Title
Simultaneous estimation of states and parameters in Newell's simplified kinematic wave model with Eulerian and Lagrangian traffic data

Authors
Sun, Zhe
Jin, Wen-Long
Ritchie, Stephen G

Publication Date
2017-10-01

DOI
10.1016/j.trb.2017.06.012

Peer reviewed

Transportation Research Part B 104 (2017) 106–122

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Simultaneous estimation of states and parameters in Newell's simplified kinematic wave model with Eulerian and Lagrangian traffic data

Zhe Sun[a], Wen-Long Jin[b,*], Stephen G. Ritchie[a]

[a] Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Irvine, CA, 92697, USA
[b] Department of Civil and Environmental Engineering, California Institute for Telecommunications and Information Technology,Institute of Transportation Studies Irvine, CA, 92697-3600, USA

ARTICLE INFO

Article history:
Received 28 August 2015
Revised 19 June 2017
Accepted 21 June 2017
Available online 4 July 2017

Keywords:
Traffic state estimation
Newell's simplified kinematic wave model
Eulerian and Lagrangian traffic data
Optimization problem
NGSIM dataset

ABSTRACT

The traffic state estimation process estimates various traffic states from available data in a road network and provides valuable information for travelers and decision makers to improve both travel experience and system performance. In many existing methods, model parameters and initial states have to be given in order to estimate traffic states, which limits the accuracy of the results as well as their transferability to different locations and times. In this paper, we propose a new framework to simultaneously estimate model parameters and traffic states for a congested road segment based on Newell's simplified kinematic wave model (Newell, 1993). Given both Eulerian traffic count data and Lagrangian vehicle reidentification data, we formulate a single optimization problem in terms of the initial number of vehicles and model parameters. Then we decouple the optimization problem such that the initial number of vehicles can be analytically solved with a closed-form formula, and the model parameters, including the jam density and the shock wave speed in congested traffic, can be computed with the Gauss-Newton method. Based on Newell's model, we can calculate individual vehicles' trajectories as well as the average densities, speeds, and flow-rates inside the road segment. We also theoretically show that the optimization problem can have multiple solutions under absolutely steady traffic conditions. We apply the proposed method to the NGSIM datasets, verifying the validity of the method and showing that this method yields better results in the estimation of average densities than existing methods.

1. Introduction

Traffic information is essential for travelers and decision makers to improve both travel experience and system performance. The traffic estimation process estimates various traffic states, e.g. density, flow rate, from available data in a road network. Ideally, an estimation method should provide a complete picture of the traffic states based on limited available data (Wang and Papageorgiou, 2005).

* Corresponding author.
E-mail address: wjin@uci.edu (W.-L. Jin).

http://dx.doi.org/10.1016/j.trb.2017.06.012
0191-2615/

Two types of traffic data can be collected through various sensors. The first one is the Eulerian data collected at fixed locations. The inductive loop detector system is the most commonly used sensing system to collect Eulerian data. These detectors collect traffic counts and occupancy, and aggregate them at certain sampling intervals (usually 30 s) for each lane. This type of data has been used for many traffic applications and studies, including traffic management and control and traffic state estimation. The second type is the Lagrangian data collected for individual vehicles. This type of data can be generated by a vehicle reidentification system, which matches vehicles passing different locations (Sun et al., 1999). A vehicle reidentification system can be implemented based upon different sensing technologies, such as video cameras, AVI (automatic vehicle identification) tags, and loop detectors (Jeng, 2007). This type of data has been used for travel time estimation (Coifman and Cassidy, 2002), performance evaluation (Jeng, 2007; Oh et al., 2005), and O/D trip estimation (Oh et al., 2002).

Traditionally, traffic states are estimated with Eulerian traffic data. Coifman (2002) proposed a method to reconstruct vehicle trajectories from speed measurements of dual loop detectors. This method exploits the propagation of characteristic waves based on the Lighthill–Whitham–Richard (LWR) model (Lighthill and Whitham, 1955; Richards, 1956). Sun et al. (2003) developed a mixture Kalman filter to estimate traffic states based on the Switching-Mode Model (SMM) and tested it with Eulerian traffic counts and occupancy data through the PeMS database. Wang and Papageorgiou (2005) described traffic dynamics with a second-order continuum model, and estimated traffic densities and speeds with Kalman filters from Eulerian flow and occupancy measurements of loop detectors.

Recently, there has been increased interest in assimilating both Eulerian and Lagrangian data for traffic state estimation. Nanthawichit et al. (2003) adopted Payne's traffic flow model (Payne, 1971) for traffic state estimation, and a Kalman filtering estimation framework was implemented to combine data from both loop detectors and probe vehicles. Herrera and Bayen (2008) developed a similar Kalman filtering framework to estimate traffic states from both mobile sensor and loop detector data, but with the SMM (Muñoz et al., 2003). In addition, the study also introduced the nudging method (Newtonian relaxation) for data fusion. The two methods were evaluated with real-world traffic data, and it was found that Kalman filtering slightly outperformed the nudging method at the cost of being more complicated to tune and implement. The study by Claudel and Bayen (2010a, 2010b) reformulated the traffic state estimation problem using the Hamilton–Jacobi formula of the LWR model. This approach guarantees an exact solution with piece-wise linear initial and boundary data. In the study by Tossavainen and Work (2013), the Markov chain Monte Carlo (MCMC) method is used to estimate parameters in the fundamental diagram from GPS data. Using synthetic data, the authors verified the accuracy of parameter estimation and examined the absolute error of the velocity field obtained using the parameters. The study by Deng et al. (2013) proposed a stochastic three-detector (STD) method to incorporate heterogeneous data sources and formulated the estimation problem as an optimization problem. In a recent study by Bekiaris-Liberis et al. (2016), a "linear parameter-varying system", which is closely related to SMM, is proposed for traffic state estimation with both Eulerian and Lagrangian data in the Kalman filter framework. Nantes et al. (2016) proposed the incremental Extended Kalman filter for traffic estimation on arterial corridors using data from multiple sensors. Most existing studies support that probe sensor data, complementary to fixed location sensor data, improve data availability and enhance estimation accuracy. It is worth noting that Kalman filter (and its extensions) is a widely used technique for data assimilation in most of the traffic estimation literatures. It provides a natural and principal way of combining difference data sources with the error models, within a solid decision framework. However, misleading results can be generated if the prior distribution and error models are misspecified. Also, the computational cost is usually quite high, especially for traffic networks with large number of parameters.

Almost all traffic flow models used by existing methods involve a number of model parameters, especially those in the fundamental diagram (Cassidy, 1998), including the free-flow speed, jam density, and shock wave speed. In addition, traffic states at a time also depend on the initial conditions. Despite the substantial progress made in traffic estimation, existing estimation methods are limited since either the model parameters or the initial states have to be predetermined. In most existing studies,[1] the fundamental diagram is assumed to be given, or calibrated before estimating the initial and other traffic states. In other studies the initial states are to assumed be empty (Deng et al., 2013) or known (Sun et al., 2003) in advance. Such assumptions can limit the accuracy of the results as well as their transferability to different locations and times.

In this paper, we attempt to fill the gap by proposing a new framework to simultaneously estimate model parameters and traffic states for a homogeneous road segment with no internal ramps based on Newell's simplified kinematic wave model (Newell, 1993). Note that, by simultaneous estimation, we mean that model parameters in the fundamental diagram, the initial states, and later traffic states can all be estimated within the same framework, but calculations can still be sequentially ordered. This is substantially different from existing methods where model parameters and initial states have to be observed or estimated separately with other frameworks. The simultaneous estimation of both parameters and states is achieved by formulating a single optimization problem in terms of the initial number of vehicles and model parameters. Here we assume that both Eulerian traffic count data and a portion of Lagrangian vehicle trajectory data are available through loop detectors and vehicle reidentification systems or GPS devices, respectively. We then decouple the optimization problem such that the initial number of vehicles can be calculated with a closed-form formula, and the model parameters, including the jam density and shock wave speed in congested traffic, can be computed with the Gauss-Newton method. Further, based on

[1] There are exceptions. e.g. Canepa and Claudel (2012).

Table 1
Notations.

l	Length of a road segment
Δt	Time step size
$F(t)$	Cumulative traffic counts at the upstream boundary from 0 to t
$G(t)$	Cumulative traffic counts at the downstream boundary from 0 to t
$f(t)$	Upstream flow-rate at t
$g(t)$	Downstream flow-rate at t
I	Number of reidentified vehicles
r_i	Entry time of vehicle i
s_i	Exit time of vehicle i
n_0	The initial number of vehicles within the segment
V	Free-flow speed
$-W$	Shock wave speed in congested traffic
K	Jam density
$n(t, x)$	The cumulative count at location x from time 0 to t
$X_i(t)$	Location of vehicle i at time t
$k(t; x_1, x_2)$	Average density at time t within subsegment x_1 to x_2 ($0 < x_1 < x_2 < l$)

Newell's model, we can calculate other traffic state variables, including individual vehicles' trajectories as well as the average densities, speeds, and flow-rates inside the road segment.

The proposed framework has some similarities with that in Deng et al. (2013): both are based on Newell's simplified kinematic wave model and use the same types of data, including Eulerian traffic counts and Lagrangian vehicle trajectories. However, the two methods are fundamentally different, since Deng et al. (2013) assumes predetermined model parameters and initial states, and the resulting optimization problems are substantially different. Furthermore, with the simultaneous framework in this study, we show that the new method yields better results in the estimation of average densities. This is an essential benefit of the simultaneous framework. Table 1 lists the notation.

The paper is organized as follows. In Section 2, we review Newell's simplified kinematic wave model (Newell, 1993) and the STD method (Deng et al., 2013). In Section 3, we present the new estimation framework based on Newell's model for a congested road segment and formulate an optimization problem in the initial traffic states and model parameters. In Section 4, we present the solution method and discuss some properties of the optimization problem. In Section 5, we use the NGSIM data (USDOT, 2008) to test the validity of our method with different correct matching rates and compare our method with the STD method. We conclude the whole study in Section 6.

2. Literature review

In this section we first review Newell's simplified kinematic wave model, pointing out its relation with the LWR model, and then summarize the STD traffic estimation method developed by Deng et al. (2013). These reviews are essential for our study, as our estimation method is also based on Newell's simplified kinematic wave model, and we will carefully compare our new framework with the STD method both conceptually and empirically.

2.1. Newell's simplified kinematic wave model

The LWR model (Lighthill and Whitham, 1955; Richards, 1956) describes the spatial-temporal evolution of flow-rate, $q(t, x)$, and density, $k(t, x)$, at time t and location x by the following scalar conservation law,

$$\frac{\partial k}{\partial t} + \frac{\partial Q(k)}{\partial x} = 0, \tag{1}$$

where $q = Q(k)$ is the fundamental diagram (Greenshields, 1935). In the LWR model, the initiation, propagation, and dissipation of traffic queues are associated with such kinematic waves as shock and rarefaction waves.

In Newell (1993) a simplified kinematic wave model was introduced with a new state variable, the cumulative flow, $n(t, x)$, which is the number of vehicles passing t and x after a reference vehicle. Since the flow-rate is the time-derivative of the cumulative flow; i.e.,

$$q(t, x) = \frac{\partial n(t, x)}{\partial t}, \tag{2}$$

and the density is the negative space-derivative; i.e.,

$$k(t, x) = -\frac{\partial n(t, x)}{\partial x}, \tag{3}$$

the simplified kinematic wave model is equivalent to the Hamilton–Jacobi formulation of the LWR model:

$$\frac{\partial n(t, x)}{\partial t} = Q\left(-\frac{\partial n(t, x)}{\partial x}\right), \tag{4}$$

with a triangular fundamental diagram

$$Q(k) = \min\{Vk, W(K-k)\}, \tag{5}$$

where V is the free-flow speed, $-W$ the shock wave speed in congested traffic, and K the jam density. Daganzo (2005) employed variational theory to solve (4). In Evans (1998), the Hopf-Lax formula and other methods were introduced to solve general Hamilton–Jacobi equations.

In this study, we are interested in applying Newell's simplified kinematic wave model, (4), to estimate traffic states on a homogeneous road segment of length l with no internal ramps. Here we denote $x \in [0, l]$, where $x = 0$ and $x = l$ are respectively the upstream and downstream boundaries of the road segment. We denote the cumulative traffic counts from 0 to t at the upstream and downstream boundaries by $F(t)$ and $G(t)$, respectively. If we consider the vehicle passing the downstream boundary at $t = 0$ as the reference vehicle, then

$$G(t) = n(t, l), \tag{6}$$

$$F(t) + n_0 = n(t, 0), \tag{7}$$

where $F(0) = G(0) = 0$, and n_0 is the initial number of vehicles on the road segment.

If initially there exist no transonic rarefaction wave on the road segment; i.e., if the initial traffic condition is one of the three types: (i) the whole road is uncongested, (ii) the whole road is congested, or (iii) an upstream part of the road is uncongested and the downstream part is congested, it was shown in Jin (2015) that a simple closed-form formula can be derived for Newell's simplified kinematic wave model from the Hopf-Lax formula:

$$n(t,x) = \min\left\{F\left(t - \frac{x}{V}\right) + n_0, G\left(t - \frac{l-x}{W}\right) + K \cdot (l-x)\right\}. \tag{8}$$

This equation indicates that the cumulative flow on a homogeneous road is determined by the wave emitted from either upstream (first term on the right-hand side) or the downstream (second term on the right-hand side) boundaries based on a minimization principle (Newell, 1993).

From (8), we can see that, given cumulative traffic counts, $F(t)$ and $G(t)$, the initial number of vehicles on the road segment, n_0, and the model parameters, V, W, and K, one can compute the cumulative flow, $n(t, x)$, and therefore estimate other traffic states at any time and location inside the road segment. Among these variables, traffic counts $F(t)$ and $G(t)$ can be obtained in real time through loop detectors installed at both boundaries of the road segment, but the initial number of vehicles and model parameters have to be estimated.

2.2. The STD traffic estimation method

In Deng et al. (2013), the stochastic version of (8) was referred to as the stochastic three-detector model (STD), based on which a traffic estimation method was developed. Thus it was referred to the STD traffic estimation method. The method makes the following assumptions:

- The initial number of vehicles, n_0, is given.
- The model parameters, V and W, are given.
- The jam density, K, is time-dependent and estimated together with later traffic states.

The traffic states to be estimated are the cumulative traffic counts at the upstream and downstream boundaries.

In this method, Eulerian traffic count data, $F(j\Delta t)$ and $G(j\Delta t)$ for $j = 0, \cdots, J$, are available through loop detectors at the two ends of a road segment; and Lagrangian vehicle data, the entry and exit times of vehicles, are available through vehicle reidentification technologies. Then a multiple-objective optimization problem is formulated to minimize the absolute difference between estimated and observed cumulative traffic counts as well as the absolute difference in the cumulative numbers of vehicles with respect to a reidentified vehicle at the upstream and downstream boundaries. In addition, three types of constraints are introduced: (i) (8) is used as constraints on the upstream and downstream cumulative traffic counts; (ii) the estimated cumulative traffic counts have to be non-negative and non-decreasing in time; and (iii) the estimated upstream and downstream flow-rates cannot be greater than the capacity, $C = \frac{VW}{V+W}K$.

With NGSIM data for vehicle trajectories on a segment of I-80 in California (USDOT, 2008), it was shown that the method is valid and can be used to estimate cumulative traffic counts and the average densities on the segment. Therefore, by explicitly considering the stochasticity of different data sources, the STD method is a powerful traffic estimation framework and can be potentially applied to solve many traffic problems. However, the method assumes that the initial number of vehicles and model parameters, V and W, have to be estimated separately with other methods, in advance.

3. A new framework for simultaneous estimation of states and parameters

In this study, we present a new traffic estimation method based on Newell's simplified kinematic wave model (8). As in the STD method, Eulerian traffic count data, $F(t)$ and $G(t)$, are available through loop detectors at the two boundaries

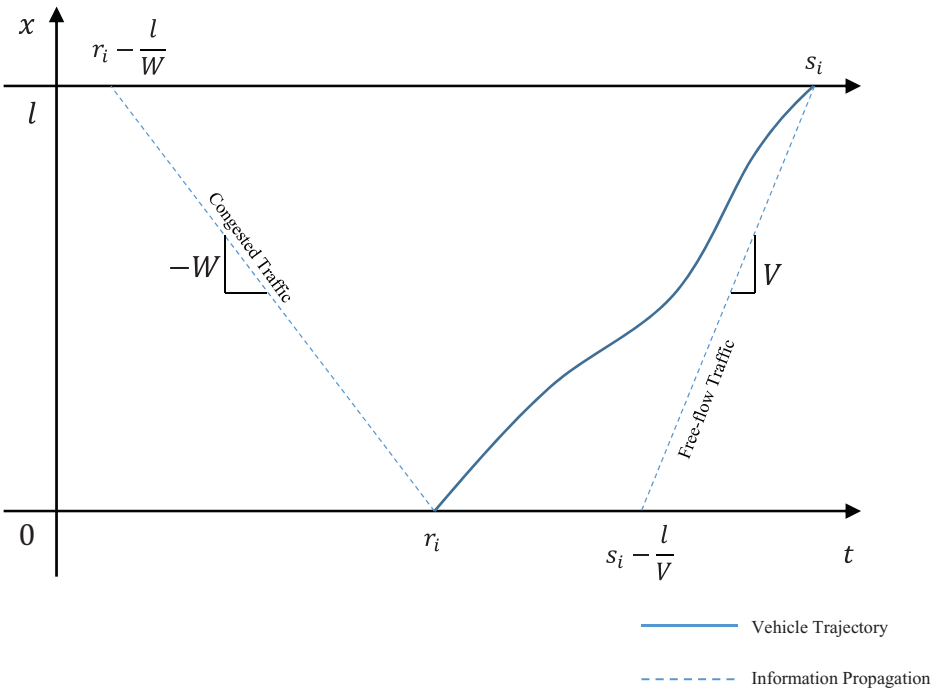


Fig. 1. Illustration of a vehicle trajectory and kinematic waves on a road segment.

of a road segment; in addition, Lagrangian vehicle data, the entry and exit times of vehicles, are available through vehicle reidentification technologies. We assume that I vehicles are reidentified. If vehicle i ($i = 1, \cdots, I$) is reidentified at the downstream boundary, we denote its entry and exit times in a road segment by r_i and s_i, respectively. If we denote $X_i(t)$ as the location of vehicle i at t, then $X_i(r_i) = 0$ and $X_i(s_i) = l$. Here we assume only a portion of vehicles are reidentified. In Fig. 1, we illustrate the trajectory of vehicle i (the solid curve), whose entry and exit times are known, and the kinematic waves in uncongested and congested traffic (the dashed lines).

Different from the STD method, the new method can estimate the initial number of vehicles, model parameters, and traffic states by solving one optimization problem. In this sense, this is a simultaneous estimation framework.

3.1. An optimization problem in initial states and model parameters

Assume that vehicles follow the First-In-First-Out (FIFO) principle on the whole road segment, then the contour lines of $n(t, x)$ are vehicle trajectories. In particular, the cumulative numbers of vehicles should be equal when a reidentified vehicle passes the upstream and downstream boundaries; i.e.,

$$F(r_i) + n_0 = G(s_i). \tag{9}$$

From (7) and (8), we have

$$F(r_i) + n_0 = n(r_i, 0) = \min\left\{F(r_i) + n_0, G\left(r_i - \frac{l}{W}\right) + Kl\right\},$$

$$G(s_i) = \min\left\{F\left(s_i - \frac{l}{V}\right) + n_0, G(s_i)\right\}.$$

Combining the above equations with (9) we further have

$$G(s_i) = \min\left\{F(r_i) + n_0, G\left(r_i - \frac{l}{W}\right) + Kl\right\}, \tag{10}$$

$$F(r_i) + n_0 = \min\left\{F\left(s_i - \frac{l}{V}\right) + n_0, G(s_i)\right\}. \tag{11}$$

In particular, when traffic is congested, traffic information propagates upstream with a speed of $-W$, and we have

$$G(s_i) = G\left(r_i - \frac{l}{W}\right) + Kl; \tag{12}$$

when traffic is uncongested, traffic information propagates downstream with a speed of V, and we have

$$F(r_i) = F\left(s_i - \frac{l}{V}\right). \tag{13}$$

In this study, we focus on congested traffic conditions, as the uncongested traffic condition is less of a concern in practice. Therefore, we will use ((12)) for further discussion, but the results for (13) will be similar.

Note that the equalities in (9) and ((12)) can be altered by the following four types of errors:

1. Errors caused by the measurement of cumulative traffic counts. Cumulative traffic counts are usually collected with loop detectors and can be corrupted by noise and errors, due to a variety of problems including pavement/saw-cut failures, intermittent communications, double counting of lane-changing vehicles, and so on (Coifman, 2006; Lee and Coifman, 2011). According to PeMS, which is a widely used data source for the freeway sensor system in California, only 67% of the detectors were working properly in May, 2014. Some districts (e.g. district 7 in Los Angeles County) have even lower proportions of working detectors.
2. Errors caused by the vehicle reidentification technology. The entry and exit times of reidentified vehicles can be corrupted by time shifts and vehicle mismatching in the vehicle reidentification technology. The time shift is mainly caused by the spatial and temporal inaccuracy of the underlying detection technology, including the Bluetooth technology or the GPS technology (Wang et al., 2011). Mismatch errors can be caused by a variety of algorithmic and technology limitations. Jeng (2007) defined the correct matching rate (CMR) as the ratio between total number of correct matched vehicles and total number of vehicles, and obtained CMR of 69.76% to 73.59% in off-peak conditions and 50.68% to 54.20% in peak conditions. The same CMR measure is also used in this study.
3. Errors caused by the violation of the FIFO principle. The FIFO principle is a fundamental assumption to derive both (9) and ((12)). In reality, however, the principle is only approximately followed on a multi-lane road, as travel speeds of different types of vehicles on different lanes are usually different in the same longitudinal location at the same time. In the research by Jin and Li (2007), it was confirmed that the FIFO principle is generally violated, and the level of FIFO violation was empirically studied with an NGSIM database.
4. Errors associated with Newell's simplified kinematic wave theory. Newell's simplified kinematic wave theory, (8), is valid only under a number of assumptions: (i) there exists an equilibrium relation between flow-rate and density; (ii) the fundamental diagram is triangular; and (iii) initially there exists no transonic rarefaction wave. In reality, however, (i) the existence of an equilibrium flow-density relation is only approximately true, as the observed data are be quite scattered (Hall et al., 1986); (ii) the shape of the fundamental diagram is also debatable, due to impacts of lane changes, multi-class vehicles, and so on (Jin, 2013); and (iii) incidents could create some transonic rarefaction waves on a road.

To carefully study the properties of the aforementioned errors can be fundamentally important, but is beyond the scope of this study. Here we introduce error terms into both (9) and ((12)), such that

$$n_0 = G(s_i) - F(r_i) + \epsilon_i, \tag{14}$$

$$G(s_i) = G\left(r_i - \frac{l}{W}\right) + Kl + \xi_i, \tag{15}$$

In particular, ϵ_i is related to the first three types of errors, and ξ_i all four types of errors. We propose to estimate n_0, W, and K by minimizing the sum of squared errors in (14) and (15):

$$\min_{\hat{n}_0,\hat{W},\hat{K}} \sum_{i=1}^{I} \epsilon_i^2 + \xi_i^2 = \sum_{i=1}^{I} [G(s_i) - F(r_i) - \hat{n}_0]^2 + \left[G(s_i) - G\left(r_i - \frac{l}{\hat{W}}\right) - \hat{K}l\right]^2. \tag{16}$$

where $\hat{n}_0$ is the estimated number of vehicles initially on the road segment, and $\hat{W}$ and $\hat{K}$ are respectively estimated shock wave speed and jam density.

We do not specify any statistical distribution for the error terms ϵ_i and ξ_i. Thus, this can be considered a least squares estimation method (Charnes et al., 1976). The least square estimation framework should be considered only as a method of "curve fitting" instead of statistical inference since we have yet to gather enough evidence to justify the distribution of the error terms. The least square estimates are mathematically equivalent to the maximum likelihood estimation only when the error terms are uncorrelated and normally distributed (Charnes et al., 1976). However, it still provides reasonable approximates in engineering practice.

3.2. Estimation of traffic states

Once the initial number of vehicles and the three model parameters are estimated by solving the optimization problem, (16), we can estimate all traffic states for $x \in (0, l)$ and $t > 0$:

The cumulative flows can be calculated by using Newell's simplified kinematic wave model, (8), as follows:

$$\hat{n}(t,x) = \min\left\{F\left(t-\frac{x}{\hat{V}}\right)+\hat{n}_0, G\left(t-\frac{l-x}{\hat{W}}\right)+\hat{K}\cdot(l-x)\right\}, \tag{17}$$

where $\hat{V}$ is the estimated free-flow speed which can be estimated in uncongested traffic by solving a similar optimization problem as (16).

From the estimated cumulative flows, we can then estimate the trajectories of individual vehicles by assuming that vehicles follow the FIFO principle. For vehicle j, its location at t is denoted by $X_j(t)$, which satisfies the following equation:

$$\hat{n}(t, X_j(t)) = j. \tag{18}$$

That is, the trajectory of an individual vehicle is a contour line in the cumulative flow.

From the estimated cumulative flows, we can also estimate the average density inside a subsegment from x_1 to x_2 at time t, denoted by $\hat{k}(t; x_1, x_2)$:

$$\hat{k}(t;x_1,x_2) = \frac{1}{x_2-x_1}\left[G\left(t-\frac{l-x_1}{\hat{W}}\right)-G\left(t-\frac{l-x_2}{\hat{W}}\right)+\hat{K}(x_2-x_1)\right]. \tag{19}$$

In particular, for the whole road segment from 0 to l, the average density can be estimated by

$$\hat{k}(t;0,l) = \frac{1}{l}[F(t)-G(t)+\hat{n}_0]. \tag{20}$$

Note that the estimated average density for the whole road segment only involves the estimated initial number of vehicles, and does not directly depend on $\hat{K}$ or $\hat{W}$.

We can also estimate the flow-rate at any location $x \in [0, l]$ and time t. Since we just consider congested traffic, the flow-rate at location x is determined by the cumulative flow at the downstream:

$$\hat{q}(t,x) = \frac{n(t+\Delta t,x)-n(t)}{\Delta t} = \frac{G(t+\Delta t-\frac{l-x}{\hat{W}})-G(t-\frac{l-x}{\hat{W}})}{\Delta t}, \tag{21}$$

where Δt is the time-step size. In particular, the upstream flow-rate, $f(t) = q(t, 0)$, can be estimated as

$$\hat{f}(t) = \frac{G(t+\Delta t-\frac{l}{\hat{W}})-G(t-\frac{l}{\hat{W}})}{\Delta t}. \tag{22}$$

Alternatively, $f(t)$ can be calculated from observed cumulative traffic counts at the upstream boundary

$$f(t) = \frac{F(t+\Delta t)-F(t)}{\Delta t}. \tag{23}$$

By comparing $\hat{f}(t)$ and $f(t)$ we can measure the accuracy of the new estimation method.

In addition to the aforementioned variables, we can also estimate travel speeds, travel times, and other traffic information from the estimated cumulative flows, $\hat{n}(t, x)$.

4. Solution of the optimization problem

From the preceding section we can see that the traffic estimation problem, (16), is solved. In this section, we present a solution method and discuss some properties for it.

4.1. A decoupling method

We first observe that the optimization problem, (16), can be decoupled as follows.

Theorem 1. *The optimization problem, (16), can be decoupled into the following two optimization problems:*

$$\min_{\hat{n}_0}\sum_{i=1}^{I}[G(s_i)-F(r_i)-\hat{n}_0]^2, \tag{24}$$

$$\min_{\hat{W},\hat{K}}\sum_{i=1}^{I}[G(s_i)-G(r_i-\frac{l}{\hat{W}})-\hat{K}l]^2. \tag{25}$$

Proof. This is straightforward, since, in the objective function of (16), $\hat{n}_0$ appears only in the first term, and $\hat{W}$ and $\hat{K}$ appear only in the second term. Thus the objective function of (16) is minimized if and only if both terms are minimized. □

Therefore we can solve (24) and (25) separately. In particular, since the objective function in (24) is quadratic in $\hat{n}_0$, it has the following closed-form analytical solution.

Corollary 1. *The initial number of vehicles on the road segment can be estimated by*

$$\hat{n}_0 = \frac{\sum_{i=1}^{I}[G(s_i) - F(r_i)]}{I}. \tag{26}$$

That is, the initial number of vehicles equals the mean value of the differences in the cumulative counts at the entry and exit times of reidentified vehicles.

Proof. Taking the derivative of the objective function in (24) with respect to $\hat{n}_0$ and setting it to zero, we can easily obtain (26). Furthermore we can verify that the objective function attains its minimum, since the second-order derivative is positive at the optimal point. □

Although the objective function in (25) is quadratic in $\hat{K}$, its relation with $\hat{W}$ is not in a simple functional form. Thus (25) cannot be analytically solved. Since the objective function in (25) is quadratic in $\hat{K}$, with known $\hat{W}$ the optimal value of $\hat{K}$ can be analytically solved as

$$\hat{K} = \sum_{i=1}^{I} \frac{G(s_i) - G(r_i - \frac{l}{\hat{W}})}{Il}. \tag{27}$$

In addition, if the flow-rate is relatively constant over time; i.e., when traffic is in a steady state (Cassidy, 1998) with an average flow-rate of $\bar{q}$, we have

$$G(s_i) - G\left(r_i - \frac{l}{\hat{W}}\right) \approx \left(s_i - r_i + \frac{l}{\hat{W}}\right)\bar{q}. \tag{28}$$

In this case, the objective function in (25) is also convex in $\hat{W}$. This property justifies the usage of a gradient-based Gauss–Newton method (Milliken, 1990).

We denote $\boldsymbol{\theta}$ as the column vector of $(\hat{W}, \hat{K})^T$. Starting from an initial guess $\boldsymbol{\theta}^{(0)} = (\hat{W}^{(0)}, \hat{K}^{(0)})^T$, the Gauss-Newton method updates the results by iterating

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - [\boldsymbol{J}(\boldsymbol{\theta}^{(j)})^T \boldsymbol{J}(\boldsymbol{\theta}^{(j)})]^{-1} \boldsymbol{J}(\boldsymbol{\theta}^{(j)})^T \boldsymbol{\xi}(\boldsymbol{\theta}^{(j)}),$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Jacobian matrix of $\boldsymbol{\xi}(\boldsymbol{\theta}) = (\xi_1(\boldsymbol{\theta}), \xi_2(\boldsymbol{\theta}), \ldots, \xi_I(\boldsymbol{\theta}))^T$ and is defined as

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \xi_1(\boldsymbol{\theta})}{\partial \hat{W}} & \frac{\partial \xi_1(\boldsymbol{\theta})}{\partial \hat{K}} \\ \vdots & \vdots \\ \frac{\partial \xi_I(\boldsymbol{\theta})}{\partial \hat{W}} & \frac{\partial \xi_I(\boldsymbol{\theta})}{\partial \hat{K}} \end{bmatrix}. \tag{29}$$

The essential idea is to approximate the Hessian matrix of the objective function, $\boldsymbol{\xi}^T(\boldsymbol{\theta}^{(j)})\boldsymbol{\xi}(\boldsymbol{\theta}^{(j)})$, with its first order approximation $\boldsymbol{J}(\boldsymbol{\theta}^{(j)})^T\boldsymbol{J}(\boldsymbol{\theta}^{(j)})$ at time-step j. The iteration stops when the solution converges. We consider that the algorithm has converged when the second norm of the error $||\boldsymbol{\theta}^{(j+1)} - \boldsymbol{\theta}^{(j)}||$ is smaller than a predefined error bound (we used 10^{-4} as it is a typical setting for engineering practice).

In this problem, the elements in $\boldsymbol{J}(\boldsymbol{\theta})$ are $(i = 1, \cdots, I)$

$$\frac{\partial \xi_i(\boldsymbol{\theta})}{\partial \hat{W}} = -\frac{l}{\hat{W}^2}\frac{dG(t)}{dt}\Big|_{t=r_i-\frac{l}{\hat{W}}} = -\frac{l}{\hat{W}^2} g\left(r_i - \frac{l}{\hat{W}}\right), \tag{30}$$

$$\frac{\partial \xi_i(\boldsymbol{\theta})}{\partial \hat{K}} = -l. \tag{31}$$

The downstream flow-rate in (30), $g(r_i - \frac{l}{\hat{W}})$, is approximated by the forward finite difference,

$$g\left(r_i - \frac{l}{\hat{W}}\right) \approx \frac{G(r_i - \frac{l}{\hat{W}} + \Delta t) - G(r_i - \frac{l}{\hat{W}})}{\Delta t}, \tag{32}$$

where Δt is the time-step size. For simplicity, we fix the time-step size to be 30 s.

4.2. Potential issue with near-steady traffic

When traffic is absolutely steady such that the trajectories of vehicles are parallel lines, we denote the density, speed, and flow-rate by k, v, and q, respectively. Here $q = kv$. Assuming that there are no measurement errors in the cumulative traffic counts, then

$$G(s_i) - G\left(r_i - \frac{l}{\hat{W}}\right) = \left(s_i - r_i + \frac{l}{\hat{W}}\right)q,$$

Further, assuming that there are no errors caused by reidentification technologies, the travel time for reidentified vehicle i is constant

$$s_i - r_i = \frac{l}{v}.$$

Thus the optimization problem, (25), can be simplified as

$$\min_{\hat{W},\hat{K}} l^2 \cdot \sum_{i=1}^{I} \left(k + \frac{q}{\hat{W}} - \hat{K}\right)^2,$$

which is analytically solved by

$$\hat{K} = k + \frac{q}{\hat{W}}. \tag{33}$$

In this case, $\hat{W}$ and $\hat{K}$ follow a hyperbolic relationship, and cannot be determined uniquely. That is, if vehicles have parallel trajectories in steady traffic, there are multiple (infinite) solutions for $\hat{W}$ and $\hat{K}$.

In reality, it is impossible for traffic to be absolutely steady. However, it is possible that the traffic is near-steady. In this case, the objective function in (25) becomes relatively flat and its value is almost constant for a large range of values in $\hat{W}$ and $\hat{K}$. Thus, the parameters are hardly identifiable. This insight was also used in calibrating car-following models (Yang et al., 2011). Therefore, such limitations are actually caused by a lack of variation in the observed data, not the estimation framework itself.

5. Applications to NGSIM data

In this section, the performance of the proposed method is evaluated in terms of its ability to estimate the initial traffic state (n_0), the average density for the whole road segment, and the upstream flow-rate. The data are chosen from the *Next Generation Simulation* (NGSIM) database (USDOT, 2008) for the US 101 freeway in Los Angeles, CA, from 7:50 a.m. to 8:35 a.m. on June 15, 2005. The study site is a segment between the Ventura Blvd and the Cahuenga Blvd off-ramps on the southbound US 101 freeway as shown in Fig. 2. The segment is about 0.13 miles long with five lanes and one auxiliary lane.

The 45 min period is split into three 15 min intervals. In this study, we only consider the through traffic which passed both upstream and downstream of the road segment. There are 1894 vehicle trajectories in the first time period, 1842 in the second time period, and 1698 in the third period. We deploy virtual detectors at both ends of the road segment and generated volume counts and vehicle entry/exit times based on vehicles' longitudinal coordinates ("LocalY" field in the NGSIM trajectory dataset). While there has been some discussion about the NGSIM data accuracy in recent years (Choi, 2014; Punzo et al., 2011; Thiemann et al., 2008), most of the discovered issues are related to speed and acceleration. The longitudinal coordinates are unbiased and it is acceptable to use it as the actual vehicle path (Punzo et al., 2011).

On average, it took around 25 s for vehicles to go through this segment under congested conditions. Only the left five lanes of data were used since the auxiliary lane had much lower flow-rates and was significantly under-utilized, which violates the congested traffic state assumption. We use a 3 min warm-up period such that all vehicle trajectories are captured in the spatial-temporal domain (see Fig. 2).

To study the impact of the sample size of vehicle reidentification, we used three different CMR (100%, 20%, 5%) for each time interval; i.e. we assume respectively 100%, 20%, and 5% of all vehicle are correctly reidentified at both upstream and downstream boundaries and their entry/exit times are used in the estimation. To consider the errors in the observed cumulative flow, an error model similar to the one by Laval et al. (2012) is used. Assume that, at location x and time t, the correct counts $n_1(t, x)$, the double counts $n_2(t, x)$, and the missed counts $n_3(t, x)$ are generated from a multinomial distribution with total trials $n_1(t,x) + n_2(t,x) + n_3(t,x)$, and the corresponding probabilities $p_1 = 1 - p_2 - p_3, p_2, p_3$, respectively, with $p2$, $p3 \in [0, 0.5]$. The observed cumulative count at location x is then $n_1(t,x) + 2n_2(t,x)$. In case 4 and 5, the cumulative count errors at both upstream and downstream locations are assumed to be unbiased, i.e. $p_2 = p_3$, with two different double/miss counting rate (2% and 5%). In case 6, the upstream sensor is assumed to have systematic error so it double counts 5% of the vehicles, i.e., $p_2 = 0.05, p_3 = 0$. The downstream detector has 5% unbiased random error. In case 7, the upstream detector is assumed to miscount vehicles with 5% probability, i.e., $p_2 = 0, p_3 = 0.05$ while the downstream detector has 5% unbiased random errors. In summary, we tested this method using seven different scenarios as shown in Table 2.

5.1. Data preparation

The sampling frequency of NGSIM is 10 Hz. For vehicle i, its location at time step j is $X_i(j\Delta t)$, where $\Delta t = 0.1s$. We prepared two datasets for the estimation method, namely, the vehicle reidentification measurements and the flow measurements.

The vehicle reidentification data contain the entry/exit times of individual vehicles at the study site. Let x_0 and x_l denote the entry and exit point of the segment respectively. We used a linear interpolation function to find the entry time s_i, formally,

$$s_i = j\Delta t + \frac{x_0 - X_i(j\Delta t)}{X_i(j\Delta t + \Delta t) - X_i(j\Delta t)}\Delta t,$$

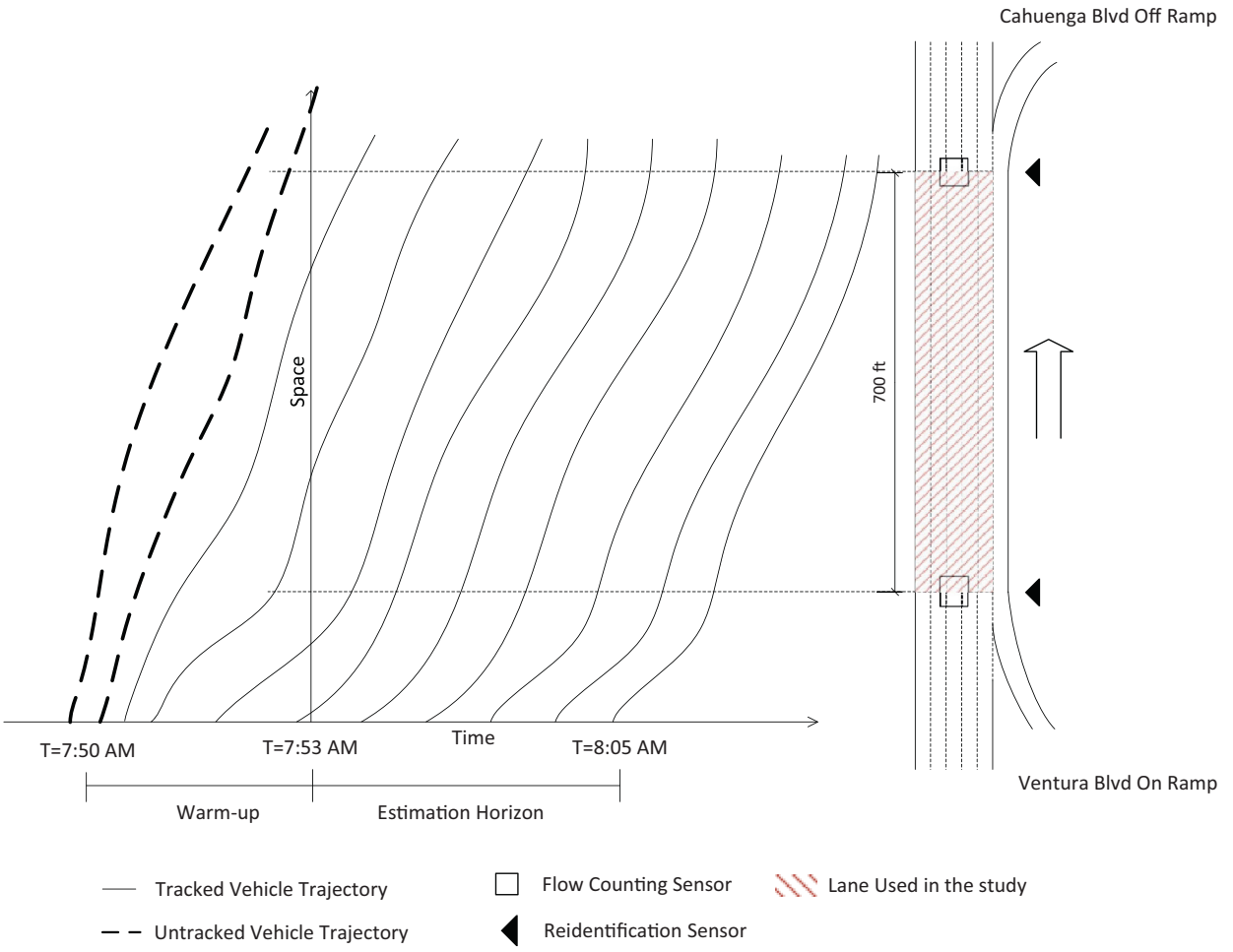


Fig. 2. Illustration of the study site: Using the interval 7:50 a.m. – 8:05 a.m. as an example, the first 3 min are used as the warm-up period. After 7:53 a.m., all the vehicle trajectories are used in estimation.

Table 2
Test cases.

Case	CMR	Upstream sensor error		Downstream sensor error	
		Double-count	Miscount	Double-count	Miscount
1	100%	0%	0%	0%	0%
2	20%	0%	0%	0%	0%
3	5%	0%	0%	0%	0%
4	100%	2%	2%	2%	2%
5	100%	5%	5%	5%	5%
6	100%	5%	0%	5%	5%
7	100%	0%	5%	5%	5%

where j satisfies $X_i(j\Delta t) \le x_0 \le X_i(j\Delta t + \Delta t)$. Similarly, the exit time is

$$r_i = j'\Delta t + \frac{x_l - X_i(j'\Delta t)}{X_i(j'\Delta t + \Delta t) - X_i(j'\Delta t)}\Delta t,$$

where j' is chosen such that $X_i(j'\Delta t) \le x_l \le X_i(j'\Delta t + \Delta t)$.

We can interpolate the cumulative flow based on the entry and exit times, s_i and r_i, for vehicle i. The cumulative flow functions $F(t)$ and $G(t)$ are practically step functions. However, it is preferable to approximate them with continuous functions which are easier to evaluate. In particular, we can find the instantaneous upstream flow-rate at time t by evaluating $f(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t}$ at almost all points, and the same for downstream flow-rate $g(t)$. Here, we use the approximation method in Daganzo (1997). The method approximates the step function with a piecewise linear curve passing through

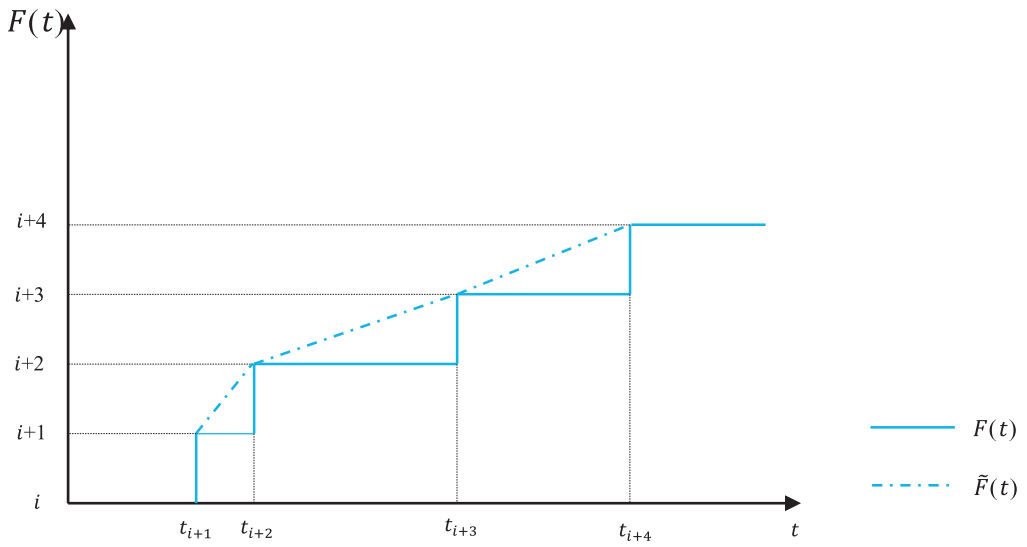


Fig. 3. Piecewise linear interpolation of cumulative flows.

Table 3
Parameter estimation.

7:50 a.m. – 8:05 a.m.			
Case No.	$\hat{n}$ (True: 39)	$\hat{W}$(mph)	$\hat{K}$(vpm)
1	38.31(0.00)	20.00(0.00)	156.51(0.00)
2	38.25(0.49)	23.19(3.29)	145.99(9.08)
3	38.32(1.02)	23.39(5.11)	146.78(12.98)
4	38.57(4.50)	21.94(2.22)	149.27(7.65)
5	39.07(13.85)	22.28(2.45)	147.50(7.84)
6	16.50(11.12)	24.35(1.98)	144.63(7.47)
7	60.38(14.04)	21.44(2.27)	150.55(6.92)
8:05 a.m. – 8:20 a.m.			
Case No.	$\hat{n}$ (True: 39)	$\hat{W}$(mph)	$\hat{K}$(vpm)
1	38.92(0.00)	25.67(0.00)	141.46(0.00)
2	39.05(0.53)	21.71(2.01)	153.52(6.41)
3	38.98(1.18)	22.46(3.05)	151.60(9.21)
4	39.26 (8.44)	24.52 (2.08)	142.68 (5.39)
5	39.71 (15.20)	24.06 (2.06)	143.65 (5.56)
6	17.13 (13.91)	24.78(2.39)	144.72 (4.87)
7	54.99 (15.48)	22.80(2.53)	142.43 (5.23)
8:20 a.m. – 8:35 a.m.			
Case No.	$\hat{n}$ (True: 51)	$\hat{W}$(mph)	$\hat{K}$(vpm)
1	51.32(0.00)	21.15(0.00)	157.98(0.00)
2	51.36(0.62)	20.98(1.14)	158.80(3.80)
3	51.32(1.34)	21.45(2.15)	157.52(6.97)
4	50.30(8.52)	21.20(0.64)	156.20(2.21)
5	50.87(12.44)	20.99(0.65)	157.12(2.72)
6	22.83(11.96)	22.94(0.47)	159.58(2.75)
7	69.74(12.59)	20.64(0.84)	158.87(2.53)

the crests as illustrated in Fig. 3. In the plot, the cumulative flow function at the upstream, $F(t)$, is approximated by $\tilde{F}(t)$. The resulting function is differentiable almost everywhere except at the transition points.

5.2. Estimation of initial states and model parameters

We first estimate the initial number of vehicles on the road segment, $\hat{n}_0$, and the model parameters, $\hat{W}$ and $\hat{K}$, with methods presented in Section 4.1. In the Gauss–Newton method, the initial guess of $\hat{W}$ is 20 miles per hour and that of $\hat{K}$ is 200 vehicles per mile per lane.

The results in Table 3 provides the mean values of $\hat{n}_0$, $\hat{W}$ and $\hat{K}$ followed by their standard deviation for 100 runs each case in Table 2. The true values of n_0 are shown in the parentheses in the header line. First, the mean values of the estimates are insensitive to CMRs and measurement errors. Second, as expected, the variance of the estimation increases

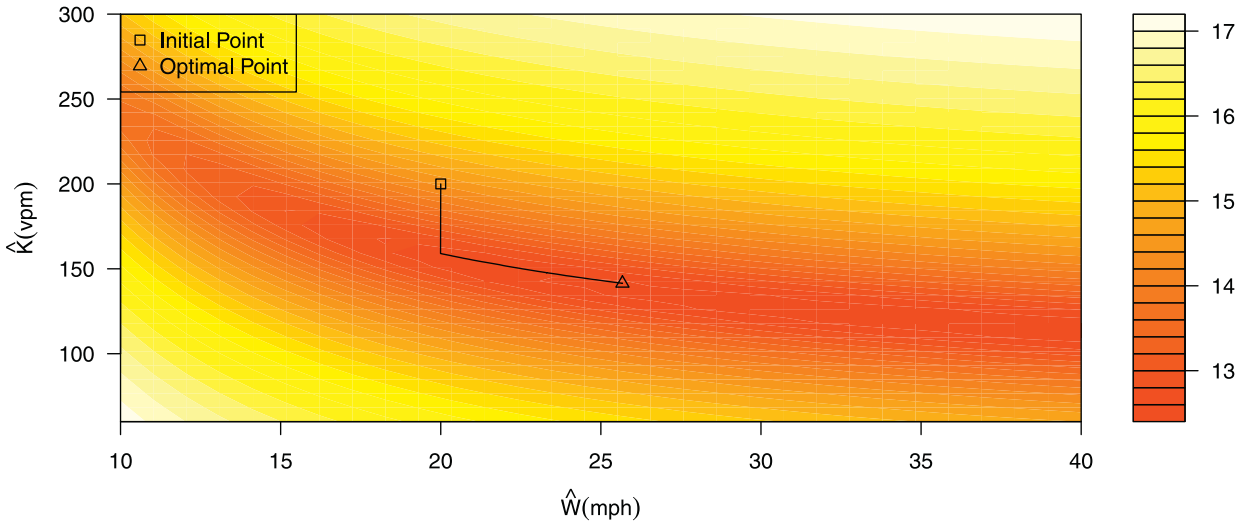


Fig. 4. The objective function in (25).

as the CMR decreases and measurement error increases. This trend indicates that the estimation precision is lower with fewer reidentification observations and higher errors. Compared with the true values, the estimation of the initial states is very accurate when there is no measurement error in the cumulative flows (first three cases). The measurement error in the cumulative flow seems to have a huge impact (cases four to seven) on initiate state estimation. In particular, when the measurement errors are unbiased (case four and five), the estimation mean of n_0 is still close the ground truth while the standard deviation increases with the error rate. The estimated mean of n_0 is much lower than the ground truth when the upstream sensor over-counts vehicles (case six) and is higher when the upstream sensor miscounts vehicles (case seven). The method yields incorrect estimation of the initial states as flow conservation assumption is violated when the flow sensor suffers from systematic errors. Moreover, as the n_0 is estimated based on the difference of cumulative flows, the error increases infinitely with time. Thus biased errors on cumulative flow counts could be fatal for initial state estimation over long period of time. On the other hand, the impact of such measurement errors is very limited on the mean of traffic parameters estimations (jam density and shock wave speed), although the standard deviation of the estimates still suffer from random errors.

The estimated shock wave speed is around 22 mph for all time periods, which is higher than that calibrated from loop detector data in PeMS (around 15 mph). At the same time, the estimated jam density is lower than that from PeMS (around 150 vpm vs 200 vpm). This discrepancy may be due to the fact that the parameters estimated here are for a road segment, while the PeMS results are mostly based on point measurements.

We also notice that the variances of the estimated parameters are smaller for the later time periods, indicating higher precisions. A possible reason is: the estimation method is derived under the assumption of congested traffic states, so the modeling error should be relatively low for more congested traffic data. Since the traffic condition in this dataset is getting increasingly congested, the estimation of the later time periods are more accurate.

In Fig. 4, we show the contour plot of the objective function in (25) of case 1 in the second time period. Starting from the initial value $(\hat{W}^{(0)}, \hat{K}^{(0)}) = (20, 200)$, the solution is found to be (25.67, 141.46) after 20 iterations with the Gauss–Newton method. The contour plot of the objective function is constructed based on 900 sample points in the parameter space (30 points on each axis). From the figure we can see that the contour lines are roughly hyperbolic as in (33). This suggests that traffic is close to steady, but still features enough oscillations so that the optimal point is identifiable.

5.3. Validation with respect to the average density and upstream flow-rate

To examine the performance of the estimation method and evaluate the reliability of the estimated values, the average density within the whole segment and the estimated upstream flow-rate are compared with their true values for all seven cases in Table 2 during each time period. The average density within the segment is estimated as in (20). The estimation equation for upstream flow-rate during time interval Δt is shown in (22). To ensure a sufficient number of counts during the aggregation interval, Δt is set to be 30 s.

We use the mean absolute percentage error (MAPE) as the measure of accuracy. The MAPE for the average density, k, is calculated as:

$$MAPE_k = \frac{\sum_{j=1}^{J} |k(j\Delta t) - \hat{k}(j\Delta t)|}{J\bar{k}}, \tag{34}$$

Table 4
MAPE of estimated average density and upstream flow-rate.

Traffic states	Case	MAPE (%): Mean (Std dev)		
		7:50 a.m.–8:05 a.m.	8:05 a.m.–8:20 a.m.	8:20 a.m.–8:35 a.m.
Average density	1	1.47 (0.00)	0.15 (0.00)	0.57 (0.00)
	2	1.43 (0.86)	0.81 (0.57)	0.95 (0.64)
	3	1.90 (1.48)	1.72 (1.23)	1.80 (1.51)
	4	14.19(12.95)	15.41 (15.64)	7.87 (6.56)
	5	16.90(14.08)	18.1(13.55)	9.82 (7.41)
	6	37.20(12.47)	30.39(11.48)	28.32(6.95)
	7	47.15(12.84)	45.82(12.81)	49.48(6.82)
Upstream flow-rate	1	13.99 (0.00)	10.87 (0.00)	14.50 (0.00)
	2	14.54 (0.49)	11.76 (0.96)	14.52 (0.14)
	3	14.39 (0.79)	11.84 (1.22)	14.47 (0.31)
	4	13.67(0.85)	10.92 (1.42)	12.84 (0.85)
	5	14.07(1.02)	10.97 (1.55)	13.41 (0.88)
	6	13.09(0.97)	9.80 (1.27)	13.03(0.71)
	7	12.56(1.06)	10.45 (1.38)	13.17(0.91)

where J is the number of time-steps, and $\bar{k} = \sum_{j=1}^{J} k(j\Delta t)/J$. The MAPE for the upstream flow-rate, f, is calculated as:

$$MAPE_f = \frac{\sum_{t=1}^{J} |f(j\Delta t) - \hat{f}(j\Delta t)|}{J\bar{f}},$$

where $\bar{f} = \sum_{j=1}^{J} f(j\Delta t)/J$.

In Fig. 5 we demonstrate the estimated and ground truth upstream flow-rates of seven cases in Table 2 for the three time intervals compared with the true values. From the plots we can see that the proposed estimation method is able to capture the general trends in the upstream flow-rates. Therefore, the estimated model parameters are reliable. Note that, for the case with CMR lower than 100%, the line is just the result of one run. To better study the estimation accuracy, we conducted repeated sampling for all cases and the results are shown in Table 4.

Table 4 summarizes the MAPEs of estimated average density and upstream flow-rate over three time periods under seven difference cases. The means followed by the standard deviations of the MAPEs are calculated based on the results of 100 runs each. When there are no measurement errors (the first three cases), the mean MAPE of the estimated average density is much smaller than that of the upstream flow-rate. The mean MAPE of the estimated average density is much larger when there are measurement errors (case four to seven). In particular, when the upstream flow suffers from biased error (case six and seven), the mean error of the cumulative flow increases dramatically. This is due to the fact that the estimated average density involves the estimated initial number of vehicles, which was estimated with high accuracy in the first three cases, but poorly in case four to seven, especially case six and seven.

On the other hand, the measurement errors have very limited impacts on the MAPE of the estimated upstream flow-rate. This is because the upstream flow-rate is estimated based on estimated traffic parameters, which are relatively stable across all cases according to Table 3.

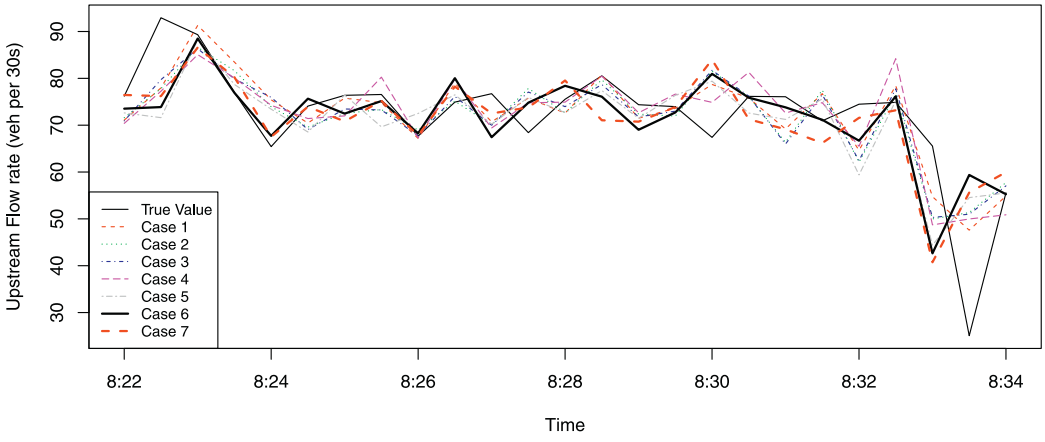
As expected, the variance of MAPE is higher with smaller CMR and higher measurement errors. Both estimated traffic states have smaller mean MAPEs in the second time period (8:05 AM-8:20 AM). This is probably because the second dataset contains fewer errors from the sources mentioned in Section 3.1.

5.4. A comparison with the STD method

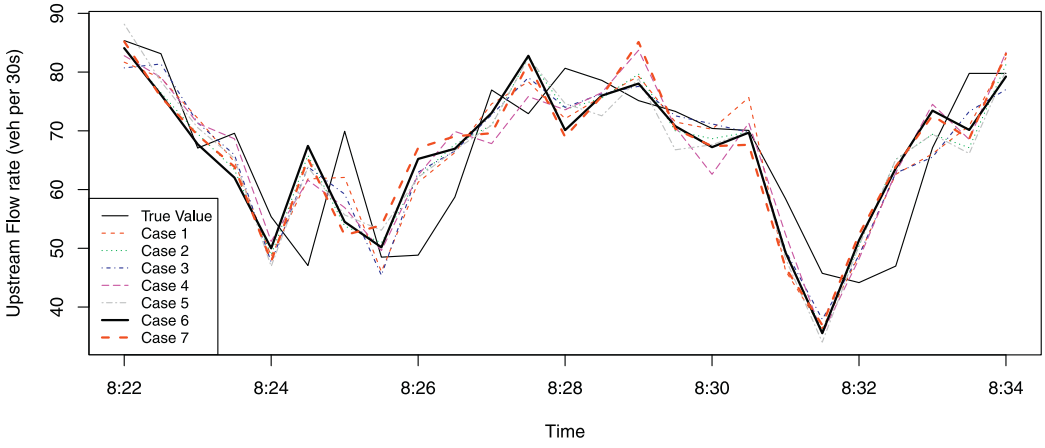
In this subsection, we compare the proposed estimation method with the STD method in Deng et al. (2013) using the same NGSIM I-80 dataset from 5:15 p.m. to 5:30 p.m. The warm-up period is set to be 5 min to match the experiment setting. We also set the initial guess of shock wave speed to be 12 mph and jam density to be 220 vpm to match the parameter settings. Three different CMRs,[2] 1%, 2%, and 5% were used in Deng et al. (2013), which is essentially the proportion of reidentification pairs used in the study. We estimate the average density according to (20) given different combinations of CMRs and flow counting intervals. Each combination is repeated 100 times, and the means and standard deviations of the MAPEs for different CMRs of reidentified vehicles and sampling intervals (the time-step size Δt) are shown in Fig. 6.

From Fig. 6 we can see that the CMR of the reidentified vehicles can impact the estimation results: the mean and variance of the MAPE are both higher with smaller CMR. But the effect of the time counting interval is trivial: in the case when the CMR is 1%, the mean increases slightly while the variance stays almost the same with larger counting interval.

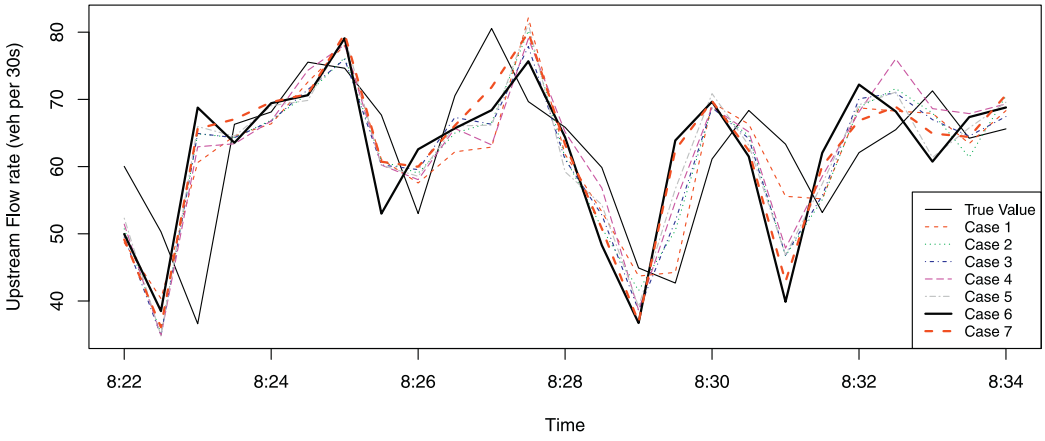
[2] In Deng et al. (2013), the entry/exit time of vehicles are assumed to be measured by AVI technology. Then the CMR is equivalent to the market penetration rate of AVI vehicles.



(a) First Time Period: 7:50 AM-8:05 AM

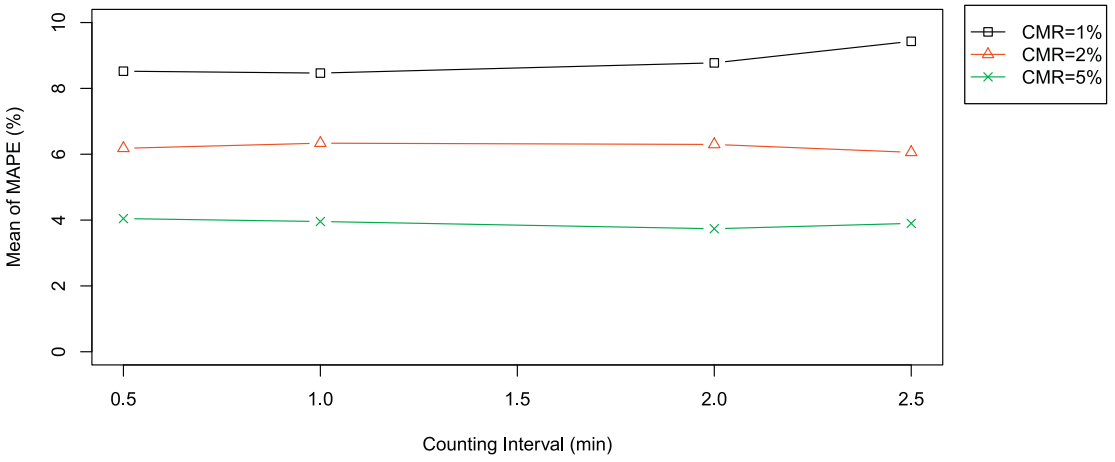


(b) Second Time Period: 8:05 AM - 8:20 AM

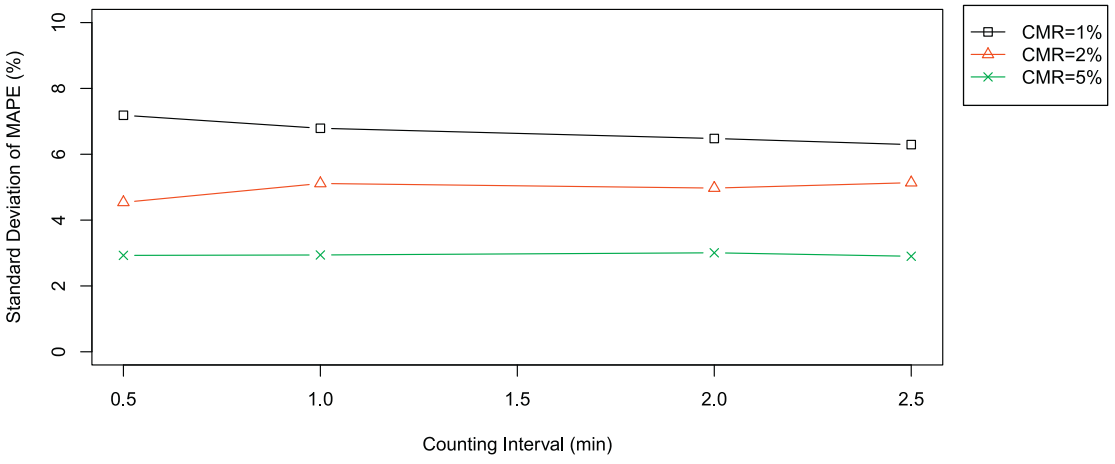


(c) Third Time Period: 8:20 AM- 8:35 AM

Fig. 5. Estimated upstream flow-rate v.s. ground truth.



(a) Mean MAPE by CMR and sampling interval



(b) SD of MAPE by CMR and sampling interval

Fig. 6. Impact of flow counting interval and CMR on average density estimation.

Comparing the results of Fig. 6 with those of Deng et al. (2013),[3] our method outperforms the STD method for all scenarios. For example, when the CMR is 1%, and the sampling interval is 1 min, the MAPE of the average density is about 24% by the STD method, but 8.5% by our method. Note that in Deng et al. (2013), the initial number of vehicles was assumed given and accurate, and W was predefined. In contrast, we estimate n_0, W, and K simultaneously. Thus, the simultaneous estimation framework yields better results without determining the initial number of vehicles and the shock wave speed separately.

6. Conclusion

This research proposed a new framework to simultaneously estimate states and parameters with Eulerian traffic count data from loop detectors and Lagrangian vehicle trajectory data from reidentification technologies. The method was developed based on Newell's simplified kinematic wave theory, which uses cumulative flow to describe the traffic states. We first formulated an optimization problem in terms of the initial number of vehicles and model parameters for a road segment, from which other traffic states can be calculated accordingly. We showed that the optimization problem can be decoupled and can be solved either analytically or with the Gauss–Newton method. We also pointed out a potential issue under steady traffic conditions. Finally we applied the method to the NGSIM data and demonstrated its validity based on the mean abso-

[3] Note that is Deng et al. (2013) only provided the result of one run.

lute percentage errors of both the average density and the upstream flow-rate. A comparison with the STD method in Deng et al. (2013) further highlighted the advantage of the simultaneous estimation framework.

The most important contribution of this study is the simultaneous framework to estimate both traffic states and model parameters. In particular, we formulated a single optimization problem, (16), from which one can analytically solve the initial number of vehicles and numerically calculate model parameters. This method is simpler than existing methods both conceptually and computationally. In addition, it offers improvements in the estimation accuracy compared to the STD method in Deng et al. (2013).

The proposed estimation method still suffers from several limitations. First, as shown in the real world example, the initial state estimation is sensitive to the measurement errors, especially when the flow conservation is violated due to measurement errors. However, the proposed estimation method does not have an explicit error model. Several recent studies (Laval et al., 2012; Blandin et al., 2012) have shown that unimodal uncertainty in initial/boundary condition would result in bimodal distribution of traffic states due to the nonlinearity of traffic. A possible way to address this issue is estimate using Bayesian inference with carefully chosen prior. But the same traffic estimation model proposed in this paper can still be used. Second, the current method only accepts cumulative flow counts and vehicle reidentification data. It would be ideal if we can extend the current method to accommodate other data types (for example, GPS trajectory data). Furthermore, the current method only applies to linear links with congested traffic state. The next step is to incorporate junction models and extend the method to fixed traffic states.

Acknowledgment

The authors would like to thank the Multi-Campus Research Program and Initiative on Sustainable Transportation of the University of California for its support on the research.

References

Bekiaris-Liberis, N., Roncoli, C., Papageorgiou, M., 2016. Highway traffic state estimation with mixed connected and conventional vehicles. IEEE Trans. Intell. Transp. Syst. 17 (12), 3484–3497.
Blandin, S., Couque, A., Bayen, A., Work, D., 2012. On sequential data assimilation for scalar macroscopic traffic flow models. Physica D 241 (17), 1421–1440.
Canepa, E.S., Claudel, C.G., 2012. Exact solutions to traffic density estimation problems involving the Lighthill–Whitham–Richards traffic flow model using mixed integer programming. In: In Intelligent Transportation Systems (ITSC), 15th International IEEE Conference on. IEEE, pp. 832–839.
Cassidy, M.J., 1998. Bivariate relations in nearly stationary highway traffic. Transp. Res. Part B 32, 49–59.
Charnes, A., Frome, E., Yu, P.-L., 1976. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. J. Am. Stat. Assoc. 71, 169–171.
Choi, J., 2014. Estimating Emissions by Modeling Freeway Vehicle Speed Profiles Using Point Detector Data. University of California, Irvine.
Claudel, C.G., Bayen, A.M., 2010. Lax–Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation. Part I: theory. IEEE Trans. Autom. Control 55, 1142–1157.
Claudel, C.G., Bayen, A.M., 2010. Lax—Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation. Part II: computational methods. IEEE Trans. Autom. Control Vol. 55, 1158–1174.
Coifman, B., 2002. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. Transp. Res. Part A 36 (4), 351–364.
Coifman, B., 2006. Vehicle level evaluation of loop detectors and the remote traffic microwave sensor. J. Transp. Eng. 132 (3), 213–226.
Coifman, B., Cassidy, M., 2002. Vehicle reidentification and travel time measurement on congested freeways. Transp. Res. Part A 36, 899–917.
Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations, pp. 26–27.
Daganzo, C.F., 2005. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transp. Res. Part B 39, 187–196.
Deng, W., Lei, H., Zhou, X., 2013. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: a three detector approach. Transp. Res. Part B 57, 132–157.
Evans, L., 1998. Partial differential equations. American Mathematical Society.
Greenshields, B.D., 1935. A Study of Traffic Capacity. Highway Research Board Proceedings, 14, pp. 448–477.
Hall, F.L., Allen, B.L., Gunter, M.A., 1986. Empirical analysis of freeway flow-density relationships. Transp. Res. Part A 20 (3), 197–210.
Herrera, J.C., Bayen, A.M., 2008. Traffic flow reconstruction using mobile sensors and loop detector data. Transportation Research Board 87th Annual Meeting.
Jeng, S.-T., 2007. Real-time Vehicle Reidentification System for Freeway Performance Measurements. UC Irvine. Dissertation
Jin, W.-L., 2013. A multi-commodity Lighthill–Whitham–Richards model of lane-changing traffic flow. Transp. Res. Part B 57, 361–377.
Jin, W.-L., 2015. Continuous formulations and analytical properties of the link transmission model. Transp. Res. Part B 74, 88–103.
Jin, W.-L., Li, L., 2007. First-in-first-out is violated in real traffic. In: Proceedings of Transportation Research Board Annual Meeting.
Laval, J., He, Z., Castrillon, F., 2012. Stochastic extension of Newell's three-detector method. Transp. Res. Rec. 2315, 73–80.
Lee, H., Coifman, B., 2011. Identifying and correcting pulse-breakup errors from freeway loop detectors. Transp. Res. Rec. 2256, 68–78.
Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. A 229, 317–345.
Milliken, G.A., 1990. Nonlinear regression analysis and its applications. Technometrics 32, 41–43.
Muñoz, L., Sun, X., Horowitz, R., Alvarez, L., 2003. Traffic density estimation with the cell transmission model. In: Proceedings of the 2003 American Control Conference, Vol. 5. IEEE, pp. 3750–3755.
Nantes, A., Ngoduy, D., Bhaskar, A., Miska, M., Chung, E., 2016. Real-time traffic state estimation in urban corridors from heterogeneous data. Transp. Res. Part C 66, 99–118.
Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transp. Res. Rec. 1855, 49–59.
Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: general theory. Transp. Res. Part B 27, 281–287.
Oh, C., Ritchie, S.G., Oh, J.-S., Jayakrishnan, R., 2002. Real-time origin-destination (OD) estimation via anonymous vehicle tracking. In: Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems. IEEE, pp. 582–586.
Oh, C., Tok, A., Ritchie, S.G., 2005. Real-time freeway level of service using inductive-signature-based vehicle reidentification system. IEEE Trans. Intell. Transp. Syst. 6, 138–146.
Payne, H. J., 1971. Models of Freeway Traffic and Control. Simulation Councils Proceedings series. Mathematical Models of Public Systems 1 (1), 51–61.
Punzo, V., Borzacchiello, M.T., Ciuffo, B., 2011. On the assessment of vehicle trajectory data accuracy and application to the next generation SIMulation (NGSIM) program data. Transp. Res. Part C 19 (6), 1243–1262.

Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4, 42–51.
Sun, C., Ritchie, S.G., Tsai, K., Jayakrishnan, R., 1999. Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways. Transp. Res. Part C 7, 167–185.
Sun, X., Muñoz, L., Horowitz, R., 2003. Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control. In: Proceedings of the 42nd IEEE Conference on Decision and Control, Vol. 6, pp. 6333–6338.
Thiemann, C., Treiber, M., Kesting, A., 2008. Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data. Transp. Res. Rec. 2008, 90–101.
Tossavainen, O.P., Work, D.B., 2013. Markov chain Monte Carlo based inverse modeling of traffic flows using GPS data. Netw. Heterogen. Media 8, 3.
USDOT, 2008. NGSIM–Next Generation Simulation. http://www.ngsim.fhwa.dot.gov/.
Wang, Y., Malinovskiy, Y., Wu, Y.-J., Lee, U.K., Neeley, M., 2011. Error modeling and analysis for travel time data obtained from bluetooth mac address matching. Department of Civil and Environmental Engineering. University of Washington.
Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transp. Res. Part B 39, 141–167.
Yang, H., Gan, Q., Jin, W.-L., 2011. Calibration of a family of car-following models with retarded linear regression methods. In: Proceedings of the 90th Annual Meeting of Transportation Research Board.