Traffic state estimation on highway: A comprehensive survey

Toru Seo, Alexandre M. Bayen, Takahiko Kusakabe, Yasuo Asakura

Tokyo Institute of Technology, 2-12-1-M1-20, O-okayama, Meguro, Tokyo 152-8552, Japan
University of California, Berkeley, 109 McLaughlin Hall, Berkeley CA 94720-1720, United States
Center for Spatial Information Science, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

Abstract
Traffic state estimation (TSE) refers to the process of the inference of traffic state variables (i.e., flow, density, speed and other equivalent variables) on road segments using partially observed traffic data. It is a key component of traffic control and operations, because traffic variables are measured not everywhere due to technological and financial limitations, and their measurement is noisy. Therefore, numerous studies have proposed TSE methods relying on various approaches, traffic flow models, and input data. In this review article, we conduct a survey of highway TSE methods, a topic which has gained great attention in the recent decades.

We characterize existing TSE methods based on three fundamental elements: estimation approach, traffic flow model, and input data. Estimation approach encompasses methods that estimate the traffic state, based on partial observation and a priori knowledge (assumptions) on traffic dynamics. Estimation approaches can be roughly classified into three according to their dependency on a priori knowledge and empirical data: model-driven, data-driven, and streaming-data-driven. A traffic flow model usually means a physics-based mathematical model representing traffic dynamics, with various solution methods. Input data can be characterized by using three different properties: collection method (stationary or mobile), data representation (disaggregated or aggregated), and temporal condition (real-time or historical).

Based on our proposed characterization, we present the current state of TSE research and proposed future research directions. Some of the findings of this article are summarized as follows. We present model-driven approaches commonly used. We summarize the recent usage of detailed disaggregated mobile data for the purpose of TSE. The use of these models and data will raise a challenging problem due to the fact that conventional macroscopic models are not always consistent with detailed disaggregated data. Therefore, we show two possibilities in order to solve this problem: improvement of theoretical models, and the use of data-driven or streaming-data-driven approaches, which recent studies have begun to consider. Another open problem is explicit consideration of traffic demand and route-choice in a large-scale network; for this problem, emerging data sources and machine learning would be useful.

Keywords: highway operation, traffic state estimation, traffic flow model, numerical scheme, traffic data

1. Introduction

Traffic state estimation (TSE) refers to the process of inference of traffic state variables, namely flow (veh/h), density (veh/km), speed (km/h), and other equivalent variables, on road segments, using partially observed and noisy...
TSE plays an important role in traffic operations and planning. For example, traffic control, such as ramp metering, pricing, and information provision, requires precise traffic state information in order to mitigate congestion effectively. Strategic transportation planning such as infrastructure improvements also require traffic state information. These operation and planning tasks can be greatly improved by an efficient and accurate observation of the traffic state. However, the traffic state is not observed everywhere and practical measurements are usually noisy. Thus, the traffic state in unobserved areas needs to be estimated; and that in observed areas needs to be improved (denoised) as well.

In this article, TSE is defined as the simultaneous estimation of flow, density, and speed on road segments with high spatiotemporal resolution, based on partially observed traffic data and a priori knowledge of traffic. Fig. 1 illustrates a conceptual procedure of TSE.

This article focuses on TSE on highways motivated by the following four reasons. First, highways play a significant role in road transportation systems, with high service capability in terms of the volume and speed. Second, they exhibit some controllability (Papageorgiou et al., 2003) because of the nature of vehicular traffic. Third, recent technology developments enable new applications with the use of various and heterogeneous traffic data. Fourth, because of the aforementioned reasons, numerous highway TSE methods with various features have been proposed to date, which are worth considering and summarizing.

For our best knowledge, there is no comprehensive survey on TSE, although more than a decade has passed from the early and influential work on highway TSE by Wang and Papageorgiou (2005) and there have been considerable advances in this field. This makes it difficult to assess available approaches, respective benefits, and potential improvements.

The aims of this article are to provide a comprehensive and systematic summary of highway TSE, to contribute toward a better understanding of state-of-the-art methods, and to identify future research directions. Note that this article does not determine which methodology outperforms others in terms of general performance metrics. Respective advantages or disadvantages of a TSE method need to be discussed separately, by considering the specification of the application. One of the aims of this article is to provide the fundamental materials for such discussion. To achieve these aims, this article characterizes existing TSE methods based on their three fundamental elements: estimation approaches, traffic flow models, and input data. The following briefly describes these elements (we will discuss the details in the corresponding sections later).

---

1“High spatiotemporal resolution” loosely means spatial discretization on the order of few hundreds meters, and temporal on the order of a few minutes for practical purposes in transportation engineering.

2Several review articles from specific perspectives have been provided: a comprehensive review in 2005 (Wang and Papageorgiou, 2005, section 1) (we avoid overlapping with this one), data fusion in generic intelligent transportation system (ITS) research (El Faouzi et al., 2011), emerging traffic data collection methods in 2011 (Antoniou et al., 2011), coordinate system in traffic flow models for TSE (Yuan, 2013, chapter 2), TSE for specific ITS applications (Timotheou et al., 2015b, section 3), and application of vehicular ad hoc networks (Song and Lee, 2013; Darwish and Abu Bakar, 2015).
The estimation approach consists of methods that estimate the traffic state, based on a priori knowledge of traffic and partial observation. A priori knowledge (i.e., assumptions) could be physical traffic flow models and data-driven models, which is usually obtained by abstracting actual traffic by employing physical principles and statistical/machine-learning (ML) methods, respectively. In this study, the approaches are grouped into three categories, namely, model-driven, data-driven, and streaming-data-driven, according to types of a priori knowledge and input data they rely on. In short, model-driven ones rely on physical models of traffic which is characterized by empirical relation. Data-driven ones rely on dependence in historical-data and statistical/ML methods. Streaming-data-driven ones do not rely on these two previous elements. Because the assumptions vary greatly among the approaches, they have different advantages and disadvantages.

Traffic flow models describe the physical and theoretical aspects of traffic dynamics in the spatiotemporal domain. Therefore, they are used by model-driven TSE methods to infer the traffic state in an unobserved time–space region. Various models along with various solution methods have been proposed; and they have totally different advantages and disadvantages in terms of TSE.

The input data for TSE is the partial observation of traffic and is essential for TSE. Because of recent technology advances, the availability of novel data types (e.g., global positioning system (GPS), call detail record (CDR), on board diagnostics second generation (OBD2), etc.) is rapidly increasing in terms of both quality and quantity. This has resulted in the emergence of various TSE methods.

The remainder of this article is organized as follows. In Section 2, we briefly present highway traffic and TSE, along with fundamental terms and definitions. In Section 3, we summarize existing traffic models commonly used for TSE. In Section 4, we describe available measurements and data for traffic. In each sub- and sub-sub-section in Sections 3 and 4, general introduction on the corresponding topic and its application in TSE are discussed sequentially. In Section 5, we review TSE approaches, which mainly use models described in Section 3 and data described in Section 4. In Section 6, we summarize the survey results and propose future research directions.

Note that there are over 100 different methods to the TSE problem; thus, it is not practical to enumerate and explain them all in the core of this article. Therefore, only studies with substantial originality in terms of the aforementioned characteristics (i.e., approach, model, data) are referred in the main text. The remaining studies are summarized in Tables 1–4 together with those explained in the main text. Case studies which did not propose new TSE methods are not reviewed in this article.

2. Highway traffic and state estimation

In this section, we provide preliminary information about highway traffic and basic introduction of TSE. For systematic introduction to state-of-art of traffic flow and its theory, see, for example, Treiber and Kesting (2013) and Garavello et al. (2016).

2.1. Link

First, traffic on a link is considered. A link is defined as a road segment without an internal merge/diverge section, and is directional in the case of a highway.

At a microscopic scale, traffic on a link can be represented as a set of vehicle trajectories. A vehicle trajectory can be represented as $X(t, n)$ and $L(t, n)$, a position and lane-index, respectively, of vehicle $n$ at time $t$. The position is often defined as the distance along the road from its origin (i.e., postmile). Fig. 2d shows examples of vehicle trajectories in a freeway (US Department of Transportation, 2006) (sampling ratio: 1/33).

At a macroscopic scale, traffic on a link is often represented by a traffic state, which can be a subset of the following variables: flow $q$, density $\rho$, and average speed $v$.

$^3$Technically, a set of two variables from among flow or headway, density or spacing, and speed or pace (Laval and Leclercq, 2013) is a sufficient definition for a traffic state, since $q = \rho v$ holds.
\[ q(A) = \frac{d(A)}{|A|} \]  
\[ \rho(A) = \frac{t(A)}{|A|} \]  
\[ v(A) = \frac{\frac{d(A)}{t(A)}}{\frac{t(A)}{|A|}} \]  

where \( d(A) \) is the total distance traveled by all the vehicles in the region \( A \) (veh km), \( t(A) \) is the total time spent by all the vehicles in region \( A \) (veh h), and \( |A| \) is the time–space area of the region \( A \) (km h).

For practical scenarios, specific definitions are often used, such as point flow in space, space density, and space mean speed at a time point. The definition can be applied to either a single lane or multiple lanes in a link. Figs. 2a–2c show examples of traffic state variables, corresponding to the trajectories of Fig. 2d. The definition of \( q, \rho, v \) and the way they relate to particles (i.e., individual vehicles) are historically inspired by hydrodynamics.

Cumulative flow (also known as accumulated vehicle count, cumulative surface, Moskowitz function, etc.) is a unified representation of traffic where lane-changing is ignored. Cumulative flow \( N(t,x) \) is defined as the number of vehicles that passed position \( x \) by time \( t \). Continuous cumulative flow shapes a continuous “surface” in the vehicle–time–space \((n-t-x)\) domain, in which information about all the aforementioned traffic states are included (Moskowitz, 1965; Makigami et al., 1971). For example, a contour line of a cumulative surface, where the \( n \)-axis is the height, is a vehicle trajectory \( \bar{X}(t,n) \) if there is no lane-changing. In addition, partial derivatives of \( N \) represent the traffic state at the corresponding time–space point:

\[ q(t,x) = \partial_t N(t,x) \]  
\[ \rho(t,x) = -\partial_x N(t,x) \]  
\[ v(t,x) = \frac{q(t,x)}{\rho(t,x)} \]  

By taking limit of Eqs. (1) and (2) as size of \( A \) to be zero, they lead Eqs. (4) and (5). Fig. 2e shows an example of a cumulative flow for freeway data (contour interval: 33). Note that it slightly differs from Fig. 2d, as heterogeneity among lanes is averaged.

Fig. 2 shows an example of traffic state, vehicle trajectories, and cumulative flow in time–space diagrams based on the same traffic in NGSIM dataset (US Department of Transportation, 2006). Specific to this data, congestion (high density and low speed) was observed around \( t = 200, 300, \) and \( 700 \) (s) at \( x = 600 \) (m) and propagates spatially backwards as time progresses.

2.2. Network

A network is a set of links and nodes, in which the links are connected by the nodes. At a node, traffic may merge and/or diverge. Additionally, there are source and sink nodes (i.e., boundaries) where traffic demand is generated and attracted (except for a closed loop network, which is not practical). Fig. 3 illustrates the relation between links and nodes.

In a highway network, merge/diverge nodes correspond to junctions between multiple highway routes, and source/sink nodes correspond to on/off-ramps from/to arterial roads or boundaries of the spatial area of interest. At a diverge node, a driver of a vehicle has to determine which route s/he uses. At a source node, a driver has to determine when s/he enters the mainline. At a sink node, a driver has to determine whether s/he exits from the network. All these choices depend on her/his preferences, the travel itinerary, and traffic situation, which are not necessarily observable.

2.3. Traffic state estimation

2.3.1. Overview

In this article, TSE methods are characterized according to three elements: estimation approach, traffic flow model, and input data. The conceptual relation among these elements is shown in Fig. 4.
Figure 2: Time–space diagram of traffic state, vehicle trajectories, and cumulative flow. The data is based on US Department of Transportation (2006).

Figure 3: Sample road network.

Estimation approaches are grouped into three categories, namely, model-driven, data-driven, and streaming-data-driven, depending on assumptions (i.e., models) and input data. Missing/noisy data imputation like TSE always requires exogenous assumptions. Such assumptions can be “strong” or “weak” in a relative context. An assumption is “stronger” if it cannot be justified without extensive empirical validation. Examples of “strong” assumptions are traffic flow models which are developed based on empirical observation, and values of the model parameters which are subject of empirical calibration. A “strong” assumption is highly useful to estimate traffic state accurately, if the assumption is valid. On the other hand, excessive error will be caused if the assumption becomes invalid (e.g., a standard traffic flow model in a presence of a traffic accident). The problem is that, in general, it is not always possible to determine whether the assumption is valid or not under the practical TSE situation. It means that a TSE method with “strong” assumption can be vulnerable against unpredictable or uncertain traffic phenomena. Therefore, in general, it would be practically preferable if accurate TSE is possible based on “weaker” assumptions. In this article, traffic flow models with empirical relation and use of (statistical) dependency in historical-data are considered as “strong”
The model-driven approach are defined as an approach for TSE that uses a physical model describing traffic dynamics, namely, a traffic flow model. They estimate traffic states in unobserved areas using the model, and real-time data as its input. The models are usually based on empirical relation; the parameters describing the relation are either exogenously calibrated by using historical-data or endogenously estimated within the methods. The model-driven approaches have been widely used by existing TSE methods.

The data-driven approach are defined as an approach for TSE that extensively relies on historical-data, instead of physical traffic flow models. The approach extracts dependence between data from historical-data by using statistical/ML methods, and then estimates the traffic state based on real-time data and the dependence. This means that it does not require a priori knowledge of traffic being modeled explicitly as in physical traffic flow models. The approach generally requires large amount of historical-data.

The streaming-data-driven approach are defined as an approach for TSE that uses real-time (i.e., streaming) data only and is not strongly characterized by empirical relation which can be found in common traffic flow models (e.g., the fundamental diagram). It means that the approach less relies on a priori knowledge on traffic compared to the other two approaches (i.e., it only uses “weaker” assumptions). On the other hand, it may requires large amount of streaming data, and its accuracy may not be as high as that of model-driven or data-driven approaches if traffic was predictable.

Traffic flow models have been used by model-driven TSE methods as mentioned earlier. They describe traffic dynamics, such as propagation of congestion, mainly according to some physical principles and empirical relation. Various traffic flow models have been proposed to date and formulated under different assumptions. Various numerical

---

4A typical example of “strong” assumption is the existence of the fundamental diagram (FD) (c.f., Sections 3.1 and 3.2) of traffic, which consists of a functional form and parameter value of the FD. It is often assumed that functional form and parameter value of the FD are constant. If these assumption are valid, traffic dynamics can be well reproduced and accurate TSE is possible. In other words, accuracy of a TSE method with “strong” assumption may be high under ordinary traffic conditions. However, the parameter values can be changed by uncertain or unpredictable events such as so-called capacity drop and traffic accidents. Under such situations, accuracy of a TSE method with the constant FD (i.e., “strong” assumption) will be lower than a method with variable FD or without FDs (i.e., methods with “weaker” assumption). Therefore, the overall performance of a TSE method with “strong” assumptions is not necessarily better than that with “weak” assumptions; it depends on actual traffic situations.

Contrary, a typical example of “weak” assumptions is the conservation law of vehicles, which essentially means vehicles do not appear nor disappear without good reasons.
schemes to compute theoretical models have been proposed as well, and each scheme has unique advantages and disadvantages in terms of TSE.

Data is essential for all TSE methods. It is used as real-time data and/or historical-data by a TSE method. The real-time data is data collected at the moment when the unobserved traffic state is estimated, and is utilized for all TSE. The historical-data is data collected at the past for a long term, and is often used for model calibration or identification in model-driven and data-driven TSE. Fig. 5 shows the conceptual relation between TSE, real-time data, and historical-data. Because of recent advances in information and communication technology (ICT), a wide variety of data is available at the present time. Data types are grouped into two categories according to the measurement methodology: stationary and mobile. Stationary data is collected by fixed sensors, such as inductive loops, and can be considered as conventional. Mobile data is collected by mobile sensors, such as on-vehicle GPS devices, and is relatively new. With the emergence of connected and automated vehicles, this kind of data is becoming an increasing source of traffic data. Additionally, we introduce another independent categorization: aggregated and disaggregated. In aggregated data, information from multiple vehicles is aggregated (e.g., 5 min time interval) and stored. In disaggregated data, it is not aggregated and is stored as collected; therefore, it contains more information than the aggregated data, and may require advanced techniques to be used. Various techniques have been developed to use this data.

In the following sections, traffic flow models are first discussed in Section 3. Then, data is discussed in Section 4, as the use of specific data requires some knowledge about the models. Finally, approaches are discussed in Section 5, because they integrate models and data to estimate the traffic state.

2.3.2. Application

TSE methods are generally applied to traffic operations. The most notable application is traffic control mainly to mitigate traffic congestion. Examples of control on the highway are ramp metering, variable speed limit, pricing, and information provision for route choice guidance. Each of these requires precise information about traffic states (Papageorgiou et al., 2003). Results of TSE can be also used for strategic planning, such as road construction.

3. Traffic flow models

In this section, we introduce the different frameworks used for traffic flow models and review their applications to TSE. Each section describes general introduction on each topic first, and then reviews its application to TSE. In

Figure 5: Traffic data and its spatiotemporal relation.
Section 3.1, we explain a fundamental concept in traffic flow theory: the fundamental diagram (FD). In Section 3.2, we present the models that describe traffic in a link, and then in Section 3.3, we briefly describe the models for nodes.

3.1. Fundamental diagram

The FD, also known as a speed–density relation, flow–density relation, or flux function, is one of the most fundamental concepts in traffic flow theory. It describes the empirical relation among traffic state variables in stationary traffic (also referred as steady or equilibrium traffic), which is an ideal state in which all vehicles have the same constant speed and spacing. The FD appears in almost all theories of traffic flow as it relates the traffic state variables and contains remarkable information about traffic characteristics, such as free-flow speed and capacity (Greenshields, 1935) and is inherently linked to flux functions in hydrodynamic theory (Lamb, 1932; Evans, 1998).

Fig. 6 shows a flow–density relation in actual traffic. The typical FD theory assumes that the traffic state \((q, \rho, v)\) in stationary traffic satisfies

\[
v = V(\rho),
\]

or equivalently,

\[
q = \rho V(\rho) = Q(\rho),
\]

where \(V\) represents the speed–density FD and \(Q\) represents the flow–density FD. The FD has two essential elements: functional form and parameters values. The elements depend on various factors, such as spatial conditions (e.g., road geometric alignment and road surface), temporal conditions (e.g., time of day, weather, accident), and vehicle conditions (e.g., hardware performance and driver characteristics).

Various functional forms for FDs have been proposed. A triangular FD (Newell, 1993a) is one of the most popular forms because of its simplicity, theoretically preferable features (c.f., Section 3.2), and some empirical evidence (Cassidy, 1998). Fig. 7 shows examples of a triangular FD in various planes: the blue curves represent free-flowing regimes and red curves represent congested regimes. Other forms have been proposed and used as well, including differentiable, noncontinuous, or multivalued forms. For the details, see recent reviews by Traffic Flow Theory and Characteristics Committee (2011); Carey and Bowers (2012); van Wageningen-Kessels et al. (2014).

In fact, traffic characteristics may vary depending on time, vehicle, and location. This is called traffic heterogeneity. Therefore, technically, the FD function can be expressed as \(Q(\rho, t, n, x)\). For example, lane-drops can be represented by a space\((x)\)-varying FD. Difference between passenger vehicles and heavy trucks can be represented by a vehicle\((n)\)-varying FD. Weather change can be represented by a time\((t)\)-varying FD. Capacity drop at sags is known to be affected by spatial and driver conditions; therefore, it can be represented by a vehicle-space\((n, x)\)-varying FD. However, depending on importance of heterogeneity to the application, some of the dependencies are simplified or neglected to enable efficient calibration and estimation.
Identifying the values of the FD parameters has attracted a significant amount of interests over the past decades. Most existing calibration methods are based on stationary sensors (e.g., Dervisoglu et al., 2009; Chiabaut et al., 2009; Coifman, 2014; Zhong et al., 2015), and some studies have considered mobile data, such as the GPS probe vehicles with some assumptions (Herrera et al., 2010; Seo et al., 2016a) and probe vehicles with additional sensors (Seo et al., 2015b).

**Application to TSE:** Because the FD relates traffic state variables to each other and is a core of traffic flow theory, it plays a significant role in most of TSE methods. If a TSE method requires an FD, its functional form is given considering its requirement and a priori knowledge of the fitness to empirical data. Then, its parameters are either calibrated based on a priori knowledge/data (i.e., offline calibration) or jointly estimated with the traffic state (i.e., online calibration, adaptive estimation). The former makes the problem simple and easily solved; therefore, many studies used this approach. However, such methods can be disrupted if the FD changes dynamically. The latter is robust against such phenomena. For example, Wang et al. (2009) performed adaptive TSE by considering the time and space dependencies of an FD (i.e., $V(\rho, t, x)$) to capture the effect of traffic accidents.

Conversely, some of the data-driven methods and streaming-data-driven methods do not rely on FDs at all (e.g., Chen et al., 2003; Coifman, 2003; Seo et al., 2015a; Bekiaris-Liberis et al., 2016); thus, they require less—a priori knowledge. This means that they can be robust against uncertain phenomena and unpredictable incidents, and require less—or no—calibration.

3.2. **Link models**

Traditionally, traffic on a link is modeled using partial differential equation(s) (PDE) representing a system of conservation law (CL) in hydrodynamic flow modeling, which describes aggregated behavior of traffic by approximating it as continuum fluid (and its discretization). The model uses the concept of the FD (c.f., Section 3.1) as a basis of its flux functions. Because of its continuum fluid approximation, the model is often called a macroscopic
Fig. 8 shows a conceptual example of traffic flow, its continuous PDE model, and discretized numerical scheme. 

In Section 3.2.1, we introduce typical (seminal) traffic flow models. In Section 3.2.2, we briefly describe the extension of the basic models. In Section 3.2.3, we explain the solution methods (i.e., numerical schemes) for the models.

3.2.1. One-dimensional flow models

The Lighthill–Whitham–Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956), also known as a first-order model, is the most seminal work in macroscopic traffic flow modeling. The LWR model assumes a CL and that the traffic state is always equilibrated following an FD. It can be formulated as:

\[
\begin{align*}
\partial_t \rho + \partial_x (\rho v) &= 0, \\
v &= V(\rho)
\end{align*}
\]

or equivalently,

\[
\partial_t \rho + \partial_x Q(\rho) = 0.
\]

There are other models, such as microscopic models (e.g., car-following models and cellular automata), mesoscopic models (e.g., gas-kinetic models), and others (Helbing, 2001; Hoogendoorn and Bovy, 2001). However, they have not been used for TSE in the past because of their complexity and high computational cost, with some exceptions that are, in fact, equivalent to macroscopic models.
Eq. (9a) represents a CL and Eq. (9b) represents the FD. The model can reproduce basic phenomena of traffic, such as a distinction between a free-flowing regime and congested regime, and simple behavior of a traffic jam and shockwave. Additionally, the model can be calibrated and solved efficiently because of its simplicity. However, several limitations have been pointed out. The equilibrium assumption (9b) means that only traffic states following the FD are allowed to exist. In addition, this model makes vehicles change their speed instantaneously; obviously, this is a crude approximation of vehicle dynamics. These limitations and the mathematical features of this model make it impossible for the model to reproduce well-known phenomena, such as the capacity drop, hysteresis effect, traffic instability, stop-and-go waves, etc.

Higher-order models have been developed to overcome the limitation of the LWR model. To represent shifts (also referred as perturbation, non-equilibrium state) from the equilibrium traffic state, these models generally use an additional equation (referred as the momentum equation) that describes the dynamics of $v$, instead of the FD equation (9b) in the LWR model. The Payne–Whitham (PW) model (Payne, 1971; Whitham, 1974) is the first well-known attempt at higher-order modeling. It can be formulated as

$$\partial_t \rho + \partial_x (\rho v) = 0, \quad (11a)$$
$$\partial_t v + v \partial_x v = -\frac{v - V(\rho)}{\tau} - \frac{c_0^2}{\rho} \partial_x \rho, \quad (11b)$$

where $\tau$ is the relaxation time and $c_0^2$ is a parameter related to driver anticipation. Eq. (11b) is the momentum equation. The PW model and its extensions (e.g., Papageorgiou et al., 1989) succeed in reproducing complex traffic phenomena to some extent. However, it may produce physically inappropriate behavior, such as negative speed and information propagating faster-than-vehicle speed (for the details, see the discussion by Del Castillo et al. (1994); Daganzo (1995); Papageorgiou (1998); Hoogendoorn and Bovy (2001)).

The Aw–Rascle–Zhang (ARZ) model (Aw and Rascle, 2000; Zhang, 2002) is another well-known higher-order model. It allows traffic states to shift from the equilibrium, and it overcomes the aforementioned limitation of the PW model. It can be formulated as

$$\partial_t \rho + \partial_x (\rho v) = 0, \quad (12a)$$
$$\partial_t (v - V(\rho)) + v \partial_x (v - V(\rho)) = -\frac{v - V(\rho)}{\tau}. \quad (12b)$$

Eq. (12b) is the momentum equation. The ARZ model converges to the LWR model at $\tau \to 0$; therefore, it can be regarded as a generalization of the LWR model (in the physical sense). Further generalization and extension of the ARZ model have been proposed, such as generic second order models (Lebacque et al., 2007), phase transition models (PTM) (Colombo, 2003; Blandin et al., 2011), and the generalized ARZ model (Fan et al., 2014).

Recently, there have been theoretical advances in the interpretation of traffic flow models. That is, the connection between the some of the models and the Hamilton–Jacobi (HJ) PDE, which is well studied in PDE and physics communities (Evans, 1998). For example, the conventional LWR model (10), in which the state variable is $\rho(t, x)$, can be transformed into following equation where the state variable is the cumulative flow, $N(t, x)$ (Newell, 1993a,b,c):

$$\partial_t N - Q(-\partial_x N) = 0. \quad (13)$$

This equation is an HJ PDE whose Hamiltonian is the FD. Therefore, for example, the model can be efficiently solved by using techniques in HJ PDE theory (Aubin et al., 2008; Claudel and Bayen, 2010a,b). In addition, because of the nature of the cumulative flow (c.f., Section 2.1), the model can describe vehicle trajectories and travel time explicitly. It shows an essential connection between the macroscopic traffic flow model and microscopic vehicle behavior (e.g., car-following model); this can be considered as a preferable feature for a physical model. This is also a preferable feature to exploit mobile data (i.e., trajectories of probe vehicles) in a straightforward manner (for the details, see Section 4.2). Based on this theory, new coordinate system called Lagrangian coordinates (the conventional is Eulerian coordinate)

\footnote{For the proper mathematical link to HJ and viability framework for Eq. (13), see Aubin et al. (2008); Claudel and Bayen (2010a,b).}
system based on \((t,x)\) is proposed to use the cumulative flow and vehicle trajectories in macroscopic traffic flow model (e.g., Newell, 2002; Daganzo, 2006; Leclercq et al., 2007; Laval and Leclercq, 2013). Similar topics for other traffic flow models have also been discussed (Aw et al., 2002; Jin, 2016).

**Application to TSE:** The aforementioned models have been extensively applied for model-driven TSE. The LWR model is the most widely applied (e.g., Haj-Salem and Lebacque, 2002; Muñoz et al., 2003). One of the reasons for this is that the model can be calibrated and solved efficiently. Another reason is that the model can explicitly manage vehicle trajectory data, which has received a great deal of attention recently (e.g., Claudel and Bayen, 2008); in fact, this feature is also because of the model’s simplicity (c.f., Sections 3.2.3, 4.2). Some studies have used part of the LWR model’s implications (i.e., propagation speed of a shockwave) for TSE (e.g., Coifman, 2002; Treiber and Helbing, 2002). Regarding the HJ PDE version of the LWR model, some recent studies have applied the theory of HJ PDE or the Lagrangian coordinate system (e.g., Coifman, 2002; Claudel and Bayen, 2008; Yuan et al., 2012) in order to utilize disaggregate vehicle trajectory data. Besides, higher-order models are used in TSE studies to estimate the non-equilibrium traffic state and/or acceleration rates, which are significantly important for some traffic operations and ignored by the LWR model. For example, the PW model and its extension (e.g., Nanthawichit et al., 2003; Wang et al., 2006), higher-order model modified from the LWR model (e.g., Mihaylova and Boel, 2004), ARZ model and its extension (e.g., Wilkie et al., 2013; Fan et al., 2014), and several PTMs (e.g., Blandin et al., 2013; Piccoli et al., 2013) have been used.

3.2.2. Extensions

Several extensions of one-dimensional models have been proposed. In this section, we describe multi-lane, multi-class, and stochastic modeling.

Actual traffic is often multi-lane and multi-class; however, the models described in Section 3.2.1 are based on single-lane and single-class. To account for this problem, some link models that explicitly consider multi-lane and/or multi-class behavior of traffic have been developed. These extensions are quite advanced and technical topics; therefore, refer, for example, to Wong and Wong (2002), Daganzo (2002), Laval and Daganzo (2006), van Wageningen-Kessels et al. (2014), and Costeseque and Duret (2016) and references therein for the case of the LWR model. One of the main challenges for such extensions is modeling lane-changing behavior, in which human behavior is essentially important. Nevertheless, it should be noted that the ARZ model and its extension can be considered as a multi-class LWR model because it can be interpreted as the LWR model with vehicle-specific FDs.

Another important aspect of traffic flow modeling is the stochastic factors. Stochastic models have been developed to explain the uncertainty of traffic dynamics and effect of the input data’s uncertainty on the model’s output. For example, Sumalee et al. (2011) developed a stochastic model based on the LWR model that considered both factors. In models of Jabari and Liu (2012) and Jabari et al. (2014), the source of stochasticity in traffic dynamics accounted for vehicle heterogeneity, which implied a relation between multi-class and stochastic models. Wada et al. (2017) developed a stochastic model based on the HJ PDE version of the LWR model.

**Application to TSE:** Some studies have investigated multi-class and/or multi-lane TSE. For example, van Lint et al. (2008); Ngoduy (2008) proposed multi-class TSE methods, Chapter 4 of Yuan (2013) proposed a multi-class TSE method in Lagrangian coordinates, and Lovisari et al. (2015) developed a multi-lane TSE method. Coifman (2003) proposed a multi-lane TSE method that is not model-driven. Note that single-class single-lane TSE models are not necessarily impractical in multi-class multi-lane environments; because their results can be interpreted as an “average state” of multi-class multi-lane traffic (Papageorgiou, 1998). However, for the case of mobile data (c.f., Section 4.2), ignoring multi-class behavior can cause a biased estimation because data might be sampled from biased classes. Additionally, rich disaggregated data (c.f., Section 4.2) is possibly problematic for single-class single-lane models because of, paradoxically, its richness (Seo and Kusakabe, 2015). For example, if a probe vehicle overtakes another, a singularity will arise in the single-lane Lagrangian coordinate.\(^8\) To the best of the authors’ knowledge, no studies have solved this problem explicitly.

Several TSE studies have used stochastic models to capture estimation uncertainty explicitly. For example, Sumalee et al. (2011), Laval et al. (2012), and Jabari and Liu (2013) proposed TSE methods based on their stochastic

\(^7\)e.g., other HJ PDEs such as \(\partial_t X(t,n) - V(-\partial_n X(t,n)) = 0\) which is equivalent to Eq. (13). Notice that \(X(t,n)\) is a trajectory of vehicle \(n\), \(\partial_t X(t,n)\) is its speed, and \(\partial_n X(t,n)\) is its spacing (headway distance).

\(^8\)If the number of probe vehicles is small, this problem rarely occurs; thus, ad hoc data cleansing can work. Otherwise, it can be troublesome.
models. Meanwhile, instead of analytically deriving stochastic models from their sources, many studies have constructed stochastic models by simply adding random noise into the numerical schemes of the deterministic traffic flow models. In the latter studies, Monte Carlo simulation is often used to compute the model easily (Section 5.1.2), so that analysis on theoretical properties of the models—which is generally challenging—are not required.

### 3.2.3. Numerical schemes

Numerical schemes such as finite difference methods (FDMs) are common to solve traffic PDEs numerically. Typically, state variables in a domain within given boundaries can be computed based on given boundary conditions using the proper numerical schemes. For proper approach to this problem, one needs to consider the appropriate discretization of the corresponding boundary value problem (BVP) (Evans, 1998; Garavello et al., 2016).

Standard numerical schemes based on FDMs (c.f., LeVeque, 1992) has been widely used for computing numerical solutions to these problems, such as the Godunov scheme (Lebacque, 1996), an upwind scheme (Leclercq et al., 2007), Lax–Friedrichs scheme (i.e., first order scheme) (Wong and Wong, 2002; Göttlich et al., 2013), and Lax–Wendroff scheme (i.e., second-order scheme) (Michalopoulos et al., 1993). In particular, to compute PW-like models, specific FDMs for the models have been extensively used (Papageorgiou et al., 1989). However, it is known that the first-order schemes are diffusive in general, which can be problematic for accuracy when reproducing heads and ends of traffic jams; and the second-order schemes oscillate. Nevertheless, the Godunov scheme (Godunov, 1959) has been widely applied for traffic flow models, such as the LWR model (Daganzo, 1994; Lebacque, 1996) and ARZ model (Mammar et al., 2009) because of its accuracy and generality. In particular, a simplified case of the scheme for the LWR model, namely one with triangular or trapezoidal FDs and discretized with the Courant–Friedrichs–Lewy (CFL) number equal to 1, is often referred to cell transmission model (CTM) (Daganzo, 1994) in transportation engineering and has been widely used because of its reasonable computational cost, accurate shock wave representation, and its solid theoretical basis. However, it is not differentiable in general and therefore may cause technical problems for some TSE approaches (c.f., Section 5.1.2), while other schemes such as upwind and Lax–Friedrichs are differential (Blandin et al., 2012a). For second-order models, Delis et al. (2014) proposed high-resolution methods, which are more accurate and do not oscillate.

Switching mode models (SMMs) are efficient computation methods for solution of the LWR model (e.g., Muñoz et al., 2006). They are piecewise linearization of the CTM and consist of a discrete set of “modes” that represents the current regime (free-flowing or congested) of traffic on road segments. By using an SMM, traffic dynamics can be efficiently computed once the current mode is identified in some manner. Several extensions have been proposed following this idea, such as considering uncertainty (e.g., Sumalee et al., 2011; Morbidi et al., 2014) and reducing the size of a set of modes to enable large-scale computation (e.g., Canudas de Wit et al., 2012; Thai and Bayen, 2015).

For the LWR model, the exact solution can be efficiently obtained by using techniques in the HJ PDE theory, such as the Hopf–Lax (HL) formula (Evans, 1998). Several methods have been developed in a similar manner, for example, exact methods with triangular FD, known as Newell’s minimization principle and variational theory (VT) within the transportation community (Newell, 1993a,b,c; Daganzo, 2005a,b), a method that can manage internal boundary conditions based on the HL formula (Claudel and Bayen, 2010a,b), and an exact method with arbitrary FD (Mazaré et al., 2011). Solution methods in Lagrangian coordinates have also been developed (Newell, 2002; Leclercq et al., 2007; Han et al., 2012), which are, in fact, equivalent to VT/HL in Eulerian coordinates (Daganzo, 2006; Laval and Leclercq, 2013). For the details on these methodologies, see Daganzo (2006), Aubin et al. (2008), Claudel and Bayen (2010a,b), and Laval and Leclercq (2013).

Regarding boundary conditions, the most common boundary of a domain is a rectangle in the model’s coordinate system, for example, two locations and two time instants in an Eulerian coordinates system, and two vehicles’ trajectories and two locations or two time instants in Lagrangian coordinates systems. More complex boundaries, such as an arbitrarily closed loop (Daganzo, 2005a) and internal boundaries (Claudel and Bayen, 2010a), are sometimes defined if necessarily. If the domain is on a road network, the flow from sink/source nodes can be also be a boundary condition (c.f., Section 3.3).

**Application to TSE:** The most commonly used schemes in TSE studies are as follows: SMMs for efficient computation (e.g., Muñoz et al., 2003), the Godunov scheme (e.g., Haj-Salem and Lebacque, 2002; Mihaylova and

---

9In the case of the LWR PDE and extensions, FDMs have been than more prevalent than finite volume methods. An example of few exceptions is Tchrakian and Zhuk (2015) which applied a finite volume method to LWR-based TSE.
Boel, 2004), the FDM for PW-like models (e.g., Nanthawichit et al., 2003; Wang and Papageorgiou, 2005), and HL/VT for the exact computation of the LWR model (e.g., Claudel and Bayen, 2008; Laval et al., 2012). Choosing a numerical scheme requires considering estimation approaches and input data. For example, some accurate numerical schemes (e.g., the Godunov scheme) cannot be incorporated into certain computationally efficient estimation approaches (Blandin et al., 2012a). Using disaggregated mobile data requires specific schemes (e.g., HL, VT, Lagrangian coordinates) because of their different nature compared with stationary data. For the details, see Sections 4.2 and 5.1.2. Some recent studies have begun to consider new options, such as the Fourier–Galerkin (FG) method and discontinuous Galerkin (DG) method for minimax estimation (Tchrakian and Zhuk, 2015; Tchrakian et al., 2015) and spectral method (Xia et al., 2017).

3.3. Node models

Node models, also referred to as junction and intersection models, are required to model traffic on a network. They connect links and determine merge and diverge flows that depend on the states of the links. In the case of highways, junctions connecting multiple highway routes and on/off-ramps are considered as nodes. Additionally, a multi-lane link is sometimes modeled as a network of single-lane links (e.g., Laval and Daganzo, 2006). Several models that correspond to different link models have been proposed. Essentially, a node model has to determine and integrate two factors: supply (i.e., receiving flow) of the downstream links and demand (i.e., sending flow) of the upstream links (Lebacque, 1996; Jin et al., 2009). Supply is the maximum flow that is received by the downstream link. It is determined by the geometric structure of the node and links and traffic situation in the downstream link. Demand is the maximum flow that is sent from the upstream link. It is determined by the geometric structure, traffic situation in the upstream link, and driver behavior such as route choice. If a node is at the boundary of the network, boundary demand (e.g., on-ramp flow from arterial street) should be considered; its explicit modeling requires driver behavior models, again. Therefore, this problem is still an open question because the many factors are dominated by human behavior, which is uncertain due to possible irrationality and incomplete information for both of driver and road administrator (Bliemer et al., 2017). For the details on node models and networks, see Garavello et al. (2016) and references therein.

Application to TSE: In the context of TSE, the node model has been paid little attention. Although several studies have investigated TSE in a large-scale network using certain node models (e.g., Wang et al., 2006; Zuurbier et al., 2006), the importance of node models for TSE has not been investigated fully and is still the focus of ongoing work, with the changing landscape of traffic information (Garavello et al., 2016). In particular, node models that describe route choice at nodes that are responsive to the traffic situation and information (e.g., drivers avoiding congestion) have not been incorporated into TSE. This might be because such route choice behavior is uncertain and makes the problem complex (Tampère and Immers, 2007). Regarding boundary demand, some studies have incorporated demand estimation/prediction in their TSE methods, and the importance of demand estimation has been pointed out (e.g., Wang and Papageorgiou, 2005; Wang et al., 2006; Zhang and Mao, 2015).

4. Data

In this section, we introduce available traffic measurement data used for TSE. In general, they can be grouped into two categories based on their nature: stationary and mobile data. In addition, we introduce another independent categorization: disaggregated and aggregated. Fig. 9 illustrates this categorization. These types of data are utilized in either streaming and historical manner in TSE. For practical perspectives on traffic data collection, see Leduc (2008).

4.1. Stationary data

Stationary sensors (e.g., inductive loop detectors, ultrasonic detectors, radar detectors, and closed-circuit television cameras) can collect traffic data that consists of every vehicle at their installed locations. Using typical loop detectors, traffic counts and vehicle occupancy at the location can be measured. The instantaneous vehicle speed at the location is also measured, or inferred from a given vehicle’s body length information. Dual detectors and cameras can measure vehicle trajectories on short road segments directly. These raw data are generally disaggregated data, in which an individual vehicle’s information (i.e., trajectory or cumulative flow in the observable area) is contained. Therefore, it can be represented as \( N(t, x) \) at a certain \( x \) for all \( t \) (Fig. 10a). However, because of limitations on data handling (e.g., storage and transmission), the data provided is often converted to aggregated data, in which the average traffic state...
at a certain temporal resolution (e.g., 1–5 min) is contained. Such aggregated data can be regarded as “irreversibly compressed” from corresponding disaggregated data. It can be represented as \( q(A) \), \( k(A) \), and \( v(A) \) of Eqs. (1)–(3) where the size of rectangular (on time–space diagram) \( A \) is the aggregation unit Fig. 10b. Stationary data is also known as Eulerian data. These stationary sensors are widely used to collect highway traffic data by road administrators and researchers since the era of Greenshields (1935). However, the amount of data they provide is not always sufficient for traffic control,\(^{10}\) due to sparse deployment of the sensors (which is due to high installation and maintenance costs of the sensors). For example, usual detector spacing varies from several hundred meters to several kilometers. Moreover, some highways do not have any detectors, especially in rural areas and developing countries. Additionally, their accuracy and precision may be not reliable, for example, because of frequent misses and/or double counting of loop detectors (Chen et al., 2003).

Application to TSE: Because stationary data are conventional and commonly available on highways in practice, the vast majority of TSE studies use this data. The Eulerian nature of the data is beneficial to be incorporated to the Eulerian traffic flow models which are conventional and common, too. In addition, stationary sensors are often able to measure all traffic state variables; therefore, TSE based only on these data has been well studied (e.g., Muñoz et al., 2003; Wang and Papageorgiou, 2005). Most of the studies have used aggregated data, whereas few have used disaggregated data (e.g., Coifman, 2002; Deng et al., 2013). Coric et al. (2012) proposed methods of reconstructing high-resolution data from aggregated detector data, and then investigated effects of reconstruction to TSE accuracy.

4.2. Mobile data

Mobile data is data associated with specific vehicles that usually provides measurements along vehicle trajectory. A typical method to collect mobile data is to use on-vehicle sensors (e.g., GPS, OBD2, and most recently video cameras, radars, etc).\(^{11}\) Such vehicles with sensors are often referred to as probe vehicles or floating cars. Mobile data collection has been widely enabled by recent advances in ICT (e.g., navigation systems and smartphones). Typical disaggregated mobile data can be represented as \( X(t, n) \) for all or specific \( X \), \( n \), and \( t \). Therefore, it is also known as Lagrangian data. Although the use of probe vehicles has received a great deal of attention from the mid-1990s (Sanwal and Walrand, 1995; Zito et al., 1995), they have been re-recognized as “connected vehicles” in the context of the Internet of Things and vehicle automation at the present time (Bekiaris-Liberis et al., 2016).

A promising feature of mobile data is that it is possible to collect data from a wider spatiotemporal domain, compared with stationary sensors (Zito et al., 1995; Herrera et al., 2010), because probe vehicles ultimately have

\(^{10}\)E.g., usual traffic control measures require accurate traffic state information with space resolution of a few hundred meters (Wang and Papageorgiou, 2005).

\(^{11}\)Use of airplanes and satellites (e.g., Coifman et al., 2006) is notable alternative approach to collect vehicle trajectory in a large area. However, data collected by such approach is not considered as “mobile data” in this article—in fact, such data is similar to “fixed data”, in the sense that it contains trajectories of all vehicles (i.e., it is not sampled data associated with vehicles).
Figure 10: Illustration of traffic data on time–space diagrams.
broader coverage of the road network. The data is continuously collected while the probe vehicles are driving. Two important characteristics of probe vehicle data are penetration rate and temporal sampling rate. The penetration rate is defined as the ratio of probe vehicles to all traffic. The temporal sampling rate is defined as the time interval for the consecutive reporting of data. In general, it is preferable if penetration rate (in proportion) and temporal sampling rate (in frequency) are high.

GPS (i.e., global navigation satellite system) is the most typical on-vehicle sensor (Herrera et al., 2010). Its raw data is generally disaggregated data, which is an individual probe vehicle’s continuous trajectory (Fig. 10c). However, aggregated data, which consists of average speed with a certain spatiotemporal resolution (Fig. 10d), is widely used in practical situations. Fig. 11 illustrates this kind of mobile data. The sampling and aggregation policies are in general driven by privacy consideration (Krumm, 2009; Hoh et al., 2012).

Other advanced on-vehicle sensors, such as OBD2, can also be used to collect mobile data. An example is a range finder (e.g., radar, lidar, and a stereo camera), which is now being equipped for vehicle automation purposes, known as advanced driver assistance system (ADAS) (Bengler et al., 2014) and autonomous vehicles. It provides novel information about traffic; for example, it can observe local traffic density, which is approximately the inverse of bumper-to-bumper distance, near a probe vehicle (Seo et al., 2015a). These new data types, in addition to conventional positioning data, have been named extended floating car data (xFCD) (Huber et al., 1999).

The connection to vehicular ad hoc networks (VANET), by which a probe vehicle can count and report the number of nearby connected vehicles (Darwish and Abu Bakar, 2015), is utilized to collect and transmit traffic data. If all the vehicles are equipped with VANET and their connectivity is high, the traffic density can be directly determined (e.g., Panichpapiboon and Pattara-atikom, 2008).\footnote{In this kind of VANET-based TSE, an essential problem is the connectivity of VANET, rather than traffic flow models or estimation approaches which are the focus of this article. Such connectivity is a highly technical problem; thus, it is out of scope for this article. For the details, see survey articles by Song and Lee (2013); Darwish and Abu Bakar (2015), dedicated to this and other VANET-specific problems.}

Travel time information (TTI) of a vehicle between distant locations can be considered as disaggregated probe vehicle data with a long temporal sampling interval. It is collected by using either automatic vehicle identification system\footnote{It matches a vehicle at distant locations based on the vehicle’s features, such as license plate number (Kanayama et al., 1991), electromagnetic features of the vehicle’s body (Cofman, 1998), and visual features (Sumalee et al., 2012)} or mobile communication instruments.\footnote{E.g., media access control (MAC) address (Wasson et al., 2008; Barceló et al., 2010), CDR (call digital records) (Caceres et al., 2007; Wu et al., 2015) and similar data from mobile phones (Asakura and Hato, 2004)} Especially, the CDR and MAC address data has been paid great deal of attention because of wide spread of mobile phones in these days and their less dependency to infrastructure. The travel time is measured between either within-link locations (i.e., multiple dots in a link) or inter-link locations (i.e., one or zero dot in a link). The latter has unique feature compared with other data. That is, travel route is not necessarily identified and therefore required to be inferred (Wu et al., 2015). Because of the TTI’s wide-ranging...
nature, it has been used for large-scale estimation problems, such as travel time measurement (e.g., Coifman, 1998), origin-destination matrix estimation (e.g., Asakura et al., 2000; White and Wells, 2002; Caceres et al., 2007; Barceló et al., 2010), and route choice estimation (e.g., Wu et al., 2015; Yang and Sun, 2015).

A problem with mobile data is that its sampling and potential biases. For example, if probe vehicles belong to a logistic fleet, they may travel at a slower speed than average. In addition, vehicles with recent advanced driving technologies (e.g., ADAS and connected vehicles), which may tend to act as probe vehicles, may show different driving characteristics compared with fully manual vehicles, and progressively change driving and traffic patterns.

**Application to TSE:** Mobile data has received a great deal of attention in TSE recently. Some studies have used both mobile and stationary data, others have mainly used mobile data together with limited stationary data, and a few studies have used mobile data only. Because of its nature (i.e., sampled data), mobile data is used in notably different manners compared with conventional TSE methods using stationary data.

For model-driven TSE, incorporating some of the mobile data (i.e., GPS probe vehicle and TTI) requires special techniques. This is because such mobile data only contain average speed or sampled trajectories data. This means that other traffic state variables (i.e., flow and density) cannot be determined from it without additional assumptions. Moreover, in conventional traffic flow models (e.g., the LWR model with an Eulerian coordinate system), the state variable is density; this may be due to that the conventional models have been developed based on empirical observation from the stationary sensors. Unfortunately, it is known that mapping from speed to density via an FD can be erroneous, especially in free-flowing regimes (Herrera and Bayen, 2010). To account for this problem, several approaches have been proposed. For aggregated GPS probe vehicle data, Nanthawichit et al. (2003) used a second-order traffic flow model to combine probe vehicle data with stationary data. Work et al. (2008, 2010) developed a first-order traffic flow model whose state variable is speed. Kesting and Treiber (2009) introduced a shock wave detection method based on probe vehicle data. van Lint and Hoogendoorn (2010) and Treiber et al. (2011) used an adaptive smoothing filter (ASF) to combine probe vehicle data with stationary data. Herrera and Bayen (2010) implemented the Newtonian relaxation method to use the erroneous FD-based mapping. For disaggregated GPS probe vehicle data, Claudel and Bayen (2008, 2011) developed an HL-based incorporation method, which considered probe vehicle data as internal boundary conditions. Yuan et al. (2012) used the LWR model with the Lagrangian coordinate system. Deng et al. (2013) and Kuwahara et al. (2013) defined cumulative flow as state variables using VT. Piccoli et al. (2013, 2015) developed mapping methods from trajectories to traffic states based on PTMs. Xia et al. (2017) formulated a method that could explicitly consider vehicle trajectories in the Eulerian LWR model. Bucknell and Herrera (2014) investigated effects of penetration rate and temporal sampling rate of GPS probe vehicles to accuracy of a TSE method.

Some data-driven methods have used aggregated GPS probe vehicle data. Essentially, they first extracted an FD or similar relations from historical stationary data, and then estimated traffic states from mobile data and the FD-like relation (e.g., Blandin et al., 2012b; Neumann et al., 2013). Wilby et al. (2014) proposed a TSE method using xFCD with a similar approach.

All the aforementioned methods require calibrated FDs or similar relations. Therefore, even if they mainly use mobile data, they essentially require stationary data for FD calibration (c.f., Section 3.1). This implies that the application of these studies in wide-ranging areas is still difficult. To account for this problem, FD-free TSE using mobile data has been investigated. Particularly, streaming-data-driven TSE methods using conventional mobile data and limited stationary data (e.g., detectors at boundaries) have been proposed by Coifman (2003); Astarita et al. (2006); Qiu et al. (2010); Bekiaris-Liberis et al. (2016). They rely on the CL of traffic, which is likely to be satisfied under common traffic situations. Other studies have investigated the use of advanced mobile data with no stationary data. Wardrop and Charlesworth (1954) proposed the use of xFCD (manually collected at that age) to estimate the traffic state—in fact, this might be the most ancient TSE study using mobile data. After recent technological advances, some applications of advanced mobile data have been proposed. Seo et al. (2015a) proposed a streaming-data-driven TSE method using xFCD. Then, Seo and Kusakabe (2015) incorporated the method into a CL. To the best of the authors’ knowledge, these are the only methods that implement TSE with mobile data only. Seo et al. (2015b) extended the method as a model-driven method, where its FD was endogenously calibrated.

---

15In fact, this implies that the equilibrium assumption in the LWR model is not very accurate.

16Exceptions are VANET-based TSE methods, which assume that all vehicles report their position via VANET (e.g., Punichpapiboon and Pattara-atikom, 2008).
Subject to TTI, within-link matching information has been used in some studies. Cheng et al. (2006) used CDR to derive traffic speed; however, they reported it caused excessive error. Gundlegard et al. (2015) developed a method of combining mobile datasets with various temporal sampling rate, such as TTI and GPS, for accurate estimation. Deng et al. (2013) combined TTI with other mobile and stationary data, considering the heterogeneity of these different data. He et al. (2016) used CDR to determine the average speed on a long highway route. Coifman (2003) used TTI in his streaming-data-driven TSE method. To the best of the authors’ knowledge, inter-link matching information has not been used in TSE studies.

Regarding the bias in mobile data specific to the TSE problem, to the best of the authors’ knowledge, no studies have investigated this problem.

5. Estimation approaches

TSE approaches, which may be based on the models and data that we discussed in previous sections, are summarized in this section.

5.1. Model-driven approaches

In this article, a model-driven TSE method is defined as a TSE method which is based on physical traffic flow models (c.f., Section 3). It relies on explicit a priori knowledge of traffic dynamics to estimate the traffic state based on partial observation.

The advantages are as follows: first, if the model is representative of the physics of traffic, it should add value to the observation. Therefore, the method may estimate an accurate traffic state with less input data. In addition, it has high explanatory power. This means that even if the estimation is inaccurate, it would be possible to identify the reasons as well as confidence intervals. Moreover, it can be integrated with traffic control operations directly, for example, by using model predictive control. The disadvantage is that, in the case of poor model or poorly calibrated models, such models can lead to poor performance of the TSE. Therefore, model-driven TSE requires careful selection and calibration of models, depending on the application. In some cases, checking the validity of a model or calibrating a model requires a large amount of data.

5.1.1. BVP

As mentioned in Section 3.2.3, numerical schemes have been developed to solve BVPs numerically. In many model-driven TSE studies, the boundary conditions are observed or given. Therefore, solving BVPs can be regarded as TSE where the boundary conditions and model are assumed to be correct (obviously, this is forward simulation rather than estimation; this distinction is not always clear in some of the TSE literatures).

Several studies have proposed TSE simply based on BVPs. Coifman (2002) proposed a calculation method for vehicle trajectories from disaggregated stationary data based on application of the LWR model. Laval et al. (2012) developed a stochastic extension to Newell’s principle to estimate traffic states with their confidence intervals. Kuwahara et al. (2013) used VT to combine stationary and mobile data. Blandin et al. (2013) and Fan et al. (2014) used PTM and ARZ-like models, respectively, to estimate non-stationary traffic.

5.1.2. Data assimilation

Data assimilation (DA) or inverse modeling are two terms sometimes interchangeably used to refer to the problem of estimation (or inference) and model calibration (or system identification) respectively (Chen, 2003; Evensen, 2009). For DA-based TSE, traffic flow models and data are no longer considered as perfect. Instead, the approach estimates “the most probable state” which may not be identical to both model prediction and observation. In other words, it allows observation to correct the model’s prediction. This approach has the benefit of integrating the modeling and measurement error which we discussed in previous sections. Therefore, the DA approach is widely employed by TSE studies. To the authors’ knowledge, use of DA for TSE is firstly proposed by Szeto and Gazis (1972).

The Kalman filter (KF) and its extensions and variations (KFTs: KF-like techniques) are among the most well-known implementation of DA. In general, the KFTs are based on the state-space model:

\[
x_t = f_t(x_{t-1}, \nu_t),
\]

(14)
Figure 12: DA and TSE.

\[ y_t = h_t(x_t, \omega_t), \]  

where \( x_t \) is a state vector, \( f_t \) is a system model, \( \nu_t \) is system noise, \( h_t \) is an observation model, and \( \omega_t \) is observation noise at time \( t \). Eq. (14) is referred to as a system or process equation and represents the dynamics of the system. Eq. (15) is referred to as an observation or measurement equation and represents the observation of the system. The main objective of the KFTs is to obtain the \( x_t \) that maximizes \( p(x_t|y_1, y_2, \ldots, y_t) \). In other words, the KFTs estimates the most probable state variables with respect to an available observation, assumed system model, and noises. For the details on DA and KFTs, see Chen (2003); Evensen (2009); Higuchi (2011); Van Leeuwen et al. (2015) (description on KFTs in this section is mainly based on Higuchi (2011)). Fig. 12 illustrates a framework of DA in TSE.

In the TSE context, the state vector \( x_t \) often corresponds to a discretized traffic flow or cumulative flow that is the subject of the TSE. Then, \( f_t \) generally represents the numerical scheme used for the continuous PDE traffic flow model (Section 3.2.3), to which model noise is added. The model parameters, such as the FD parameters, are either endogenously estimated together with the traffic state or exogenously assumed, as discussed in Section 3.1 (this problem is often referred to as system identification or model calibration (Ljung, 1998)). The former is performed using various techniques, such as defining a state vector that includes model parameters (e.g., Wang and Papageorgiou, 2005; Wang et al., 2009) and applying a dual filter (e.g., van Lint et al., 2008). The latter makes the problem simple; however, it requires an independent (offline) parameter calibration procedure, which may cause the aforementioned problems. The system noise \( \nu_t \) encompasses the modeling error or uncertainty. Some studies have explicitly modeled noise from its source (e.g., Sumalee et al., 2011; Laval et al., 2012; Kurzhanskiy and Varaiya, 2012), whereas many others have assumed that the noise is simple white noise with a given deviation, which means that it may require another calibration. The observation vector \( y_t \) corresponds to traffic data; and the observation model \( h_t \) corresponds to a mapping from the traffic state to observed traffic data. A linear observation model, which means that the state variables are directly observed, is often employed. For the case of stationary data, either aggregated or disaggregated, it is easy to construct an observation model for a conventional traffic flow model (i.e., a Eulerian coordinate system-based model with density or cumulative flow as the state variable) in a straightforward manner. For the case of mobile data, some techniques are required, as discussed in Section 4.2, and can pose some problems when penetration rate or temporal sampling rate is low. Observation noise \( \omega_t \) represents measurement errors, which are usually known from the technical specification of sensors. There are various implementations of the KFTs that assume different conditions to Eqs. (14) and (15). The following explains the most representative implementations together with their applications and features for TSE.

The KF is the most basic KFT as the name suggests. It assumes that both the system and observation models are linear (e.g., \( x_t = F_t x_{t-1} + G_t \nu_t \)). The KT is efficient to compute; however, as standard traffic flow models are...
clearly nonlinear, KT has not been widely used in TSE. Some studies have used the KF to exploit its advantage; for example, Sun et al. (2003) and Thai and Bayen (2015) developed an SMM in which the system model $f_i$ behaves linearly and applied the KF (or mixture KF).

The extended Kalman filter (EKF) can use a nonlinear system model to some extent. This is made possible by linearizing the model around the current state (i.e., $F_t = \partial_x f_i(x_{t-1}, \nu_t)$). Because of this beneficial feature, it has been widely used in TSE studies (e.g., Nanthawichit et al., 2003; Wang and Papageorgiou, 2005; Schreiter et al., 2010). Note that a common mistake perpetrated in the literature (knowingly or not, ignored on purpose or not) is to use EKF technique on models that are not differentiable, such as the Godunov discretization of the (Eulerian) LWR model, including CTM (Blandin et al., 2012a). The linearization can be based on either an analytical or numerical differential. An analytical differential approach is preferable as a numerical differential induces numerical error. However, it is known that some of accurate numerical schemes, such as the Godunov scheme for the Eulerian LWR model again, are not differentiable in general (Blandin et al., 2012a). Yuan et al. (2012) avoided this shortcoming by using the fact that the Godunov scheme can be differentiable in Lagrangian coordinates. A notable example of accurate and differential schemes is the Lax–Friedrichs scheme used by, for example, Wong and Wong (2002) and Göttlich et al. (2013).

The unscented Kalman filter (UKF) overcomes the shortcomings of the EKF. It can use a nonlinear system model well and does not require an analytical differential. Therefore, it has been used early on for some TSE studies, such as those using Godunov schemes (e.g., Mihaylova et al., 2006).

The ensemble Kalman filter (EnKF) overcomes the shortcomings of the EKF by employing Monte Carlo simulation. It can use a nonlinear and/or non-differentiable system models well (Blandin et al., 2012a). Therefore, it has been used often in recent studies (e.g., Work et al., 2008). The problem is that it can be computationally costly because of the Monte Carlo simulation. Meanwhile, another costliest operations in the aforementioned filters, namely, KF, EKF, UKF, and EnKF, is matrix inversion. To account for this problem, several studies have investigated localization methods (van Hinsbergen et al., 2012; Sun and Work, 2014; Yuan et al., 2015), so that the size of the matrix can be reduced.

The particle filter (PF) extensively uses Monte Carlo simulation to represent nonlinear phenomena precisely. It has been used in several studies (e.g., Mihaylova and Boel, 2004). Naturally, it is very computationally costly. In particular, the curse of dimensionality makes it essentially difficult to apply to large-scale estimation. However, at the same time, the PF has a promising feature in terms of computational cost: unlike other filters, it does not require matrix inversion. It means that the PF can be parallelized easily and is therefore scalable (Heyi et al., 2007; Mihaylova et al., 2012). Wright and Horowitz (2016) applied Rao–Blackwellized PF for efficient computation.

Other methods that are not based on the KFTs have been also proposed. The ASF (adaptive smoothing filter) developed by Treiber and Helbing (2002) specifically for TSE has been used to combine multiple sensing data. It relies on the constant wave speed (i.e., information propagation speed) in each of the free-flowing and congested regimes; therefore, it can be considered to be based on the LWR model with triangular FD. Treiber and Helbing (2002) used ASF to estimate the traffic state using stationary data. van Lint and Hoogendoorn (2010) and Treiber et al. (2011) developed extensions of ASF to incorporate mobile data as well.

Some studies have reduced a TSE problem to an optimization problem, leveraging mathematical properties of the problem formulation. Convex optimization problems representing TSE have been formulated using theories in HJ PDEs (e.g., Claudel and Bayen, 2008, 2011; Han et al., 2012). Deng et al. (2013) and Lei and Zhou (2014) formulated optimization problems representing offline17 TSE. Sun et al. (2015) formulated a least square method that jointly estimates the traffic state and FD parameters.

There are other methods which might not fall in aforementioned conventional class of estimation. Abouaissa et al. (2008) proposed their original method to estimate model parameters jointly. Herrera and Bayen (2010) used the Newtonian relaxation method, which is a heuristic DA technique, to incorporate erroneous mapping from probe vehicle data by changing the model. Kurzhanskiy and Varaiya (2012) developed a TSE method with detailed consideration of uncertainty in models and observation. Canaud et al. (2013) used probability hypothesis density filtering (PHDF) to capture an uncertainty properly like PF, while maintaining a low computational cost. Piccoli et al. (2013, 2015) have developed mapping methods based on features of PTMs to derive traffic state and other variables

---

17Offline means “after the data”, when all data has been collected, as opposed to online (i.e., using streaming data). This is sometimes also referred to as smoothing in the machine learning community.
from probe vehicle data. Vivas et al. (2015) proposed a distributed consensus-based method to enable large-scale estimation. By nature of the model equations, spectral methods are less common in the field, but have been used; for example, Tchrakian and Zhuk (2015) and Tchrakian et al. (2015) used minimax estimation to implement Galerkin methods as numerical schemes instead of conventional FDMs in order to define flexible resolution. Nguyen et al. (2016) used the adjoint method to estimate model parameters jointly. Zheng and Su (2016) applied compressed sensing (CS) and Markov random field to represent noises more appropriately.

5.2. Data-driven approaches

In this article, a data-driven TSE method is defined as a TSE method that directly relies on historical-data rather than explicit traffic flow models. It mainly uses statistical and ML approaches, and infers real-time traffic states based on dependence found in historical-data. Compared with the model-driven approach, the data-driven approach does not require explicit theoretical assumptions, such as an underlying PDE with FD and a well posed BVP with proper discretization scheme. Additionally, the data-driven approach often rely on relatively simple models; therefore, their computation cost for estimation can be remarkably low. However, a dependency on historical-data means that the methods can fail if irregular events or a long-term trend occurred. The computation cost for training and learning can be significantly high. Moreover, the method can be considered as a “black box,” which means that it is difficult to obtain deductive insights.

Imputation methods have developed to complement missing data caused by malfunctions of detectors and communication. Statistical approaches were often used in early studies. These approaches relied on data from neighboring detectors and time periods. Smith et al. (2003) compares several heuristic methods based on historical-data with a statistical method using a data augmentation technique. A linear regression model using data from spatial neighbor detectors (Chen et al., 2003), and autoregressive integrated moving average (ARIMA) using time series dataset (Zhong et al., 2004) are developed. To consider greater complexity of traffic data, several studies have developed methodologies that depend more substantially on historical-data, such as ML and Bayesian statistics. Ni and Leonard II (2005) proposed time series based model incorporating to Bayesian network (BN) to improve bias and robustness of estimation. Kernel regression (KR) (Yin et al., 2012), fuzzy c-means (FCM) (Tang et al., 2015), k-nearest neighbors (kNN) (Tak et al., 2016), and probabilistic principal component analysis (PPCA) (Li et al., 2013) have been used to incorporate more spatial-temporal information. Tan et al. (2014) developed a method based on robust principal component analysis (RPCA) to consider capacity of traffic flow as well as the temporal correlation. Tensor-based methods using Tucker decomposition (TD) (Tan et al., 2013) have also been used to consider spatial-temporal information and other information, such as week, day, and hour. Xu et al. (2015) adopted CS to estimate traffic state in large-scale networks. Ran et al. (2016) proposed tensor completion algorithm (HalRTC) to improve imputing performance. Duan et al. (2016) applied deep learning (DL) (c.f., Polson and Sokolov, 2017). Bayesian state-space modeling has also been applied to estimate the hidden parameters of the traffic state. Polson and Sokolov (2016) used the Bayesian particle filter (BPF) to estimate traffic state corresponding to free-flowing, breakdown, and recovery regimes.

Several studies have attempted to estimate flow using probe vehicle data. These studies have explored statistical relations between flow and other variables. Essentially, the parameters of TSE models (i.e., concepts similar to an FD) are estimated from historical stationary data, and then streaming mobile data are used for TSE. For example, speed (Anuar et al., 2015; Neumann et al., 2013), variance of speed (Blandin et al., 2012b; Bulteau et al., 2013), and other variables in xFCD (Wilby et al., 2014) are used for such estimation.

It is known that ML can be good at predicting nonlinear phenomena often found in the transportation field (Karlaftis and Vlahogianni, 2011). However, as shown above, relatively few studies have used ML for TSE. This might be because the physics of link traffic flow has been well understood in previous studies as shown in Section 3.2, and the available models’ performance is relatively high for estimation and subsequent control compared with data-driven methods.

Meanwhile, several ML-based approaches can consider some indirectly observed variables of state, such as demand and route choice, which are difficult to be modeled explicitly as mentioned in Section 3.3. For example, this advantage is exploited by short-term future prediction of traffic state (van Lint and van Hinsbergen, 2012; Vlahogianni et al., 2014; Ermagun and Levinson, 2016). Moreover, for an arterial road (which is a large-scale network connected with unobserved roads and difficult to described by a physical model), such a hybrid of a traffic flow model and ML
works well for travel time prediction (Herring et al., 2010; Hofleitner et al., 2012a,b). Therefore, such use of ML might be beneficial for particular highway TSE.

5.3. Streaming-data-driven approaches

In this article, a streaming-data-driven TSE method is defined as a TSE method that rely on streaming data and “weak” assumptions, such as random sampling condition and CL. It does not rely on “strong” assumptions characterized by empirical relation, such as PDE models and FDs (as in model-driven approaches), nor dependency in historical-data (as in data-driven approaches). This means that the method requires less a priori knowledge and no historical-data; thus, it is robust against uncertain phenomena and unpredictable incidents. At an age of near ubiquitous sensor (e.g., cell phone) penetration, and with the massive emergence of connected vehicles, this approach might become prevalent in the near future. The limitation of streaming-data-driven approach is that they may require massive streaming data in order to perform accurate estimation. In addition, the approach itself does not have any future prediction capability.

Wardrop and Charlesworth (1954) proposed the moving observer method, which is TSE using vehicles that counts nearby vehicles. Florin and Olariu (2017) proposed a modern variant of the moving observer method. By assuming a random sampling condition only, Seo et al. (2015a) proposed TSE methods using vehicles that measure spacing (i.e., xFCD).

The CL is a principle in traffic flow theory and considered as physically reasonable without empirical justification (Papageorgiou, 1998). Vehicle trajectories (i.e., disaggregated probe vehicle data) are useful information to be incorporated with the CL, because the number of vehicles between two vehicles will be conserved because of the CL if there is no lane-changing. Although this is not exactly the case in practice due to lane-changing, it can be a reasonable assumption if a sufficiently wide area is considered where effects of lane-changing can be ignored; in this sense, it can be considered as a “weak” assumption. Several studies have proposed TSE methods based on this idea. Coifman (2003) used disaggregated data from TTI and detectors to estimate density. Its notable feature is that it considers and estimates lane-changing flow as well. Astarita et al. (2006) and Qiu et al. (2010) used aggregated and disaggregated, respectively, probe vehicle data and a limited number of detectors (e.g., at the entrance and exit). Bekiaris-Liberis et al. (2016) combined this approach with the KF, whose system model represents the CL. Seo and Kusakabe (2015) combined disaggregated spacing probe vehicle data with the CL.

6. Summary of the survey

In this article, we have provided a comprehensive survey of existing methods for highway TSE problem. Existing TSE methods were characterized according to three elements which are fundamental to TSE: estimation approach, traffic flow model, and input data. Then their basic features, advantages, and disadvantages were discussed. The results would be useful for assessing the applicability of a TSE method and determining future research directions.

6.1. Current state

Tables 1–4 show comprehensive lists of existing TSE methods with their main characteristics, namely estimation approach, traffic flow model, and input data. For the notation, see Appendix A.

From these tables, progress and the current state of TSE can be surveyed at a glance. In summary:

1. The model-driven approach has been extensively studied. It is predominated both in the traffic engineering and in the control/optimization literature. This might be because of decades of extensive work of modelers.

2. For the model-driven approach, the LWR model is widely used. This might be because of its simplicity and compactness. Use of Godunov scheme for the LWR model is also popular.

3. For the model-driven approach, the KFTs are widely used probably because it leads itself well to discretization of PDE models. Recent studies have begun to incorporate several new frameworks, such as convex optimization.

4. The use of mobile data has received a considerably amount of attention recently, since the era of so-called Internet of Things, the emergence of the mobile internet, and crowd sourcing traffic information data. However, most of the mobile data-based TSE methods today still requires stationary data.
5. The use of disaggregated data has received some attention recently. This might because of theoretical advances in traffic flow models, combined with the aforementioned technology progress.

6. Data-driven and streaming-data-driven approaches appear to have received more attention recently, probably because of high data availability in these days.

6.2. Future research directions

Based on the survey results, considerable future research directions can be suggested.

The use of disaggregated data has begun to receive significant attention. Simultaneously, data availability continues to increase in terms of both quantity (e.g., penetration rate of probe vehicles) and quality (e.g., xFCD and TTI), making TSE a beneficiary of this trend. However, this may result in some theoretical challenges for model-driven TSE because of discrepancies between the single-lane single-class traffic flow models, which have been widely used in current studies, and actual disaggregated data, which is collected from a multi-lane multi-class environment (c.f., Sections 3.2.2 and 4.2). To account for this problem, the development and use of Lagrangian multi-lane multi-class models (e.g., Yuan, 2013; Costeseque and Duret, 2016), which are challenging tasks, for TSE could be considered. The use of data-driven or streaming-data-driven approaches (which recent studies have begun to consider) would be also valuable, as they do not require explicit traffic flow models. Regarding application of TSE, such high-resolution traffic states will be valuable for using automated vehicle technologies, such as lane-changing and merging assistance, which require lane-scale detailed traffic state information (Diakaki et al., 2015).

The bias in mobile data is an important problem (c.f., Sections 3.2.2 and 4.2). To solve this problem, Lagrangian multi-lane multi-class models would be useful because they can track specific vehicles by considering their characteristics.

Traffic demand, both at boundaries and junctions (c.f., Section 3.3), is often neglected by TSE studies (e.g., simply assumed to be observed or inferred from historical-data). This is partially accounted for by existing studies, such as the estimation of boundary flow (e.g., Wang et al., 2006; Zhang and Mao, 2015); however, demand responsive to traffic has not been well investigated in TSE. This problem would be significant for TSE in large-scale networks. One of the reasons for a lack of such studies might be the modeling difficulty caused by the uncertainty of human behavior. To account for this problem, the incorporation of ML could be considered because it works well for the case of arterial roads (Herring et al., 2010; Hofleitner et al., 2012a,b). The use of inter-link TTI data, which has not been used by current TSE (c.f., Section 4.2), would be valuable data for this problem, since the data represent route choice behavior of large number of road users. Additionally, to enable such large-scale estimation, it is necessary to extensively develop and implement computationally efficient techniques, such as numerical schemes (e.g., Muñoz et al., 2006), localization (e.g., van Hinsbergen et al., 2012), and parallelization (e.g., Hegyi et al., 2007).
<table>
<thead>
<tr>
<th>Study</th>
<th>Approach</th>
<th>Model Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Num. scheme</td>
<td>external FD?</td>
</tr>
<tr>
<td>Note</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pourmoallem et al. (1997)</td>
<td>NKF</td>
<td>PW</td>
</tr>
<tr>
<td>Cofman (2002)</td>
<td>BVP</td>
<td>LWR-(N)-like</td>
</tr>
<tr>
<td>Treiber and Helbing (2002)</td>
<td>ASF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Muñoz et al. (2003)</td>
<td>SMM</td>
<td>LWR</td>
</tr>
<tr>
<td>Nanthawichit et al. (2003)</td>
<td>EKF</td>
<td>PW</td>
</tr>
<tr>
<td>Sun et al. (2003)</td>
<td>MKF</td>
<td>LWR</td>
</tr>
<tr>
<td>Mihaylova and Boel (2004)</td>
<td>PF</td>
<td>higher-order Godunov</td>
</tr>
<tr>
<td>Wang and Papageorgiou (2005)</td>
<td>EKF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Cheng et al. (2006)</td>
<td>PF</td>
<td>higher-order Godunov-like</td>
</tr>
<tr>
<td>Hegyi et al. (2006)</td>
<td>EKF/UKF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Mihaylova et al. (2006)</td>
<td>UKF</td>
<td>higher-order Godunov</td>
</tr>
<tr>
<td>Wang et al. (2006)</td>
<td>EKF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Zuurbier et al. (2006)</td>
<td>EKF</td>
<td>LWR</td>
</tr>
<tr>
<td>Hegyi et al. (2007)</td>
<td>PF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Mihaylova et al. (2007)</td>
<td>PF/UKF</td>
<td>higher-order Godunov</td>
</tr>
<tr>
<td>Sau et al. (2007)</td>
<td>PF</td>
<td>LWR</td>
</tr>
<tr>
<td>Tampère and Immers (2007)</td>
<td>EKF</td>
<td>LWR</td>
</tr>
<tr>
<td>Abouaissa et al. (2008)</td>
<td>other</td>
<td>LWR (-N)</td>
</tr>
<tr>
<td>Claudel and Bayen (2008)</td>
<td>EKF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>van Lint et al. (2008)</td>
<td>EKF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Ngoduy (2008)</td>
<td>UKF/UPF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Work et al. (2008)</td>
<td>EnKF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Kesting and Treiber (2009)</td>
<td>other</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Wang et al. (2009)</td>
<td>EKF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Herrera and Bayen (2010)</td>
<td>nudging</td>
<td>LWR</td>
</tr>
<tr>
<td>Schreiter et al. (2010)</td>
<td>EKF</td>
<td>LWR</td>
</tr>
<tr>
<td>van Lint and Hoogendoorn (2010)</td>
<td>ASF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Work et al. (2010)</td>
<td>EnKF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Chen et al. (2011)</td>
<td>PF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Claudel and Bayen (2011)</td>
<td>other</td>
<td>LWR (-N)</td>
</tr>
<tr>
<td>Moraescu and Canudas de Wit (2011)</td>
<td>SMM</td>
<td>LWR</td>
</tr>
<tr>
<td>Ngoduy (2011)</td>
<td>PF-like</td>
<td>higher-order</td>
</tr>
<tr>
<td>Sumalee et al. (2011)</td>
<td>SMM</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Treiber et al. (2011)</td>
<td>ASF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Canepa and Claudel (2012)</td>
<td>other</td>
<td>LWR (-N)</td>
</tr>
<tr>
<td>Canudas de Wit et al. (2012)</td>
<td>SMM</td>
<td>LWR</td>
</tr>
<tr>
<td>Han et al. (2012)</td>
<td>other</td>
<td>LWR (-N)</td>
</tr>
<tr>
<td>Kurzhanskiy and Varaiya (2012)</td>
<td>other</td>
<td>LWR</td>
</tr>
<tr>
<td>Laval et al. (2012)</td>
<td>BVP</td>
<td>LWR-(N)-like</td>
</tr>
<tr>
<td>van Hinsbergen et al. (2012)</td>
<td>EKF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Mihaylova et al. (2012)</td>
<td>PF-like</td>
<td>PW-like</td>
</tr>
<tr>
<td>Yuan et al. (2012)</td>
<td>EKF</td>
<td>LWR (-N)</td>
</tr>
<tr>
<td>Study</td>
<td>Approach</td>
<td>Model Data</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>Blandin et al. (2013)</td>
<td>BVP</td>
<td>PTM</td>
</tr>
<tr>
<td>Canaud et al. (2013)</td>
<td>PHDF</td>
<td>LWR</td>
</tr>
<tr>
<td>Deng et al. (2013)</td>
<td>other</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Kuwahara et al. (2013)</td>
<td>BVP</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Pascale et al. (2013)</td>
<td>PF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Piccoli et al. (2013)</td>
<td>mapping</td>
<td>PTM</td>
</tr>
<tr>
<td>Wilkie et al. (2013)</td>
<td>EnKF</td>
<td>ARZ</td>
</tr>
<tr>
<td>Yuan (2013)</td>
<td>EKF</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Claudel et al. (2014)</td>
<td>other</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Fan et al. (2014)</td>
<td>BVP</td>
<td>ARZ-like</td>
</tr>
<tr>
<td>Lei and Zhou (2014)</td>
<td>other</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Li et al. (2014)</td>
<td>other</td>
<td>multiple</td>
</tr>
<tr>
<td>Morbidi et al. (2014)</td>
<td>SMM</td>
<td>LWR</td>
</tr>
<tr>
<td>Polson and Sokolov (2015)</td>
<td>PF</td>
<td>LWR</td>
</tr>
<tr>
<td>Sun and Work (2014)</td>
<td>KF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Wang and Work (2014)</td>
<td>EnKF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Yuan et al. (2014)</td>
<td>EKF</td>
<td>LWR-N</td>
</tr>
<tr>
<td>Gundiegard et al. (2015)</td>
<td>EnKF</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Kuwahara (2015)</td>
<td>BVP</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Lovisari et al. (2015)</td>
<td>other</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Majid and Abouaissa (2015)</td>
<td>SMM</td>
<td>LWR</td>
</tr>
<tr>
<td>Piccoli et al. (2015)</td>
<td>mapping</td>
<td>PTM</td>
</tr>
<tr>
<td>Seo et al. (2015b)</td>
<td>EnKF</td>
<td>LWR</td>
</tr>
<tr>
<td>Sun et al. (2015)</td>
<td>other</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>Tchrakian and Zhuk (2015)</td>
<td>Minimax</td>
<td>LWR</td>
</tr>
<tr>
<td>Tchrakian et al. (2015)</td>
<td>Minimax</td>
<td>LWR</td>
</tr>
<tr>
<td>Thai and Bayen (2015)</td>
<td>KF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Timotheou et al. (2015a)</td>
<td>KF</td>
<td>LWR</td>
</tr>
<tr>
<td>Vivas et al. (2015)</td>
<td>other</td>
<td>LWR</td>
</tr>
<tr>
<td>Zhang and Mao (2015)</td>
<td>KF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Canepa and Claudel (2016)</td>
<td>other</td>
<td>LWR-N HL</td>
</tr>
<tr>
<td>Duret et al. (2016)</td>
<td>other</td>
<td>LWR-N VT</td>
</tr>
<tr>
<td>He et al. (2016)</td>
<td>EKF</td>
<td>PW-like</td>
</tr>
<tr>
<td>Lovisari et al. (2016)</td>
<td>other</td>
<td>LWR-like</td>
</tr>
<tr>
<td>Nguyen et al. (2016)</td>
<td>adjoint</td>
<td>LWR</td>
</tr>
<tr>
<td>Seo et al. (2016b)</td>
<td>Minimax</td>
<td>LWR</td>
</tr>
<tr>
<td>Wang et al. (2016a)</td>
<td>PF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Wang et al. (2016b)</td>
<td>PF-like</td>
<td>LWR</td>
</tr>
<tr>
<td>Wright and Horowitz (2016)</td>
<td>PF-like</td>
<td>higher-order Godunov-like yes S no pseudo-second-order model</td>
</tr>
<tr>
<td>Zheng and Su (2016)</td>
<td>other</td>
<td>LWR</td>
</tr>
<tr>
<td>Xia et al. (2017)</td>
<td>EnKF/PF</td>
<td>LWR-like</td>
</tr>
</tbody>
</table>
### Table 3: Summary of data-driven TSE methods.

<table>
<thead>
<tr>
<th>Study</th>
<th>Approach</th>
<th>Model</th>
<th>Data</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Name</td>
<td>Num. scheme</td>
<td>external FD?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al. (2003)</td>
<td>regression</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Smith et al. (2003)</td>
<td>regression</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Zhong et al. (2004)</td>
<td>ARIMA</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Ni and Leonard II (2005)</td>
<td>regression</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Blandin et al. (2012b)</td>
<td>regression</td>
<td>FD-like</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Yin et al. (2012)</td>
<td>Kernel regression</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Bulteau et al. (2013)</td>
<td>regression</td>
<td>FD-like</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Li et al. (2013)</td>
<td>PPCA</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Neumann et al. (2013)</td>
<td>BN</td>
<td>FD</td>
<td>-</td>
<td>yes</td>
</tr>
<tr>
<td>Ran et al. (2016)</td>
<td>HaLRTC</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Tan et al. (2013)</td>
<td>TD</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Tan et al. (2014)</td>
<td>RPCA</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Wilby et al. (2014)</td>
<td>regression</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Tang et al. (2015)</td>
<td>FCM</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Xu et al. (2015)</td>
<td>CS</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Anuar et al. (2015)</td>
<td>regression</td>
<td>FD</td>
<td>-</td>
<td>yes</td>
</tr>
<tr>
<td>Asif et al. (2016)</td>
<td>multiple</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Duan et al. (2016)</td>
<td>DL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Ran et al. (2016)</td>
<td>TD</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Tak et al. (2016)</td>
<td>kNN</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Polson and Sokolov (2016)</td>
<td>BPF</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
</tbody>
</table>

### Table 4: Summary of streaming-data-driven TSE methods.

<table>
<thead>
<tr>
<th>Study</th>
<th>Approach</th>
<th>Model</th>
<th>Data</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Name</td>
<td>Num. scheme</td>
<td>external FD?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wardrop and Charlesworth (1954)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Cofman (2003)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Astarita et al. (2006)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Qu et al. (2010)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Seo et al. (2015a)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Seo and Kusakabe (2015)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Bekiaris-Liberis et al. (2016)</td>
<td>CL</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Florin and Olariu (2017)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>no</td>
</tr>
</tbody>
</table>

Note: FD = feature detection, S = seasonal, M = moving, xFCD = exact feature detection, CL = clustering, TD = tracking dictionary.
Acknowledgements

The part of this research is financially supported by the Japan Society for the Promotion of Science (KAKENHI Grant-in-Aid for JSPS Fellows 14J10218, Young Scientists (B) 16K18164, and Scientific Research (S) 26220906). The first author would like to thank Dr. Kentaro Wada (The University of Tokyo), Dr. Wataru Nakanishi (Tokyo Institute of Technology), Dr. Masami Yanagihara (Tokyo Metropolitan University), and Mr. Kouki Satsukawa (The University of Tokyo) for insightful discussion on traffic flow models.

Appendix A. Notation for Tables 1–4

In the “Approach” column, NKF denotes the neural Kalman filter, MKF denotes the mixture Kalman filter, and UPF denotes the unscented particle filter. In the “Model/Name” column, LWR-N indicates that a method uses the LWR model with explicit consideration of vehicle trajectories using, for example, HJ PDE or Lagrangian coordinates. The “Model/external FD?” column indicates whether a method requires external calibration of FD parameters. The “Data/Type” column indicates data used by a method: S denotes stationary data, M denotes mobile data, and (S) indicates that a method requires minimum stationary data (e.g., boundary data) although it mainly uses mobile data. The “Data/Disaggr.” column indicates whether a method exploits disaggregated data.

References


