An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices

Daniel B. Work *Student Member, IEEE*, Olli-Pekka Tossavainen, Sébastien Blandin, Alexandre M. Bayen *Member, IEEE*, Tochukwu Iwuchukwu, Kenneth Tracton

***Abstract*—Traffic state estimation is a challenging problem for the transportation community due to the limited deployment of sensing infrastructure. However, recent trends in the mobile phone industry suggest that GPS equipped devices will become standard in the next few years. Leveraging these GPS equipped devices as traffic sensors will fundamentally change the type and the quality of traffic data collected on large scales in the near future. New traffic models and data assimilation algorithms must be developed to efficiently transform this data into usable traffic information.**

In this work, we introduce a new *partial differential equation* (PDE) based on the *Lighthill-Whitham-Richards* PDE, which serves as a flow model for velocity. We formulate a Godunov discretization scheme to cast the PDE into a *Velocity Cell Transmission Model* (CTM-v), which is a nonlinear dynamical system with a time varying observation matrix. The *Ensemble Kalman Filtering* (EnKF) technique is applied to the CTM-v to estimate the velocity field on the highway using data obtained from GPS devices, and the method is illustrated in microsimulation on a fully calibrated model of I880 in California. Experimental validation is performed through the unprecedented 100-vehicle *Mobile Century* experiment, which used a novel privacy-preserving traffic monitoring system to collect GPS cell phone data specifically for this research.

I. Introduction

A. Highway Traffic Monitoring at the Age GPS-Enabled Mobile Devices

A common feature of the current highway traffic monitoring infrastructure is the heavy investment required to develop, deploy, and maintain it. In California, the two principle means of monitoring include *inductive loop detector*s (ILD) used in the PeMS system [1], and *in-vehicle transponders* (IVTs) such as FasTrak. Recently, cellular phone based highway traffic monitoring has shown promise for obtaining cheap, reliable real time traffic information, without the high infrastructure and maintenance costs incurred by either the State or the Federal Departments of Transportation. This emergence began with the *Wireless Communications and Public Safety Act of 1999*, which required US telephone carriers to develop an *Enhanced 911* service with the ability to locate the position of the cellular phones. Since the mid 1990's a significant amount of experimental research [2], [3], [4], [5] has attempted to address the practicality of these systems, with limited success for estimation of travel times due to the poor quality position measurement accuracy of the trilateration based methods. However, the convergence of communication and multi-media platforms (Nokia N95, iPhone, Android platform) has enabled a key new component in monitoring: mobility monitoring via GPS. Business plans of most major cellular phone manufacturers such as Nokia include the embedding of GPS receivers in most cell phones in the next few years, subsequently leading to a very high penetration rate of GPS equipped travelers on freeways in the near future.

B. A New Data Source for Inverse Models

The potential availability of high quality position and speed data at a high penetration rate provides motivation for the development of new inverse modeling techniques to reconstruct highway traffic flows, densities and speeds from these mobile measurements. Inverse modeling techniques use a flow model and available measurements to provide an optimal estimate of the state variables of the model based on the measurements. This reconstruction of the state of a system using data is also called data assimilation [6]. Unlike ILDs which produce occupancy counts and limited quality speed measurements for all vehicles at a single point on the highway, GPS based cell phone sensing provides accurate position and speed information from a population of equipped vehicles. Because the density is not measured and cannot easily be extrapolated from speed measurements, traditional highway traffic theory such as the *Lighthill-Whitham-Richards* (LWR) *partial differential equation* (PDE) [7], [8] and related density

D. Work is a Ph.D. student in Systems Engineering, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710 USA (e-mail: dbwork@berkeley.edu). Corresponding author.

O. P. Tossavainen is a Post Doctoral researcher in Systems Engineering, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710 USA. and with the Department of Physics, University of Kuopio, 70211 Kuopio, Finland (e-mail: optossav@berkeley.edu).

S. Blandin is a M.S. student, Département Ville Environnement Transport, Ecole Nationale des Ponts et Chaussées, Cité Descartes Champs-sur-Marne, 77455 Marne-la-Vallée cédex 2, France (e-mail: sebastien.blandin@polytechnique.org).

A. Bayen is an Assistant Professor, Systems Engineering, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710 USA (e-mail: bayen@berkeley.edu).

T. Iwuchukwu is a Member of Research Staff with Nokia Research Center Palo Alto, CA 94304-1003 USA (e-mail: toch.iwuchukwu@nokia.com).

K. Tracton is a Principal Scientist with Nokia Research Center Palo Alto, CA 94304-1003 USA (e-mail: ken.tracton@nokia.com).

based algorithms such as [9] cannot be used as such for data assimilation. The principle objective of this work is to develop new inverse modeling techniques specifically designed to use velocity measurements as inputs, incorporate them into a flow model for velocity, and to produce an optimal estimate of the velocity field on a highways.

C. Related Work

Kalman Filtering (KF) has been widely used for traffic state estimation in earlier studies in its various forms. In [10], [11], *Mixture Kalman Filtering* (MKF) was applied to the *Cell Transmission Model* (CTM) to estimate traffic densities for ramp metering. The nonlinear CTM was transformed into a switching state space model, which enabled the use of a set of linear equations to describe the state evolution for the distinct flow regimes on the highway (e.g. highway is in free-flow or congestion). In [12], a Kalman Filter was used to incorporate Lagrangian velocity trajectories into a density based CTM for highway traffic. A real time algorithm for traffic estimation based on the *Extended Kalman Filter* (EKF) using second order PDE as a flow model was used in [13]. Other treatments of traffic monitoring include adjoint based data assimilation in [14], [15], *Unscented Kalman Filtering* (UKF) in [16] and *Particle Filtering* (PF) in [16], [17], [18], [19].

A common factor for the CTM based methods [10], [11], [12] described above is that the evolution of traffic state (typically density, not velocity) relies on a set of linearized equations which are needed in order to use the KF or EKF techniques. On the other hand, the fully nonlinear PF technique is more accurate, but has a higher computational cost. The approach proposed in this work employs *Ensemble Kalman Filtering* (EnKF), which enables the use of fully nonlinear evolution equations such as the discretization of the new flow model proposed in this article, while exploiting its linear observation equation. Unlike UKF, which uses a deterministic sampling technique, EnKF uses Monte Carlo integrations to maintain the nonlinear features of error statistics. Furthermore, by employing a fully nonlinear velocity evolution model, no highway-mode-selection-algorithms or simplifications to the equations are needed in this work.

This work is organized as follows. In section II, we propose a new velocity flow model for highways based on the LWR PDE and discretize it using a Godunov scheme. We detail the EnKF algorithm which enables us to maintain the nonlinearities of the velocity CTM-v in section III. Finally, in section IV, this new data assimilation algorithm for highway velocity field reconstruction is implemented in microsimulation, and as a real-time application through the *Mobile Century* experiment which occurred on February 8, 2008.

II. Highway Traffic Flow Model

A. Speed Evolution Equation

To address the problem of reconstruction of the velocity field on the highway, we introduce a new first order hyperbolic PDE similar to the LWR PDE, which models the evolution of speed on the highway. This PDE can be shown to be equivalent to the kinematic wave theory for the Greenshields flux function. This PDE is referred to as LWR-v PDE ("v" for velocity). Proper weak boundary conditions are defined to formulate the well posedness of an initial boundary value problem.

The seminal LWR equation proposed in [7] and [8] to model traffic on highways reads:

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{1}$$

where $q(x,t)$ and $\rho(x,t)$ respectively denote the flow of vehicles and their density at location x and time t. Additionally we denote $v(x,t)$ the velocity field on the highway. Equation (1) is derived from hydrodynamics theory and expresses the conservation of mass for a fluid of density ρ and of flux q and is considered relevant to model traffic on highway (see [20], [21] for more background).

In order to express the flow q as a function of the density ρ traffic theory uses an empirical relation called the *fundamental diagram*:

$$q(x,t) = Q(\rho(x,t)) \tag{2}$$

where Q is the flux function which is assumed to be independent from time and space. One of the seminal flux functions used is the *Greenshields* flux function [22] which expresses a linear relation between ρ and v as:

$$v = v_{\max}\left(1 - \frac{\rho}{\rho_{\max}}\right) \tag{3}$$

where $v_{\max}$ and $\rho_{\max}$ denote respectively the maximal velocity and the maximal density allowed by the model.

When the flux function is a Greenshields flux function, it is possible to invert the speed-density function and express ρ as a function of v, namely $\rho = \rho_{\max}\left(1 - \frac{v}{v_{\max}}\right)$. Thus inserting this expression of ρ in (1), one can re-write the LWR PDE on ρ as a LWR-v PDE:

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}\left(R(v)\right) = 0 \tag{4}$$

where $R(v) = (v)^2 - v_{\max}\, v$, using the notation $(v)^2 := vv$ to avoid confusion with discretization indices introduced later.

The simple variable change $v = v - \frac{v_{\max}}{2}$ transforms equation (4) into:

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(v)^2 = 0 \tag{5}$$

on the domain $(x,t) \in \,]a,b[\times]0,T[$, which is a Burgers equation with a factor $\frac{1}{2}$ omitted (see [23]).

The initial condition and boundary conditions of (4) in weak form read:

$$\forall x \in \,]a,b[\; v\,(x,0) = v_0(x) \tag{6}$$

and

$$\begin{cases} v(a,t) = v_a(t) \text{ or} \\ R'\left(v(a,t)\right) \le 0 \text{ and } R'\left(v_a(t)\right) \le 0 \text{ or} \\ R'\left(v(a,t)\right) \le 0 \text{ and } R'\left(v_a(t)\right) \ge 0 \text{ and } R\left(v(a,t)\right) \ge R\left(v_a(t)\right) \end{cases} \quad (7)$$

and

$$\begin{cases} v\left(b,t\right) = v_b\left(t\right) \text{ or} \\ R'\left(v\left(b,t\right)\right) \ge 0 \text{ and } R'\left(v_b(t)\right) \ge 0 \text{ or} \\ R'\left(v\left(b,t\right)\right) \ge 0 \text{ and } R'\left(v_b(t)\right) \le 0 \text{ and } R\left(v\left(b,t\right)\right) \ge R\left(v_b(t)\right) \end{cases} \quad (8)$$

where $v_a(t)$ and $v_b(t)$ denote the boundary conditions that are applied but which are not always active, as described by the above equations. As demonstrated in [24], the PDE (4) with the initial condition (6) and the weak boundary conditions (7)-(8) admits a unique entropy solution in the space $BV(]a,b[\times]0,T[)$.

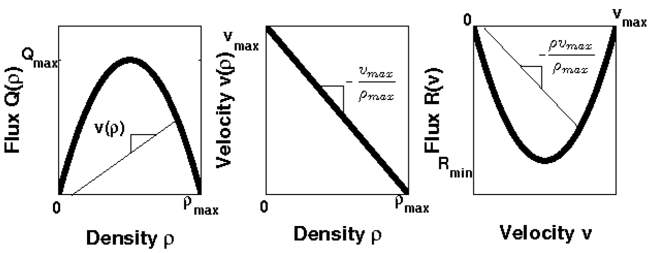

Figure 1. Greenshields model. **Left**: Classical fundamental diagram (parabolic). **Center**: Linear relation between speed and density. **Right:** Flux function for the LWR-v PDE (4). The flux is parabolic with negative values.

The fundamental properties of the LWR PDE (1) are conserved in the LWR-v PDE (4). First, the speed of a characteristic, given for a state (ρ_0, v_0) by the derivative of the flux function in (ρ_0, v_0) is the same for both PDEs. Indeed, for the LWR PDE it is:

$$Q'(\rho_0) = v_{\max} - 2\,\rho_0\,\frac{v_{\max}}{\rho_{\max}} \quad (9)$$

whereas for the LWR-v PDE it is:

$$R'(v_0) = 2\,v_0 - v_{\max} \quad (10)$$

and these two expressions are equivalent given relation (3). Second, the Rankine-Hugoniot relation giving the speed of shocks for conservation laws is also preserved. Indeed, the speed of a shock between a state (ρ_1, v_1) and a state (ρ_2, v_2) is given for the LWR PDE by:

$$\frac{Q(\rho_1) - Q(\rho_2)}{\rho_1 - \rho_2} = v_{\max} - \frac{v_{\max}}{\rho_{\max}}\,(\rho_1 + \rho_2) \quad (11)$$

whereas for the LWR-v PDE it is given by:

$$\frac{R(v_1) - R(v_2)}{v_1 - v_2} = (v_1 + v_2) - v_{\max} \quad (12)$$

Given the speed-density relation (3) expressed by the Greenshields model, these two expressions can also be checked to give the same speed for a shock.

B. Numerical Discretization

For practical implementation, the LWR-v PDE is discretized using a Godunov numerical scheme to obtain a *velocity cell transmission model* (CTM-v) [20], [21], [25]. In this section we detail the use of the Godunov scheme (see [25], [26]) for equation (5). This scheme is known to be convergent for convex and concave flux functions ([25], [26], [27]) such as the Greenshields flux function.

Let $N, M \in \mathbb{Z}$ the set of integers, we discretize time and space in N time steps J_n $(0 \le n \le N)$ of length $\Delta t = \frac{T}{N}$ and M space cells I_i $(0 \le i \le M)$ of length $\Delta x = \frac{b-a}{M}$. We call v_i^n the discrete value of v on $I_i \times J_n$. According to the Godunov scheme, at each time step v_i^{n+1} is computed from the previous time step by the following formula:

$$v_i^{n+1} = v_i^n - \frac{\Delta t}{\Delta x}\,\left(g\left(v_i^n, v_{i+1}^n\right) - g\left(v_{i-1}^n, v_i^n\right)\right) \quad (13)$$

where the numerical flow g is defined, as follows:

$$g\left(v_1, v_2\right) = \begin{cases} R\left(v_2\right) & \text{if } v_1 \le v_2 \le v_c \\ R\left(v_c\right) & \text{if } v_1 \le v_c \le v_2 \\ R\left(v_1\right) & \text{if } v_c \le v_1 \le v_2 \\ \max\left(R\left(v_1\right), R\left(v_2\right)\right) & \text{if } v_1 \ge v_2 \end{cases} \quad (14)$$

with v_c defined to be the minimum of the convex flux function from equation (4) (for example: $v_c = 0$ for $R(v) = v^2$, and $v_c = \frac{v_{\max}}{2}$ for $R(v) = (v)^2 - v_{\max} v$). For stability of the numerical discretization, the spatial and temporal step sizes must obey the *Courant-Friedrichs-Lewy* (CFL) condition:

$$\left|\frac{\Delta t}{\Delta x}\max\left(R'(v)\right)\right| \le 1 \quad (15)$$

In order to implement the weak boundary conditions defined in the previous section, we use ghost cells placed at each side of the domain defined by the strong boundary conditions we would like to be satisfied, namely:

$$v_{-1}^n = \frac{1}{\Delta t}\int_{J_n} v_a\left(t\right)dt \text{ and } v_{M+1}^n = \frac{1}{\Delta t}\int_{J_n} v_b\left(t\right)dt \quad (16)$$

with $J_n = \left[n\,\frac{T}{N}, (n+1)\,\frac{T}{N}\right]$. The choice of the Godunov scheme to solve a first order scalar hyperbolic conservation law such as (4) is standard in literature. Note that equations (13)-(14)-(16) could also be viewed as a counterpart of the cell transmission model for speed.

This model is thus a nonlinear dynamical system, in which the state of the system $v^n = [v_0^n, v_2^n, \cdots, v_M^n]$ is the vector of velocities in all cells at time step n. Letting $\mathcal{M}$ represent the nonlinear discrete time dynamical system model (13)-(14) for the full state vector v^n, and η^n the state noise, the state dynamics of the system can be written as:

$$v^{n+1} = \mathcal{M}[v^n] + \eta^n \quad (17)$$

Here the state noise η^n represents the modeling error introduced by discretization and uncertain boundary conditions.

For more information on modeling errors and state noise covariance estimation, see for example [28], [29].

The observation equation can be written as follows:

$$y^n = \mathbf{H}^n[v^n] + \varepsilon^n = \mathbf{H}^n v^n + \varepsilon^n \tag{18}$$

where $\mathbf{H}^n \in \{0,1\}^{p^n \times M}$ encodes the p^n discrete cells on the highway for which the velocity is observed during discrete time step n, and ε^n is the Gaussian observation noise with covariance $\mathbf{R}^n$ associated with the observation. In the event no equipped vehicles are in the spatial domain during a particular timestep, $\mathbf{H}^n$ reduces to the zero matrix.

III. Speed Estimation

A. Ensemble Kalman Filter

A common approach to solving the inverse modeling problem for *linear time invariant* (LTI) systems is to implement a KF algorithm [30], which is particularly well suited for real-time algorithms because of its recursive nature. Only the previous state of the system is needed to optimally integrate new measurements in the *Minimum Mean Square Error* (MMSE) sense. Due to the nonlinearity of the CTM-v, the standard KF can not be used. Furthermore, due to the nondifferentiability in the flux function of CTM-v the EKF has limited applicability. Hence, we extend this framework by implementing an *Ensemble Kalman Filter* (EnKF) algorithm [31], first introduced in [32], for the velocity data assimilation problem.

EnKF is a sequential data assimilation method, which uses Monte Carlo or ensemble integrations. By integrating an ensemble of model states forward in time, it is possible to compute the mean and error covariances needed at analysis times (measurement update) [33], [34]. The analysis scheme in the EnKF uses traditional update equations of the KF, except that the Kalman gain is computed using the error covariances provided by the ensemble of model states.

The nonlinearities introduced by CTM-v equation are captured well by EnKF because of the sample based computation of covariance matrices in contrast to tangent linear models (Jacobian matrix) used in EKF.

The EnKF algorithm can be broken into three phases [33]:

1) First, we generate and ensemble of model states by drawing K samples $\xi_k^0 \in \mathbb{R}^{M+1}$, $k = 1, \ldots, K$, from a Gaussian prior distribution to initialize the algorithm. These samples represent our prior knowledge of the initial velocity field v^0 on the highway. We assume that the initial velocity profile on the highway is smooth, without shocks. The construction of a smooth prior is made using a framework proposed in [35]. With this approach we can generate initial model states which maintain the same correlation properties regardless of the discretization level (number of cells) of the CTM-v model.
2) Next, a prediction $\hat{\xi}_k^n$ of ξ_k^n is made from the CTM-v model:

$$\hat{\xi}_k^n = \mathcal{M}[\xi_k^{n-1}] + \eta_k^{n-1} \tag{19}$$

We then compute the mean of the ensembles:

$$v^n = \frac{1}{K}\sum_{k=1}^{K} \hat{\xi}_k^n \tag{20}$$

from which the covariance $\mathbf{P}^n$ of the predicted state can be computed as:

$$\mathbf{P}^n = \frac{1}{K-1}\mathbf{E}^n \left(\mathbf{E}^n\right)^T \tag{21}$$

where matrix $\mathbf{E}^n$ is defined as:

$$\mathbf{E}^n = [\hat{\xi}_1^n - v^n, \cdots, \hat{\xi}_K^n - v^n] \tag{22}$$

3) Next we compute the Kalman gain $\mathbf{G}^n$:

$$\mathbf{G}^n = \mathbf{P}^n \left(\mathbf{H}^n\right)^T \left[\mathbf{H}^n \mathbf{P}^n \left(\mathbf{H}^n\right)^T + \mathbf{R}^n\right]^{-1} \tag{23}$$

Finally, the ensemble ξ_k^n is updated with new measurements y^n as follows:

$$\xi_k^n = \hat{\xi}_k^n + \mathbf{G}^n[y^n - \mathbf{H}^n\hat{\xi}_k^n + \varepsilon_k^n] \tag{24}$$

Note that since the state of the dynamical system (in this case, velocity) is directly observed, the observation equation (18) is linear. This feature enables the Kalman gain to be computed explicitly in the EnKF algorithm, which circumvents the need for more computationally expensive particle filters.

The presence of ε^n in (24) is important both from a physical interpretation, as well as for the stability of the convergence of the EnKF routine. When the state of the highway is observed from GPS measurements, ε^n accounts for the GPS position and speed error. During field testing with the Nokia N95 mobile device, a mean velocity error of 3 mph has been observed giving approximate lane level position accuracy. From an algorithmic viewpoint, the random error is shown in [36] to be necessary to maintain sufficient variance in the ensemble and to prevent filter divergence.

IV. Implementation and Validation

We have implemented the EnKF on the LWR-v PDE as a traffic estimation algorithm in a privacy-preserving architecture described in detail in earlier work [37]. While the EnKF algorithm can be used for assimilating velocity data collected from a variety of sources, we test it using a special sampling system which collects speed and position data of vehicles which have GPS enabled mobile devices (such as a cell phone) on-board. When the car passes prescribed points along the highway known as *Virtual Trip Lines* (VTLs), the mobile device records, encrypts, and transmits the data to a proxy server for further scrubbing of personal information. The virtual trip lines can be thought of as virtual ILDs, which record the speed of selected vehicles, instead of density. A second server unlocks the encrypted traffic information before it is sent to the EnKF CTM-v model server for data assimilation. Using this framework, we validate the EnKF CTM-v algorithm both in simulation, and through a field experiment.

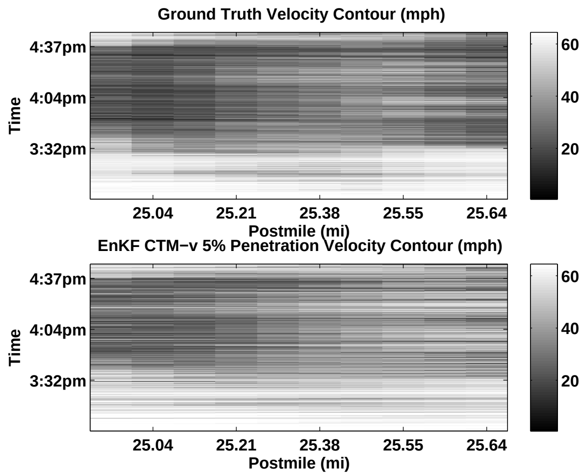

Figure 2. Paramics velocity contours. **Top:** Ground truth velocity contour average across all vehicles. **Bottom:** Estimated velocity contour from the EnKF CTM-v algorithm (19)-(24) at 5% penetration rate. X-axis: position along highway in postmile; Y-axis: time of day.

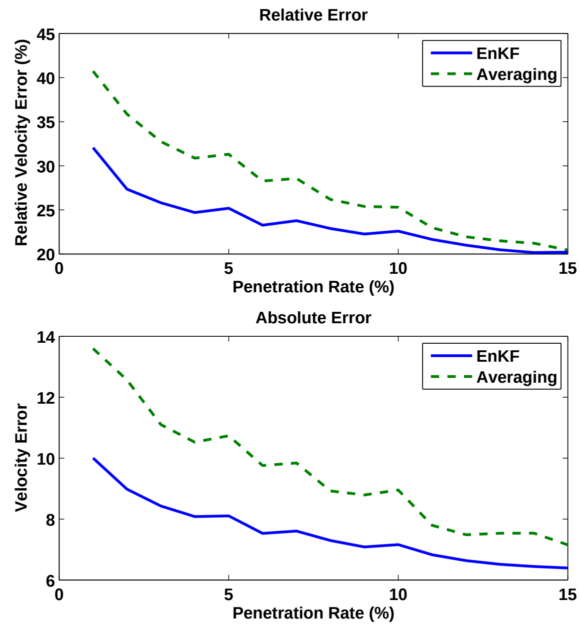

Figure 3. Error comparison of the EnKF CTM-v scheme (19)-(24) (solid) and the averaging scheme (25) (dashed) using Paramics. **Top:** Relative error computed from (26) as a function of penetration rate. **Bottom:** Absolute error computed from (27) as a function of penetration rate.

A. Paramics Microsimulation

We analyze the performance of the EnKF CTM-v algorithm using the Paramics Microscopic Traffic Simulation software [38], calibrated for highway I880 south of Oakland, CA. The I880 calibrated model produces individual trajectories on the highway for each vehicle in the simulation, and it has previously been used for bottleneck identification in [39]. For this experiment, a subset of the vehicles are randomly selected as vehicles which are equipped with GPS phones. The percentage of equipped vehicles relative to the total traffic flow is known as the penetration rate. Ten VTLs are placed evenly between Industrial Parkway (postmile 24.917) and Tennyson Rd. (postmile 25.767) which cause equipped northbound vehicles to report speeds and positions after crossing the VTL.

The highway is discretized into ten spatial cells, and we take a timestep of two seconds in order to maintain stability in the Godunov numerical scheme (15). A maximum speed of 70 mph is assumed for the Greenshields equivalent velocity flux function. The simulation is run for two hours, over which time congestion increases and the speed of flow decreases. Just after 3:30 pm in the simulation, the four lane averaged speeds decrease from the 65-70 mph free flow speed to speeds ranging between 20-40 mph. The congestion and corresponding slowdown is captured in the ground truth velocity contour shown at the top of Fig. 2.

In order to compare the performance of the EnKF CTM-v algorithm, a simple averaging-based estimation scheme is introduced. For this scheme, the velocity v_i^n in the discrete cell $I_i \times J_n$ is computed from the average of all measurements observed $\left(v_{\text{obs}}\right)_i^n$ in each discrete cell as:

$$v_i^n = \begin{cases} v_i^{n-1} & \text{if } v_{\text{obs}} = \emptyset \\ \left(v_{\text{obs}}\right)_i^n & \text{otherwise} \end{cases} \tag{25}$$

The mean point-wise $\mathcal{L}_1$ relative error v_{re} between the estimated velocity v_{est} and the ground truth velocity v_{gt} is computed by:

$$v_{\text{re}} = \frac{1}{MN} \sum_{i=0}^{M} \sum_{n=0}^{N} \frac{\left| \left(v_{\text{est}}\right)_i^n - \left(v_{\text{gt}}\right)_i^n \right|}{\left(v_{\text{gt}}\right)_i^n} \tag{26}$$

and the mean point-wise $\mathcal{L}_1$ absolute error v_{ae} of the discrete density field is computed as:

$$v_{\text{ae}} = \frac{1}{MN} \sum_{i=0}^{M} \sum_{n=0}^{N} \left| \left(v_{\text{est}}\right)_i^n - \left(v_{\text{gt}}\right)_i^n \right| \tag{27}$$

In Fig. 2 the error plot of the estimate of the EnKF plot relative to the ground truth shows that the main features of the shock wave are captured, even with the relatively low penetration rate of 5%. It is worth noting that the highway exhibits lane shearing, where vehicles in each lane have different mean speeds. The result is that vehicles sampled from the same discrete space and time cell, but from different lanes may have significant variance relative to the lane averaged mean speed. As the penetration rate increases (see Fig. 3), the sampled vehicles become more representative of the the flows on each lane, and thus more accurately predict the lane averaged mean speed.

The added value of the EnKF CTM-v algorithm is shown in Fig. 3, relative to a simple averaging estimate based on tracking. For this comparison, the complete trajectories of the equipped vehicles are observed and used in the simple averaging scheme (25), representing a privacy intrusive method in which the complete vehicle path is known. Alternatively, the EnKF CTM-v algorithm only integrates the velocities observed as a result of the equipped vehicle crossing the VTL. Even by assimilating less data, the EnKF CTM-v estimate has less error than the averaging scheme. At low penetration rates, the EnKF CTM-v algorithm reduces the relative error by 8% or three miles per hour, and as the penetration rate increases, the simple averaging estimate converges towards the EnKF CTM-v estimate. Although the relative error remains large at low penetration rates, the error occurs in the congested regime where the ground truth speed is slower and absolute errors are magnified.

B. February 8, 2008: Large Scale Field Test

Nicknamed the *Mobile Century* experiment, on February 8, 2008, the new privacy-preserving data collection system was built and used to estimate traffic conditions for a day on I-880 near San Francisco, CA. With the help of 165 UC Berkeley students, 100 vehicles carrying Nokia N95 phones drove repeated loops of six to ten miles in length continuously for eight hours. These vehicles represented approximately 5% of the total volume of traffic on the highway during the experiment.

This section of highway was selected specifically for its complex traffic properties, which include alternating periods of free-flowing, uncongested traffic, and slower moving traffic during periods of heavy congestion. The section is also covered with existing ILDs feeding into the PeMS system, which will be used to further assess the quality of the EnKF estimates in future work.

Qualitatively, the algorithm was validated around 10:30 am when a multiple car accident created significant unanticipated congestion on northbound traffic south of CA 92. The EnKF algorithm, running in real-time during the experiment, detected the accident's resulting bottleneck and corresponding shock wave. Fig. 4 shows the slowdown as estimated by the EnKF CTM-v model and broadcast during the experiment, compared to the live 511 traffic service [40] which includes data from ILDs.

Figure 4. I880 traffic report, 10:52 am, Feb. 8, 2008. A multi-car accident on the northbound lanes caused unexpected congestion. Foreground: 511 traffic information system; Background: EnKF CTM-v algorithm output. Numbers in circles correspond to speed in mph.

V. Conclusion and Future Work

This work presents a method for assimilating GPS speed and position data into a new velocity model derived from the LWR PDE. By working directly with the velocity PDE, conversions to density for data assimilation and back to velocity for travel time computations are eliminated. At low penetration rates, the method implemented using the VTL framework outperforms trajectory averaging based on tracking, despite using less data. Given that all data must be transmitted across a cellular network, optimal data assimilation methods will be increasingly important to efficiently use limited data streams. Furthermore, the recursive structure of the EnKF method is well suited for real-time applications such as demonstrated during the February 8, 2008 *Mobile Century* experiment.

In addition to improving estimates at low penetration rates, the EnKF CTM-v algorithm has additional features which we intend to highlight in future work. The framework can be run forward in time to produce forecasts of the traffic state, in addition the current state estimates presented here. This will be important to compute dynamic travel times that account for changes in the traffic state as the vehicle travels on the highway.

Finally, as historic data is collected, the model accuracy can be improved, by computing more accurate state noise and observation noise covariance matrices used to model the system dynamics. In a way similar to the accumulation of historical data for the PeMS system, the availability of training data will become large in the future as the penetration rates of GPS-equipped cellular phones on the highway increases.

Acknowledgment

The authors wish to thank Dr. David Sutter, Dr. Murali Annavaram, and Dr. Quinn Jacobson of Nokia Research Center Palo Alto, Baik Hoh and Prof. Marco Gruteser of Winlab at Rutgers University, and Ryan Herring, Juan Carlos Herrera and Dr. Jeff Ban of UC Berkeley for their invaluable contributions to develop, build, and deploy the traffic monitoring system implemented as part of the Mobile Century experiment. We thank the staff of the California Center for Innovative Transportation for the Mobile Century logistics planning and and successful implementation. The authors thank Dr. Jeff Ban for assistance with the Paramics simulations and for capturing in real time the images in Fig. 4 during the field test on Feb. 8, 2008. This research was supported by Caltrans, Nokia, the Center for Information Technology Research in the Interest of Society, the Finnish Funding Agency for Technology and Innovation (Tekes), and the National Science Foundation under contract CNS-0615299.

REFERENCES

[1] C. CHEN, P. VARAIYA, and J. KWON, "An empirical assessment of traffic operations," in *International Symposium on Transportation and Traffic Theory*, College Park, MD, July 2005, pp. 19–21.

[2] UNIVERSITY OF MARYLAND TRANSPORTATION STUDIES CENTER, *Final evaluation report for the CAPITAL-ITS operational test and demonstration program*. Transportation Studies Center, University of Maryland, 1997.

[3] J. YGNACE, C. DRANE, Y. YIM, and R. DE LACVIVIER, *Travel time estimation on the San Francisco Bay area network using cellular phones as probes*. California PATH Program, Institute of Transportation Studies, University of California, Berkeley, CA, 2000, no. UCB-ITS-PWP-2000-18.

[4] B. SMITH, H. ZHANG, M. FONTAINE, and M. GREEN, *Cell phone probes as an ATMS tool*. Center for Transportation Studies, University of Virginia, June 2003, no. UVACTS-15-5-79.

[5] Y. YIM and R. CAYFORD, *Investigation of vehicles as probes using global positioning system and cellular phone tracking: field operational test.* California PATH Program, Institute of Transportation Studies, University of California, Berkeley, CA, 2001, no. UCB-ITS-PWP-2001-9.

[6] G. EVENSEN, *Data Assimilation – The Ensemble Kalman Filter*. Springer, 2007.

[7] M. LIGHTHILL and G. WHITHAM, "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*, vol. 229, no. 1178, pp. 317–345, 1955.

[8] P. I. RICHARDS, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[9] H. Haj-Salem and J. Lebacque, "Reconstruction of false and missing data with first-order traffic flow model," *Transportation Research Record*, vol. 1802, pp. 155–165, 2002.

[10] X. SUN, L. MUNOZ, and R. HOROWITZ, "Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control," in *Proc. of the 42nd IEEE Conference on Decision and Control*, vol. 6, Maui, HI, 2003.

[11] ——, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proc. of the American Control Conference*, vol. 3, Boston, MA, 2004, pp. 2098 – 2103.

[12] J.-C. HERRERA and A. M. BAYEN, "Traffic flow reconstruction using mobile sensors and loop detector data," in 87^{th} *TRB Annual Meeting*. Washington D.C.: Transportation Research Board, Jan. 12-17 2008.

[13] Y. WANG and M. PAPAGEORGIOU, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B*, vol. 39, no. 2, pp. 141–167, 2005.

[14] D. JACQUET, C. CANUDAS DE WIT, and D. KOENIG, "Traffic control and monitoring with a macroscopic model in the presence of strong congestion waves," in *Proc. of the 44th IEEE Conference on Decision and Control, and European Control Conference*, Sevilla, Spain, 2005, pp. 2164–2169.

[15] D. JACQUET, M. KRSTIC, and C. CANUDAS DE WIT, "Optimal control of scalar one-dimensional conservation laws," in *Proc. of the 25th American Control Conference*, Minneapolis, MN, 2006, pp. 5213–5218.

[16] R. BOEL. L. MIHAYLOVA and A. HEGYI, "Freeway traffic estimation within recursive bayesian framework," *Automatica*, vol. 43, no. 2, pp. 290–300, 2007.

[17] L. MIHAYLOVA and R. BOEL, "A particle filter for freeway traffic estimation," in *Proc. of the 43rd IEEE Conference on Decision and Control*, vol. 2, 2004, pp. 2106–2111.

[18] J. SAU, N. EL FAOUZI, A. BEN ASSA, and O. DE MOUZON, "Particle filter-based real-time estimation and prediction of traffic conditions," *Applied Stochastic Models and Data Analysis*, vol. 12, 2007.

[19] R. BOEL. Z. LENDEK. A. HEGYI, L. MIHAYLOVA, "Parallelized particle filtering for freeway traffic state tracking," in *Proc. of the European Control Conference*, Kos, Greece, July 2007, pp. 2442–2449.

[20] C. F. DAGANZO, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B*, vol. 28, no. 4, pp. 269–287, 1994.

[21] C. DAGANZO, "The cell transmission model, part II: network traffic," *Transportation Research Part B*, vol. 29, no. 2, pp. 79–93, 1995.

[22] B. GREENSHIELDS, "A study of traffic capacity," *Highway Research Board*, vol. 14, pp. 448–477, 1935.

[23] L. C. EVANS, *Partial Differential Equations*. Providence, RI: American Mathematical Society, 1998.

[24] C. Bardos, A. Y. Leroux, and J. C. Nedelec, "First order quasilinear equations with boundary conditions," *Communications in partial differential equations*, vol. 4, no. 9, pp. 1017–1034, 1979.

[25] S. GODUNOV, "A difference method for the numerical calculation of discontinuous solutions of hydrodynamic equations," *Mathematics Sbornik*, vol. 47, no. 3, pp. 271–306, 1959.

[26] R. LEVEQUE, *Numerical Methods for Conservation Laws*. Basel, Switzerland: Birkhäuser Verlag, 1992.

[27] I. STRUB and A. BAYEN, "Weak formulation of boundary conditions for scalar conservation laws: An application to highway traffic modelling," *Int. J. Robust Nonlinear Control*, vol. 16, pp. 733–748, 2006.

[28] J. KAIPIO and E. SOMERSALO, *Statistical and Computational Inverse Problems*. New York, NY: Springer, 2005.

[29] A. Seppänen, M. Vauhkonen, E. Somersalo, and J. Kaipio, "State space models in process tomography – approximation of state noise covariance," *Inverse Problems in Engineering*, vol. 9, pp. 561–585, 2001.

[30] R. KALMAN, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[31] G. EVENSEN, "The ensemble kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[32] ——, "Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99, no. C5, pp. 10 143–10 162, 1994.

[33] A. HEEMINK, M. VERLAAN, and A. SEGERS, "Variance reduced ensemble Kalman filtering," *Monthly Weather Review*, vol. 129, pp. 1718–1728, 2001.

[34] G. BURGERS, P. VAN LEEUWEN, and G. EVENSEN, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review*, vol. 126, pp. 1719–1724, 1998.

[35] J. KAIPIO and E. SOMERSALO, "Statistical inverse problems: Discretization, model reduction and inverse crimes," *Journal of Computational and Applied Mathematics*, vol. 198, no. 2, pp. 493–504, 2007.

[36] G. BURGERS, P. JAN VAN LEEUWEN, and G. EVENSEN, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724, 1998.

[37] B. HOH, M. GRUTESER, R. HERRING, J. BAN, D. WORK, J. HERRERA, A. M. 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*, Breckenridge, CO, June 17-18 2008, pp. 15–28.

[38] QUADSTONE LIMITED, "PARAMICS Modeller Version 5.1 Reference Manual," 2004.

[39] X. BAN, L. CHU, and H. BENOUAR, "Bottleneck identification and calibration for corridor management planning," *Transportation Research Record*, vol. 1999, pp. 40–53, 2007.

[40] "511," publically available as a service provided to Bay Area commuters by the Metropolitan Transportation Commission, http://www.511.org.