Parameter estimation for stochastic hybrid model applied to urban traffic flow estimation

Herman Y Sutarto [a*], René K. Boel [a] and Endra Joelianto [b]

[a] *SYSTeMS Research Group, EESA, University of Ghent, B-9052 Zwijnaarde, Belgium;*

E-mail : *Herman.Sutarto@ugent.be* *;* *Rene.Boel@ugent.be*

[b] *Instrumentation and Control Research Group, Bandung Institute of Technology,*

Bandung 40132, Indonesia. E-mail : *ejoel@tf.itb.ac.id*

Abstract

This paper proposes a novel data-based approach for estimating the parameters of a stochastic hybrid model describing the traffic flow in an urban traffic network with signalized intersections. The model represents the evolution of the traffic flow rate, measuring the number of vehicles passing a given location per time unit. This traffic flow rate is described in this paper using a mode-dependent first order autoregressive (AR) stochastic process. The parameters of the AR-process take different values depending on the mode of traffic operation - free flowing, congested, or faulty – making this a hybrid stochastic process. Mode switching occurs according to a first-order Markov chain, and hence we call this hybrid process a jump Markov process. This paper proposes an expectation-maximization (EM) technique for estimating the transition matrix of this Markovian mode process and the parameters of the AR models for each mode. The technique is applied in this paper to actual traffic flow data from the city of Jakarta, Indonesia. The model thus obtained is validated by using the smoothed inferences algorithms and an online particle filter (PF). We also develop an EM parameter estimation that, in combination with a time window shift technique, can be useful and practical for periodically updating the parameters of hybrid model leading to an adaptive traffic flow state estimator. The proposed parameter estimation technique can thus be used as part of an adaptive model-based filter for feedback control of traffic lights.

Keywords: Hybrid system, parameter estimation, particle filter, stochastic system, fault detection, urban traffic networks.

1

*) First author is currently with Institut Teknologi Harapan Bangsa, Bandung, Indonesia

1. Introduction

Urban traffic congestion is a problem that significantly affects many aspects of the quality of life. A more efficient use of the existing road infrastructure, using advanced traffic control strategies, can lead to reduced congestion, reduced emissions, reduced fuel consumption, and improved safety. The model based control strategies that are needed in order to achieve such an improvement depend strongly on the quality and the accuracy of the dynamic model of the system, on the availability of reliable online data and also on the ease of implementation of the control strategy. The model must describe the variability over time of the traffic flow, allowing model based estimation of the current state, fusing noisy traffic data from various traffic sensors. This estimation together with the model in turn allows probabilistic prediction of future traffic flows, so that control actions – selecting switching times of traffic lights in the application we have in mind – can properly anticipate future traffic flow. In this paper we use a stochastic hybrid model to effectively describe the evolution over time of the arrival and flow rates of vehicles in an urban traffic network as stochastic processes. We define the flow rate as the number of vehicles that pass a location per red/green cycle, divided by the length of this cycle (veh./sec). The random value of the continuous variables defining the flow rates evolve according to an autoregressive model (AR) with time varying parameters, while some other events like changes in the environment, the inflow from outside the area under consideration, or the occurrence of unexpected incidents, influence the parameters of the AR-flow rate model. This stochastic hybrid models represents uncertainties in the evolution of the traffic flow in probabilistic form, for time scales of the order several red/green cycles of the traffic light. This corresponds to the prediction horizon that should be taken into account for a real-time model-prediction based feedback controller of the traffic lights in an urban network.

A great deal of recent work has considered traffic flow state estimation for freeway traffic and for arterial roads [1,2,3,4]. Our focus on urban networks requires different models and different estimation procedures, due to the stop and go phenomenon resulting from the green/red switching. The traffic flow state estimation algorithms typically rely on the knowledge of a stochastic dynamic model that describes the evolution of the traffic flow variables. This model depends on the noise characteristics in between mode changes, as well as on the transition probabilities from one discrete mode of operation to another mode. We identify or estimate the parameters of this hybrid system model from data observed during the normal operation of the plant. Since the discrete event and discrete time dynamics of such models are coupled, partially observable and stochastic, this is a very challenging problem in the field of data driven estimation and control.

Previous work of the authors [5] uses a single ARX model and applies a particle filter and kernel smoothing method to estimate the parameters of the ARX model. However, the disadvantage of this model is that it fails to accurately classify the mode-dependent variance of the data which deteriorates the performance of the estimator. The disadvantage is due to the fact that the model consists of only one single ARX model, keeping the variance for the whole given measurement data set constant. Off course this assumption is quite unrealistic due to that the variance of the video-type sensor usually depends on the intensity of traffic flow, and to the influence of external factors on the traffic flow evolution. This is the another reason for developing the stochastic hybrid model which basically consists of a multiple model. It has also been shown in [6] that using mode-dependent models and identifying these modes leads to better estimation and prediction results for the traffic flow. By using different models in different modes of traffic behavior, we significantly improve the performance of the estimator in terms of the accuracy both in free-flow and congested flow regime.

The stochastic hybrid models for which we estimate the parameters can be used for estimating and predicting traffic flow rates, and thus also for predicting queue lengths at signalized intersections. The variability of the traffic flow during successive cycles of the traffic light must be estimated with sufficient accuracy in order to predict the expected queue-length resulting from control decisions. Therefore estimating parameters of the traffic flow model, the

topic of this paper, is a prerequisite for implementing a good real-time controllers of the traffic lights in an urban environment.

In traffic measurement data we have to deal with many types of noise and uncertainty. In the case study in section 5 we use traffic data obtained from a video-type sensor that detects vehicles passing several locations near two neighboring intersections in the capital city of Indonesia, Jakarta. Our past experience with this type of video data has indicated that there are a lot of missed detections (due to lack of detail in the image, or due to one vehicle obscuring another) as well as false detections (due to reflections, or other errors in the image processing). A model based state estimator is therefore necessary in order to filter the noise from the useful data, and to obtain a sufficiently accurate estimate of the flow rates at the different locations. Our stochastic hybrid model must also take into account errors due to incidents (like accidents) that may occur in the measured lanes. An incident may block traffic flow causing a traffic jam. In order to build a highly accurate estimate, a reliable fault detection system, and a good real-time description all of these phenomena should be considered in the dynamic models.

In this paper, we present a novel approach to parameter estimation for a stochastic hybrid model of traffic flow. Traffic flow is described by a continuous random variable $\beta(t)$ which expresses that the number of vehicles passing a given location in the urban network during the time interval $[t, t + dt)$ is $\beta(t).dt$ (in order to avoid working with large integers we fluidize our traffic variables, approximating integer numbers of vehicles by real number, see [7]). We observe the traffic flow $\beta_k = \beta(t_k)$ at an increasing sequence of discrete time t_k (the easiest case is $t_k = k.\Delta t$, but in our application these t_k's are successive switching times of traffic lights that may not operate with a fixed cycle). So we define as traffic flow the ratio $\beta_k = N_k \big/ (t_{k+1} - t_k)$ where N_k counts the number of vehicles that pass the given location in the interval $[t_k, t_{k+1})$. The fluid flow approximation implicitly assumes that the traffic flow remains constant during the interval $[t_k, t_{k+1})$, i.e. $\beta_k = \beta(t)$ for $t \in [t_k, t_{k+1})$. Of course this assumption that vehicles travel approximately at equal distances during the interval $[t_k, t_{k+1})$ is an approximation that is only acceptable for sufficiently small values of the time increments t_{k+1}-t_k, but it reduces the computational complexity of our algorithms a lot since we do not have to consider individual vehicles. Moreover in our application we look at flow rates averaged over a typical red or green cycle period of the traffic light, and traffic flow is fairly regular during such a period.

As long as the mode of operation of the traffic remains unchanged we make the very reasonable assumption that β_k evolves as a first order autoregressive (AR) process. The parameters of this AR-process however can take different values depending on the mode of traffic operation - free flowing, congested or faulty. Mode switching can be seen as a discrete event and we assume that these events occur after geometrically distributed random times, so that the modes are described by a first-order Markov chain. This hybrid stochastic process is called a jump Markov model. The transition probability matrix π of the mode switching system must also be estimated using the online data available from the video cameras.

The problem considered in this paper thus is this following :

Determine those parameters that maximize the likelihood of the proposed stochastic hybrid model, best describing the stochastic dynamic model of the traffic flow rates at the given location, using a sufficiently long sequence of data $\beta_k, k = 1, 2, ..., K$ of traffic flow measurements, observed at a that location and use the model thus obtained for estimating and predicting the traffic flow.

Note that these hybrid model parameters include both the parameters of the AR-process (including the noise variance) and the entries of the transition matrix π. In this paper, the parameter estimation is performed as an offline technique using a given batch of observations, a

window of time during which the model is valid. Based on those parameters, we apply the model to be used in estimating/ predicting the traffic flow for the next time-window. This time window shift technique can be useful and practical for periodically updating the parameters of hybrid model leading to an adaptive traffic flow state estimator.

The remainder of this paper is organized as follows. Section 2 explains the motivation for the stochastic hybrid model used in this paper. This jump Markov model representing the traffic flow rates via an autoregressive equation with mode dependent parameters is explained in detail in section 3. Section 4 describes the expectation maximization (EM) technique as a tool for estimating the parameters of this model while section 5 presents the experimental lay-out for the data used in our validation of the EM technique. Model validation is described in section 6. Finally, section 7 summarizes the findings of this paper and draws some conclusions on its practical usefulness.

2. Problem formulation

In this section, we concentrate on stochastic hybrid modelling of traffic flow along one particular approach route to a signalized intersection, and indicate how this model is useful in controlling the operation of a signalized intersection. This explains how the parameter estimation problem treated in this paper fits into the overall traffic control problem. We validate our traffic flow model, and the parameter estimation algorithm based on this model, using data collected at 2 neighboring intersections, called 313 and 314, in the traffic network of Jakarta, Indonesia as shown in Fig.1. At each of the location indicated in Figure 1 by a bar, identified by their names D_F_ij, a video camera tries to detect all passing vehicles using a computer based counting algorithm. While false detections and missed vehicles cause significant errors in the data, the output $y(t)$ of the sensors approximately counts the number of vehicles that cross an entrance location D_F_ij during one cycle of the traffic light.

The data set also lists the number of vehicles that cross a location per cycle of the local traffic light, as well as the duration of each cycle (between 120-200 seconds). This provides us with a data set of traffic flow rates averaged per cycle. By using measurement data obtained from video cameras at the locations at the intersections described above we build a model describing the dynamics of the traffic flow β(t), use the EM-parameter estimation technique to estimate the model parameters, and validate the model thus obtained by using smoothed inference for predicting future values of β(t). A good model should have the capability to cope with variation of urban traffic flow enabling queue length based traffic control. Indeed the evolution of the queue-length at approach route ℓ is given by

$$Q_\ell(t) = Q_\ell(t_0) + \int_{t_0}^{t} (\lambda_\ell(\tau) - \mu_\ell(\tau))\, d\tau \tag{1}$$

where $\lambda_\ell(t)\ (\mu_\ell(t))$ is the traffic flow rate (generically called β(t) earlier) of the arrival stream of vehicles (resp. the departure stream of vehicles) for a location along approach route ℓ, both flow rates measured in veh/sec, averaged over one measurement cycle corresponding usually to one cycle of the traffic lights. As mentioned above we use the fluid-flow approach, aggregating the number of vehicles during a cycle period, implicitly assuming that during the k-th cycle, between t_k and t_{k+1}, the flow rate remains constant, equal to λ_k, resp. μ_k. Using our hybrid model equation (1) therefore also allows us to obtain probabilistic predictions for the queue lengths.

Figure 1 shows that intersection I-313 has arrival flow λ_{L11} which is determined by summing the number of passing vehicles detected at sensor location 9,6,7 and 12 divided by the cycle length of the intersection I-314. The value of this the arrival flow rate as measured during the k-th cycle of the traffic lights is further on called λ_k. The departure rate for intersection 314, called μ_{L11}, is defined by counting the vehicles that pass the sensor location 8,9,10 at downstream intersection I-313, divided by the green period of lane L_{11} at intersection I-313 (the time delay corresponding to the travel time between intersection 313 and 314 must of course be taken into account when using the flow rate λ_{L11} as inflow rate at the downstream intersection I-

313). The value of this departure flow rate as measured during the k-th green period is further on denoted μ$_k$. Of course this is under the assumption that drivers of vehicles that pass the sensor locations 9,6,7 and 12 follow the traffic rules.

For each intersection the traffic cycle is divided into 2 separate periods, called the red and the green period. In each of these periods only some traffic flows are allowed to cross the intersection, so that the departure rate $\mu_{L11}(t) = 0$ whenever traffic in the North-South direction sees red at intersection 313. As shown in Figure 2 there are, for each intersection and for each cycle, two decision variables: $T_g = t_{2k+1} - t_{2k}$ (the green period) representing the time duration between switching times t_{2k} and t_{2k+1}, and $T_r = t_{2k+2} - t_{2k+1}$ (the red period between t_{2k+1} and t_{2k+2}). The traffic signal control problem determines for each cycle, a good value for these two decision variables, minimizing some cost function, typically depending on the expected evolution of the queue-length.

The design of a good traffic controller therefore requires that a good model is available describing the evolution of the arrival and departure traffic flows, λ_k (a non-negative random process defined such that $\lambda_k.(t_{2.k+2} - t_{2.k})$ vehicles enter the intersection during the k-th cycle, from t$_{2k}$ to t$_{2(k+1)}$) and μ_k (a non-negative random process such that μ$_k$.(t$_{2k+1}$ − t$_{2k}$) vehicles cross the intersection during the k-th green period from t$_{2k}$ till t$_{2k+1}$). Note that the departure flow rate takes the value 0 whenever the traffic light is red for the corresponding direction. The purpose of this paper is to identify the parameters of the stochastic hybrid model describing the evolution of λ$_k$ and of μ$_k$, k = 1,2,.... Our parameter estimation algorithm uses real data from sensors at the location where the model is to be applied.
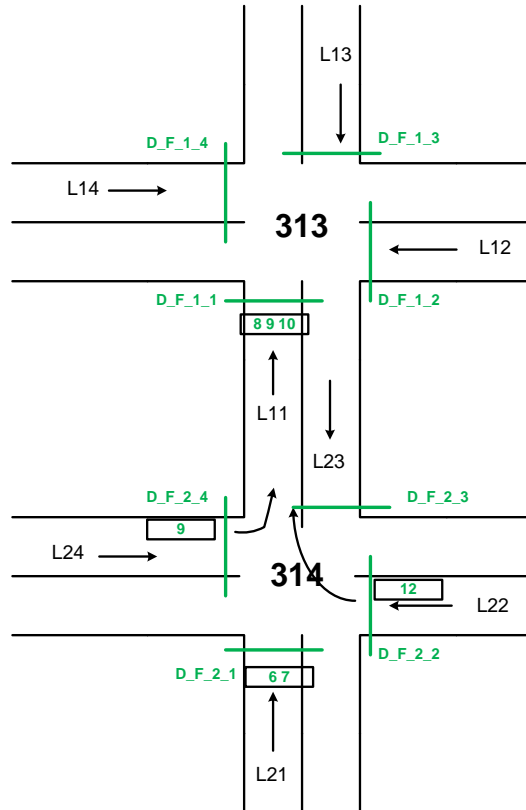

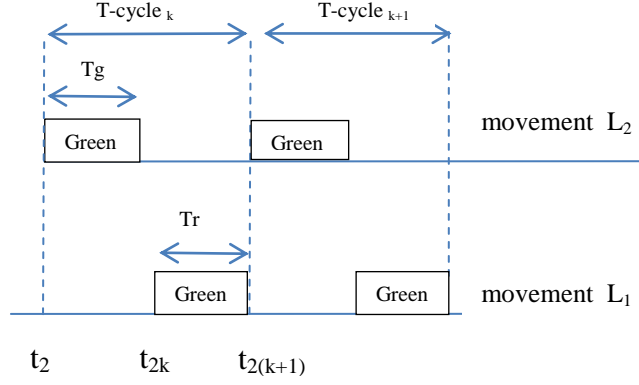Figure.1 Configuration of area under investigation

Figure 2. Signal Traffic Sequence

Consider 4 queues $\ell = 1,\ldots,4$ at 4 approach roads to an intersection (like I-313 or I-314 in fig. 1), and assume that the evolution of arrival rate $\lambda_{\ell,k}$ and departure rate $\mu_{\ell,k}$ is well defined by the parameters of the stochastic hybrid model identified with sufficient accuracy by the methods proposed in section 4. Further on we consider only one queue at a time, and hence we drop the subscript ℓ to simplify the notation. Define a 'clock' state variable $z_i(t) = [\, z_1(t), z_2(t)\,]$, associated with the GREEN light cycle for queues $\{1,3\}$ and for $\{2,4\}$ respectively; $z_1(t)$ measures the time since the last switch from RED to GREEN of the traffic light for queues $\{1,3\}$. It is reset to 0 as soon as the GREEN cycle length θ_1 is reached and remains at this value while the light is GREEN for queues $\{2,4\}$; as soon as the cycle ends, i.e., $z_2(t) = \theta_2$, then $\overset{\bullet}{z}_1(t)$ is reset to 1 and the process repeats. The same applies to $z_2(t)$, but for determining the evolution of queues $\{2,4\}$. This provides the control designer with a complete model for predicting the queue sizes $x_j(t)$ for the j-th queue :

$$\overset{\bullet}{x}_j(t) = \begin{cases} \lambda_j(t) & \text{if } z_1(t) = 0,\ j = 1,3 \text{ or} \\ & \qquad z_2(t) = 0,\ j = 2,4 \\ 0 & \text{if } x_j(t) = 0 \text{ and } \lambda_j(t) \le \mu_j(t) \\ \lambda_j(t) - \mu_j(t) & \text{otherwise} \end{cases} \tag{2}$$

The operation of the intersection can be modelled as a stochastic hybrid system as shown in Fig.3 with the time-driven dynamics described by equation (2) and event-driven dynamics dictated by GREEN-RED light switches and by events causing some $x_j(t)$ to switch from positive to zero or vice versa. We can describe the evolution of queue-length $x_j(t)$ for example by a Stochastic Hybrid Automaton as shown in Figure 3, equivalent to equation (2). The event set that affects the evolution of queue $x_j(t)$ is $E = \{e_1, e_2, e_3, e_4, e_5\}$ (dropping again subscripts j in order not to overload the notation), where : (a) e_1 corresponds to the value of $\lambda_j(t) - \mu_j(t)$ becoming strictly positive (after a period where it was non-positive), (b) e_2 corresponds to a switch of $\lambda_j(t)$ from 0 to strictly positive, (c) e_3 corresponds to the queue-length becoming empty, i.e., $x_j = 0$, (d) e_4 switches a traffic light from RED to GREEN, (e) e_5 switches a traffic light from GREEN to RED.
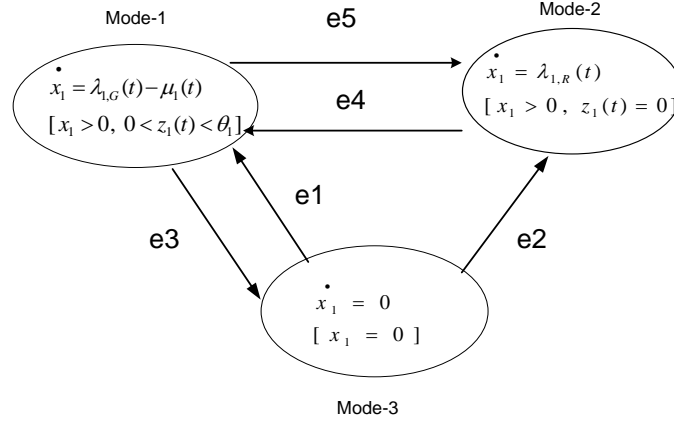
Figure 3. Stochastic Hybrid Automata

3. Jump Markov Model Structure

In this section we describe in detail the stochastic hybrid model, also called further on the jump Markov model (JMM), for a generic traffic flow rate β_k (which could represent the arrival flow rate λ_k or the departure flow rate μ_k). During a period of time when the traffic conditions, also called the mode s of operation of the traffic, remains unchanged, this variable β_k is modelled by a first-order autoregressive (AR) model:

$$\beta_{k+1} = \alpha(s) + \gamma(s)\beta_k + w_k \qquad (3)$$

(where $\alpha(s)$ and $\gamma(s)$ are mode-dependent parameters to be identified, and w_k, $k = 1, 2, ...$ is an independently identically distributed sequence of zero mean random variables with variance $\sigma^2(s)$, with $\sigma^2(s)$ also a mode dependent parameter to be identified). The fluid flow assumption of this paper implies that the traffic flow is constant during each interval $[t_k, t_{k+1})$: $\beta(t) = \beta_k$ for $t \in [t_k, t_{k+1})$. The traffic flow implication is that during any interval $[\theta, r) \subseteq t \in [t_{2k} - t_{2(k+1)})$ the number of vehicles that cross the location where this model is valid is on the average $\beta_k \cdot (r - \theta)$, and that these vehicles are approximately uniformly distributed over this interval.

From time to time the mode $s(t)$ of operation will change, due to external changes of the inflow rate, due to incidents that make the operation more or less efficient, or due to randomness. The mode variable s used above therefore should also be considered as a time varying random process. We assume that the mode changes only occur at times t_k when the value of the traffic flow rate β_k is updated by the AR equations. Therefore, we denote the mode as s_k in the interval $[t_k, t_{k+1})$. In the traffic model we consider 3 different values for the mode of operation $s_k \in \{1, 2, 3\}$:

- $s_k = 1$ denotes the desirable mode of operation where traffic is flowing normally without too much interference between successive vehicle;
- $s_k = 2$ denotes the congested mode where vehicles hinder each other significantly, and the system operates inefficiently;
- $s_k = 3$ corresponds to a faulty state, describing outliers in the behaviour that may be due to incidents (like accidents, bad weather conditions).

We assume that the mode process s_k can be modelled by a Markov chain, i.e. at each time t_k the mode variable $s_{k-1} = i$ changes randomly to the value $s_k = j$ with a probability $\pi_{ij} =$

Prob $(s_k = j | s_{k-1} = i, s_{k-2}, s_{k-3}, ...)$ which only depends on the most recent mode (or Markov state) s_{k-1}, not on states further in the past.

Equation (3) with its interpretation $\beta(t) = \beta_k$ for $t \in [t_k, t_{k+1})$, together with the Markov chain model for the mode s_k provides us with a complete mathematical model of traffic flow. The parameters of this model will be estimated in the next section according to the EM method. In total there are 15 parameters to be estimated.:

- For each mode $s \in \{1, 2, 3\}$ the AR model (3) has 3 parameters, $\alpha(s)$, $\gamma(s)$ and $\sigma^2(s)$

- The transition matrix $\Pi = (\pi_{ij})$ of the Markov chain describing the mode process has 3 rows of 3 elements, satisfying the normalization condition $\sum_{j=1,2,3} \pi_{ij} = 1, \forall i$

The observations, providing the data used in the parameter estimation described in section 5 and section 6, which are also the data that an online traffic controller could use, are obtained by counting the number of vehicles y_k passing the location to which the model is applied during the interval $t \in [t_k, t_{k+1})$. This output variables y_k is defined as :

$$y_k = \beta_k \cdot (t_{k+1} - t_k) + n_k$$

where n_k, $k = 1, 2, ...$ is a sequence of independent zero mean random variables with known distribution (describing the probabilities of missed vehicles and false detections in the video cameras). Note that the output y_k only becomes available at time $t_{2(k+1)}$. This is not a problem for our offline parameter estimation algorithm, but it should be taken into account for an online control algorithm.

4. Parameter estimation: EM approach

Estimation of the parameters ($(\alpha(s), \gamma(s), \sigma^2(s), s = 1, 2, 3)$ of the AR-models, and of the transition probabilities (π_{ij}, $i, j = 1, 2, 3$) of the JMM can be performed using an iterative two-step EM procedure. Therefore the EM algorithm allows us to completely identify the JMM stochastic hybrid model proposed in the previous sections as a model for traffic flow. In this section, we develop the application of the EM approach, originally proposed by [8] and further extended in [9,10], to switching systems as our JMM, based on forward-backward recursion or 'smoothing'. A good introduction and survey paper on EM can be found in [11]. This EM approach is formulated in batch or offline form, i.e. it uses a given number of observations obtained over a time interval [0,T) to iteratively find better and better estimates of the unknown parameters of a model that is valid over the period [0,T). This offline approach needs significant memory requirements and processing power for storing and processing large datasets, but this approach is shown to be useful and applicable further on in this paper.

In this paper we do reformulate the EM approach proposed by [9] in order to get a simpler and easier algorithm, changing the cost-function in the M-step by adding a weighted term with corresponding smoothed inferences about the currently active mode. The smoothed inferences are the probabilities $P\left(S(t) = i | \mathbf{y}_1^{t-1}; \theta\right)$ that the t-th observation comes from mode i given all the information y_1^{t-1} available at time t, where we use the notation $\mathbf{y}_1^{t-1} = (y_1, y_2, ..., y_{t-1})$. The reason for this modification is that each observation $y(t)$ belongs to the i-th mode with probability $P\left(S(t) = i | y(t); \theta\right)$. By using these smoothed inferences we find a better maximum likelihood estimate. If we transform our AR model to state-space (SS) form then we notice that the SS model can be classified as an observable canonical form (it can

be seen as an output feedback form) and the assumption that the mode $S(t)$ depends on past observation $\mathbf{y}_1^{t-1}$ only through the value of $S(t-1)$, implies that the calculation of the distribution $P\left(S(t)=i\big|\mathbf{y}_1^{t-1};\theta\right)$ is a problem of hidden state estimation of hybrid systems, where $\theta=\left\{\alpha_0,\varphi_0,\sigma_0^2,....,\alpha_i,\varphi_i,\sigma_i^2,\pi\right\}$. The key idea of this calculation is to perform a forward-backward or 'smoothing' filter recursion for each possible mode sequence.

The algorithm starts with an arbitrarily chosen vector of 15 independent parameters (see section 3), chosen randomly but using prior information in order to speed up convergence of the algorithm and to increase the probability of converging to the global maximum. The EM technique for hybrid systems requires a method for recursively approximating the conditional probability $P\left(S(t)=i\big|\mathbf{y}_1^{t-1};\theta\right)$, for each index m counting the number of EM iterations. Moreover an algorithm for finding the parameter values that achieves the maximum likelihood must be available. This differs from the approach proposed in [12] for a general non-linear system, where the objective is to find the best estimate of the conditional density $p_\theta\left(y(t)\,|\,\mathbf{y}_1^t\right)$. In the case of hybrid systems, as in this paper, the E-step, indirectly infers about the modes. Since the mode $S(t)$, is unobservable, only the expected values of the modes at successive time t, given the observation vector, can be calculated.

Since each measurement $y(t)$ evolves according to the i-th mode with probability $P\left(S(t)=i\big|\mathbf{y}_1^t;\theta\right)$ (given all the data $y(t)$, t = 1, 2...., T), it is reasonable if each component on the right-hand side of equation (4) below has to be weighted with the corresponding smoothed inferences (see Appendix for further details). Therefore, the weighted log-likelihood function is given by (assuming that the noise terms n_k have a normal distribution; similar formulae can be written down for other noise distributions):

$$L(\theta)=-\log\ell(\theta)=-\log\ell(\alpha_i,\varphi_i,\sigma_i,\pi)=-\sum_{t=2}^{T}P(S(t)=i\,|\,y_1^t;\theta^m)$$

$$*\left[\frac{\ln\left(\sqrt{2\pi}\sigma_i\right)+\left(y_t-\alpha_i-\varphi_i y_{t-1}\right)^2}{2\sigma_i^2}\right] \tag{4}$$

A detailed explanation of the EM-algorithm is given below:
The E-Step
The E-step consists of forward filtering and backward filtering/smoothing and this step aims to calculate the conditional probabilities $P\left(S(t)=i\big|y(t);\theta^m\right)$. Assume that θ^m is the parameter vector calculated in the M-step during the *m*-th iteration and $\rho_i^{(0)}\equiv P(S(1)=i)$. The algorithm starts with an arbitrarily (but sensibly, as explained above) chosen vector of initial parameters

$$\theta^0=\left\{\alpha_1^0,\varphi_1^0,\sigma_1^{2^0},\alpha_2^0,\varphi_2^0,\sigma_2^{2^0},\alpha_3^0,\varphi_3^0,\sigma_3^{2^0},\pi^0\right\}$$

i) **Step E-1:** *Forward Filtering* : for t = 1,2...., T iterate:

$$P\left(S(t)=i\big|\mathbf{y}_1^t;\theta^m\right)=\frac{P\left(S(t)=i\big|\mathbf{y}_1^{t-1};\theta^m\right)g_i\left(y(t)\,|\,S(t);\mathbf{y}_1^{t-1};\theta^m\right)}{\sum_i^3 P\left(S(t)=i\big|\mathbf{y}_1^{t-1};\theta^m\right)g_i\left(y(t)\,|\,S(t);\mathbf{y}_1^{t-1};\theta^m\right)} \tag{5}$$

where $\mathbf{y}_1^t=\left(y_1,y_2,...,y_t\right)$ is the vector of traffic flow measurements available at time t and $g_i\left(y(t)\,|\,S(t);\mathbf{y}_1^{t-1};\theta^m\right)$ is the conditional probability density function at of y(t) time t assuming that the mode *i* is active at time t:

$$g_i\left(y(t)\,|\,S(t);\mathbf{y}_1^{t-1};\theta^m\right)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{\left(y(t)-\alpha_i-\varphi_i y(t-1)\right)^2}{2\sigma_i^2}\right)$$

and
$$P\big(S(t+1)=i\big|\mathbf{y}_1^t;\theta^m\big)=\sum\nolimits_{j=1}^{3}\pi_{ji}^m\,P\big(S(t)=i\big|\mathbf{y}_1^t;\theta^m\big)$$

until $P\big(S(T)=i\big|y(T);\theta^m\big)$ is calculated.

The starting point for the iteration is chosen as : $P\big(S(1)=i\big|y(1);\theta^m\big)=\rho_i^m$ where ρ_i^m is the likelihood of being in mode i obtained in the previous iteration m

The Proof of Step E-1 is straightforward using Bayes' Rules.

ii) **Step E-2:** *Backward filtering* : for $t = T-1, T-2, \dots,$ 1 iterate on

$$P\big(S(t)=i\big|y(T);\theta^m\big)=\sum_{j=1}^{3}\frac{P\big(S(t)=i|\mathbf{y}_1^t;\theta^m\big)P\big(S(t+1)=j\mid y(T);\theta^m\big)\pi_{ij}^m}{P\big(S(t+1)=j\mid\mathbf{y}_1^t;\theta^m\big)}\tag{6}$$

See [9] for the proof (taking into account that the structure and terminology used there is completely different from what is used in this paper).

The M-Step

In the M-step of the EM algorithm, new and more exact maximum likelihood estimates θ^{m+1} for all model parameters are calculated. As mentioned previously each component of the log-likelihood function has to be weighted with the corresponding smoothed inference. In particular, for the model defined by equation (3) explicit formulas for the estimates can be derived by setting the partial derivatives of the log-likelihood function to zero and solving the resulting system of non-linear equations by a substitution technique:

$$\alpha_i=\frac{\sum_{t=2}^{T}P\big(S(t)=i\mid y(T),\theta^m\big)\big(y(t)-\varphi_i y(t-1)\big)}{\sum_{t=2}^{T}P\big(S(t)=i\mid y(T),\theta^m\big)}\tag{7}$$

$$\varphi_i=\frac{\sum_{t=2}^{T}[P\big(S(t)=i\mid y(T),\theta^m\big)y(t)y(t-1)]-\alpha_i[\sum_{t=2}^{T}P\big(S(t)=i\mid y(T),\theta^m\big)y(t-1)]}{\sum_{t=2}^{T}P\big(S(t)=i\mid y(T),\theta^m\big)y(t-1)^2}\tag{8}$$

$$\sigma_i^2=\frac{\sum_{t=2}^{T}[P\big(S(t)=i\mid y(T),\theta^m\big)\big(y(t)-\alpha_i-\varphi_i y(t-1)\big)^2]}{\sum_{t=2}^{T}P\big(S(t)=i\mid y(T),\theta^m\big)}\tag{9}$$

Because equations (7-8-9) are nonlinear, it is not possible to solve analytically for θ^m as a function of $\{y_1,y_2,..,y_T\}$. However, these equations do suggest an appealing iterative gradient ascent algorithm for finding the maximum likelihood estimate (or in practice a gradient descent for the loglikelihood since the relation between likelihood and log-likelihood is monotonely decreasing).

Iterating back from T to 1 we obtain $\rho_i^{m+1}=P\big(S(1)=i\big|\mathbf{y}_T;\theta^m\big)$ and the transition probabilities are estimated by using equation (10) (for a proof see [8]).The transition probabilities are restricted only by the condition that $\pi_{ij}\geq 0$ and $(\pi_{i1}+\pi_{i2}+\pi_{i3})=1$, $i=1,2,3$:

$$\pi_{ij}^{m+1} = \frac{\sum_{t=2}^{T} P\big(S(t) = j \mid y(T); \theta^m\big) \dfrac{\pi_{ij}^m P\big(S(t-1) = i \mid \mathbf{y}_1^t; \theta^m\big)}{P\big(S(t) = j \mid \mathbf{y}_1^{t-1}; \theta^m\big)}}{\sum_{t=2}^{T} P\big(S(t-1) = i \mid y(T); \theta^m\big)} \tag{10}$$

where π_{ij}^m is the transition probability obtained during the m-th iteration.

All values obtained in the M-step are then used as a new parameter vector $\theta^{m+1} = \big(\alpha_i^{m+1}, \varphi_i^{(m+1)}, \sigma_i^{2\,m+1}, \pi^{m+1}\big)$ and ρ_i^{m+1}, $i = 1, 2, 3$ in the next iteration of the E-step. The algorithm is terminated when $\big|\theta^m - \theta^{m-1}\big| \leq \delta$ for some preset accuracy δ. It means that one continues iterating in this fashion until the change between θ^m and θ^{m-1} is smaller than some specified convergence criterion. In the algorithm that we implemented we used the L_2 norm, but other norms could also be used.

The EM-algorithm can be implemented in 2 different forms, EM and GEM (Generalized EM) [8]. The EM algorithm maximizes the conditional expectation at every iteration, while the GEM only ensures that the likelihood increases at each step. Our implementation follows the GEM form. We avoid oscillations between different modes with identical likelihoods by updating the model only when the likelihood increases. The algorithm implementing this Generalized (G) EM-iteration procedure is shown in Table 1 by referring equations in both E-step and M-step above.

Table.1 : GEM Algorithm

1). **Initialization**	Set m=1. Select $\varepsilon \in \mathbf{R}$ arbitrarily small
	Initialize θ^1 to initial guess
2). **Expectation (E) Step**	Set m = m+1
	Given θ^m calculate $L\big(\theta^m\big)$ using equation (4), (5) and (6)
3).**Maximization (M) Step**	Set θ^{m+1} to value of θ by using equation (7), (8), (9) and (10) that maximize $L\big(\theta^{m+1}\big)$
4). **Convergence check**	Evaluate $L\big(\theta^{m+1}\big)$. If $\mid L\big(\theta^{m+1}\big) - L\big(\theta^m\big)\mid < \varepsilon,$, stop. Otherwise go to 2.

Related to the algorithm, the two main problems with the its implementation are:

(1) the conditional expectation is difficult to compute, and therefore we replace it by smoothed inferences. Thus we use the whole data (batch) obtaining more accurate results, especially for the estimate of the mode of operation at each time t. This the reason why in E-step we need the backward filtering as shown in equation (6).

(2) since the convergence may be to a local maximum that is not the global maximum, a good choice of initial conditions is necessary. In our traffic case, we simply choose the initial conditions based on intuitive guesses corresponding to similar historical real measurement data.

Since we use a stochastic model with Gaussian noise, the likelihood functions are bounded and the Hessian is always negative definite [20]. Therefore, the sequence converges to a local maximum (of the likelihood, a local minimum of the log-likelihood), proving convergence of our parameter estimation, assuming identifiability of the model (see [19,20]). In practice we found that the urban traffic model propose in this paper always leads to an identifiable system.

5. Experimental layout

This section focuses on modeling urban traffic flow, estimating parameter values based on the GEM-algorithm of section 4, using the data over a time window [0,T] for the experiment

layout shown in Figure 1. The time-window size is important to define the model that will be used for estimating traffic flow for the next time-window.

In this section, we consider as an example the arrival flow and departure flow λ_{L11} and μ_{L11} at the intersection I-313, and apply of it the parameter estimation techniques of section 4, using the data obtained during one single day by the video cameras that are part of a currently implemented SCATS traffic control system. The data obtained from the sensors of the SCATS system include measurements of the traffic flow as well as the period of each cycle of each intersection. Let N_k^s be the number of vehicles passing sensor location s (s=8,9,10 at intersection n=313) during the time intervals $[t_k^n, t_k^{n+1})$, n=313, then $\lambda_{L11}^k = (N_{7,314}^k + N_{8,314}^k + N_{9,314}^k + N_{12,314}^k)/(t_{k+1}^{314} - t_k^{314})$, while $\mu_{L11}^k = (N_{8,313}^k + N_{9,313}^k + N_{10,313}^k)/(t_{k+1}^{313} - t_k^{313})$. Note that the k-th sample for arrival and for departure flows does not in general correspond to the same physical time t. Due to this limitation in our experimental setup, for the time being, we will only focus on the study of the development and validation of a model of traffic flow (not a queue-length) as a hybrid system. Online control applications will be treated in a future paper.

Aim of the current experiment is to check the practical implementability of our algorithm by identifying the parameters $\theta = \{\alpha_1, \varphi_1, \sigma_1^2, \alpha_2, \varphi_2, \sigma_2^2, \alpha_3, \varphi_3, \sigma_3^2, \pi\}$ of the JMM model for arrival and departure traffic flows. For this purposes we use as input for the EM-algorithm described above data for one-day (0 pm – 24 am) obtained on September 1, 2012, from the video cameras installed for the operation of the SCATS traffic control system.. The data is taken from the area of Thamrin Street (data courtesy of the Newtel Pte Ltd). Table 2 shows the 18 parameters of the model for traffic flow thus obtained. It turns out to be feasible to identify all possible modes: free flow mode, congested mode and faulty mode. By using these parameters we can characterize the AR model of the traffic flow in each of the 3 modes, and the transition probability matrix $\Pi = (\pi_{ij}, I,j = 1,2,3)$ describing the Markovian mode transition process. Since for all estimated values we find that $|\varphi_i| < 1$, the stationary value $E_i(y(t))$ can

be defined $E_i(y(t)) = \dfrac{\alpha_i}{1 - \varphi_i}$, where $i = 1$, 2 and 3. The values of stationary value $E_i(y(t))$ can be found in Table.2.

It is clear from the results of parameter estimation in Table 2, that the EM technique is able to identify the modes. In the case of traffic flow μ_{L11}, the first and third modes have an average value $E_i(y(t))$ that is almost the same but with very different values of the variance. This might make the results sensitive to how outliers –unlikely or abnormal events or observations - in the data set are treated. In estimation techniques, a standard approach to deal with this outlier issue is to reject any measurement that is at least three standard deviations away from 'the normal' measurement. This means that the definition of outlier depends on what we consider the normal variation. A value for the variance that is almost tripled, is a strong indication that the third mode is an outlier mode. This is also obvious from Figure 4 when considering the results for the estimated probabilities for those measurement data $y(t)$ marked with a circle, indicating that they belong to the third mode. These points in time correspond to events like traffic incidents, persistent counting errors, night counting error. It is important to know the system dynamics at each time, including at these rare events, and to detect when a change in the AR model parameters occurs.
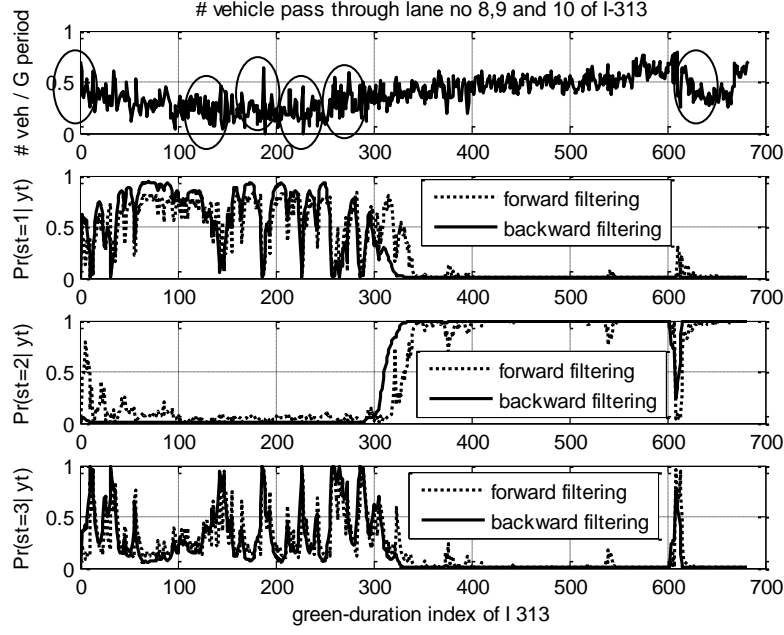
Figure 4. One-day (0 pm – 24 am) measurement data of μ_{L11} (top graph), JMM based estimated probability of mode $s = 1,2,3$ (second, third and bottom graph)

Table 2. Parameter estimation results

Traffic flow μ_{L11}			Traffic flow λ_{L11}		
$S_n=1$	$S_n=2$	$S_n=3$	$S_n=1$	$S_n=2$	$S_n=3$
$\varphi_1=$ 0.6130	$\varphi_2=$ 0.9205	$\varphi_3=$ 0.4579	$\varphi_1=$ 0.6123	$\varphi_2=$ 0.8163	$\varphi_3=$ 0.5033
$\alpha_1=$ 0.0950	$\alpha_2=$ 0.0330	$\alpha_3=$ 0.1459	$\alpha_1=$ 0.0608	$\alpha_2=$ 0.0180	$\alpha_3=$ 0.0373
$\sigma_1^2=$ 0.0070	$\sigma_2^2=$ 0.0078	$\sigma_3^2=$ 0.0222	$\sigma_1^2=$ 0.0068	$\sigma_2^2=$ 0.0292	$\sigma_3^2=$ 0.0045
$E_1(y_t)=$ 0.245	$E_2(y_t)=$ 0.415	$E_3(y_t)=$ 0.269	$E_1(y_t)=$ 0.157	$E_2(y_t)=$ 0.098	$E_3(y_t)=$ 0.075
Transition Probabilities $\pi = \begin{bmatrix} 0.8836 & 0 & 0.1164 \\ 0 & 0.9979 & 0.0021 \\ 0.1788 & 0.0142 & 0.8071 \end{bmatrix}$			Transition Probabilities $\pi = \begin{bmatrix} 0.9843 & 0.0071 & 0.0086 \\ 0.0054 & 0.9946 & 0 \\ 0.0122 & 0 & 0.9878 \end{bmatrix}$		

Figure 4 shows both forward and backward filtering results and it seems that both approaches give similar results in terms of predicting the outliers. Overall, the EM algorithm is able to correctly identify clusters of mode-1, mode-2 or mode-3 operation, meaning that the EM technique is applicable for modelling free flow and the congested modes as well as the faulty/outlier modes.

Analysis of Figure.4 and Table 2 shows that the transition probability to the outlier mode for traffic flow μ_{L11} is very high. By analyzing the green duration data (not shown here) at times when the probability of an outlier is high, it becomes clear that this is due to the fact that when the traffic volume increases, the green duration is decreasing compared to the previous one. This is an undesirable effect of the control system that can be classified as an error in the timing of the green switches. This implies that traffic flow at those times increases drastically, as shown in fig. 4 with a circle marking. In the case of traffic flow λ_{L11}, the EM parameter estimation is also able to identify the modes as shown in Table 2.

The performance of this EM algorithm against the weighted log-likelihood function $L(\theta^m)$ can be evaluated and does provide insight in how close the convergence to the true maximum likelihood is feasible. We first analyse the convergence of the algorithm in terms of the evaluation of $L(\theta^m)$. Figure 5.d shows that, after a short transient period (< 40 iterations), the EM-algorithm decreases the value of the weighted log-likelihood function $L(\theta^m)$ - thus also increases the likelihood - as the iterations of the parameter estimation proceed, indicating convergence of the estimated parameters (assuming identifiability).
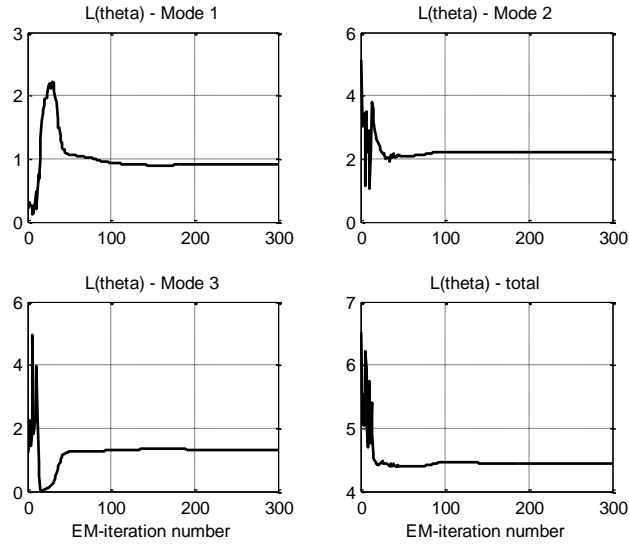

Figure.5 Convergence $L(\theta^k)$ of the EM algorithm for μ_{L11} : (a) $L(\theta^k)$ of mode 1 (b) $L(\theta^k)$ of mode 2 (c) $L(\theta^k)$ of mode 3 (d) $L(\theta^k)$

6. Model Validation and Estimation

In the previous section we analysed the convergence of the log-likelihood parameter estimation to an equilibrium. To validate whether this equilibrium is useful for model based state estimation and for online control we compare in this section the evolution of measurement $y(t)$ and the estimated values $\hat{y}(t)$, based on a simulation that implements the JMM model of equation (3), with parameter values shown in table 2, identified by EM-algorithm. In other word, model validation is performed by using the model based estimator using the identified model (obtained via EM parameter estimation) and checking how accurately this estimator predicts the traffic flow. As mentioned in the previous section, modes are not directly observable and hence, *'the smoothed inferences'* about the mode state process are used.

There are two different ways in which the smoothed inferences can be used for the model based prediction:
   (a) equally weighted empirical distributed function (*ewedf*),
   (b) weighted empirical distribution function (*wedf*).

Below we compare the performance of the *ewedf* and the *wedf* approach as part of the algorithm for the prediction of the traffic flow rates. For the *ewedf* approach, we use the natural choice of relating each observation with the most probable mode by letting $S(t) = i$ if $P\left(S(t) = i \middle| \mathbf{y}_1^{t-1}; \theta\right) > 0.5$. The evolution of the modes is shown in Figure 6. The mode classification using *ewedf* does not always work well, especially for cycle-indices around 300 (around 10 am, marked by a red oval). Around this time the probabilities of mode-1, mode-2 and mode-3 are all less than 0.5. The probabilities for all modes can be found in Figure 4. This implies that during this period around 10 a.m., observations cannot be unambiguously classified to one of the modes. The first graph of Figure 7, which shows more detailed results around the cycle time 300, shows that, during this period (again marked by a red oval and represented as 'mode-0') the simulation results also show a big difference with the measurement data. In order to overcome this problem we developed the w*edf* approach where we define weights $wf_i = P\left(S(t) = i \middle| \mathbf{y}_1^{t-1}; \theta\right)$, and based on these weights, we predict the measurement by multiplying the weights wf_i and the right-side of equation (3): $\alpha_i + \varphi_i\, y(t-1) + w(t)$ and then summing the 3 products (as symbolized by $\sum$) which amounts to averaging over modes:

$$\hat{y}(t) = \sum_{i=1}^{3} wf_i\, (\alpha_i + \varphi_i\, y(t-1) + w(t)),$$

The 2nd graph of Figure 7 shows that *wedf* gives better predictions results than *ewedf*, especially during those periods when observations cannot be classified unambiguously to one of the modes.

It should be emphasized that both approaches *ewedf* and *wedf* use *the smoothed inference* that was generated by the iterated EM algorithm. These approaches are not easily implementable in real time due to the need for a bigger memory to save the whole available information $P\left(S(t) = i \middle| \mathbf{y}_1^{t-1}; \theta\right)$, but the approach is quite useful for validating the model.

To overcome the limitation of the offline approach, we developed a method that can be used both for validating the model and for estimating traffic flow in an online algorithm. For the online method, we only use the parameter θ of the JMM in order to estimate the traffic flow for the next time-window. In this paper we use a type of particle filter technique called *observation and transition-based most likely modes tracking particle filter* (OTPF) technique for hybrid systems as developed by [14]. We will not discuss in detail the OTPF algorithm that has been implemented in this experiment with N=500.

We will use a simulation scenario with different time-window shift in order to study the performance of the JMM based prediction :

Simulation is tested in the period of index 350-450 with a hybrid model, which is obtained from the EM-estimated parameters using measurement data from the period with index 1: 350 (corresponding in real time to midnight till approximately 10.30 am).

Figure.8 shows the simulation of JMM based prediction w.r.t the scenario above. Simulation with different simulation scenario shows a similar performance (not shown here).
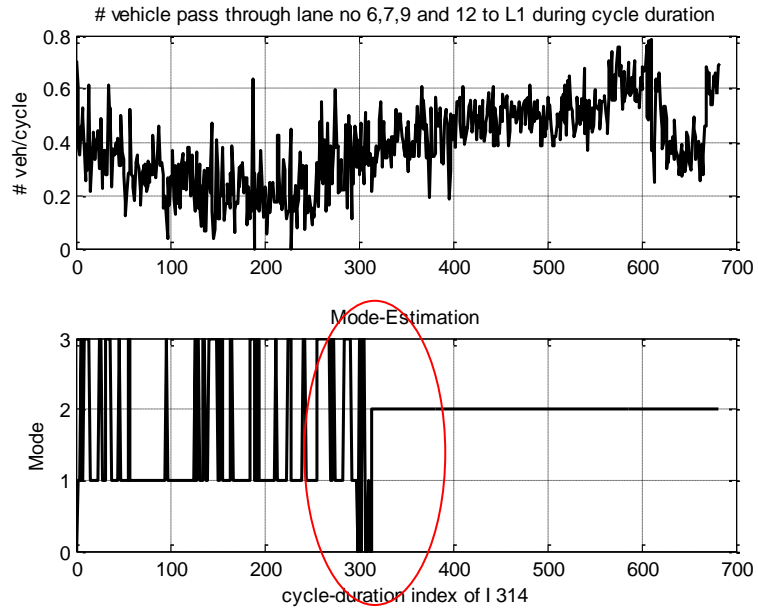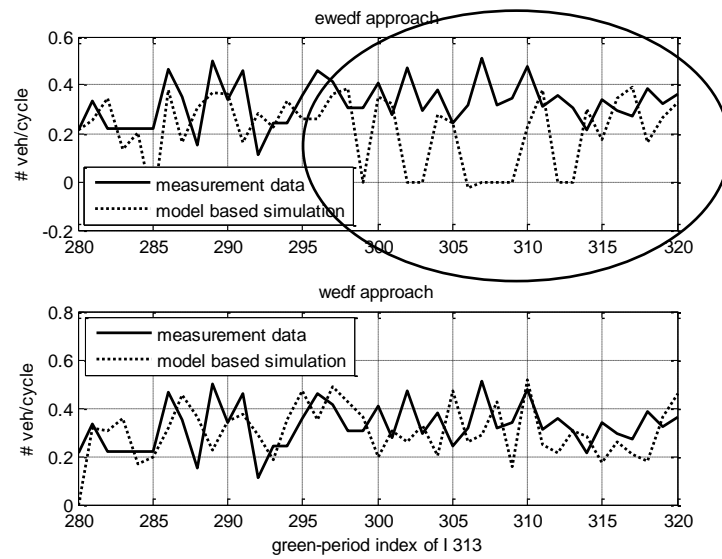
Figure.6 Mode-evolution

Figure.7 Measurement data vs model based simulation : ewedf approach (first graph) and wedf (second graph)
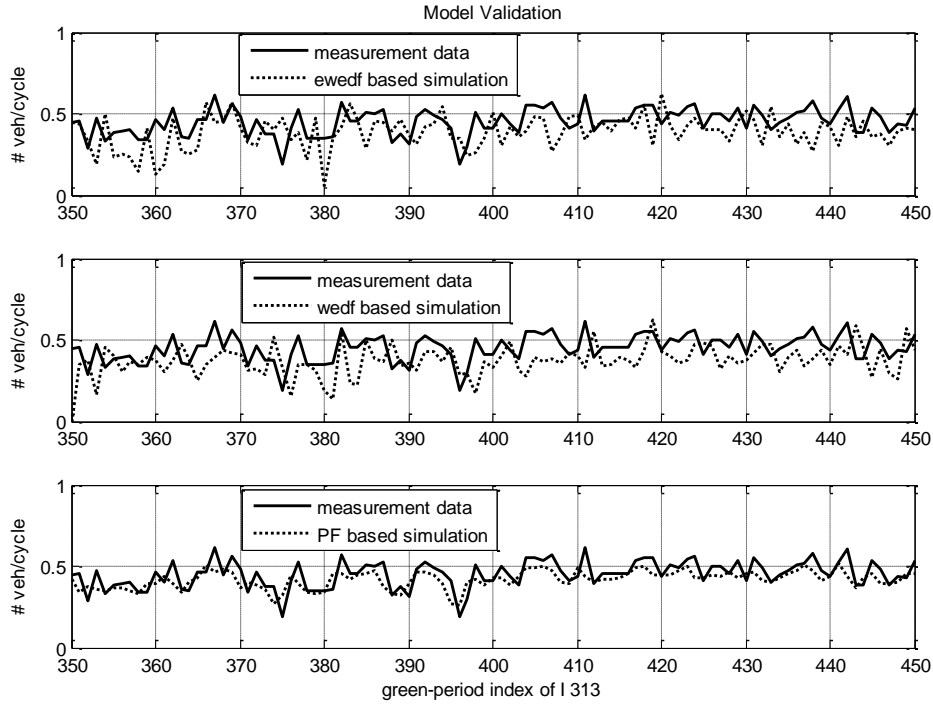
Figure.8 the simulation of JMM based prediction

The results of the scenario above shows that:

(1) EM offline parameter estimation along with a time-window shift technique can be useful and practical for updating the parameters of the JMM hybrid model. In the future we will investigate what choice of window size W is best for estimating and for prediction in the sense that it leads to the most useful results for online traffic light control. One approach to consider is a sliding window method, where the JMM model parameters are updated at times m.W, using measured data y_k over the interval $k \in [(m - n).W, m.W-1]$ to identify the model parameters, and these model parameters are then used over the interval $[m.W, (m+1).W)$ for predicting the traffic flow rate.

(2) Based on the update of JMM hybrid model, online PF estimator for the next time-window can be applied and may be a good candidate as a traffic flow state estimator for coordinating traffic light control in small network. This may of course be combined with the sliding window approach.

The results above confirm that a good model is pre-requisite for predicting traffic flow. In the context of designing feedback control at a signalized intersection, the prediction is an important factor to anticipate the variability of traffic flow coming from neighbouring intersections. This variability should be anticipated by traffic control in terms of green/red switching in order to reduce the congestion and to reduce the delay time. Many works in traffic engineering consider the prediction of traffic flow in designing traffic light control. Interested readers may refer to paper [15,16]. The data-driven modeling approach of this paper fits into the currently very active research domain trying to take full advantage of the huge amounts of available process data [17,18].

However, this EM offline parameter estimation along with the time-window shift technique has a disadvantage that the performance of state estimator strongly depends on the length of the time window shift. Certainly, the optimal choice of W is a compromise. EM algorithm will estimate the parameters more accurately if the time-window W is long enough. This is reasonable since the EM algorithm needs a certain amount of measurement data for accurate

parameter identification. But of course this will affect the ability of adaptive filters to detect a change of the model parameters sufficiently quickly. In the context of the coordinated control, the online parameter estimation approach is a critical part of an adaptive filter and finding the best window size W becomes an interesting research topic for further research.

7. Conclusions

This paper proposes a hybrid dynamic system model as a powerful approach for capturing the complicated dynamics of urban traffic flow, including many sources of uncertainty. The model is appealing as traffic flow conditions can be classified into three-modes and the switch between these 3 modes is controlled by a first-order Markov chain. The model is characterized by a set of parameters to be estimated using measured data (e.g. from a video camera overlooking traffic) and it is shown that a time-window shift technique may lead to a useful real time state estimation algorithm, that can be part of a feedback control loop.

The study reported in this paper investigated the proposed approach by using actual traffic flow data and confirmed its validity by showing that the *smoothed inferences* technique and a particle filter based on the identified model provide satisfactory state estimation and correctly capture the random variation of the traffic flow. The proposed hybrid model along with the particle filter estimator will be applied in a paper in preparation for queue length estimators as crucial part for the synthesis of traffic light control.

Acknowledgments

8. References

[1].Wang, Yibing., Papageorgiou, Markos., Messmer Albert., Coppola, Pierlugi., Tzimitsi, Atina and Nuzollo, Agustino, " An adaptive freeway traffic estimator",Automatica Vol 45 ,issue 1,January 2009.

[2].Mihaylova, Lyudmila., Boel, Rene and Hegyi, Andreas ,"Freeway Traffic Estimation within Particle Filtering Framework",Automatica, vol 43,pp 290-300,2007

[3]. Anthony.S and Karlaftis, Matthew G, "Multivariate state space approach for urban traffic flow modeling and prediction," Transportation Research Part C vol 11, pp. 271–350, 2003

[4]. Yasdi, R ,"Prediction of road traffic using a neural network approach", Neural Computing and Application 8, 135-142,1999.

[5]. Sutarto, Herman Y., Joelianto, Endra and Sumardi S,Taufik," Estimation and prediction of road traffic flow using particle filter for real-time traffic control", IEEE Conference on ICCSII, Bandung, 2013

[6]. Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. "Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach." *Computer-Aided Civil and Infrastructure Engineering* 23.7 (2008): 536-548.

[7]. Julves,J. and Boel, R." A Continuous Petri Net Approach for Model Predictive Control of

Traffic Systems", IEEE Transactions on Systems Man and Cybernetics – Part.A Systems and Humans, August, 2010.

[8]. Dempster,A., Laird,N.,Rubin,D.B ,"Maximum Likelihood from incomplete data via the EM algorithm.Journal of Royal Statistics Society,39,1-38,1977.

[9]. Hamilton J,"Analysis of time series subject to changes in regime", Journal of Economics,45,39-70,1990

[10].Kim,C.J."Dynamics Linear Model with Markov-Switching", Journal of Economics,60,1-22,1994.

[11]. Moon, Todd K," The Expectation-maximization algorithm", Signal Processing Magazine, IEEE 13, no.6 (1996) : 47-60.

[12]. Schon, Thomas B., Wills, Adrian and Ninness, Brett ,"System Identification of Nonlinear State-Space Models", Automatica, vol 47(1), January 2011.

[13]. Roads and Traffic Authority, SCATS NSW Government,   http://www.scats.com.au/ accessed Jan 2014

[14]. Tafazoli,Siamak and Sun, Xuchong,"Hybrid System State Tracking and Fault Detection using Particle Filters", IEEE Transactions on Control System Technology,Vol 14,No 6,Nov 2006.

[15]. Lucas Barcelos de Oliveira, Eduardo Camponogara, " Multi-agent model predictive control of signaling split in urban traffic networks", *Transportation Research Part C* ,18(1), 120-139,2010

[16]. K. Aboudolas, M. Papageorgiou, A. Kouvelas, E.Kosmatopoulos, " A rolling-horizon quadratic programming approach to the signal control problem in large-scale congested urban road networks", *Transportation Research Part C*, 18(5),680-694, 2010.

[17]. S.Yin, S.X. Ding, X.Xie and H.Lao. "A Review on Basic Data-Driven Approaches for Industrial Process Monitoring", *IEEE Transactions on Industrial Electronics*, Vol.61.No.11,Nov,2014.

[18]. S.Yin, X.Li,H.Guo and O.Kaynak, " Data Based Technique Focused on Modern Industry: An Overview", *IEEE Transactions On Industrial Electronics*, Vol. 61, No 2,Feb,2014.

[19]. C. F. WU, "On the convergence properties of the EM algorithm", *The Annals of Statistics*, 11, pp. 95–103, 1983

[20]. L.Ljung, "System Identification: Theory for the User", *Prentice Hall Inc*, New York, 1999.