International Journal of Informatics Information System and Computer Engineering

Modeling Traffic Flows with Fluid Flow Model

*Paulus Setiawan Suryadjaja *, Maclaurin Hutagalung **, Herman Yoseph Sutarto****

Department of Information Technology, Institut Teknologi Harapan Bangsa, Jl. Dipatiukur No. 80-84, Bandung, Indonesia

E-mail: *paulus1945s@gmail.com **maclaurin@ithb.ac.id ***hytotok@gmail.com

A B S T R A C T S

This Research presents a macroscopic model of traffic flow as the basis for making Intelligent Transportation System (ITS). The data used for modeling is The number of passing vehicles per three minutes. The traffic flow model created in The form of Fluid Flow Model (FFM). The parameters in The model are obtained by mixture Gaussian distribution approach. The distribution consists of two Gaussian distributions, each representing the mode of traffic flow. In The distribution, intermode shifting process is illustrated by the first-order Markov chain process. The parameters values are estimated using The Expectation-maximization (EM) algorithm. After The required parameter values are obtained, traffic flow is estimated using the Observation and transition-based most likely estimates Tracking Particle Filter (OTPF). To Examine the accuracy of the model has been made, the model estimation results are compared with the actual traffic flow data. Traffic flow data is collected on Monday 20 September 2017 at 06.00 to 10.00 on Dipatiukur Road, Bandung. The proposed model has accuracy with MAPE value below 10% or falls into highly accurate categories.

A R T I C L E I N F O

Article History:
Received 1 Nov 2020
Revised 20 Nov 2020
Accepted 25 Nov 2020
Available online 26 Dec 2020

Keywords:
Fluid Flow Model,
Expectation-maximization,
Particle Filter,
OTPF,
Intelligent Transportation Systems,
Gaussian Distributions,
Mixture Gaussian,
Markov Chain.

1. INTRODUCTION

Traffic jam in Bandung city has been at an alarming rate. The survey results of the Ministry of Transportation showed that Bandung is the seventh most congested city in Indonesia. The average vehicle speed in the city of Bandung is only 14.3 kilometers per hour and the value of Volume to Capacity (VC) ratio streets in the city of Bandung for 0.85 (Wiyono, A, S., 2014). For that, it needs to be applied more efficient traffic engineering system by utilizing modern technology, for example Intelligent Transportation System (ITS).

ITS is a utilization of information technology elements such as sensors, traffic models, software, and telecommunications networks to improve the security and efficiency of traffic systems. ITS was created to address the common problems of transportation systems that are less efficient, including air pollution in the streets, fuel wastage, transportation security, and waste of time due to traffic congestion (Ali, M., 2011). The control system that it applies requires an accurate model describing the behavior of traffic on the road network and the intersection for optimal performance. The model that it takes ITS system is a model of traffic flows and a long model of queuing at the intersection (Sutarto, Y, H., 2016). The behavior of traffic in an area is not the same as other regions, so it is necessary to create a model separately for each region that applies ITS. If the city of Bandung wants to implement ITS effective, it needs to be made accurate traffic model based on traffic behavior in Bandung city.

The traffic flow model created in this study is a model of the traffic flow with a macroscopic scale in the form of the Fluid Flow Model (FFM). FFM has the advantage of other macroscopic traffic flow models in terms of the ease of gathering and processing of traffic flow data. As a comparison, the model of the Spatio-Temporal Random Effects (STRE) , which was designed on research (Yao-Jan., et al, 2012) requires a large number of traffic flow data, i.e. two-week traffic flow data to be able to produce Satisfactory estimation. In addition, the model can produce an estimate with a high accuracy of MAPE 8% to 15% for traffic flow data on the main road between cities. As for traffic flow data in the center of the crowd, its accuracy is dropped to the MAPE number gained to 20% to 29.6%. The platoon-based traffic flow Model created in research (Marinica, N., and Boel, R., 2012) can be used to efficiently regulate the light length of red and green lights at the traffic lights. However, because the model modeled the movement of a group of vehicles, traffic flow data and the capacity of the road segments and the intersection are interconnected on the modelled street area. Traffic flow models can also be created based on the queue theory, as is done in research (Vandaele, N., et al., 2000). However, the steps that need to be done are more complicated. There are three variables that need to be collected data for later estimated namely the arrival rate (Arrival Rate), the service rate (service rate), and the maximum density (Maximum traffic density). Once these three variables are estimated, the maximum traffic flow values that can pass the modelled roads can be calculated.

The formulation of the problem with the estimation process of the parameter and state is described in part II of this paper. Part III explains the process of testing the accuracy of models that have been made using traffic flow data on Dipatiukur road, Bandung. Part IV contains the conclusion of this study as well as further research advice on development in the field of traffic flow modeling.

2. MAIN CONTENT

This section describes the process of creating a FFM traffic flow model. The process of creating models in the research consists of several stages namely the formulation of problems, estimation of parameters, and estimates State.

2.1. PROBLEM FORMULATION

Fluid Flow Model (FFM) analyses the flow of traffic as fluid flow. FFM described the flow with a random variable which indicates the number of vehicles crossing a section tk y of a road at a time interval t_k , $t_{k\ 1}$. Mathematically, y_{tk} can be defined as:

$$y_{tk} = \frac{N_{tk}}{(t_{k+1} - t_k)} \qquad (1)$$

With N_{tk} represents the number of vehicles passing through the observation point at the time interval t_k , $t_{k\ 1}$. Based on that definition, the value of y_{tk} is a non-negative number *real*. Because the definition states that at the time interval t_k , $t_{k\ 1}$ there is a vehicle as much as y_{tk} that crosses the observation point, so implicitly there is an assumption that the traffic flow at the time interval t_k , $t_{k\ 1}$ of unchanged (constant) and distance The vehicles passing through at that time interval were uniform (Sutarto, Y, H., 2016), (Sutarto, Y, H., and Joelianto, E., 2015).

In this study, to form the FFM, the random variable y_{tk} was described with the mixed Gaussian distribution approach (*Mixture Gaussian*). Gaussian distribution of the mixture is a system that is illustrated with some *modes* which are each a Gaussian distribution. on that distribution, the inter- *mode* is set with the process Markov chain First order. In this research, traffic flow is categorized into two *modes*. *The first* Mode represents normal traffic flow with a lower y_{tk} value and a second mode that represents crowded traffic with a higher y_{tk} value. With the Gaussian distribution of such mixtures, it is expected that the traffic flow can be described by *mode* which each has different characteristics (Sutarto, Y, H., and Joelianto, E., 2015).

The Model is made to form a non-Gaussian system, since the mixed Gaussian distribution not only contains a Gaussian distribution, but also contains a Markov chain component. Each *mode* traffic flow is a random variable with a Gaussian distribution. Thus, the individual *mode* has Gaussian distribution parameters, i.e. average and variance. To explain the inter-mode displacement process, the model also needs to contain a Markov chain component. The Markov chain (*Markov chain*) is used to estimate conditions in the present time based on known past conditions. The Markov chain process is generally described with *Transition Probability Matrix* (TPM) which indicates the probability of displacement from a state to another. In this study, the Markov chain was used to regulate displacement probability between existing *mode*. The system made assumed to follow the Markov chain first order. That is, to estimate *mode* that occurs at the present time only needs *mode* is going to take place in one previous unit.

The process *estimation parameter* is done with the help of the algorithm help *Expectation – Maximization* (EM) to look for the distribution parameters Gaussian and TPM chain Markov for each *mode* traffic flow. There are ten parameters that need to be estimated to compose FFM in this study, namely;

- Two Gaussian distribution parameters for each *mode*, namely the average μ and the variances σ^2. Since the model made in this research consists of two *modes*, then altogether there are four Gaussian distribution parameters that need to be estimated.
- The π Parameter indicating the probability of data flow traffic at a time is in the specified *mode of*. Since the model made in this research consists of two *modes*, then altogether there are two parameters π that need to be estimated.
- Transition Probability Matrix (TPM) P which indicates the probability of migrations from a mode to other modes. Since the model created on this research will classify the traffic flow into two modes, the TPM used in this study measures 2 × 2.

After the required parameters are obtained from the process 0659 estimation, it can be done state estimation. Estimation State is done to get the approximate value of the traffic flow based on the model compiled from the parameters that have been obtained. In this study, an estimate was made using the estimator Observation and transition-based most likely modes Tracking Particle Filter (OTPF).

To test the model made, the model is applied to estimate traffic flow data on Dipatiukur Road, Bandung City. The estimated result is compared to traffic flow data of the observation result, then calculated the value of error the estimate with method Mean Absolute Percentage Error (MAPE).

2.2. ESTIMATION PARAMETERS

The process estimation parameter is done with the algorithm Expectation – Maximization (EM). Broadly, the EM algorithm is an algorithm to suspect a parameter in a function that contains incomplete data using Maximum Likelihood Estimation (MLE). The algorithm has many advantages including this simple algorithm, based on common theories, and can be used in a wide range of applications (Kusuma, A, T., and Suparman., 2014), (Dempster P, A., et al 2014).

The EM algorithm consists of two stages, namely Expectation (E-Step) and Maximization (M-Step). At the Estep performed the expected expectation value for the likelihood function based on variables on the system being observed. Then the MLE value for the parameters is searched by maximizing the likelihood expectation resulting from Estep. The parameters generated from M-step will be reused for E-step in the next iteration, and this step will be repeated until it delivers a convergent value (Kusuma, A, T., and Suparman., 2014). The EM algorithm can generally be written as follows:

- **start** , $m = 0$: Estimate the initial value θ^m
- **Repeat** $m \rightarrow m+1$ (up to convergent),
  - **E Step:** Calculate

    $$Q(\theta \mid \theta^m) = E\left\{\log\left[f_x(x \mid \theta) \mid y = y\right]; \theta^m\right\}$$

  - **M Step:** Finalise

    $$\theta^{m+1} = \arg\max_{\theta \in \Theta} Q(\theta \mid \theta^m)$$

Here M is an index iteration, θ is a space vector parameter, y is the vector data, and $f_x\ x \mid$ is the *Probability Density Function*

(PDF) of the complete data (Moon, K, T., 1996).

In this study, the equation of PDF to complete data was taken from the equation PDF distribution Gaussian:

$$f\left(x,\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right) \quad (2)$$

Where x is the observed data, μ is the average, and Σ2 is a variance. By substituted the symbol on a Gaussian PDF with the symbol used in this study, a PDF function generated a data flow of traffic:

$$f\left(y_t \mid S_t = j,\theta\right) = \frac{1}{\sigma_j\sqrt{2\pi}}\exp\left(\frac{-(y_t-\mu_j)^2}{2\sigma_j^{\,2}}\right) \quad (3)$$

With y_t is the number of vehicles passing in minutes to- t, St is mode traffic flow in the minute to-t with *j*as its index θ is a space vecto R parameter that contains $\left\{\mu_j,\sigma_j^2,\pi_j\right\}$. π_j is the average in mode *J* and σ_j^2 is a variance on mode *j.*

The parameter π_j which is the probability of data flow at a time is at mode J based on the stochastic theory Markov Chain. The probability of data flow traffic y is in the mode J formulated with the equation.

$$P(S_t = j,\theta) = \pi_j \quad (4)$$

By multiplying the equation (3) and (4), then summing the results for all existing state, obtained the probability equation of the data flow y_t against the vector space ϑ:

$$\begin{aligned} P\left(y_t,S_t = j,\theta\right) &= f(y_t \mid S_t = j,\theta)P(S_t = j,\theta) \\ &= \frac{\pi_n}{\sigma_j\sqrt{2\pi}}\exp\left(\frac{-(y_t-\mu_j)^2}{2\sigma_j^{\,2}}\right) \quad (5) \\ f\left(y_t,\theta\right) &= \sum_{j=n}^{N} P\left(y_t,S_t = j,\theta\right) \\ &= \sum_{j=n}^{N}\frac{\pi_n}{\sigma_j\sqrt{2\pi}}\exp\left(\frac{-(y_t-\mu_j)^2}{2\sigma_j^{\,2}}\right) \quad (6) \end{aligned}$$

To find the estimated value of each parameter in each iteration of the EM algorithm, used Lagrange Multiplier. The Lagrange Multiplier specifies the maximum or minimum value relative of a function that is delimited by a constraint (constrain the conditions). By using Lagrange Multiplier, then in each iteration of the EM algorithm it can calculate the estimated Gaussian and TPM Markovian distribution parameters at once. This can be done by making TPM constraints as a delimiter function (Constraint) in the Help function Lagrange Multiplier.

$$G\left(x,y,z\right) = F\left(x,y,z\right) + \lambda\cdot\phi\left(x,y,z\right) \quad (7)$$

With is the auxiliary function, is the G (x, y, z) F (x, y, z) function that will be searched for the maximum or minimum value, and is the delimiter function ϕ (x, y, z) constraint). Optimal value is obtained by resolving the following differential equations.

$$\frac{\partial G}{\partial x} = \frac{\partial G}{\partial y} = \frac{\partial G}{\partial z} = 0.$$

In this research, the function that is searched for its optimal value is the likelihood function of the $f\left(y_t,\theta\right)$. Likelihood function which is suitable for use in the form of the log-likelihood function because the logarithmic function helps to simplify the exponential component of the $f\left(y_t \mid S_t = j,\theta\right)$.

$$L(\theta) = \sum_{t=1}^{T}\log\left(f\left(y_t,\theta\right)\right) \quad (8)$$

Where $L(\theta)$ is the function log-likelihood which is searched for the minimum value there are two limitations that need to be applied to the estimated calculation of the value 0659 п.

- $0 \le \pi_j \le 1$ Any component in the TPM must be greater than or equal to zero and less than or equal to one.
- $\pi_1 + \pi_2 + ... + \pi_N = 1$ The entire probability value of each mode occurrence (π_j) must be 1.

The Help function the Lagrange Multiplier used is:

$$J(\theta) = L(\theta) + \lambda(1 - \pi_1 - \pi_2 - ... - \pi_N) \quad (9)$$

The equations used to estimate the value of each parameter in each iteration of the EM algorithm are derived from the decline in auxiliary function Lagrange Multiplier $J(\theta)$. The auxiliary function Lagrange Multiplier $J(\theta)$ needs to be derived against the variables μ_j, σ_j^2, and π which is found in the function $f(y_t, \theta)$ that is searched for its optimal value. The process of deriving these equations is described as follows:

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{1}{f(y_t, \theta)} \frac{\partial f(y_t, \theta)}{\partial \theta} \quad (10)$$

Based on equations (3) and (5), then:

$$\frac{\partial f(y_t, \theta)}{\partial \pi_j} = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left\{ \frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right\} = f(y_t \mid S_t = j, \theta) \quad (11)$$

While the decline of $f(y_t, \theta)$ against μ_j and σ_j^2 generates:

$$\frac{\partial f(y_t, \theta)}{\partial \mu_j} = \frac{\partial}{\partial \mu_j} \left\{ \frac{\pi_j}{\sigma_j \sqrt{2\pi}} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right) \right\}$$

$$= \frac{\pi_j}{\sigma_j \sqrt{2\pi}} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right) \left(\frac{2y_t - 2\mu_j}{2\sigma_j^2} \right)$$

$$= P(y_t, S_t = j, \theta) \left(\frac{y_t - \mu_j}{\sigma_j^2} \right)$$

$$= \frac{y_t - \mu_j}{\sigma_j^2} P(y_t, S_t = j, \theta), \quad (12)$$

$$\frac{\partial f(y_t, \theta)}{\partial \sigma_j^2} = \frac{\partial}{\partial \sigma_j^2} \left\{ \frac{\pi_j}{\sigma_j \sqrt{2\pi}} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right) \right\}$$

$$= \frac{\partial}{\partial \sigma_j^2} \left\{ \frac{\pi_j}{\sqrt{2\pi} \sqrt{\sigma_j^2}} \right\} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right)$$

$$+ \frac{\pi_j}{\sigma_j \sqrt{2\pi}} \cdot \frac{\partial}{\partial \sigma_j^2} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right)$$

$$= \frac{\pi_j}{\sqrt{2\pi}} \cdot \left(-\frac{1}{2} \cdot \frac{1}{\sigma_j^3} \right) \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right)$$

$$+ \frac{\pi_j}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_j} \cdot \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right) \cdot \frac{-(y_t - \mu_j)^2}{2(-\sigma_j)^4}$$

$$= \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right) \cdot \left(\frac{\pi_j}{\sigma_j \sqrt{2\pi}} \right) \cdot \left(-\frac{1}{2\sigma_j^2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4} \right)$$

$$= \left\{ -\frac{1}{2} \sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4} \right\} P(y_t, S_t = j, \theta). \quad (13)$$

So, the equation (10) can be written as:

$$\frac{\partial L(\theta)}{\partial \pi_j} = \sum_{t=1}^{T} \frac{1}{f(y_t, \theta)} f(y_t \mid S_t = j, \theta), \quad (14)$$

$$\frac{\partial L(\theta)}{\partial \mu_j} = \sum_{t=1}^{T} \frac{1}{f(y_t, \theta)} \frac{y_t - \mu_j}{\sigma_j^2} P(y_t, S_t = j, \theta), \quad (15)$$

$$\frac{\partial L(\theta)}{\partial \sigma_j^2} = \sum_{t=1}^{T} \frac{1}{f(y_t, \theta)} \left\{ -\frac{1}{2} \sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4} \right\} \cdot P(y_t, S_t = j, \theta). \quad (16)$$

According to the theorem of Bayes depicting conditional opportunities between the two occurrences of A and B are as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \quad (17)$$

Then it can be done forward filtering ie counting chances of mode $j$based on traffic flow data y_t based on the following equation:

$$P(S_t = j \mid y_t, \theta) = \frac{P(y_t, S_t = j, \theta)}{f(y_t, \theta)}$$

$$= \frac{\pi_j f(y_t \mid S_t = j, \theta)}{f(y_t, \theta)}. \quad (18)$$

Next, backward filtering is performed, namely smoothing process on the probability function that is generated from the forward filtering process. Calculated value $P(S_{t+1} = j \mid y_t, \theta)$ until the value of $P(S_{t+1} = j \mid y_t, \theta)$ is obtained. The equation is used as follows [8]:

$$P(S_{t+1} = j \mid y_t, \theta) = \sum_{j=1}^{2} p_{kj}^{(m)} P(S_t = j \mid y_t, \theta) \quad (19)$$

Then the following computed probability function backward filtering: for $n = N-1, N-2, \dots, 1$ count [8]:

$$P(S_t = j \mid y_T, \theta) = \sum_{j=1}^{2} \left\{ \frac{P(S_t = j \mid y_t, \theta)}{P(S_{n+1} = k \mid y_t, \theta)} \right.$$

$$\left. \cdot P(S_{t+1} = k \mid y_T, \theta)\, p_{jk}^{(m)} \right\} \quad (20)$$

Based on the equation (18), then the equation (14), (15), and (16) can be written to:

$$\frac{\partial L(\theta)}{\partial \pi_j} = \pi_j^{-1} \times \sum_{t=1}^{T} P(S_t = j \mid y_t, \theta) \quad (21)$$

$$\frac{\partial L(\theta)}{\partial \mu_j} = \sum_{t=1}^{T} \frac{y_t - \mu_j}{\sigma_j^2} P(S_t = j \mid y_t, \theta) \quad (22)$$

$$\frac{\partial L(\theta)}{\partial \sigma_j^2} = \sum_{t=1}^{T} \left\{ -\frac{1}{2}\sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4} \right\} \quad (23)$$

$$\cdot P(S_t = j \mid y_t, \theta)$$

The optimal value of the equation Lagrange Multiplier can be found by creating the first instance of the auxiliary function against each delimiter function variable equal to zero. As a result, equations (21), (22) and (23) need to be made equal to zero. By making the equation (22) equal to zero, then:

$$\sum_{t=1}^{T} y_t \times P(S_t = j \mid y_t, \theta) = \mu_j \sum_{t=1}^{T} P(S_t = j \mid y_t, \theta)$$

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} y_t P(S_t = j \mid y_t, \hat{\theta})}{\sum_{t=1}^{T} P(S_t = j \mid y_t, \hat{\theta})}. \quad (24)$$

By making the equation (23) equal to zero, then:

$$\sum_{t=1}^{T} (y_t - \hat{\mu}_j)^2 \times P(S_t = j \mid y_t, \theta) = \sigma_j^2 \sum_{t=1}^{T} P(S_t = j \mid y_t, \theta)$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} (y_t - \hat{\mu}_j)^2 P(S_t = j \mid y_t, \hat{\theta})}{\sum_{t=1}^{T} P(S_t = j \mid y_t, \hat{\theta})} \quad (25)$$

based on equations (9) and (21), then:

$$\frac{\partial J(\theta)}{\partial \pi_j} = \pi_j^{-1} \sum_{t=1}^{T} P(S_t = j \mid y_t, \theta) - \lambda = 0,$$

Or can also be written as:

$$\sum_{t=1}^{T} P(S_t = j \mid y_t, \theta) = \lambda \pi_j. \quad (26)$$

With a sum of equations (26) as much as $j=1,2,\cdots,N,$ generates:

$$\sum_{t=1}^{T}\left[P\left(S_t=j\mid y_t,\theta\right)+\ldots+P\left(S_t=N\mid y_t,\theta\right)\right]=\lambda\left[\pi_1+\ldots+\pi_N\right]$$

or:

$$j=1: \quad \sum_{t=1}^{T}\left[P\left(S_t=1\mid y_t,\theta\right)\right]=\lambda\pi_1$$

$$j=2: \quad \sum_{t=1}^{T}\left[P\left(S_t=2\mid y_t,\theta\right)\right]=\lambda\pi_2$$

$$\ldots$$

$$j=n: \quad \sum_{t=1}^{T}\left[P\left(S_t=n\mid y_t,\theta\right)\right]=\lambda\pi_n \quad +$$

$$\sum_{t=1}^{T}\sum_{j=1}^{n}P\left(S_t=j\mid y_t,\theta\right)=\lambda(\pi_1+\pi_2+\ldots+\pi_n)$$

Because $\sum_{j=1}^{n}P\left(S_t=j\mid y_t,\theta\right)$ and $\pi_1+\pi_2+\ldots+\pi_n$ are worth 1, then:

$$\sum_{t=1}^{T}\{1\}=\lambda(1)$$

$$\underbrace{1+1+\ldots+1}_{T}=\lambda(1)$$

$$T\cdot 1=\lambda(1)$$

$$T=\lambda$$

By substituted λ with T on the equation (26), it generates the equation:

$$\hat{\pi}_j=T^{-1}\sum_{t=1}^{T}P\left(S_t=j\mid y_t,\hat{\theta}\right) \qquad (27)$$

Transition Probability Matrix (TPM) P created in this study was a 2 × 2-sized matrix, indicating the probability of switching between mode. To look for the TPM value, it takes the probability value generated from the backward filtering process $P(S_{t+1}=j\mid y_t,\theta)$ and $P(S_t=j\mid y_T,\theta)$ for each of the entire data generated from the observation process, except for the data from the first minute. It also takes TPM values from previous iterations of the EM algorithm. With this information, in each iteration of the EM algorithm it is possible to estimate the value of the TPM element in row j and column k with the following equation taken from reference (Sutarto, Y, H., and Joelianto, E., 2015) with minor modifications after discussion with the reference author:

$$p_{jk}^{(m)}=\frac{\sum_{t=2}^{T}P\left(S_t=k\mid y_T,\theta\right)\dfrac{p_{jk}^{(m-1)}P\left(S_{t-1}=j\mid y_t,\theta\right)}{P\left(S_t=k\mid y_{t-1},\theta\right)}}{\sum_{t=2}^{T}P\left(S_{t-1}=j\mid y_T,\theta\right)}. \qquad (28)$$

The process estimation parameter with the EM algorithm that has been described is done with the help of the computer. At each iteration, the first calculation of the function log likelihood $L(\theta)$ by equation (8). Specifically, for the first iteration, used the parameters μ_j, σ_j^2, and π_j. Which is initial condition. Initial condition to μ_j and σ_j^2 is determined from the average and the variance of data from the observation result of the first observation point for each mode. Initial condition for π_j is considered balanced for both modes, so π_j respectively, that the respective mode is 0.5. Initial condition for The Matrix P is considered balanced for the whole possible displacement of mode, so the Initial condition for the entire matrix element p is worth 0.5. The Initial condition that is used in the process of the estimation parameter is shown in table 1. Performed forward filtering with equations (18) and backward filtering with equations (20). Then, it is done with the calculation to find the values μ_j, σ_j^2, and π_j, new using the equation (24), (25), and (27). Also calculated TPM values with equations (28). Then the value is recalculated and calculated the difference with the value $L(\theta)$ from the previous iteration. The iteration process is repeated continuously until the difference in value $L(\theta)$ with the value $L(\theta)$ (threshold). This study used the threshold of 0.00001 due to multiple attempts, after a difference in value of $L(\theta)$ is no larger than the threshold value of the estimated parameters no longer experiencing a change in the value Significant.

2.3. STATE ESTIMATION

After getting the parameters of each mode and making the initial estimate without estimator based on those parameters, the next step that needs to be done is the state estimation process (state estimation). State estimation is the process of looking for the best aproxation regarding State of a system based on known information. In this study, State estimation was used to get estimated traffic flows based on the value of parameters that have been generated by the EM algorithm. The process state estimation is generally done with the help of estimator. The estimator of used in this study is a sequence of Particle Filter Observation and transition-based most likely modes Tracking Particle Filter (OTPF). OTPF is a Particle Filter variant that is suitable for a system consisting of several modes. OTPF considers mode to as an unknown parameter. Therefore, on OTPF there is a stage that does not exist in the usual Particle Filter, which is the stage to do the estimation of mode that is most likely to occur from any Particle. Once the approximate of is being obtained is most likely to occur, the system is assumed to follow the dynamics of the corresponding mode. The next stages of Particle Filter are done as usual (Kusuma, A, T., and Suparman., 2014).

Mathematically, the OTPF algorithm can be written as follows (Kusuma, A, T., and Suparman., 2014):

1) Initialization

- $t=0$. The mode at time 0 is describe as m_0. For $i=1,\ldots,N$, the ustron $x_0^{(1)}$ from the initial condition *(initial condition)* in the m_0 mode then set $t=1$.

2) Prediction

- *For each mode* m_t^j *which has a displacement probability* $T_{m_{t-1}m_t^j}$ *from mode* m_{t-1} *to* m_t^j *that value is not equal to zero (* $j=1,\ldots,K$ *with* K *represent the number of modes), sampling* $\tilde{x}_t^{(i)} \sim p_{m_t^j}\left(x_t \mid x_{t-1}^{(1)}\right)$*, for* $i=1,\ldots,N$.
- *For each mode* m_t^j*, it is calculated importance weight* $\tilde{w}_t^{(i)} \sim p_{m_t^j}\left(y_t \mid \tilde{x}_t^{(1)}\right)$*, for* $i=1,\ldots,N$.

3) Mode Selection

- Calculated the average *particle* weight in each *mode* m_t^j then multiply by the probability of *mode* switching: $\tilde{w}_t^{m_t^j} = T_{m_{t-1}m_t^j}\sum_{i=1}^{N}\tilde{w}_t^{(i)}/N$.
- Find the most likely *mode*: $m_t = \arg\max_{m_t^j}\left\{\tilde{w}_t^{m_t^j}\right.$:for all $\left. m_t^j, j=1,\ldots,K\right\}$.

- Normalize the weight of particle in *mode* m_t^j.

4) Resampling

- *Resample* N new *particle* $\{x_t^{(i)},$ to $i=1,\dots,N\}$ replace the *particle* in *mode* $m_t\{\tilde{x}_t^{(i)},$ to $i=1,\dots,N\}$ base on *importance weight.*
- Set $t=t+1$ then return to stage 2.

3. RESULTS AND DISCUSSION

This section explains the testing of the accuracy of the model by applying the model to estimate the actual traffic flow data on Dipatiukur Road, Bandung. To find out the accuracy of the resulting estimation, the value of error estimation is calculated using the MAPE method

3.1. RETRIEVAL OF DATA

At this stage traffic flow data is collected on the observed road section. The data collection process was carried out on Monday 20 September 2017 at 06.00-10.00. Data is collected by counting the number of vehicles that pass each minute manually with the help of the application counter contained on Android smartphone. Observation of the traffic flow is carried out for both directions of traffic on the observed road, i.e. the flow from south to north and flow from north to south. Observations are made without distinguishing the types of vehicles that pass, so that all vehicles that pass the observation point are included in the calculation. Observations were made at two ends of the road that were observed simultaneously. The traffic flow data per minute is collected into traffic flow data per three minutes. By selecting a longer data interval, it can be minimized the occurrence of random noise in the data so as to reduce the value of error resulting in the estimation process (Yao-Jan., et al, 2012), (Moon, K, T., 1996). The first observation point is in front of the ITHB campus while the second observation point is in front of Indomaret Point Dipatiukur. Location of the observation of traffic flow can be seen in Fig. 1.

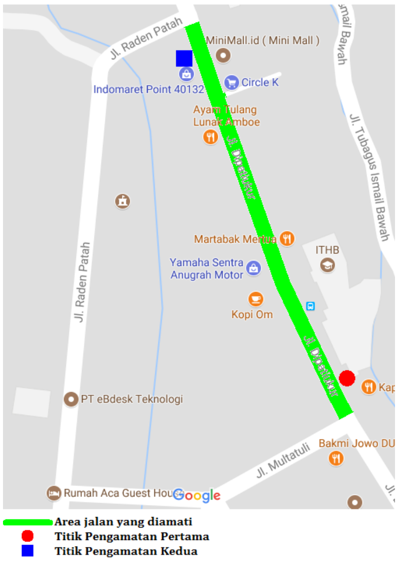

Fig. 1. Location of Traffic flow observations

3.2. MODEL TESTING

After the traffic flow data on the observed road sections have been successfully collected, it can be tested on the modeling methods designed in this study. The EM algorithm is applied to estimate the parameters of traffic flow data to the north and south. The parameters generated are shown in Table 1. After the required parameters have been obtained from the parameter estimation process, then we can estimate the traffic flow using OTPF with 500 particles. The estimation results are shown in Figs. 2 and 3. Visually it can be seen that the resulting estimation results are quite close to the traffic flow data from observations. In this study, to determine the numerical accuracy of estimation values, the calculation of the error value estimate using the method Mean Absolute Percentage Error (MAPE). In general, the amount of MAPE can be used to classify the accuracy of a forecast into four levels, namely (Kumar, S, V., 2017):

- If MAPE is less than 10%, the accuracy is highly accurate
- If MAPE is between 11% and 20%, the accuracy is good.
- If MAPE is between 21% to 50%, the accuracy is reasonable.
- If MAPE is more than 51%, the accuracy is included in inaccurate category.

The estimated traffic flow generated in this study has a MAPE value of 7.7978% for traffic flow from south to north and 5.8547% for traffic flow from north to south. Both of the MAPE values are below 10%, so the estimation accuracy generated by the model belongs to the category highly accurate.

Table 1. Parameter values generated from the EM algorithm.

Parameter	Traffic Flow to the North		Traffic Flow to the South	
	***Mode* 1**	***Mode* 2**	***Mode* 1**	***Mode* 2**
μ_j	83,4794	90,3488	76,0913	78,1023
σ_j^2	183,9877	1208,9447	171,9612	17,0912
π_j	0,7040	0,2960	0,7719	0,2281
p_{jk}	$\begin{bmatrix} 0{,}9995 & 0{,}0005 \\ 0{,}9897 & 0{,}0103 \end{bmatrix}$		$\begin{bmatrix} 0{,}9893 & 0{,}0107 \\ 1{,}0000 & 0{,}0000 \end{bmatrix}$	

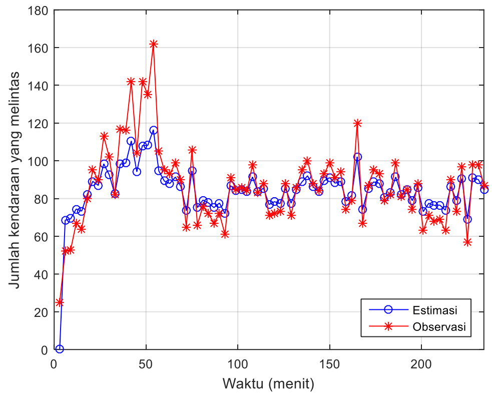

Fig. 2. Comparison of the results of the estimation experiment using OTPF with the observed data from south to north traffic flow

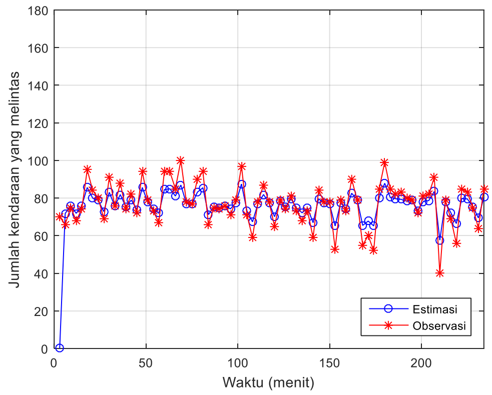

Fig. 3. Comparison of the results of the estimation experiment using OTPF with the observed data from north to south traffic flow

4. CONCLUSION

Based on parameter estimation, estimation state and system testing that has been carried out in this research, it can be concluded that the model compiled from mixed Gaussian distribution with parameters generated by parameter estimation with EM algorithm can already accurately describe the traffic flow. With an estimated MAPE value of 7.7978% and 5.8547%, it can be said that the estimate has a high accuracy, and includes the category highly accurate. The MAPE value is even lower than the traffic flow model produced by other studies that are more complex and require more data.

There are still a few gaps that allow even more accurate modeling. Some of them are by increasing the number of modes, extending data intervals, and increasing the amount of data used in the model testing process. Models can also be drawn that describe dynamic systems such as the Autoregressive (AR) model to get estimates with higher accuracy.

REFERENCES

A.S. Wiyono. (2014, Okt 22). Bandung kota termacet ketujuh, ini penyebabnya [Online]. Available: https://www.merdeka.com/peristiwa/bandung-kota-termacet-ketujuh-ini-penyebabnya.html

Ali, M. (2011, April). Intelligent Transport System (ITS). In *10 th Research Seminar Series Workshop*.

Clark, S. (2003). Traffic prediction using multivariate nonparametric regression. *Journal of transportation engineering*, 129(2), 161-168.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22.

Kumar, S. V. (2017). Traffic flow prediction using Kalman filtering technique. *Procedia Engineering*, *187*, 582-587.

Kusuma, T, A., & Suparman. (2014) Algoritma Expectation-Maximization (EM) Untuk Estimasi Distribusi Mixture. *Jurnal Konvergensi*, 4(2), 65-83.

Marinică, N., & Boel, R. (2012, June). Platoon based model for urban traffic control. In *2012 American Control Conference (ACC),* 6563-6568.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6), 47-60.

Salim, T., Hutagalung, M., & Sutarto, H. Y. (2019). Simulasi Arus Lalu Lintas Menggunakan Perangkat Lunak Simulink. *Jurnal Telematika*, 13(1), 7-12.

Sutarto, H. (2016). *Adaptive control for traffic signals using a stochastic hybrid system model* (Doctoral dissertation, Ghent University).

Sutarto, H., & Joelianto, E. (2015). Expectation-maximization based parameter identification for HMM of urban traffic flow. *International Journal of Applied Mathematics & Statistics*, 53(2), 90-101.

Tafazoli, S., & Sun, X. (2006). Hybrid system state tracking and fault detection using particle filters. *IEEE Transactions on Control Systems Technology*, 14(6), 1078-1087.

Vandaele, N., Van Woensel, T., & Verbruggen, A. (2000). A queueing based traffic flow model. *Transportation Research Part D: Transport and Environment*, 5(2), 121-135.

Wu, Y. J., Chen, F., Lu, C., Smith, B., & Chen, Y. (2012). Traffic flow prediction for urban network using spatio-temporal random effects model. In *91st Annual Meeting of the Transportation Research Board (TRB)*.