Bayesian Particle Tracking of Traffic Flows

Nicholas Polson
Booth School of Business, University of Chicago
and
Vadim Sokolov
Argonne National Laboratory

December 1, 2014

Abstract

In this paper, we develop a model of traffic flows that is capable of capturing non-linearities and discontinuities present in flow dynamics. Our model includes a hidden state variable that captures sudden regime shifts between traffic free flow, breakdown and recovery. We construct an efficient particle learning algorithm for real time online inference and prediction. This requires a two step algorithm to resample current particles with a mixture predictive distribution and then to propagate states using the conditional posterior. Particle learning of parameters is achieved via updating recursions for conditional sufficient statistics. We also describe how our model can be adapted to special events and severe weather conditions. To illustrate our methodology, we analyze measurements of daily traffic flow from the Illinois interstate I-55 highway system. Finally, we conclude with directions for future research.

Keywords: traffic flow, particle filter, data analysis
1 Introduction

Modeling traffic dynamics of a transportation system that includes highways, arterial rods and public transit, is an important task for effectively managing traffic flow. This task requires estimation of traffic flow conditions on highways from field measurements such as in-ground loop detectors or GPS probes. Traffic managers use model-based forecasts to regulate ramp metering, apply speed harmonization, and to change road pricing as congestion mitigation strategies. The general public uses model-based predictions for making decisions on departure times, and travel route choices among other things.

This paper proposes a state-space model that incorporates a hidden switching variable together with a traffic flow state to capture the non-linearities and discontinuities in traffic flows. Our dynamic model allows for three types of traffic flow regimes; free flow, breakdown and recovery. A physical interpretation of the change in the flow regime is a traffic queue with congestion inside the queue and free flow outside. In addition to switching and traffic flow states we track a state that measures the rate of degradation or recovery. Traffic flow data measurements are sparse as both fixed and moving sensors are located at a small number of locations at any given point in time. Statistical estimation is further complicated by noisy non-Gaussian observations; for example, Rasschaert (2003) show that the error distribution is a mixture of Poissons for video cameras.

Real-time sequential inference, and prediction is achieved with an efficient particle filter and learning algorithm. This allows us to track in real-time the speed of traffic flow together with the latent variables that describe regime switches and the rate of degradation (recovery). To illustrate our methodology we use data from a network of in-ground loop detectors on Chicago’s Interstate I-55 with measurements on speed, counts and occupancy of traffic flow. The sequential nature of particle filtering makes frequent updating feasible. Our methodology, therefore, provides an alternative to Markov Chain Monte Carlo (MCMC) simulation methods whose computational cost grows linearly with then number of observations. Particle filters have previously been applied to traffic flow problems, see Mihaylova et al. (2007) who use the evolution dynamics as a proposal distribution before re-sampling, the so-called bootstrap or sampling/importance resampling (SIR) filter. We improve the efficiency of inference and prediction with a Rao-Blackwellized filter and we
also incorporate particle learning.

Our approach builds upon existing statistical methods. Tebaldi and West (1998) infer network route flows and Westgate et al. (2013) develop MCMC methods to estimate travel time reliability for ambulances using noisy GPS for both path travel time and an individual road segment travel time distributions. Anacleto et al. (2013) propose a dynamic Bayesian network to model external intervention techniques to accommodate situations with suddenly changing traffic variables.

Small changes in consecutive speed measurements can be explained by sensor noise but significant changes require estimation of probability of a flow regime switch, which represented by a discontinuity in traffic flow speed. Previous work on estimating traffic flows use extensions of the Kalman filter and relies heavily on Gaussianity assumptions, see Gazis and Knapp (1971); Schreiter et al. (2010); Wang and Papageorgiou (2005); Work et al. (2008). Blandin et al. (2012); Polson and Sokolov (2014) recently show that dynamic properties of traffic flow such as discontinuities (or shock waves) lead to forecast distributions that are mixtures. Our non-Gaussian state-space model will explicitly captures such behavior.

Our approach builds on the current literature in a number of ways:

1. Provides a hierarchical model rather than, conservation law based model. Our approach avoids boundary condition estimation

2. A predictive likelihood particle filter provides an efficient estimation strategy. Our filter is less sensitive to measurement outliers and less prone to particle degeneracy

3. Tracks traffic flows variable such as flow speed and occupancy together with additional latent variables for regime switching and degeneracy (recovery) rate.

The rest of the paper is outlined as follows. Section 2 develops our state-space model for tracking traffic flows speed. We discuss several ways to model a dynamic switching state variable that models flow regimes. Section 3 describes applications to special events and severe weather conditions. Section 4 develops a particle filtering and learning algorithm for inference and prediction that allows on-line inference. Section 5 illustrates our methodology with study of traffic flow on Chicago’s I-55. Finally, Section 6 concludes with directions for future research.
2 Bayesian Modeling of Traffic Flow Speed

Traffic flow speed data often relies on sparse and noisy measurements. Sparseness occurs with a fixed grid of sensors or dynamically changing data source such as are GPS probes. Our model is designed to be applicable in both scenarios. There are discontinuities in the traffic flow dynamics which need to be accounted for. We build a multi-process Dynamic Linear Model (DLM) [West and Harrison (1997); Carlin et al. (1992)] for the traffic state during three regimes: free flow, breakdown and recovery.

For demonstration purposes, we use data from sensor with id N-6041. The sensor is located eight miles from the Chicago downtown on I-55 north bound (near Cicero Ave), which is part of a route used by many morning commuters to travel from southwest suburbs to the city. Figure 1 illustrates a typical day traffic flow pattern on Chicago’s I-55 highway where sudden breakdowns followed by a recovery to the free flow regime. This traffic pattern is recurrent and very similar to one observed on any other work day. We can see a breakdown in flow speed during the morning peak (region 2) period followed by speed recovery (region 3). The free flow regimes (regions 1, 4, and 5 on the plot) are usually of little interest to traffic managers. This data motivates the choice of the statistical model developed in this section. Our goal is to build a statistical model that is capable of capturing the sudden regime changes such as change from free flow to congestion at the beginning of the morning rush hour (regions 1 and 2) and change decline in speed to the recovery regime at the end of the rash hour (regions 2 and 3).
A key feature of our modeling strategy is the inclusion of a switching state variable $\alpha_{t+1} \in \{-1, 0, 1\}$ to identify different flow regimes. Trends during the break-down periods and recovery periods can be modeled using the first order polynomial component.

Specifically, we use a state-space model of the form

Observation: $y_{t+1} = Hx_{t+1} + \gamma^Tz_{t+1} + v_{t+1}, \; v_{t+1} \sim N(0, V_{t+1})$ \hspace{1cm} (1)

Evolution: $x_{t+1} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} F_{\alpha_{t+1}} & \alpha_{t+1} \\ 0 & 1 \end{pmatrix} x_t + w_{t+1}, \; w_{t+1} \sim N(0, W_{t+1})$ \hspace{1cm} (2)

Switching Evolution: $\alpha_{t+1} \sim p(\alpha_{t+1}|\alpha_t, Z_t)$ \hspace{1cm} (3)

Our hidden state variable $x_t = (\theta_t, \beta_t)^T$, where $\theta_t$ is traffic flow speed and $\beta_t$ is rate of change. Measurements are given by $y_t = (speed, count, density)$, and we incorporate a switching variable with tree states $\alpha_t \in \{breakdown, free\; flow,\; recovery\}$. The observation matrix $H$ allows for partial observation of the state vector $x_t$.

The parameter $\mu$ measures the average free flow speed. We allow for the possibility of regressors, $z_{t+1}$, in the observation equation which effect the sensor measurement model, $\gamma$.
are regressors parameters. The switching coefficient $F_{\alpha_t}$ is defined by

$$F_{\alpha_t} = \begin{cases} 
1, & \alpha_t \in \{1, -1\} \\
F < 1, & \alpha_t = 0.
\end{cases}$$

The dynamics $p(\alpha_{t+1}|\alpha_t, Z_t)$ of the switching evolution depends on the set of variables $Z_t$, where $Z_t$ is the set of variables that explain regime shifts. Several approaches to model switching process $p(\alpha_{t+1}|\alpha_t, Z_t)$ and choosing $Z_t$ are described in Section 2.1.

### 2.1 Modeling $p(\alpha_{t+1}|\alpha_t, Z_t)$

The discrete state $\alpha_{t+1} \in \{-1, 0, 1\}$ models breakdown ($\alpha = -1$), free flow ($\alpha = 0$), and recovery ($\alpha = 1$) regimes of a traffic system. We need to specify a transition kernel that describes the evolution of this state given a set of exogenous predictor variables, denoted by $Z_t$, and current state $\alpha_t$. Given $Z_t$, the set of probabilities, $p(\alpha_{t+1}|\alpha_t, Z_t)$, will then form a $3 \times 3$ matrix which will be combined with the evolution of the hidden state vector $x_t$.

Most traffic breaks downs at a given location happen more or less the same time during morning or evening peak periods on a work day. It is therefore natural to introduce an exogenous variable $Z_t = (\text{period}, \text{day of week})$, $\text{period} \in \{1, 2, 3\}$ where the three periods correspond to morning, evening peak period and rest of the day. Incorporating additional considerations beyond period of the day and day of week, such as weather, month, spatial event leads to

$$Z_t = (\text{day of week}, \text{month}, \text{weather}, \text{month}, \text{event}, \text{accident})^T.$$

One way to build such a model is to identify $3 \times 3$ transition matrix $p(\alpha_{t+1}|\alpha_t)$ for each combination of the parameters in $Z_t$ based on the historic observations.

Figure 2(a) shows the average speed for a fixed location on the network for the April 4 - March 3 period in 2009. Weekend days identified by blue dots. We see that the average speed on weekend is very close to the free flow speed, roughly 63 mi/h, which means there is no congestion on those days. On the other hand average speed on a work day is significantly lower as a result if congestion during rush hour.

Figure 2(b) shows a scatter plot of measured flow speed for all non-holiday Wednesdays in 2009. The measurements taken every five minutes. We can see that congestion starts
roughly at the same time at around six in the morning and lasts roughly for three hours. The breakdown in the evening happens rarely.

Figure 2: Recurrent traffic flows. The left panel (a) shows average speed as measured by the sensor N-6041 for each day during the April 4 - March 3 of 2009 time period. Weekends are marked by blue dot. The right panel (b) shows raw speed measurements from the sensor N-6041 for each five minute interval of every Wednesday in 2009.

Another observation is that traffic congestion is a recurrent event and is very similar from one week to another. Figure 3 shows the recurrent traffic conditions grouped by the day of the week. This observation leads us to another approach to choose $Z_t$ In particular, approach based on non-parametric regression that uses historical traffic flow data. For example, Smith et al. (2002) and Chiou et al. (2013) showed that the last three measured speed values, perform very well for predicting traffic flows. In this case we can write $Z_t$ as

$$Z_t = (\alpha_t, \hat{x}_t, \hat{x}_{t-1}, \hat{x}_{t-2}, \text{time of day})^T,$$

where $\hat{x}_t$ is the filtered value of state vector $x_t$. Then the transition dynamics $p(\alpha_{t+1}|\alpha_t Z_t)$ is generated by computing a weighted average of those points from a historic database that fall within a neighborhood of $Z_t$. To calculate the neighborhood, we calculate distance between $Z_t$ and each of the point from the data base, and then choose $k$ points with smallest distance. The weights are proportional to the distance.
Figure 3: Daily Traffic Patterns, measured in 2009. Each plot shows raw speed measurements averaged over five minute intervals from the sensor N-6041 for a given day of the week, with holidays and days with erroneous measurements removed from data.

3 Tracking Traffic Flows during Special and Weather Events

A useful feature of our model is that it can easily handle special events or severe weather that can upset a typical traffic patterns and thus need to be accounted for to properly estimate and forecast traffic conditions on those days. To show the empirical effect, Figure 4 compares the expected travel time (red line), which is calculated based on historical data for the last 150 days, with the travel time on a snow day (green line), for December 11, 2013. There were 1.8 inches of snow on this day, with snow starting at midnight and continuing till noon. There were no traffic accidents on this road segment on this day. As we can see, even a light snow in a region, where drivers are used to drive during snow days can cause major delays. The yellow regions are 70% confidence interval based on historical data.
Figure 4: Impact of light snow on travel times on I-55 near Chicago on December 11, 2013. Both plots show travel time on a 27 mile stretch of highway I-55 between I-355 and I-94 in both north and south directions. The north bound direction is from southwest suburbs to the city. The red line is an travel time averaged over previous 150 days, the yellow area show 70% confidence interval for the data and the green line is the travel time on the day of the event.

Special events is another potential cause of unusual traffic conditions. Figure 5 shows impacts of special events on travel times on interstate I-55 north bound (towards the city). The weekday football game, which takes place at Soldier Field stadium in Chicago downtown, combined with typical commute traffic has very significant impact on travel times. Weekend special events have relatively minor negative impact. On the other hand, the NATO summit that was hold in Chicago’s McCormick Place located slightly south of downtown, on Monday had positive impact on travel times. This can be explained by regular commuters, who knew about the event, changing there departure times, using commuter rail or simply working from home on this day.
Figure 5: Impact of special events on I-55 north bound travel times. All plots show travel time on a 27 mile stretch of highway I-55 between I-355 and I-94 in both north (towards the city) and south directions. All plots compare travel times on the day of the special event (red line) with the travel time averaged over the previous 150 days (blue line).

4 Building a Particle Filter

Given our state-space mixture model developed above, the goal is to provide an algorithm for finding, in an on-line fashion, the set of joint filtered posterior distributions \( p(x_t, \alpha_t | y_{1:t}) \) for both \( x_t, \alpha_t \), the traffic flow state and switching variable at each time point. The advantage of particle filtering is that we simply re-sample from the predictive distribution the current posterior and then propagate to the next set of particles to generate the approximation to the conditional posterior update. Appendix A provides a review of particle filtering and learning methods.
From a distributional viewpoint, we can re-write our model (1) - (2) as a hierarchical model

\( (y_{t+1}|x_{t+1}) \sim N(Hx_{t+1} + z_t, V_{t+1}) \),

\( (x_{t+1}|x_t, \alpha_{t+1}) \sim N(G_{\alpha_{t+1}}x_t + (I - G_{\alpha_{t+1}})\mu, W_{t+1}) \),

where the evolution matrix is given by

\[
G_{\alpha_{t+1}} = \begin{pmatrix}
F_{\alpha_{t+1}} & \alpha_{t+1} \\
0 & 1
\end{pmatrix}.
\]

Now, marginalizing out \( x_{t+1} \), this implies a marginal distribution is given by

\( (y_{t+1}|x_t, \alpha_{t+1}) \sim N(HG_{\alpha_{t+1}}x_t + H(I - G_{\alpha_{t+1}})\mu, H^TW_{t+1}H + V_{t+1}) \)

Here \( V_{t+1} \) and \( W_{t+1} \) are variance-covariance matrices.

We can further marginalize out \( \alpha_{t+1} \) using the transition kernel \( p(\alpha_{t+1}|\alpha_t, Z_t) \). This leads to 3-component mixture predictive model of the form

\[
p(y_{t+1}|x_t, \alpha_t) = \sum_{\alpha_{t+1} \in \{0, 1, -1\}} \phi(y_{t+1} | HG_{\alpha_{t+1}}x_t + H(I - G_{\alpha_{t+1}})\mu, H^TW_{t+1}H + V_{t+1}) p(\alpha_{t+1}|\alpha_t, Z_t)
\]

where \( \phi(x|\mu, C) \) denotes a normal density with mean \( \mu \) and covariance \( C \)

\[
\phi(x|\mu, C) = \frac{1}{(2\pi)^\frac{d}{2} |C|^\frac{1}{2}} \exp \left( -\frac{1}{2}(x - \mu)^T C^{-1}(x - \mu) \right).
\]

Our model, therefore allows for heavy-tails and non-Gaussianity in the traffic flow evolution.

We now show how this can be used to implement a particle filter and learning algorithm and track the filtered posterior distributions of the hidden state \( p(x_t|y_{1:t}) \) over time as new data \( y_{t+1} \) arrives. The goal is to find the next filtered posterior \( p(x_{t+1}|y_{1:t+1}) \), which is obtained from the marginal of the joint posterior \( p(x_{t+1}, \alpha_{t+1}|y_{1:t+1}) \).

Suppose that we are currently at time \( t \). We assume a particle approximation for the joint posterior of \( (x_t, \alpha_t) \) of the form

\[
p^N(x_t, \alpha_t|y^t) = \sum_{i=1}^N w_t^{(i)} N(m_t^{(i)}, C_t^{(i)}) \delta_{\alpha_t^{(i)}}(\alpha_t)
\]
where $\delta_\alpha(\cdot)$ is a Dirac measure and $w_i^{(i)}$ is a set of particle weights, which we will provide recursive updates for. We denote the Kalman moments by $s_t = (m_t, C_t)$ which form a set of conditional sufficient statistics for the state. We describe the recursions later.

To assimilate the next measurement, we need to find an updated posterior for $(x_{t+1}, \alpha_{t+1})$, with approximate weights $w_{t+1}^{(i)}$ and particles $(x_{t+1}, \alpha_{t+1})$ of the form

$$p^N(x_{t+1}, \alpha_{t+1}|y_{t+1}) = \sum_{i=1}^{N} w_{t+1}^{(i)} N(m_{t+1}^{(i)}, C_{t+1}^{(i)}) \delta_{\alpha_{t+1}^{(i)}}(\alpha_{t+1}),$$

where $y_{t+1} = (y_1, ..., y_{t+1})$. Our weights will be updated using the predictive likelihood, $p(y_{t+1}|s_{t}^{(i)})$, which are re-normalized. We aim to provide a posterior with the same number of particles $N$ with mixture weights of the form

$$w_{t+1}^{(i)} = \frac{w_{t}^{(i)} p(y_{t+1}|(\alpha_{t}, s_{t})^{(i)})}{\sum_{i=1}^{N} w_{t}^{(i)} p(y_{t+1}|(\alpha_{t}, s_{t})^{(i)})}.$$

We now provide recursive updates.

### 4.1 Recursive Updating

At time zero, we set an initial state distribution $p(x_0, \alpha_0)$, which has a form of a mixture multivariate normal

$$p(x_t|\alpha_t) = \sum_{i=1}^{N} w_{\alpha_t}^{(i)} \phi(x_t, \mu_{\alpha_t}^{(i)}, C_{\alpha_t}^{(i)})$$

We take $p(\alpha_0)$ as a discrete uniform distribution.

Conditional on the $\alpha_t$, the Kalman filter recursions, imply that filtered posterior distribution at time $t$ is also mixture multivariate normal, i.e

$$p^N(x_t|\alpha_{1:t}, y_{1:t}) = \sum_{i=1}^{N} w_{t}^{(i)} \phi(x_t, \mu_{t}^{(i)}, C_{t}^{(i)})$$

where $(m_t, C_t)^{(i)}$ are functions of the whole path $\{\alpha_{1:t}, y_{1:t}\}$.

To implement this algorithm, first compute the predictive likelihood of the next observation $y_{t+1}$ given $\alpha_{t+1}, s_t^x$, where $s_t^x$ is the sufficient statistics, we have

$$p(y_{t+1}|\alpha_{t+1}, s_t^x) = \int p(y_{t+1}|\alpha_{t+1}, x_{t+1}) p(x_{t+1}|s_t^x, \alpha_{t+1}) dx_{t+1}$$

$$= \int \phi(y_{t+1}, Hx_{t+1}, V) \phi(x_{t+1}, G_{\alpha_{t+1}}^\mu_t + (1 - |\alpha_{t+1}|)\mu, G_{\alpha_{t+1}}^CG_{\alpha_{t+1}}^T + W) dx_{t+1}$$

$$= \phi(y_{t+1}, H(G_{\alpha_{t+1}}^\mu_t(|\alpha_t| - 1)\mu, V + H^T(G_{\alpha_{t+1}}^CG_{\alpha_{t+1}}^T + W_{\alpha_{t+1}}) H).$$
The inner predictive distribution is given by
\[
p(y_{t+1}|\alpha_{t+1}, s_t^x) = \int_{\mathbb{R}^n} p(y_{t+1}|x_{t+1}) p(x_{t+1}|\alpha_{t+1}, s_t^x) dx_{t+1}
\]
where \(p(x_{t+1}|\alpha_{t+1}, s_t^x) = \phi(x_{t+1}, G_{\alpha_{t+1}} \mu_t + (I - G_{\alpha_{t+1}}) \mu, G_{\alpha_{t+1}} C_t G_{\alpha_{t+1}}^T + W_{\alpha_{t+1}})\).

Given \((\alpha_{t+1}, y_{t+1})\), we need to update \(s_t = (\mu_t, C_t)^T\), where we suppress the index \(i\) for clarity.

These updates are given by Kalman recursion operator, \(\mathcal{K}\), which is given by
\[
\mu_t^f = G_{\alpha_{t+1}} \mu_t + (I - G_{\alpha_{t+1}}) \mu, \quad C_t^f = G_{\alpha_{t+1}} C_t G_{\alpha_{t+1}}^T + W_{\alpha_{t+1}}
\]
\[
\mu_t = \mu_t^f + K_t (y_t - H \mu_t^f), \quad C_t = (I - K_t H_t) C_t^f
\]
with Kalman gain matrix \(K_t = C_t^f H_t (H C_t^f H_t + V)^{-1}\).

Now we are in a position to find the predictive density in equation (4), namely
\[
p(y_{t+1}|\alpha_t, s_t^x) = \sum_{\alpha_{t+1} \in \{0, 1, -1\}} p(y_{t+1}|\alpha_{t+1}, s_t^x) p(\alpha_{t+1}|\alpha_t)
\]
This is a 3-component model.

We now have the following efficient Rao-Blackwellised particle filter

**Algorithm. Bayesian Particle Filtering for traffic flows:**

*Step 1* (Draw) an index \(k_t(i) \sim Multi(w_{t+1})\)-distribution with weights
\[
w_{t+1}^{(i)} = p(y_{t+1}|s_t^{(i)}) / \sum_{i=1}^N p(y_{t+1}|s_t^{(i)}). \tag{4}
\]

*Step 2* (Propagate) switching state \(\alpha_{t+1}^{(i)} \sim p(\alpha_{t+1}|k_t^{(i)})\)

*Step 3* (Propagate) sufficient statistics \(s_t^{k_t^{(i)}}\) using assimilated data and Kalman filter recursion
\[
s_t^{k_t^{(i)}} = \mathcal{K}\left(s_t^{k_t^{(i)}}, \alpha_{t+1}^{(i)}, y_{t+1}\right) \tag{5}
\]

The weights are updated according to the following rule
\[
w_{t+1}^{(i)} = \frac{w_t^{(i)} p(\alpha_{t+1}|\alpha_t^{(i)})}{\sum_{i=1}^N w_t^{(i)} p(\alpha_{t+1}|\alpha_t^{(i)})}
\]
Finally, we draw new state vector \(x_{t+1}\) from its mixture multivariate normal distribution.
5 Tracking Traffic Flow on Interstate I-55

5.1 Dataset Description

The data was provided by the Lake Michigan Interstate Gateway Alliance (http://www.travelmidwest.com/) formally Gary-Chicago-Milwaukee Corridor (GCM). The data is measurements from the loop-detector sensors installed on interstate highways. A loop detector is a very simple presence sensor that senses when a vehicle is on top of it and generates an on/off signal. There are slightly more than 900 loop-detector sensors that cover a large portion of the Chicago metropolitan area. Every 5 minutes a report for each of the loop detector sensors is recorded. The report contains averaged speed, count, and occupancy.

5.2 Numerical Experiments

Consider a single road segment. We use a measured data for a 24 hour period taken on a week day. The segment we consider is a part of highway I-55 north bound. This part of the highway is heavily congested during the morning rush hour, mostly due to commuters, who travel from south-west suburbs to the central business district of Chicago. There were no special events on that day and the weather was clear, thus a very similar congestion pattern can be observed on any “usual” work day on this road segment. The measurements are made by a single loop detector. For this example we calculate the filtering distribution for the travel speed, flow regime, and rate of change variables on this segment. The time series of measured speeds is of length $N = 288 \ (24 \times 12)$.

The state $x_t = (\theta_t, \beta_t)^T \in \mathbb{R}^2$ is a true travel speed and associated trend coefficient. The parameters of the observation and evolution models are set to:

$$H = (1 \ 0), \ V_t = 4, \ F_0 = 0.5, \ W_t = \begin{pmatrix} 1.9 & 0 \\ 0 & 4.5 \end{pmatrix}.$$  

In our simulation study, we define a transition matrix $P$ with equilibrium probabilities $\pi P = \pi$ that calibrate well to the three states in practice. Thus, we construct a Markov-
switching process with the following probability transition matrix

\[ P_{\alpha_t} = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.15 & 0.7 & 0.15 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}, \]

where \( \alpha_t \in \{\text{breakdown, free flow, recovery}\} \). The \( F_0 \) was fitted using the data from the times when the traffic is stationary, and \( W_t \) was fitted using the data from both the stationary and non-stationary regimes.

Figure 6 shows the filtered speed and its quantiles, along with the measured data. We see that the filtered state curve more-or-less follows the measurement curve. The evolution model proposed in this paper is very general and allows large changes in the speed state. Thus, the jittering behavior of measurement get mostly explained by statistical model and does not get filtered out.

![Figure 6: Filtered traffic speed given loop detector measurements.](image)

Figure 6 shows the filtered probability of \( \alpha \) for each of the values (0,1,-1). We can see that algorithm accurately captures the changes in the flow regime and assigns high probability to the free flow regime before the morning peak and the shifts probability to the breakdown regime.

![Figure 7: Filtered probability of \( \alpha \) for each of the values (0,1,-1).](image)
Figure 7: Measured and filtered traffic flow speed with quantiles

Figure 8 shows the filtered values for rate of change of speed during recovery and breakdown regimes. The algorithm captures all of the changes in traffic flow change rates. The algorithms captures “fast” breakdown little after 6am and “slow” breakdown at around 10am. It also, captures, for example, the recovery between 2pm and 4pm.

Figure 8: Rate of change of traffic flow

The results of numerical experiments illustrate the following features. The filtered speed plot follows the measurement plot, which is an expected results. We have chosen a well-
behaving sensor for the study and there are no outliers in the measurements. The filtering algorithm properly identifies rate of change during break down and recovery regime. We can see that breakdown happens faster. This is a well known fact, that it takes less time for a queue to build up than to dissipate. This difference in time can be explained by driver’s behavior and the fact that vehicles acceleration rate is lower than deceleration rate. However, the results for filtered probabilities of the switching variable $\alpha_t$ are less intuitive. Filter properly identifies free flow regime in the morning, but then gets “slightly confused” during morning rush hour by assigning very close probabilities to recovery and breakdown regimes. We interpret this as saying that the Markov switching process model used to model $P(\alpha_t|\alpha_{t+1}, Z_t)$ is misspecified and needs to be refined. This is a topic for future research.

6 Discussion

This paper proposes a mixture state-space model together with an particle filter and learning for tracking the state of traffic flow and other hidden variables such as flow regime and rate of change in the flow speed. The proposed method is very flexible in a sense that it does not require the state-space model to be Gaussian. Our approach does not rely on blind particle proposal for estimating the forecast distribution, instead draws are taken from a smoothed distribution, that takes the measurement into consideration. Thus, our filter is fully adapted with exact samples from the filtering distribution are drawn. We used the sufficient statistics representation for state particles lead to a computationally efficient method. We formulate the traffic flow evolution equation as a hierarchal Bayesian model that is capable of capturing traffic flow discontinuities.

Although we have focused on representation of state-space dynamics using multi-process DLM, the same approach will work for the kinematic wave theory based approach, when the evolution equation is given by the classical macroscopic traffic flow model, namely Lighthill-Whitham-Richards (LWR) partial differential equation \cite{Lighthill:55, Richards:56}. A Bayesian analysis of traffic flows that using LWR model is described in \cite{Polson:14}. The analysis based on LWR model requires a estimation of boundary conditions for a road segment which is not always available. The
statistical model described here has the flexibility to avoid assumptions on locations of sensors.

While we demonstrated an efficient solution to the filtering problem, the fact that we provided exact sampling from also allows us to use the same approach for particle learning and smoothing Carvalho et al. (2010). It works by augmenting the particle space to \( \{x_t, \phi_t\} \), where \( \phi_t \) is the sufficient statistic for the parameter space \( \gamma \), i.e. \( p(\gamma|x_t, y_t) = p(\gamma|\phi_t) \). In the DLM setting we can learn \( \beta \), by removing the evolution equation for this parameter and rather learning it directly from data, without relying on any model. Further research into predictive performance of such models is warranted.

### A Particle Filtering and Learning Methods

Particle filtering and learning methods are designed to provide state and parameter inference via the set of joint posterior filtering distribution obtained in an on-line fashion Gordon et al. (1993), Carpenter et al. (1999), Pitt and Shephard (1999), Storvik (2002), Carvalho et al. (2010).

Let \( y_t \) denote the data, and \( \theta_t \) the state variable. Let \( \phi \) denote the unknown parameters. For the moment, we suppress the conditioning on the parameters \( \phi \). We will show how to update state variables and sufficient statistics for \( \phi \). First, we factorize the joint posterior distribution of the data and state variables both ways as

\[
p(y_{t+1}, \theta_{t+1}|\theta_t) = p(y_{t+1}|\theta_{t+1})p(\theta_{t+1}|\theta_t) = p(y_{t+1}|\theta_t)\, p(\theta_{t+1}|\theta_t, y_{t+1})
\]

The goal is to obtain the new filtering distribution \( p(\theta_{t+1}|y_{t+1}) \) from the current \( p(\theta_t|y_t) \). A particle representation of the previous filtering distribution is a random histogram of draws. It is denoted by

\[
p^N(\theta_t|y_t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\theta^{(i)}_t}.
\]

where \( \delta \) is a Dirac measure. As the number of particles increases \( N \to \infty \) the law of large numbers guarantees that this distribution converges to the filtered distribution \( p(\theta_t|y_t) \).
In order to provide random draws of the next distribution, we first resample \( \theta_t \)'s using the smoothing distribution
\[
p(\theta_t|y^{t+1}) \propto p(y_{t+1}|\theta_t)p(\theta_t|y^t)
\]
obtained by Bayes rule. Thus, we draw \( \theta^{k(i)}_t \) by drawing the index \( k(i) \) from a multinomial distribution with weights
\[
w^{(i)}_t = \frac{p(y_{t+1}|\theta^{(i)}_t)}{\sum_{j=1}^N p(y_{t+1}|\theta^{(j)}_t)}.
\]
We set \( \theta^{(i)}_t = \theta^{k(i)}_t \) and “propagate” to the next time \( t+1 \) using
\[
p(\theta_{t+1}|y_{1:t+1}) = \int p(\theta_{t+1}|\theta_t, y_{t+1})p(\theta_t|y_{1:t+1})d\theta_t.
\]
Given a particle approximation \( \{\theta^{(i)}: 1 \leq i \leq N\} \) to \( p^N(\theta_t|y^t) \), we can use Bayes rule to write
\[
p^N(\theta_{t+1}|y^{t+1}) \propto \sum_{i=1}^N p\left(y_{t+1}|\theta^{(i)}_t\right) \left(\theta_{t+1}|\theta^{(i)}_t, y_{t+1}\right)
\]
\[
= \sum_{i=1}^N w^{(i)}_t \left(\theta_{t+1}|\theta^{(i)}_t, y_{t+1}\right),
\]
where the particle weights are given by
\[
w^{(i)}_t = \frac{p(y_{t+1}|\theta^{(i)}_t)}{\sum_{i=1}^N p(y_{t+1}|\theta^{(i)}_t)}.
\]
This mixture distribution representation leads to a simple simulation approach for propagating particles to the next filtering distribution.

The algorithm consists of two steps:

**Step 1.** (Resample) Draw \( \theta^{(i)}_t \sim Mult_N \left(w^{(1)}_t, \ldots, w^{(N)}_t\right) \) for \( i = 1, \ldots, N \).

**Step 2.** (Propagate) Draw \( \theta^{(i)}_{t+1} \sim p\left(\theta_{t+1}|\theta^{(i)}_t, y_{t+1}\right) \) for \( i = 1, \ldots, N \).

To implement this algorithm, we need the predictive likelihood for the next observation, \( y_{t+1} \), given the current state variable \( \theta_t \). It is defined by
\[
p(y_{t+1}|\theta_t) = \int p(y_{t+1}|\theta_{t+1})p(\theta_{t+1}|\theta_t)d\theta_{t+1}.
\]
We also need the conditional posterior for the next states $\theta_{t+1}$ given $(\theta_t, y_{t+1})$. It is given by

$$p(\theta_{t+1}|\theta_t, y_{t+1}) \propto p(y_{t+1}|\theta_{t+1}) p(\theta_{t+1}|\theta_t).$$

This algorithm has several practical advantages. First, it does not suffer from the problem of particle degeneracy which plagues the standard sample-importance resample filtering algorithms. This effect is heightened when $y_{t+1}$ is an outlier. Second, it can easily be extended to incorporate sequential parameter learning. It is common to also require learning about other unknown static parameters, denoted by $\phi$. To do this, we assume that there exists a conditional sufficient statistic $s_t$ for $\phi$ at time $t$, namely

$$p(\phi|\theta_{1:t}, y_{1:t}) = p(\phi|s_t)$$

where $s_t = s(\theta_{1:t}, y_{1:t})$. Moreover, we can propagate these sufficient statistics by the deterministic recursion $s_{t+1} = S(s_t, \theta_{t+1}, y_{t+1})$. This will lead to efficient inference for all model parameters.

Given particles $(\theta_t, \phi, s_t)^{(i)}$, $i = 1, \ldots, N$. First, we resample $(\theta_t, \phi, s_t)^{(k(i))}$ with weights proportional to $p(y_{t+1}|(\theta_t, \phi)^{(k(i)})$. Then we propagate to the next filtering distribution $p(\theta_{t+1}|y_{1:t+1})$ by drawing $\theta_{t+1}^{(i)}$ from $p(\theta_{t+1}|\theta_t^{(i)}, \phi^{(i)}, y_{t+1}), i = 1, \ldots, N$. We next update the sufficient statistic for $i = 1, \ldots, N$,

$$s_{t+1} = S(s_t^{(k(i))}, \theta_{t+1}^{(i)}, y_{t+1}),$$

This represents a deterministic propagation. Parameter learning is completed by drawing $\phi^{(i)}$ using $p(\phi|s_{t+1}^{(i)})$ for $i = 1, \ldots, N$. We now track the state, $\theta_t$, and conditional sufficient statistics, $s_t$, which will be used to perform off-line learning for $\phi$.

The algorithm now consists of four steps:

Step 1. (Resample) Draw Index $k_t(i) \sim Mult_N \left(w_t^{(1)}, \ldots, w_t^{(N)}\right)$ for $i = 1, \ldots, N$

The weights are proportional to $p(y_{t+1}|(\theta_t, s_t)^{(i)})$

Step 2. (Propagate) Draw $\theta_{t+1}^{(i)} \sim p(\theta_{t+1}|(\theta_t, s_t)^{(k(i)}, y_{t+1})$ for $i = 1, \ldots, N$.

Step 3. (Update) Deterministic $s_{t+1}^{(i)} = S(s_t^{(k(i))}, \theta_{t+1}^{(i)}, y_{t+1})$ for $i = 1, \ldots, N$.

Step 4. (Learning) Offline $\phi^{(i)} \sim p(\phi|s_{t+1}^{(i)})$ for $i = 1, \ldots, N$. 

References

Anacleto, O., C. Queen, and C. J. Albers (2013). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62(2), 251–270.

Blandin, S., A. Couque, A. Bayen, and D. Work (2012). On sequential data assimilation for scalar macroscopic traffic flow models. *Physica D: Nonlinear Phenomena* 241(17), 1421–1440.

Carlin, B. P., N. G. Polson, and D. S. Stoffer (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association* 87(418), 493–500.

Carpenter, J., P. Clifford, and P. Fearnhead (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation* 146(1), 2–7.

Carvalho, C. M., M. S. Johannes, H. F. Lopes, and N. G. Polson (2010). Particle learning and smoothing. *Statistical Science* 25(1), 88–106.

Chiou, Y.-C., L. W. Lan, and C.-M. Tseng (2013). A novel method to predict traffic features based on rolling self-structured traffic patterns. *Journal of Intelligent Transportation Systems* (just-accepted).

Gazis, D. C. and C. H. Knapp (1971). On-line estimation of traffic densities from time-series of flow and speed data. *Transportation Science* 5(3), 283–301.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, Volume 140, pp. 107–113. IET.

Lighthill, M. J. and G. B. Whitham (1955). On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229(1178), 317–345.
Mihaylova, L., R. Boel, and A. Hegyi (2007). Freeway traffic estimation within particle filtering framework. *Automatica* 43(2), 290–300.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association* 94(446), 590–599.

Polson, N. and V. Sokolov (2014). Bayesian analysis of traffic flow on interstate i-55: The lwr model. *arXiv preprint arXiv:1409.6034*.

Rasschaert, G. (2003). Analysis of traffic data. Master’s thesis, Department of Electrical Energy, Systems and Automation, SYSTeMS, University of Ghent, Belgium.

Richards, P. I. (1956). Shock waves on the highway. *Operations research* 4(1), 42–51.

Schreiter, T., C. van Hinsbergen, F. Zuurbier, H. van Lint, and S. Hoogendoorn (2010). Data-model synchronization in extended kalman filters for accurate online traffic state estimation. In *2010 Proceedings of the Traffic Flow Theory Conference, Annecy, France*, Volume 86.

Smith, B. L., B. M. Williams, and R. Keith Oswald (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* 10(4), 303–321.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *Signal Processing, IEEE Transactions on* 50(2), 281–289.

Tebaldi, C. and M. West (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association* 93(442), 557–573.

Wang, Y. and M. Papageorgiou (2005). Real-time freeway traffic state estimation based on extended kalman filter: a general approach. *Transportation Research Part B: Methodological* 39(2), 141 – 167.

West, M. and J. Harrison (1997, February). *Bayesian Forecasting and Dynamic Models (Springer Series in Statistics)*. Springer-Verlag.
Westgate, B. S., D. B. Woodard, D. S. Matteson, S. G. Henderson, et al. (2013). Travel time estimation for ambulances using bayesian data augmentation. *The Annals of Applied Statistics* 7(2), 1139–1161.

Work, D. B., O.-P. Tossavainen, S. Blandin, A. M. Bayen, T. Iwuchukwu, and K. Tracton (2008). An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pp. 5062–5068. IEEE.