Title
Traffic flow reconstruction using mobile sensors and loop detector data

Permalink
https://escholarship.org/uc/item/6v40f0bs

Authors
Herrera, Juan C
Bayen, Alexandre M

Publication Date
2007-01-01
Traffic flow reconstruction using mobile sensors and loop detector data

Juan C. Herrera∗
Institute of Transportation Studies, Department of Civil and Environmental Engineering
University of California, Berkeley.

Alexandre M. Bayen†
Systems Engineering, Department of Civil and Environmental Engineering
University of California, Berkeley.

July 31, 2007

Submitted to the 87th TRB Annual Meeting (currently under review)

PLEASE DO NOT REDISTRIBUTE

Word count: 5715 words in text + 5 figures + 1 table = 7215 words

∗Corresponding author: 116 McLaughlin Hall, Berkeley, CA 94720; jcherrera@berkeley.edu; Tel:510-642-7390.
†Assistant Professor. 711 Davis Hall, Berkeley, CA 94720; bayen@berkeley.edu; Tel:510-642-2468; Fax:510-643-5264.
Abstract

In order to develop efficient control strategies to improve traffic conditions on freeways, it is necessary to know the state of the freeway at any point in time and space. Using data collected from stationary detectors—such as loop detector stations—the density field can be currently reconstructed to a certain accuracy. Unfortunately, deploying this type of infrastructure is expensive, and its reliability varies. This article proposes and investigates new algorithms that make use of data provided by mobile sensors, in addition to that collected by stationary detectors, to reconstruct traffic flow. Two approaches are proposed and evaluated with traffic data. The first approach is based on data assimilation methods (so-called nudging method) and the second is based on Kalman filtering. These approaches are evaluated using traffic data. Results show that the proposed algorithms appropriately incorporate the new data, improving significantly the accuracy of the estimates that consider loop detector data only.
1 INTRODUCTION

In order to develop efficient control strategies to improve traffic conditions on freeways, it is necessary to know the state of the freeway at any point in time and space. In particular, the knowledge of the density field or the velocity field is useful to determine bottleneck locations, physical extent of queues, and travel time. Using data collected from stationary detectors—such as loop detector stations—these fields can be currently reconstructed to a certain accuracy. Unfortunately, deploying this type of infrastructure is expensive, and its reliability varies.

New technology developments such as the Global Positioning System (GPS) allow cellular phones to broadcast their position and/or velocity in real time. If the cellular phone is located in a moving vehicle, the information collected from the phone can be used—in addition to (or instead of) the information provided by static sensors (e.g., loop detectors)—in traffic state estimation.

This new sensing technology has greatly impacted the transportation field, constituting a novel and rich source of traffic data. Maintenance problems common with stationary sensors (loop detectors mainly) do not appear with this technology. For public transportation agencies, there is no cost of deploying this new sensing technology, which is market driven. Personal devices (cellular phones) that people already have act as sensors, which has brought private companies into a field that has historically been monitored and managed by public agencies (encouraging partnerships between the public and private sector). Given the high penetration rate of cellular phones across the population—particularly in developing countries—this new sensing technology has the potential of being deployed almost everywhere.

The problem of interest consists of finding new algorithms to make use of the data obtained from mobile sensors, in particular vehicles acting as probes—in addition to the data collected from stationary detectors—to estimate traffic state on freeways. Our goal in this article is to incorporate this type of information into traffic flow models so better estimates of the traffic state can be obtained.

The rest of the article is organized as follows. Background material is provided in Section 2, where studies making use of data from loop detectors or probe vehicles are presented. Section 3 describes the proposed methods to address the problem of interest. These approaches were evaluated using traffic data, and the results are discussed in Section 4. Finally, some preliminary conclusions are presented in Section 5.

2 BACKGROUND

Depending on the way data is collected, there are two types of measurements used in traffic flow observation. The first one consists in observing traffic flow at fixed locations (Eulerian measurements).

This is the case of the data collected from loop detectors stations. The second way is by observing the trajectories of specific vehicles, which yields Lagrangian measurements (i.e., trajectory-based). Lagrangian data can be obtained with GPS technology inside cellular phones, or any tracking device/technology providing information (position, velocity) about individual vehicles.

Different methods are appropriate to reconstruct the density of vehicles on the freeway depending on the nature of collected data (Eulerian or Lagrangian). Traffic models using data collected from loop detectors (i.e., Eulerian data) are able to reconstruct the density field in a fairly reasonable way. These models, however, have no explicit way of incorporating Lagrangian data into the modeling. The next sections summarize how Eulerian and Lagrangian sensors have been used in other studies to estimate traffic
2.1 Eulerian sensing

 Numerous freeways around the world, in particular in the US, have loop detector stations embedded in the pavement to collect traffic data. For each lane, these detectors aggregate information at a sample time (usually 20 or 30 seconds) for each lane. Vehicle counts, occupancy are among the information that a detector collects. This type of data has been used in different ways for several traffic applications and studies. In particular, some studies have used this type of data to perform traffic state estimation. Gazis and Knapp in [1] and Szeto and Gazis in [2] are among the first studies that make use of Kalman filtering techniques to estimate the density on a section of freeway. The corresponding algorithms are based on the conservation of vehicle principle and they use vehicle counts and speeds at the boundaries of the section to perform the estimation.

 Lin and Ahanotu in [3] validate the Cell Transmission Model (CTM) using data collected from loop detectors. The CTM is a traffic flow model proposed by Daganzo in [4] and [5]. It is a discretization of the Partial Differential Equation (PDE) proposed in the 1950’s by Lighthill and Whitham [6] and Richards [7] (known as the LWR PDE) to describe the evolution of the traffic flow over time and space. The CTM divides the highway in “cells” of length $\Delta x$ and every $\Delta t$ units of time the number of vehicles on each cell is updated according to the conservation of vehicles principle. The CTM transforms the nonlinear flux function into a nonlinear discrete operator. Muñoz et al. in [8] propose the Switching-Mode Model (SMM), which is a piecewise linearized version of the CTM. Using the framework of hybrid systems [9], SMM combines discrete event dynamics (mode identification) with nonlinear continuous dynamics (density estimation). In [8] and [10], the authors test the proposed model using data collected from detectors on a stretch of highway I-210W in Pasadena, California.

 Treiber and Helbing in [11] propose and test a method that “filters” data from detectors (density or velocities) to reconstruct traffic flow. In contrast with previous methods, this method does not take into account information about the traffic dynamics (such as the LWR PDE). It only considers the main propagation direction of the information to perform the estimation (downstream under free flow conditions and upstream when congestion arises).

2.2 Lagrangian sensing

 Unlike Eulerian sensors, the use of Lagrangian sensors to reconstruct traffic flow has not been widely studied in transportation. Several studies have focused on technical and operational aspects of a probe system used to estimate travel times. Other studies have dealt with methodological aspects of travel time estimation and incident detection. Sanwal and Walrand in [12] propose an algorithm to estimate the average speed on a given link from probe speeds using a Bayesian approach (i.e. local speed estimation using Lagrangian speed measurement) in order to estimate travel time. The speeds estimated correlate well with the ones collected by loop detectors. Westerman et al. in [13] discuss the idea of fusing infrastructure-based loop detector data and non-infrastructure based probe vehicle data (i.e Eulerian and Lagrangian data) to obtain travel time estimations and to detect incident automatically. Ygnace et al. in [14] present results of analytical and simulation models to estimate travel time from individual speed measurements. Chu et al. in [15] use Kalman filtering techniques to combine point detection data and

conditions.
\( k_i^h \) is density on cell \( i \) at time \( h \), \( q_i^h \) is the flow into cell \( i \) between time \( h \) and \( h+1 \), and \( r \) is the inverse of the speed needed to travel one cell in exactly one time step (i.e. \( r = \frac{\Delta t}{\Delta x} \)). \( q_i^h \) depends nonlinearly on \( k_i^h \) and \( k_{i-1}^h \). The Godunov scheme [22] will be used to solve (7), and \( q_i^h = q_G (k_{i-1}^h, k_i^h) \) where \( q_G \) is the Godunov flux defined as follows:

\[
q_G(k_1, k_2) = \begin{cases} 
q(k_2) & \text{if } k_c < k_2 < k_1, \\
q(k_c) & \text{if } k_2 < k_c < k_1, \\
q(k_1) & \text{if } k_2 < k_1 < k_c, \\
\min(q(k_1), q(k_2)) & \text{if } k_1 \leq k_2.
\end{cases}
\]

\( k_c \) is the critical density, and it corresponds to the point at which the flux reaches its only maximum (the flux function is concave). The discretization of the nudging term needs to be added to Equation (7). The final expression for the model is as follows:

\[
k_i^{h+1} = k_i^h - r \left( q_G \left(k_i^h, k_{i+1}^h\right) - q_G \left(k_{i-1}^h, k_i^h\right) \right) - \Delta t \sum_{j=1}^{J} \Delta t \sum_{t_p^j \in S_i^h} \lambda \left(x_i - x_j (t_p^j), h \Delta t\right) \cdot \left[ \hat{\rho}^{-} \left(x_j (t_p^j), t_p^j\right) - \rho^{0} \left(x_j (t_p^j), t_p^j\right) \right] \quad i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1. \quad (8)
\]

The notation introduced in Equation (8) is explained below:

- \( J \): total number of equipped vehicles,
- \( x_i \): location of the beginning of cell \( i \), \( x_i = x_0 + (i - 1) \cdot \Delta x \), for \( i = 1, 2, \ldots, I \), where \( x_0 \) is the beginning of the section of interest,
- \( t_p^j \): time when the \( p \)-th observation from vehicle \( j \) occurs (in time units),
- \( S_i^h \): set of all the observations times of vehicle \( j \) that are less than or equal to the time elapsed until the \( h \)-th time step, i.e. \( S_i^h = \{ t_p^j : 0 < t_p^j \leq h \cdot \Delta t \} \), and
- \( \hat{\rho}^{-} \left(x_j (t_p^j), t_p^j\right) = \left(1 - \theta_p^j\right) \cdot k_{i_pj}^h + \theta_p^j \cdot k_{(i_pj+1)}^h \), where:
  - \( h_p^j \): time step when the \( p \)-th observation from vehicle \( j \) occurs (in time step units), i.e. \( h_p^j = \left\lceil \frac{t_p^j}{\Delta t} \right\rceil \) (because of roundoff error of \( h_p^j \), the algorithm might exhibit some lag of duration less than \( \Delta t \) in sensing),
  - \( i_pj \): cell where vehicle \( j \) is located at time \( t_p^j \), \( i_pj = \left\lfloor \frac{x_j(t_p^j) - x_0}{\Delta x} \right\rfloor \), where \( x_j(t_p^j) > x_0 \), and
  - \( \theta_p^j = \frac{x_j(t_p^j) - x_{i_pj}}{\Delta x} \), where \( x_{i_pj} \leq x_j(t_p^j) \leq x_{(i_pj+1)} \).

The last term in the RHS of Equation (8) is the discrete version of the nudging factor times the difference between estimated and measured density (terms inside the square brackets). The density estimate at location \( x_j(t_p^j) \) is based on linear interpolation between the density at the closest points –on the grid– to
x_j(t_p^j). The nudging factor is given by Equation (6), after substituting \( x - x_j(t) \) by \( x_i - x_j \left( t_p^j \right) \) and \( t \) by \( h \Delta t \).

As pointed out earlier on, weak boundary conditions are required for a proper characterization of the LWR PDE in (2). For implementation purposes, one ghost cell is inserted at each boundary of the section. The ghost cells contain the boundary conditions, and they allow the first and last cells of the computational domain to be updated depending on the existing traffic conditions (free flow, congested, or a mix of the two). This approach was successfully implemented and tested by Strub and Bayen in [19] with traffic data collected from loop detectors. The equations for the ghost cells (\( i = 0 \) and \( i = I + 1 \)) and for the initial conditions (\( h = 0 \)) are given by Equation (9) and (10), respectively:

\[
k^h_0 = \frac{1}{\Delta t} \int_{(h-1)\Delta t}^{h\Delta t} \rho_a(t)dt \quad \text{and} \quad k^h_{I+1} = \frac{1}{\Delta t} \int_{(h-1)\Delta t}^{h\Delta t} \rho_b(t)dt \quad h = 0, 1, \ldots, H - 1 \quad (9)
\]

\[
k^0_i = \frac{1}{\Delta x} \int_{x_{i-1}}^{x_i} \rho_0(x)dx \quad i = 1, 2, \ldots, I. \quad (10)
\]

### 3.3 Kalman filtering

Kalman filtering is a recursive method to estimate the state of a discrete process governed by a linear stochastic dynamical system. The estimate is such that its covariance is minimized. This section describes how Kalman filtering techniques can be used for the purposes of this work.

**State space representation.** The state space representation of the system consists in two equations: the dynamic (or state) equation and the measurement (or observation) equation. The dynamic equation describes how the state of the process (density in this case) evolves and is given by Equation (7), with \( q^h_i = q_G \left( k^h_{i-1}, k^h_i \right) \). Since \( q_G \) is nonlinear, the SMM is used. Depending on the mode detected, the dynamic equation is chosen. This constitutes one of the main differences with respect to the nudging method described before, where no mode identification is performed (the method does not assume anything about freeway conditions).

If only the situations in which the whole section is either under free flow or congested are considered, the dynamic equation reads as follows (using the same notation as before):

\[
k^h_{i+1} = \begin{cases} k^h_i & \text{if free flow} \\ (1 - \delta) k^h_i + \delta k^h_{i+1} & \text{if congested} \end{cases} \quad i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1. \quad (11)
\]

where \( \delta = w/v_f \), and \( w \) is the backward shockwave speed. Equation (11) assumes that every cell has the same length \( \Delta x = v_f \Delta t \) and the same fundamental diagram, and that there is no creation or loss of vehicles at middle points (i.e. no ramps). For a more general formulation for these two cases (and three other cases that consider a mix of free flow and congested modes), the reader is referred to [23]. In matrix form, Equation (11) can be written as:

\[
k_{h+1} = A_h \cdot k_h + B_h \cdot u_h + w_h \quad (12)
\]
Bold letters represent vectors. $k_h = [k^h_1 \ k^h_2 \ \ldots \ k^h_I]^T$ is the state vector at time step $h$, and matrices $A_h$ and $B_h$ are time-varying in the sense that they depend on the mode $m$ at time step $h$ (for instance, $m = 1$ if free flow, $m = 2$ is congested). The term $w_h$ represents the error of the state equation (caused by false or missed counts – because of lane changing manoeuvres for instance – at boundaries). The input $u_h$ is a scalar and it is related to the measurements obtained at the boundaries of the computational domain (used to compute the density on the first or last cell under free flow or congested mode, respectively). In fact, if the whole section is either under free flow or congested mode:

$$u_h = \begin{cases} k^h_0, & \text{if free flow (} m = 1 \text{) on the section during } h; \\ k^h_{I+1}, & \text{if congested (} m = 2 \text{) on the section during } h. \end{cases} \quad (13)$$

where $k^h_0$ and $k^h_{I+1}$ come from Equation (9). Initial conditions $k^0_i$ are the same as in Equation (10).

The measurement or observation equation compares the observed values of the state variables with the ones predicted by the model. This equation is as follows:

$$y_h = C_h \cdot k_h + v_h \quad (14)$$

$y_h$ is the observed vector, and $v_h$ is the measurement error (which constitutes another difference with respect to the nudging method, in which the observed density is assumed to be perfect or error free).

Kalman filtering also assumes that $w_h$ and $v_h$ are independent and normally distributed with zero mean.

The matrix $C_h$ is time dependent. The size and elements of $C_h$ depend on where the Lagrangian observations at $h$ are coming from, and it contains only zeros and ones (it is assumed that the local density can be observed from Lagrangian sensors). Lagrangian observations are coming from different cells, and also the number of observations between consecutive time steps might be different. Thus, observed vector $y_h$ in Equation (14) will have different dimensions at different time steps. The latter implies that matrix $C_h$ will also have different dimensions at every time step.

In summary, the dynamic or state equation and the measurement or observation equation of the system are given by Equation (12) and (14), respectively.

**Solution method: Kalman filtering.** Let us use the following notation:

- $\hat{k}_h^-$: a-priori state estimate of $k_h$,
- $\hat{k}_h$: a-posteriori state estimate of $k_h$,
- $P_h^-$: a-priori estimate error covariance, where $e_h^- = k_h - \hat{k}_h^-$ is the a-priori estimate error,
- $P_h$: a-posteriori estimate error covariance, where $e_h = k_h - \hat{k}_h$ is the a-posteriori estimate error.
Kalman filtering provides a set of recursive equations to estimate the vector state. The equations are as follows:

\[
\hat{k}_{h+1} = A_h \cdot \hat{k}_h + B_h \cdot u_h \tag{15}
\]

\[
P_{h+1} = A_h \cdot P_h \cdot A_h^T + Q \tag{16}
\]

\[
F_{h+1} = P_{h+1} \cdot C_{h+1}^T \left[ C_{h+1} \cdot P_{h+1} \cdot C_{h+1}^T + R \right]^{-1} \tag{17}
\]

\[
\hat{k}_{h+1} = \bar{k}_{h+1} + F_{h+1} \left( y_{h+1} - C_{h+1} \cdot \bar{k}_{h+1} \right) \tag{18}
\]

\[
P_{h+1} = (I - F_{h+1} \cdot C_{h+1}) \cdot P_{h+1} \tag{19}
\]

Initial conditions \( \hat{k}_0 \) and \( P_0 \) are assumed to be known. \( Q \) and \( R \) are the covariance matrix of the process and measurement error, respectively. \( F_h \) is known as the Kalman filter gain at time step \( h \). The matrix \( C_h \) have different dimensions at every time step, which implies variable dimension of covariance matrix \( R \).

**Implementation issues.** As before, the section is divided in \( I \) cells of length \( \Delta x \) and the simulation period is divided in \( H \) time steps of length \( \Delta t \) each, such that \( \Delta x \) and \( \Delta t \) satisfy the CFL stability condition. The state vector contains the density on each cell. At the beginning of each time step the traffic conditions on the network need to be identified (in order to determine the set of matrices \( A \) and \( B \) to use). Once the mode has been identified at the beginning of time step \( h+1 \), Equation (15) and (16) are used to obtain the a-priori density estimate and covariance, respectively (using the Eulerian data provided from detector stations). At this point, Lagrangian data becomes available to the model, i.e. the observed density at time \( h+1 \) will be known for some cells (the quantity and position of the Lagrangian sensors at \( h+1 \) will determine how many and for which cells the density is observed). With this information, observed vector \( y_{h+1} \) and matrix \( C_{h+1} \) can be constructed. Then, the Kalman filter gain is computed using Equation (17). Finally, the a-posteriori density estimate and its covariance are obtained using Equation (18) and (19), respectively. In case no observation is available at time \( h+1 \), the matrix \( C_{h+1} \) is set equal to zero, which implies that the Kalman gain is also zero, and therefore the a-posteriori density estimate is equal to the a-priori density estimate.

4 RESULTS

Approaches described in Section 3 are implemented with traffic data to evaluate their performance. The performance is evaluated in terms of the ability of the models to estimate three variables derived from the density field: number of vehicles between two points in space, density profile at middle location, and travel time to travel through the entire section.

Traffic data from the Next Generation Simulation (NGSIM) project [24] is used to evaluate the proposed approaches. The data consist of the trajectory for every vehicle that enter a stretch of US Highway 101S in Los Angeles, CA, during 45 minutes (from 7:50am to 8:35am). Figure 1 illustrates the location of interest. The site is approximately 0.4 miles in length, with 5 mainline lanes. An auxiliary lane exists between the on-ramp and the off-ramp. Transition from free flow to congested mode happens during the first 12-15 minutes (i.e. part of the section was in the free flow mode while the rest was in the congested mode).
For implementation purposes, only the two left-most lanes were considered. The whole section (2040 ft) has been divided into $I = 17$ cells of $\Delta x = 120$ ft each, and the total simulation time (2634 sec) has been divided into $H = 2195$ time steps of $\Delta t = 1.2$ seconds each. Parameters of the fundamental diagram were extracted from PeMS$^1$.

\[ q_{max} = 2040 \text{ vphpl}, \quad k_f = 205 \text{ vpmpl}, \quad k_c = 30 \text{ vpmpl}, \quad v_f = 68 \text{ mph}, \quad w = -11.6 \text{ mph}. \]

For simplicity, no mode identification was performed in the Kalman filtering implementation. The whole section was assumed to be congested all the time. Four different scenarios were tried depending on how Lagrangian data is created. Some proportion $p$ of the trajectories are chosen (randomly) as equipped vehicles. These vehicles are reporting their position and speed every certain time $T$, and the reported speed is the average speed over the last $\tau$ seconds. The values chosen for $p$, $T$, and $\tau$ determine the total number of Lagrangian measurements created for each of the four cases investigated (see Table 1).

Travel time for the section of interest is about 2 minutes under congested conditions. Thus, cases 1 and 3 assume that each equipped vehicle sends only one report while it is traveling the section. Continuous tracking for equipped vehicles is assumed in cases 2 and 4. Because of privacy issues, these two cases might not be a realistic situation in the future. They were investigated, however, to evaluate the potential of the methods proposed.

For comparison purposes, a model that uses Eulerian data only (EDO) was also implemented according to the numerical scheme proposed by Strub and Bayen in [19]. Figures 2, and 3 compare the true or observed evolution of the number of vehicles on the section with the estimates from the approaches.

Figure 2 compares the results using EDO with those from the proposed approaches under case 4 (i.e. assuming that 20% of the vehicles are equipped and their speed is sampled continuously). The inclusion of Lagrangian data in the estimation process improves the accuracy of the estimates, regardless of the method used (Kalman filtering at the top and nudging method at the bottom). Both approaches appropriately incorporate the Lagrangian data into the model.

When both approaches use the same Lagrangian data, Kalman filtering yields more accurate estimates. The same results are found when the approaches are compared under any of the other three cases. Since the observed density has some error –as stated in Section 3.1, the observed density is actually an estimate, based on the fundamental diagram–, Kalman filtering’s capability of handling an error in the observed density –see Equation (14)– might explain its slightly better performance (the nudging method assumes that the observed density is perfect, i.e. it has no error). However, the advantage of making no assumption about the freeway state (i.e. no mode identification needs to be performed) makes the results from the nudging method encouraging.

\footnote{Assuming a triangular shape: $q_{max} = 2040 \text{ vphpl}, \quad k_f = 205 \text{ vpmpl}, \quad k_c = 30 \text{ vpmpl}, \quad v_f = 68 \text{ mph}, \quad w = -11.6 \text{ mph}.}$

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{NGSIM_site.png}
  \caption{NGSIM site.}
\end{figure}
Figure 2: Estimation improvement when Lagrangian data is available.

Figure 3 compares the Kalman filtering approach for the four different cases (i.e. assuming different penetration rates and sampling rates). The graph at the top contains the estimation for cases 2 and 4. The graph at the bottom contains the same curves but for cases 1 and 3. For low penetration rates (cases 1 and 2), the estimates are accurate as long as the sampling rate is high. For low sampling rate (cases 1 and 3), the estimates are not as accurate even when the penetration rate is high. As expected, the best results are achieved under case 4, when the amount of Lagrangian data is the highest.

The fifth column in Table 1 shows the Root Mean Square Error\(^2\) (RMSE) of the estimation of the number of vehicles for both approaches and the four cases used. The table is sorted according to the seventh column.

In terms of the density profile and travel time estimation, the qualitative results are the same as before. Figure 4 shows the true or observed density profile at a middle point (extracted from vehicle trajectories and then smoothed) and the estimated profile for three different cases. The sixth column in Table 1 presents the RMSE’s in terms of the density profile at a specific point in the middle of the section. Again, high

\(^2\)The RMSE is defined as: \(\text{RMSE} = \sqrt{\frac{\sum_{h}(\hat{z}_h - z_h)^2}{H}}\) where \(\hat{z}_h\) and \(z_h\) are the estimator and its actual value at time step \(h\), respectively.
Acknowledgements

The authors would like to thank NGSIM for providing the vehicle trajectories used in this study and Saneesh Apte from CCIT for processing the data and computing sensor occupancies and speeds from the NGSIM vehicle trajectories. Also thanks to Issam Strub, Eric Lew and Shuo Yang for their involvement on different stages of this study. Finally, JD Margulici for providing CCIT support and fruitful discussion.

References


