A Method for Estimation of Traffic States Considering Floating Car Data in a Large Network

Kaifeng DUAN[1], Ryuichi TANI[2], Kenetsu UCHIDA[3],
Keita ZUSHII[4] and Osamu NAGAOKA[5]

[1]Student member of JSCE, Graduate School of Engineering., Hokkaido University
(5 Chome Kita 8 Jonishi, Kita Ward, Sapporo, Hokkaido 060-0808, Japan)
E-mail: kaifeng.duan.k5@elms.hokudai.ac.jp (Corresponding Author)
[2]Member of JSCE, Assistant Professor, Faculty of Engineering., Hokkaido University
(5 Chome Kita 8 Jonishi, Kita Ward, Sapporo, Hokkaido 060-0808, Japan)
E-mail: r-tani@eng.hokudai.ac.jp
[3]Member of JSCE, Professor, Faculty of Engineering., Hokkaido University
(5 Chome Kita 8 Jonishi, Kita Ward, Sapporo, Hokkaido 060-0808, Japan)
E-mail: uchida@eng.hokudai.ac.jp
[4]Nonmember of JSCE, DOCON CO., LTD.
(6,Kita 1 Jo Nishi 1-Chome, Chuo Ward, Sapporo, Hokkaido 060-0001, Japan)
E-mail: kz1899@docon.jp
[5]Member of JSCE, DOCON CO., LTD.
(6,Kita 1 Jo Nishi 1-Chome, Chuo Ward, Sapporo, Hokkaido 060-0001, Japan)
E-mail: on1255@docon.jp

This study proposes a computationally efficient method for the traffic states estimation by using floating car data. The traffic states in a road network are obtained by solving a full information maximum likelihood estimation (FIMLE) problem under user equilibrium (UE) constraints. Solving the FIMLE problem under equilibrium constraints requires much computations time. Therefore, a solution algorithm based on the sensitivity analysis for UE is proposed in this study. We estimate the traffic states in the Asahikawa network in Hokkaido, Japan by applying the method proposed in this study.

Key Words : *ETC2.0 data, Full information maximum likelihood estimation, Sensitivity analysis*

1. INTRODUCTION

The field of traffic planning has witnessed a significant impact due to the advancements in information and communication technology. Previously, traffic data collection was carried out through manual means, but the development of traffic data collection technology such as traffic detectors and ETC2.0 has facilitated the collection of comprehensive road traffic states. The development of a precise system capable of processing large traffic data is crucial for accurately estimating road traffic states.

Several studies have been conducted to estimate traffic states in networks. Hara (2018) estimated link speeds following a multivariate normal distribution from accumulated and observed probe car data and developed a method to complement link speeds on road links where probe car data is not available.

Tani and Uchida (2020) constructed a maximum likelihood estimation model with equilibrium constraints based on a traffic assignment model that considers the uncertainty of the road network and proposed a method to estimate the joint distribution of link traffic flow and delay time for the entire network for multiple time periods.

Additionally, Bayesian inference has been used to estimate traffic states. Hazelton (2010), and Parry and Hazelton (2013) estimate traffic flows by network-based models using Bayesian inference. Wei and Asakura (2013) present a Bayesian method for estimating road network traffic flows. The traffic flows are computed using a stochastic user equilibrium (SUE) traffic assignment model.

Kawamura et al. (2021) proposed a method for estimating the posterior distribution of traffic conditions. The probability distribution of traffic conditions is estimated from the maximum likelihood estimation model constructed by Tani and Uchida (2020) as the prior distribution, and then Bayesian estimation using traffic flow data obtained from traffic detectors and travel time obtained from probe data as the likelihood. The proposed method is based on the Bayesian estimation of the posterior distribution of the traffic condition.

Zheng and Van Zuylen (2013) offer an artificial neural network approach based on probe car data for link travel time estimation. Tani et al. (2018) proposed a method to estimate stochastic path travel times by using both traffic counter data and probe car data.

Tobin and Frizes (1988) proposed a sensitivity analysis method for an equilibrium (UE) traffic assignment problem. Yang and Bell (2009) modify the method by Tobin and Frizes (1988).

In this study, a method for the estimation of traffic states using link delay data is proposed. The link delay data collected by Electronic Toll Collection Systems 2.0 (ETC2.0) is used in the method. The ETC2.0 can collect the travel times only on links where vehicles with ETC2.0 pass. The market penetration rate of the ETC2.0 is not high so far. Therefore, there are some links in a road network in Japan where no travel time is collected. To cope with a road network where some of the link data is unavailable, the traffic states estimation is formulated as a full information maximum likelihood estimation (FIMLE) problem with UE constraints in this study. An efficient algorithm for solving the FIMLE problem based on the sensitivity analysis for UE is also presented.

The rest of this paper is as follows. Section 2 formulates a network model including the definition of traffic flows, travel times, and a UE traffic assignment problem. Section 3 introduces the FIMLE problem UE constraints. Section 4 proposes an efficient solution algorithm based on sensitivity analysis for UE. Section 5 demonstrates and verifies the proposed method in a real network. Finally, the concluding remarks and future directions are shown in the final section.

2. NETWORK MODELS

(1) Notations

The notations below are adopted in this study. Random variables are expressed in capital letters, and lowercase letters are used to denote the expected values of the corresponding random variables.

A	Set of links in the network
I	Set of OD pairs
J_i	Set of path
Q	Total OD demands
Q_i	Demand between OD pair i
T_a	Travel time on the link a
T'_a	Delay time on the link a
C_a	Traffic capacity on the link a
p_i	The ratio of demand for OD pair i to the total OD demands
p_{ij}	Probability of selecting path j in OD pair i
Ξ_{ij}	Travel time on path j between OD pair i
F_{ij}	Traffic flow on path j between OD pair i
V_a	Traffic flow on the link a
t_a^0	Free flow travel time on the link a
α, β	Parameters of BPR function
c_{ij}	Generalized path travel cost on path j between OD pair i
μ_i	Minimum path travel cost between OD pair i
Δ	Link/path incidence matrix
$\boldsymbol{F}, \boldsymbol{\mu}, \boldsymbol{\Xi}, \boldsymbol{Q}, \boldsymbol{C}$	Vector of path flows, minimum path travel costs, path travel times, OD demands, and generalized travel costs
$\boldsymbol{f}, \boldsymbol{c}, \boldsymbol{q}$	Mean vector of path flows, generalized travel costs, and OD demands

(2) Traffic flow

Consider a road network G(N,A) with a set of nodes N and a set of links A. The total demand Q in this network that follows a log-normal distribution is

$$Q \sim LN(\mu, \sigma^2) \tag{1}$$

where μ and σ^2 are the parameters of the log-normal distribution. The demand Q_i for OD pair $i \in I$ is expressed as

$$Q_i = p_i \cdot Q \ \forall i \in I \tag{2}$$

where the ratio p_i for OD pair i holds $\sum_{i \in I} p_i = 1$. OD demand Q_i follows the following log-normal distribution.

$$Q_i \sim LN\left(\mu_{Q_i}, \sigma_{Q_i}^2\right) \ \forall i \in I \tag{3}$$

Its parameters are

$$\mu_{Q_i} = \mu + \ln(p_i) \ \ \forall i \in I \tag{4}$$

$$\sigma_{Q_i}^2 = \sigma^2 \ \ \forall i \in I \tag{5}$$

The vector of mean OD demand is represented as follows:

$$\boldsymbol{\mu}_Q = \left(\mu_{Q_1}, \dots, \mu_{Q_i}, \dots, \mu_{Q_{|I|}}\right) \tag{6}$$

The path flow F_{ij} is calculated as the product of the OD demand Q_i and p_{ij} which is the probability of selecting route j between OD pair i.

$$F_{ij} = p_{ij} \cdot Q_i = \ p_{ij} \cdot p_i \cdot Q \quad \forall\ i \in I, \forall j \in J_i \tag{7}$$

The path flow follows the following log-normal distribution.

$$F_{ij} \sim LN\left(\mu_{F_{ij}}, \sigma^2_{F_{ij}}\right) \quad (8)$$
$$\forall\, i \in I, \forall j \in J_i$$

where

$$\mu_{F_{ij}} = \mu + \ln\left(p_{ij} \cdot p_i\right) \quad (9)$$
$$\forall\, i \in I, \forall j \in J_i$$
$$\sigma^2_{F_{ij}} = \sigma^2 \ \forall\, i \in I, \forall j \in J_i \quad (10)$$

The link flow is expressed as the sum of path flows that pass through the link:

$$V_a = \sum_{i \in I} \sum_{j \in J_i} F_{ij} \cdot \delta_{aj} \quad (11)$$
$$= \sum_{i \in I} \sum_{j \in J} p_{ij} \cdot p_i \cdot Q \cdot \delta_{aj} = \hat{p}_a \cdot Q$$
$$\forall a \in A$$

where

$$\hat{p}_a = \sum_{i \in I} \sum_{j \in J_i} \delta_{aj} \cdot p_{ij} \cdot p_i \quad (12)$$
$$\forall a \in \mathrm{A}$$

Its parameters are given as follows.

$$V_a \sim LN\left(\mu_{V_a}, \sigma^2_{V_a}\right) \forall a \in A \quad (13)$$
$$\mu_{V_a} = \mu + \ln\left(\hat{p}_a\right) \ \forall a \in A \quad (14)$$
$$\sigma^2_{V_a} = \sigma^2 \ \forall a \in A \quad (15)$$

(3) Travel time

In this study, the link travel time is defined by the BPR function. Using the previously defined link flow and the link capacities $C_a \sim LN\left(\mu_{C_a}, \sigma^2_{C_a}\right)$ $(\forall a \in A)$, the link travel time follows the following shifted log-normal distribution.

$$T_a(V_a, C_a) = t_a^0 \cdot \left(1 + \alpha \left(\frac{V_a}{C_a}\right)^{\beta}\right) \quad (16)$$
$$\forall a \in A$$

Link travel time is expressed as follows.

$$T_a(V_a, C_a) = t_a^0 + \frac{t_a^0 \cdot \alpha_a}{{C_a}^{\beta_a}} \cdot (V_a)^{\beta_a} \quad (17)$$
$$= t_a^0 + \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a} \forall a \in A$$
$$\gamma_a = t_a^0 \cdot \alpha_a \ \ \forall a \in A \quad (18)$$

The link travel time shown above can be divided into a deterministic term (free flow travel time) and a random term (link delay time).

The random term is affected by changes in link flow and link capacity. Since the link flow and link capacity follow mutually independent log-normal distributions, the link delay time also follows a log-normal distribution. The link delay time follows the following log-normal distribution.

$$T_a' = \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a} \sim LN\left(\mu_{T_a'}, \sigma^2_{T_a'}\right) \quad (19)$$
$$\forall a \in A$$

where

$$\mu_{T_a'} = \ln(\gamma_a) + \beta_a \cdot \ln(\hat{p}_a)$$
$$+\beta_a \cdot \left(\mu - \mu_{C_a}\right) \forall a \in A \quad (20)$$
$$\sigma^2_{T_a'} = (\beta_a)^2 \cdot \left(\sigma^2 + \sigma^2_{C_a}\right) \quad (21)$$
$$\forall a \in A$$

The link travel time follows a shifted log-normal distribution. Thereby, the link travel time is

$$T_a = t_a^0 + T_a' \ \ \forall a \in A \quad (22)$$

In other words, the distribution of link travel time is a positive parallel shift to the distribution of link delay; hence, the traffic states were estimated using the link delay, which follows the log-normal distribution, and not link travel time. Then, the mean of link travel time is given by

$$E[T_a] = t_a^0 + E(T_a') \quad (23)$$
$$= t_a^0 + \exp\left(\mu_{T_a'} + \frac{1}{2}\sigma^2_{T_a'}\right)$$
$$\forall a \in A$$

The path travel time is expressed as the sum of the link travel time as follows.

$$\Xi_{ij} = \sum_{a \in A} T_a \cdot \delta_{aj} \forall i \in I, \forall j \in J_i \quad (24)$$

where δ_{aj} is the variable that equals 1 if path j uses link and 0 otherwise. Here the mean of path travel time is calculated as follows,

$$E\left[\Xi_{ij}\right] = \sum_{a \in A} E[T_a] \cdot \delta_{aj} \quad (25)$$
$$\forall i \in I, \forall j \in J_i$$

the vector of path travel times is expressed as follows.

$$\boldsymbol{\Xi} = \left(\Xi_{11}, \dots, \Xi_{ij}, \dots, \Xi_{|I||J_i|}\right)^T \quad (26)$$

(4) User equilibrium assignment model

The generalized path travel cost is calculated by the sum of the mean of link travel time.

$$c_{ij} = \sum_{a \in A} E[T_a] \cdot \delta_{aj} \quad (27)$$
$$\forall i \in I, \forall j \in J_i$$
$$\boldsymbol{c} = \left(c_{11}, c_{12} \dots c_{ij} \dots c_{|I||J_i|}\right)^T \quad (28)$$

where δ_{aj} is the variable that equals 1 if path j uses link and 0 otherwise. The vectors of path flows, OD demands, and the mean vectors of path traffic flows and OD demands are respectively represented by

$$\boldsymbol{F} = \left(F_{11}, \dots, F_{ij}, \dots, F_{|I||J_i|}\right)^T \quad (29)$$

$$\boldsymbol{f} = \left(f_{11}, \dots, f_{ij}, \dots, f_{|I||J_i|}\right)^T \quad (30)$$

$$\boldsymbol{Q} = \left(Q_1, \dots, Q_i, \dots, Q_{|I|}\right)^T \quad (31)$$

$$\boldsymbol{q} = \left(q_1, \dots, q_i, \dots, q_{|I|}\right)^T \quad (32)$$

When mean vector of path flows is $\boldsymbol{f} \geq 0$, the minimum path travel cost between OD pair i is expressed as follows.

$$\mu_i = min\{c_{ij}, \forall j \in J_i\} \ \forall i \in I \quad (33)$$

Here, the vector of minimum path travel costs is shown below.

$$\boldsymbol{\mu} = \left(\mu_1, \dots, \mu_i, \dots, \mu_{|I|}\right)^T \quad (34)$$

For any path $j \in J_i$ between OD pair $i \in I$, the mean vector of path flows $\boldsymbol{f}$ is in Wardropian equilibrium (User Equilibrium) shown below.

$$f_{ij} > 0 \rightarrow c_{ij} = \mu_i \ \forall i \in I, \forall j \in J_i \quad (35)$$

$$f_{ij} = 0 \rightarrow c_{ij} \geq \mu_i \ \forall i \in I, \forall j \in J_i \quad (36)$$

The equilibrium condition for User Equilibrium (UE) assignment for a given OD pair is that the generalized costs of the used routes are all equal and are smaller than or equal to the costs of the unused routes.

Given this equilibrium condition, the user equilibrium condition can be expressed as the following variational inequality problem:

$$\sum_{i \in I} \sum_{j \in J_i} c_{ij}(f_{ij} - {f_{ij}}^*) \geq 0 \quad (37)$$

$$\sum_{i \in I} \sum_{j \in J_i} f_{ij} = q_i \quad (38)$$

$$f_{ij} \geq 0 \ \forall i \in I, \forall j \in J_i \quad (39)$$

where ${f_{ij}}^*$ is the path flow in User Equilibrium state.

In addition, (37)-(39) can be expressed as follows,

$$\boldsymbol{c} \cdot (\boldsymbol{f} - \boldsymbol{f}^*)^T \geq 0$$
$$\boldsymbol{f} \geq 0 \quad (40)$$

3. FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATION

The preceding sections showed the definition of the network models, and a user equilibrium assignment model, which forms equilibrium constraints for the objective function. This section presents the FIMLE problem, referring to Tani and Uchida (2021).

The vector of link delay times is

$$\boldsymbol{T'} = \left(T'_1, \dots, T'_a, \dots, T'_{|A|}\right) \quad (41)$$

that follows the following multivariate log-normal distribution.

$$\boldsymbol{T'} \sim MVLN\left(\boldsymbol{\mu}_{T'_a}, \boldsymbol{\Sigma}_{T'_a}\right) \quad (42)$$

where

$$\boldsymbol{\mu}_{T'_a} = \left(\mu_{T'_1} \ \mu_{T'_2} \dots \mu_{T'_{|A|}}\right)^T \quad (43)$$

$$\boldsymbol{\Sigma}_{T'_a} = \begin{pmatrix} \sigma^2_{T'_1} & \cdots & \sigma_{T'_1, T'_{|A|}} \\ \vdots & \ddots & \vdots \\ \sigma_{T'_{|A|}, T'_1} & \cdots & \sigma^2_{T'_{|A|}} \end{pmatrix} \quad (44)$$

$$\sigma_{T'_a, T'_b} = \beta_a \cdot \beta_b \cdot \sigma^2 \quad (45)$$

The method for estimating traffic conditions in this study is based on the method proposed by Tani and Uchida (2020).

Given G days of traffic observation data, define the observed state vector $\boldsymbol{m}_g$ and the observed value vector $\widehat{\boldsymbol{d}}_g$ for the observation on day $g \in \{1, \dots, G\}$. $\widehat{\boldsymbol{d}}_g$ is a vector representing the traffic conditions observed across the road network on day $g \in \{1, \dots, G\}$, defined as

$$\widehat{\boldsymbol{d}}_g = \left(\hat{t}_1 \dots \ \hat{t}_a \dots \hat{t}_{|A|}\right) \ g \in \{1, \dots, G\} \quad (46)$$

where $\hat{t}_a$ is the logarithm of the observed delay time of link a. For instance, the link delays observed on day $g \in \{1, \dots, G\}$ in a simple network which is comprised of three links are 10 min at link1 and 5min at link 2, then the $\widehat{\boldsymbol{d}}_g$ is expressed as follows.

$$\widehat{\boldsymbol{d}}_g = (\ \ln(10) \ \ \ln(5) \ \ -\infty) \quad (47)$$

If there is no observed value or the observed value is less than or equal to zero, the corresponding value is assumed to be $-\infty$. Since the link delay time is defined as the increase in travel time relative to free flow travel time. The link delay time cannot be defined if the link travel time is less than or equal to the free flow travel time.

Observed state vector $\boldsymbol{m}_g$ shown by (48) reflects the traffic state on links, and its entry $m_{a,g}$ equals 1 if the link delay time is observed on link a and 0 otherwise.

$$\boldsymbol{m}_g = \left[m_{a,g}, \forall a \in A\right] \quad (48)$$

where

$$m_{a,g} = \begin{cases} 1, & if \ \ \hat{t}_a \neq -\infty \\ 0, & otherwise \end{cases} \quad (49)$$

The $\boldsymbol{m}_g$ corresponding to (49) is

$$\boldsymbol{m}_g = (1 \ 1 \ 0) \quad (50)$$

$\boldsymbol{M}_g$ is defined as the diagonal matrix of $\boldsymbol{m}_g$ from which each row containing only zero entries is removed. This indicates that only data from observed links are used to calculate the likelihood. For example, $\boldsymbol{m}_g = (1 \ 1 \ 0)$, as shown in (50), $\boldsymbol{M}_g$ can be expressed as:

$$\boldsymbol{M}_g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (51)$$

the number of data observed on day $g \in \{1, \dots, G\}$ is $n(g)$.

Link delay for the entire road network follows a multivariate Log-normal distribution, as mentioned before, it is denoted as $\boldsymbol{T}' \sim MVLN(\boldsymbol{\mu}_{T_a'}, \boldsymbol{\Sigma}_{T_a'})$, and the natural logarithm of link delays follows a multivariate normal distribution, denoted as $\ln(\boldsymbol{T}') \sim MVN(\boldsymbol{\mu}_{T_a'}, \boldsymbol{\Sigma}_{T_a'})$. Here, the parameters are the same for both. Therefore, the estimates obtained by full information maximum likelihood estimation using link delay that follow a multivariate log-normal distribution as likelihoods are the same as those obtained by full information maximum likelihood estimation of a multivariate normal distribution. The method used in this study takes advantage of this property to estimate the parameters by using full information maximum likelihood estimation of a multivariate normal distribution.

The FIMLE problem in this study is defined as:

$$\max \quad L(\mu, \sigma^2 | \widehat{\boldsymbol{d}}_g \forall g \in \{1, \dots, G\}) = \prod_{g \in \{1,\dots,G\}} \frac{\exp\left(-\frac{1}{2}\left(\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g(\mu)\right)^T \boldsymbol{\Sigma}_g^{-1}(\sigma^2)(\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g(\mu))\right)}{2\pi^{\frac{n(g)}{2}} |\boldsymbol{\Sigma}_g(\sigma^2)|^{\frac{1}{2}}} \tag{52}$$

s.t. (37)-(39), where

$$\boldsymbol{\Sigma}_g(\sigma^2) = \boldsymbol{M}_g \boldsymbol{\Sigma}_{T_a'}(\sigma^2) \boldsymbol{M}_g^T \quad \forall g \in \{1, \dots, G\} \tag{53}$$

$$\boldsymbol{\mu}_g(\mu) = \boldsymbol{M}_g \boldsymbol{\mu}_{T_a'}(\mu) \ \forall g \in \{1, \dots, G\} \tag{54}$$

$$\widehat{\boldsymbol{d}}_g = \boldsymbol{M}_g \widehat{\boldsymbol{d}}_g^T \ \forall g \in \{1, \dots, G\} \tag{55}$$

4. SOLUTION ALGORITHM BASED ON SENSITIVITY ANALYSIS FOR UE FLOWS

When solving the likelihood maximization problem mentioned in section 3 to estimate the traffic state, it is difficult to find the gradient of the objective function with equilibrium constraints.

Therefore, this chapter introduces the sensitivity analysis proposed by Yang and Bell (2009) to the FIMLE problem for converting the equilibrium constraints to linear constraints. Yang and Bell (2009) considered that the perturbation parameters exist in the link cost function $\boldsymbol{t}(\boldsymbol{V}, \boldsymbol{\varepsilon_t})$ and the OD traffic demand vector $\boldsymbol{q}(\boldsymbol{\varepsilon_q})$. Here, the perturbation parameters are assumed to be the perturbations $\boldsymbol{\varepsilon_t}$ and $\boldsymbol{\varepsilon_q}$ for link toll and OD demand. In this study, the perturbation of the link toll $\boldsymbol{\varepsilon_t}$ is not considered and the unknown variable is the OD demands. In the following, the perturbation parameter for OD demands will be denoted by $\boldsymbol{\varepsilon}$. The $\boldsymbol{q}(\boldsymbol{\varepsilon})$ is represented by

$$\mathbf{q}(\boldsymbol{\varepsilon}) = \boldsymbol{q} + \boldsymbol{\varepsilon} \tag{56}$$

The path flow calculated at $\mathbf{q}(\boldsymbol{\varepsilon})$ can be approximated by

$$\tilde{\boldsymbol{f}}(\boldsymbol{q}(\boldsymbol{\varepsilon})) = \boldsymbol{f}^*(\boldsymbol{q}) + \boldsymbol{\nabla_\varepsilon f} \cdot \boldsymbol{\varepsilon} \tag{57}$$

where $\boldsymbol{f}^*$ was defined in Equation (40), $\boldsymbol{\nabla_\varepsilon f}$ is calculated by the method of sensitivity analysis. Thus, $\tilde{\boldsymbol{f}}(\boldsymbol{q}(\boldsymbol{\varepsilon}))$ is the vector of path flows calculated at $\boldsymbol{q}(\boldsymbol{\varepsilon})$. The link flows can be defined as

$$\boldsymbol{v}(\boldsymbol{\varepsilon}) = \tilde{\boldsymbol{f}}(\boldsymbol{q}(\boldsymbol{\varepsilon})) \cdot \Delta \tag{58}$$

The ratio of link flows to mean OD demands in (12) is calculated as

$$\hat{p}_a(\boldsymbol{\varepsilon}) = \frac{v_a(\boldsymbol{\varepsilon})}{q(\boldsymbol{\varepsilon})} \tag{59}$$

It is complex to solve the likelihood function with equilibrium constraints directly and which requires a large amount of computation time. However, the method of sensitivity analysis converts the equilibrium constraints to linear constraints, which makes the objective function finding the optimal solution easier. The maximum likelihood estimation problem can be written as follows.

$$\max L(\mu, \sigma^2, \boldsymbol{\varepsilon} | \widehat{\boldsymbol{d}}_g \forall g \in \{1, \dots, G\}) = \prod_{g \in \{1,\dots,G\}} \frac{\exp\left(-\frac{1}{2}\left(\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g(\boldsymbol{\varepsilon}, \mu)\right)^T \boldsymbol{\Sigma}_g^{-1}(\sigma^2)\left(\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g(\boldsymbol{\varepsilon}, \mu)\right)\right)}{2\pi^{\frac{n(g)}{2}} |\boldsymbol{\Sigma}_g(\sigma^2)|^{\frac{1}{2}}} \tag{60}$$

s.t.

$$\sum_{i \in I} q_i(\boldsymbol{\varepsilon}) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{61}$$

Step 0 Initialization	Set $n = 1$ and the initial settings for parameters μ and σ^2 of total OD demand and p_i
Step 1 UE assignment	User Equilibrium assignment is performed to obtain the traffic conditions.
Step 2 Sensitivity analysis	Calculate the $\boldsymbol{\varepsilon}$ based on the results of traffic conditions calculated in step 1.
Step 3 Full information likelihood maximization Update of μ and σ^2	Solve the full information likelihood maximization problem to optimize $\boldsymbol{\varepsilon}$. Each OD demand and new parameters μ and σ^2 are updated by optimized $\boldsymbol{\varepsilon}$.

Step 4 Convergence	If convergence is achieved, the calculation is terminated. If not, set $n = n + 1$ and return to Step 1

5. NUMERICAL CALCULATIONS

(1) Network information

The road network in Asahikawa, Japan, shown in Figure 1, is used for the numerical calculation. The network has 340 directed links, 135 nodes, and 182 OD pairs. And the red nodes represented in the figure show centroid nodes from/to which OD demand is generated/absorbed.

Figure 1. Asahikawa network

(2) Observation data preparation (ETC2.0)

The link delay times collected by ETC2.0 in October 2021 were used in the calculation. The data was collected from 9:00 am to 10:00 am for 31 days.

Table 1 shows an example of observation data where '-' indicates no observation data is available. From this table, it is shown that there are some links in the network where no observation data is available.

Table 1. Observed link delay times for the Asahikawa network

Day \ Link No.	104	112	113	114	115	116	117	118
10	714.4	126.4	-	-	-	76.4	130.4	-
11	1576.4	105.4	718.4	239.4	82.4	68.4	164.4	23.4
12	743.4	129.4	-		255.4	180.4	112.4	9.4
13	1152.4	128.4	-	221.4	92.4	82.4	216.4	72.4
14	744.4	122.4	-	-	-	57.4	127.4	0.4
15	1661.4	144.4	-	179.4	65.4	65.4	211.4	35.4
16	821.4	102.4	950.4	156.4	0.0	72.4	101.4	12.4
17	6808.4	184.4	-	-	92.4	72.4	88.4	-
18	924.4	119.4	728.4	224.4	96.4	74.4	155.4	14.4
19	1363.4	131.4	-	-	-	68.4	237.4	7.4

It is obvious from Table 1 that there are some outliers, such as a delay time of 6808.4 seconds, for which data cleansing is required before the estimation of traffic states.

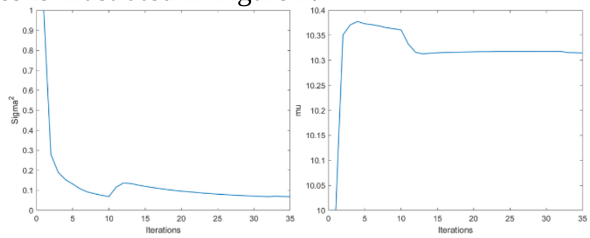
In this study, an outlier is defined as a value that is more than three scaled median absolute deviations (MAD) from the median. For a finite-length vector $\boldsymbol{B}$ made up of N scalar observations, the median absolute deviation (MAD) is defined as

$$MAD = median(|B_i - median(\boldsymbol{B})|) \quad i = 1,2 \dots N \tag{62}$$

Find the outlier using the above procedure, assign it the value zero, and then execute the numerical calculation.

(3) Results for the Asahikawa network

The estimated μ and σ^2 for the Asahikawa network are 10.3580 and 0.0562 and their convergence process is illustrated in Figure 2.



(a) (b)

Figure 2. Convergence process of μ (a) and σ^2 (b) for the Asahikawa network.

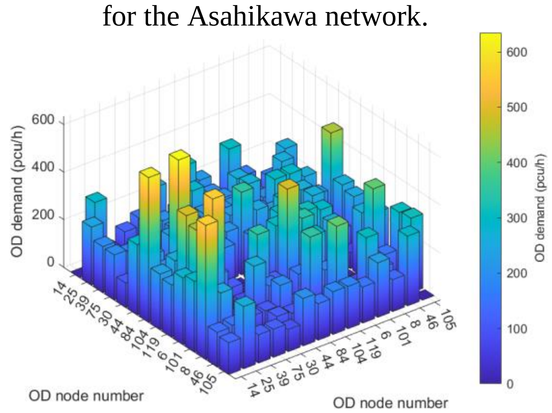


Figure 3. Demand for each OD pair for the Asahikawa network.

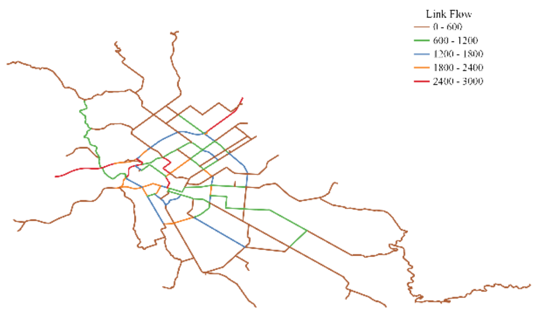
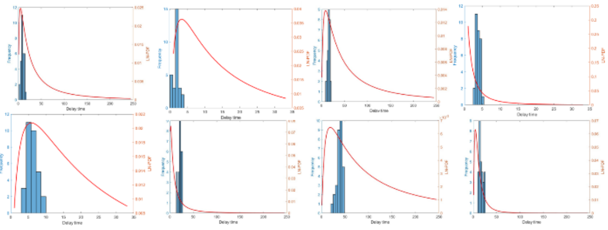


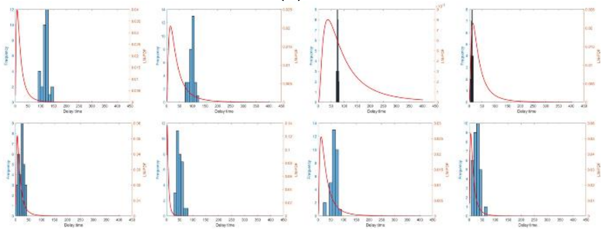
Figure 4. Estimated average link flow.

Figure 3 shows the estimated mean demand for each OD pair (there are a total of 182 OD pairs). The total OD demand estimated is 3.2406×10^4[pcu/hour]. The estimated OD demands for OD pairs 44-14, 44-6, 104-6, and 84-14 are significantly larger than those for other OD pairs, given that the centroid node of 84 is located near a primary school and the centroid nodes of 44, 104 and 14 are located near residential areas. In the vicinity of centroid node 6, there is the expressway from the outskirts to the center of Asahikawa City.

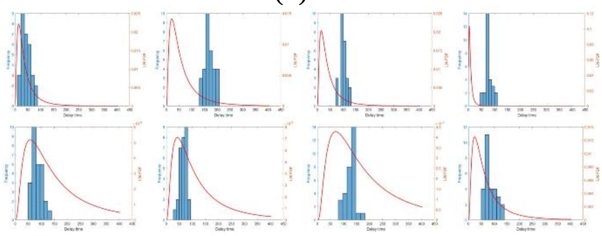
Figure 4 shows the estimated average link flow for the Asahikawa network. In this study, a road section in the network is comprised of opposing two directed links, and since the flows of opposing two directed links are essentially the same. The link flow shown in Figure 4 is the average of mean link flows for the two opposing directed links. The average link flow in the central part is around 600-1,200 [pcu/hour], and the links in red have flows between 2,400-3,000 [pcu/hour]. The reason for this is that the links are located near the centroid nodes.



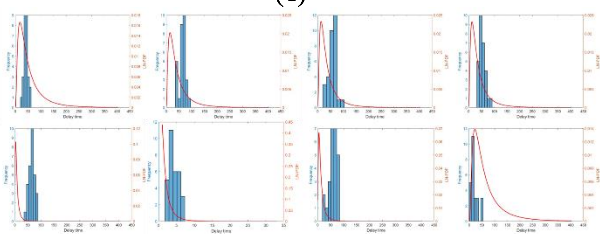
(a)



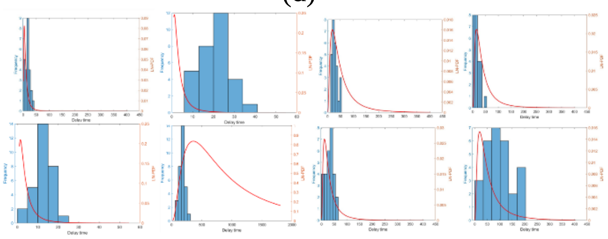
(b)



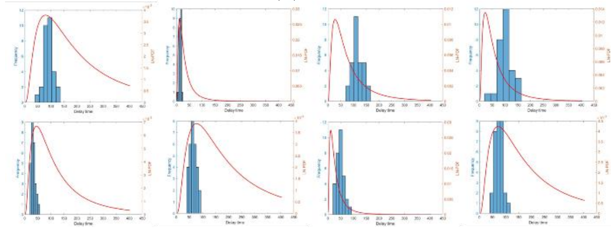
(c)



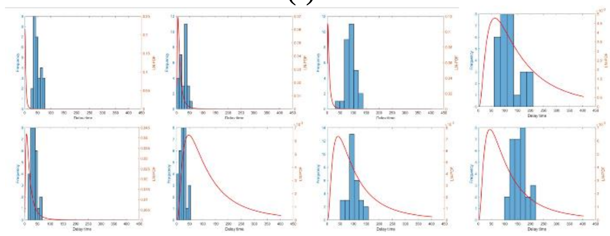
(d)



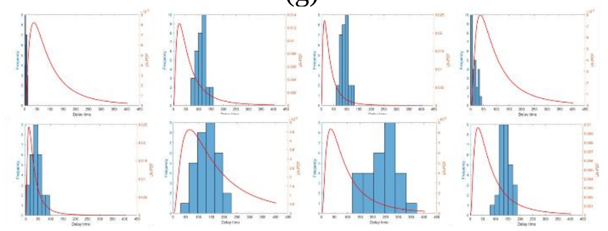
(e)



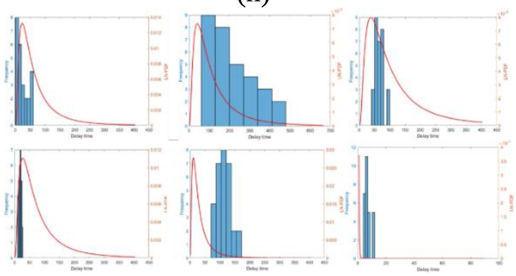
(f)



(g)



(h)



(i)

Figure 5. Histograms of delay time and PDF of log-normal distribution (a) to (i).

Figure 5 shows the relationship between the observed link delay times and the probability density function of the estimated lognormal distribution function. Only the links with more than 25 observed link delays are addressed in the figure. Since the number of such links is large, we choose one of two opposing links for a road section in the network.

Based on the results, the majority of estimated link delays are similar to the observed delays.

(4) KL divergence

The KL divergence was used for comparing the real observed link delays with the estimated results of link delays by the proposed method, and the definition of KL divergence is shown as:

$$KL(P\|Q) = \sum_{i=1}^{n} P_i ln \frac{P_i}{Q_i} \tag{63}$$

where P_i and Q_i are the probability densities of the i th bin in the two histograms.

The link delay times collected by ETC2.0 in October 2021 over a period of 31 days were used in the calculation, and Figure 6 shows the corresponding KL divergence results.

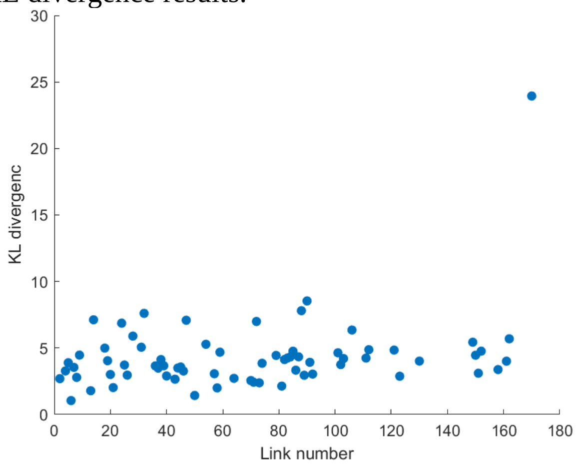


Figure 6. Results of KL divergence

As evidenced in Figure 6, 80% of the KL divergence values fall within the range of 0 and 5, and 98% of the KL divergence values fall within the range of 0 and 10, except for link 170, which has a significantly higher KL divergence value as it is a dummy link. The results of the KL divergence analysis demonstrate that the proposed method accurately estimated link delays in this study by achieving a strong alignment between the true distribution and the approximation.

6. CONCLUSIONS

This study developed a method to estimate the traffic states by the FIMLE problem with UE constraints. In addition, a solution algorithm for the FIMLE problem based on the sensitivity analysis for UE is also presented.

Since the market penetration rate of the ETC2.0 system is not high, therefore the fully observed data cannot be obtained. The proposed method in this study can address such incomplete traffic data.

Numerical calculations are carried out in the real Asahikawa network using real observed link delays. Especially, the method proposed in this study estimated relatively reasonable link delays in the Asahikawa network. The accuracy of the estimated links delays was verified by KL divergence and the issue of traffic condition estimation for multiple time periods in a large network will be addressed in the future study.

REFERENCES

1) Hara, Y., Suzuki, J., and Kuwahara, M. : Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. *Transportation Research Part C.* 86, 622-638, 2018.
2) Hazelton, M.L.: Bayesian inference for network-based models with a linear inverse structure. *Transportation Research Part B.* 44(5), 674–685, 2010.
3) Parry, K., Hazelton, M.L. : Bayesian inference for day-to-day dynamic traffic models. *Transportation Research Part B.* 50, 104–115, 2013.
4) Tani, R., Uchida, K. : An Estimation Method for Spatiotemporal Traffic States Based on Incomplete Traffic Observation. *The 8th International Symposium on Dynamic Traffic Assignment*, 2021.
5) Tani, R., Owada, T., Uchida, K. : Path travel time estimating method by incomplete traffic data. *International Journal of Intelligent Transportation Systems Research, 2018.*
6) Tobin, R.L., Frizes, T.L. : Sensitivity analysis for equilibrium network flows. *Transportation Science*, 22, 242-250,1988.
7) Wei, C., Asakura, Y.: A Bayesian approach to traffic estimation in stochastic user equilibrium networks. *Transportation Research Part C.* 36, 446–459, 2013.
8) Yang, H., Michael, GH. Bell. : Sensitivity analysis of network traffic equilibrium revisited: the corrected approach. *4th IMA International Conference on Mathematics in Transport Institute of Mathematics and its Applications*, September, London, United Kingdom, 2009.
9) Zheng, F., Van Zuylen, H. : Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C.* 31, 145–157,2013.
10) 峪 龍一，内田 賢悦：プローブデータと感知器データを組み合わせた交通状態の時空間推定，*第62回土木計画学研究発表会*，2020
11) 川村 雄斗，峪 龍一，内田 賢悦：不完全交通データを活用した道路ネットワーク交通状態のリアルタイム更新，*第63回土木計画学研究発表会*，2021