

COMPUTING TRAVEL TIMES FROM FILTERED TRAFFIC STATES

PIERRE-EMMANUEL MAZARÉ

Telenav, Inc
950 De Guigne Dr
Sunnyvale, CA 94085, USA

OLLI-PEKKA TOSSAVAINEN

Microsoft
1065 La Avenida St
Mountain View, CA 94043, USA

DANIEL B. WORK

University of Illinois at Urbana-Champaign
1203 Newmark Civil Engineering Laboratory, 205 N. Mathews Ave
Urbana, IL 61801, USA

ABSTRACT. This article experimentally assesses the influence of sensor data rates on travel time estimates computed from filtered traffic speed estimates. Using velocity data obtained from GPS smartphones and inductive loop detector data collected during the *Mobile Century* experiment near Berkeley, CA, and an evolution equation for average velocity along the roadway, an estimate of the traffic state is obtained via ensemble Kalman filtering. A large-scale batch of computations is run to produce estimates of traffic velocity with varying degrees of input data, and instantaneous and a posteriori dynamic travel times are compared to travel times recorded using license plate re-identification. We illustrate that dynamic travel time estimates can be computed with less than 10% error regardless of the data source, and that existing inductive loop detector data can significantly improve the accuracy of travel time estimates when GPS data is sparse.

1. Introduction.

1.1. Objective. Probe data will likely become ubiquitous in the not too distant future, due to the rapid expansion of consumer generated probe data from cell-phones, personal navigation devices, and intelligent vehicles. As the use of probe data for traffic monitoring increases, so does the need to understand the benefits and trade-offs between GPS data and conventional data sources. Yet, a complete analysis of the trade-offs between probe data and fixed sensors is difficult, because the value of the data from any sensor (GPS equipped probe vehicles, inductive loop detectors, etc.), is dependent on the specifics of the sensing technology, the method used to process the data, and the specific traffic monitoring objective in question.

2010 *Mathematics Subject Classification.* Primary: 90B20, 35L65; Secondary: 62L12.

Key words and phrases. Traffic estimation, traffic flow, vehicular traffic, scalar conservation laws, experimental data.

As navigation devices in vehicles become more common, either on smartphones or through integrated in-vehicle systems, the need to estimate travel times between any two arbitrary points on the network in real-time will become increasingly important. One approach to solving this problem is to construct explicit travel time estimators for all origin destination pairs on the network. Obviously, this approach quickly becomes intractable as the size of the network grows to scales relevant to commercial traffic monitoring companies. A computationally efficient alternative is to estimate the traffic speed throughout the network. Then, with the speeds on each link known, travel time estimates can be computed as requested by the navigation devices.

The latter approach to travel time estimation motivates our work, which is to study the following question. *To what degree can GPS probe data act as a substitute for conventional traffic monitoring technologies such as inductive loop detectors, for the purpose of estimating travel times?* Specifically, we focus our attention on estimating travel times by integrating various data volumes from inductive loop detectors and GPS equipped probe vehicles into a velocity flow model equivalent to the *Cell Transmission Model* [11, 12], using an estimation technique known as *ensemble Kalman filtering* (EnKF). A preliminary version of this work was presented in conference form in [23].

This article emphasizes the experimental performance of a flow-model based estimator using data collected during a one day field experiment known as *Mobile Century* [17]. The data set collected during this experiment is unique because of the large number of GPS equipped probe vehicles representing 2-5% of the traffic flow, the dense coverage of working inductive loop detectors on the experiment site, and the availability of travel time data obtained from video license plate re-identification. Thus, although the results presented in this article are still limited in geographic scope and in time, to the best of the authors' knowledge, they are based on the most comprehensive publicly available GPS data set to date [30].

1.2. Related work. Several field experiments have been conducted to assess the applicability of cell phone-based measurements for traffic monitoring [1, 2, 3, 8, 22, 25, 28, 29], including data generated from cell phone towers, which produces less accurate vehicle position and speed measurements compared to GPS. Bar-Gera [3] compared several months of network data from cellphones to inductive loop detector data on a 14 km freeway segment in Israel, and found them to be in good agreement. Liu et al. [22] evaluated a different network-based cell phone system in Minnesota, and compared travel times to license plate re-identification, and found the system generated results with varying accuracies. A summary of the major network-based cell phone experiments to date can be found in Liu et al. [22].

Several studies have also been conducted to assess the trade-offs between inductive loop detector data and data collected from GPS equipped probe vehicles. In Kwon et al. [20], it is shown that annual estimates of total delay, average duration of congestion, and average spatial extent of congestion can be made with less than 10% error by using either inductive loop detectors placed with half-mile spacing, or by using probe vehicle runs at a rate of about three vehicles an hour. Approximately four to six days of data is needed for reliable estimates from either data source.

The work of Herrera et al. [16] compares a nudging algorithm and a mixture Kalman filtering algorithm to examine how the addition of probe vehicle measurements sampled at a fixed time interval can decrease errors in estimating traffic velocity. On a 0.4 mile stretch of roadway, sampling 5% the traffic at 150 second intervals with inductive loops at both ends of the domain leads to a 16% improvement

over the inductive loop detector data alone. The article also uses the *Mobile Century* experiment data to compare three scenarios of time-based sampling of probe vehicles, finding that probe data outperforms inductive loop detector data for estimating traffic velocity if a sufficient number of measurements can be obtained from probe vehicles. This work uses the same data set from *Mobile Century*, but we now consider nearly one thousand scenarios to compare probe data to inductive loop detector data.

The works [4, 7, 10] are also closely aligned with the present work. In Cristiani et al. [10], experimental mobile sensor data is used to calibrate a flux function and to determine an initial condition for a traffic flow model instantiated for a stretch of highway in Rome, Italy. The model is then simulated forward in time to make accurate travel time estimates. Blandin et al. [4] use forward simulation with boundary data obtained from sensors to compare estimates of the traffic state from a calibrated scalar traffic model with a calibrated 2×2 phase transition model [9]. Instead of open loop simulation with a calibrated model, the traffic prediction in [7] is updated by assimilating historical data in the forecast.

1.3. Methodology overview and organization of the article. In order to assess the trade-offs between velocity data collected from GPS smartphones and velocity data obtained from inductive loop detectors, it is necessary to define the process by which the data is transformed into an estimate of travel time. In this article, we rely on a velocity estimation algorithm developed as part of the *Mobile Millennium* project [30]. The algorithm combines velocity measurements from GPS smartphones or inductive loop detectors with a model of traffic evolution, using a technique known as *ensemble Kalman filtering* (EnKF) to produce an improved estimate of the velocity field, from which the travel time is computed. The resulting travel time computed from this process is then compared to the travel times recorded from the license plate re-identification video data.

With the data processing algorithm determined, we create a number of scenarios in which the volume of probe data and number of inductive loop detectors made available to the processing algorithm are adjusted. For example, this allows us to compare the accuracy of computing travel times when all of the probe data is made available, to travel times which are computed when only some of the probe data is available, to travel times when some probe data is available and some inductive loop detector data is available. In this way, we can quantify the trade-offs of various amounts of data from probes and inductive loop detector data in terms of increased or decreased accuracy of the computed travel times.

In order to describe and quantify what probe data is made available to the travel time processing algorithm, we detail two metrics of importance to probe data, namely the *penetration rate* and the *sampling rate*. The penetration rate is defined as the percentage of cars on the roadway reporting probe data compared to the overall traffic flow, including the vehicles which do not send data. In addition to increasing the number of measurements, as the penetration rate increases, the sample of vehicles which generate measurements are more likely to be representative of the total traffic flow. The sampling rate refers to the frequency at which data is collected from the probe vehicles, and can be used to increase or decrease the number of measurements made available for estimating travel times from the same vehicles. The probe data collection technique used in this work collects data from probe vehicles at fixed points in space using a technique known as *Virtual Trip Lines*

(VTLs) [18] invented by Nokia. By decreasing spacing between the VTLs, the probe vehicles will send more measurements, with smaller spacing between measurements.

In order to modify the amount of data obtained from inductive loop detectors, the number of inductive loop detectors which are made available to the processing algorithm is adjusted. Because this article focuses on a real highway, it is not possible to modify the location of the inductive loop detectors. Instead, given a fixed number of inductive loop detectors to include for a given scenario, we select the specific loop detectors such that they achieve as uniform of a spacing along the highway as is possible.

The work presented in this article summarizes the findings of a large scale computational study in which an ensemble Kalman filtering estimation algorithm is used to produce traffic estimates, and to characterize the dependency of the solution in the amount of sensing data used for the computation. The contribution of the article is thus the method used to assess the potential gains provided by probe and inductive loop data.

The remainder of this article is organized as follows. The key features of the processing algorithm used for velocity estimation are given in Section 2, and the methods for computing travel times from the velocity field are described. The data collected from the *Mobile Century* experiment is detailed in Section 3.1, and in Section 3.2, the techniques for generating scenarios with various amounts of input data from probe vehicles and inductive loops are presented. In Section 4, the results of nearly one thousand scenarios using various amounts of inductive loop detector data and probe data for travel time estimation are analyzed. Finally, the discussion in Section 5 summarizes the results.

2. Algorithm for estimating travel times. The processing algorithm used in this work is based on a velocity estimation algorithm developed in the *Mobile Millennium* system. The algorithm takes velocity data from inductive loop detectors and probe vehicles as input, combines the data with a physical model of traffic evolution, and produces an improved estimate of the velocity along the corresponding stretch of roadway. Using this improved estimate of velocity, an estimated travel time is computed using an instantaneous method and a dynamic method, to compare against the travel times recorded from video data. A brief overview of this process is described in this section.

2.1. Mobile Millennium highway traffic estimation algorithm. The velocity estimation algorithm developed in the *Mobile Millennium* system is based on a discretization of a traffic flow model known as the *Lighthill-Whitham-Richards* (LWR) partial differential equation [21, 24] which describes the evolution of traffic density on the highway. In its discrete form, this model is also known as the *Cell Transmission Model* [11, 12]. In order to simplify the velocity estimation problem, this model is transformed into an equivalent velocity evolution equation [27] operating on a 30 second time step. The complete mathematical details of the employed traffic velocity evolution equation, and the fusion of velocity measurement data with the evolution equation using *ensemble Kalman filtering* (EnKF) are presented in [27]. We give an overview of the model and filtering algorithm next.

2.1.1. Velocity traffic dynamics. The seminal LWR equation, as proposed in [21, 24], reads:

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial q(x, t)}{\partial x} = 0, \quad (x, t) \in (0, L) \times (0, T) \quad (1)$$

where $\rho(x, t)$ and $q(x, t)$ respectively denote the density and flow at location x and time t . Additionally, we denote $v(x, t)$ the velocity field on the highway, and assume the velocity can be specified as a function V of the density only. This allows the flux q to be defined as a function of the density:

$$q = \rho \times V(\rho) . \tag{2}$$

The widely used triangular flux function assumes a constant velocity in free-flow and a hyperbolic velocity in congestion:

$$v = V(\rho) = \begin{cases} v_{\max} & \text{if } \rho \leq \rho_c \\ -w_f \left(1 - \frac{\rho_{\max}}{\rho}\right) & \text{otherwise} \end{cases}$$

where v_{\max} , ρ_{\max} , ρ_c and w_f are respectively the maximum velocity, maximum density, critical density at which the flow transitions from free-flow to congested, and the backwards propagating wave speed, respectively. Because the triangular velocity function is not strictly monotonic in free-flow, it cannot be inverted.

In order to use the triangular model in a velocity setting, we approximate it by the Smulders velocity function [26], with a linear expression in free-flow and a hyperbolic expression in congestion:

$$v = V(\rho) = \begin{cases} v_{\max} \left(1 - \frac{\rho}{\rho_{\max}}\right) & \text{if } \rho \leq \rho_c \\ -w_f \left(1 - \frac{\rho_{\max}}{\rho}\right) & \text{otherwise.} \end{cases} \tag{3}$$

For continuity of the flux at the critical density ρ_c , the additional relation $\frac{\rho_c}{\rho_{\max}} = \frac{w_f}{v_{\max}}$ must be satisfied.

The Smulders velocity function (3) can be inverted to obtain the velocity as a function of density:

$$\rho = V^{-1}(v) = \begin{cases} \rho_{\max} \left(1 - \frac{v}{v_{\max}}\right) & \text{if } v \geq v_c \\ \rho_{\max} \left(\frac{1}{1 + \frac{v}{w_f}}\right) & \text{otherwise} \end{cases} \tag{4}$$

where v_c is the critical velocity: $v_c = V(\rho_c)$.

The LWR PDE is discretized using a Godunov numerical scheme to obtain a discrete density evolution equation. After discretization, we can apply a variable change on the densities using the inverse velocity function V^{-1} . Let K and i_{\max} be two positive integers, we discretize time (indexed by k) and space (indexed by i) in K time steps of length $\Delta T = \frac{T}{K}$ and $i_{\max} + 1$ space cells of length $\Delta x = \frac{L}{i_{\max} + 1}$. Then, according to the Godunov scheme, the velocity value $v_{k+1,i}$ of cell i at time step $k + 1$ can be computed as:

$$v_{k+1,i} = V \left(V^{-1}(v_{k,i}) - \frac{\Delta T}{\Delta x} (g(v_{k,i}, v_{k,i+1}) - g(v_{k,i-1}, v_{k,i})) \right) \tag{5}$$

where $g(v_1, v_2)$ represents the numerical flux between consecutive cells with respective velocities v_1 and v_2 . In the case of the Smulders model, we obtain:

$$g(v_1, v_2) = \begin{cases} v_2 \rho_{\max} \left(\frac{1}{1 + \frac{v_2}{w_f}}\right) & \text{if } v_c \geq v_2 \geq v_1 \\ v_c \rho_{\max} \left(1 - \frac{v_c}{v_{\max}}\right) & \text{if } v_2 \geq v_c \geq v_1 \\ v_1 \rho_{\max} \left(1 - \frac{v_1}{v_{\max}}\right) & \text{if } v_2 \geq v_1 \geq v_c \\ \min(V^{-1}(v_1) v_1, V^{-1}(v_2) v_2) & \text{if } v_1 \geq v_2 . \end{cases} \tag{6}$$

We note that the evolution of the velocity field at each discrete point on an edge is well defined by (5) and (6), except at the boundary points $v_{k,0}$ and $v_{k,i_{\max}}$. At these boundaries, the equations

$$v_{k+1,0} = V \left(V^{-1}(v_{k,0}) - \frac{\Delta T}{\Delta x} (g(v_{k,0}, v_{k,1}) - g(v_{k,-1}, v_{k,0})) \right) \quad (7)$$

$$v_{k+1,i_{\max}} = V \left(V^{-1}(v_{k,i_{\max}}) - \frac{\Delta T}{\Delta x} (g(v_{k,i_{\max}}, v_{k,i_{\max}+1}) - g(v_{k,i_{\max}-1}, v_{k,i_{\max}})) \right) \quad (8)$$

contain references to the ghost cells $v_{k,-1}$ and $v_{k,i_{\max}+1}$, which are points which do not lie in the physical domain. The values of $v_{k,-1}$ and $v_{k,i_{\max}+1}$ are given by the prescribed boundary conditions to be imposed on the left and right side of the domain respectively. Extension of the model to networks follows a standard approach [15], and is detailed in [27].

2.1.2. *Traffic state estimation.* Given the velocity field $v_k = [v_{k,0}, v_{k,1}, \dots, v_{k,i_{\max}}]^T$ on all space cells at time step k , the velocity field at time $k+1$ can be obtained by applying:

$$v_{k+1} = f(v_k) + w_{k+1} \quad (9)$$

where f represents the update algorithm as described in (5), (6), (7) and (8). The term $w_k \sim (0, \mathbf{Q})$ is the white, zero mean model noise with covariance \mathbf{Q} .

At every time step k , the measurements y_k are related to the traffic state through the observation equation:

$$y_k = \mathbf{H}_k v_k + \nu_k \quad (10)$$

where $\mathbf{H}_k \in \{0, 1\}^{z_k \times (i_{\max}+1)}$ encodes the location of the z_k data sources that sent measurements during that time interval, and ν_k represents the measurement errors, which are assumed to follow a zero mean distribution with a measurement error covariance matrix \mathbf{R}_k . Note that \mathbf{H}_k depends on the time due to the fact that the location where measurements are recorded changes over time [27].

The recursive state estimation problem for linear time invariant systems is commonly solved using the *Kalman filtering* [19] algorithm. New measurements can be integrated optimally at every time step, using only the estimate at the previous time step. An a priori estimate at time k given data through time $k-1$, denoted $v_{k|k-1}$, is computed by evolving the a posteriori estimate $v_{k-1|k-1}$ through a linear evolution equation. Then, measurements at time k are collected and used to construct a corrected estimate $v_{k|k}$ by minimizing the trace of the posterior error covariance matrix $\mathbf{P}_{k|k}$.

Due to the nonlinearity of (9), a Kalman filter cannot be used, and an *ensemble Kalman filter* (EnKF) approach is chosen instead, as described in [13] and [14]. The EnKF circumvents the need for linearizing the evolution equation to fit the Kalman filtering framework, and instead approximates the evolution of the error covariance matrix $\mathbf{P}_{k|k}$ through an ensemble (sample) approximation. The algorithm is detailed in Algorithm 1.

A few remarks on the performance of the velocity estimation algorithm described above are in order. First, it is noted that the algorithm was designed as part of the *Mobile Millennium* system, where it is not possible to track probe vehicles for privacy reasons, and thus no continuous GPS records from probes are assumed to

Algorithm 1 Ensemble Kalman filter

- 1: *Initialization*: Draw J ensemble realizations $v_{0|0}^j$ (with $j \in \{1, \dots, J\}$) from a process with a mean $\bar{v}_{0|0}$ and covariance $\mathbf{P}_{0|0}$.
- 2: *Model prediction*: Update each of the J ensemble members according to the CTM-v forward simulation algorithm f :

$$v_{k|k-1}^j = f\left(v_{k-1|k-1}^j\right) + w_k^j \tag{11}$$

Then update the ensemble mean $\bar{v}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$ according to:

$$\bar{v}_{k|k-1} = \frac{1}{J} \sum_{j=1}^J v_{k|k-1}^j \tag{12}$$

$$\mathbf{P}_{k|k-1} = \frac{1}{J-1} \sum_{j=1}^J \left(v_{k|k-1}^j - \bar{v}_{k|k-1}\right) \left(v_{k|k-1}^j - \bar{v}_{k|k-1}\right)^T \tag{13}$$

- 3: *Measurement update*: Obtain measurements, compute the Kalman gain \mathbf{G} , and update the estimate of the state:

$$\mathbf{G}_k = \mathbf{P}_{k|k-1} \left(\mathbf{H}_k\right)^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1} \left(\mathbf{H}_k\right)^T + \mathbf{R}_k\right)^{-1} \tag{14}$$

$$v_{k|k}^j = v_{k|k-1}^j + \mathbf{G}_k \left(y_k + \nu_k^j - \mathbf{H}_k v_{k|k-1}^j\right) \tag{15}$$

- 4: Return to 2.
-

be available for the estimation algorithm. In practice, it is expected that the performance of the estimation algorithm could be improved when tracking of individual probe vehicles is allowed.

Second, it should be noted that the flow model requires some historical flow information to calibrate the model. In this study, historical inductive loop detector data from PeMS was used to estimate constant mainline inflow and outflows, and thus all results presented use inductive loop data in this way.

Next, the methods for computing the instantaneous and dynamic travel times from an estimated velocity field are described.

2.2. Methods for computing travel times. The position of a vehicle $p(t)$ moving at the average speed of traffic evolves according to:

$$\begin{cases} \dot{p}(t) &= v(p(t), t) \\ p(t_0) &= p_0 \end{cases} \tag{16}$$

where $v(x, t) = V(\rho(x, t))$. Thus, the *dynamic travel time* $\tau_{\text{dyn}}(t_0)$ of a vehicle starting a trip at time t_0 at position p_0 is computed by solving (16) for the vehicle position, then solving:

$$p(t_0 + \tau_{\text{dyn}}(t_0)) = L - p_0$$

for τ_{dyn} , where L is the distance of the trip. Note that in general, the dynamic travel time may be infinite if the velocity drops to zero and remains zero indefinitely, although this is not an issue in practice.

A limitation of (16) is that it relies on knowledge of the velocity field up to τ_{dyn} into the future. This means that the velocity field requires a short term estimate into the future, which further introduces computational overhead. Moreover, when

the future forecast is an open loop model prediction without data to correct the estimate, the error in the velocity estimate increases. It is also worth noting that the dynamic travel time for individual vehicles is measurable, for example via license plate re-identification.

As an alternative, a common simplifying assumption is to assume the velocity field does not evolve during the travel time computation. This simplification gives rise to the *instantaneous travel time*, τ_{inst} . To compute the instantaneous travel time, the position of the vehicle starting at p_0 at time t_0 is assumed to evolve according to:

$$\begin{cases} \dot{p}(t) &= v(p(t), t_0) \\ p(t_0) &= p_0. \end{cases} \quad (17)$$

Then, the instantaneous travel time can be computed by solving (17) for the vehicle position, then solving:

$$p(t_0 + \tau_{\text{inst}}(t_0)) = L - p_0$$

for τ_{inst} .

The instantaneous travel time has several advantages, most notably that it does not require the overhead of computing the velocity field forward in time, which can be intensive for large network predictions with an accurate discretization. This comes at the cost of decreased accuracy when the traffic state is changing quickly. Moreover, the issue of an infinite travel time is a more serious issue for the instantaneous travel time, as it can be generated if at any point the velocity drops to zero. To keep the estimates bounded, a minimum velocity greater than zero can be used during the integration of the vehicle's position.

Since the velocity field is the a posteriori filtered estimate, the instantaneous and dynamic travel times are approximated by integrating through the discrete velocity field $v_{k|k,i}$. Specifically, the velocity field in (16) or (17) is approximated by:

$$v(p(t), t) = v_{k|k,i}$$

where $i = \lfloor p(t)/\Delta x \rfloor$ with $\lfloor \cdot \rfloor$ denoting the floor operator. When computing the dynamic travel time, the time index k is updated according to $k = \lfloor t/\Delta T \rfloor$, and fixed as $k = \lfloor t_0/\Delta T \rfloor$ for the instantaneous travel time.

Because the estimated velocity field is piecewise constant in each cell over each time step, the solution of (16) or (17) can be computed exactly. Assuming the current position of the vehicle is in cell i , the vehicle updates its position each ΔT by computing the time τ required to reach the boundary between cells i and $i + 1$ when traveling at the speed $v_{k|k,i}$. If $\tau < \Delta T$, the vehicle reaches the boundary within the time step. The vehicle travels at the speed $v_{k|k,i}$ until the boundary is reached, and travels at the speed $v_{k|k,i+1}$ afterward. If the boundary is not reached, the vehicle travels at $v_{k|k,i}$ for the full time step. The position update is summarized as

$$p(t + \Delta T) = \begin{cases} p(t) + v_{k|k,i}\Delta T & \text{if } \tau > \Delta T \\ p(t) + v_{k|k,i}\tau + v_{k|k,i+1}(\Delta T - \tau) & \text{if } \tau \leq \Delta T. \end{cases}$$

After ΔT , the position of the vehicle is updated, k is incremented, the new cell index i is computed, and the process is repeated.

Note that higher resolution numerical reconstructions are available for the dynamic travel time reconstructions [5] if discretization errors become significant.

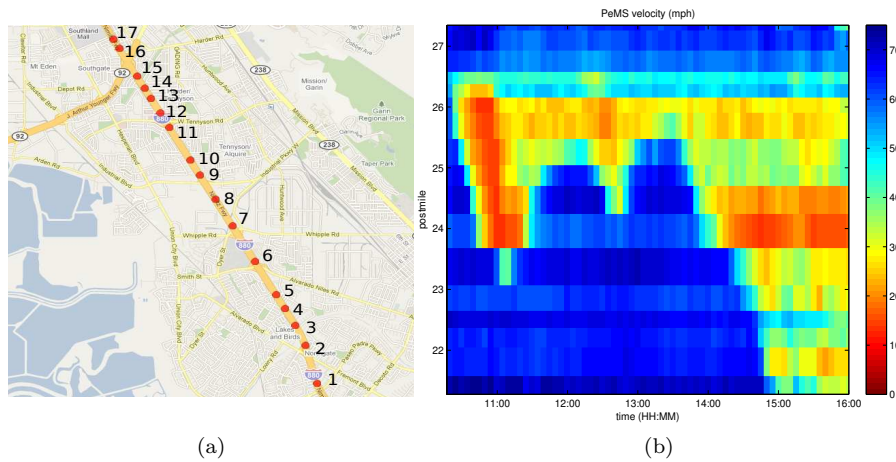


FIGURE 1. I-880N experiment inductive loops. (a) Location of the northbound inductive loop detector stations on the area where travel times are estimated. (b) Velocity contour plot from the PeMS system. Color denotes speed in mph. x -axis: time of day. y -axis: postmile (northbound from station one (a)).

3. Experimental data.

3.1. The Mobile Century experiment. In this section, the key features of the data collected during the *Mobile Century* experiment are presented. For a complete description of the experiment, the interested reader is referred to [17].

The *Mobile Century* field experiment was a one-day test in the San Francisco Bay Area which collected GPS data from cell phones in probe vehicles, inductive loop detector data, and travel time data from license plate re-identification video data. The experiment took place on February 8th, 2008, and involved 100 probe vehicles equipped with Nokia N95 cell phones which repeatedly drove a stretch of the I-880 freeway near Union City, CA, generating 2,200 vehicle trajectories.

The experiment site is also covered with 17 working inductive loop detector stations which feed measurements into the PeMS system [6]. The inductive loop detectors record the sensor occupancy and vehicle counts every 30 seconds, which is processed by a *Mobile Millennium* filtering algorithm in order to obtain the 30 second average velocity at the sensor. At 5 minute intervals, the PeMS system produces an estimate of the 5 minute average velocity at the sensor, which is shown in Figure 1b for the northbound traffic. The locations of the inductive loop detector stations are shown in Figure 1a.

Finally, as part of the experiment, high definition video cameras were temporarily installed to record license plates of northbound traffic. The travel times recorded from the re-identified vehicles traveling northbound is shown in Figure 2. During the morning, a five-car accident caused significant delay, and some drivers experienced travel times in excess of 20 minutes around 10:48 AM. Between 11:50 AM and 1:20 PM, vehicles experienced travel times between 8 and 10 minutes on the same stretch of roadway, which steadily increased from 1:20 PM to 3:20 PM. By 3:20 PM,

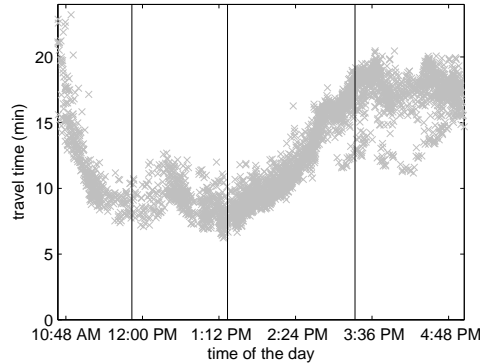


FIGURE 2. *Mobile Century* northbound travel times divided into four time bins from left to right: morning accident (10:00am-11:50am), free flow (11:50am-1:20pm), congestion building (1:20pm-3:20pm), and full congestion (3:20pm-). The travel times obtained from the license plate reidentification video recordings are marked with crosses.

most re-identified drivers experienced heavy evening congestion with travel times increasing to 15–20 minutes.

3.2. Algorithms for data selection. To assess the trade-offs between different amounts of probe data and inductive loop detector data for the purpose of estimating travel times, we algorithmically select different subsets of loop and GPS probe data from the *Mobile Century* experiment [17], and use these subsets as inputs to the estimation process described in the previous section. We now describe the scenarios which modify the type and amount of the data which is made available for estimation, and the selection criteria which are used to generate the scenarios.

3.2.1. Selection of inductive loop detector data. In order to modify the number of inductive loop detector stations which are made available for computing travel times, a simple selection criterion is developed. Specifically, given a fixed number of stations to include, the loops are selected in order to minimize the variance of the distance between consecutive sensors, resulting in approximately uniformly spaced sensors.

We consider a stretch of highway of length L , starting at $x = 0$ and ending at $x = L$, with n inductive loop detector stations located at x_1, x_2, \dots, x_n , as shown in Figure 3.

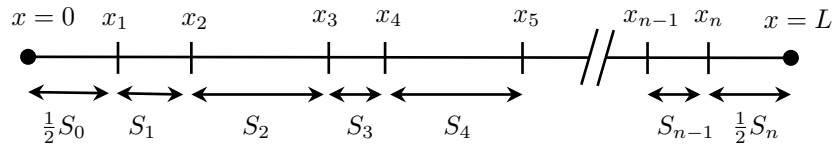


FIGURE 3. Highway segment of length L , with n inductive loop detector stations located at x_i .

Let S_i denote the spacing between sensor i and $i + 1$. In order to treat the boundaries without explicit knowledge of sensors outside the domain $x \in [0, L]$, it is assumed that only half of the first inter-station spacing S_0 and of the last inter-station spacing S_n lie in the domain of interest. The weighted average spacing between the n sensors is given by:

$$\bar{S}(n) = \frac{1}{n} \left(\frac{1}{2}S_0 + S_1 + S_2 \cdots + S_{n-1} + \frac{1}{2}S_n \right) = \frac{L}{n} \tag{18}$$

where the first and last spacings have a weight $\frac{1}{2}$, since only half of these spacings actually lie within the $[0, L]$ domain. Note that the average spacing is independent of the specific locations of the sensors x_i , and consequently cannot be used as a selection criterion.

Instead, we use a selection criterion which explicitly takes the uniformity of the inter-station distances S_i into account. This is achieved by minimizing the variance σ^2 of the inter-station spacings S_i , given by:

$$\sigma^2 = \frac{1}{2n} (S_0 - \bar{S})^2 + \frac{1}{n} \sum_{i=1}^{n-1} (S_i - \bar{S})^2 + \frac{1}{2n} (S_n - \bar{S})^2. \tag{19}$$

Again, the first and last spacings have a weight $\frac{1}{2}$, since only half of these spacings actually lie within the $[0, L]$ domain.

In practice, rather than minimizing the variance σ^2 , it is convenient to minimize an equivalent loop detector placement criterion denoted \tilde{S} :

$$\tilde{S}(x_1, x_2, \dots, x_n) = 2 \sqrt{\frac{x_1^2}{2n} + \frac{(L - x_n)^2}{2n} + \frac{1}{n} \sum_{i=1}^{n-1} \left(\frac{x_{i+1} - x_i}{2} \right)^2} \tag{20}$$

which shares the same minimizer as σ^2 . The best set of m inductive loop detector stations is then given by:

$$U^*(m) = \operatorname{argmin}\{\tilde{S}(U) \mid U \subset \{x_1, x_2, \dots, x_n\} \text{ and } |U| = m\} \tag{21}$$

where $|U|$ represents the number of elements in the set U . The resulting selections for the inductive loop detector stations are shown in Table 1.

In the case when the chosen inductive loop detector stations are uniformly spaced within the section of interest, the criterion \tilde{S} is equal to the average spacing \bar{S} . Because \bar{S} serves as a lower bound for \tilde{S} , the difference between \tilde{S} and \bar{S} indicates the degree of non uniformity of the sensor spacings caused by the fixed set from which the sensors are selected. Table 1 shows the difference between the inductive loop detector placement criterion $\tilde{S}(U^*(m))$ and its lower bound, the average inductive loop detector spacing $\bar{S}(m)$, is small, indicating that the sensor spacing is relatively uniform.

3.2.2. Penetration rate for probe data. A few remarks about probe penetration rates are important to make, before criteria to modify the penetration rate are discussed. In general, the penetration rate is difficult to determine for probe vehicles specifically because it depends on the number of equipped probe vehicles, the total traffic flow, and the evolution of the traffic flow in space and time. Typically, only the total number of equipped probe vehicles is known to probe data providers. Similarly, the total traffic flow can only be estimated from counts recorded by inductive loop detectors at predefined locations. Finally, because the evolution of the traffic flow

TABLE 1. Inductive loop detector selection results. Given a number m , the selection algorithm returns the set $U^*(m)$ of m inductive loop detector stations which minimizes the inductive loop detector placement index $\tilde{S}(U^*(m))$. The labels in $U^*(m)$ correspond to the labels of the inductive loop detectors in Figure 1a.

| m | $\tilde{S}(m)$ (mi) | $\tilde{S}(U^*(m))$ (mi) | $U^*(m)$ |
|-----|---------------------|--------------------------|--|
| 0 | ∞ | ∞ | \emptyset |
| 1 | 6.50 | 6.51 | { 8 } |
| 2 | 3.25 | 3.25 | { 4, 11 } |
| 3 | 2.17 | 2.17 | { 3, 8, 14 } |
| 4 | 1.62 | 1.63 | { 2, 6, 9, 15 } |
| 5 | 1.30 | 1.33 | { 2, 6, 8, 11, 16 } |
| 6 | 1.08 | 1.11 | { 1, 3, 6, 8, 11, 16 } |
| 7 | 0.93 | 0.95 | { 1, 3, 6, 8, 10, 13, 16 } |
| 8 | 0.81 | 0.83 | { 1, 3, 6, 7, 9, 11, 14, 16 } |
| 9 | 0.72 | 0.73 | { 1, 2, 4, 6, 7, 9, 11, 14, 16 } |
| 10 | 0.65 | 0.66 | { 1, 2, 4, 6, 7, 8, 10, 11, 14, 16 } |
| 11 | 0.59 | 0.60 | { 1, 2, 3, 5, 6, 7, 8, 10, 11, 14, 16 } |
| 12 | 0.54 | 0.55 | { 1, 2, 3, 5, 6, 7, 8, 10, 11, 13, 15, 16 } |
| 13 | 0.50 | 0.51 | { 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 15, 16 } |
| 14 | 0.46 | 0.48 | {1, \dots , 11, 13, 15, 16} |
| 15 | 0.43 | 0.46 | {1, \dots , 11, 13, 15, 16, 17} |
| 16 | 0.41 | 0.43 | {1, \dots , 13, 15, 16, 17} |
| 17 | 0.38 | 0.41 | {1, \dots , 17} |

is not under the control of the probe vehicles, it is nearly impossible to a priori specify a penetration rate which is both uniform in space, and uniform in time.

Because of the inherent difficulty in specifying the penetration rate a priori, we instead elect to directly modify the number of equipped probe vehicles as a proxy for modifying the penetration rate. The number of equipped probe vehicles in this study varies from 0% to 100% of the 2,200 *Mobile Century* vehicle trajectories, increasing by increments of 10%. Over the eight hour experiment, this corresponds to an average rate of probe vehicles between 27.5 veh/hr and 275 veh/hr. At a probe rate of 275 veh/hr, the 20 minute average penetration rate at the center of the experiment site ranges between 1.5% and 3% (Figure 4b). When a subset of the vehicle trajectories is required, the subset is determined by selecting the trajectories at random before the simulation.

3.2.3. Space-based sampling. In order to modify the number of measurements used from each probe vehicle trajectory under spatial sampling, the number of locations where measurements are collected are modified. The locations where measurements are obtained are encoded through the placement of virtual trip lines (VTLs), which can be viewed as virtual geographic markers which trigger vehicles to send measurements when the vehicle trajectory intersects the VTL. A complete description of the VTL sampling strategy is described in detail in Hoh et al. [18].

Because the VTLs are virtual, it is possible to place them anywhere on the experiment site. The determination of optimal VTL placement is complex, so instead

we elect to place the VTLs uniformly across the experiment site. The number of VTLs, denoted by n_{VTL} , tested in our scenarios varies from 9 VTLs to 99 VTLs, increasing by increments of 10 VTLs. This corresponds to an average spacing between 0.79 to 8.68 VTL/mi. Note the number of VTLs used on the experiment site is significantly higher than the number of inductive loop detector stations. This is possible because unlike inductive loop detector stations, the marginal cost of virtual trip lines is small. The number of probe vehicle measurements used in each simulation is shown in Figure 4a.

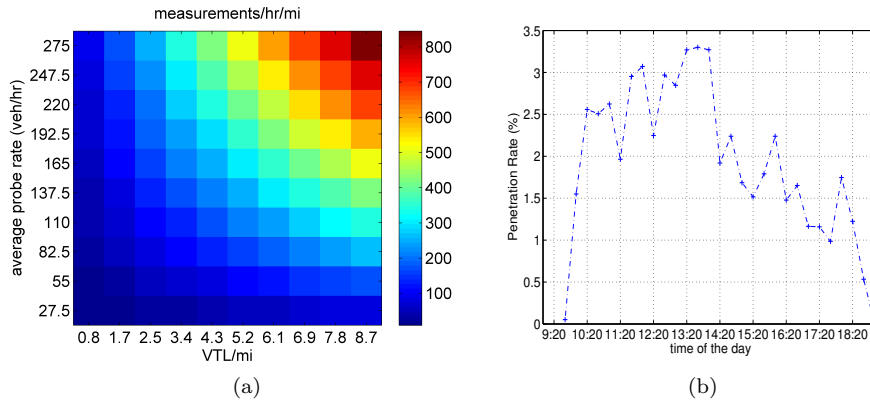


FIGURE 4. (a) Number of probe vehicle measurements used in the simulations when using VTL data; (b) 20-minute average penetration rate in the center of the *Mobile Century* experiment site on I-880 NB.

3.3. Summary of scenarios considered. Below, we give a summary of the various combinations of input data used for computing travel times in this work.

- *Number of inductive loop detectors.* Nine sets of inductive loop detector data, ranging from scenarios with zero to 16 stations, increasing by increments of two.
- *Number of probe data measurements.* The amount of probe data is modified in two ways.
 - *Penetration rate.* Eleven penetration rates are considered, ranging from scenarios with no probe data, to scenarios when 100% of the 2,200 probe vehicle trajectories are used, increasing by increments of 10%. This corresponds to an average rate of probe vehicles between 27.5 veh/hr and 275 veh/hr.
 - *Number of measurements per vehicle.* Ten sets of VTL locations are considered, ranging from scenarios with 9 (0.79 VTL/mi) to 99 VTLs (8.68 VTL/mi), increasing by increments of 10 VTLs.

In total, 917 scenarios are created by instantiating scenarios with all combinations of the 9 sets of inductive loop detector data sets, 11 probe penetration rates, and the 10 VTL sets. In the remainder of this section, we describe the specific algorithms which select the data for each scenario. The scenarios tested are summarized in Table 2.

TABLE 2. A subset of runs used in the study.

| Run | ILD stations | Probe rate (veh/hr) | VTL/mi |
|----------|--------------|---------------------|----------|
| 1 | 1 | 0 | 0 |
| 2 | 2 | 0 | 0 |
| 3 | 3 | 0 | 0 |
| \vdots | \vdots | \vdots | \vdots |
| 101 | 0 | 13.75 | 0.79 |
| 102 | 0 | 13.75 | 1.67 |
| 103 | 0 | 13.75 | 2.54 |
| 104 | 0 | 13.75 | 3.42 |
| \vdots | \vdots | \vdots | \vdots |
| 478 | 6 | 110 | 6.93 |
| 479 | 6 | 110 | 7.81 |
| 480 | 6 | 110 | 8.68 |
| 481 | 6 | 123.75 | 0.79 |
| 482 | 6 | 123.75 | 1.67 |
| \vdots | \vdots | \vdots | \vdots |
| 915 | 16 | 275 | 6.93 |
| 916 | 16 | 275 | 7.81 |
| 917 | 16 | 275 | 8.68 |

4. Results and discussion. In this section, we present the results of the 917 runs with varying amounts of probe and inductive loop detector data. We also vary the type of travel time computed (instantaneous or dynamic). First, the quantification of error is described, then the computational results are presented.

4.1. Error metric. Because validation data is available for dynamic travel times via license plate re-identification, an error metric is used to compare the velocity estimation algorithm output that has been converted to travel times with the travel time measured from video recordings. By using the travel time error as a performance metric, estimation algorithm results can be compared with the results obtained when using different types and quantities of the input data.

Since the license plate re-identification data provides a distribution of individual vehicle travel times (see Figure 2), we define the true travel time as a one minute moving average of the recorded travel times. Figure 2 also shows the division of the experiment into four time periods that represent the different phases of the traffic during the experiment. These periods are (i) the *morning accident*, where travel times are decreasing as an incident clears, (ii) a *free flow* period during the middle of the day when travel times are low, (iii) a *congestion building* period before the evening rush hours, and (iv) *full congestion* during the evening rush hours. Because of the different traffic conditions present in these time intervals, in addition to computing the error across the full day, the error is also computed for each time interval.

The travel time error is computed as follows. Let K be the number of estimates given in a period for which the error is to be computed, with each estimate indexed by k . Let $\tau_v(t_k)$ be the mean travel time from the video data for vehicles entering at

time t_k , and let $\tau_{\text{inst}}(t_k)$ and $\tau_{\text{dyn}}(t_k)$ be the estimated mean travel time computed with the instantaneous and dynamic methods at time t_k , respectively. The *mean absolute percent error* (MAPE) for the travel time computed with the instantaneous method is:

$$\varepsilon_{\text{inst,MAPE}} = \frac{1}{K} \sum_{k=1}^K \left| \frac{\tau_v(t_k) - \tau_{\text{inst}}(t_k)}{\tau_v(t_k)} \right| \quad (22)$$

while the MAPE for the travel time computed with the dynamic method is computed similarly.

4.2. Computational results.

4.2.1. *Implementation.* The estimates were computed using the existing *Mobile Millennium* [27] highway model summarized in Section 2, and implemented in the Java programming language. The model was run 917 times with various data inputs. Each run consisted in the computation of the mean speed field evolution and computation of both instantaneous and dynamic travel time every 30 seconds, from 10 am to 6 pm. The runs took 320 CPU-hours, and were distributed on 8 servers equipped with 2.2 GHz dual core AMD Opteron CPUs and 8 GB RAM, which reduced the computation time to 39 hours.

Despite taking nearly two weeks of CPU time, we note that our computational implementation is actually quite efficient, which enables computational experiments of this scale. Specifically, each simulation requires propagation of 100 ensembles through the discretized network flow model, and correcting the model predictions with the EnKF algorithm repeatedly over a period of eight hours, before the instantaneous and dynamic travel time estimates are computed. In total, each eight-hour numerical experiment takes 20 min to complete, which means it runs about 24 times faster than real-time.

4.2.2. *Using only inductive loop detector data in the model.* The first analysis of the traffic estimates is based on the results obtained when using inductive loop detector data as the only input to the model. These results give us a baseline for the comparison between probe and loop detector data. A total number of 17 runs were conducted based only on the inductive loop detector data, by varying the number of sensors according to the selection algorithm in Section 3.2. Both instantaneous and dynamic travel times were computed. The labels of the inductive loop detector stations used in the estimation are presented in Table 1 as a function of the number of stations selected (see also Figure 1a).

The results of these runs are shown in Figure 5. The subfigures show the estimation error broken down by time of the day, as defined in Figure 2. During the morning accident (Figure 5a), the dynamic travel times converge to estimates with 7% error, while the instantaneous estimates remain above 20% error. The instantaneous and dynamic estimates have between 6% and 7% error during the free flow and congestion building periods (Figure 5b and 5c), and 13% error during the full congestion period (Figure 5d), with the instantaneous and dynamic estimates performing similarly.

The number of inductive loop detector stations used tends to have a positive impact on the quality of the estimate when less than eight inductive loop detector stations are used. Note that the curve is not monotonically decreasing. This is because when only a few sensors are deployed, the error becomes highly dependent on the placement of the sensors. It is expected that an optimal sensor placement

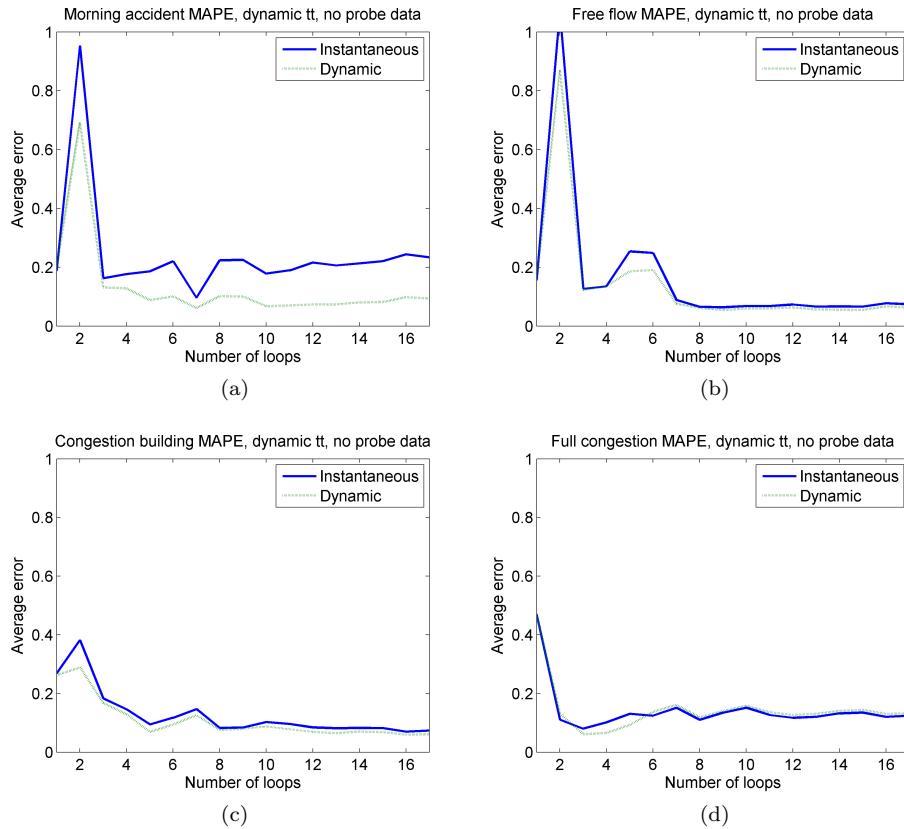


FIGURE 5. MAPE computed using inductive loop detector data only, no probe data. Travel time is computed using the dynamic method (green dash) and instantaneous method (solid blue). x -axis: number of inductive loop detector sensors, y -axis: MAPE (a) morning incident; (b) free flow; (c) afternoon as congestion increases; (d) evening congestion.

algorithm would reduce the error. The threshold of eight inductive loop detector stations corresponds to the inductive loop detector placement index $\tilde{S}(U^*(8)) = 0.83$ mi (Table 1). However, using data from more than eight inductive loop detector stations does not improve the quality of the estimates. If fewer than three inductive loop detectors are used, the estimation error is unacceptably high, at some points reaching as high as 100% error.

4.2.3. *Using only VTL data in the model.* The second part of the analysis consists of the travel time estimates obtained when using VTL data only. The changing parameters of the input data are the number of VTLs deployed on the experiment site and the rate of the probe vehicles used to produce speed measurements at the locations of VTLs. The estimation errors of the travel times obtained with the dynamic method are shown in Figure 6.

In each of the time periods, estimates of the travel time can be achieved with less than 6% MAPE, with sufficient probe vehicles and virtual trip lines. However,

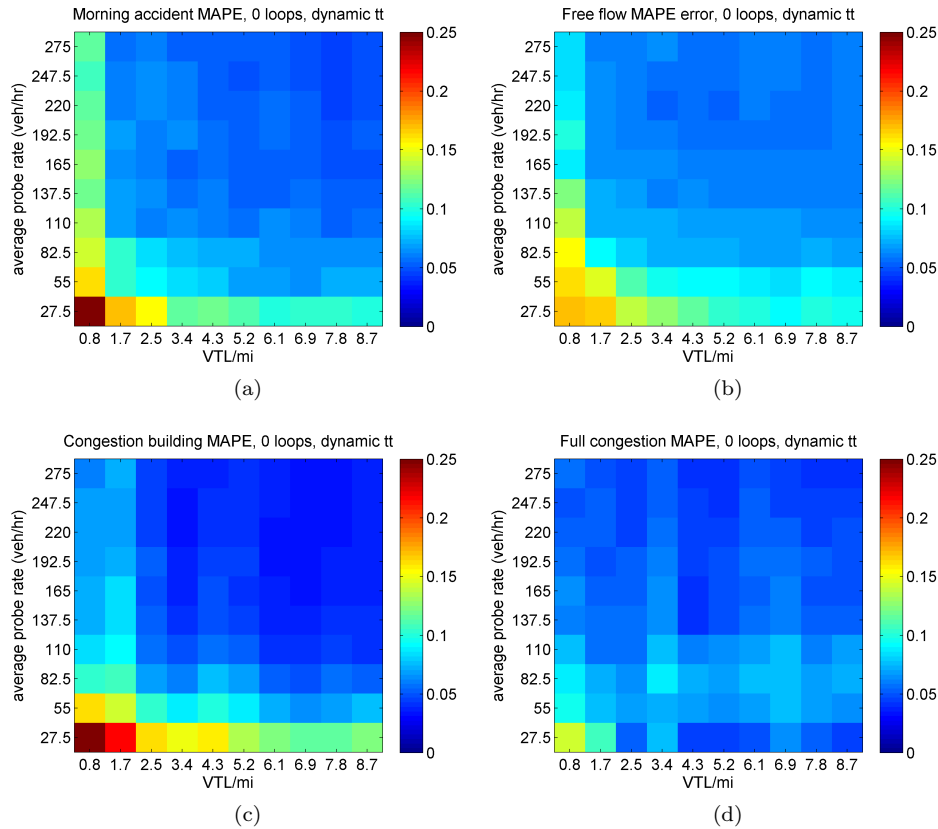


FIGURE 6. MAPE contours computed using VTL data only, no inductive loop detector sensors. Travel time is computed using the dynamic method. x -axis: number of VTLs, y -axis: average probe data rate (a) morning accident; (b) free flow; (c) congestion building; (d) full congestion. Color scale limited to 0.25.

when more than 137.5 veh/hr are used with more than 2.54 VTL/mi, only small improvements in the accuracy of the estimates can be achieved. When compared with inductive loop detectors, using 137.5 veh/hr and 2.54 VTL/mi performs as well as the estimates using more than eight inductive loop detector stations during the morning accident, free flow, and congestion building periods, but has less than half the error of inductive loops during the full congestion period.

4.2.4. *Fusing VTL and loop detector data.* The dynamic travel time estimation error using both VTL and loop detector data simultaneously is assessed in Figure 7, where the change in the dynamic travel time MAPE due to the addition of data from six inductive loop detectors is computed. The results shown are a representative subset of all the runs performed when mixing the two data types.

At low probe data rates during the morning accident, free flow, and congestion building periods, adding inductive loop detector data increases the accuracy of the dynamic travel time estimates. For example, during the morning accident (Figure 7a), with a probe rate of 27.5 veh/hr and a VTL spacing of 0.79 VTL/mi,

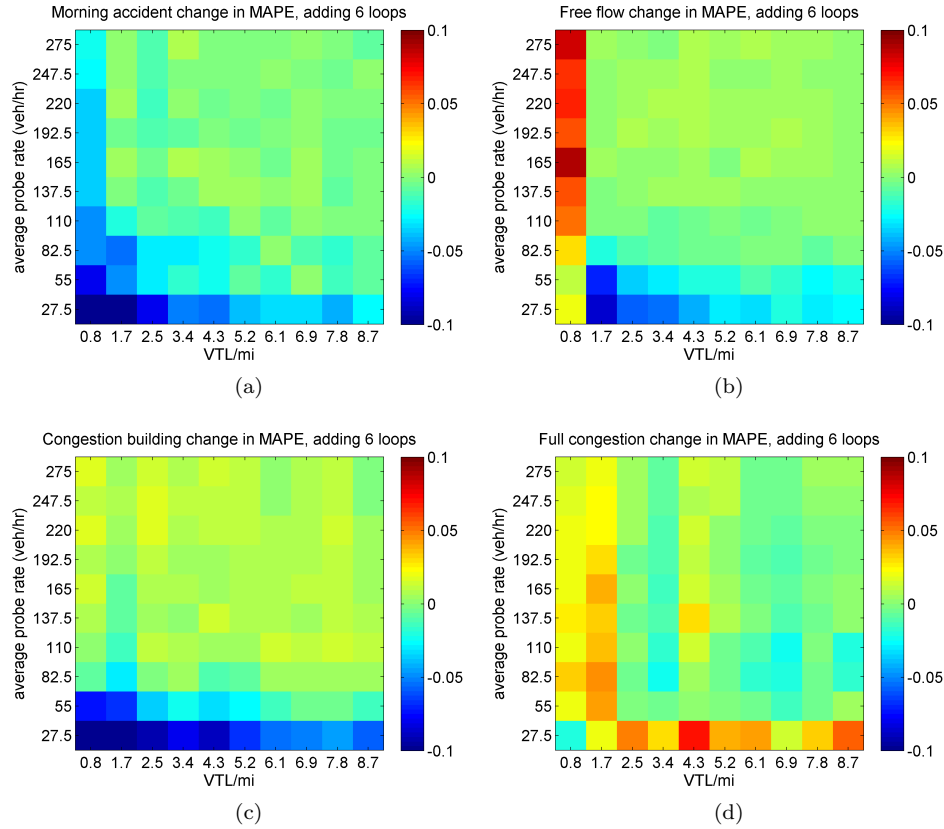


FIGURE 7. Change in MAPE contours when adding six inductive loop detectors to VTL data. x -axis: number of VTLs, y -axis: average probe data rate (a) morning accident; (b) free flow; (c) congestion building; (d) full congestion. Color scale limited to ± 0.1 .

adding inductive loop detector data reduced the error from 29% to 8%. During the full congestion period, the dynamic travel time estimate accuracy decreased when inductive loop detector data was added at low probe rates (27.5 veh/hr). This is likely due to the fact that the estimates based on virtual trip line data only were unusually accurate, even performing better than simulations with more probe vehicles.

At higher penetration rates (above 137.5 veh/hr) adding data from the six inductive loops has negligible effect, increasing or decreasing the accuracy only slightly. The exception is during the free flow period, when the MAPE increased (between 0.05 and 0.08) even at high probe rates, when 0.79 VTL/mi were used. The errors in the free flow period are magnified due to the small base travel time, which is under 10 minutes, and it is in fact not constant during the period (see Figure 2). Moreover, it is clear from Figure 1 that there is an area of heavy congestion around postmile 26 even during the free flow period, which is difficult to capture correctly with sparse sampling.

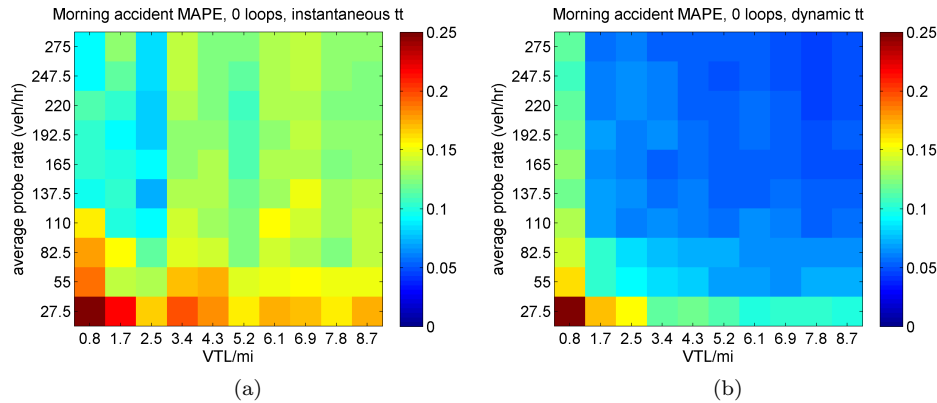


FIGURE 8. MAPE contours computed for the morning accident using VTL data and inductive loop detector sensors. x -axis: number of VTLs, y -axis: average probe data rate (a) 0 inductive loop detector sensors, instantaneous travel time; (b) 0 inductive loop detector sensors, dynamic travel time. Color scale limited to 0.25.

4.2.5. *Using instantaneous travel time as an estimate for dynamic travel time.* Figure 8 shows a comparison of the estimation errors when using instantaneous and dynamic travel times for the morning accident period. Instantaneous travel times can be determined at any time on any route using the speed estimates, and used as a proxy for dynamic travel times. As was shown for the inductive loop detector data in Figure 5b and 5d, instantaneous and dynamic travel time estimates are very similar when traffic conditions change sufficiently slowly. The same holds when estimating travel times from probe data.

By looking at the instantaneous travel time errors in Figure 8a, an interesting result can be seen. The results suggest that adding more probe data results in an increased travel time estimation error. However, this result is expected, and can be explained by focusing on the scenarios (in Figure 8a) in which the penetration rate of the probe vehicles is low and no loop detectors are used. Here, the instantaneous travel time estimate performs well, and may seem like a valid estimate of the true travel time during the incident. However, this gives a misleading indication of the quality of these travel time estimates. The good performance of the instantaneous estimate is caused by the fact that the current state of the traffic (speed field) is very poorly captured in the underlying scenario and the speed of the traffic is heavily overestimated. This causes the instantaneous travel time estimate to perform as a good predictor of the future traffic conditions, namely, as a predictor of the clearing incident. When the number of probe measurements increases, the speed field estimate is captured more accurately, and the increased error in the travel time estimate is caused by the instantaneous approximation.

5. **Summary.** In this study, trade-offs between velocity data collected from GPS smartphones in probe vehicles, and velocity data obtained from inductive loop detectors, for the purpose of computing travel times on a stretch of highway were studied.

This work used experimental probe data obtained from the *Mobile Century* field experiment and loop detector data from PeMS. The measurements were combined with a mathematical traffic model in a highway traffic estimation algorithm using a data assimilation technique called ensemble Kalman filtering, developed as a part of the *Mobile Millennium* project. The results of the algorithm were compared against the true travel times experienced by the drivers, obtained through license plate re-identification. A number of scenarios were created in which the volume of the probe data and number of inductive loop detector stations available for the estimation algorithm could be adjusted.

There are several practical results which were obtained from the extensive experimental studies including:

- *Achieving 10% error for dynamic travel times.* In this study, it was found that the dynamic travel time estimates can be achieved with less than 10% error when using a flow model with data assimilation, by using either inductive loop detector data, probe data, or a mixture of both inductive loop detector data and probe data. Moreover, the estimates from virtual trip line-based probe data can achieve a higher degree of accuracy when all available probe data is used compared to estimates from inductive loop detectors when all inductive loops on the experiment site are used, although in general the performance is similar.
- *Minimum loop detector spacing for travel time estimation.* In this study, using data from more than eight inductive loop detector stations (average spacing 0.83 miles) did not give additional benefit in the travel time estimation. The error remains constant between 6–13% depending on the time of day, regardless of the added loop detector stations.
- *Diminishing travel time accuracy improvement.* When sampling probe vehicles at a rate of 137.5 veh/hr with more than 2.54 VTL/mi, increasing the number of probe measurements by adding more probe vehicles or additional trip lines causes only a small improvement on the travel time accuracy.
- *A mixture of probe and loop detector data in travel time estimation.* It was found that when complementing loop detector data with probe vehicle data, better estimates for travel times are obtained, especially at low penetration rates. For example, if using loop detectors spaced more than 2.11 miles apart, probe data can give over 50% increase in the travel time accuracy.

The above findings may assist practitioners to identify the value of flow model based traffic monitoring systems for computing travel times.

Acknowledgments. The authors are grateful to Alex Bayen and the California Center for Innovative Transportation at UC Berkeley for supporting this work.

REFERENCES

- [1] A. Alessandri, R. Bolla and M. Repetto, Estimation of freeway traffic variables using information from mobile phones, in *Proc. of the American Control Conference*, 2003, 4089–4094.
- [2] V. Astarita and M. Florianz, [The use of mobile phones in traffic management and control](#), in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2001, 10–15.
- [3] H. Bar-Gera, [Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel](#), *Transportation Research Part C: Emerging Technologies*, **15** (2007), 380–391.
- [4] S. Blandin, G. Bretti, A. Cutolo and B. Piccoli, [Numerical simulations of traffic data via fluid dynamic approach](#), *Applied Mathematics and Computation*, **210** (2009), 441–454.

- [5] G. Bretti and B. Piccoli, [A tracking algorithm for car paths on road networks](#), *SIAM Journal on Applied Dynamical Systems*, **7** (2008), 510–531.
- [6] Caltrans, *Performance Measurement System (PeMS)*, <http://pems.dot.ca.gov/>.
- [7] H. Chen and H. A. Rakha, [Prediction of dynamic freeway travel times based on vehicle trajectory construction](#), in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2012, 576–581.
- [8] P. Cheng, Z. Qiu and B. Ran, Particle filter based traffic state estimation using cell phone network data, in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2006, 1047–1052.
- [9] R. Colombo, [Hyperbolic phase transitions in traffic flow](#), *SIAM Journal on Applied Mathematics*, **63** (2002), 708–721.
- [10] E. Cristiani, C. de Fabritiis and B. Piccoli, A fluid dynamic approach for traffic forecast from mobile sensor data, *Communications in Applied and Industrial Mathematics*, **1** (2010), 54–71.
- [11] C. F. Daganzo, [The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory](#), *Transportation Research Part B: Methodological*, **28** (1994), 269–287.
- [12] C. F. Daganzo, [The cell transmission model, part II: network traffic](#), *Transportation Research Part B: Methodological*, **29** (1995), 79–93.
- [13] G. Evensen, [Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics](#), *Journal of Geophysical Research*, **99** (1994), 10143–10162.
- [14] G. Evensen, [The ensemble Kalman filter: Theoretical formulation and practical implementation](#), *Ocean Dynamics*, **53** (2003), 343–367.
- [15] M. Garavello and B. Piccoli, *Traffic Flow on Networks*, American Institute of Mathematical Sciences, Springfield, MO, 2006.
- [16] J.-C. Herrera and A. Bayen, [Incorporation of Lagrangian measurements in freeway traffic state estimation](#), *Transportation Research Part B: Methodological*, **44** (2010), 460–481.
- [17] J.-C. Herrera, D. Work, R. Herring, J. Ban, Q. Jacobson and A. Bayen, [Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century experiment](#), *Transportation Research Part C: Emerging Technologies*, **18** (2010), 568–583.
- [18] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. Bayen, M. Annavaram and Q. Jacobson, [Virtual trip lines for distributed privacy-preserving traffic monitoring](#), in *6th International Conference on Mobile Systems, Applications, and Services*, 2008, 15–28.
- [19] R. E. Kalman, [A new approach to linear filtering and prediction problems](#), *Transactions of the ASME Journal of Basic Engineering*, **82** (1960), 35–45.
- [20] J. Kwon, K. Petty and P. Varaiya, [Probe vehicle runs or loop detectors?](#), *Transportation Research Record*, **2012** (2007), 57–63.
- [21] M. Lighthill and G. Whitham, [On kinematic waves. II. A theory of traffic flow on long crowded roads](#), *Proc. Roy. Soc. London. Ser. A*, **229** (1955), 317–345.
- [22] H. Liu, A. Danczyk, R. Brewer and R. Starr, [Evaluation of cell phone traffic data in minnesota](#), *Transportation Research Record*, **2086** (2008), 1–7.
- [23] P.-E. Mazaré, O.-P. Tossavainen, A. Bayen and D. Work, Trade-offs between inductive loops and GPS probe vehicles for travel time estimation: A Mobile Century case study, in *Proc. of the Transportation Research Board 91st Annual Meeting*, 2012.
- [24] P. I. Richards, [Shock waves on the highway](#), *Operations Research*, **4** (1956), 42–51.
- [25] B. Smith, H. Zhang, M. Fontaine and M. Green, *Cell Phone Probes as an ATMS Tool*, Technical Report UVACTS-15-5-79, University of Virginia Center for Transportation Studies, 2003.
- [26] S. Smulders, [Control of freeway traffic flow by variable speed signs](#), *Transportation Research Part B: Methodological*, **24** (1990), 111–132.
- [27] D. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli and A. Bayen, A traffic model for velocity data assimilation, *Applied Mathematics Research Express*, **2010** (2010), 1–35.
- [28] J. L. Ygnace, C. Drane, Y. B. Yim and R. de Lacvivier, [Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes](#), Technical Report UCB-ITS-PWP-2000-18, University of California Berkeley Institute of Transportation Studies, 2000.

- [29] Y. Yim and R. Cayford, *Investigation of Vehicles as Probes Using Global Positioning System and Cellular Phone Tracking: Field Operational Test*, Technical Report UCB-ITS-PWP-2001-9, University of California Berkeley Institute of Transportation Studies, 2001.
- [30] Mobile Millennium, <http://traffic.berkeley.edu/>.

Received July 2013; revised August 2013.

E-mail address: pe.mazare@gmail.com

E-mail address: olli-pekka.tossavainen@microsoft.com

E-mail address: dbwork@illinois.edu