



## MODELING, IDENTIFICATION, ESTIMATION, AND SIMULATION OF URBAN TRAFFIC FLOW IN JAKARTA AND BANDUNG

Herman Y. Sutarto<sup>a,\*</sup>, Endra Joelianto<sup>b</sup>

<sup>a</sup>SYSTeMS Research Group, Universiteit Gent  
Technologie park 913, B-9052 Zwijnaarde, Belgium

<sup>b</sup>Instrumentation and Control Research Group, Institut Teknologi Bandung  
Jalan Ganesha 10, Bandung 40132, Indonesia

Received 6 December 2014; received in revised form 6 April 2015; accepted 8 April 2015  
Published online 30 July 2015

### Abstract

This paper presents an overview of urban traffic flow from the perspective of system theory and stochastic control. The topics of modeling, identification, estimation and, simulation techniques are evaluated and validated using actual traffic flow data from the city of Jakarta and Bandung, Indonesia, and synthetic data generated from traffic micro-simulator VISSIM. The results on particle filter (PF) based state estimation and Expectation-Maximization (EM) based parameter estimation (identification) confirm the proposed model gives satisfactory results that capture the variation of urban traffic flow. The combination of the technique and the simulator platform assembles possibility to develop a real-time traffic light controller.

Keywords: intelligent transportation system; stochastic hybrid system; state/parameter estimation; expectation-maximization; particle filter.

### I. INTRODUCTION

The application of advanced technologies including computers, electronics, and communications can contribute to traffic conditions improvement, environmental quality enhancement, and economic productivity increase. These technologies, collectively known as intelligent transportation system (ITS), are rapidly being accepted by transport authorities as a viable alternative to reliance on building more roads to reduce congestion. Advanced traffic management systems (ATMS) in particular have been shown in many experiments around the world to decrease travel times, to improve networks speeds, and to reduce environmental emissions and congestion. However, the benefits of these systems have also been found to be a function of the accuracy and the robustness of the underlying computer algorithms and optimization techniques that provide various levels of intelligence to traffic signal control and operations.

One of the ITS technologies is a coordinated control of traffic lights in an urban area. It

proposes good prospects for reducing travel time and pollution due to traffic. The control action that can be used in order to improve the behavior of the system consists of selecting the red-to-green switching times of traffic signals. However, the design of a good (model based) control system, typically, needs a dynamic model that relies on the prior knowledge of a mathematical model that describes the evolution of states and the noise characteristics. Therefore, it is necessary to identify or to estimate the parameters of the system model, both offline and online, from the data observed during the operation.

The stochastic model must be able to describe the variability over time of the traffic flow, allowing probabilistic prediction of future traffic flows so that control actions by switching traffic lights can anticipate future traffic. Anticipation (including prediction) can improve the performance by properly coordinating traffic lights. This requires models that depict the evolution of traffic flow both in time and in space. Plant modeling can be either data-driven or based on first principles. Data-driven plant modeling uses techniques such as system identification (parameter estimation). With

\* Corresponding Author. Phone: +62-82130271090  
E-mail: hytotok@gmail.com

system identification, the plant model is identified by acquiring and processing raw data from a real-world system and selects a mathematical algorithm that is used to identify a mathematical model. Various kinds of analysis and simulations can be performed using the identified model before it is utilized to design a model-based controller.

The need for accurate traffic flow has long been proposed in the international scientific literatures [1,2,3]. Numerous methods have been developed especially for freeway traffic state estimation. These methods aim to estimating traffic variables such as flow, speed, and densities for considered freeway network with an adequate time-spatial resolution. This estimation is performed with an assumption based on a limited amount of available noisy measurement data. The methods were almost exclusively based on the macroscopic traffic flow modeling such as by using (extended) Kalman filter [4] and particle filter [5]. The macroscopic model depends on some important parameters such as free speed, critical density and capacity, etc. The values of some considered parameters are set as unknown parameters that should be determined by using previous off-line model calibration.

The stochastic models with estimated parameters can be used for estimating and predicting traffic flow, and thus also for estimating queue lengths at signalized intersections. The variability of the traffic flow during successive cycles of the traffic light is a prominent factor in predicting the expected queue-length, therefore estimating parameters. Establishing model that represents this variability correctly is a challenging problem due to strong coupling between traffic light (discrete events) and traffic flow (discrete time). The discrete event and the discrete time dynamics of such model are coupled, called the stochastic hybrid model.

In traffic measurement data, we have to deal with many types of errors. This paper uses data from a video-type sensor to detect the vehicles on the road network of two cities of Indonesia, namely Jakarta and Bandung. Noise for this type of sensor includes both false detections and missed vehicles (due to of traffic congestion, counting errors due to night time and obstacles in an urban environment).

The paper is aimed to introduce the general overview of modeling, identification/estimation and simulation aspects of urban traffic flow which is then validated using actual traffic flow data. In the simulation parts, VISSIM is used to

simulate traffic flow through the analysis of queuing and fluid flow model.

## II. PROBLEM FORMULATION

In this section, we consider stochastic hybrid modeling 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, explaining why the identification/parameter estimation problem treated in this paper fits into the overall traffic control problem.

We may define the evolution of the queue-length by:

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

where  $\lambda_\ell(t)$  ( $\mu_\ell(t)$ ) is the traffic flow rate (generally called  $\alpha(t)$  so far) of the arrival stream of vehicles (respectively the departure stream of vehicles), measured in veh/sec. As mentioned above, we use the fluid-flow approach, aggregate the number of vehicles during a cycle period, and implicitly assume that during the  $k$ -th cycle, between  $t_k$  and  $t_{k+1}$ , the flow rate remains constant, equals to  $\lambda_k$  respectively  $\mu_k$ .

Figure 1 shows that intersection I-313 has arrival flow  $\lambda_{L11}$  which is determined by using sensor (video camera) output that counts the

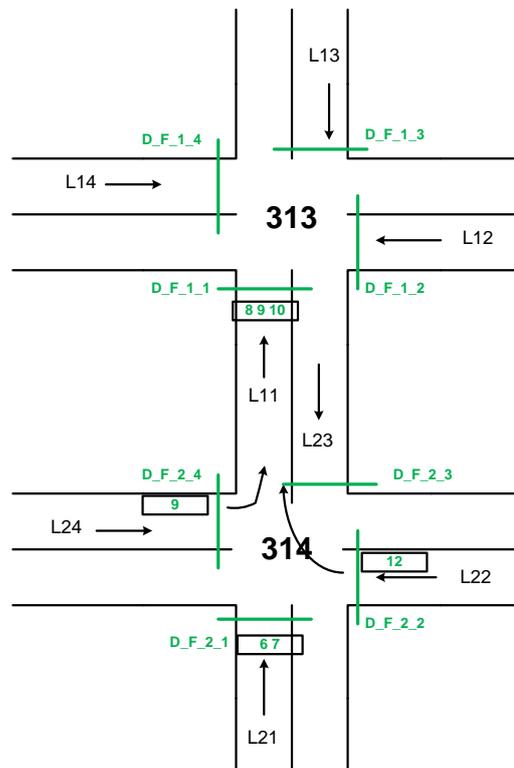


Figure 1. Configuration of arterial study

number of vehicles that passes the sensor location 9, 6, 7, and 12 divided by the cycle length of the intersection I-314. This approach defines the arrival rate  $\lambda(t)$ . The departure rate for intersection 314, called  $\mu_{L11}$ , is defined by counting the vehicles that pass the sensor location 8, 9, 10 at downstream intersection I-313, divided by 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 here). Of course this is under the assumption that drivers of vehicles that pass the sensor location 9, 6, 7, and 12 follow the traffic rules.

For each intersection, the traffic cycle is divided into two separate periods, called the red and the green periods. For each of the period, 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. There are, for each intersection, two decision variables  $T_g = t_{2k+1} - t_{2k}$  (green period) representing the time duration between switching times  $t_{2k}$  and  $t_{2k+1}$  while  $T_r = t_{2k+2} - t_{2k+1}$  (red period) acting between  $t_{2k+1}$  and  $t_{2k+2}$ . The traffic control signal problem is to determine, for each cycle, good values for these two decision variables by minimizing some cost functions, typically depending on the expected queue-length.

The design of a good traffic controller therefore requires that a good model is available to represent the evolution of the arrival and the departure traffic flows,  $\lambda_k$  (a non-negative

random process) and  $\mu_k$  (a non-negative random process that takes the value 0 whenever the traffic light is red for the corresponding direction). The stochastic hybrid model provides the required model. The goal of this paper is to develop an efficient algorithm for estimating the parameters of such model using real data from sensors at the location where the model is to be applied.

Referring to the simple intersection in Figure 1, with one single lane per approach road, let assume that a stochastic hybrid model for the arrival and the departure traffic flows  $\lambda(t) = \lambda_k$  for  $t$  in  $k$ -th cycle, and  $\mu(t) = \mu_k$  if  $t$  in  $k$ -th cycle and traffic light green,  $\mu_k = 0$  otherwise.

The evolution of  $\lambda_k$  and  $\mu_k$  is well defined as soon as the parameters of the stochastic hybrid model proposed in this paper have been identified with sufficient accuracy. A complete model for the queue sizes  $x_j(t)$  for the  $j$ -th queue is described in the following equation

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

The operation of the intersection can be modeled as a stochastic hybrid system as shown in Figure 2 with the time-driven dynamics described by (1) and the 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. The evolution of

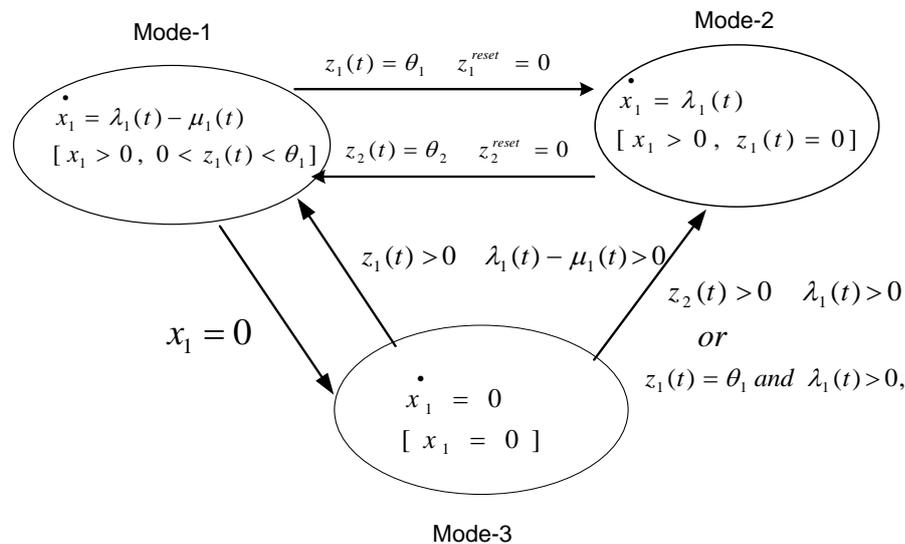


Figure 2. Stochastic hybrid automaton

queue-length with  $n=1$  can be described, for example, by a Stochastic Hybrid Automaton as shown in Figure 2. The event set that affects the evolution of queue is  $E = \{e_1, e_2, e_3, e_4, e_5\}$ , where: (a)  $e_1$  corresponds to the value of  $\lambda_j(t) - \mu_j(t)$  becoming strictly positive (after a period of non-positive), (b)  $e_2$  switches in the sign of  $\alpha_j(t)$  from 0 (zero) to strictly positive or a switch from mode-3 to mode-2, (c)  $e_3$  is the queue-length becoming empty, i.e.,  $x_j = 0$  or a switch from mode-1 to mode-3, (d)  $e_4$  switches a traffic light from RED to GREEN or a switch from mode-2 to mode-1, (e)  $e_5$  switches a traffic light from GREEN to RED or a switch from mode-1 to mode-2.

In this paper, we assume that  $\lambda_k$  and  $\mu_k$  can be approximated by hidden Markov model (HMM). In the next section, we further develop this model in detail. In later section, it is then shown how to estimate the parameters of this model by using data-driven approaches.

### III. MODEL OF TRAFFIC FLOW

In this approach, we use fluid flow model [6,7]. The traffic flow is described by a continuous random variable  $\beta(t)$  which expresses 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 by approximating integer numbers of vehicles by the real number. Hence, it can be defined as traffic flow with the ratio:

$$\beta_k = \frac{N_k}{(t_{k+1} - t_k)} \quad (3)$$

where  $N_k$  counts the number of vehicles 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  $t \in (t_k, t_{k+1})$ . This paper uses macroscopic approach to simplify the complexity of traffic flow. The hidden Markov model (HMM) model is also considered in this study. In this section, we model a generic traffic with flow rate  $\beta_k = \alpha_k$  (which could represent the arrival flow rate  $\lambda_k$  or the departure flow rate  $\mu_k$ ).

This HMM with Gaussian distribution is one of the simplest types of hybrid model. This type of HMM will be studied and used to show that this simple model with the transition probabilities is able to capture the complex phenomena of

urban traffic flows. The underlying idea behind HMM is to describe the observed phenomena by two (or more) separate regimes with different statistical properties. Here, we use the Gaussian distribution to describe the traffic flow distribution (called  $\alpha(t)$ ) [8]:

$$\alpha_s : \begin{cases} \alpha_1(t) \sim N(\mu_1, \sigma_1^2) & \text{if } s=1 \\ \alpha_2(t) \sim N(\mu_2, \sigma_2^2) & \text{if } s=2 \end{cases} \quad (4)$$

where the traffic flow measurement is described by two separates states (modes) with different underlying properties. The variable denotes the state variable for traffic flow condition "1" is the non-congestion mode, and "2" is the congestion mode. The random variable traffic flow is driven by the normal distribution with mean  $\mu$  and variance  $\sigma^2$  as parameters. Equation (4) implies that statistical properties of the measurement  $H_n$  depend on the actual state  $s$ . The jumping/switching mechanism between states is assumed to be governed by an unobserved Markov chain. The discrete transitions are assumed Markovian which means the current discrete mode  $s$  depends only on the mode in the preceding period  $s-1$ . This model is completed by defining the transition probabilities of moving from one mode to another denote.

$$\Pr(s_n = i | \xi_{n-1}) = \Pr(s_n = i | s_{n-1} = j) = \pi_{ij} \quad (5)$$

where  $\xi_{n-1}$  is a vector representing all the information available at the time  $n-1$ . These transitions can be used to represent the transition matrix  $P = [p_{ij}]$  controlling the behavior of traffic flow. Markovian assumption is reasonable since systems frequently exhibit probabilistic patterns in their switching behavior. To illustrate, urban traffic flow changes abruptly during the transition phases of entering or leaving in the rush hours. The HMM system has a framework of multiple modes corresponding to different traffic flow conditions. By defining the parameters  $\theta_s = \{\mu_s, \sigma_s, \pi\}$  of HMM, the objective is to estimate or to identify the parameters of HMM system (4) by using Expectation-Maximization (EM) technique described in the next section.

### IV. IDENTIFICATION/PARAMETER ESTIMATION

The parameter estimation can be carried out using an iterative two step EM procedure. In this case, the EM algorithm allows us to completely identify the HMM stochastic hybrid models proposed in the previous sections as a model for traffic flow. In this section, it is described the

application of the EM approach, originally proposed in [9] and then refined in [8] for switching systems (HMM) based on forward-backward recursion or 'smoothing'. The EM approach is formulated in batch or off-line form, i.e. it uses a given number of observations obtained over a time interval  $(0, T)$  to iteratively find better and better estimate of the unknown parameters of a model that is valid over the period  $(0, T)$ . This offline approach needs a significant memory requirement and processing power for storing and processing large data sets. However, this approach is shown to be useful and applicable further in this paper.

Estimation of the parameters  $\theta$  of the traffic flow model can be defined for HMM. The parameters of the HMM model are given by  $\theta = \{\mu_1, \sigma_1, \mu_2, \sigma_2, \pi_{ij}, i, j = 1, 2\}$  and the transition probabilities  $\pi_{ij}, i, j = 1, 2$ .

In this section, it will be briefly discussed the application of the EM algorithm for HMM due to limited space. The algorithm starts with an arbitrarily chosen vector of initial parameters  $\theta = \{\mu_1, \sigma_1, \mu_2, \sigma_2, \pi\}$  for  $s \in \{1, 2\}$ . In the E-step inferences about the state process are derived. Since  $S_n$  is unobservable, only the expected values of the state process given by the observation vector can be calculated. In Bayesian inferences, these expectations result in the so called smoothed inferences that represent the conditional probabilities of the process being in modes  $s$  at time  $n$ . Next step is the second step (M-step) maximum likelihood estimates of the parameters vector  $\theta$  based on the smoothed inferences in the E-step. The both steps (E and M steps) are repeated until the maximum of the likelihood function is reached with certain accuracy. The  $i$ -th mode weighted log-likelihood function is given by the following formula:

$$\ln[L(\mu_i, \sigma_i, \pi)] = - \sum_{n=2}^T P(S_n = i | \alpha_n; \theta^{(k)}) \left[ \frac{\ln(\sqrt{2\pi}\sigma_i) + (\alpha_n - \mu_i)^2}{2\sigma_i^2} \right] \quad (6)$$

The two main issues with the implementation of the EM algorithm are: (1) The conditional expectation is sometimes difficult to compute. This is the reason for using smoothed inferences which means that we use the whole data (batch) to find more accurate results. (2) The parameter vector converges to a local maximum, a good choice of initial conditions is necessary. In our traffic case, it is simply done by choosing the initial conditions based on the real measurement data.

A detailed explanation of the algorithm is given in the following steps.

### A. The E-Step

The E-step consists of forward filtering and backward filtering/smoothing. This step aims to calculate the conditional probabilities  $P(S_n = i | \alpha_{n-1}; \theta^{(k)})$ . Assume that  $\theta^n$  is the parameter vector calculated in the M-step during the previous iteration and  $\rho_i^{(0)} \equiv P(S_1 = i)$ .

#### 1) Forward Filtering:

For  $n = 1, 2, \dots, N$  iterate on this equation:

$$P(S_n = i | \alpha_n; \theta^{(k)}) = \frac{P(S_n = i | \alpha_{n-1}; \theta^{(k)}) g_i(\alpha_n; \theta^{(k)})}{\sum_i P(S_n = i | \alpha_{n-1}; \theta^{(k)}) g_i(\alpha_n; \theta^{(k)})} \quad (7)$$

where  $\alpha_n = (\alpha_1, \alpha_2)$  is the vector of traffic flow measurements and  $g_i(\alpha_n; \theta^{(k)})$  is the probability density function at time  $n$  conditional on the measured value coming from mode  $i$  (here,  $g_1$  and  $g_2$  are Gaussian PDFs as shown in equation (4)); and

$$P(S_{n+1} = i | \alpha_n; \theta^{(k)}) = \sum_{j=1}^2 p_{ji}^{(k)} P(S_n = j | \alpha_n; \theta^{(k)})$$

until  $P(S_N = i | \alpha_N; \theta^{(k)})$  is calculated.

The starting point for the iteration is chosen as:  $P(S_1 = i | \alpha_0; \theta^{(k)}) = \rho_i^{(k)}$ .

#### 2) Backward filtering

for  $n = N-1, N-2, \dots, 1$  iterate on

$$P(S_n = i | \alpha_n; \theta^{(k)}) = \sum_{j=1}^2 \frac{P(S_n = i | \alpha_n; \theta^{(k)}) P(S_{n+1} = j | \alpha_n; \theta^{(k)}) p_{ij}^{(k)}}{P(S_{n+1} = j | \alpha_n; \theta^{(k)})} \quad (8)$$

### B. The M-Step

The new parameter estimates  $\theta^{(k+1)}$  can be derived by maximizing the log-likelihood function as shown in equation (6). It is straight forward in finding the parameters  $\mu_i$  and  $\sigma_i^2$ .

$$\mu_i^{(k+1)} = \frac{\sum_{n=1}^N f_{i,n} P(S_n = i | \alpha_N; \theta^{(k)})}{\sum_{n=1}^N P(S_n = i | \alpha_N; \theta^{(k)})} \quad (9)$$

$$(\sigma_i^{(k+1)})^2 = \frac{\sum_{n=1}^N g_{i,n} P(S_n = i | \alpha_N; \theta^{(k)})}{\sum_{n=1}^N P(S_n = i | \alpha_N; \theta^{(k)})} \quad (10)$$

where  $f_{i,n} = E_{n,i}$  and  $g_{i,n} = [E_{n,i} - \mu_i^{(k+1)}]^2$ ,  $i = 1, 2$ .

Finally, we have  $\rho_i^{(k+1)} = P(S_1 = i | \alpha_N; \theta^{(k)})$  and the transition probabilities are estimated by using equation (11), with the transition probabilities are restricted only by the condition that  $p_{ij} \geq 0$  and  $(p_{i1} + p_{i2} + p_{i3} + \dots + p_{iN}) = 1$  as shown in [9].

$$p_{ij}^{(k+1)} = \frac{\sum_{n=2}^N P(S_n = j | \alpha_N; \theta^{(k)}) p_{ij}^{(k)} P(S_{n-1} = i | \alpha_{n-1}; \theta^{(k)})}{\sum_{n=2}^N P(\alpha_n = j | \alpha_{n-1}; \theta^{(k)})} \quad (11)$$

where  $p_{ij}^{(k)}$  is the transition probability from the previous iteration. All values obtained in the M-step are then used as a new parameter vector  $\theta^{(k+1)} = (\mu_i^{(k+1)}, \sigma_i^{(k+1)}, P^{(k+1)}, \rho_i^{(k+1)})$ ,  $i=1,2$  in the next iteration of the M-step. The algorithm is terminated when  $|\theta(k) - \theta(k-1)| \leq \delta$  for some small  $\delta$ .

## V. STATE ESTIMATION

As mentioned previously, the paper concerns with hybrid systems. Referring to equation (2) to (4) as a complete hybrid model of queue-length evolution, it is clear that the model depends on the arrival and departure flows:  $\lambda$  and  $\mu$ . These flows can then be approximated by HMM whose parameters evolve through time according to the mode  $s_k$ . These parameters of HMM are resulted from EM parameter estimation.

The state estimation/filtering problem means to look for an estimate of the state  $x$  based on the measurements up to the time instant  $k$ , which will be denoted by  $Y_k = \{y_1, y_2, \dots, y_k\}$ . Due to the stochastic nature of the system, the state is a random variable described by the conditional PDF  $p(x|Y_k)$ . It is important to note that in state estimation, we consider the state not as parameters. In this state estimation, we use PF technique but the PF algorithm is not discussed in detail. The tutorial paper by [11] is a very good introduction to study PF. In this PF, we use parameters of the HMM model identified in the EM parameter estimation.

In this paper, we develop particle filter (PF) which is one of the approximations of Bayesian estimation. PF state estimation of a hybrid system with computationally efficient is the observation and transition-based most likely mode tracking particle filter (OTPF) technique proposed in [10].

In state estimation based on OTPF, one needs to compute the PDF of the hybrid system:  $p(x_k, s_k | Y_k)$ , where  $s_k$  is the mode of the system at time  $k$ . The basic idea of OTPF algorithm is to look at the mode as an unknown system parameter. Since at each time step the system only follows one mode, and then it is reasonable to let it follows the most-likely mode denoted by  $\hat{s}_k$ , therefore:

$$\begin{aligned} p(x_k, s_k | Y_k) &= p(s_k | Y_k) p(x_k | s_k, Y_k) \\ &\sim p(\hat{s}_k | Y_k) p(x_k | \hat{s}_k, Y_k) \\ &= p(\hat{s}_k | Y_k) p_{\hat{s}_k}(x_k | Y_k) \end{aligned} \quad (12)$$

where  $p(\hat{s}_k | Y_k)$  is a constant and  $Y_k = \{y_1, y_2, \dots, y_k\}$  are the measurements up to the time instant  $k$ . From the derivation of equation (12), it can be noted that  $p(\hat{s}_k | Y_k)$  will be absorbed in the normalization constant and will not affect the basic PF algorithm. Thus, the estimation problem for the hybrid system can be solved by using the basic PF [11] after the most likely mode is substituted.

The PF in combination with the EM offline parameter estimation is a crucial part of the predictor system of the traffic flow. It can be useful and practical for periodically updating the parameters of hybrid model leading to an adaptive traffic flow state estimator/predictor. The interested reader may refer to references [8] for the detail.

## VI. EXPERIMENTAL LAYOUT

This section focuses on modeling urban traffic flows using parameter estimation techniques. The data set of the experiment layout is shown in Figure 1 over a time window  $[0, T]$ . The time-window size is important in order to define the model that will be used for estimating traffic flow for the next time-window. The effect of varying time-window size and its practical implication on the performance of the estimator will be investigated in the next section. The accuracy of estimated traffic flow is critical for achieving good performance of traffic control strategies.

An example of the arrival flow  $\lambda_{L11}$  and the departure flow  $\mu_{L11}$  of the intersection I-313 using the data with a one day time-window size to identify the parameters of the flow model as a hybrid system with 3 modes is considered as introduced in the section 2 and 3. The data set was obtained from the sensors of the the sydney coordinated adaptive traffic system (SCATS) system on Jalan Thamrin in Jakarta. The system includes measurement of the traffic flow as well as the period of the cycle of each intersection. Note that the intersections (313 and 314) have a different red/green cycle time. This implies that there is a difference in the time intervals  $[t_k^n, t_k^{n+1})$ ,  $n=313$  or  $n=314$ , which are used for measuring  $\lambda_{L11}$  and  $\mu_{L11}$ . Let  $N_k^m$  be the number of vehicles passing sensor location  $m$  ( $m=6, 7, 9, 12$  at intersection  $n=314$ ;  $m=8, 9, 10$  at intersection  $n=313$ ) during the time intervals

$[t_k^n, t_k^{n+1})$ ,  $n = 313$  or  $n = 314$ , then  $\lambda_{L11}^k = (N_{7,314}^k + N_{8,314}^k + N_{9,314}^k + N_{12,314}^k) / (t_{k+1}^{314} - t_k^{314})$ , and  $\mu_{L11}^k = (N_{8,313}^k + N_{9,313}^k + N_{10,313}^k) / (t_{k+1}^{313} - t_k^{313})$ . Note that the  $k$ -th sample for arrival and for departure flows does not in general correspond to the same physical time  $t$ .

The paper only focuses on the study of the development and the validation of a model of traffic flow as a hybrid system. The experiment is aimed to model the traffic flows using HMM characterized by parameters shown in equation (2) to (4).

The objective of next experiment is to model the considered traffic flow using HMM. The model is characterized by parameters  $\theta_s = \{\mu_s, \sigma_s, \pi\}$  which are mean and variance of traffic flow in each mode, and transition probabilities. In this experiment, there are two cases: Case 1 is a case of an urban traffic flow of the City of Jakarta with the configuration and location is shown in Figure 1. The data are taken from September 1 to September 8, 2012. It is important to note that the data consist of not only the traffic flow but also the duration of cycle of each intersection. In this case, we consider a departure flow L11 of I-313. In this case, the traffic flow is defined as a number of vehicles passing to the lane number 8, 9 and 10 of L11/green duration of I-313. Whereas case 2 is a case of incoming flow to urban network on the City of Bandung is considered. The data were taken at 15 minutes time interval during June 11 to June 13, 2012. The data were gathered from 0:00 AM to 24 PM on each day. In this case, the traffic flow is defined as a number of vehicles passing the specific lanes during 15 minutes.

Using the EM-technique as shown in previous section, we can find the parameters of the model for the traffic flow of the two cases as shown in Table 1. It is clear that the EM technique is able to identify the modes. Although this HMM is a simple model, as shown in Figure 3, the model can capture the complexity of the traffic flow. A capability to identify the mode is represented by the results of estimated probabilities for that measurement data  $H_n$  being in specific mode  $j$  at time  $n$ , such as mode 1 or mode 2. This capability is very important to capture the transition phenomena in traffic flow. Overall, the EM algorithm with HMM is able to correctly identify clusters of mode-1 or mode-2.

The parameters in Table 1 are used to carry out one-step-ahead prediction, but we only do validation for the case 2 with 15 minutes time update. The particle filter (PF) technique will be applied to validate the model and to estimate the

traffic flow under the assumption that the parameters of the underlying HMM model are known correctly. As shown in previous section, the objective of optimization in the EM algorithm is to find the parameters of the HMM through maximization the log-likelihood function defined in equation (6) with respect to the underlying unknown parameters.

Case 2 is the case for 15 minute updated traffic flow and compared with measurement/observations on June 13, 2012. It is important to note that the parameters in Table 1 were built based on the measurement data on June 11-12. For prediction, we used measurement data on June 13. Figure 4 shows that the proposed model is able to capture the complex phenomena with satisfactory results except around index 80<sup>th</sup>. This result confirms that the model is applicable to this case.

## VII. TRAFFIC SIMULATION

The increasing power of computer technologies, the evolution of software engineering and the advent of the intelligent transport systems (ITS) have prompted traffic simulation to become one of the most used approaches for traffic analysis in support of the design and evaluation of traffic systems. The ability of traffic simulation to emulate the time variability of traffic phenomena makes it a unique tool for capturing the complexity of traffic systems. Microscopic traffic simulation based on the emulation of traffic flows from the dynamics of individual vehicle is becoming one the most attractive approaches.

The paper is not intended to explain the simulator in detail. The interested reader may refer to documents of VISSIM. However, in this paper, traffic simulator VISSIM will be used to analyze traffic flow in urban network, especially in the signalized intersection context.

We use VISSIM microscopic traffic simulator to create a hypothetical network to test the proposed Bayesian PF estimator (part of control strategies) at a network level. The studied network is a single grid type intersection. The length of each link is 300 meters with two-way roads exist in this network to simulate the typical situation of an urban network. The features or the characteristics of the simulation are described below. The desired speed distribution used in this exercise was 35 to 65 kph. This value was chosen based on prior experience with VISSIM and by default, we considered one vehicle occupies the full width of one lane in VISSIM.

The signals in the network are set to fit the configuration of the intersection. There are two-

phased signals. The signal is modeled as fixed controlled with duration of each phase is fixed for minimum green time to a maximum green time. The cycle length is set to 80 seconds with green duration is 35 seconds, amber duration is 3 seconds and all-to-red is 2 seconds. The traffic demand of this network is an input in the form of OD matrices. The vehicle flow of generation origin is 0.16 vehicle/second and the simulation time interval is 3,600 seconds. There are 'zones' located at the periphery of the network, and the OD matrix is defined as zone to zone demand. From the links associated with these zones, vehicles are released into the network.

In VISSIM simulator: (a). A vehicle is in queue condition if its speed is less than the *Begin* speed and has not exceeded the *End* speed yet;

(b). Current queue length is measured upstream every time step. From these values, the arithmetical average is computed for every time interval. By that definition and hybrid model of the queue-length in equation (2, 3, 4), the cumulative numbers of vehicles traversing sensor locations (e.g. AF-4 and DF-4) are recorded w.r.t traffic signal sequence which therefore refer to  $\lambda_{k,G} \cdot T_g$ ,  $\lambda_{k,R} \cdot T_r$ , and  $\mu_{k,G} \cdot T_g$ . Based on that, the PF algorithm can then be applied to estimate the queue length. The 'actual' queue length (by VISSIM) can be recorded by putting the queue counter in the front of stop-line *at the end of every red light* (cycle-by-cycle). In this case, we set and used type of 'average' of the queue-length. Figure 5 shows that the particle filter estimator gives a similar result compare to the

Table1.  
Parameter estimation results

Case 1:		Case 2:	
$S_n = 1$	$S_n = 2$	$S_{n=1}$	$S_{n=2}$
$\mu = 0.2174$	$\mu = 0.6806$	$\mu = 115.78$	$\mu = 471.16$
$\sigma^2 = 0.014$	$\sigma^2 = 0.044$	$\sigma^2 = 6.3762e+003$	$\sigma^2 = 4.3072e+003$
Transition Probabilities		Transition Probabilities	
$\pi = \begin{bmatrix} 0.9868 & 0.0132 \\ 0.0125 & 0.9875 \end{bmatrix}$		$\pi = \begin{bmatrix} 0.9761 & 0.0239 \\ 0.0192 & 0.9808 \end{bmatrix}$	

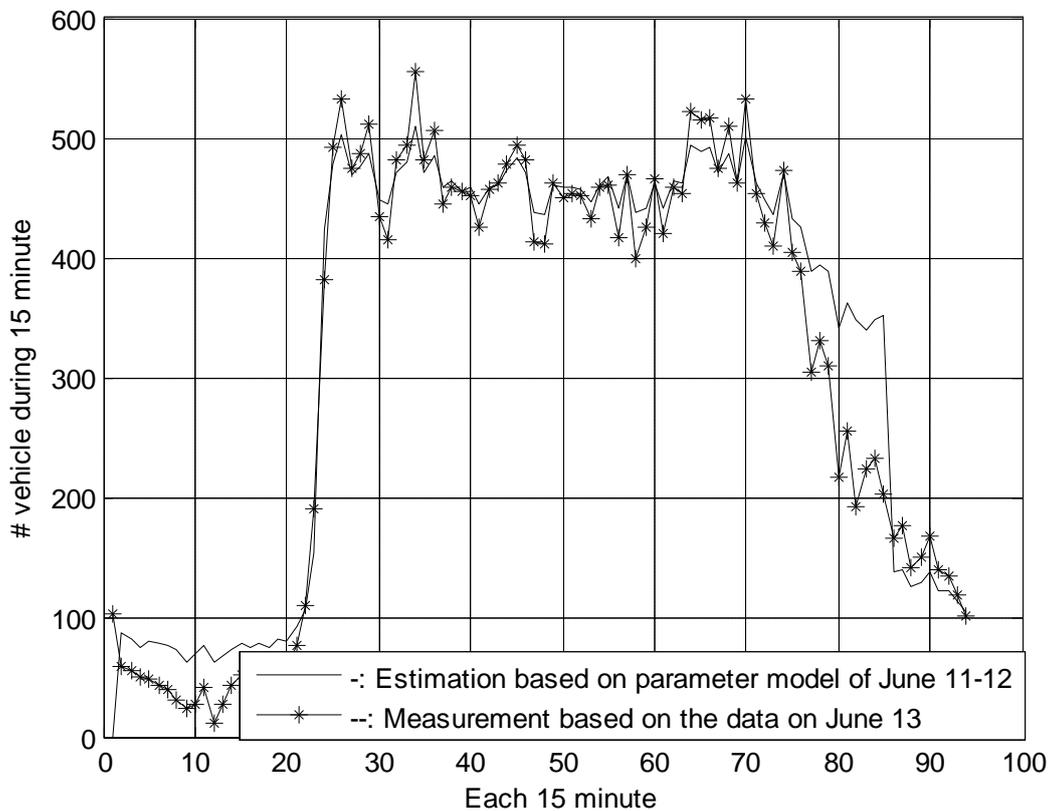


Figure 4. Predicted flow versus observed flow

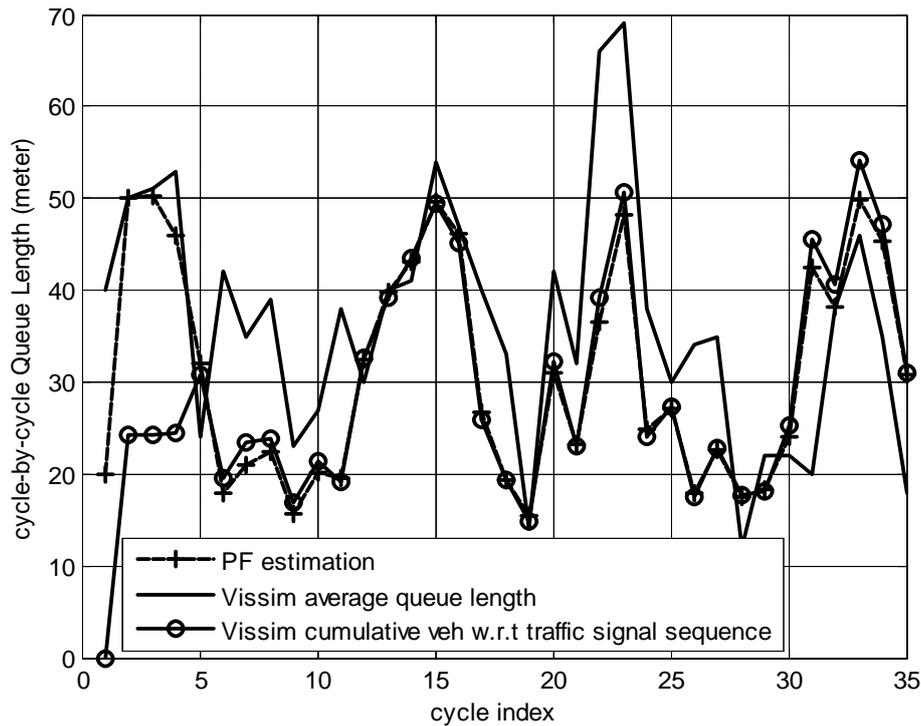


Figure 5. Queue length estimation in VISSIM



Figure 6. Visualization in VISSIM

both approach of VISSIM. Figure 6 shows the example of the visualization of signal time table and measuring data (veh) in each detector in VISSIM.

## VIII. CONCLUSION

The paper presents an overview of the application of modeling, identification,

estimation, and simulation techniques for urban traffic flows. The application confirmed that the proposed model gave satisfactory results which captures the variation of traffic flow. The technique together with simulator platform create the possibility to develop a real-time traffic light controller. By describing the urban traffic flow in the perspective of system and control theory, we encourage the control community to explore the

ideas and perspectives for ITS community. The big problem in ITS is how to develop the coordinated control for large-scale urban network and for that the control community has a big opportunity to offer a better solution.

### ACKNOWLEDGEMENT

The work of the first author was supported in part by European Community's Seventh Framework Programmed under Project DISC (Grant Agreement n.INFSO-ICT-224498). The authors would like to gratefully acknowledge Ir. Agoes Suwardi and Ir. Rosyad from P.T. Newtel, Bandung, and also Ir. Zulkarnaen Zakaria, M.Eng. from P.T. General Intelligent Technology, Bandung for their helpful discussion and assistance of the measurement data. The first author would also like to specially acknowledge Em.Prof. René K. Boel at Universiteit Gent for deep and helpful discussion and having an eye for the details and also Aditya Nugroho, Senior Engineer at Surbana International Consultans Pte. Ltd., Singapore and also Prof. Budi Yulianto from *Universitas Sebelas Maret*, Solo for their assistance and cooperation in running VISSIM simulator. This work is also supported by the Research and Innovation Program, Institut Teknologi Bandung, 2012.

### REFERENCES

- [1] W.Y. Ming, *et al.*, "User needs in green ITS: results of a questionnaire survey and proposal for green ITS design," *Inter. J. Intell. Transport. Syst.*, 10, 47–55, 2012.
- [2] W. Liu, *et al.*, "Decision-making analysis on urban transportation planning based on improved GRA model," *Inter. J. App. Math. Stat.*, 48(18), 136-142, 2013.
- [3] J. Zheng, "Simulation model for transportation flows and its nexus to economic growth under small sample," *Inter. J. App. Math. Stat.*, 52(3), 90-99, 2014.
- [4] Y.B. Wang, *et al.*, "An adaptive freeway traffic estimator," *Automatica*, 45(1), 10-24, 2009.
- [5] L. Mihaylova, *et al.*, "Freeway traffic estimation within particle filtering framework," *Automatica*, 43, 290-300, 2007.
- [6] H.Y. Sutarto and R.K. Boel, "Hybrid automata model approach for coordinating traffic signal control," *29<sup>th</sup> Benelux Meeting on Systems and Control*, Heeze-Netherland, 2010.
- [7] C.R. Vasquez, *et al.*, "Hybrid petri net model of a traffic intersection in a urban network," *IEEE Multi-conference on Systems and Control*, Yokohama, Japan, Sept 2010.
- [8] H.Y. Sutarto and E. Joelianto, "Expectation maximization based parameter identification for hidden markov model of urban traffic flow," *Inter. J. App. Math. Stat.*, 53(2), 90-101, 2015.
- [9] J.D. Hamilton, "Analysis of time series subject to change in regime," *Journal of Econometrics*, 45,39-70,1990.
- [10] S. Tafazoli and X. Sun, "Hybrid system state tracking and fault detection using particle filter," *IEEE Trans. on Control System Technology*, 14(6), 2006.
- [11] B. Ristic, *et al.*, "Beyond the kalmanfilter," *Artech House*, 2004.