Second-order models and traffic data from mobile sensors

Benedetto Piccoli\textsuperscript{a}\textsuperscript{*} Ke Han\textsuperscript{b}\textsuperscript{†} Terry L. Friesz\textsuperscript{c}\textsuperscript{‡} Tao Yao\textsuperscript{c}\textsuperscript{§}

\textsuperscript{a}Department of Mathematics, Rutgers University, NJ 08102, USA
\textsuperscript{b}Department of Civil and Environmental Engineering, Imperial College London, London SW7 2BU, UK
\textsuperscript{c}Department of Industrial and Manufacturing Engineering, Pennsylvania State University, PA 16802, USA

Abstract

Mobile sensing enabled by on-board GPS or smart phones has become the primary source of traffic data. For sufficient coverage of the traffic stream, it is important to maintain a reasonable penetration rate of probe vehicles. From the standpoint of estimating higher-order traffic quantities such as acceleration/deceleration, emission rate and fuel consumption rate, it is desirable to examine the effectiveness of the sampling frequency of current sensing technology in capturing higher-order variations inherent in traffic stream. Of the two concerns raised above, the latter is rarely studied in the literature.

In this paper, we study the two characteristics of mobile sensing: penetration rate and sampling frequency, and their impacts on the quality of traffic estimation and reconstruction. A computational method is presented that integrates vehicle trajectory data into a second-order hydrodynamic model known as the phase transition model (Colombo, 2002a). We utilize the Next Generation SIMulation (NGSIM, 2006) dataset containing high time-resolution vehicle trajectories. It is demonstrate through extensive numerical study that while first-order traffic quantities can be accurately estimated using prevailing sampling frequency at a reasonably low penetration rate, higher-order traffic quantities tend to be misinterpreted due to insufficient sampling frequency of current mobile devices. We propose, for estimating emission and fuel consumption rates, a correction factor approach which is proven to yield improved accuracy via statistical validation.

Keywords: mobile sensing, GPS, phase transition model, higher-order traffic quantity, emission rate, fuel consumption rate

1 Introduction

1.1 Mobile sensing

In the era of mobile internet, mobile sensors such as global positioning system (GPS) and smart phones have become the primary means of collecting traffic-related information (Herrera et al., 2010). Attractive features of GPS-based mobile sensing include potentially complete spatial and temporal coverage of the traffic network and high positioning accuracy. Traffic data on
first-order quantities, including velocity, density and travel time, are often processed in connection with first-order traffic flow models such as the Lighthill-Whitham-Richards (LWR) model or its discrete versions such as the cell transmission model (Daganzo, 1994, 1995), with an incomplete list of references including Claudel and Bayen (2010a,b); Wang and Papageorgiou (2005); Work et al. (2010); Yuan et al. (2011). Another important class of data, namely, higher-order traffic quantities including acceleration/deceleration, emission rate and fuel consumption rate, are not immediately available through traditional mobile sensing technologies and are insufficiently captured by first-order traffic models.

Moreover, existing studies on mobile sensing are primarily concerned with the penetration rate of probe vehicles (Demers et al., 2006; Kwon et al., 2007; Yim and Cayford, 2001); little attention is raised on the impact of sampling frequency on the quality of estimation, especially regarding higher-order traffic quantities. Typical mobile sensors such as GPS devices report the location of a probe vehicle every three to four seconds, thus derived information on velocity and acceleration etc. is averaged in the same time period. However, there exist higher-order variations inherent in traffic stream that take place on a much smaller time scale, especially when the traffic is congested or unstable. Many of those higher-order variations are unaccounted for by existent sensing paradigms.

1.2 Data description

In order to address research issues related to higher-order traffic quantities, we focus on dataset that contain high time-resolution vehicle trajectories provided by the Next Generation SIMulation (NGSIM) (NGSIM, 2006). The NGSIM program collected high-quality primary traffic and trajectory data intended to support the research and testing of the traffic models and open behavioral algorithms. It collected and processed traffic data on a segment of Interstate 80 located in Emeryville, California on April 13, 2005. A total of 45 minutes of data are available, segmented into three 15 minute periods. The dataset contains vehicle trajectory recorded at a high precision of every 1/10 second. Derived information on instantaneous velocity and acceleration is also provided. A detailed description of the NGSIM field experiment will be provided in Section 4.

1.3 Contribution and findings

This paper addresses the issue of estimating both first-order and higher-order traffic quantities mentioned above, by employing a second-order traffic flow model known as the phase transition model (PTM) (Colombo, 2002a) as well as its variations. The dataset under consideration is provided by the NGSIM program which contains high time-resolution vehicle trajectories on a segment of I-80. Such detailed vehicle trajectories provide unique information, especially on higher-order variations, of traffic stream that is unavailable through traditional mobile or fixed sensors. We propose several computational schemes for reconstructing Eulerian and Lagrangian traffic quantities using vehicle trajectory data as input. We will demonstrate how the estimation quality deteriorates with less frequent sampling of vehicle locations and with lower probe penetration rates. In particular, we will evaluate the performance of existent sensing devices (GPS) in reconstructing first- and higher-order quantities.

The main contributions and findings made by this paper are summarized as follows.

- A computational scheme for integrating vehicle trajectory data into the second-order phase transition model is proposed. The scheme takes into account different assumptions
on the traffic state, and produces estimation results of first- and higher-order quantities along each vehicle trajectory (Lagrangian estimation) or at a fixed location (Eulerian estimation).

- Along each vehicle trajectory, the estimation of higher-order traffic quantities deteriorates significantly when the sampling frequency decreases, while the estimation of first-order quantities remain relatively accurate with the same sampling frequency.

- For the estimation with fixed locations, we provide numerical results on reconstructed first- and higher-order Eulerian quantities when both the sampling frequency and the probe penetration rate vary. Both first- and higher-order quantities enjoy an improved estimation accuracy due to the effect of averaging multiple measurements; but the latter still suffer from under sampling.

- As an application of the proposed data fusion methodology, we perform the estimation of automobile emission and fuel consumption rates on a road segment using appropriate sampling frequencies and probe penetration rates. The true values tend to be underestimated due to the negligence of higher-order variations in velocity and acceleration. A correction factor approach is proposed to fix such estimation bias, which is shown to improve the estimation accuracy under statistical validation.

This paper takes the unique advantage of complete coverage of the traffic stream at a high time-resolution provided by the NGSIM dataset. We provide insights on the performance of existent mobile sensing paradigms as well as supporting numerical results that serve the practical purpose of advising practitioners in assessing and deploying sensing infrastructure.

1.4 Organization of this paper

The rest of this paper is organized as follows. Section 2 describes several numerical approximations of various Eulerian and Lagrangian traffic quantities, based only on vehicle trajectory data. Section 3 provides a review on the phase transition model (PTM), followed by several computational schemes for fusing vehicle trajectory data into the PTM under different assumptions on the traffic stream. A modal model for emission and fuel consumption estimation is also illustrated in this section. In Section 4 we calibrate the PTM using the NGSIM data. Section 5 assesses the estimation quality for first- and second-order quantities along vehicle trajectories, when the sampling frequency varies. Section 6 performs and evaluates the estimation of Eulerian (cell-based) quantities. We also propose a correction factor approach for improving the accuracy and reliability of estimating vehicle emission and fuel consumption rates.

2 Estimating basic traffic quantities

Onboard sensors such as GPS measure the location of a moving car every $\delta t$ seconds, where $\delta t$ is related to the device’s characteristics such as desired precision and transmission capacity. For a given vehicle, we denote by $x(t)$ its location and by $v(t)$ its velocity at time $t$. Assume that the location is recorded at three consecutive time shots $t_1, t_2$ and $t_3$ with $t_2 - t_1 = t_3 - t_2 = \delta t$. From these measurements one can deduce the approximate velocities in the time intervals $[t_1, t_2]$ and $[t_2, t_3]$ as

$$v_{1,2} = \frac{x(t_2) - x(t_1)}{\delta t}, \quad v_{2,3} = \frac{x(t_3) - x(t_2)}{\delta t} \quad (2.1)$$
The velocity at time $t_2$ is approximated as
\[ v(t_2) \approx \frac{v_{1,2} + v_{2,3}}{2} = \frac{x(t_3) - x(t_1)}{2\delta t} \]  
(2.2)

One also gets
\[ \frac{D}{Dt}v(t_2) \approx \frac{v_{2,3} - v_{1,2}}{\delta t} = \frac{x(t_3) - 2x(t_2) + x(t_1)}{\delta t^2} \]  
(2.3)

where $D/Dt = d/dt + v \cdot d/dx$ denotes the material derivative in the Eulerian coordinates corresponding to the acceleration of the car in the Lagrangian coordinate.

Another important quantity to estimate is the spatial variation of velocity in the Eulerian coordinate. For notation convenience, we set $x_i = x(t_i)$. Assuming a mild variation in time of the Eulerian velocity $v(t, x)$, we write:
\[ v\left(t_2, \frac{x_2 + x_1}{2}\right) \approx v_{1,2}, \quad v\left(t_2, \frac{x_3 + x_2}{2}\right) \approx v_{2,3} \]
from which we get, by setting $\delta x = \frac{x_3 + x_2}{2} - \frac{x_2 + x_1}{2}$, that
\[ \frac{\partial}{\partial x} v\left(t_2, \frac{x_3 + x_2}{2} \right) \approx \frac{v_{2,3} - v_{1,2}}{\delta x} \]
\[ \frac{\partial}{\partial x} v\left(t_2, \frac{x_3 + 2x_2 + x_1}{4}\right) \approx \frac{v_{2,3} - v_{1,2}}{\frac{x_3 - x_1}{2}} = \frac{2}{\delta t} \frac{x_3 - 2x_2 + x_1}{x_3 - x_1} \]  
(2.4)

Clearly such approximation is acceptable as long as the variation between $v_{1,2}$ and $v_{2,3}$ is not too large.

3 Model fitting using vehicle trajectory data

Throughout this paper, we consider the phase transition model (PTM) as well as its variations to represent traffic dynamics. Meanwhile, we employ a modal model (Post et al., 1984) to calculate vehicle emission and fuel consumption rates. These models will be elaborated in this section. We will also discuss how first- and higher-order traffic quantities may be estimated with the PTM and vehicle trajectory data.

3.1 The phase transition model

The hyperbolic phase transition model (PTM) for vehicular traffic flow is introduced by Colombo (2002a) and Colombo (2002b), and studied subsequently by Colombo and Corli (2002); Colombo et al. (2007) and Blandin et al. (2012). The PTM belongs to a model class known as second-order, since it captures second-order variations of traffic in addition to the average velocity and density. Other second-order models include the Payne-Whitham model proposed independently by Payne (1971, 1979) and Whitham (1974), and the Aw-Rascle-Zhang model developed by Aw and Rascle (2000) and Zhang (2002). The phase transition model is motivated by the empirical observation that when the vehicle density exceeds certain critical value, the density-flow pairs are scattered in a two-dimensional region, instead of
forming a one-to-one relationship. This is in contrast to the first-order kinematic wave models such as the classical Lighthill-Whitham-Richards (LWR) model \cite{Lighthill1955, Richards1956}.

As the name suggests, the phase transition model consists of two phases: the uncongested phase and the congested phase. In the uncongested phase, the dynamic is governed by the LWR model

\begin{equation}
\begin{aligned}
\rho_t + [\rho \cdot v]_x &= 0 \\
v &= v(\rho)
\end{aligned}
\end{equation}

where \(\rho\) denotes density, and velocity \(v\) is expressed as a function of density only. On the other hand, the congested phase is governed by the following system of conservation laws:

\begin{equation}
\begin{aligned}
\rho_t + [\rho \cdot v]_x &= 0 \\
q_t + [(q - q^*) \cdot v]_x &= 0 \\
v &= v(\rho, q)
\end{aligned}
\end{equation}

where the velocity \(v(\rho, q)\) depends not only on the local density \(\rho\), but also on \(q\) which describes the perturbation or deviation from the equilibrium state. \(q^*\) is a given constant. One form of \(v(\rho, q)\) is

\begin{equation}
\begin{aligned}
v(\rho, q) &= \left(1 - \frac{\rho}{\rho_{jam}}\right) \cdot \frac{q}{\rho}
\end{aligned}
\end{equation}

for some positive parameters \(A\), \(B\) and \(\rho_{jam}\).

One has a lot of flexibility in choosing the system of equations for the congested phase. For instance, \cite{Goatin2006} proposes a phase transition model which employs the Aw-Rascle-Zhang equations \cite{Aw2000, Zhang2002} for the congested phase. Furthermore, one may consider a version of the PTM by taking into account the reaction time of drivers. More specifically, following \cite{Siebel2006a}, we write

\begin{equation}
\begin{aligned}
q_t + [(q - q^*) \cdot v]_x &= \frac{q^* - q}{T - \tau}
\end{aligned}
\end{equation}

as the second equation for the congested phase. The right hand side is called the Siebel-Mauser type source term. Here \(\tau\) (in second) is a reaction time and typically varies within [0.5, 2] \cite{Koppa1997}. In this paper, the value of \(\tau\) is set to be 1 (second). Following \cite{Siebel2006a} and \cite{Kuene1997}, we choose \(T = 2/3\) (second) and therefore \(T - \tau = -1/3\). Finally \(T - \tau\) is modeled as a factor which can be positive (for very small or very high densities) or negative (for intermediate densities). One has to notice that a negative factor gives rise to stable traffic, while a positive ones produces instabilities \cite{Siebel2006a}.

### 3.2 Estimating traffic quantities associated with the PTM

The main variables of the PTM are the vehicle density \(\rho\) (first-order) and the perturbation \(q\) (second-order). It is possible to estimate these quantities using nothing but vehicle trajectory data, as we demonstrate below. Note that subsequent derivations are established based on the congested phase of the PTM for two reasons. One reason is that the NGSIM dataset suggest
the presence of medium to heavy congestion during the time of study; another is that it is mainly in the congested phase of traffic where higher-order variations take place.

Subsequent derivations are dependent on the type of assumptions made on the model and on the traffic condition.

### 3.2.1 Phase transition model with source and strongly stable traffic

We consider the PTM with velocity in the congested phase given by

\[ v(\rho, q) = A(\rho_{jam} - \rho) + B(q - q^*)(\rho_{jam} - \rho) \]  

(3.10)

In addition, we add a Siebel-Mauser type source term. Thus the equation for the congested phase is

\[ q_t + [(q - q^*) \cdot v]_x = \frac{q^* - q}{T - \tau} \]  

(3.11)

Assume that the traffic is strongly stable, that is, during the data measurement we may assume \( \rho_t, \rho_x, q_x \approx 0 \) with only \( q_t \) being non-vanishing. By such strong stability assumption, (3.11) reduces to

\[ q_t = \frac{q^* - q}{T - \tau} \]  

(3.12)

Then we can derive

\[
\frac{Dv}{Dt} = \partial_t v(\rho(t, x), q(t, x)) + v \cdot \partial_x v(\rho(t, x), q(t, x)) \\
= v_\rho \cdot \rho_t + v_q \cdot q_t + v \cdot (v_\rho \cdot \rho_x + v_q \cdot q_x) \\
\approx v_q \cdot q_t \\
= B(\rho_{jam} - \rho) \cdot \frac{q^* - q}{T - \tau} \\
= \frac{1}{T - \tau} B(\rho_{jam} - \rho)(q^* - q) = \frac{1}{T - \tau} (A(\rho_{jam} - \rho) - v) 
\]  

(3.13)

Taking into account only the measurements of \( v \) and \( Dv/Dt \), and by introducing variables \( \hat{\rho} \doteq \rho_{jam} - \rho, \hat{q} \doteq q - q^* \), we deduce from (3.13) that

\[
\hat{\rho} = (\rho_{jam} - \rho) = \frac{1}{A} \left( v + (T - \tau) \frac{Dv}{Dt} \right) 
\]

(3.14)

\[
\hat{q} = (q - q^*) = - \frac{A(T - \tau)}{B} \frac{Dv}{Dt} \frac{Dv}{Dt} 
\]

(3.15)

Following our discussion at the end of Section 3.1, we take \( T - \tau = -\frac{1}{3} \). The velocity \( v \) is estimated according to (2.2), and the acceleration \( Dv/Dt \) is computed from (2.3).

### 3.2.2 Phase transition model with source and less stable traffic

In this case, we rely on the less strongly stable assumption on traffic. In other words, we no longer assume that \( \rho_x \) vanishes, while still neglecting \( \rho_t \) and \( q_x \). Then equation (3.11) becomes

\[ q_t + v_x(q - q^*) = \frac{q^* - q}{T - \tau} \]  

(3.16)
We can now write
\[ \frac{Dv}{Dt} = v_t + vv_x \approx v_q q_t + vv_x = v_q \left( \frac{q^* - q}{T - \tau} - (q - q^*)v_x \right) + vv_x \]  (3.17)

Recalling the variables \( \dot{\rho} \approx \rho_{jam} - \rho \), \( \dot{q} = q - q^* \), we obtain
\[ v = A(\rho_{jam} - \rho) + B(q - q^*)(\rho_{jam} - \rho) = \dot{\rho} (A + B\dot{q}) \]  (3.18)

and deduce from (3.17)–(3.18) that
\[ \frac{Dv}{Dt} - vv_x = -v_q \left( \frac{\dot{q}}{T - \tau} + \dot{q}v_x \right) = -B\dot{\rho} \left( \frac{\dot{q}}{T - \tau} + \dot{q}v_x \right) \]  (3.19)

From (3.18) and (3.19) we immediately get the expressions for \( \dot{\rho} \) and \( \dot{q} \) in terms of \( v \), \( v_x \) and \( \frac{Dv}{Dt} \).
\[ \dot{\rho} = \frac{1}{A} \left( v + (T - \tau) \frac{Dv}{Dt} \right) \]  (3.20)
\[ \dot{q} = \frac{A}{B} \left( T - \tau \right) \left( vv_x - \frac{Dv}{Dt} \right) \]  (3.21)

The quantities \( \frac{Dv}{Dt}, v_x \) are given by (2.3) and (2.4) respectively. Regarding velocity \( v \), notice that if one approximates \( v \) with (2.1), then
\[ \frac{Dv}{Dt} - vv_x \approx \frac{x(t_{i+1}) - 2x(t_i) + x(t_{i-1})}{\Delta t^2} - \frac{2}{\Delta t} \left( \frac{\Delta}{\delta t} \right) \frac{x(t_{i+1}) - x(t_{i-1})}{\Delta \delta t} \]  (3.22)

which renders (3.21) identically zero. To avoid such a trivial case, one should instead approximate \( v \) by
\[ v(t_2) \approx v_{1,2} = \frac{x(t_2) - x(t_1)}{\Delta t}, \quad \text{or} \quad v(t_2) \approx v_{2,3} = \frac{x(t_3) - x(t_2)}{\Delta t} \]

### 3.2.3 Phase transition model without source term

We consider the PTM without the Siebel-Mauser type source term:
\[
\begin{aligned}
\rho_t + \mathbf{q} \cdot \mathbf{v}_x &= 0 \\
q_t + ((q - q^*) \cdot v)_x &= 0 \\
\mathbf{v} \cdot \mathbf{q} &= A(\rho_{jam} - \rho) + B(q - q^*)(\rho_{jam} - \rho)
\end{aligned}
\]  (3.23)

Computation in this case is less straightforward than the previous two. We start with the identity
\[ \frac{Dv}{Dt} = v_t + v \cdot v_x = v_{\rho} \rho_t + v_q q_t + v(v_{\rho} \rho_x + v_q q_x) \]

Using (3.23), we have
\[ \frac{Dv}{Dt} = v_{\rho}(-\rho_x v - \rho v_x) + v_q(-q_x v - (q - q^*)v_x) + v(v_{\rho} \rho_x + v_q q_x) \]
\[ = -v_x(v_{\rho} \rho + v_q(q - q^*)) \]
\[ = v_x(A \rho + B(q - q^*)(2 \rho - \rho_{jam})) \]  (3.24)
Recall the variables $\dot{\rho} = \rho_{jam} - \rho$ and $\hat{q} = q - q^*$. Combining (3.24) with the expression of $v(\rho, q)$ and solving for $\rho$ we get

$$\rho = \frac{1}{B\hat{q}} \left( v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} \right) \quad (3.25)$$

One immediate observation from (3.25) is that $\hat{q}$ and $v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam}$ always have the same sign. Substituting (3.25) into (3.24), we get

$$B^2\rho_{jam}\hat{q}^2 + B \left( 2A\rho_{jam} - \frac{1}{v_x} \frac{Dv}{Dt} - 2v \right) \hat{q} - A \left( v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} \right) = 0 \quad (3.26)$$

which is a quadratic equation in the variable $\hat{q}$. The discriminant of such quadratic equation is

$$\Delta = 4B^2 \left( v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} \right)^2 + 4AB^2\rho_{jam} \left( v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} \right)$$

$$= 4B^2 \left( v + \frac{1}{v_x} \frac{Dv}{Dt} \right)^2 - A\rho_{jam} \sqrt{A\rho_{jam} v} \quad (3.27)$$

In order for any meaningful real root of (3.26) to exist, a necessary condition is that $\Delta$ is nonnegative, that is,

$$\left| v + \frac{1}{2v_x} \frac{Dv}{Dt} \right| \geq \sqrt{A\rho_{jam} v} \quad (3.28)$$

If the equality holds in (3.28), a real solution of $\dot{q}$ exists if and only if

$$v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} < 0 \quad (3.29)$$

In the case where the strict inequality of (3.28) holds, two distinct roots $\hat{q}_1$ and $\hat{q}_2$ exist. We distinguish between two cases:

- If

  $$v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} > 0 \quad (3.30)$$

  Then $\hat{q}_1\hat{q}_2 < 0$. Equation (3.26) has one positive root and one negative root. By (3.25), one should choose the positive root since $\rho$ must be non-negative.

- If

  $$v + \frac{1}{v_x} \frac{Dv}{Dt} - A\rho_{jam} \leq 0 \quad (3.31)$$

  Then $\hat{q}_1\hat{q}_2 \geq 0$. Equation (3.26) has too roots with the same sign. In view of (3.25) and (3.31), to ensure that $\rho$ is nonnegative, both $\hat{q}_1$ and $\hat{q}_2$ should be nonpositive and at least one root is negative. This in turn requires that

  $$2A\rho_{jam} - \frac{1}{v_x} \frac{Dv}{Dt} - 2v > 0 \quad (3.32)$$

  In view of (3.31), a sufficient condition for (3.32) to hold is $\frac{1}{v_x} \frac{Dv}{Dt} > 0$.

\[^{1}\text{zero is considered to have both positive sign and negative sign.}\]
It turns out that the above analysis can be more easily interpreted in a discrete-time setting. Let us consider three consecutive time points \( t_i-1, t_i, t_{i+1} \). First, notice that

\[
\frac{1}{v_x} \frac{Dv}{Dt} \approx \frac{\delta t}{2} \cdot \frac{x(t_{i+1}) - x(t_{i-1})}{2} \cdot \frac{x(t_{i+1}) - 2x(t_i) + x(t_{i-1})}{\delta t^2}
\]

\[
= \frac{x(t_{i+1}) - x(t_{i-1})}{2\delta t} \approx v(t_i) \geq 0
\] (3.33)

In light of this calculation, the feasibility condition (3.28) becomes

\[
\frac{3}{2} v \geq \sqrt{A \rho_{jam} v}, \quad \text{or} \quad v \geq \frac{4}{9} A \rho_{jam}
\] (3.34)

Using (3.33), the decision rules following (3.28) can be explicitly summarized as follows.

**Algorithm 1 PTM without source term**

```plaintext
if \( v < \frac{4}{9} A \rho_{jam} \) then the system has no real solutions

else if \( v = \frac{4}{9} A \rho_{jam} \) then \( \rho = \frac{\rho_j}{3}, \hat{q} = -\frac{A}{3B} \)

else if \( v > \frac{1}{2} A \rho_{jam} \) then

\[ \rho = \frac{2 \rho_{jam}(2v - A \rho_{jam})}{3v - 2A \rho_{jam} + \sqrt{9v^2 - 4A \rho_{jam}v}}, \quad \hat{q} = \frac{3v - 2A \rho_{jam} + \sqrt{9v^2 - 4A \rho_{jam}v}}{2B \rho_{jam}} \]

else

\[ \rho = \frac{2 \rho_{jam}(2v - A \rho_{jam})}{3v - 2A \rho_{jam} - \sqrt{9v^2 - 4A \rho_{jam}v}}, \quad \hat{q} = \frac{3v - 2A \rho_{jam} - \sqrt{9v^2 - 4A \rho_{jam}v}}{2B \rho_{jam}} \]

end if

end if
```

**Remark 3.1.** We note that the above computational procedure does not produce any result if the velocity is below \( \frac{4}{9} A \rho_{jam} \), which is around 22 (foot/second) according to our calibration result presented later. Thus such method is quite restrictive in application, despite the fact that it is consistent with the original phase transition model.

### 3.3 Estimating emission and fuel consumption rates

We consider a model of emission and fuel consumption rates that is based on the modal operation of a vehicle. That is, emission/fuel consumption is directly related to the vehicle operating modes such as idle, steady-state cruise, acceleration, deceleration and so forth. Such model requires detailed vehicle trajectory data available in the NGSIM dataset. According to the power demand based emission/fuel consumption model proposed by Post et al. (1984), the instantaneous total power demand \( Z \) of a vehicle with mass \( m \) (in kg) is given by

\[
Z = (0.04 v + 0.5 \times 10^{-3} v^2 + 10.8 \times 10^{-6} v^3) + \frac{m v}{1000 \ 3.6} \left( \frac{a}{3.6} + 9.81 \sin \theta \right)
\] (3.35)
where $Z$ is in kilowatt, the velocity $v$ is in km/h, the acceleration/deceleration $a$ is in km/h per second. $\theta$ denotes road degree. [Post et al. (1984)] also propose the following model of hydrocarbon (HC) emission rate for spark ignition vehicles based on field experiments

$$r_{HC}(t) = \begin{cases} 52.8 + 4.2Z & Z > 0 \\ 52.8 & Z \leq 0 \end{cases}$$

(3.36)

where the hydrocarbon emission rate $r_{HC}(t)$ is in gram/hour. The instantaneous fuel consumption (FC) model is

$$r_{FC}(t) = \begin{cases} 2.35 + 0.55Z & Z > 0 \\ 2.35 & Z \leq 0 \end{cases}$$

(3.37)

where the fuel consumption rate $r_{FC}(t)$ is in liter/hour.

4 Model Calibration

Before we implement the computational procedures proposed in the previous section, it remains to estimate the values for $A$, $B$ and $\rho_{jam}$ appearing in (3.10) and subsequent calculations. We will make such estimation dependent on the dataset to be utilized for our numerical experiment.

To this end, we consider the northbound of I-80 located in Emeryville, CA. The highway segment of interest spans 1650 feet in length with an on-ramp at Powell Street and an off-ramp at Ashby Avenue. The highway segment has six lanes with the leftmost lane being a high-occupancy vehicle (HOV) lane, and the rightmost one being a merge lane. Data were collected using several video cameras. Digital video images were collected over an approximate five-hour period from 2:00 pm to 7:00 pm on April 13, 2005. Complete vehicle trajectories transcribed at a resolution of 1 frame per 0.1 second, along with vehicle type, lane identification and so forth were recorded and processed over three time slots: 4:00 pm - 4:15 pm, 5:00 - 5:15 pm, and 5:15 - 5:30 pm. The layout of the study area is shown in Figure 1.

4.1 Estimating the density-flow relationship

We begin with estimating the density-flow relationship needed for the congested phase represented by Eqn. (3.6). For modeling accuracy we exclude data collected on the HOV lane (# 1) and the merging lane (# 6), since these lanes are not well-represented by the congested phase of the PTM.

Unlike the LWR model where the density-flow relation is expressed as a single-valued function, the fundamental diagram corresponding to the congested phase of PTM is a multi-valued map. This means that a given density $\rho$ corresponds to a continuum range of velocity $v(\rho, q)$, $q \in [q_{min}, q_{max}]$. In order to identify all the possible values of $v(\rho, q)$ in the NGSIM dataset for a given density value $\rho$, we partition the temporal-spatial domain into small bins $C_{ij} = [t_{i-1}, t_i] \times [x_{j-1}, x_j]$, $i = 1, \ldots, N_T$, $j = 1, \ldots, N_X$ where $i$ and $j$ indicate the time step and the spatial step respectively. The average density associated with $C_{ij}$ is estimated by the number of vehicles whose trajectories indicate their presence in the road segment $[x_j, x_{j+1}]$ during time interval $[t_i, t_{i+1}]$. The velocity inside $C_{ij}$ is calculated as the mean of all velocity measurements collected within this bin. Figure 2 illustrates such a procedure. The dimension of the bins used to construct the density-flow relation in the congested phase is 4 (seconds) $\times$ 400 (feet).
Figure 1: The study area spans 1650 feet in length in the northbound of Interstate 80 located in Emeryville, CA.

Figure 2: Estimation of density and velocity inside a bin. Curves represent vehicle trajectories where the locations are recorded at the solid dots. For the depicted scenario, the occupancy of bin $C_{ij}$ is three vehicles; the average velocity is taken as the mean of velocities measured/calculated at the four dots inside $C_{ij}$. 
The flow within \( C_{ij} \) is calculated as the product of the density and the average velocity. The density-flow data plots for the time periods 4:00 pm - 4:15 pm and 5:00 - 5:15 pm are shown in Figure 3. The combined data plots for the whole study period (4:00 pm - 4:15 pm, and 5:00 pm - 5:30 pm) are shown in Figure 4.

![Figure 3: The fundamental diagram for the PTM expressed as a set-valued function of density.](image)

![Figure 4: The fundamental diagram for the PTM expressed as a set-valued function of density.](image)

### 4.2 Constructing the congested region in the fundamental diagram

Equation (3.10) suggests an affine density-velocity relation when the perturbation \( q - q^* \) is zero. Notice that the perturbation can have different signs. We normalize \( q - q^* \) such that \( q - q^* \in [-1, 1] \). Therefore the upper and lower envelopes of the congested domain in the density-velocity relationship depicted in the right half of Figure 4 are respectively

\[
\begin{align*}
A(\rho_{jam} - \rho) + B(\rho_{jam} - \rho) & \quad \text{(upper envelop)} \\
A(\rho_{jam} - \rho) - B(\rho_{jam} - \rho) & \quad \text{(lower envelop)}
\end{align*}
\]

Accordingly, the upper and lower envelops of the congested domain in the density-flow relationship depicted in the left half of Figure 4 are

\[
\begin{align*}
A(\rho_{jam} - \rho)\rho + B(\rho_{jam} - \rho)\rho & \quad \text{(upper envelop)} \\
A(\rho_{jam} - \rho)\rho - B(\rho_{jam} - \rho)\rho & \quad \text{(lower envelop)}
\end{align*}
\]

We choose \( A = 350, B = 160, \) and \( \rho_{jam} = 0.14 \) (vehicle/foot). The resulting congested domains are shown as the areas formed by the black curves in Figure 4.
5 Estimating traffic quantities along vehicle trajectories

In this section, we present the estimation results associated with various first- and higher-order traffic quantities along vehicle trajectories. Those quantities include: velocity and acceleration given by (2.2)-(2.3), vehicle density given by (3.14), (3.20) or Algorithm 1, perturbation given by (3.15), (3.21) or Algorithm 1 and the emission and fuel consumption rates given by (3.36)-(3.37). Different sampling frequencies will be used to investigate the deterioration of estimation quality due to under sampling. Recall that the vehicle locations in the NGSIM raw dataset are provided every 1/10 second. In order to accommodate different sampling frequencies, we introduce integer $N$ and extract data from the raw dataset every $N$ points. For example, $N = 30$ implies a sampling period of $1/10 \times 30 = 3$ seconds. We then compute the relative $L^1$ error between the quantities obtained with $N > 1$ and the ones obtained with $N = 1$, i.e., the ground truth.

All the numerical results reported below are based on data collected during 4:00 pm - 4:15 pm and 5:00 pm - 5:30 pm. The total numbers of vehicles involved in these time periods are over 9000.

5.1 Velocity and acceleration

Vehicle speed and acceleration are the most fundamental quantity of our numerical study. Following the steps explained immediately above, we compute and summarize the mean and standard deviation of the relative $L^1$ errors, and present them in Table 1. It is apparent that the acceleration estimation is more susceptible to under sampling than velocity estimation. This is because much of the higher-order variation in the acceleration profile takes place on a small time scale; this can be confirmed by the highly oscillatory pattern in the acceleration profile, which cannot be sufficiently captured by under sampling.

| $N$ | Mean (%) | Standard deviation (%) |
|-----|----------|------------------------|
|     | $\|v_{true} - v_N\|_{L^1}$ | 3.27 | 4.72 | 6.36 | 7.58 |
|     | $\|v_{true}\|_{L^1}$ | 1.50 | 1.81 | 2.31 | 2.87 |
|     | $\|a_{true} - a_N\|_{L^1}$ | 88.12 | 98.77 | 102.23 | 102.53 |
|     | $\|a_{true}\|_{L^1}$ | 9.52 | 3.10 | 2.04 | 2.04 |

Table 1: Relative errors of velocity and acceleration estimation using different sampling periods. $v_{true}$ and $a_{true}$ denote the ground-truth velocity and acceleration; $v_N$ and $a_N$ denote the reconstructed velocity and acceleration with different values of $N$.

5.2 Vehicle density and perturbation

The density and the perturbation are estimated by (3.14)-(3.15) when the traffic is assumed to be strongly stable, by (3.20)-(3.21) when the traffic is less stable, and by Algorithm 1 when there is no source term in the PTM.

The mean and standard deviation of the relative $L^1$ error are summarized in Table 2 for the strongly stable case, in Table 3 for the less stable case, and in Table 4 for the no source case. Here $\hat{\rho}_{true}$ (or $\rho_{true}$) and $\hat{q}_{true}$ denote the ground-truth density and perturbation; $\hat{\rho}_N$ (or $\rho_N$) and $\hat{q}_N$ denote the reconstructed density and perturbation using under sampling.
We observe from all three cases that the estimation of the first-order quantity $\hat{\rho}$ or $\rho$ is far more accurate than the estimation of the second-order quantity $\hat{q}$, when the sampling frequency decreases. Such observation coincides with our earlier assertion that higher-order variations in the traffic stream are overlooked by lower sampling frequency.

Note that compared to the PTM with source, the without-source case has a much lower error in estimating both density and perturbation. We explain such result by first recalling that the computational scheme for the no-source case works only when $v > \frac{4}{9} A \rho_{jam} \approx 21.8$ (foot/second). Thus in our computation we only utilize those vehicle trajectories with speed above 21.8 (foot/second), which constitute a small portion of the dataset since Figure 4 shows that most vehicle speeds are below such threshold. When the speed is relatively high it is more likely that the driving condition is stable, this may contribute to the improvement in the estimation quality. In addition, the exact PTM with no source term makes no ad hoc assumptions on the traffic condition, which may yield a better representation of traffic dynamics.

5.3 Emission and fuel consumption rates

The power demand function, emission rate and fuel consumption rate are estimated according to (3.35), (3.36) and (3.37) respectively, where the velocity $v$ and acceleration $a$ are approximated by (2.2) and (2.3).

Table 5 summarizes the relative errors in estimating the power demand $Z$, hydrocarbon
Table 2: Estimation of $\hat{\rho}$ and $\hat{q}$ based on the phase transition model with source and assuming strongly stable traffic. $\rho_{\text{true}}$ and $q_{\text{true}}$ denote the ground-truth density and perturbation; $\hat{\rho}_N$ and $\hat{q}_N$ denote the reconstructed density and perturbation using under sampling.

|       | Mean (%) | Standard deviation (%) |
|-------|----------|------------------------|
|       |          |                        |
|       | Mean (%) | Standard deviation (%) |
|       |          |                        |
|       | Mean (%) | Standard deviation (%) |
|       |          |                        |

Table 3: Estimation of $\hat{\rho}$ and $\hat{q}$ based on the phase transition model with source and assuming less stable traffic. $\rho_{\text{true}}$ and $q_{\text{true}}$ denote the ground-truth density and perturbation; $\hat{\rho}_N$ and $\hat{q}_N$ denote the reconstructed density and perturbation using under sampling.

emission rate $r^{HC}$ and fuel consumption rate $r^{FC}$, when different sampling frequencies are employed. It is expected in general that estimation of these higher-order quantities suffer from under sampling. However, there are a few qualitative differences: the estimation of $r^{HC}$ deteriorate less severely compared to the other two as the sampling period increases. We explain this using (3.36) and (3.37): the affine power-emission relationship for hydrocarbon show in (3.36) has a large intercept and relatively small slope, making the emission rate less sensitive to the error in estimating $Z$, while the other two cases are more susceptible to the error in $Z$, which is caused by overlooking the higher-order variations in velocity and
The previous section is mainly concerned with traffic quantities attached to a Lagrangian particle (moving vehicle). It would be desirable to further explore the effect of under sampling in an Eulerian framework, that is, we will look at traffic quantities with fixed locations. We are also prompted to examine how the estimating error depends on the penetration rate of probe vehicles (mobile sensors). The Eulerian traffic quantities of interest in this section are: vehicle density, and aggregated emission and fuel consumption rates.

### 6 Estimating Eulerian quantities

The previous section is mainly concerned with traffic quantities attached to a Lagrangian particle (moving vehicle). It would be desirable to further explore the effect of under sampling in an Eulerian framework, that is, we will look at traffic quantities with fixed locations. We are also prompted to examine how the estimating error depends on the penetration rate of probe vehicles (mobile sensors). The Eulerian traffic quantities of interest in this section are: vehicle density, and aggregated emission and fuel consumption rates.

Table 4: Estimation of $\rho$ and $\dot{q}$ based on the phase transition model without source. $\rho_{\text{true}}$ and $\dot{q}_{\text{true}}$ denote the ground-truth density and perturbation; $\rho_N$ and $\dot{q}_N$ denote the reconstructed density and perturbation using under sampling.

|       | Mean (%) | Standard deviation (%) |
|-------|----------|------------------------|
|       | 0.96     | 1.27                   |
|       | 1.50     | 1.65                   |
|       | 1.10     | 1.49                   |
|       | 1.59     | 1.51                   |
| 4:00-4:15 | 6.17     | 9.29                   |
|       | 12.37    | 14.47                  |
|       | 3.09     | 4.58                   |
|       | 6.04     | 7.11                   |
| 5:15-5:30 | 0.65     | 1.43                   |
|       | 2.82     | 3.23                   |
|       | 0.63     | 1.61                   |
|       | 3.26     | 3.65                   |
| 5:00-5:15 | 3.93     | 6.82                   |
|       | 11.35    | 14.30                  |
|       | 1.77     | 3.52                   |
|       | 3.26     | 3.65                   |

Table 5: Estimation of emission and fuel consumption along vehicle trajectories. $Z$ denotes power demand; $r^{HC}$ denotes hydrocarbon emission rate; $r^{FC}$ denotes fuel consumption rate.

|       | Mean (%) | Standard deviation (%) |
|-------|----------|------------------------|
|       | 78.78    | 94.49                  |
|       | 99.08    | 99.49                  |
|       | 8.91     | 3.31                   |
|       | 2.20     | 2.08                   |
|       | 30.42    | 35.88                  |
|       | 37.10    | 36.69                  |
|       | 16.09    | 16.97                  |
|       | 17.15    | 17.10                  |
|       | 47.27    | 56.29                  |
|       | 58.67    | 58.41                  |
|       | 16.03    | 16.11                  |
|       | 16.21    | 16.30                  |

6 Estimating Eulerian quantities

The previous section is mainly concerned with traffic quantities attached to a Lagrangian particle (moving vehicle). It would be desirable to further explore the effect of under sampling in an Eulerian framework, that is, we will look at traffic quantities with fixed locations. We are also prompted to examine how the estimating error depends on the penetration rate of probe vehicles (mobile sensors). The Eulerian traffic quantities of interest in this section are: vehicle density, and aggregated emission and fuel consumption rates.
Figure 6: Ground-truth and reconstructed power demands (top), hydrocarbon emission rates (middle), and fuel consumption rates (bottom).

### 6.1 Vehicle densities

#### 6.1.1 Computational procedure

The procedure of estimating Eulerian density based on vehicle trajectory data and the phase transition model is illustrated as follows. We consider the temporal-spatial bins $C_{kl}$, $k = 1, \ldots, N_T$, $l = 1, \ldots, N_X$, each expressed as a product of intervals $[\tau_{k-1}, \tau_k] \times [x_{l-1}, x_l]$. Recall that given a discrete-time trajectory of a vehicle:

$$\ldots, x(t_{i-1}), x(t_i), x(t_{i+1}), \ldots$$

where $\{t_i\}$ is a fixed time grid, one can estimate the velocity $v(t_i)$ and acceleration $a(t_i)$ as well as the Lagrangian density $\hat{\rho}(t_i, x(t_i))$ using techniques elaborated in Section 2 and Section 3.

In order to estimate $\hat{\rho}$ inside $C_{kl}$, we search for the probe vehicles whose trajectories intersect $C_{kl}$. Then the quantity $\hat{\rho}_{kl}$ associated with $C_{kl}$ is estimated as the mean of all $\hat{\rho}(t_j, x(t_j))$ such that $(t_j, x(t_j)) \in [\tau_{k-1}, \tau_k] \times [x_{l-1}, x_l]$.

An example of such calculation is presented in Figure 7 where we utilize all vehicle trajectories to perform the estimation of bin-based densities, which is then compared with the ground-truth densities. The ground-truth densities are obtained simply by counting the number of vehicles present in a given bin. The proposed method captures several backward propagating shock waves displayed as red curves in the figure. Note that such result is based on a 100 % probe penetration rate.

In the rest of this section we shall examine the performance of the proposed computational procedure in the presence of much lower penetration rates. In the mean time we will also consider the effect of under sampling of locations associated with the same probe vehicle. The
The corresponding procedure of estimating densities is similar to that described above except with fewer and less time-resolved vehicle trajectories. Notice that one should choose the dimension of a bin reasonably large to ensure having at least one measurement inside each cell. In our numerical experiment, the dimension of a bin is chosen to be 4 (seconds) × 400 (feet).

![Image of 4:00 pm - 4:15 pm]

Figure 7: Reconstruction of cell densities (in vehicle/foot) during 4:00 pm - 4:15 pm. The upper panel shows the ground truth density obtained by counting the number of cars present in each cell during each time interval. The lower panel is the computational result with techniques described in Section 3 and 6.1.

6.1.2 Performance of density estimation with insufficient data coverage

In this section we will evaluate the performance of the aforementioned estimation method in the presence of insufficient data coverage, that is, when the penetration rate and the sampling frequency are low. The goal of this presentation of results is to identify certain range of penetration rates and sampling frequencies such that the domain of study is sufficiently covered and the estimation error remains reasonably low.

We first calculate the ground-truth densities with a sampling period of 0.1 (second) and a 100 % penetration rate. This is compared with the estimated ones with a combination of lower sampling frequencies and lower penetration rates. The relative errors among all active bins are averaged and shown in Table 6 for strongly stable traffic, and in Table 7 for less stable traffic.

From both tables, we notice that the penetration rate of probe vehicles has a substantial effect on the accuracy of estimation; while the sampling period only plays a minor role. This

---

2 Active bins refer to those having at least one measurement in it so that the density therein can be computed. The ratio between the number of active bins and the total number of bins is indicated as “Coverage Rate” in Table 6 and 7.
again confirms the intuition that under sampling does not affect significantly the estimation of first-order quantities. The modeling assumptions of strongly and less strongly stable traffic seems to yield qualitatively similar outcomes. We also observe that a penetration rate around 10% leads to both good coverage rate (≥97%) and satisfactory accuracy (error ≤20%), regardless of the sampling frequency.

A visualization of the density estimation is presented in Figure 8 when the penetration rate is 10% and the sampling period is 3 seconds. Figure 8 confirms that the proposed estimation scheme yields relatively accurate results under a low penetration rate and produces only a few inactive cells appearing as dark blue in the figure.

| Probe Vehicle Penetration Rate | Sampling Period | 100 % | 20 % | 10 % | 5 % | 2 % |
|-------------------------------|----------------|-------|------|------|-----|-----|
| 4:00-4:15                     | N = 10         | 0.65  | 9.49 | 13.20| 18.69| 23.63|
|                               | N = 20         | 1.29  | 9.23 | 12.86| 18.32| 23.62|
|                               | N = 30         | 2.21  | 9.70 | 13.35| 18.31| 21.98|
| Coverage Rate                 |                | 100.00%| 99.42%| 97.69%| 87.50%| 45.38%|
| 5:00-5:15                     | N = 10         | 0.72  | 11.47| 19.77| 30.14| 40.03|
|                               | N = 20         | 1.53  | 11.65| 19.19| 29.90| 38.91|
|                               | N = 30         | 2.75  | 11.90| 19.61| 29.13| 37.28|
| Coverage Rate                 |                | 100.00%| 100.00%| 98.79%| 88.91%| 49.40%|
| 5:15-5:30                     | N = 10         | 0.73  | 11.98| 17.86| 25.31| 39.33|
|                               | N = 20         | 1.58  | 11.72| 17.69| 25.33| 38.75|
|                               | N = 30         | 2.93  | 11.96| 17.47| 24.53| 36.60|
| Coverage Rate                 |                | 100.00%| 100.00%| 99.85%| 96.22%| 59.88%|

Table 6: Results of bin-based density estimation using different sampling periods and penetration rates, when the traffic is assumed to be strongly stable.

6.2 Emission and fuel consumption rates

In Section 5.3, it is shown that the emission and fuel consumption estimation along vehicle trajectories are largely deteriorated by under sampling. Such significant error is caused by the high variation in the acceleration profile. However, we note that if assessed from an average (integral) sense, the estimation of emission and fuel consumption rates may have improved performance even with larger sampling periods. To confirm such hypothesis, we re-examine the estimation error associated with each bin. It is our expectation that, by averaging several data points in the same bin, one could come up with a better estimation of these higher order quantities.

6.2.1 Computational procedure

The bin-based emission and fuel consumption rates can be estimated as follows. We first construct density \( \rho_{kl} \) for bin \( C_{kl} \) as illustrated before. We then use vehicle trajectories to compute \( r^{HC}(t_i, x(t_i)) \) and \( r^{FC}(t_i, x(t_i)) \) as shown in Section 5.3. Finally, estimations of \( r^{HC} \) or \( r^{FC} \) that fall within \( C_{kl} \) are collected to come up with the average, which is then multiplied by the bin density.

19
| Sampling Period | Probe Vehicle Penetration Rate | Coverage Rate |
|-----------------|-------------------------------|---------------|
|                 | 100 % | 20 % | 10 % | 5 % | 2 % | 100.00 % | 99.42 % | 97.69 % | 87.31 % | 45.19 % |
| 4:00-4:15       |       |      |      |      |      |           |          |          |          |        |
| Average         | N = 10 | 1.68 | 9.69 | 13.54 | 18.56 | 23.47 |
| Error (%)       | N = 20 | 2.89 | 10.09 | 13.67 | 18.76 | 24.02 |
|                 | N = 30 | 4.31 | 11.17 | 14.68 | 19.56 | 23.82 |
| Coverage Rate   |       |      |      |      |      |           |          |          |          |        |
| 5:00-5:15       |       |      |      |      |      |           |          |          |          |        |
| Average         | N = 10 | 2.71 | 11.03 | 18.85 | 27.68 | 36.46 |
| Error (%)       | N = 20 | 4.42 | 11.64 | 18.97 | 28.49 | 36.40 |
|                 | N = 30 | 6.04 | 12.51 | 19.96 | 29.06 | 36.84 |
| Coverage Rate   |       |      |      |      |      |           |          |          |          |        |
| 5:15-5:30       |       |      |      |      |      |           |          |          |          |        |
| Average         | N = 10 | 4.83 | 12.66 | 17.84 | 24.38 | 34.98 |
| Error (%)       | N = 20 | 6.88 | 13.47 | 19.01 | 25.60 | 36.24 |
|                 | N = 30 | 8.75 | 14.78 | 19.78 | 26.39 | 35.76 |
| Coverage Rate   |       |      |      |      |      |           |          |          |          |        |

Table 7: Results of bin-based density estimation using different sampling periods and penetration rates, when the traffic is assumed to be less strongly stable.

Figure 8: A comparison of the ground-truth densities (left column) and the reconstructed densities (right column) with 10% penetration rate and three-second sampling period. The three rows represent, from top to bottom, periods 4:00-4:15 pm, 5:00-5:15 pm and 5:15-5:30 pm, respectively.
6.2.2 Performance of emission/fuel consumption estimation with insufficient data coverage

Similar to the previous section, the emission and fuel consumption rates are estimated with lower sampling frequencies and penetration rates, which are then compared with the ground true. The average errors are summarized in Table 8, assuming the traffic is strongly stable. The case with less stability is qualitative similar, and is thus omitted from this paper. The results show improved estimation accuracy compared to that along individual vehicle trajectories; the reader is referred to Table 5 for a comparison. This confirms our earlier expectation that the estimation of emission or fuel consumption rates should be more immune to under sampling when averaged inside a bin.

| Sampling Period | Probe Vehicle Penetration Rate |
|-----------------|-------------------------------|
|                 | 100 % | 50 % | 20 % | 10 % |
| 5:00-5:30       |       |      |      |      |
| \(r_{HC}\)      |       |      |      |      |
| \(N = 5\)       | 8.31  | 10.97| 16.86| 24.02|
| \(N = 10\)      | 11.66 | 13.44| 17.94| 24.09|
| \(N = 20\)      | 14.49 | 15.95| 19.45| 24.35|
| \(N = 30\)      | 15.97 | 17.26| 20.74| 25.54|
| Coverage Rate   |       |      |      |      |
|                 | 100.00 % | 100.00 % | 99.42 % | 97.69 % |

| 5:00-5:30       |       |      |      |      |
| \(r_{FC}\)      |       |      |      |      |
| \(N = 5\)       | 15.66 | 17.29| 23.41| 31.47|
| \(N = 10\)      | 21.89 | 22.64| 26.38| 32.57|
| \(N = 20\)      | 26.97 | 27.54| 29.94| 34.04|
| \(N = 30\)      | 29.28 | 29.77| 31.98| 35.97|
| Coverage Rate   |       |      |      |      |
|                 | 100.00 % | 100.00 % | 99.42 % | 97.69 % |

Table 8: Estimation of \(r_{HC}\) and \(r_{FC}\) using different sampling periods and penetration rates, when the traffic is assumed to be strongly stable.

To further aggregate the emission and fuel consumption rates, we compute, for the whole road segment of interest, the total emission and fuel consumption rates. These time-dependent rates are shown in Figure 9, where the ground-truth quantity is compared to one with a three-second sampling period and 10% probe penetration rate. From these figures we see that the proposed method accurately captures the overall trends of emission and fuel consumption rates, although it tends to underestimate the true value since the three-second sampling period ignores higher-order variations in velocity and acceleration.

6.2.3 Correction factors

The previous numerical results reveal discernible errors for emission and fuel consumption estimations, despite that the overall time-varying trends of these quantities are captured using the proposed method. As commented before, these errors are in a way inevitable since the current sensing technology has insufficient sampling frequency. One obvious way out of this is to increase the sampling frequency of existing mobile sensors or to deploy more probe sensors in the traffic stream, which requires technological advancement and is beyond the scope of this paper. Instead, we propose in this section a correction factor approach for calibrating our estimation results.

We employ a linear regression approach that finds an appropriate affine relationship between the ground truth value and the estimated value. To show the validity of the proposed
Figure 9: Time-dependent emission and fuel consumption rates on the whole study area (1600 feet in length) for time period 5:00 - 5:30 pm. $r^{HC}$ denotes emission rate of hydrocarbon in gram/hour; $r^{FC}$ denotes fuel consumption rate in liter/hour. Red curve denotes the estimation based on a three-second sampling period and 10% penetration rate.

We denote by $r_{gt}$ the ground truth quantity, and by $r_{est}$ the estimated one. The linear regression model assumes that $r_{gt} = \beta_0 + \beta_1 r_{est}$ where $\beta_0, \beta_1 \in \mathbb{R}$. Data utilized for such linear regression is treated as training data. To be assured of the validity and robustness of the proposed affine adjustment, we employ a $k$-fold cross validation to be discussed below.

### 6.2.4 Cross validation

We partition available NGSIM vehicle trajectories into $k$ equal size subsets. Of the $k$ subsets, a single subset is used as the training data to obtain the linear coefficients $\beta_0$ and $\beta_1$. The rest of the $k - 1$ subsets are used to validate the affine adjustment. Such procedure is repeated $k$ times, each with distinct training dataset. Since we are focusing on a 10% penetration rate, it is natural to choose $k = 10$. 
The unsigned relative error $\epsilon_{\text{uns}}$ and the signed relative error $\epsilon_{\text{sgn}}$ are computed as

$$\epsilon_{\text{uns}} = \frac{|r_{\text{est}} - r_{\text{gt}}|}{|r_{\text{gt}}|}, \quad \epsilon_{\text{sgn}} = \frac{r_{\text{est}} - r_{\text{gt}}}{|r_{\text{gt}}|},$$

where $r_{\text{est}}$ denotes the estimated quantity, $r_{\text{gt}}$ denotes the ground truth. Results of the 10-fold cross validation is summarized in Table 9. It is quite clear from these results that applying the correction factor yields improved estimation accuracy even with lower penetration rate and sampling frequency. Figure 11 shows the corrected estimations of hydrocarbon and fuel consumption rates given by the test data, with correction factors obtained from the training data. Results of an informal normality test of the relative errors are shown in Figure 12. The errors roughly follow a normal distribution centered around zero.

|                      | Hydrocarbon | Fuel consumption |
|----------------------|-------------|------------------|
|                      | Mean (%)    | Std (%)          | Mean (%)    | Std (%)    |
| $\epsilon_{\text{uns}}$ with correction factor | 13.70       | 12.33            | 23.34       | 21.75      |
| $\epsilon_{\text{sgn}}$ with correction factor  | -3.62       | 18.07            | -10.06      | 30.28      |
| $\epsilon_{\text{uns}}$ without correction factor | 19.11       | 10.89            | 32.12       | 14.80      |

Table 9: 10-fold Cross validation results: mean and standard deviation of estimation errors with and without correction factors.

7 Conclusion

In GPS-enabled mobile sensing paradigm, two factors are of pivotal importance: the probe penetration rate and the sampling frequency. The former has been widely emphasized in the existing literature. On the other hand, it is our contention, established and proven in this paper, that sampling frequency of mobile devices also has a major impact on the quality of estimating higher-order traffic quantities such as acceleration, emission rate and fuel consumption rate. In particular, we have demonstrated the following:
Figure 11: Emission and fuel consumption rates after applying the correction factors. Red curves represent the corrected estimations, based on three-second sampling period and 10% penetration rate.

Figure 12: Histograms of the signed estimation error $e_{sgn}$.

- Various Eulerian (fixed-location) and Lagrangian (fixed-car) traffic quantities can be estimated via vehicle trajectory data, when the phase transition model is employed.
- For both Lagrangian and Eulerian estimation, first-order quantities are accurately reconstructed by the proposed method.
For second-order traffic quantities, under-sampling has a great impact on the estimation quality. The error is inevitable since the prevailing three-second sampling period overlooks higher-order variations.

The underestimation of hydrocarbon emission and fuel consumption rates can be corrected by imposing an affine correction factor that yields relatively accurate and robust estimation results.

Notes should be taken on the following important facts:

Some numerical results presented herein should be interpreted only in comparison with each other. They do not necessarily reflect the real-world situation. For example, the estimation of emission and fuel consumption depends on each vehicle’s characteristic such as type, make, etc., not just on its driving modes. It is for the purpose of extracting qualitative insights that simplified formulae such as (3.35)-(3.37) are employed.

The traffic reconstruction approach presented by this paper does not directly address data error. However, qualitative results remain valid.

The proposed correction factor for emission and fuel consumption is a potentially practical method that enhances GPS-enabled estimation. This method needs to be carefully calibrated in close view of road characteristics, type of vehicles, probe penetration rate and other relevant factors.

References

Federal Highway Administration. Next Generation Simulation, 2006. http://ngsim-community.org/

Aw, A., Rascle, M., 2000. Resurrection of “second order” models of traffic flow. SIAM Journal on Applied Mathematics 60, 916-938.

Barth, M., An, F., Norbeck, J., Ross, M., 2007. Modal emission modeling: A physical approach. Transportation Research Record 1520, 81-88.

Blandin, S., Argote, J., Bayen, A., Work, D., 2012. A phase transition model of non stationary traffic flow: Definition, properties and solution method. Transportation Research Part B 52, 31-55.

Chan, C.C., 1997. An overview of power electronics in electric vehicles. IEEE Transactions on Industrial Electronics 44 (1), 3-13.

Claudel, C.G., Bayen, A.M., 2010. Lax-Hopf based incorporation of internal boundary conditions into Hamilton-Jacobi equation. Part I: Theory. IEEE Transactions on Automatic Control 55 (5), 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 Transactions on Automatic Control 55 (5), 1158-1174.

Colombo, R.M., 2002a. Hyperbolic phase transitions in traffic flow. SIAM Journal on Applied Mathematics 63 (2), 708-721.
Colombo, R.M., 2002b. On a $2 \times 2$ hyperbolic traffic flow model. Math. Comput. Modeling 35, 683-688.

Colombo, R.M., Corli, A., 2002. Sonic hyperbolic phase transition and Chapman-Jouguet detonations, Journal of Differential Equations, 184, 321-347.

Colombo, R.M., Goatin, P., Priuli, F.S., 2007. Global well posedness of traffic flow models with phase transitions. Nonlinear Analysis 66, 2413-2426.

Daganzo, C.F., 1994. The cell transmission model. Part I: A simple dynamic representation of highway traffic. Transportation Research Part B 28 (4), 269-287.

Daganzo, C.F., 1995. The cell transmission model. Part II: Network traffic. Transportation Research Part B 29 (2), 79-93.

Demers, A., List, G.F., Wallace, W.A., Lee, E.E., Wojtowicz, J.M., 2006. Probes as path seekers: a new paradigm. Transportation Research Record: Journal of the Transportation Research Board 1944, 107-114.

Goatin, P., 2006. The Aw-Rascle vehicular traffic flow model with phase transitions. Mathematics of Computation 44, 287-303.

Kwon, J., McCullough, B., Petty, K., Varaiya, P., 2007. Evaluation of PeMS to Improve the Congestion Monitoring Program. California PATH Research Report UCB-ITS-PRR-2007-6, Institute of Transportation Studies, University of California, Berkeley, CA.

Lighthill, M., Whitham, G., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 229 (1178), 317-345.

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. Transportation Research Part C 18 (4), 568-583.

Koppa, R.J., 1997. Human factors, in Monograph on Traffic Flow Theory, U.S. Department of Transportation, Federal Highway Administration, Washington, DC, 1997, (3)1-32; available online at [http://www.tfhrc.gov/its/tft/tft.htm](http://www.tfhrc.gov/its/tft/tft.htm).

Kühne, R., Michalopoulos, P., 1997. Continuum flow models, in Monograph on Traffic Flow Theory, U.S. Department of Transportation, Federal Highway Administration, Washington, DC, 1997, pp. (5)1-51; available online at [http://www.tfhrc.gov/its/tft/tft.htm](http://www.tfhrc.gov/its/tft/tft.htm).

Payne, H.J., 1971. Models of freeway traffic and control. Simulation Council.

Payne, H.J., 1979. FREFLO: a macroscopic simulation model of freeway traffic. Transportation Research Record, 722 (1979), 68-75.

Post, K., Kent, J.H., Tomlin, J., Carruthers, N., 1984. Fuel consumption and emission modelling by power demand and a comparison with other models. Transportation Research Part A 18 (3), 191-213.

Richards, P.I., 1956. Shockwaves on the highway. Operations Research 4 (1), 42-51.
Siebel, F., Mauser, W., 2006a. On the fundamental diagram of traffic flow, SIAM Journal on Applied Mathematics 66 (4), 1150-1162.

Siebel, F., Mauser, W., 2006b. Synchronized flow and wide moving jams from balanced vehicular traffic, Physical Review E, DOI: 10.1103/PhysRevE.73.066108.

Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. Transportation Research Part B 39 (2), 141-167.

Whitham, G.B., 1974. Linear and Nonlinear Waves, Pure and Applied Mathematics, Wiley–Interscience, New York.

Work, D. B., Blandin, S., Tossavainen, O. P., Piccoli, B., Bayen, A.M. (2010). A traffic model for velocity data assimilation. Applied Mathematics Research eXpress, 2010 (1), 1-35.

Yim, Y., Cayford, R., 2001. Investigation of Vehicles as Probes using Global Positioning System and Cellular Phone Tracking: Field Operational Test. California PATH Working Paper UCB-ITS-PWP-2001-9, Institute of Transportation Studies, University of California, Berkeley, CA.

Yuan, Y., Van Lint, J.W.C., Hoogendoorn, S. P., Vrancken, J. L. M., Schreiter, T. (2011). Freeway traffic state estimation using extended Kalman filter for first-order traffic model in Lagrangian coordinates. In Networking, Sensing and Control (ICNSC), 2011 IEEE International Conference on (pp. 121-126). IEEE.

Zhang, H.M., 2002. A non-equilibrium traffic model devoid of gas-like behavior. Transportation Research Part B 36 (3), 275-290.