Fundamental Diagram Estimation Using GPS Trajectories of Probe Vehicles

Yutaka Kawasaki, Toru Seo, Takahiko Kusakabe, and Yasuo Asakura

***Abstract*—A fundamental diagram (FD), which determine relation among flow, destiny, and speed of traffic, is one of the most important principles in traffic flow theory. Conventionally, the FD has been estimated by measurements from detectors which are fixed on roads. Therefore, it has been difficult to estimate FDs over wide spatial domain, due to high installation and maintenance cost of detectors. In this study, a method of estimating FD using measurements from GPS-equipped probe vehicles (also known as connected vehicles) is developed. Specifically, the proposed method estimate values of free-flow speed and backward wave speed of so-called triangular FD based on trajectories of multiple probe vehicles which are randomly sampled. Since probe vehicles can collect data from wide spatial domain, it can be expected that the proposed method enable to estimate FD continuously in large area. The proposed method was verified by using actual traffic data in an urban expressway.**

I. Introduction

A *fundamental diagram (FD)* is one of the most important principles in traffic flow [1]. It is theoretical relation between traffic flow and density in stationary traffic, and determines important characteristics of traffic, such as free-flow speed and traffic capacity. Conventionally, the FD has been estimated by measurements from cameras or detectors which are fixed on roads [2]. Therefore, it has been difficult to estimate FDs over wide spatial domain, due to high installation and maintenance cost of detectors. If an FD is estimated based only on data obtained by *probe vehicles* (which typically equip global positioning system (GPS) and measure their own trajectories), an FD can be estimated continuously wherever probe vehicles traveled. Since the availability of probe vehicles keeps increasing these days because of technology advances such as mobile phones and connected vehicles [3], we can expect that such probe-vehicle-based FD estimation method would be used to estimate FDs over almost entire road network. Furthermore, such method can also promote novel applications of probe vehicle data which require FDs, such as traffic state estimation [4]–[7]

Only limited studies investigated the FD estimation problem based on probe vehicle data. Herrera et al. [3] calculated values of variables of FD from particular vehicles trajectories by manual operation by assuming the value of jam density. Sun et al. [8] and Seo et al. [9] proposed methods to simultaneously estimate an FD and traffic state based on probe vehicles and additional sensors, namely, fixed detectors and spacing information, respectively. Seo et al. [10] proposed an algorithm to estimate an FD based on GPS probe vehicle data by assuming the value of jam density; however, the method has some theoretical limitation (i.e., the method's parameter estimation method was heuristic) and had not been validated by actual data. Zhu and Ukkusuri [11] investigated an estimation problem of congested part of Lagrangian FD of connected vehicles.

The aim of this paper is to develop an algorithm of estimating FDs based only on probe vehicle trajectory data by improving the preliminary method of Seo et al. [10], and to validate the algorithm by using actual traffic data. The rest of this paper is organized follows. Section II develops the algorithm of estimating FD based only on probe vehicle trajectories. Section III validates the proposed algorithm using actual traffic data. Section IV concludes this paper.

II. Estimation algorithm

A. Basic idea

The basic idea of the proposed algorithm, which is almost identical to our preliminary study [10], is as follows.[1] The subjects of estimation are values of free-flow speed u and backward wave speed w of a triangular FD[2] (the black curve in Figure 1). The jam density κ is assumed exogenously, as it can be easily inferred by external knowledge such as average vehicle body length.

Suppose that two probe vehicles are driving the same road with homogeneous FD; and $c-1$ vehicles exist between the probe vehicles. If traffic in a time–space region between two probe vehicles is *stationary* (c.f., stationary traffic is defined as traffic in which all the vehicles have the same constant speed and spacing [1]), a traffic state (i.e., flow, density, and speed) in the region follows the FD by the definition. If the region is stationary and the distance traveled by the probe vehicles are the same, a traffic state of probe vehicles only in the region (i.e., *probe traffic state: PTS*) follows a curve which is similar to the FD with $1/c$ scale (i.e., *probe-FD*; the red curve in Figure 1).

If such measurement data are collected by multiple probe vehicles which have driven various traffic conditions (i.e., free-flowing and congested traffic), the probe-FD can be identified. Finally, since the probe-FD is similar to the FD with $1/c$ scale and the value of the FD's jam density is already known, the

Y. Kawasaki, T. Seo and Y. Asakura are with Tokyo Institute of Technology. T. Kusakabe is with the Center for Spatial Information Science, at the University of Tokyo. Email: y.kawasaki@plan.cv.titech.ac.jp (Y. Kawasaki) t.seo@plan.cv.titech.ac.jp (T. Seo)

Part of this work was financially supported by the Japan Society for the Promotion of Science (KAKENHI Grant-in-Aid for Young Scientists (B) 16K18164).

[1] The difference between the current method and the previous method is explained in II-B.

[2] Note that the proposed methodology is easy to extend to consider other FD's functional forms which is continuous. Non-continuous FD functions (e.g., one with capacity drop), however, would be difficult because they often have curves which are not equality constrained by the jam density.

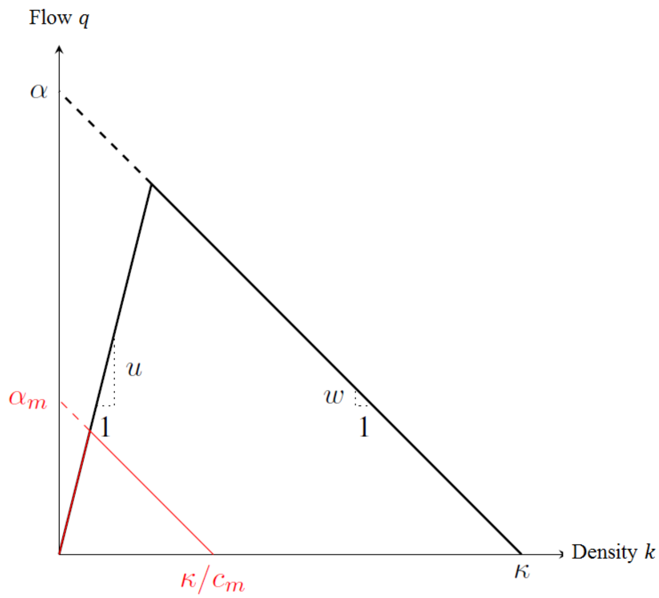

Fig. 1. A triangular FD and its parameters.

free-flow speed and the critical density can be identified. For the details on this idea with mathematical representations, see Section 2 of Seo et al. [10].

B. Overview of the proposed algorithm

The proposed algorithm estimate FDs based on the aforementioned basic idea. Specifically, the algorithm estimates value of free-flow speed u and backward wave speed w of a triangular FD, based on trajectories of randomly sampled vehicles and some parameters. An example of triangular FD is shown in Figure 1. It has three independent parameters, namely, u, w, and the inverse of driver reaction time α, which is q-intercept of the backward wave curve of the FD ($\alpha = w\kappa$).

The algorithm consists of two steps: namely, PTS calculation step (II-C) and FD estimation step (II-D). The former step construct an input data based on probe vehicle data, and the latter step estimates the FD based on the input data.

Compared to the method of Seo et al. [10], the proposed algorithm has two notable improvements. First, because of the improved PTS calculation step, the current algorithm is expected to extract more information from probe vehicle data than the previous method. Second, the current FD estimation step is based on an Expectation–Maximization (EM) algorithm, which is a significantly sophisticated algorithm compared with heuristic residual minimization algorithm employed by the previous method. For example, convergence to a local optimum is guaranteed. Moreover, distribution of the observation error can be properly considered.

C. Probe traffic state calculation step

This step calculates PTS probe vehicles' trajectory following the procedure described below:

Step 1 Define a time–space region $\mathbf{A}$ whose FD is to be estimated.

Step 2 Select probe m from probe set $\mathbf{P}(\mathbf{A})$.

Step 3 Do:

Step 3.1 Select probe $m - \Delta m$ from $\mathbf{P}(\mathbf{A})$. Define an empty set $\mathbf{A}'(m, \Delta m, \mathbf{A})$.

Step 3.2 At time t, define a time–space region $\mathbf{a}^{(t)}_{m,\Delta m}$ which is surrounded by trajectories of probe m and $m-\Delta m$, the strait line $(t, X(t,m))$–$(t', X(t', m-\Delta m))$ whose slope is ϕ, and the strait line $(t, X(t,m))$–$(t', X(t', m-\Delta m))$ whose slope is ϕ (Figure 2).

Step 3.3 Determine whether traffic in $\mathbf{a}^{(t)}_{m,\Delta m}$ is stationary based on following equations:

$$\frac{\sigma(\mathbf{a}^{(t)}_{m,\Delta m})}{\overline{v}(\mathbf{a}^{(t)}_{m,\Delta m})} \leq \theta \tag{1a}$$

$$q_m(\mathbf{a}^{(t)}_{m,\Delta m}) \geq 0 \tag{1b}$$

$$k_m(\mathbf{a}^{(t)}_{m,\Delta m}) \geq 0 \tag{1c}$$

If (1) holds true, then it is stationary; go to Step 3.4. If not, go to Step 3.5.

Step 3.4 Add PTS $S_m(\mathbf{a}^{(t)}_{m,\Delta m}) = (q_m(\mathbf{a}^{(t)}_{m,\Delta m}), k_m(\mathbf{a}^{(t)}_{m,\Delta m}))$ to set $\mathbf{A}'(m, \Delta m, \mathbf{A})$.

Step 3.5 If $\mathbf{a}^{(t)}_{m,\Delta m}$ can be defined based for another t, go back to Step 3.2. If not, go to Step 4.

Step 4 If another m can be selected, go back to Step 2. If not, terminate the PTS calculation step.

The notation is as follows:

$X(t, m)$ is a position of probe m at time t.
$\mathbf{A}$ is a time–space region whose FD is to be estimated.
$\mathbf{P}(\mathbf{A})$ is a set of all the probe in $\mathbf{A}$.
m is a probe vehicle. The preceding probe of m is denoted as $m-1$.
$q_m(\mathbf{a}^{(t)}_{m,\Delta m}), k_m(\mathbf{a}^{(t)}_{m,\Delta m}), v_m(\mathbf{a}^{(t)}_{m,\Delta m})$ are PTS, defined as (2) and (3)
$d_m(\mathbf{a})$ is total distance traveled by probe m in region $\mathbf{a}$.
$t_m(\mathbf{a})$ is total time spent by probe m in region $\mathbf{a}$.
$c_{m,\Delta m}$ is the total number of (probe and non-probe) vehicles between m and $m-\Delta m$ in $\mathbf{A}$ which include m but don't include $m - \Delta m$.
$q(\mathbf{a}), k(\mathbf{a})$ are actual traffic state in region $\mathbf{a}$, defined by the Edie's definition.
$S_m(\mathbf{a})$ is the PTS vector $(q_m(\mathbf{a}), k_m(\mathbf{a}))$.
$\acute{\mathbf{A}}(m, \Delta m, \mathbf{A})$ is a set of subregions $\mathbf{a}^{(t)}_{m,\Delta m}$ for appropriate t and m which are stationary in $\mathbf{A}$.
$\mathbf{a}^{(t)}_{m,\Delta m}$ is a time–space region surrounded by trajectories of probe m and $m - \Delta m$, the strait line $(t, X(t,m))$–$(t', X(t', m-\Delta m))$ whose slope is ϕ, and the strait line $(t, X(t,m))$–$(t', X(t', m-\Delta m))$ whose slope is ϕ (Figure 2).
$|\mathbf{a}|$ is the area of time–space region $\mathbf{a}$.
$\overline{v}(\mathbf{a}^{(t)}_{m,\Delta m})$ is a mean velocity of probe vehicles m and $m-\Delta m$ in $\mathbf{a}^{(t)}_{m,\Delta m}$.
$\sigma(\mathbf{a}^{(t)}_{m,\Delta m})$ is a standard deviation of velocity of probe vehicles m and $m - \Delta m$ in $\mathbf{a}^{(t)}_{m,\Delta m}$.

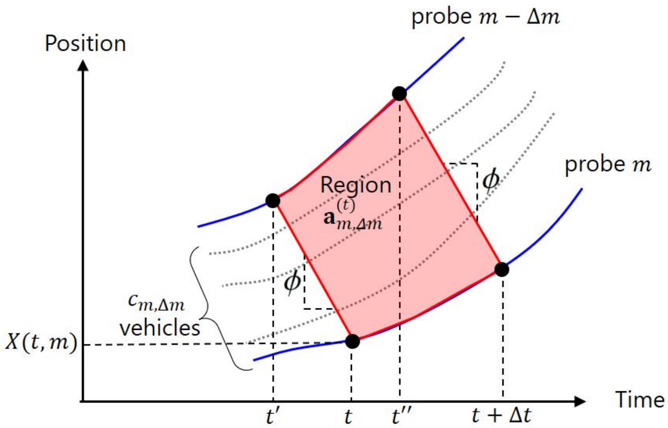

Fig. 2. Illustrated definition of time–space region $\mathbf{a}_{m,\Delta m}^{(t)}$ for probe m and time t.

The PTS is originally defined as

$$q_m(\mathbf{a}_{m,\Delta m}^{(t)}) = q(\mathbf{a}_{m,\Delta m}^{(t)})/c_{m,\Delta m}, \tag{2a}$$

$$k_m(\mathbf{a}_{m,\Delta m}^{(t)}) = k(\mathbf{a}_{m,\Delta m}^{(t)})/c_{m,\Delta m}, \tag{2b}$$

$$v_m(\mathbf{a}_{m,\Delta m}^{(t)}) = q_m(\mathbf{a}_{m,\Delta m}^{(t)})/k_m(\mathbf{a}_{m,\Delta m}^{(t)}), \tag{2c}$$

which is identical to the following Edie's definition [12]

$$q_m(\mathbf{a}_{m,\Delta m}^{(t)}) = d_m(\mathbf{a}_{m,\Delta m}^{(t)})/|\mathbf{a}_{m,\Delta m}^{(t)}|, \tag{3a}$$

$$k_m(\mathbf{a}_{m,\Delta m}^{(t)}) = t_m(\mathbf{a}_{m,\Delta m}^{(t)})/|\mathbf{a}_{m,\Delta m}^{(t)}|, \tag{3b}$$

in a stationary $\mathbf{a}$. Notice that (2) cannot be known because of unknown $q(\mathbf{a})$, $k(\mathbf{a})$, $c_{m,\Delta m}$, whereas (3) can be calculated based on the available probe data.

This step requires four (hyper-)parameters: θ, Δm, Δt, and ϕ. θ is used in Step 3.3 to determine whether stationary. Δm is used in Step 3.1 to select probe pair. Δt is used in Step 3.2 to determine size of $\mathbf{a}_{m,\Delta m}^{(t)}$. ϕ is used in Step 3.2 to determine slope of $\mathbf{a}_{m,\Delta m}^{(t)}$. Among these, only θ would be essential to the algorithm, meaning that the value should be determined carefully. This is because it is possible to assume several values for other parameters, and execute the PTS calculation step to generate one dataset to feed one FD estimation step. It can be expected that if θ is small, more accurate PTS in terms of stationarity will be obtained; however, amount of data will be small.

D. FD estimation step

This step estimate u and w by using the PTS calculated by the previous step. It is based on an EM algorithm, which maximizes a likelihood function based on some observable variables where some of variables are latent (unobservable) by iteratively executing the expectation step (E-step) and maximization step (M-step) [13]. For the current problem, the actual traffic states and FD parameters correspond to the latent variables for the EM algorithm, whereas the PTS corresponds to the observable variables.

We assume that likelihood l of state s (whether free-flowing F or congested C) in $\mathbf{a}_{m,\Delta m}^{(t)}$ from probe pair m and $m-\Delta m$ follows a normal distribution whose variance is σ:[3]

$$l(m, \mathbf{a}, s) = \begin{cases} \mathcal{N}(q_{m\mathbf{a}}, uk_{m\mathbf{a}}, \sigma) & \text{if } s = \mathrm{F}, \\ \mathcal{N}(q_{m\mathbf{a}}, \alpha_m - wk_{m\mathbf{a}}, \sigma) & \text{if } s = \mathrm{C}. \end{cases} \tag{4}$$

where, $\mathcal{N}$ is probability density function of normal distribution, s is variable which indicates the state ($s \in \{\mathrm{F}, \mathrm{C}\}$), F represents free-flowing state, C represents Congested state, and $q_{m\mathbf{a}}$ and $k_{m\mathbf{a}}$ are identical to $q_m(\mathbf{a})$ and $k_m(\mathbf{a})$ obtained by the PTS calculation step. The likelihood L of the mixture distribution is given by

$$L(q, k|\pi, u, w, \{\alpha_m\}\sigma) = \prod_{m \in M'} \prod_{\mathbf{a} \in \mathbf{A}'_m} \sum_{s \in \{\mathrm{F},\mathrm{C}\}} \pi_s l(m, \mathbf{a}, s), \tag{5}$$

where α_m is the q-intercept of line of probe-FD's congested area of probe m ($\alpha_m = w\kappa/c_m$, see Figure 1), $\{\alpha_m\}$ is $\{\alpha_m \mid m \in M'\}$, M' is the set of all m which have a probe pair $(m, m-\Delta m)$, π_s is contribution ratio of mixture distribution ($\sum_{s \in \{\mathrm{F},\mathrm{C}\}} \pi_s = 1$). Equation (5) indicates a likelihood when q_m, k_m is observed, while parameters $\pi, u, w, \{\alpha_m\}, \sigma$ are latent. Therefore, the FD parameters $u, w, \{\alpha_m\}$ can be estimated by maximizing (5).

If the complete data set are available (i.e., latent variables $z_{m\mathbf{a}s}$ which equal to 1 if $s = (\text{state of } S_{m\mathbf{a}})$ and 0 otherwise are available), the likelihood can be expressed as

$$\begin{aligned} &L(q, k, z|\pi, u, w, \{\alpha_m\}, \sigma) \\ &= \prod_{m \in M'} \prod_{\mathbf{a} \in \mathbf{A}'_m} \prod_{s \in \{\mathrm{F},\mathrm{C}\}} \pi_s^{z_{m\mathbf{a}s}} l(m, \mathbf{a}, s)^{z_{m\mathbf{a}s}}. \end{aligned} \tag{6}$$

By taking the expectation of logarithm of (6), we obtain

$$\begin{aligned} &E[LL(q, k, z|\pi, u, w, \{\alpha_m\}, \sigma)] \\ &= \sum_{m \in M'} \sum_{\mathbf{a} \in \mathbf{A}'_m} \sum_{s \in \{\mathrm{F},\mathrm{C}\}} p_{m\mathbf{a}}(s) \left(\ln \pi_s + \ln l(m, \mathbf{a}, s)\right), \end{aligned} \tag{7}$$

where $p_{m\mathbf{a}}(s)$ is the expectation of $z_{m\mathbf{a}s}$ (i.e., probability of probe datum $m\mathbf{a}$ belonging state s).

Now, we can estimate u and w by maximizing (7). An EM algorithm for the maximization can be constructed as follows:

Step 5 Set initial values for $\pi_s, u, w, \{\alpha_m\}, \sigma$.
Step 6 Do:
- **Step 6.1 E-step** Calculate $p_{m\mathbf{a}}(s)$ based on current values of π_s, u, w, $\{\alpha_m\}$, σ by (8) in Section II-D1.
- **Step 6.2 M-step** Calculate π_s, u, w, $\{\alpha_m\}$, σ based on current values of $p_{m\mathbf{a}}(s)$ by (11), (12), (15), (16) in Section II-D2.
- **Step 6.3** Test the convergence. If $|x^{(n)} - x^{(n-1)}| \leq \varepsilon, \forall x \in \{p_{m\mathbf{a}}(s), \pi_s, u, w, \{\alpha_m\}, \sigma_s\}$ (n indicates the current iteration, $n-1$ indicates the previous iteration, $x^{(n)}$ indicates a parameter value of iteration n, ε is a criteria) is satisfied, then it has converged and go to Step 7. If not, go back to Step 6.1.

Step 7 Output the current values of u, w.

[3]Hereafter, we omit obvious super/subscripts for simplicity. For example, $_{m,\Delta m}$ is abbreviated as $_m$.

1) E-step: It calculates $p_{m\mathbf{a}}(s)$ from $\pi_s, u, w, \{\alpha_m\}, \sigma$ by

$$p_{m\mathbf{a}}(s) = \frac{\pi_s l(m, \mathbf{a}, s)}{\sum_{r\in\{\mathrm{F},\mathrm{C}\}} \pi_r l(m, \mathbf{a}, r)}, \quad \forall s \in \{\mathrm{F},\mathrm{C}\},\ \forall \mathbf{a} \in \mathbf{A}'_m,\ \forall m \in M'. \tag{8}$$

2) M-step: It finds values of $\pi_s, u, w, \{\alpha_m\}, \sigma$ which maximize (7). We employ the method of Lagrange multiplier, where the Lagrangian is defined as

$$\begin{aligned}
&\mathcal{L}(q, k, z|\lambda, \pi, u, w, \{\alpha_m\}, \sigma)\\
&= \sum_m \sum_{\mathbf{a}} \sum_s p_{m\mathbf{a}}(s) \left(\ln \pi_s + \ln l(m, \mathbf{a}, s)\right) - \lambda \left(\sum_s \pi_s - 1\right)\\
&= \sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{F}) \left(\ln \pi_\mathrm{F} - \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(q_{m\mathbf{a}} - uk_{m\mathbf{a}})^2\right)\\
&+ \sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) \left(\ln \pi_\mathrm{C} - \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(q_{m\mathbf{a}} - \alpha_m + wk_{m\mathbf{a}})^2\right)\\
&- \lambda\left(\sum_s \pi_s - 1\right)
\end{aligned} \tag{9}$$

and the Lagrangian multiplier is λ.

Following the method of Lagrange multiplier, we can obtain π_s by solving

$$\frac{\partial \mathcal{L}}{\partial \pi_s} = \frac{\sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(s)}{\pi_s} - \lambda = 0, \qquad \forall s \in \{\mathrm{F},\mathrm{C}\}, \tag{10}$$

thus

$$\pi_s = \frac{\sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(s)}{\sum_{r\in\{\mathrm{F},\mathrm{C}\}} \sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(r)}, \qquad \forall s \in \{\mathrm{F},\mathrm{C}\}. \tag{11}$$

Similarly, as for u, we obtain

$$u = \frac{\sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{F}) q_{m\mathbf{a}} k_{m\mathbf{a}}}{\sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{F}) k_{m\mathbf{a}}^2}. \tag{12}$$

We can obtain w and $\{\alpha_m\}$ by solving

$$\begin{cases}
\sum_m \left(\sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) k_{m\mathbf{a}}\right) \alpha_m - \left(\sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) k_{m\mathbf{a}}^2\right) w \\
\quad = \sum_m \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) q_{m\mathbf{a}} k_{m\mathbf{a}}, & \text{(13a)}\\
\left(\sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C})\right) \alpha_m - \left(\sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) k_{m\mathbf{a}}\right) w \\
\quad = \sum_{\mathbf{a}} p_{m\mathbf{a}}(\mathrm{C}) q_{m\mathbf{a}}, \qquad \forall m \in M'. & \text{(13b)}
\end{cases}$$

These $|M'| + 1$ equations with $|M'| + 1$ unknown variables can be represented by using matrix:

$$A(w,\ \alpha_1,\ \alpha_2,\ \ldots,\ \alpha_{|M'|})^\mathsf{T} = \boldsymbol{b}. \tag{14}$$

If A is not singular, we obtain

$$(w,\ \alpha_1,\ \alpha_2,\ \ldots,\ \alpha_{|M'|})^\mathsf{T} = A^{-1}\boldsymbol{b}. \tag{15}$$

As for σ, we obtain

$$\sigma^2 = \frac{\sum_m \sum_{\mathbf{a}} \left(p_{m\mathbf{a}}(\mathrm{F})(q_{m\mathbf{a}} - uk_{m\mathbf{a}})^2 + p_{m\mathbf{a}}(\mathrm{C})(q_{m\mathbf{a}} - \alpha_m + wk_{m\mathbf{a}})^2\right)}{2\sum_m \sum_{\mathbf{a}} \sum_s p_{m\mathbf{a}}(s)}. \tag{16}$$

Note that above solution does not exist if input data are insufficient. For example, if probe vehicle data only contains measurements from free-flowing traffic, A will be singular; thus w cannot be determined. If probe vehicle data only contains measurements from congested traffic, (12) will be indefinite; thus u cannot be determined. In fact, the sufficient condition for the existence of the solution is identical to that of Seo et al. [10]. It means that data from various traffic condition must be collected to utilize the proposed algorithm. In practice, such data can be obtained by collecting probe vehicle data on a road with recurrent congestion for a long time; such data collection would not be difficult in recent days [3].

This step also requires some parameter values, namely, ε and initial values for $\pi_s, u, w, \{\alpha_m\}$, and σ, as in usual EM algorithms. Since the objective function is not necessarily convex and the EM algorithm only finds local optimum, it is necessary to test multiple initial value sets and compare each result in order to find the global optimum. It means that the computation cost can be large to obtain an accurate result. The value of criteria for convergence, ε, also affects the computation cost.

III. Validation

In this section, the proposed algorithm is empirically validated by using actual trajectories collected by GPS-equipped probe vehicles on an urban expressway.

A. Data

GPS-equipped probe vehicle data collected by [14] was employed for the validation. The data contains trajectories of 86 probe vehicles, which corresponds to 3.5% penetration rate according to comparison with the detector data. The probe vehicles have recorded their position for every 5 seconds. The road segment for the validation is 3860m length section between Hamazakibashi junction to Edobashi junction on the Inner Circular route (an urban expressway in central Tokyo) where no merge/diverge sections except for on/off-ramps exist. The flow from/to on/off-ramps were almost negligible during the experiment. The speed regulation on the segment was 50

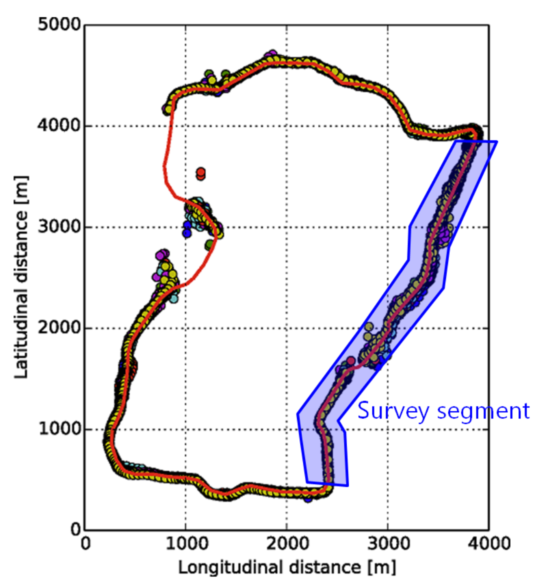

Fig. 3. Trajectories collected by GPS-equipped probe vehicles: top-view (adopted from [14]).

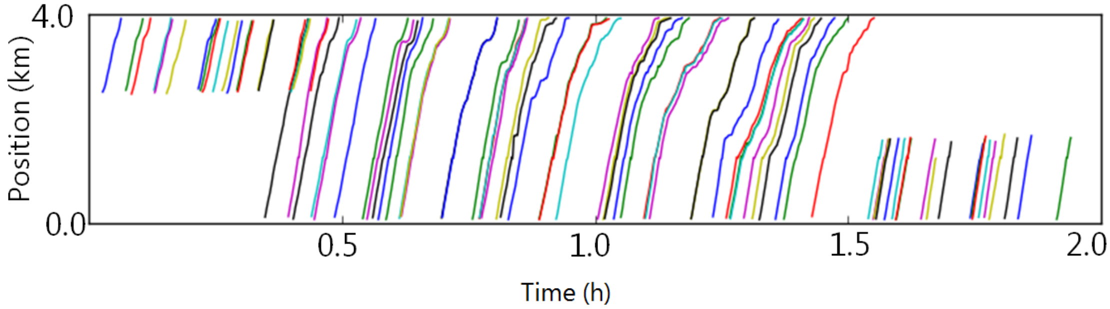

Fig. 4. Trajectories collected by GPS-equipped probe vehicles: time–space diagram on the survey segment.

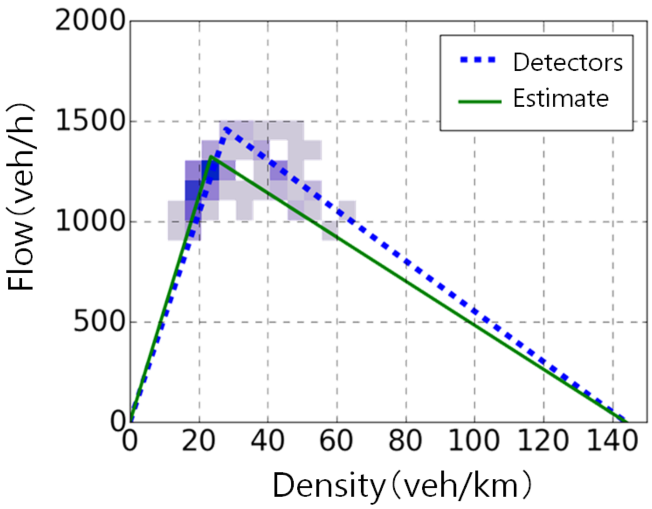

Fig. 5. An FD estimation result.

km/h. The time duration is two hours, from 14:30 to 16:30 on September 24th, 2013. The raw GPS data is shown in Figure 3 along with a map of the Inner Circular Route. The map-matched data on the survey segment is shown in Figure 4 as a time–space diagram. Detectors are densely placed on this segment (roughly 200 m spacing), and measure instantaneous speed and flow with 1-min resolution for each lane; they were used for reference in this validation. The traffic have consisted of free-flowing and congested traffic. According to the detector data, the reference FD parameters have been derived by residual regression as follows: $u = 52.15$ (km/h), $w = 12.57$ (km/h), and $\kappa = 144.0$ (veh/km).

B. Result

An estimation result of the proposed algorithm was $u = 56.21$ (km/h) and $w = 10.98$ (km/h). Figure 5 shows the estimation result along with detector measurement (blue colormap; thick color means that such data was observed frequently) and the ground truth FD. The parameter values of the algorithm for this estimation were $\theta = 0.05$, $\Delta m = 2$ (veh), $\Delta t = 15$ (s), and $\phi = 13$ (km/h). In this estimation, 43 PTSs were calculated from 12 probe pairs by the PTS calculation step.

C. Discussion

For the presented result, the relative difference between the estimate of the proposed method and the reference value provided by detectors are as follows: 7.9% for free-flow speed, −12.6% for backward wave speed, and −9.5% for capacity. This can be considered as fairly accurate, considering the fact that number of samples was limited (43 in total, collected from 2 hours probe vehicle data) and only near-critical traffic states were observed. However, it must be noted that such error around 10% in an FD estimation might be problematic for some traffic management purposes and applications. If the accuracy of the algorithm improves with the number of samples, the proposed algorithm will obtain much accurate results in actual application, since current availability of probe vehicle data is significantly higher than the experimental data employed by this validation. Moreover, such large amount of data usually contains measurements from wide variety of traffic states from sparse free-flowing state to severely congested state, meaning that it is more suitable situation for the proposed algorithm.

The algorithm requires some predetermined parameter values, such as, θ, Δm, Δt, and ϕ for the PTS calculation step. It was confirmed that value of θ, which seems to be essential for the proposed algorithm (c.f., II-C), affect strongly to the estimation result. For example, meaningful results were not obtained when $\theta = 0.01$, due to lack of PTSs; contrary, accuracy was significantly degraded when $\theta = 0.10$, due to inaccurate PTSs. However, how other parameter values affect to the algorithm's accuracy had not been clarified due to lack of data.

IV. Conclusion

This paper proposed an algorithm of estimating fundamental diagram (also known as flow–density relation) of traffic based only on GPS-equipped probe vehicle data. The algorithm estimates free-flow speed and backward wave speed of a triangular FD from randomly sampled vehicle trajectories, by extracting near-stationary traffic from probe vehicle data and applying an EM algorithm for parameter estimation. The proposed algorithm was empirically validated based on actual probe vehicle data on an urban highway. As a result, it was confirmed that the algorithm can estimate the FD fairly accurately under certain parameter setting. The proposed method would be valuable to monitor dynamic traffic performance in the wide-ranging road networks, in the coming era of connected vehicles.

The main future work is further empirical analysis. Although the algorithm successfully estimated an FD under certain conditions, the generality was not clarified as discussed in III-C. This is mainly due to lack of data. Therefore, it must be clarified that how the accuracy of the algorithm is affected by the parameter setting of the algorithm, as well as number and quality of samples.

ACKNOWLEDGEMENTS

The authors would like to thank Metropolitan Expressway Co., Ltd. for providing the detector data.

REFERENCES

[1] M. Treiber and A. Kesting, *Traffic Flow Dynamics: Data, Models and Simulation*. Springer Berlin Heidelberg, 2013.

[2] N. Chiabaut, C. Buisson, and L. Leclercq, "Fundamental diagram estimation through passing rate measurements in congestion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 355–359, 2009.

[3] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.

[4] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 49–59, 2003.

[5] D. B. Work, O.-P. Tossavainen, S. Blandin, A. M. Bayen, T. Iwuchukwu, and K. Tracton, "An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices," in *2008 IEEE 47th Conference on Decision and Control*, 2008, pp. 5062–5068.

[6] Y. Yuan, J. W. C. van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, "Real-time Lagrangian traffic state estimator for freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 59–70, 2012.

[7] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annual Reviews in Control*, vol. 43, pp. 128–151, 2017.

[8] Z. Sun, W.-L. Jin, and S. G. Ritchie, "Simultaneous estimation of states and parameters in Newell's simplified kinematic wave model with Eulerian and Lagrangian traffic data," *Transportation Research Part B: Methodological*, vol. 104, pp. 106–122, 2017.

[9] T. Seo, T. Kusakabe, and Y. Asakura, "Traffic state estimation with the advanced probe vehicles using data assimilation," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 824–830.

[10] T. Seo, T. Kusakabe, and Y. Asakura, "Calibration of fundamental diagram using trajectories of probe vehicles: Basic formulation and heuristic algorithm," in *Transportation Research Procedia*, vol. 21, 2017, pp. 6–17, (Selected paper from ISTS&IWTDCS, Jeju, Korea, July 7–8, 2016).

[11] F. Zhu and S. V. Ukkusuri, "An optimal estimation approach for the calibration of the car-following behavior of connected vehicles in a mixed traffic environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 2, pp. 282–291, 2017.

[12] L. C. Edie, "Discussion of traffic stream measurements and definitions," in *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, J. Almond, Ed., 1963, pp. 139–154.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

[14] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transportation Research Part C: Emerging Technologies*, vol. 53, pp. 134–150, 2015.