Report: Decomposition in different modes. **Authors:** ...

Abstract

abstract

1 Introduction

Current highway traffic monitoring systems rely on large amounts of data from different sources. These include *inductive loop detectors* (ILD) used in the PeMS system [4] and *in-vehicle transponders* (IVTs) such as FasTrak. Recently, the available data on traffic has increased tremendously since the development of cellular phone based highway traffic monitoring. With the cellular phone communication infrastructure in place and privacy aware smartphone sensing technology in full expansion [14], a large volume of data from mobile devices is now available [13]. Numerous traffic estimation techniques developed in the literature rely on density based traffic models such as the Lighthill-Whitham-Richards (LWR) partial differential equation (PDE) [17, 19] and its discretization using the Godunov scheme [15, 16, 20] (also known as the Cell Transmission Model (CTM) [6, 7] in the transportation literature). Large scale applications include traffic flow estimation using the *Ensemble Kalman Filter* to assimilate velocity measurements [21, 22], which is a rapidly expanding field at the heart of mobile internet services.

In particular, a switching-mode model (SMM) has been derived from the CTM in [18], which is a nonlinear discrete time dynamical system. By switching among different sets of linear difference equations, or modes, the Mixture Kalman filter algorithm [5] is employed to assimilate data in a switching state-space model. In this paper, for a family of fundamental diagrams, we decompose the Godunov scheme applied to the LWR model (described in [7]) into different modes and apply the Extended Kalman Filter (EKF) for each mode. We show that the algorithm perform well and more importantly that such a decomposition in modes unravel the linearity of the CTM (Daganzo-Newell fundamental diagram), and hence the Extended Kalman filter is exaclty a Kalman filter in this case.

2 Traffic flow modelling

2.1 The LWR Model

Lighthill and Whitham in 1955 [17] introduce a macroscopic dynamic model of traffic based on conservation of cars (2.1), using Greenshields' hypothesis [11] of a static flow/density relationship (2.2), known as the *fundamental diagram*. The model consists of the following two equations:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{2.1}$$

$$q(x,t) = Q(\rho(x,t)) \tag{2.2}$$

where $\rho(x,t)$ and $q(x,t)$ denote the density and the flow of vehicles at location x and time t respectively, and Q is the flux function which is assumed to be a function of the density only.

Equation (2.1) is the principle of conservation of mass, or in this case conservation of vehicles, from fluid dynamics. These equations can be written more compactly as:

$$\frac{\partial \rho(x,t)}{\partial t} + Q'(\rho(x,t))\frac{\partial \rho(x,t)}{\partial x} = 0 \tag{2.3}$$

This equation is commonly known as the *Lighthill-Whitham-Richards*, or LWR, model. Different fundamental diagrams have been suggested. Greenshields [11] found that freeway speed

and density could be reasonably well approximated by a straight line. The expression of the velocity and the flux are then:

$$v = V_G(\rho) = v_f(1 - \frac{\rho}{\rho_j}) \tag{2.4}$$

$$Q_G(\rho) = \rho V_G(\rho) = v_f(\rho - \frac{\rho^2}{\rho_j}) \tag{2.5}$$

where v_f is the free flow (or maximum) velocity, and ρ_j is the jam (or maximum) density. In this case, the flow is a quadratic function of the density.

Many researchers have later suggested alternative shapes that provide a better fit to the measured data. They all share the same characteristics **LWR1-6**:

LWR1. Greenshields' hypothesis of a static flow/density relationship: $q = Q(\rho(x,t))$

LWR2. $Q(0) = Q(\rho_j)$=0

LWR3. The continuous portions of $Q(\rho)$ are concave.

LWR4. $V(0) = v_f$, and $V(\rho_j) = 0$.

LWR5. A critical density ρ_c can be defined where the maximum flow q_c is attained. Then, $Q(\rho)$ is increasing for $\rho \leq \rho_c$ and decrasing for $\rho > \rho_c$.

LWR6. The critical density ρ_c separates the fundamental diagram into two regimes: *free flow* when $\rho \leq \rho_c$ and *congestion* when $\rho > \rho_c$

Many researchers have later suggested alternative shapes that provide a better fit to the measured data. For instance, the widely used Daganzo-Newell velocity function assumes a constant velocity in free-flow and a hyperbolic velocity in congestion as shown in Figure 2.1:

$$v = V_{DN}(\rho) = \begin{cases} v_f & if \quad \rho \leq \rho_c \\ -\omega_f \left(1 - \frac{\rho_j}{\rho}\right) & if \quad \rho > \rho_c \end{cases} \tag{2.6}$$

and the corresponding flux function is:

$$Q_{DN}(\rho) = \rho V_{DN}(\rho) \begin{cases} v_f \rho & if \quad \rho \leq \rho_c \\ -\omega_f (\rho - \rho_j) & if \quad \rho > \rho_c \end{cases} \tag{2.7}$$

where ω_f is the backwards propagation wave speed.

In the context of the traffic model for velocity data assimilation [23], the Daganzo-Newell velocity function is approximated by a linear-hyperbolic velocity function, with a linear expression in free-flow ($\rho \leq \rho_c$) and a hyperbolic expression in congestion ($\rho > \rho_c$) in order to have an invertible velocity function:

$$v = V_{LH}(\rho) = \begin{cases} v_f \left(1 - \frac{\rho}{\rho_j}\right) & if \quad \rho \leq \rho_c \\ -\omega_f \left(1 - \frac{\rho_j}{\rho}\right) & if \quad \rho > \rho_c \end{cases} \tag{2.8}$$

where v_c is the critical velocity $v_c = V(\rho_c)$. For continuity of the flux at the critical density ρ_c, the additional relation $\frac{\rho_c}{\rho_j} = \frac{\omega_f}{v_f}$ must be satisfied. The corresponding flux function is:

$$Q_{LH}(\rho) = \rho V_{LH}(\rho) = \begin{cases} v_f \ \rho(1 - \frac{\rho}{\rho_j}) & if \quad \rho \leq \rho_c \\ -\omega_f (\rho - \rho_j) & if \quad \rho > \rho_c \end{cases} \tag{2.9}$$

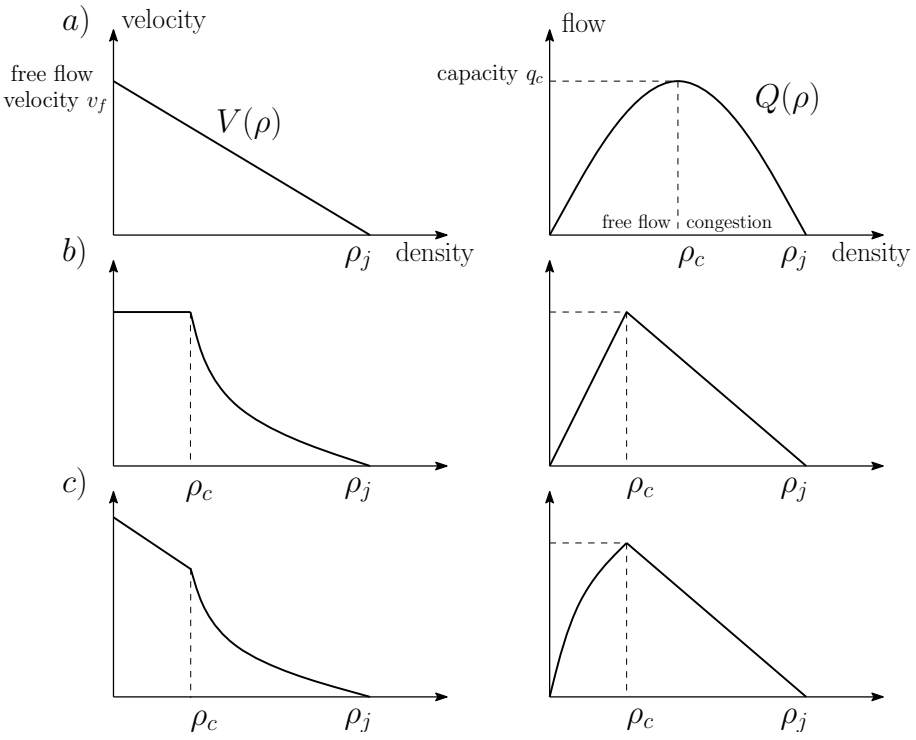

Figure 2.1: Speed and flow relationships (fundamental diagrams) for Greenshields (a), Daganzo-Newell (b), linear-hyperbolic (c).

2.2 Numerical Discretization

A good numerical method to solve the equations along roads is represented by the Godunov scheme, which is based on exact solutions to Riemann problems [9, 10]. This leads to the construction of a nonlinear discrete time dynamical system.

The Godunov discretization scheme is applied on the LWR PDE, where the discrete time step ΔT is indexed by t, and the discrete space step is indexed by i:

$$\rho_i^{t+1} = \rho_i^t - \frac{\Delta T}{\Delta X}\left(G(\rho_i^t, \rho_{i+1}^t) - G(\rho_{i-1}^t, \rho_i^t)\right) \tag{2.10}$$

where the Godunov flux $G(\rho_1, \rho_2)$ is in general defined as:

$$G(\rho_1, \rho_2) = \begin{cases} \min_{\rho\in[\rho_1,\rho_2]} Q(\rho) & \text{if} \quad \rho_1 \le \rho_2 \\ \max_{\rho\in[\rho_2,\rho_1]} Q(\rho) & \text{if} \quad \rho_2 \le \rho_1 \end{cases} \tag{2.11}$$

In order to ensure numerical stability, the time and space steps are coupled by the CFL condition [16]: $c_{max}\frac{\Delta T}{\Delta X} \le 1$ where c_{max} denotes the maximal characteristic speed.

For a family of flux functions $Q(\rho)$ that share the same characteristics **LWR1-6** listed above, the Godunov flux takes an explicit form[1]. It can also be expressed as the minimum of the *sending flow* from the upstream cell and the *receiving flow* from the downstream cell through a boundary connecting two cells of a homogeneous road (i.e. the upstream and downstream cells have the same characteristics). The *sending flow* $S(\rho)$ is equal to the upstream flow if the upstream traffic is in free flow ($\rho \le \rho_c$) or the capacity of the upstream section q_c if the upstream traffic is in congestion ($\rho > \rho_c$); on the other hand, the *receiving flow* $R(\rho)$ is equal to the capacity of

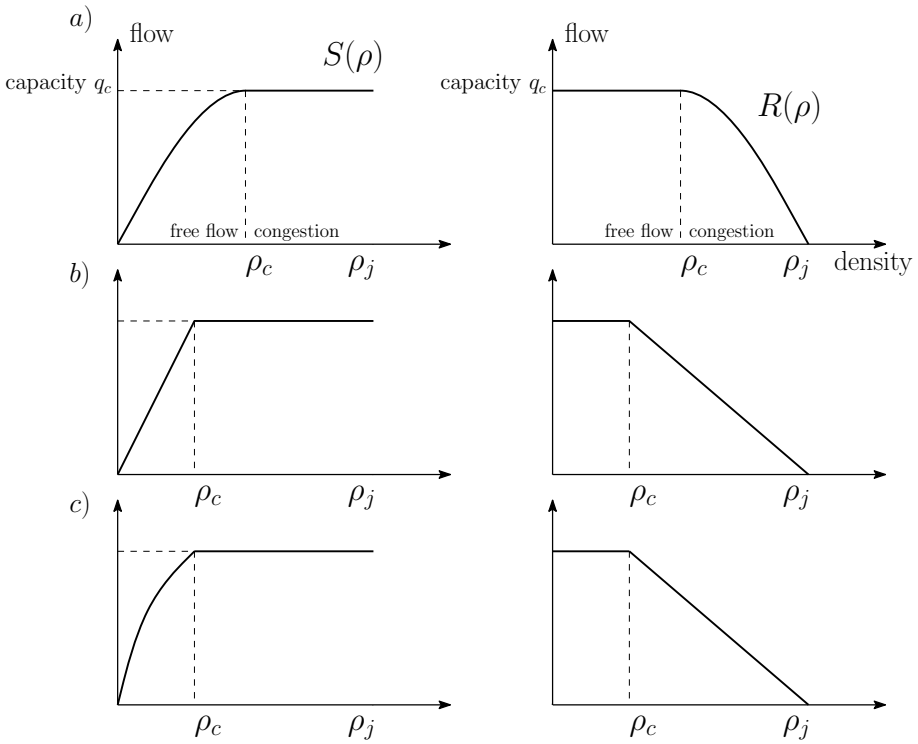

Figure 2.2: Sending and receiving flows for Greenshields (a), Daganzo-Newell (b), and linear-hyperbolic (c) velocity functions.

the downstream section if the downstream traffic is in free flow or the downstream flow if the downstream traffic is in congestion.

$$G(\rho_1, \rho_2) = \min(S(\rho_1), R(\rho_2)) \tag{2.12}$$

$$S(\rho) = \begin{cases} Q(\rho) & \text{if} \quad \rho \leq \rho_c \\ q_c & \text{if} \quad \rho > \rho_c \end{cases} \tag{2.13}$$

$$R(\rho) = \begin{cases} q_c & \text{if} \quad \rho \leq \rho_c \\ Q(\rho) & \text{if} \quad \rho > \rho_c \end{cases} \tag{2.14}$$

[1] The application of the Godunov scheme (2.11) to the fundamental diagrams is expressed in this system of equations, which is equivalent to (2.12, 2.13, 2.14) but less intuitive:

$$G(\rho_1, \rho_2) = \begin{cases} Q(\rho_2) & \text{if} \quad \rho_c \leq \rho_2 \leq \rho_1 \\ q_c & \text{if} \quad \rho_2 \leq \rho_c \leq \rho_1 \\ Q(\rho_1) & \text{if} \quad \rho_2 \leq \rho_1 \leq \rho_c \\ \min(Q(\rho_1), Q(\rho_2)) & \text{if} \quad \rho_1 \leq \rho_2 \end{cases}$$

As shown in Figure 2.2, the application of the Godunov scheme to the fundamental diagrams introduces intuitive concepts of *supply* and *demand* at the boundary connecting two cells. The upstream cell supplies the flow at the boundary up to capicity. We can note that in the discontinuous case, there is a drop in supply capacity when the upstream traffic is in congestion, as described in [1, 3, 12]. As a result, the flow through the boundary is smaller, even if the downstream cell can receive more flow. On the other hand, when the downstream traffic is congested, there is a decrease in demand from the downstream cell, limiting the flow through the boundary.

Figure 2.3 shows the explicit values taken by $G(\rho_1, \rho_2)$ in different regions of the space (ρ_1, ρ_2): $G(\rho_1, \rho_2)$ is equal to $Q(\rho_2) = R(\rho_2)$ in the *white region*, q_c in the *light grey region*, where q_c is the capacity at the interface between the two adjacent cells, and $Q(\rho_1)$ in the *dark grey region*. Specifically, the boundary between the white and grey regions follows the $(\rho_1, \rho_2) = (\rho_1, \bar{R}^{-1}(\bar{S}(\rho_1)))$[2]trajectory for $\rho_1 \leq \rho_{c1}$, where $\bar{S}$ and $\bar{R}$ denote the restrictions of the sending and receiving flows to the sub-regions $[0, \rho_c)$ and $(\rho_c, \rho_j]$ respectively, which also correspond to the left and right parts (w.r.t. ρ_c) of the fundamental diagram, as shown in the Figure 2.3.

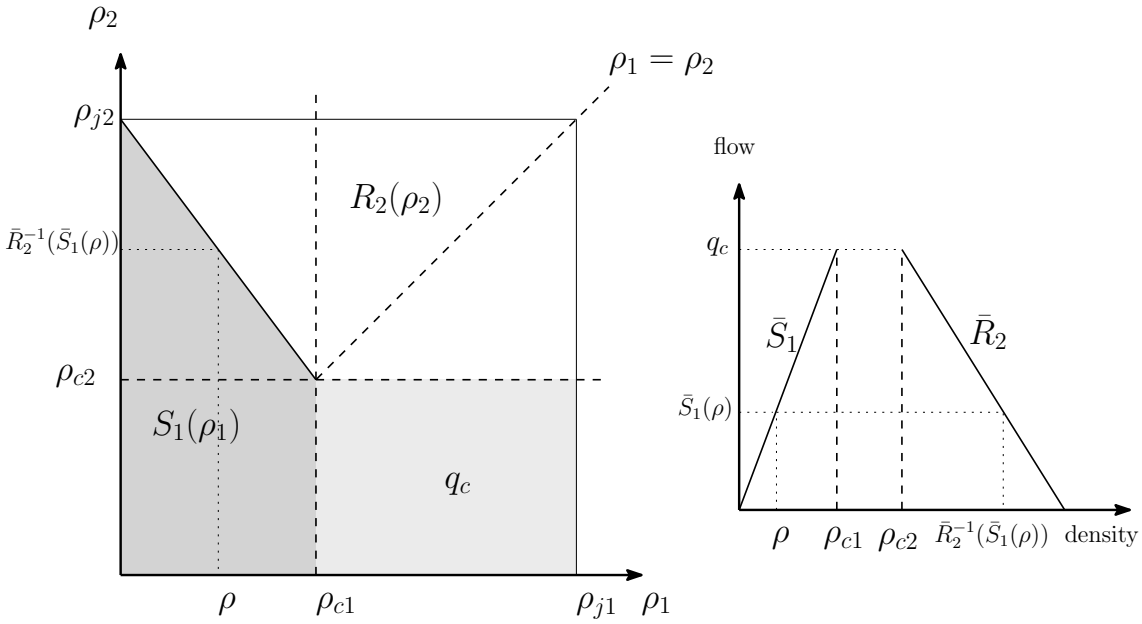

Figure 2.3: Values of $G(\rho_1, \rho_2)$ in the space (ρ_1, ρ_2) for a family of lfux functions with the characteristics **LWR1-6**.

When the velocity is the Daganzo-Newell function (2.6), the Godunov Flux becomes:

$$G_{DN}(\rho_1, \rho_2) = \begin{cases} -\omega_f (\rho_2 - \rho_j) & \text{if} \quad (\rho_1, \rho_2) \in \text{white region} \\ q_c & \text{if} \quad (\rho_1, \rho_2) \in \text{light region} \\ v_f \rho_1 & \text{if} \quad (\rho_1, \rho_2) \in \text{dark region} \end{cases} \tag{2.15}$$

When the velocity is the linear-hyperbolic function (2.8), the Godunov Flux becomes:

$$G_{LH}(\rho_1, \rho_2) = \begin{cases} -\omega_f (\rho_2 - \rho_j) & \text{if} \quad (\rho_1, \rho_2) \in \text{white region} \\ q_c & \text{if} \quad (\rho_1, \rho_2) \in \text{light region} \\ v_f \left(\rho_1 - \frac{\rho_1^2}{\rho_j}\right) & \text{if} \quad (\rho_1, \rho_2) \in \text{dark region} \end{cases} \tag{2.16}$$

[2]Here, we suppose that $\bar{R}$ is a strictly monotonic function on $(\rho_c, \rho_j]$, hence invertible, and $\bar{R}^{-1}$ denotes its inverse.

Subsequently, we will denote by **W**, **L**, and **D** the *white region*, *light region*, and *dark region* of the space (ρ_1, ρ_2) respectively.

3 Linearization of the Godunov scheme

In the Godunov scheme (2.10), the update of the density ρ_i^{t+1} at $x = i$ depends on the triplet $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$. In this section, we will refer to this triplet as (ρ_-, ρ, ρ_+), ρ_i^{t+1} as ρ^{t+1}, and the fraction $\frac{\Delta T}{\Delta X}$ as α to ease the notations. Then the Godunov scheme reads as:

$$\rho^{t+1} = f(\rho_-, \rho, \rho_+) = \rho - \alpha \left(G(\rho, \rho_+) - G(\rho_-, \rho)\right) \tag{3.1}$$

3.1 Decomposition in different "modes"

ρ^{t+1} depends on whether both pairs (ρ_-, ρ) and (ρ, ρ_+) are in **W**, **L**, or **D** via $G(\rho_-, \rho)$ and $G(\rho, \rho_+)$. So there are nine possible combinations at $x = i$, which can be reduced to seven "modes" since the pairs (ρ_-, ρ) and (ρ, ρ_+) have ρ in common. Table 3.1 list these seven possibilities, which can be easily derived from Figure 2.3.

Mode	(ρ_-, ρ)	(ρ, ρ_+)	f(ρ_-, ρ, ρ_+)	State
1	W	W	$\rho - \alpha(R(\rho_+) - R(\rho))$	congestion
2	W	L	$\rho - \alpha(q_c - R(\rho))$	congestion
3	L	W	$\rho - \alpha(R(\rho_+) - q_c)$	congestion
4	L	D	$\rho - \alpha(S(\rho) - q_c)$	free flow
5	D	W	$\rho - \alpha(R(\rho_+) - S(\rho_-))$	critical
6	D	L	$\rho - \alpha(q_c - S(\rho_-))$	free flow
7	D	D	$\rho - \alpha(S(\rho) - S(\rho^-))$	free flow

Table 3.1: Different values of $\rho^{t+1} = f(\rho_-, \rho, \rho_+)$ depending on the values of $G(\rho_-, \rho)$ and $G(\rho, \rho_+)$ in the space (ρ_1, ρ_2).

For instance, for the first mode, (ρ_-, ρ) and (ρ, ρ_+) are both in **W** (cf. Figure 2.3), thus $G(\rho_-, \rho) = R(\rho)$ and $G(\rho, \rho_+) = R(\rho_+)$, and then $\rho^{t+1} = \rho - \alpha(R(\rho_+) - R(\rho))$. By extending this result to an entire link with discrete state space indexed by $i = 1, ..., n$, where n is the number of space steps, we have a whole description of the space of "modes" along the link. Since there is a correlation between two consecutive indexes i and $i + 1$, the number of modes for the entire link reduces from 3^n to an expression in the form of $a.\beta^n + b.v\gamma^n + c.\delta^n$ which lower and upper bounds are proved to be 3.2^n and $3.(2.5)^n$ respectively (for full details, see Appendix A). Let's denote by $\mathcal{M}_n$ the space of modes ($\mathcal{M}_n \subset \{1,...,7\}^n$). For $m \in \mathcal{M}_n$, m is a vector of dimension n which $i - th$ entry $m_i \in \{1, ..., 7\}$ is the mode at cell i.

In the specific cases of the Daganzo-Newell fundamental diagram, and the linear-hyperbolic velocity function, the expression ρ^{t+1} reads:

We suppose α, ω_f, v_f, ρ_j, ρ_c, q_c constant w.r.t the time because they are only related to the geometry of the highway, independently of the current traffic on it[3]. Then, let's denote by $f_{DN}(\rho_-, \rho, \rho_+)$ and $f_{LH}(\rho_-, \rho, \rho_+)$ the vector functions for ρ^{n+1} associated with the Danganzo-Newell and the Linear-hyperbolic cases respectively, which variables are ρ_-, ρ, and ρ_+ (cf. Table

Mode	$f_{DN}(\rho_-, \rho, \rho_+)$	$f_{LH}(\rho_-, \rho, \rho_+)$
1	$(1 - \alpha\omega_f)\rho + \alpha\omega_f\rho_+$	$(1 - \alpha\omega_f)\rho + \alpha\omega_f\rho_+$
2	$(1 - \alpha\omega_f)\rho + \alpha\omega_f\rho_c$	$(1 - \alpha\omega_f)\rho + \alpha\omega_f\rho_c$
3	$\rho + \alpha\omega_f\rho_+ - \alpha\omega_f\rho_c$	$\rho + \alpha\omega_f\rho_+ - \alpha\omega_f\rho_c$
4	$(1 - \alpha v_f)\rho + \alpha v_f\rho_c$	$(1 - \alpha v_f)\rho + \alpha\frac{v_f}{\rho_j}\rho^2 + \alpha v_f\rho_c$
5	$\alpha v_f\rho_- + \rho + \alpha\omega_f\rho_+ - \alpha\omega_f\rho_j$	$\alpha v_f\rho_-(1 - \frac{\rho_-}{\rho_j}) + \rho + \alpha\omega_f\rho_+ - \alpha\omega_f\rho_j$
6	$\alpha v_f\rho_- + \rho - \alpha v_f\rho_c$	$\alpha v_f\rho_-(1 - \frac{\rho_-}{\rho_j}) + \rho - \alpha v_f\rho_c$
7	$\alpha v_f\rho_- + (1 - \alpha v_f)\rho$	$\alpha v_f\rho_-(1 - \frac{\rho_-}{\rho_j}) + (1 - \alpha v_f)\rho + \alpha\frac{v_f}{\rho_j}\rho^2$

Table 3.2: Different values of $\rho^{t+1} = f(\rho_-, \rho, \rho_+)$ for the Daganzo-Newell fundamental diagram (f_{DN}) and the Linear-hyperbolic velocity function (f_{LH}).

3.1). We define J, the Jacobian matrix of f with respect to (ρ_-, ρ, ρ_+):

$$J = \left(\frac{\partial f_i}{\partial \rho_j}\right)_{i,j} \tag{3.2}$$

f_i is the ith entry of the vector function f defined in Table 3.1.

It is useful to explicit the Jacobian matrices J_{DN} and J_{LH} of the vector functions f_{DN} and f_{LH} with respect to (ρ_-, ρ, ρ_+):

$$J_{DN} = \begin{pmatrix} 0 & 1 - \alpha\omega_f & \alpha\omega_f \\ 0 & 1 - \alpha\omega_f & 0 \\ 0 & 1 & \alpha\omega_f \\ 0 & 1 - \alpha v_f & 0 \\ \alpha v_f & 1 & \alpha\omega_f \\ \alpha v_f & 1 & 0 \\ \alpha v_f & 1 - \alpha v_f & 0 \end{pmatrix} \tag{3.3}$$

$$J_{LH} = \begin{pmatrix} 0 & 1 - \alpha\omega_f & \alpha\omega_f \\ 0 & 1 - \alpha\omega_f & 0 \\ 0 & 1 & \alpha\omega_f \\ 0 & 1 - \alpha v_f + 2\alpha\frac{v_f}{\rho_j}\rho & 0 \\ \alpha v_f(1 - 2\frac{\rho_-}{\rho_j}) & 1 & \alpha\omega_f \\ \alpha v_f(1 - 2\frac{\rho_-}{\rho_j}) & 1 & 0 \\ \alpha v_f(1 - 2\frac{\rho_-}{\rho_j}) & 1 - \alpha v_f + 2\alpha\frac{v_f}{\rho_j}\rho & 0 \end{pmatrix} \tag{3.4}$$

Since f_{DN} is a *linear function* of (ρ_-, ρ, ρ_+) as shown in Table 3.1, we can notice that J_{DN} is a constant. More notably, we will see in the next section that the decomposition in "modes" as shown in Table 3.1 leads to a linear formulation of the Godunov Scheme in the case of the Daganzo-Newell fundamental diagram, which fits into the Kalman filter framework as we will see in the next section.

For a linear-hyperbolic velocity function (2.8), the sending flow is a parabolic function of the density as shown in Figure 2.2, and so the Godunov scheme decomposed in different "modes" has some parabolic components as shown in Table 3.1. However, the calibration of the fundamental

[3] In fact, α, ω_f, v_f, ρ_j, ρ_c, q_c can vary along the link and a proper notation would be α_i, ω_{fi}, v_{fi}, ρ_{ji}, ρ_{ci}, q_{ci} for the parameters of the fundamental diagram at $x = i$. We choose not to add the subscripts i to simplify the notations.

diagram given flow measurements empirically shows that the *sending flow* component is almost linear (see Appendix B). Hence, we will see in the next section that the first order multivariate Taylor Series expansion of the prediction operator is a good linear approximation of the function, which fits into the Extended Kalman filter framework.

3.2 Kalman filtering

Let's consider a link with discrete time step indexed by $t \geq 0$ and discrete space step indexed by $i = 1, ..., n$, and let's denote $\boldsymbol{\rho}^t = (\rho_0^t, \rho_1^t, ..., \rho_n^t, \rho_{n+1}^t)$ a n+2 dimensional vector which describes the state of the link at time t. ρ_i^t is the density at time t and cell i. We can note that the ghost cells 0 and $n+1$ are included in the state of the link. $\boldsymbol{m}^t \in \mathcal{M}_n$ is the mode of the link at time t, as defined in the previous section. At each time increment, the "true" state of the link is updated through:

$$\boldsymbol{\rho}^t = F[\boldsymbol{\rho}^{t-1}] \tag{3.5}$$

with $F[.]$ a n+2 dimensional function vector such that the ith entry $F_i[\boldsymbol{\rho}^{t-1}]$ is:

$$F_i[\boldsymbol{\rho}^{t-1}] = \begin{cases} f_{m_i}(\rho_{i_-,i,i_+}^{t-1}) & \text{for} \quad i = 1, ..., n \\ \rho_0^t & \text{for} \quad i = 0 \\ \rho_{n+1}^t & \text{for} \quad i = n+1 \end{cases} \tag{3.6}$$

where m_i denotes the ith entry of $\boldsymbol{m}^t \in \mathcal{M}_n$ (we omit the subscript i to ease the notations), $f_{m_i}(\rho_{i_-,i,i_+}^{t-1})$ is the m_i-th entry of the function vector f evaluated in $(\rho_{i-1}^{t-1}, \rho_i^{t-1}, \rho_{i+1}^{t-1})$, and ρ_0^t and ρ_{n+1}^t are the boundary conditions upstream and downstream. For a Daganzo-Newell fundamental diagram and a linear-hyperbolic velocity function, the update operators of the dynamic system, namely $F_{DN}[.]$ and $F_{LH}[.]$, have the same expression with $f_{DN}(.)$ and $f_{LH}(.)$ respectively (as defined in Table 3.1). For instance, when we have a Daganzo-Newell fundamental diagram with update operator for the dynamic system $F_{DN}[.]$ (i.e. $\boldsymbol{\rho}^t = F_{DN}[\boldsymbol{\rho}^{t-1}]$), and suppose that the mode at $x = i$ is three ($m_i = 3$) then[4]:

$$F_i[\boldsymbol{\rho}^{t-1}] = f_{DN,m_i}(\rho_{i_-,i,i_+}^{t-1}) = f_{DN,3}(\rho_{i_-,i,i_+}^{t-1}) = \rho_i^{t-1} + \alpha\omega_f\rho_{i+1}^{t-1} - \alpha\omega_f\rho_c \tag{3.7}$$

In order to use the *Extended Kalman filter* to estimate the state of the link given a sequence of noisy observations, we model the process in accordance with the framework of the *Extended Kalman filter* by adding a white noise to the underlying dynamic system model. The "true" state $\boldsymbol{\rho}^t$ is then:

$$\boldsymbol{\rho}^t = F[\boldsymbol{\rho}^{t-1}] + \boldsymbol{\eta}^t \tag{3.8}$$

where $\boldsymbol{\eta}^t \sim N(0, Q_t)$ is the Gaussian zero-mean, white state noise with covariance Q_t. The estimated state at time t is denoted by $\hat{\boldsymbol{\rho}}^t$ and the estimated covariance by P_t. The *prediction step* gives the *a priori* state estimate and covariance $\hat{\boldsymbol{\rho}}^{t:t-1}$ and $P_{t:t-1}$:

$$\begin{array}{ll} \text{Predicted state estimate:} & \hat{\boldsymbol{\rho}}^{t:t-1} = F[\hat{\boldsymbol{\rho}}^{t-1}] \\ \text{Predicted covariance estimate:} & P_{t:t-1} = F_{t-1}P_{t-1}(F_{t-1})^T + Q_{t-1} \end{array} \tag{3.9}$$

where F_t is the state transition defined to be the following Jacobian:

[4] As we have seen earlier, α, ω_f, v_f, ρ_j, ρ_c, q_c can vary along the link and a proper notation would be α_i, ω_{fi}, v_{fi}, ρ_{ji}, ρ_{ci}, q_{ci} for the parameters of the fundamental diagram at $x = i$, so that:

$$F_i[\rho^{t-1}] = f_{DN,m_i}(\rho_{i_-,i,i_+}^{t-1}) = f_{DN,3}(\rho_{i_-,i,i_+}^{t-1}) = \rho_i^{t-1} + \alpha_i\omega_{fi}\rho_{i+1}^{t-1} - \alpha_i\omega_{fi}\rho_{ci}$$

$$F_t = \left(\frac{\partial\ F_i[\rho^t]}{\partial \rho_j^t} \right)_{i,j} \tag{3.10}$$

The estimated mode $\hat{\boldsymbol{m}}^t$ associated to the estimated vector state $\hat{\rho}^t$ is defined from Table 3.1. Specifically, in the context of our traffic model, the density at $x = i$ only depends on the densities at the neighbor points $x = i-1, i, i+1$. So F_t is a $(n+2) \times (n+2)$ tridiagonal matrix, such that the diagonal elements are $\{0, J_{\hat{m}_1,2}, ..., J_{\hat{m}_n,2}, 0\}$, the lower diagonal elements are $\{J_{\hat{m}_1,1}, J_{\hat{m}_2,1}, ..., J_{\hat{m}_n,1}, 0\}$, and the upper diagonal elements are $\{0, J_{\hat{m}_1,3}, J_{\hat{m}_2,3}, ..., J_{\hat{m}_n,3}\}$, where J is defined in Equation 3.10.

In the case of the Daganzo-Newell fundamental diagram, the operator $F_{DN}[.]$ defined in (3.6) is *linear* (from the linearity of f_{DN}). Along with the assumption of a white state noise, the *prediction step* (3.12) of the *Extended Kalman filter* is actually identical to the regular *Kalman filter*, and it is known from the theory that the *Kalman filter* is optimal [2].

In the case of a linear-hyperbolic velocity function, the operator $F_{LH}[.]$ is not linear if at some points of the discrete space the mode is between four and seven. Then the *Extended Kalman filter* is *near-optimal*.

We can note that the first line and first column of P_t have only zero elements because the boundary condition ρ_0^t is deterministic (i.e. $cov(\rho_0^t, \rho_i^t) = 0$ for $i = 1, ..., n$), and similarly the last line and last column of P_t are null since the boundary condition ρ_{n+1}^t is deteministic. Additionally, the observation model for the link is given by:

$$\boldsymbol{y}^t = H_t \boldsymbol{\rho}^t + \boldsymbol{\chi}^t \tag{3.11}$$

where $H_t \in \{0,1\}^{p_t \times n}$ is the linear observation observation matrix which encodes the p_t observations (each one of them being at a discrete cell on the highway) for which the density is observed during discrete time step t, and n is the number of cells along the link. The last term in (3.11) is the white, zero mean observation noise $\boldsymbol{\chi}^t \sim N(0, R_t)$ with covariance matrix R_t. The *update step* is:

$$\begin{array}{ll} \text{Kalman gain:} & K_t = P_{t:t-1} H_t^T \left(H_t P_{t:t-1} H_t^T + R_t \right)^{-1} \\ \text{Updated state estimate:} & \hat{\boldsymbol{\rho}}^t = \hat{\boldsymbol{\rho}}^{t:t-1} + K_t (\boldsymbol{y}^t - H_t \hat{\boldsymbol{\rho}}^{t:t-1}) \\ \text{Updated estimate covariance:} & P_t = (I - K_t H_t) P_{t:t-1} \end{array} \tag{3.12}$$

The Extended Kalman filter provides the distribution of $\hat{\boldsymbol{\rho}}^t$ given the sequence of observations $\boldsymbol{y}^{0:t}$, sequence of modes $\hat{\boldsymbol{m}}^{0:t} = \{\hat{\boldsymbol{m}}^0, ..., \hat{\boldsymbol{m}}^t\}$, and sequence of control parameters $\boldsymbol{u}^{0:t}$, which is exactly equal to the distribution of $\hat{\boldsymbol{\rho}}^t$ since the sequence $\hat{\boldsymbol{m}}^{0:t}$ is deterministic. Concretely, the vector of control parameters $\boldsymbol{u}^t$ contain the vector of critical densities $\boldsymbol{\rho}_c$, the vector of jam densities $\boldsymbol{\rho}_j$, and the boundary conditions ρ_0^t and ρ_{n+1}^t. Theoretically, the result is obtained by marginalizing the joint distribution of $\hat{\boldsymbol{\rho}}^t$ and $\hat{\boldsymbol{m}}^{0:t}$ as follows:

$$\begin{array}{ll} p(\hat{\boldsymbol{\rho}}^t | \boldsymbol{y}^{0:t}, \boldsymbol{u}_{0:t}) & = \int_{\mathcal{M}_n} p(\hat{\boldsymbol{\rho}}^t | \boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \boldsymbol{m}^{0:t}) p(\boldsymbol{m}^{0:t} | \boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}) d\boldsymbol{m}^{0:t} \\ & = \int_{\mathcal{M}_n} p(\hat{\boldsymbol{\rho}}^t | \boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \boldsymbol{m}^{0:t}) \mathbf{1}_{\hat{\boldsymbol{m}}^{0:t}} d\boldsymbol{m}^{0:t} \\ & = p(\hat{\boldsymbol{\rho}}^t | \boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \hat{\boldsymbol{m}}^{0:t}) \end{array} \tag{3.13}$$

4 Implementation

4.1 Algorithm for the *prediction step*

Algorithm using the structure of F_t and extension to a network

From Appendix A, In (3.12), the *a priori* state estimate $\hat{\boldsymbol{\rho}}^{t:t-1}$ is derived through the algorithm in (3.6).

4.2 Accuracy

For a Daganzo-Newell fundamental diagram, we can note that the decomposition in different modes in (3.6) is in fact a special case of *Conditional Dynamic Linear Model* (CDLM) as described in [5] with a discrete latent indicator that is deterministic. In this case, there is only one estimated deterministic sequence of modes $\hat{\boldsymbol{m}}^{0:t}$ or latent indicators, hence there is no need to use a sequential Monte-Carlo method to sample an ensemble of trajectories of modes. The Mixture Kalman Filter [5] reduces to a simple Kalman Filter. For non-triangular fundamental diagrams which have the characteristics **LWR1-6**, a first order Taylor Series expansion is applied. Such a linearization is a good approximation, as shown in the example in Appendix B. Then the Extended Kalman filter can be applied to each mode.

In our case, the estimated sequence of modes $\hat{\boldsymbol{m}}^{0:t}$ is readily infered through the Godunov scheme, and at all the cells along the link, including those where there is no observation. Contrary to some previous models as in [18], the modes are not directly sampled from density measurements along the highway. However, the mode is still indirectly infered from those measurements by assimilating the observations with the *update state* of the Kalman filter. Besides, we rely on the accuracy of the estimation of the mode provided by the Godunov scheme when we apply the Kalman filter for each mode. Such an assumption that favors the mode provided by the Godunov scheme is yet another reason why the sequential Monte-Carlo method (with the resampling step) is not used in the algorithm, because we then suppose that our sequence of modes is the one with the largest likelihoods. (to develop, do a test of likelihood)

4.3 Complexity

Comparison with the Extended Kalman filter and matlab simulations and results

5 Conclusion and Future Work

6 Acknowledgement

A Description of the state of modes

For an entire link with discrete state space indexed by $i = 1, ..., n$ we have a whole description of the space of "modes" along it. Since there is always the entry ρ_i in common for successive pairs (ρ_{i-1}, ρ_i) and (ρ_i, ρ_{i+1}), a correlation propagates along the link, reducing the number of modes to a quantity smaller than 3^n.

More precisely, suppose that the pair (ρ_0, ρ_1) is in the region **W**, then the list of possible combinations in Table 3.1 shows that (ρ_1, ρ_2) can be either in **W** or **L**. Similarly, if (ρ_0, ρ_1) is in the region **L**, (ρ_1, ρ_2) can be either in **W** or **L**, and for (ρ_0, ρ_1) in **D**, (ρ_1, ρ_2) can be either in **W**, **L**, or **D**. As an example, Table A describes all the possible sixteen combinations for the first three pairs (ρ_0, ρ_1), (ρ_1, ρ_2), and (ρ_2, ρ_3).

(ρ_0, ρ_1)	W				L					D						
(ρ_1, ρ_2)	W		L		W		D			W		L		D		
(ρ_2, ρ_3)	W	L	W	D	W	L	W	L	D	W	L	W	D	W	L	D

Table A.1: Different values of ρ^{n+1} depending on the values of $G(\rho_-, \rho)$ and $G(\rho, \rho_+)$ in the space (ρ_1, ρ_2) for the Daganzo-Newell fundamental diagram and the Linear-hyperbolic velocity function.

We can recursively compute the number of "modes" M_n with respect to n, where n is the

number of cells of the discretized link. Let's denote by w_n, l_n, and d_n the number of modes for which (ρ_k, ρ_{k+1}) is in **W**, **L**, and **D** respectively. Then we have these equations:

$$w_0 = l_0 = d_0 = 1 \tag{A.1}$$

$$\begin{array}{rcccccl} w_{k+1} & = & w_k & + & l_k & + & d_k \\ l_{k+1} & = & w_k & + & d_k & & \text{for } k \geq 0 \\ d_{k+1} & = & l_k & + & d_k & & \end{array} \tag{A.2}$$

$$n_k = w_k + l_k + d_k \quad \text{for } k \geq 0 \tag{A.3}$$

Using matrix notations, equation (A.2) reads:

$$\begin{bmatrix} w_{k+1} \\ l_{k+1} \\ d_{k+1} \end{bmatrix} = A \times \begin{bmatrix} w_k \\ l_k \\ d_k \end{bmatrix} \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{A.4}$$

Then

$$\begin{bmatrix} w_k \\ l_k \\ d_k \end{bmatrix} = A^k \times \begin{bmatrix} w_0 \\ l_0 \\ d_0 \end{bmatrix} \tag{A.5}$$

It is possible to compute A^k explicitly by diagonalizing the matrix A, to get an explicit expression for w_k, l_k, and d_k in the form of $a.\beta^k + b.\gamma^k + c.\delta^k$. However, this analytical expression is unwieldy, so we will just derive lower and upper bounds to n_k. It is easy to see that $d_k \leq n_k/2$ for $k \geq 0$, then we can prove recursively that $3.2^k \leq n_k \leq 3.(2.5)^k$.

B Calibration of the Linear-hyperbolic fundamental diagram

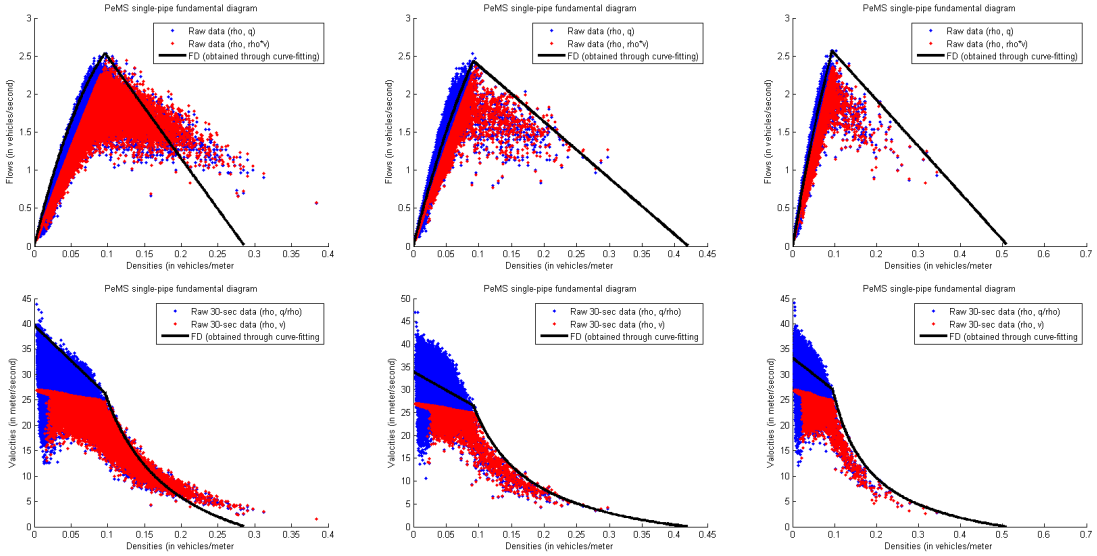

Figure B.1: Calibrated fundamental diagram.

PeMS id		400041	400137	400141	400180	400189
v_f	(m/s)	39.6631	33.9159	33.2646	30.0710	33.1679
ω_f	(m/s)	13.3431	7.3998	6.1202	6.4434	9.0381
ρ_j	$(vehicles/m)$	0.2861	0.4206	0.5139	0.6321	0.3701
q_c	$(vehicles/s)$	2.5333	2.4333	2.5667	3.2000	2.4333
ρ_c	$(vehicles/m)$	0.0963	0.0918	0.0946	0.1354	0.1008
ρ_c/ρ_j		0.3366	0.2183	0.1841	0.2142	0.2724
R^2		0.983	0.995	0.997	0.995	0.991

Table B.1: Calibrated parameters of the fundamental diagram (with a linear-hyperbolic velocity function) given PeMS measurements from five stations along the I-880 for the freeflow velocity v_f, the backwave velocity ω_f, the jam density ρ_j, the flow capacity q_c, and the density capacity ρ_c.

The California Freeway Performance Measurement System stores real-time data from 26,000 loop detectors [4]. PeMS data is accessed via an internet browser (http://pems.eecs.berkeley.edu/). Such rich data set has been used to formulate and address practical questions in freeway modelling. In [8], a method for automated, empirical calibration of freeway traffic flow characteristics is presented. The method was used to calibrate a cell transmission model of Interstate-880 in San Francisco Bay Area, California, a 40-mile long urban freeway with lots of recurrent and non-recurrent congestion and with dozens of loop detector stations. For the purpose of our study, the method was modified to calibrate the traffic flow characteristics for a linear-hyperbolic velocity function (2.8). The calibrated fundamental diagrams and velocity functions are presented in Figure B.1, and the calibrated parameters are presented in Table B.

From Equation 2.9, the maximum contribution of the quadratic component over the linear one is given by ρ_c/ρ_j, which values have been obtained empirically from the calibrated parameters in Table B. One way to test if the first order Taylor Series expansion of the prediction operator is a good linear approximation of $F_{LH}[.]$ is to fit a linear model to the parabole using the *least squares linear regression*, and compute the *coefficient of determination* R^2. As shown in Table B, R^2 is very close to one, which means that the linear model is a good fit.

References

[1] K. Agyemang-Duah and F. Hall. Some issues regarding the numerical value of freeway capacity. *International Symposium on Highway Capacity*, 1991.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Dover Publications, Inc., 2005.

[3] M. Cassidy and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research*, 1999.

[4] C. Chen, P. Varaiya, and J. Kwon. An empirical assessment of traffic operations. *International Symposium on Transportation and Traffic Theory*, 2005.

[5] Rong Chen and Jun S. Liu. Mixture kalman filters. *Royal Statistical Society*, 2000.

[6] C. F. Daganzo. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B 28, no. 4*, 1994.

[7] C. F. Daganzo. The cell transmission model, part ii: network traffic. *Transportation Research Part B 29, no. 2*, 1995.

[8] G. Dervisoglu, G. Gomes, J. Kwon, A. Muralidharan, Pravin Varaiya, and R. Horowitz. Automatic calibration of the fundamental diagram and empirical observations on capacity. *88th Annual Meeting, January 2009, Transportation Research Board*, 2008.

[9] Edwige Godlewski and Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Applied Mathematical Sciences, 1996.

[10] S.K. Godunov. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. 1959.

[11] B. D. Greenshields. A study of traffic capacity. *Proceedings of the 14th annual meeting of the Highway Research Board*, 1934.

[12] F. Hall and K. Agyemang-Duah. Freeway capacity drop and the definition of capacity. *Transportation Research Record*, 1991.

[13] J.-C. Herrera, D. B. Work, R. Herring, J. Ban, Q. Jacobson, and A. M. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: the mobile century experiment. *Transportation Research Part C*, 2009.

[14] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. *6th International Conference on Mobile Systems, Applications, and Services*, 2008.

[15] J. P. Lebacque. The godunov scheme and what it means for first order traffic flow models. *13th International Symposium on Transportation and Traffic Theory*, 1996.

[16] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser Basel, 1992.

[17] M. J. Lighthill and G. B. 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*, pages 317–345, 1955.

[18] Laura Munoz, Xiaotian Sun, Roberto Horowitz, and Luis Alvarez. Traffic density estimation with the cell transmission model. *Proceedings of the American Control Conference*, 2003.

[19] P. I. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

[20] I. S. Strub and A. M. Bayen. Weak formulation of boundary conditions for scalar conservation laws: an application to highway traffic modeling. *Int. J. Robust Nonlinear Control*, 2006.

[21] D. B. Work and A. M. Bayen. Impacts of the mobile internet on transportation cyberphysical systems: Traffic monitoring using smartphones. *National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, & Rail*, 2008.

[22] D. B. Work, A. M. Bayen, and Q. Jacobson. Automotive cyber-physical systems in the context of human mobility. *National Workshop on High-Confidence Automotive Cyber-Physical Systems*, 2008.

[23] D. B. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. M. Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 2010.