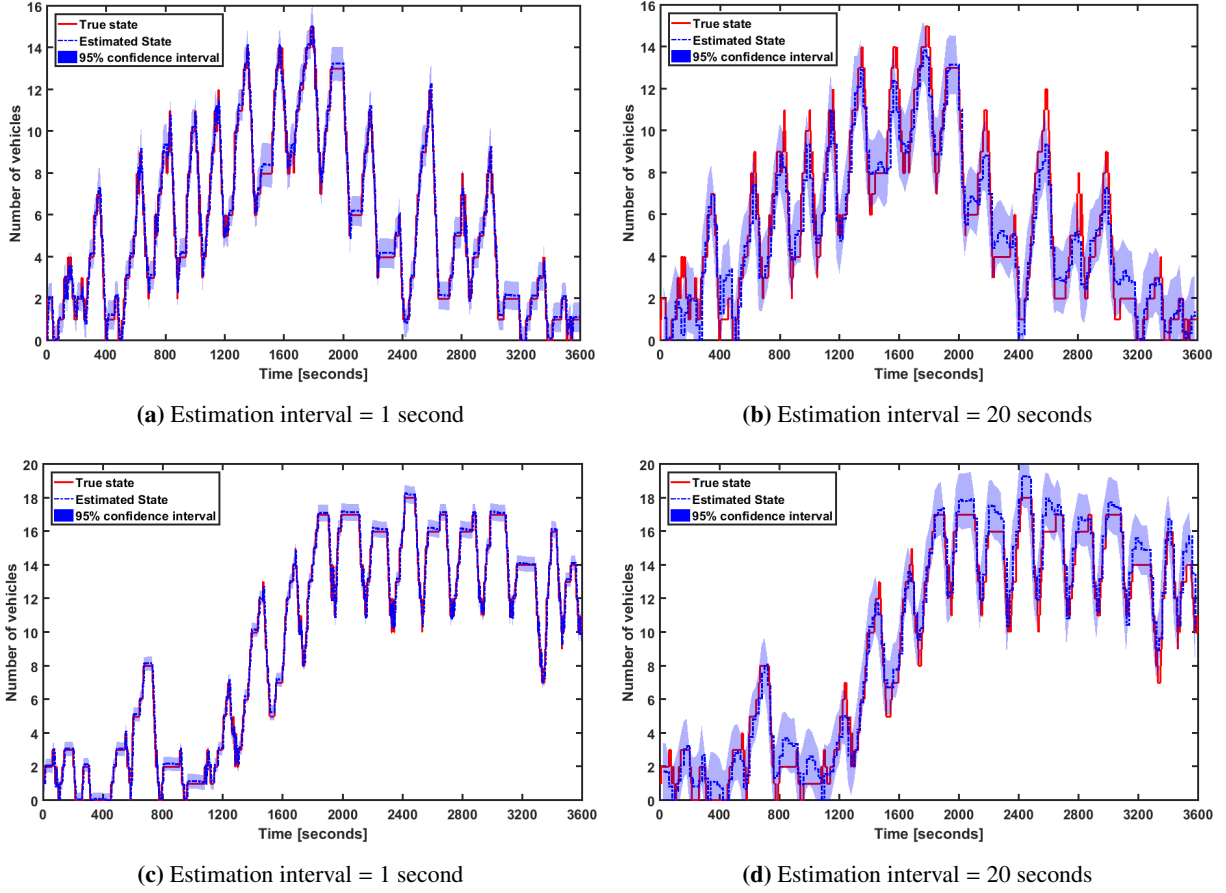
It is obvious that the estimation accuracy would be enhanced if more vehicles are involved to provide information for TSE, which is substantiated by the results of each table. Comparing the errors from the two tables with those provided in Table 4, the estimator using connected vehicle data in the case of the 50% penetration rate is able to achieve a similar level of accuracy as the case where the estimator adopts loop detector-based data.

In Table 5, the estimation accuracy does not significantly change with a reduction of the penetration rate until the penetration rate reaches 25% when the estimation interval is one second. Although the increase of WAPE can be up to 8.67% (i.e., from 5.18% to 13.85% for lane “L8”), the estimation performance is still acceptable because the mean absolute errors are all below 0.3 vehicles merely using information extracted from 25% of the vehicles at the intersection. Conversely, the estimator cannot be feasible because of the lack of vehicular information when the states are estimated every 20 seconds. For example, a large percentage error, 70.83% WAPE, is reported by the estimator performing for lane “L12” using the simulated data of the validation scenario.

Fig. 7 demonstrates the change for the estimated number of vehicles, and the 95% confidence level along with the actual number of vehicles for the same lanes (i.e., “L2” and “L8”) as presented in the previous subsection. In the plots therein, the applied penetration rate is 50%, and the estimation interval is one second or 20 seconds. As expected, the accuracy of estimation is comparatively high when the estimation interval is one second for both lanes, whereas, the accuracy is lower for the estimation interval of 20 seconds together with a larger area of the 95% confidence interval.

To discern the impacts of the level of inclusion, Fig. 8 portrays the change in estimation errors over time for the four tested penetration rates associated with lanes “L2” and “L8”, when the estimation interval is one second or 20 seconds. From Fig. 8a and Fig. 8c, most of the errors are limited within the range between -1 and 1 vehicles if the estimation interval is one second for all the four tested penetration rates.

For an estimation interval of 20 seconds, the absolute errors are below two vehicles when 90% of vehicles are



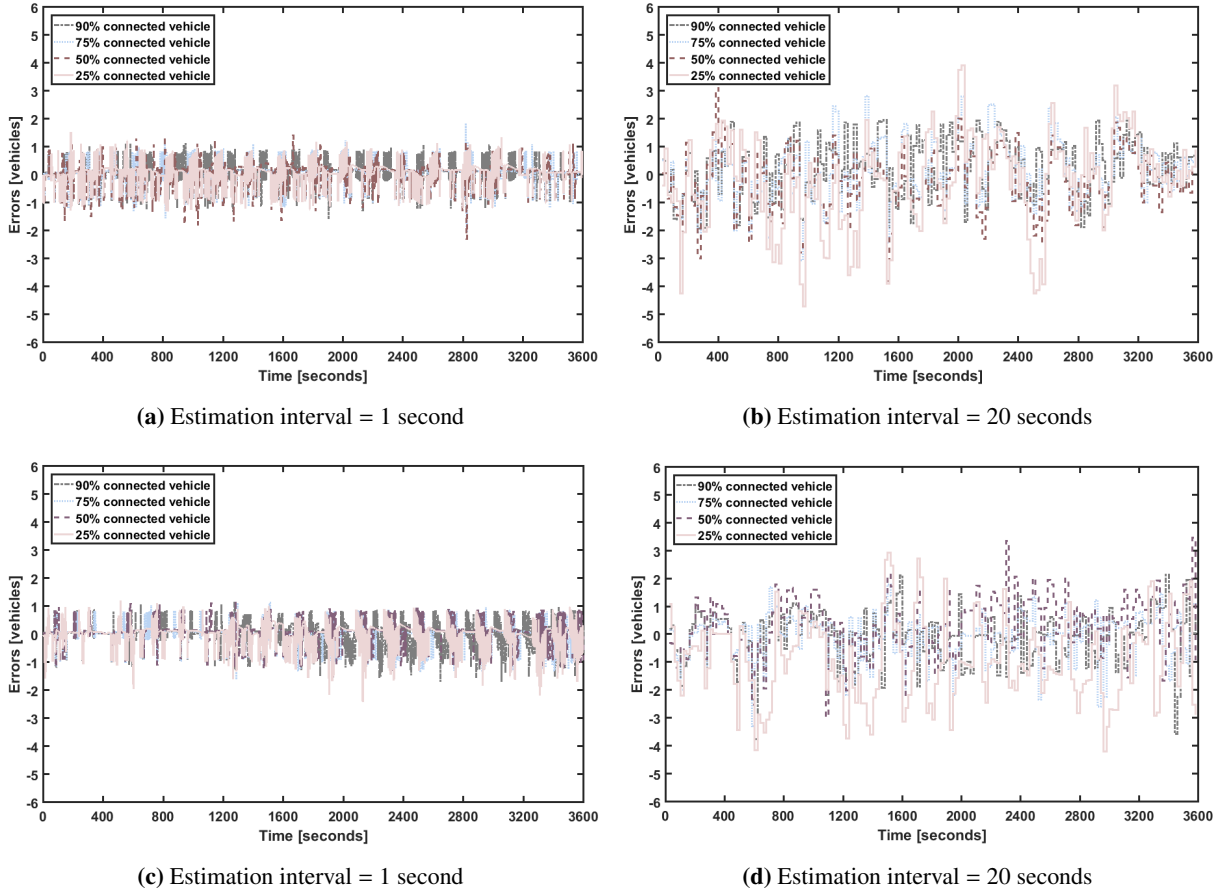
**Fig. 7.** Estimated and actual number of vehicles, and the 95% confidence interval over time for lanes “L2” and “L8” in the tested scenario with respect to two estimation intervals with 50% connected vehicles being employed

required to provide positional information. However, the error can be up to four vehicles for certain instants if less than 75% of vehicles are connected to the signal controller. At those instants, the estimator with such large errors is not reliable enough to be employed. Consequently, the proposed TSE framework is capable of providing a promising scheme for positional information on connected vehicles if the estimation interval is small (e.g., 1 second), even when the penetration rate is low. However, the effectiveness of the estimator starts to decrease when the estimation interval grows due to the lack of the transition and measurement information within the period between the two estimation time points.

## 5. Conclusions

The work presented here describes a general TSE framework for signalized intersections. The framework is not limited by external conditions that include the applied signal controller and available data sources. Moreover, any effort to model traffic systems at the signalized intersections is not necessary. In particular, the framework is based on the theory of BFs and is incorporated with a data-driven modeling approach, i.e., GP regression model.

The methodological framework is partitioned into two parts: off-line training and on-line estimation. With regards to off-line training, the GP regression models aim to build the essential models of BFs (i.e., transition and measurement models) with the acquired historical data set. The hyper-parameters employed in the kernel function of GPs are optimally determined according to the training data. The mathematical derivations of the recursive state estimation approach, incorporated with the predictive approach of GP regression models (BFGP), are also presented in detail.



**Fig. 8.** The errors (difference between the actual state and estimated state ) over time for four penetration rates applied that are associated with lanes “L2” and “L8”, respectively, when the estimation interval was 1 second or 20 seconds

When the BFGP model is ready, the states and uncertainties can be recursively computed when new measurements are collected on-line. Inspired by the idea of the general EKF method, a practical algorithm, namely GPEKF, is introduced and its procedure is depicted in pseudo-codes.

Various numerical experiments were carried out to verify the proposed modeling framework using data generated from an open-source microscopic traffic simulation tool. In the experiments, the number of vehicles in a lane was defined as an element of the traffic state for an isolated and signalized intersection. The control was defined to be accessible by any type of signal controller, which was related to the durations of traffic light indications. More importantly, two types of data, including loop detector-based data and connected vehicle data, were employed to derive measurements.

Three datasets were used in each of the experiments, specifically the training, validation, and test sets. Six traffic flow scenarios, covering typical traffic conditions at the intersection, were used to generate the training data. The validation set was randomly generated and utilized to compare model’s accuracy with different settings. In the test data set, the traffic flow was varied to investigate the versatility of the models in terms of various traffic conditions.

The estimation results using both loop detector data and connected vehicle data show the effectiveness of the estimation approach incorporated with different settings of the estimation interval. In particular, unlike the other proposed estimation approaches, such as that of Vigos et al. [27], the proposed framework is capable of providing a promising scheme for short-term TSE (e.g., one second), which is considered a more challenging task for other model-based TSE approaches. The relatively long-term TSE estimation (e.g., cycle by cycle) is intuitively obtained if the states can be estimated every second in a near precise manner. In addition, a sensitivity analysis is conducted

through the utility of different penetration rates of connected vehicles. The proposed approach manages to distinguish itself as a competitive candidate without deploying costly loop detectors to estimate traffic states, even if a minority of vehicles (25%) within the entire intersection are connected vehicles.

Further studies will continue with the development of a modeling approach based on the existing framework. In particular, an on-line learning process will be facilitated to adaptively enhance the models when streaming data is received. In addition, the approach will be integrated into a group-based intelligent signal control system proposed by Jin and Ma [8] with a proper understanding of traffic system.

## Acknowledgments

The study is supported by the Swedish Innovation Agency (Vinnova), Swedish Transport Administration (Trafikverket) and China Scholarship Council (CSC) (#201407930010) The financial supports are greatly acknowledged.

## A. Derivation of the recursive state estimation by BFs

The belief distribution over a state variable,  $\mathbf{x}_t$ , can be derived as

$$\begin{aligned} \text{bel}(\mathbf{x}_t) &= \frac{P(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})P(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})}{\int P(\mathbf{z}_t|\mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})P(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})d\mathbf{x}_t} \\ &= \frac{P(\mathbf{z}_t|\mathbf{x}_t)P(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})}{\int P(\mathbf{z}_t|\mathbf{x}_t)P(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})d\mathbf{x}_t}. \end{aligned} \quad (56)$$

According to the definition of the initial belief distribution,  $\overline{\text{bel}}(\mathbf{x}_t)$ , in Equation 7, Equation 56 can be rewritten as

$$\text{bel}(\mathbf{x}_t) = \frac{P(\mathbf{z}_t|\mathbf{x}_t)\overline{\text{bel}}(\mathbf{x}_t)}{\int P(\mathbf{z}_t|\mathbf{x}_t)\overline{\text{bel}}(\mathbf{x}_t)d\mathbf{x}_t}. \quad (57)$$

By expanding the initial belief distribution, the following equalities hold:

$$\begin{aligned} \overline{\text{bel}}(\mathbf{x}_t) &= \int P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})P(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-1})d\mathbf{x}_{t-1} \\ &= \int P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1})P(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{u}_{0:t-2})d\mathbf{x}_{t-1} \\ &= \int P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1})\overline{\text{bel}}(\mathbf{x}_{t-1})d\mathbf{x}_{t-1}. \end{aligned} \quad (58)$$

Equation 58 indicates that the internal belief distribution can be updated by processing the previous belief distribution together with the transition function and measurement function.

## B. Derivation of the conditional distribution for the predictive variable

From the product rule of probability, the corresponding conditional distribution over  $y_{N+1}$  in Equation 27 is a Gaussian distribution. The mean and covariance of a Gaussian distributed variable can be more efficiently obtained by considering the quadratic form in the exponent and then reinstating the normalization coefficient at the end of the calculation. The explanation is given as follows. For a general Gaussian distribution,  $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , the exponent can be written as

$$-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \text{const}, \quad (59)$$

where “const” refers to terms being independent of  $\mathbf{x}$ . If we take our general quadratic form and express it in the form given by the right-hand side of Equation 59, then the inverse covariance matrix,  $\boldsymbol{\Sigma}^{-1}$ , can be immediately equated by the coefficient of the second-order term entering  $\mathbf{x}$ . When  $\boldsymbol{\Sigma}^{-1}$  is calculated, the mean function can be obtained using the coefficient of the linear term,  $\boldsymbol{\Sigma}^{-1}\mathbf{x}$ , in  $\boldsymbol{\mu}$ .

The conditional distribution,  $P(y_{N+1}|\mathbf{X}, \mathbf{x}_{N+1}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta})$ , can be written as

$$P(y_{N+1}|\mathbf{X}, \mathbf{x}_{N+1}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{P(\mathbf{y}, y_{N+1}|\mathbf{X}, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \boldsymbol{\eta})}{P(\mathbf{y}|\mathbf{X}, \mathbf{x}_{N+1}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta})} \quad (60)$$

As the conditional distribution is completely characterized by its mean and variance and there does not exist a  $y_{N+1}$  term in  $P(\mathbf{y}|\mathbf{X}, \mathbf{x}_{N+1}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta})$ , the mean and variance of the conditional distribution can be directly evaluated from the joint distribution,  $P(\mathbf{y}, y_{N+1}|\mathbf{X}, \mathbf{x}_{N+1}, \boldsymbol{\theta}, \boldsymbol{\eta})$ , by fixing the resulting expression to relate with the variable,  $y_{N+1}$ .

By noting Equation 24, the exponent is derived by

$$-\frac{1}{2} \begin{bmatrix} \mathbf{y} \\ y_{N+1} \end{bmatrix}^T \begin{bmatrix} \mathbf{C} & \mathbf{c}_{N+1} \\ \mathbf{c}_{N+1}^T & C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ y_{N+1} \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} y_{N+1} \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) & \mathbf{c}_{N+1}^T \\ \mathbf{c}_{N+1} & \mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} y_{N+1} \\ \mathbf{y} \end{bmatrix} \quad (61)$$

Based on the identity of the inverse of a partitioned matrix, it is obtained that

$$\begin{bmatrix} C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) & \mathbf{c}_{N+1}^T \\ \mathbf{c}_{N+1} & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda^a & \Lambda^b \\ \Lambda^c & \Lambda^d \end{bmatrix} \quad (62)$$

where

$$\begin{aligned} \Lambda^a &= [C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} \\ \Lambda^b &= -[C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \\ \Lambda^c &= -\mathbf{C}^{-1} \mathbf{c}_{N+1} [C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} \\ \Lambda^d &= \mathbf{C}^{-1} + \mathbf{C}^{-1} \mathbf{c}_{N+1} [C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \end{aligned}$$

By revisiting the functional dependence of Equation 61 on  $y_{N+1}$ , in which  $\mathbf{y}$  is regarded as a constant, if all terms that are second order in  $y_{N+1}$  are selected, what is yielded is

$$-\frac{1}{2} y_{N+1}^T [C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} y_{N+1} \quad (63)$$

and all of the terms in Equation 61 that are linear in  $y_{N+1}$  are

$$y_{N+1}^T [C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{c}_{N+1}]^{-1} \mathbf{c}_{N+1}^T \mathbf{C}^{-1} \mathbf{y} \quad (64)$$

Consequently, by comparing Equation 63 and Equation 64 to Equation 59, the mean and variance of the variable  $y_{N+1}$  are derived as Equation 28 and Equation 29, respectively.

## References

- [1] Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2009. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C: Emerging Technologies* 17 (2), 163–174.
- [2] Ban, X. J., Hao, P., Sun, Z., 2011. Real time queue length estimation for signalized intersections using travel times from mobile sensors. *Transportation Research Part C: Emerging Technologies* 19 (6), 1133–1156.
- [3] Cheng, Y., Qin, X., Jin, J., Ran, B., 2012. An exploratory shockwave approach to estimating queue length using probe trajectories. *Journal of Intelligent Transportation Systems* 16 (1), 12–23.
- [4] Comert, G., 2013. Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals. *Transportation Research Part B: Methodological* 55, 59–74.
- [5] Gregoire, J., Qian, X., Frazzoli, E., De La Fortelle, A., Wongpiromsarn, T., 2015. Capacity-aware backpressure traffic signal control. *IEEE Transactions on Control of Network Systems* 2 (2), 164–173.
- [6] Hao, P., Ban, X., 2015. Long queue estimation for signalized intersections using mobile data. *Transportation Research Part B: Methodological* 82, 54–73.
- [7] Hofleitner, A., Herring, R., Bayen, A., 2012. Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning. *Transportation Research Part B: Methodological* 46 (9), 1097–1122.
- [8] Jin, J., Ma, X., 2017. A group-based traffic signal control with adaptive learning ability. *Engineering Applications of Artificial Intelligence* 65, 282–293.
- [9] Jin, J., Ma, X., Kosonen, I., 2017. A stochastic optimization framework for road traffic controls based on evolutionary algorithms and traffic simulation. *Advances in Engineering Software* 114, 348–360.

- [10] Ko, J., Fox, D., 2009. Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. *Autonomous Robots* 27 (1), 75–90.
- [11] Krajzewicz, D., Erdmann, J., Behrisch, M., Bieker, L., 2012. Recent development and applications of SUMO—simulation of urban mobility. *International Journal On Advances in Systems and Measurements* 5 (3 and 4), 128–138.
- [12] Lee, S., Wong, S., Li, Y., 2015. Real-time estimation of lane-based queue lengths at isolated signalized junctions. *Transportation Research Part C: Emerging Technologies* 56, 1–17.
- [13] Liu, H. X., Wu, X., Ma, W., Hu, H., 2009. Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies* 17 (4), 412–427.
- [14] Louati, A., Darmoul, S., Elkosantini, S., ben Said, L., 2017. An artificial immune network to control interrupted flow at a signalized intersection. *Information Sciences* 433-434, 70–95.
- [15] Memoli, S., Cantarella, G. E., de Luca, S., Di Pace, R., 2017. Network signal setting design with stage sequence optimisation. *Transportation Research Part B: Methodological* 100, 20–42.
- [16] Mirchandani, P. B., Zou, N., 2007. Queuing models for analysis of traffic adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems* 8 (1), 50–59.
- [17] Møller, M. F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks* 6 (4), 525–533.
- [18] Murphy, K. P., 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.
- [19] Ramezani, M., Geroliminis, N., 2015. Queue profile estimation in congested urban networks with probe data. *Computer-Aided Civil and Infrastructure Engineering* 30 (6), 414–432.
- [20] Rasmussen, C. E., 2006. *Gaussian Processes for Machine Learning*. MIT press.
- [21] Seo, T., Bayen, A. M., Kusakabe, T., Asakura, Y., 2017. Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*.
- [22] Skabardonis, A., Geroliminis, N., 2008. Real-time monitoring and control on signalized arterials. *Journal of Intelligent Transportation Systems* 12 (2), 64–74.
- [23] Thrun, S., Burgard, W., Fox, D., 2005. *Probabilistic Robotics*.
- [24] Tiaprasert, K., Zhang, Y., Wang, X. B., Zeng, X., 2015. Queue length estimation using connected vehicle technology for adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems* 16 (4), 2129–2140.
- [25] Treiber, M., Kesting, A., 2012. *Traffic flow dynamics: data, models and simulation*. Springer Science & Business Media.
- [26] Vigos, G., Papageorgiou, M., 2010. A simplified estimation scheme for the number of vehicles in signalized links. *IEEE Transactions on Intelligent Transportation Systems* 11 (2), 312–321.
- [27] Vigos, G., Papageorgiou, M., Wang, Y., 2008. Real-time estimation of vehicle-count within signalized links. *Transportation Research Part C: Emerging Technologies* 16 (1), 18–35.
- [28] Viti, F., Van Zuylen, H. J., 2010. A probabilistic model for traffic at actuated control signals. *Transportation Research Part C: Emerging Technologies* 18 (3), 299–310.
- [29] Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended kalman filter: a general approach. *Transportation Research Part B: Methodological* 39 (2), 141–167.
- [30] Zhan, X., Li, R., Ukkusuri, S. V., 2015. Lane-based real-time queue length estimation using license plate recognition data. *Transportation Research Part C: Emerging Technologies* 57, 85–102.