Development of efficient solution algorithms based on sensitivity analysis for network design problems with user equilibrium constraints

Kaifeng Duan

Graduate School of Engineering

Hokkaido University

A thesis submitted for the degree of

Doctor of Engineering

Abstract

Network Design Problems (NDPs) are essential for optimizing infrastructure and resource allocation in transport networks. A key challenge in solving NDPs arises from equilibrium constraints, e.g., the user equilibrium (UE) principle. The inclusion of UE constraints in NDPs reflects the real-world dynamics of travelers' behavior in transport networks, where travelers adjust their path choices in response to changes in path cost in the network, seeking to minimize their travel costs. The incorporation of UE constraints makes NDPs computationally intensive to solve, as each network adjustment requires reevaluation of the equilibrium state.

To address this issue, this thesis proposes efficient solution algorithms based on sensitivity analysis to improve computational efficiency in solving NDPs. Using sensitivity analysis, the proposed algorithms effectively evaluate how changes in network parameters influence the equilibrium state. The algorithms simplify the iteration process, reducing the computational burden while maintaining the accuracy of the results.

This thesis is organized into 5 chapters. Chapter 1 describes the background and purpose of this thesis. Chapter 2 describes the literature review and basic models used in this study.

Chapter 3 proposes a computationally efficient algorithm based on the sensitivity analysis of UE for solving a traffic state estimation problem 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 UE constraints. Solving the FIMLE problem under equilibrium constraints requires much computation time. Therefore, an efficient solution algorithm is needed. The traffic states in the real Asahikawa network in Hokkaido, Japan estimated by the algorithm are presented in Chapter 3.6.

Chapter 4 proposes a bi-level programming model to solve an optimization problem of bus route settings and service frequencies, in road networks where connected and autonomous buses (CABs) are introduced as a public transport system. The upper-level problem is formulated as a discrete optimization problem, which is then relaxed into a continuous form. A nonlinear optimization algorithm is applied to solve this problem. This optimization problem is subject to user equilibrium constraints at the lower level, represented by the UE principle. An algorithm based on the sensitivity analysis method is developed to solve the bi-level model. Numerical examples and results are shown in Chapter 4.9 and Chapter 4.10.

In conclusion, this thesis develops innovative algorithms to enhance the computational efficiency and practical applicability of NDPs. These algorithms were successfully applied in Chapter 3 to estimate traffic states in the real-world network of Asahikawa, Hokkaido, Japan, and in Chapter 4 to solve the optimization

problem of bus route settings and service frequencies in Sioux-Falls network. These findings demonstrate the capability of the proposed algorithm to effectively address the challenges of computational efficiency and practical implementation in NDPs.

Contents

1. Introduction

1.1 Background

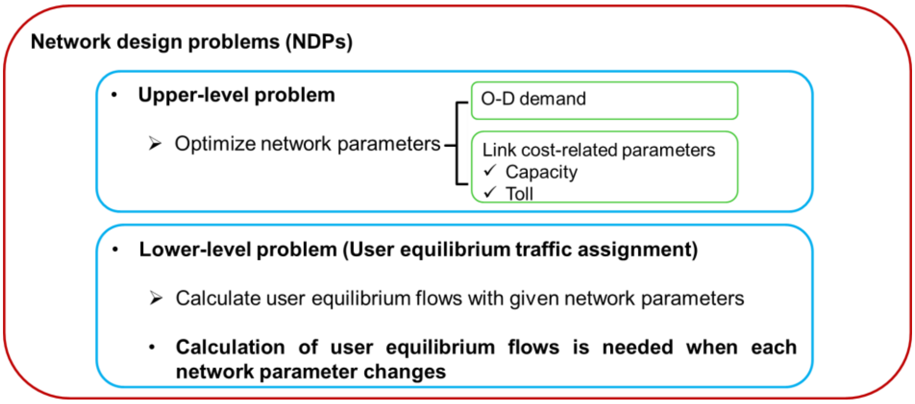


Fig.1.1 Definitions of Network design problems

Network design problems (NDPs) optimize the network parameters that are expressed as Upper-level problem and the network parameters include O-D demand and link cost-related parameters, such as capacity, toll, and so on. The network parameters are treated as fixed inputs in the Lower-level problem (basically formulated as UE traffic assignment) to calculate the UE flows. However, the calculation of user equilibrium flows is needed when each network parameter changes, which requires large amounts of computational time.

NDPs are the fundamental optimization problems in transportation that focus on improving the performance of a transportation network. The goal is to determine the optimal network configuration, such as adding new links or upgrading existing ones, to achieve objectives like minimizing total travel time or reducing congestion. A key challenge in solving NDPs results from equilibrium constraints, e.g., the user equilibrium (UE) principle. The inclusion of UE constraints in NDPs reflects the real-world dynamics of travelers' behavior in transport networks, where travelers adjust their path choices in response to changes in path cost in the network, seeking to minimize their travel costs. The incorporation of UE constraints makes NDPs computationally intensive to solve, as each network adjustment requires reevaluating the equilibrium state.

Since Abdulaal and LeBlanc (1979) first formulated NDPs as a bi-level programming problem, NDPs have become one of the most difficult problems in transportation research (Yang and Bell, 1998, 2001). Over the years, extensive research has been conducted on NDPs, leading to the development of various methods to obtain optimal or near-optimal solutions. Abdulaal and LeBlanc (1979) introduced the Hooke-

Jeeves direct search algorithm as one of the initial approaches to solve NDPs. Around the same time, Tan et al. (1979) proposed expressing the traffic equilibrium problem as a set of nonlinear, nonconvex, yet differentiable constraints in terms of path flow variables. Suwansirikul et al. (1987) introduced the equilibrium decomposed optimization algorithm, a heuristic approach that approximates the derivative of the upper-level objective function. Marcotte (1983) and Marcotte and Marquis (1992) further contributed to this field by developing heuristic methods and analyzing their efficiency. Sensitivity analysis-based algorithms were also advanced by researchers, e.g., Friesz et al. (1990) and Yang and Yagar (1994, 1995).

In more recent developments, Meng et al. (2001) and Yang et al. (2004) reformulated NDPs into a single-level, continuously differentiable optimization problem by utilizing the marginal function of the traffic equilibrium problem. Building on this, Gao et al. (2007) leveraged the relationship between system optimum and UE to develop a solution algorithm based on the method of successive averages (MSA). Additionally, Chiou (2005, 2009) implemented various subgradient-based methods to address NDPs. Ban et al. (2006) approached NDPs through mathematical programs with complementary constraints (MPCC), further enriching the field with novel solution techniques.

With advancements in autonomous driving technology and the growing demand for urban transportation, the estimation of traffic states and the development of efficient networks have become essential considerations. Historically, traffic data was gathered manually, but in recent years it has been obtained easily because of the development of technology (e.g., detectors, known as ETC2.0 in Japan). Autonomous driving technology is expected to become widespread in the near future. Autonomous vehicles play an essential role in the future of urban mobility, offering substantial benefits to transportation systems worldwide.

Autonomous vehicles present several advantages over human-driven vehicles, different from human driving vehicles, computer-controlled autonomous vehicles can reduce operational costs by eliminating the need for drivers. Additionally, autonomous vehicles enhance operational efficiency by collecting and processing real-time information about surrounding vehicles and the environment compared to human driving vehicles.

When humans drive vehicles, differences in their perception of danger result in different reaction times, which negatively impact the efficient use of road capacity. In contrast, autonomous vehicles exhibit much shorter reaction times and maintain more consistent, shorter headways than human driving vehicles, thereby increasing the road capacity. These characteristics position autonomous vehicles as a transformative solution for addressing the challenges of modern urban transportation systems.

However, the introduction of autonomous driving technologies brings changes to travelers' behaviors and link capacity due to the shorter reaction times of autonomous vehicles. Traditional NDPs usually assume that the study targets are human driving vehicles and may not handle the problem caused by autonomous vehicles. Therefore, it becomes crucial to develop NDPs that account for changes in link capacity and the impact on travelers' behaviors resulting from the implementation of autonomous vehicles.

Thereby, NDPs have to be developed in the face of changes in traveler's behaviors due to the development of autonomous driving technologies and increasing demand. This thesis proposes efficient solution algorithms for NDPs.

1.2 Objectives

This thesis focuses on computationally efficient solution algorithms for NDPs in large-scale networks. The inclusion of UE constraints ensures that the solutions account for travelers' behavior in response to network changes, e.g., changes in origin-destination (O-D) demand, toll level, and so on, making the NDPs closer to real-world scenarios. However, the incorporation of UE constraints makes NDPs computationally intensive to solve, as each network adjustment requires reevaluation of the equilibrium state. This thesis proposes the method of sensitivity analysis to approximate UE link (or path) flows so that reduce the computational burden associated with solving large-scale NDPs while maintaining solution accuracy.

The objective of this thesis is to develop efficient sensitivity analysis-based solution algorithms for NDPs. The network parameters of network design problems include O-D demand and link cost-related parameters. The traffic state estimation problem with respect to the network parameter of O-D demand by using the incomplete floating car data is addressed in Chapter 3.

In the multimodal networks, changes in service frequencies will influence the link cost and then change the user equilibrium flows, the calculation of user equilibrium flows is needed if there is a change in service frequencies. Therefore, we develop an efficient sensitivity analysis-based algorithm for solving the network design problem of optimizing the bus route network and service frequencies in Chapter 4.

The sensitivity analysis-based efficient algorithms proposed in this study solve the network design problems mentioned above and can be basically applied to any network design problems.

1.3 Organization of this paper

This thesis is organized into 5 chapters as shown in Fig.1.2. Chapter 1 describes the background and purpose of this thesis. Chapter 2 describes the existing research and basic models used in this study.

Chapter 3 proposes a computationally efficient algorithm based on the sensitivity analysis of UE for solving a traffic state estimation problem by using floating car data. The traffic states in the Asahikawa network in Hokkaido, Japan estimated by the algorithm are also presented.

Chapter 4 proposes a bi-level programming model to solve a bus network design problem in which both bus routes and the corresponding service frequencies are optimized. Connected and Autonomous Buses are introduced as a public transport system to solve the problem. An algorithm based on the sensitivity analysis method is developed for solving the bi-level model. Numerical examples and results are shown. Finally, the conclusions are presented in Chapter 5.

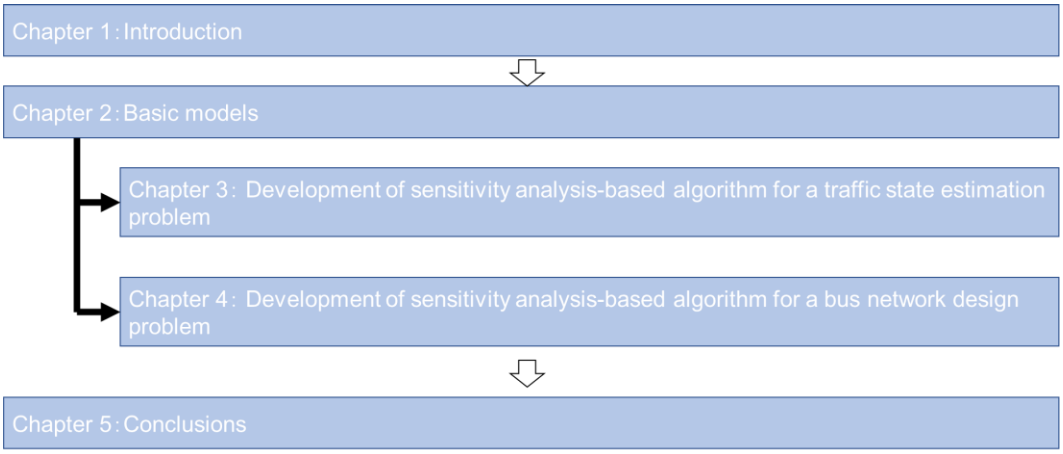


Fig.1.2 Contents of this study

2. Literature review and basic models

Traditional traffic demand forecasting, based on the four-stage model, assumes distinct stages of trip generation, trip distribution, modal choice, and traffic assignment, with demand being forecasted sequentially through these stages. The equilibrium assignment model was initially introduced to the traffic assignment stage in demand forecasting. It is important to note that much of the information outlined in Chapter 2.1 is derived from the Japan Society of Civil Engineers (JSCE, 1998). This foundational approach has served as a cornerstone in transportation planning, providing a structured framework for understanding and predicting traffic patterns within urban environments.

2.1 UE traffic assignment model

With the growing population the need for transportation facilities is increasing day by day, proper transportation planning is needed so that limited resources can be efficiently distributed. However, the core of transportation planning is the prediction of future travel demands. For this reason, the four-step model is frequently used. The trip assignment is the fourth and last step of the four-step model, which predicts the route choice of drivers for the given origin and destination. Among many trip assignment techniques (Ben-Akiva and Lerman, 1985; Meyer and Miller, 2001), the UE method is the most widely used in the practical field.

The UE assignment is based on Wardrop's first principle proposed by Wardrop (1952), which states that:

No driver can unilaterally reduce his/her travel costs by shifting to another route.

Assumptions in UE assignment

1. The drivers have perfect knowledge of the path cost.
2. Travel time on a given link is a function of the flow on that link.
3. Travel time is positive and increasing.

Now a mathematics model was used to express the UE assignment. A transportation network can be abstracted and represented as a graph. A traffic network consists of a set of nodes N and a set of directed links A connecting them, denoted by $G(N, A)$. A path between two origin-destination points (O-D pair) can be represented by a set of consecutive links between the origin and destination nodes. In the following, the set of O-D pairs is represented as I and the set of paths between O-D pairs $i \in I$ is represented as $j \in J_i$. The amount of flow between an O-D pair is allocated to the paths available between the O-D pairs without excess or deficiency, and the link flow is expressed as the sum of all path flows passing through the link (traffic flow conservation law). Traffic demands on O-D pairs, paths, and links are non-negative (non-negative traffic flow conditions). These conditions can be expressed as follows,

$$\sum_{j \in J_i} f_{i,j} = q_i \tag{2.1}$$

$$\sum_{i \in I} \sum_{j \in J_i} f_{i,j} \cdot \delta_{a,j} = v_a \;\; \forall a \in A \tag{2.2}$$

$$f_{i,j} \geq 0 \; \forall i \in I, \forall j \in J_i \tag{2.3}$$

where, q_i is demand between O-D pair $i \in I$, $f_{i,j}$ means traffic flow on path j between O-D pair $i \in I$, v_a is traffic flow on the link a and $\delta_{a,j}$ is the variable that equals 1 if path j uses link a and 0 otherwise.

The travel time on link a is assumed to be a function of the link flow $t_a(v_a)$. For simplicity, we assume that the link travel time is determined only by the amount of traffic flow over the link and that the travel time is monotonically increasing with respect to the amount of traffic flow. The path travel time is expressed as the sum of the link travel times as follows:

$$\pi_{i,j} = \sum_{j \in J_i} \delta_{a,j} \cdot t_a(v_a) \;\; \forall i \in I, \forall a \in A \tag{2.4}$$

Typically, in a static assignment model, the Bureau of Public Roads (BPR) function (Bureau of Public Roads, 1964) is used to calculate a link travel time.

$$t_a(v_a) = t_a^0 \cdot \left(1 + \alpha \cdot \left(\frac{v_a}{c_a}\right)^{\beta}\right) \forall a \in A \tag{2.5}$$

where the t_a^0 means the free flow travel time, v_a and c_a represent the link flow and capacity on link a, α and β are the parameters of BPR function, respectively.

The equilibrium condition based on Wardrop's first principle is formulated as follows:

$$f_{i,j} > 0, \pi_{i,j} = \mu_i \;\; \forall i \in I, \forall j \in J_i \tag{2.6}$$

$$f_{i,j} = 0, \pi_{i,j} \geq \mu_i \;\; \forall i \in I, \forall j \in J_i \tag{2.7}$$

where μ_i is minimum path cost between O-D pair $i \in I$. The equilibrium condition for UE assignment for a given O-D pair shows that the travel time on used paths is equal, and less than or equal to that of the travel time on the unused paths.

A problem that seeks a solution to achieve UE is called an equilibrium problem. Equilibrium problems can primarily be categorized into mathematical optimization problems (OP), nonlinear complementarity problems (NCP), variational inequality problems (VIP), and fixed-point problems (FPP). It is known that OP, NCP, VIP, and FPP have equivalent solutions under certain conditions. Furthermore, the mathematical properties of equilibrium problems, such as the uniqueness and existence of solutions, vary depending on the problem formulation.

Hereafter, the UE problem will be formulated as NCP, OP, and VIP, and the conditions under which they are equivalent will be demonstrated. Additionally, after presenting the formulations, the mathematical properties of each problem will be clarified.

First, the UE problem is formulated using the NCP. Generally, when a mapping $\mathbf{Z}: R_+^n \rightarrow R_+^n$ (R_+^n represents an n-dimensional space of nonnegative real numbers) is given, the NCP is defined as the following problem.

$$\text{Find } \boldsymbol{y}^* \in R_+^n \text{ such that } \boldsymbol{y}^* \cdot \mathbf{Z}(\boldsymbol{y}^*) = 0, \boldsymbol{y}^* \geq 0, \mathbf{Z}(\boldsymbol{y}^*) \geq 0 \tag{2.8}$$

The equilibrium conditions in (2.6) and (2.7) can also be expressed as the following complementary conditions:

$$f_{i,j} \cdot (\pi_{i,j} - \mu_i) = 0 \tag{2.9}$$

$$\pi_{i,j} - \mu_i \geq 0, f_{i,j} \geq 0 \tag{2.10}$$

From Equations (2.3) and (2.10), when Equation (2.9) is satisfied, Equations (2.6)–(2.7) are also satisfied. Therefore, these formulations are equivalent. Up to this point, the formulations have been described using only path flows and path travel times. Here, it is assumed that link travel times can be expressed as functions of the traffic flow on all links within the network based on Equation (2.5). At this time, the travel time on link a is represented as follows:

$$t_a = t_a(\mathbf{v}) \tag{2.11}$$

Where $\mathbf{v}$ is the vector of link flows. As defined in equations (2.2) and (2.4), link flows are expressed as the sum of the flows on the paths passing through the links, and path travel times are expressed as the sum of the travel times of the links composing the path. Therefore, using the vector of $\mathbf{f}$ which represents the path flows, the path travel time can be expressed as

$$\pi_{i,j} = \pi_{i,j}(\mathbf{f}) \tag{2.12}$$

However, it is important to note that the minimum travel time between O-D pairs μ_i remains constant. Since Equations (2.9) and (2.10) are defined for all O-D pairs and paths, the minimum travel time between O-D pairs and the path travel time can be expressed using the respective vectors $\boldsymbol{\pi}(\mathbf{f})$ and $\boldsymbol{\mu}$ as follows:

$$\mathbf{f} \cdot (\boldsymbol{\pi}(\mathbf{f}) - \boldsymbol{\mu}) = 0 \tag{2.13}$$

$$\boldsymbol{\pi}(\mathbf{f}) - \boldsymbol{\mu} \geq 0 \tag{2.14}$$

For simplicity, $\boldsymbol{\pi}(\mathbf{f}) - \boldsymbol{\mu}$ is represents as:

$$\mathbf{X}(\mathbf{f}) = \boldsymbol{\pi}(\mathbf{f}) - \boldsymbol{\mu} \tag{2.15}$$

Summarizing the above, the UE problem can be defined as the following NCP to find the optimal solution $\mathbf{f}^*$.

[UE-NCP]:

$$\text{Find } \mathbf{f}^* \in \Omega \text{ such that } \mathbf{f}^* \cdot \mathbf{X}(\mathbf{f}) = 0, \mathbf{X}(\mathbf{f}) \geq 0 \tag{2.16}$$

Unlike the general NCP shown in Equation (2.8), the non-negativity condition for $\mathbf{f}$ is not explicitly imposed. However, this condition is incorporated through the definition domain Ω of $\mathbf{f}$ and Ω is a convex set.

Next, the formulation of VIP is presented. Generally, when a convex set Γ and a mapping $\mathbf{Z}: \Gamma \to \Gamma$ are given, the VIP is defined as the following problem:

$$\text{Find } \boldsymbol{y}^* \in \Gamma \text{ such that } (\boldsymbol{y} - \boldsymbol{y}^*) \cdot \mathbf{Z}(\boldsymbol{y}^*) \geq 0 \quad \forall \mathbf{y} \in \Gamma \tag{2.17}$$

Thus, the VIP can be described as the problem of finding the point where the vector field $\mathbf{Z}$ intersects the convex set Γ. More precisely, since the endpoints of the convex set Γ are also included, it is the problem of finding a point $\boldsymbol{y}^*$ such that the vector $-\mathbf{Z}(\boldsymbol{y})$, as shown in Fig.2.1, is contained within the normal cone.

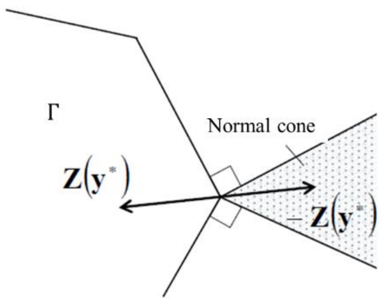


Fig.2.1 VIP and normal cone

From Equations (2.8) and (2.17), while the VIP is defined on a convex set, the NCP is defined on the nonnegative real number space. Therefore, it can be said that the VIP encompasses the NCP as a special case. Consequently, since Ω is a convex set, the NCP shown in equation (2.16) is equivalent to the following VIP:

[UE-VIP, path-based]:

$$\text{Find } \mathbf{f}^* \in \Omega \text{ such that } (\mathbf{f} - \mathbf{f}^*) \cdot \mathbf{X}(\mathbf{f}^*) \geq 0, \mathbf{X}(\mathbf{f}^*) \geq 0 \ \forall \mathbf{f} \in \Omega \tag{2.18}$$

This formulation is defined by path flow but can also be expressed in terms of link flow. $(\mathbf{f} - \mathbf{f}^*) \cdot \mathbf{X}(\mathbf{f}^*) \geq 0$ in Equation (2.18) can be expressed by following equation:

$$\sum_{i\in I}\sum_{j\in J_i}\left(\pi_{i,j}(\mathbf{f})-\mu_i\right)\cdot\left(f_{i,j}-f_{i,j}{}^*\right)\geq 0 \quad \forall\mathbf{f}\in\Omega \tag{2.19}$$

The minimum path travel time μ_i between O-D pair i is constant, and since the sum of the path flows between O-D pair i included in the convex set Ω equals the O-D demand q_i shown in Equation (2.1), the following relationship holds:

$$\sum_{i\in I}\sum_{j\in J_i}\mu_i\cdot\left(f_{i,j}-f_{i,j}{}^*\right)=\sum_{i\in I}\mu_i\cdot(q_i-q_i)=0 \quad \forall\mathbf{f}\in\Omega \tag{2.20}$$

Furthermore, as defined in Equations (2.2) and (2.4), link flows are represented as the sum of the path flows passing through the respective links, and path travel times are represented as the sum of the travel times of the links constituting the respective paths. Utilizing these definitions, the following relationship holds:

$$\begin{aligned}\sum_{i\in I}\sum_{j\in J_i}\pi_{i,j}(\mathbf{f})\cdot f_{i,j}&=\sum_{i\in I}\sum_{j\in J_i}\left(\sum_{a\in A}t_a(\mathbf{v})\cdot\delta_{a,j}\right)\cdot f_{i,j}\\&=\sum_{a\in A}t_a(\mathbf{v})\cdot v_a \quad \forall\mathbf{f}\in\Omega\end{aligned} \tag{2.21}$$

substituting equations (2.20) and (2.21) into Equation (2.19), the following inequality equivalent to Equation (2.19) is derived

$$\sum_{a\in A}(v_a-v_a{}^*)\cdot t_a(\mathbf{v}^*)\geq 0 \quad \forall\mathbf{v}\in\Omega \tag{2.22}$$

Therefore, since the path and link flows are included in the convex set Ω, we can say that the VIP shown in Equation (2.18) is equivalent to the VIP that finds the optimal link flow vector $\mathbf{v}^*$ shown below:

[UE-VIP, link-based]:

$$\text{Find } \mathbf{v}^*\in\Omega \text{ such that } (\mathbf{v}-\mathbf{v}^*)\cdot\mathbf{t}(\mathbf{v}^*)\geq 0, \mathbf{t}(\mathbf{v}^*)\geq 0\ \forall\mathbf{v}\in\Omega \tag{2.23}$$

Here $\mathbf{t}(\mathbf{v})$ is a vector containing the link travel times $t_a(\mathbf{v})$. Since the link flow $\mathbf{v}$ and the link travel times $\mathbf{t}(\mathbf{v})$ are defined within the nonnegative real space, it is evident that they are equivalent to the NCP in equation (2.23).

Next, the formulation using FPP is presented. The formulation of FPP is necessary to prove the existence of a solution to the VIP. Generally, when a convex set Γ and a mapping $\mathbf{Z}:\Gamma\rightarrow\Gamma$ are given, FPP is defined as the following problem.

$$\text{Find } \boldsymbol{y}^*\in\Gamma \text{ such that } \quad \boldsymbol{y}^*=\mathrm{Proj}_{\Gamma,\mathbf{S}}(\boldsymbol{y}^*-\mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*)) \tag{2.24}$$

where

$$\mathrm{Proj}_{\Gamma,\mathbf{S}}(\hat{\mathbf{w}}) = \underset{\mathbf{w}}{\mathrm{argmin}}\{(\mathbf{w}-\hat{\mathbf{w}})\cdot\mathbf{S}\cdot(\mathbf{w}-\hat{\mathbf{w}})^{\mathrm{T}}\} \tag{2.25}$$

here, $\mathbf{S}$ is an appropriate positive-definite matrix (for example, when it is a symmetric matrix with positive elements, it represents a scaling transformation of variables). $\mathrm{Proj}_{\Gamma,\mathbf{S}}(\hat{\mathbf{w}})$ is called the projection operator and geometrically represents the projection of the point $\hat{\mathbf{w}}$ onto the set Γ. The superscript T represents a transpose operation. Thus, the geometric meaning of FPP can be interpreted as shown in Fig. 2.2. By comparing Fig.2.1 and Fig.2.2, it is clear that their optimal solutions are identical.

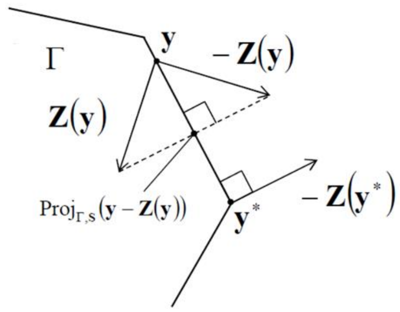


Fig.2.2 Illustration of FPP

In the UE problem, the variable ($\mathbf{f}$ or $\mathbf{v}$) is defined within the convex set Ω in the nonnegative real number space. In this case, the projection $\mathrm{Proj}_{\Gamma,\mathbf{S}}(\boldsymbol{y}^* - \mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*))$ onto Γ is expressed as follows:

$$\mathrm{Proj}_{\Gamma,\mathbf{S}}(\boldsymbol{y}^* - \mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*)) = [\boldsymbol{y}^* - \mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*)]_+ \tag{2.26}$$

where, $[\boldsymbol{y}^* - \mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*)]_+$ is an operator that applies $\max[(\boldsymbol{y}^* - \mathrm{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*)), 0]$ to each component of the vector $(\boldsymbol{y}^* - \mathbf{S}^{-1}\cdot\mathbf{Z}(\boldsymbol{y}^*))$. Thus, the UE problem can be expressed as the following FPP.

[UE-FPP, link-based]:

$$\text{Find } \mathbf{v}^* \in \Omega \text{ such that } (\mathbf{v}-\mathbf{v}^*)\cdot\mathbf{t}(\mathbf{v}^*) \geq 0, \mathbf{t}(\mathbf{v}^*) \geq 0\ \forall\mathbf{v}\in\Omega \tag{2.27}$$

The equivalence between FPP and VIP holds when path flows are used as variables.

Next, the formulation using OP is presented. The relationship between VIP and OP is derived. As is well known from the existence conditions for potentials in physics, for the mapping $\mathbf{Z}$ with respect to $\mathbf{y}$, which is included in a bounded convex set Γ, to be integrable, it is sufficient that its Jacobian $\nabla\mathbf{Z}(\mathbf{y})$ is symmetric, that is:

$$\frac{\partial Z_i}{\partial y_j} = \frac{\partial Z_j}{\partial y_i} \quad \forall i,j \tag{2.28}$$

When this condition is satisfied, the mapping $\mathbf{Z}$ has the following potential function $z(\mathbf{y})$, which satisfies the equation below:

$$z(\mathbf{y}) = \oint \mathbf{Z}(\mathbf{y}) \cdot \mathbf{dy} \quad \forall \mathbf{y} \in \Gamma \tag{2.29}$$

When a potential function $z(\mathbf{y})$as described above exists, the following lemma holds regarding the relationship between VIP and OP.

Lemma on the relationship between VIP and OP:

For the following VIP with respect to the mapping $\mathbf{Z}$ for $\mathbf{y}$ included in a bounded convex set Γ:

$$\text{Find } \boldsymbol{y}^* \in \Gamma \text{ such that } (\boldsymbol{y} - \boldsymbol{y}^*) \cdot \mathbf{Z}(\boldsymbol{y}^*) \geq 0 \quad \forall \mathbf{y} \in \Gamma \tag{2.30}$$

The solution $\boldsymbol{y}^*$ is equivalent to the solution of the following OP if the mapping $\mathbf{Z}$ has a potential function $z(\mathbf{y})$:

$$\min z(\mathbf{y}) \quad \text{s.t. } \mathbf{y} \in \Gamma \tag{2.31}$$

By using the graphical representation of the conditions for the solution of VIP and OP, it becomes almost evident that the above lemma holds. Now, the constraint $\mathbf{y} \in \Gamma$ of the OP expressed by Equation (2.26) is represented as

$$\mathbf{w}(\mathbf{y}) \geq 0 \tag{2.32}$$

where, $\mathbf{w}(\mathbf{y})$ is a vector with $\mathcal{M}$ components $w_m(y)(m \in \mathcal{M})$. The optimal conditions for such an OP are called the Karush-Kuhn-Tucker (KKT) conditions and are expressed as follows:

$$\nabla z(\boldsymbol{y}^*) = \sum_{m \in \mathcal{M}} \mathcal{p}_m \cdot \nabla w_m(\boldsymbol{y}^*) \tag{2.33}$$

$$\mathcal{p}_m \cdot w_m(\boldsymbol{y}^*) = 0, \mathcal{p}_m \geq 0, w_m(\boldsymbol{y}^*) \geq 0 \tag{2.34}$$

where, $\nabla z(\boldsymbol{y}^*)$ and $\nabla w_m(\boldsymbol{y}^*)$ represent the gradients of the objective function $z(\boldsymbol{y})$ and the constraint condition $w_m(\boldsymbol{y})$, respectively. According to equation (2.34), $\mathcal{p}_m$ is nonnegative when $w_m(\boldsymbol{y}^*) = 0$, and it takes the value 0 when $w_m(\boldsymbol{y}^*) > 0$. Thus, the solution of the OP satisfying the KKT conditions is shown in Fig 2.3. Here, the first and second constraint conditions are active.

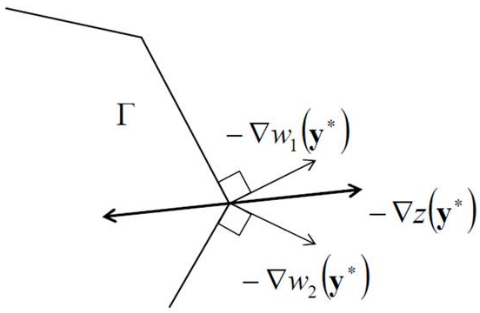


Fig.2.3 Illustration of KKT

By comparing Figures 2.1 and 2.3, it is clear that VIP and OP have the same solution.

The relationship between a general VIP and OP demonstrated can be replaced with equation (2.27), and the condition under which the link travel time $\mathbf{t}(\mathbf{v})$ has a potential function is expressed as follows:

$$\frac{\partial t_a}{\partial t_b} = \frac{\partial t_b}{\partial t_a} \quad \forall a, b \in A \tag{2.35}$$

Since the vector $\mathbf{v}$ is included in the convex set Ω, the VIP expressed in equation (2.27) satisfies the equivalent OP and is defined as:

[UE-OP, link-based]:

$$\min \oint \mathbf{t}(\mathbf{v}) \cdot \mathbf{dv} \quad \text{s.t} \quad \mathbf{v} \in \Omega \tag{2.36}$$

If path travel times cannot be expressed as the sum of the travel times of the links that constitute a path (a case not considered here), then the formulation [UE-VIP, Link-based] cannot be used. In such cases, the simplest case is often used in the UE problem, where link travel times depend only on the flow of the respective link:

$$\frac{\partial t_a}{\partial v_b} = 0 \quad \forall a \in A, \forall b \in A - a \tag{2.37}$$

In this case, the UE problem is defined by the following OP:

[UE-OP, link-independent]:

$$\min \sum_{a \in A} \int_0^{v_a} t_a(x) \cdot dx \tag{2.38}$$

In summary, the UE problem shown above is equivalent among OP, NCP, VIP, and FPP, and when the Jacobian of the link travel time mapping is symmetric, it can be shown to be equivalent to [UE-VIP, Link-based] OP.

2.2 Sensitivity Analysis

Sensitivity analysis of network traffic equilibrium has received ample attention thanks to its wide potential applicability to problems on transportation networks. Hall (1978) and Dafermos and Nagumey (1993) examined the continuity property of deterministic equilibrium link flows. Tobin and Friesz (1988), Yang (1997), and Cho et al. (2000) developed the gradient-based sensitivity analysis formulas of deterministic traffic equilibria, which have been applied to solve bi-level network optimization problems (Friesz et al., 1990, Yang and Bell, 1997); Leurent (1998) attempted to calculate the derivatives of the link flows in a bicriteria or cost versus time traffic equilibrium. Furthermore, Davis (1994), Yang et al. (2001), Yin and Miyagi (2001), Clark and Watling (2000, 2002) and Ceylan and Bell (2004) conducted sensitivity analysis of link flows in stochastic UE, with applications to network design, Origin-Destination (O-D) matrix estimation, pricing and network reliability analysis, Ying (2005) and Ying and Yang (2005) studied the sensitivity properties of a bimodal traffic network equilibrium with applications to inter-modal pricing problems. The reader may further refer to Qui and Magnanti (1989) and Patriksson (2003, 2004) for the development of sensitivity analysis by directional derivatives. These various sensitivity analysis methods are in general developed in line with the fundamental sensitivity analysis theory for nonlinear programming problems (Fiacco, 1983) or variational inequalities (Kyparisis, 1988; 1990). Amongst a wide variety of works mentioned above, Tobin and Friesz (1988) brought the classic nonlinear programming subject of sensitivity analysis to transportation research for the first time. They introduce the classical implicit function theorem to show the differentiability of equilibrium link flow solutions and develop explicit formulas for the calculation of link flow gradient in perturbation. It is well known that, due to its special structure, the formulation of traffic equilibrium problems usually includes path flow variables. However, the fact that the path flow pattern is usually not unique at equilibrium prohibits the direct application of a conventional sensitivity analysis approach to the traffic equilibrium problem. To overcome this difficulty of the non-uniqueness of equilibrium path flows, Tobin and Friesz (1988) developed an equivalent restriction approach that selects a non-degenerate extreme point in the feasible region of equilibrium path flows. The restricted UE problem has the desired unique properties required by the standard sensitivity analysis method for nonlinear programming problems. The approach is easily accessible by the rather intuitive implicit function theorem and is received as a useful tool for producing gradient information needed particularly in bi-level transportation network optimization problems. However, the gradient formula may provide a result that does not necessarily contain the information sought when the implicit function theorem is not applicable and may not provide a result when the equilibrium link flow solution is differentiable.

For the problem mentioned above, Yang (1997) shows that under the conditions that the link cost function vector is strongly monotone and once differentiable and there exists a set of strictly positive flows on the minimum cost or equilibrated paths, the classical implicit function theorem can always be applied to any set of a maximum number of linearly independent and equilibrated paths. As a result, the equilibrium link flows are differentiable under these milder conditions and their derivatives can always be obtained with the formula developed from the implicit function theorem, without dependence on network topology. Yang (1997) also pointed out that the strict complementarity conditions at equilibrium are not well defined in the literature due to the non-uniqueness of equilibrium path flows. The method proposed by Yang (1997) will be discussed next.

Considering a general situation that perturbation parameters exist in the link travel cost functions $\mathbf{t}(\mathbf{v}, \boldsymbol{\varepsilon})$ and O-D demand vector $\mathbf{q}(\boldsymbol{\varepsilon})$, where $\boldsymbol{\varepsilon}$ is a vector of all perturbations. In this case, the counterpart of the path based VI problem can be written respectively below according to Equation (2.18). Find $\mathbf{f}^*(\boldsymbol{\varepsilon}) \in \Omega_{\mathbf{f}}(\boldsymbol{\varepsilon})$ such that

$$\boldsymbol{\pi}(\mathbf{f}^*(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon})^{\mathrm{T}} \cdot \left(\mathbf{f} - \mathbf{f}^*(\boldsymbol{\varepsilon})\right) \geq 0 \ \ \forall \mathbf{f} \in \Omega_{\mathbf{f}}(\boldsymbol{\varepsilon}) \tag{2.38}$$

Now look at the general necessary conditions for the perturbed network equilibrium problem at $\boldsymbol{\varepsilon} = \mathbf{0}$ shown by Equation (2.38). The existence of equilibrium implies that there exists a solution $\mathbf{f}^*$ which is the full path flow vector to the following system equations (the first-order Karush-Kuhn-Tucker (KKT)):

$$\boldsymbol{\pi}(\mathbf{f}^*, \mathbf{0}) - \boldsymbol{\varrho} - \boldsymbol{\Lambda}^{\mathrm{T}} \cdot \boldsymbol{\mu} = \mathbf{0} \tag{2.39}$$

$$\boldsymbol{\varrho}^{\mathrm{T}} \cdot \mathbf{f}^* = \mathbf{0} \tag{2.40}$$

$$\boldsymbol{\Lambda} \cdot \mathbf{f}^* - \mathbf{q}(\mathbf{0}) = \mathbf{0} \tag{2.41}$$

$$\boldsymbol{\varrho} \geq \mathbf{0} \tag{2.42}$$

$$\mathbf{f}^* \geq \mathbf{0} \tag{2.43}$$

where $\boldsymbol{\pi}$ is the vector of path costs; $\mathbf{f}^*$ is the vector of equilibrium path flows; $\mathbf{q}$ is the vector of O-D demands; $\boldsymbol{\varrho}$ and $\boldsymbol{\mu}$ are the vectors of corresponding Lagrange multipliers respectively.

These system equations yield the derivative information of the UE problem. Yet, the aforementioned necessary condition is not fulfilled. As previously noted, the UE problem represents an asymmetric equilibrium assignment issue. While the solution to the UE problem concerning link flows is unique when the Jacobian matrix of the cost function is positive definite, the solutions related to path flows lack uniqueness. This non-uniqueness becomes an obstacle in derivative computation due to the description of the equation system in terms of path flows. Hence, when the derivative is computed unique solutions for path flows must be obtained. In this context, we employ the restricted network equilibrium approach that has been proposed by Tobin and Friesz (1988) and later enhanced by Yang and Bell (2007), to derive the expressions for the derivatives. Consequently, the equation system simplifies to:

$$\boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) - \boldsymbol{\Lambda}^{0\mathrm{T}} \cdot \boldsymbol{\mu} = \mathbf{0} \tag{2.44}$$

$$\boldsymbol{\Lambda}^0 \cdot \mathbf{f}^{0*} - \mathbf{q}(\mathbf{0}) = \mathbf{0} \tag{2.45}$$

where the superscript 0 represents the corresponding reduced vectors and matrix.

Differentiating both sides of Equations (2.44) and (2.45) with respect to the perturbations $\boldsymbol{\varepsilon}$ $(\boldsymbol{\varepsilon} = \mathbf{0})$ results in the following:

$$\begin{bmatrix} \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) & -\boldsymbol{\Lambda}^{0\mathrm{T}} \\ \boldsymbol{\Lambda}^0 & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}^0 \\ \nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} -\nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) \\ \nabla_{\boldsymbol{\varepsilon}} \mathbf{q}(\mathbf{0}) \end{bmatrix} \tag{2.46}$$

So, the Jacobian matrix of Equations (2.44) and (2.45) with respect to $(\mathbf{f}^0, \boldsymbol{\mu})$ evaluated at $\boldsymbol{\varepsilon} = \mathbf{0}$ is shown as:

$$\boldsymbol{M}_{\mathbf{f}^0,\boldsymbol{\mu}}(\mathbf{0}) = \begin{bmatrix} \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) & -\boldsymbol{\Lambda}^{0\mathrm{T}} \\ \boldsymbol{\Lambda}^0 & \mathbf{0} \end{bmatrix} \tag{2.47}$$

Suppose:

$$\boldsymbol{M}_{\mathbf{f}^0,\boldsymbol{\mu}}(\mathbf{0})^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \tag{2.48}$$

Referring to Yang and Bell (2007), the following Equations can be derived:

$$\mathbf{B}_{22} = [\boldsymbol{\Lambda}^0 \cdot \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0})^{-1} \cdot \boldsymbol{\Lambda}^{0\mathrm{T}}]^{-1} \tag{2.49}$$

$$\mathbf{B}_{12} = \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0})^{-1} \cdot \boldsymbol{\Lambda}^{0\mathrm{T}} \cdot \mathbf{B}_{22} \tag{2.50}$$

$$\mathbf{B}_{21} = -\mathbf{B}_{22} \cdot \boldsymbol{\Lambda}^0 \cdot \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0})^{-1} \tag{2.51}$$

$$\mathbf{B}_{11} = \nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0})^{-1} \cdot [\mathbf{I} + \boldsymbol{\Lambda}^{0\mathrm{T}} \cdot \mathbf{B}_{21}] \tag{2.52}$$

where

$$\nabla_{\mathbf{f}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) = \boldsymbol{\Delta}^{0\mathrm{T}} \cdot \nabla_{\mathbf{v}} \mathbf{t}^0(\mathbf{v}^*, \mathbf{0}) \cdot \boldsymbol{\Delta}^0 \tag{2.53}$$

$\mathbf{t}^0(\mathbf{v}^*, \mathbf{0})$ is the vector of link costs and $\mathbf{I}$ represents the identity matrix of appropriate dimension.

Finally, the derivative of path flows with respect to $\mathbf{f}$ at $\boldsymbol{\varepsilon} = \mathbf{0}$ holds:

$$\nabla_{\boldsymbol{\varepsilon}} \mathbf{f}^0 = -\mathbf{B}_{11} \cdot \nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) + \mathbf{B}_{22} \cdot \nabla_{\boldsymbol{\varepsilon}} \mathbf{q}(\mathbf{0}) \tag{2.54}$$

where

$$\nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\pi}^0(\mathbf{f}^*, \mathbf{0}) = \boldsymbol{\Delta}^{0\mathrm{T}} \cdot \nabla_{\boldsymbol{\varepsilon}} \mathbf{t}^0(\mathbf{v}^*, \mathbf{0}) \tag{2.55}$$

Since $\mathbf{v} = \boldsymbol{\Delta} \cdot \mathbf{f}$, the following relationship can be obtained:

$$\nabla_{\boldsymbol{\varepsilon}} \mathbf{v}^0 = \boldsymbol{\Delta}^0 \cdot \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}^0 \tag{2.56}$$

We propose an efficient method to solve the NDPs based on the sensitivity analysis method mentioned above. Which can establish the local linear approximation of the UE link flows to reduce the computation time when faced with large-scale networks.

3. Development of a sensitivity analysis-based algorithm for a traffic state estimation problem

This chapter proposes a computationally efficient algorithm for traffic estimation problems 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 UE constraints. Solving the FIMLE problem under equilibrium constraints requires much computation time. Therefore, a solution algorithm based on the sensitivity analysis for the UE is proposed in this chapter. We estimate the traffic states in the Asahikawa network in Hokkaido, Japan by applying the algorithm proposed in this chapter.

3.1 Introduction

The field of traffic planning has been significantly impacted by the progress of information and communication technology. For instance, traffic data gathering was once performed manually, but the development of traffic data collection technology enables us to collect comprehensively road traffic states. To estimate road traffic states, it is important to develop a precise system that can address big traffic data.

Indicators representing the state of traffic can be the number of cars, the queue length, and the speed of cars. Travel time (delay time) is an important indicator for both drivers and road administrators.

For travel time estimation, floating car data has been used in recent studies. It has been used for arterial travel time estimation in some studies. Liu and Ma (2009) proposed a method for the arterial travel time estimation based on the probe car model. Zheng and Van Zuylen (2013) proposed a VISSIM simulation model to estimate each link travel time using partial link data. Jenelius and Koutsopoulos (2013) estimated the travel time of a portion of the arterial route in a city by using sparse floating car data. However, the methods used in these studies cannot apply to network-level travel time estimation.

Several studies have been conducted to estimate traffic states in networks. Hara et al. (2018) estimated link speeds that follow a multivariate normal distribution from accumulated and observed floating car data and developed a method to add the speeds on road links where floating car data is not available. In their study, no traffic flows, e.g., O-D flows, path flows and link flows in a road network, were addressed. Uchida and Kato (2017) adopted a normal distribution for each O-D demand in a road network. Taylor series expansion was performed to calculate approximately both the mean travel time and the travel time variance of a link that has link flow following the normal distribution. The approximation method based on the Taylor series expansion greatly increases the computation burden. To reduce the calculation burden, Tani and Uchida (2018) adopted a lognormal distribution for each O-D demand and assumed that the corresponding link flow follows the lognormal distribution approximately. When link flow follows the lognormal distribution, the corresponding link travel time follows a shifted lognormal distribution. However, the link flow does not follow the lognormal distribution analytically because it was calculated as the summation of the path flows that follow the lognormal distribution. Therefore, Tani et al. (2021) adopted the lognormal distribution for total O-D demands so that stochastic link delay times in the network follow a multivariate lognormal distribution. However, due to this assumption, mutually and statistically correlated O-D demands which may not be realistic were generated in the network. Still, the assumption employed by Tani and Uchida (2021) brings about an advantage in estimating correctly network parameters, e.g., each O-D demand in the network, over the methods proposed in the literature. In this study, following Tani et al. (2021), the assumption that total O-D demands follow the lognormal distribution is employed to express the link delay times that follow the multivariate lognormal distribution. A method proposed in this study utilizes the mathematical property that the natural logarithm of link delay times follows a multivariate normal distribution. Thus, network parameters in this study can be estimated by using a FIMLE based on the multivariate normal distribution. When solving the FIMLE in this study, UE constraints need to be addressed.

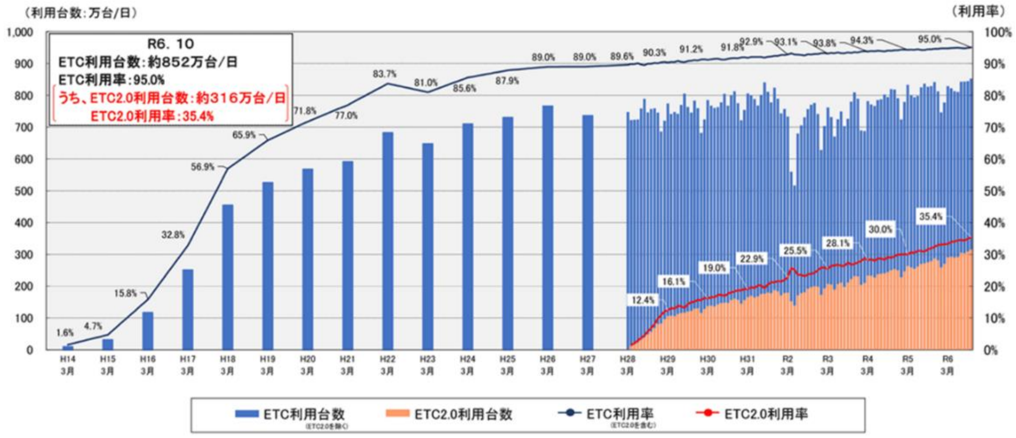


Fig. 3.1 Trends in ETC and ETC 2.0 usage in Japan.

(Source: Ministry of Land, Infrastructure, Transport and Tourism, Japan)

To estimate traffic states in a road network, floating car data is collected from vehicles equipped with probe devices. The floating car data provides us with various traffic data such as the speed, position, and direction of each vehicle. Now in Japan, floating car data can be collected and available, but there are still some problems. The market penetration rate of cars from which floating car data can be collected is still low, and therefore it is impossible to collect data for each link in the network. Fig.3.1 highlights the trends in ETC and ETC 2.0 usage in Japan. The ETC usage rate has reached 95%, with about 852 million vehicles using ETC daily. However, ETC is limited to the toll road networks, it cannot capture traffic data on nontoll roads. Compared to ETC, ETC2.0 can gather data from all roads by using advanced technologies, but the penetration rate of ETC2.0 remains low, which is only 35.4%. Therefore, there are many links in the road network where no floating data is collected. To cope with such road networks, the traffic state estimation is formulated as a FIMLE problem under 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. Chapter 3.2 formulates a network model including the definition of traffic flows, travel times, and a UE traffic assignment problem. Chapter 3.3 introduces the FIMLE problem UE constraints. Chapter 3.4 proposes an efficient solution algorithm based on sensitivity analysis for UE, and demonstrates and verifies the proposed algorithm in a test network and real network. Finally, the concluding remarks and future directions are shown in Chapter 3.5.

3.2 UE under stochastic demand

3.2.1 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.

Notation	Representation
$G(N,A)$	The network comprised of set of nodes N and set of links A
I	Set of O-D pairs
J_i	Set of paths for O-D pair i
Q	Stochastic total O-D demands
Q_i	Stochastic demand between O-D pair i
T_a	Stochastic travel time on the link a
T'_a	Stochastic delay time on the link a
C_a	Stochastic traffic capacity on the link a
p_i	The ratio of demand for O-D pair i to the total O-D demands
p_{ij}	Probability of selecting path j in O-D pair i
Ξ_{ij}	Stochastic travel time on path j between O-D pair i
F_{ij}	Stochastic traffic flow on path j between O-D pair i
V_a	Stochastic traffic flow on the link a
t_a^0	Free flow travel time on the link a

c_{ij}	Mean travel time on path j between O-D pair i
δ_{aj}	The variable that equals 1 if path j uses link a and 0 otherwise
μ_i	Minimum mean path travel time between O-D pair i
Δ	Link/path incidence matrix
$\mathbf{Q}$ (or $\mathbf{q}$)	Vector of stochastic (or mean) O-D demands
$\mathbf{F}$ (or $\mathbf{f}$)	Vector of stochastic (or mean) path flows
$\boldsymbol{\mu}$	Vector of minimum mean path travel times
$\boldsymbol{\Xi}$	Vector of stochastic path travel times
$\boldsymbol{\pi}$	Vector of mean path travel times

3.2.2 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

$$Q \sim LN(\mu, \sigma^2) \tag{3.1}$$

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

$$Q_i = p_i \cdot Q \ \forall i \in I \tag{3.2}$$

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

$$Q_i \sim LN\left(\mu_{Q_i}, \sigma_{Q_i}^2\right) \forall i \in I \tag{3.3}$$

Its parameters are:

$$\mu_{Q_i} = \mu + \ln(p_i) \ \ \forall i \in I \tag{3.4}$$

$$\sigma_{Q_i}^2 = \sigma^2 \ \ \forall i \in I \tag{3.5}$$

and vector of mean O-D demand is defined as follows:

$$\boldsymbol{\mu}_Q = \left[\mu_{Q_1} \dots \mu_{Q_i} \dots, \mu_{Q_{|I|}}\right]^{\mathrm{T}} \tag{3.6}$$

the path flow F_{ij} is calculated as the product of the O-D demand Q_i and p_{ij} which is the probability that drivers who travel between O-D pair i select route j.

$$F_{i,j} = p_{i,j} \cdot Q_i = \ p_{i,j} \cdot p_i \cdot Q \ \ \forall \ i \in I, \forall j \in J_i \tag{3.7}$$

the path flows follow the following log-normal distribution:

$$F_{i,j} \sim LN\left(\mu_{F_{i,j}}, \sigma^2_{F_{i,j}}\right)\ \forall\, i \in I, \forall j \in J_i \tag{3.8}$$

where

$$\mu_{F_{i,j}} = \mu + \ln\left(p_{i,j} \cdot p_i\right)\ \forall\, i \in I, \forall j \in J_i \tag{3.9}$$

$$\sigma^2_{F_{i,j}} = \sigma^2\ \forall\, i \in I, \forall j \in J_i \tag{3.10}$$

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

$$V_a = \sum_{i \in I} \sum_{j \in J_i} F_{i,j} \cdot \delta_{a,j} = \sum_{i \in I} \sum_{j \in J} p_{i,j} \cdot p_i \cdot Q \cdot \delta_{aj} = \hat{p}_a \cdot Q\ \forall a \in A \tag{3.11}$$

where

$$\hat{p}_a = \sum_{i \in I} \sum_{j \in J_i} \delta_{a,j} \cdot p_{i,j} \cdot p_i\ \forall a \in \mathrm{A} \tag{3.12}$$

Its parameters are given as follows:

$$V_a \sim LN\left(\mu_{V_a}, \sigma^2_{V_a}\right)\ \forall a \in A \tag{3.13}$$

$$\mu_{V_a} = \mu + \ln(\hat{p}_a)\ \forall a \in A \tag{3.14}$$

$$\sigma^2_{V_a} = \sigma^2\ \forall a \in A \tag{3.15}$$

3.2.3 Travel time

In this study, the link travel time is defined by the following BPR function below:

$$T_a(V_a, C_a) = t_a^0 \cdot \left(1 + \alpha \cdot \left(\frac{V_a}{C_a}\right)^{\beta}\right)\ \forall a \in A \tag{3.16}$$

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 is calculated as the following shifted log-normal distribution.

$$T_a(V_a, C_a) = t_a^0 + \frac{t_a^0 \cdot \alpha_a}{{C_a}^{\beta_a}} \cdot (V_a)^{\beta_a} = t_a^0 + \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a}\ \forall a \in A \tag{3.17}$$

$$\gamma_a = t_a^0 \cdot \alpha_a\ \ \forall a \in A \tag{3.18}$$

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 (Tani and Uchida, 2018).

$$T'_a = \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a} \sim LN\left(\mu_{T'_a}, \sigma^2_{T'_a}\right) \ \forall a \in A \tag{3.19}$$

where

$$\mu_{T'_a} = \ln(\gamma_a) + \beta_a \cdot \ln(\hat{p}_a) + \beta_a \cdot \left(\mu - \mu_{C_a}\right) \forall a \in A \tag{3.20}$$

$$\sigma^2_{T'_a} = (\beta_a)^2 \cdot \left(\sigma^2 + \sigma^2_{C_a}\right) \forall a \in A \tag{3.21}$$

The link travel time follows a shifted log-normal distribution shown below

$$T_a = t^0_a + T'_a \ \ \forall a \in A \tag{3.22}$$

Then, the mean of link travel time is given by:

$$E[T_a] = t^0_a + \mathrm{E}(T'_a) = t^0_a + \exp\left(\mu_{T'_a} + \frac{1}{2}\sigma^2_{T'_a}\right) \forall a \in A \tag{3.23}$$

The stochastic path travel time is expressed as the sum of the link travel times as follows:

$$\Xi_{i,j} = \sum_{a \in A} T_a \cdot \delta_{a,j} \forall i \in I, \forall j \in J_i \tag{3.24}$$

where $\delta_{a,j}$ is the variable that equals 1 if path j uses the link and 0 otherwise. Path travel time is calculated as:

$$\mathrm{E}\left[\Xi_{i,j}\right] = \sum_{a \in A} \mathrm{E}[T_a] \cdot \delta_{a,j} \ \forall i \in I, \forall j \in J_i \tag{3.25}$$

The vector of path travel times is expressed as follows:

$$\boldsymbol{\Xi} = \left[\Xi_{1,1} \ \dots \ \Xi_{i,j} \ \dots \Xi_{|I|,|J_i|}\right]^{\mathrm{T}} \tag{3.26}$$

3.2.4 UE assignment model

The mean path travel time is calculated as the sum of the mean of link travel times as follows:

$$\pi_{i,j} = \sum_{a \in A} \mathrm{E}[T_a] \cdot \delta_{a,j} \ \ \forall i \in I, \forall j \in J_i \tag{3.27}$$

$$\boldsymbol{\pi} = \left[\pi_{1,1}, \dots, \pi_{i,j}, \dots, \pi_{|I|,|J_i|}\right]^{\mathrm{T}} \tag{3.28}$$

the vectors of path flows, O-D demands, and the mean vectors of path traffic flows and O-D demands are respectively represented by

$$\mathbf{F} = \left[F_{1,1}, \dots \ F_{i,j} \ \dots F_{|I|,|J_i|}\right]^{\mathrm{T}} \tag{3.29}$$

$$\mathbf{f} = \left[f_{1,1} \ \dots \ f_{i,j}, \dots f_{|I|,|J_i|}\right]^{\mathrm{T}} \tag{3.30}$$

$$\mathbf{Q} = \left[Q_1, \dots Q_i \ \dots Q_{|I|}\right]^{\mathrm{T}} \tag{3.31}$$

$$\mathbf{q} = \left[q_1, \dots \ q_i \ \dots q_{|I|}\right]^{\mathrm{T}} \tag{3.32}$$

When the vector of mean path flows is $\mathbf{f} \geq \mathbf{0}$, the minimum path travel cost between O-D pair i is expressed as follows:

$$\mu_i = min\{\pi_{i,j}, \forall j \in J_i\} \ \forall i \in I \tag{3.33}$$

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

$$\boldsymbol{\mu} = \left[\mu_1 \dots \ \mu_i \ \dots \mu_{|I|}\right]^{\mathrm{T}} \tag{3.34}$$

For each path $j \in J_i$ between O-D pair $i \in I$, if the mean vector of path flows $\mathbf{f}$ is in Wardropian equilibrium (UE) the following conditions hold:

$$f_{i,j} > 0 \rightarrow \pi_{i,j} = \mu_i \ \forall i \in I, \forall j \in J_i \tag{3.35}$$

$$f_{i,j} = 0 \rightarrow \pi_{i,j} \geq \mu_i \ \forall i \in I, \forall j \in J_i \tag{3.36}$$

The equilibrium conditions for a given O-D pair show that generalized costs of used routes are all equal and are smaller than or equal to the costs of unused routes.

The above UE conditions can be formulated as the following VI problem in terms of path flow variables (Smith, 1979):

$$\sum_{i \in I} \sum_{j \in J_i} c_{i,j} \cdot (f_{i,j} - f_{i,j}{}^*) \geq 0 \tag{3.37}$$

$$\sum_{i \in I} \sum_{j \in J_i} f_{i,j} = q_i \tag{3.38}$$

$$f_{i,j} \geq 0 \ \forall i \in I, \forall j \in J_i \tag{3.39}$$

where $f_{i,j}{}^*$ is the path flow in the UE state.

In addition, Equations (3.37)-(3.39) can be expressed as follows:

$$\begin{gathered} \boldsymbol{\pi}(\mathbf{f}^*)^{\mathrm{T}} \cdot (\mathbf{f} - \mathbf{f}^*) \geq 0 \\ \mathbf{f} \geq 0 \end{gathered} \tag{3.40}$$

where $\mathbf{f}^*$ is the vector of path flow at UE state.

3.3 Definitions of FIMLE problem

The preceding chapters showed the definition of the network flow models, and a UE assignment model, which forms equilibrium constraints for the objective function. FIMLE can deal with missing data by using all available information, making it robust in scenarios where traditional Maximum Likelihood

Estimation (MLE) cannot handle the missing data. Therefore, the FIMLE is introduced in this chapter, referring to Tani and Uchida (2021).

Vector of link delay times is:

$$\mathbf{T}' = \left[T_1' \dots T_a' \dots T_{|A|}'\right]^{\mathrm{T}} \tag{3.41}$$

that follows the following multivariate log-normal distribution.

$$\mathbf{T}' \sim \mathrm{MVLN}\left(\boldsymbol{\mu}_{\mathrm{T}_a'}, \boldsymbol{\Sigma}_{\mathrm{T}_a'}\right) \tag{3.42}$$

where

$$\boldsymbol{\mu}_{T_a'} = \left[\mu_{T_1'}\ \mu_{T_2'} \dots \mu_{T_{|A|}'}\right]^{\mathrm{T}} \tag{3.43}$$

$$\boldsymbol{\Sigma}_{T'} = \begin{bmatrix} \sigma_{T_1'}^2 & \cdots & \sigma_{T_1', T_{|A|}'} \\ \vdots & \ddots & \vdots \\ \sigma_{T_{|A|}', T_1'} & \cdots & \sigma_{T_{|A|}'}^2 \end{bmatrix} \tag{3.44}$$

$$\sigma_{T_a', T_b'} = \beta_a \cdot \beta_b \cdot \sigma^2 \tag{3.45}$$

A 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 $\mathbf{m}_g$ and the observed value vector $\hat{\mathbf{d}}_g$ for the observation on day $g \in \{1, \dots, G\}$. $\hat{\mathbf{d}}_g$ is a vector representing the traffic conditions observed across the road network on day $g \in \{1, \dots, G\}$, defined as:

$$\hat{\mathbf{d}}_g = \left[\hat{t}_1 \dots\ \hat{t}_a \dots\ \hat{t}_{|A|}\right]^{\mathrm{T}}\ \forall\, g \in \{1, \dots, G\} \tag{3.46}$$

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 5 [min] at link 2 and no observed data at link 3, then the $\hat{\mathbf{d}}_g$ is expressed as follows:

$$\hat{\mathbf{d}}_g = [\ln(10)\ \ \ln(5)\ -\infty]^{\mathrm{T}} \tag{3.47}$$

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.

The observed state vector $\mathbf{m}_g$ shown by (3.48) 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.

$$\mathbf{m}_g = \left[m_{a,g}, \forall a \in A\right]^{\mathrm{T}} \tag{3.48}$$

where

$$m_{a,g} = \begin{cases} 1, & if\ \hat{t}_a \neq -\infty \\ 0, & otherwise \end{cases} \tag{3.49}$$

The $\mathbf{m}_g$ corresponding to Equation (3.49) is:

$$\mathbf{m}_g = [1 \quad 1 \quad 0]^{\mathrm{T}} \tag{3.50}$$

$\mathbf{M}_g$ 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, $\mathbf{m}_g = [1 \quad 1 \quad 0]^{\mathrm{T}}$, as shown in Equation (3.50), $\mathbf{M}_g$ can be expressed as:

$$\mathbf{M}_g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.51}$$

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 $\mathbf{T}' \sim \mathrm{MVLN}(\boldsymbol{\mu}_{\mathrm{T}'_{\mathrm{a}}}, \boldsymbol{\Sigma}_{\mathrm{T}'_{\mathrm{a}}})$, and the natural logarithm of link delays follows a multivariate normal distribution, denoted as $\ln(\mathbf{T}') \sim \mathrm{MVN}(\boldsymbol{\mu}_{\mathrm{T}'_{\mathrm{a}}}, \boldsymbol{\Sigma}_{\mathrm{T}'_{\mathrm{a}}})$. Here, the parameters are the same for both. Therefore, the estimates obtained by FIMLE using link delay time that follow a multivariate log-normal distribution as likelihoods are the same as those obtained by FIMLE of a multivariate normal distribution. The method used in this study takes advantage of this property to estimate the parameters by using FIMLE of a multivariate normal distribution.

The FIMLE problem in this study is defined as:

$$\max \quad L\left(\mu, \sigma^2 \middle| \hat{\mathbf{d}}_g \forall g \in \{1, \dots, G\}\right) = \prod_{g \in \{1,\dots,G\}} \frac{\exp\left(-\frac{1}{2}\left(\hat{\mathbf{d}}_g - \boldsymbol{\mu}_g(\mu)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_g^{-1}(\sigma^2)\left(\hat{\mathbf{d}}_g - \boldsymbol{\mu}_g(\mu)\right)\right)}{2\pi^{\frac{n(g)}{2}} \left|\boldsymbol{\Sigma}_g(\sigma^2)\right|^{\frac{1}{2}}} \tag{3.52}$$

s.t. (3.37)-(3.39), where

$$\boldsymbol{\Sigma}_g(\sigma^2) = \mathbf{M}_g \boldsymbol{\Sigma}_{T'_a}(\sigma^2) \mathbf{M}_g^{\mathrm{T}} \ \forall g \in \{1, \dots, G\} \tag{3.53}$$

$$\boldsymbol{\mu}_g(\mu) = \mathbf{M}_g \boldsymbol{\mu}_{T'_a}(\mu) \ \forall g \in \{1, \dots, G\} \tag{3.54}$$

$$\hat{\mathbf{d}}_g = \mathbf{M}_g \hat{\mathbf{d}}_g \ \forall g \in \{1, \dots, G\} \tag{3.55}$$

3.4 Solution algorithm based on sensitivity analysis for UE flows

When solving the likelihood maximization problem mentioned in Chapter 3.3 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. The perturbation parameter for O-D demands will be denoted by $\boldsymbol{\varepsilon}$. The perturbated O-D demand vector $\mathbf{q}(\boldsymbol{\varepsilon})$ is represented by:

$$\mathbf{q}(\boldsymbol{\varepsilon}) = \mathbf{q} + \boldsymbol{\varepsilon} \tag{3.56}$$

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

$$\tilde{\mathbf{f}}(\mathbf{q}(\boldsymbol{\varepsilon})) = \mathbf{f}^*(\mathbf{q}) + \boldsymbol{\nabla}_{\boldsymbol{\varepsilon}}\mathbf{f}^0 \cdot \boldsymbol{\varepsilon} \tag{3.57}$$

where the superscript 0 represents the corresponding reduced vectors and matrix. $\boldsymbol{\nabla}_{\boldsymbol{\varepsilon}}\mathbf{f}^0$ can be calculated by $\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}^0 = -\mathbf{B}_{11} \cdot \nabla_{\boldsymbol{\varepsilon}}\mathbf{c}^0(\mathbf{f}^*, \mathbf{0}) + \mathbf{B}_{22} \cdot \nabla_{\boldsymbol{\varepsilon}}\mathbf{q}(\mathbf{0})$ in Equation (2.28). However, $\boldsymbol{\varepsilon}$ is denoted as perturbation parameter for O-D demands in this chapter, therefore, $\nabla_{\boldsymbol{\varepsilon}}\mathbf{c}^0(\mathbf{f}^*, \mathbf{0}) = \mathbf{0}$, then $\boldsymbol{\nabla}_{\boldsymbol{\varepsilon}}\mathbf{f}^0 = \mathbf{B}_{22} \cdot \nabla_{\boldsymbol{\varepsilon}}\mathbf{q}(\mathbf{0})$. The corresponding link flow vector is defined as:

$$\mathbf{v}(\boldsymbol{\varepsilon}) = \Delta \cdot \tilde{\mathbf{f}}(\mathbf{q}(\boldsymbol{\varepsilon})) \tag{3.58}$$

It is complex to solve the likelihood function with equilibrium constraints directly. Solving this problem requires a large amount of computation time. However, the method of sensitivity analysis converts the equilibrium constraints to linear constraints and makes the problem easier to solve. The FIMLE problem based on sensitivity analysis can be written as follows.

$$\max L\left(\mu, \sigma^2, \boldsymbol{\varepsilon} \middle| \hat{\mathbf{d}}_g \forall g \in \{1, \dots, G\}\right) = \prod_{g \in \{1,\dots,G\}} \frac{\exp\left(-\frac{1}{2}\left(\hat{\mathbf{d}}_g - \boldsymbol{\mu}_g(\boldsymbol{\varepsilon}, \mu)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_g^{-1}(\sigma^2)\left(\hat{\mathbf{d}}_g - \boldsymbol{\mu}_g(\boldsymbol{\varepsilon}, \mu)\right)\right)}{2\pi^{\frac{n(g)}{2}} \left|\boldsymbol{\Sigma}_g(\sigma^2)\right|^{\frac{1}{2}}} \tag{3.59}$$

s.t.

$$\sum_{i \in I} q_i(\boldsymbol{\varepsilon}) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{3.60}$$

The solution algorithm for estimating μ, $,\sigma^2$ and $\boldsymbol{\varepsilon}$ is summarized as follows.

Step 0 Initialization	Set $n = 1$, the initial parameters μ and σ^2, and initial O-D shares p_i, i.e., $q = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ and $q_i = p_i \cdot q$.
Step 1 UE assignment	UE traffic assignment is performed to obtain equilibrium traffic flows.
Step 2 Sensitivity analysis	Calculate the $\boldsymbol{\nabla}_{\boldsymbol{\varepsilon}}\mathbf{f}^0$ in Equation (3.57) by the sensitivity analysis based on the equilibrium traffic flows.
Step 3 FIMLE	Solve the FILME problem defined in this study. Optimal parameters μ^*, σ^{2*} and $\boldsymbol{\varepsilon}^*$ are then obtained. Update mean total O-D demands and each

O-D demand as $q = \exp\left(\mu^* + \frac{\sigma^{2*}}{2}\right)$ and $q_i = q_i + \varepsilon_i^*$, respectively.

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

3.5 Numerical calculation for test network

3.5.1 Definition of test network

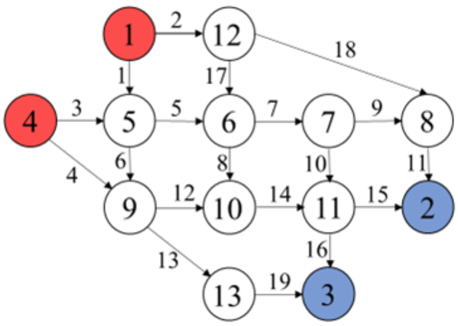


Fig. 3.2 Test network.

Table 3.1. Paths in test network.

O-D pair	Path (link sequences)
1-2	2-18-11
	1-5-7-9-11
	1-6-12-14-15
1-3	2-17-7-10-16
	2-17-8-14-16
4-2	3-5-7-10-15
	3-6-12-14-15
	4-12-14-15
4-3	3-6-12-14-16
	4-13-19

The first numerical calculation was conducted in the test network shown in Fig. 3.1. The test network is composed of 4 O-D pairs, 10 paths and 19 directed links as shown in Table 3.1. The test network is used to illustrate the FIMLE problem with equilibrium constraints. The link cost function below is used.

$$t_a(V_a, C_a) = t_a^0 \cdot \left(1 + 1 \cdot \left(\frac{V_a}{C_a}\right)^2\right) \tag{3.61}$$

The free flow travel time and link capacity for all links are 3 [min] and 1000 [pcu/h], and the coefficient of variation for each link capacity is $cv = 0.1$, where "pcu" represents the passenger car unit.

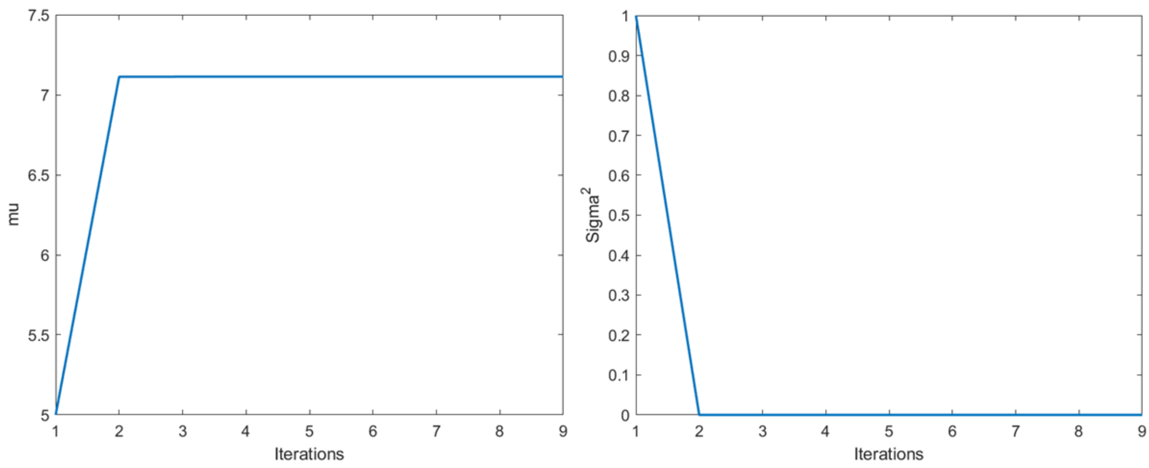
The link delays presented in Table 3.2 are assumed as observation data, where '-' indicates there is no observation data. The initial values of μ and σ^2 are 5 and 1, respectively. Thus, the initial mean total demand is calculated as 244.6919. We set $p_i = 0.25$ for each O-D pair, therefore each mean O-D demand is 61.1730.

Table 3.2. Delay time [sec] for test network.

link number	day 1	day 2	link number	day 1	day 2
1	18	16	11	20	17
2	15	20	12	-	18
3	-	14	13	17	15
4	17	15	14	18	20
5	18	-	15	15	14
6	15	18	16	15	15
7	15	20	17	18	16
8	18	-	18	20	18
9	20	16	19	15	18
10	-	18			

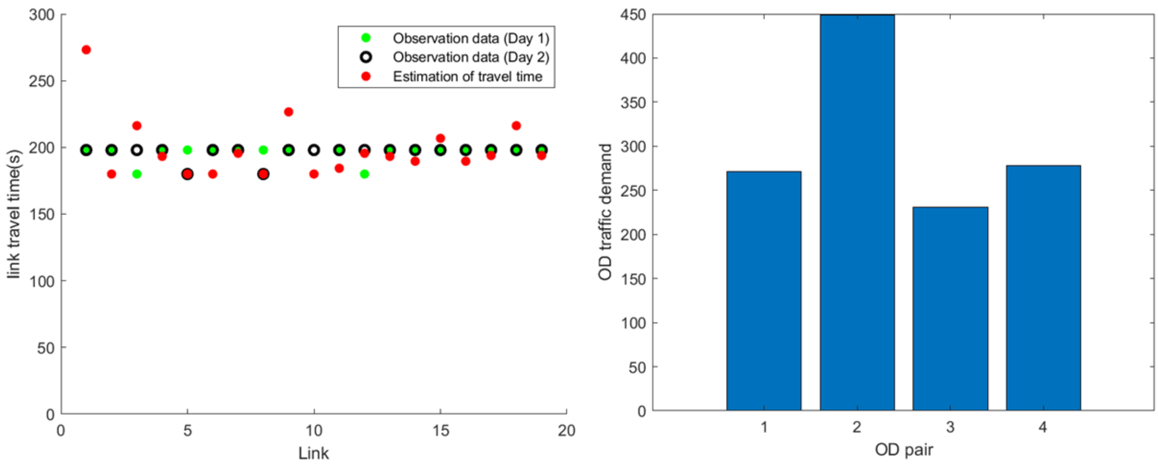
3.5.2 Results of test network

The convergence process of μ and σ^2 is shown in Fig. 3.3. The estimation results of them are 7.1133 and 1.8×10^{-8}, respectively.



(a) (b)

Fig. 3.3 Convergence process of μ (a) and σ^2 (b) for the test network.



(a) (b)

Fig. 3.4 Results of mean link travel time (a), and mean demand for each O-D pair (b).

The estimation results of mean travel time for each link are depicted in Fig. 3.4 (a). It is shown that the estimation results are in relative agreement with the observation data. As can be seen from Fig. 3.4 (a), it is not possible to estimate exactly the same travel times as the observed values for all links. This is due to the equilibrium constraints in which path flows are calculated considering O-D pair settings and drivers' route choice behavior. Fig. 3.4 (b) depicts the mean demand for each O-D pair calculated by the estimated O-D demand shares: $\mathbf{p} = [0.2205\ 0.3653\ 0.1881\ 0.2261]^{\mathrm{T}}$.

3.6 Application to a large network (Asahikawa network)

3.6.1 Network information

The road network in Asahikawa, Japan, shown in Fig. 3.5, is used for the second numerical calculation. The network has 340 directed links, 135 nodes, and 182 O-D pairs. The nodes in red in Fig. 3.5 show centroid nodes from/to which O-D demand is generated/absorbed.

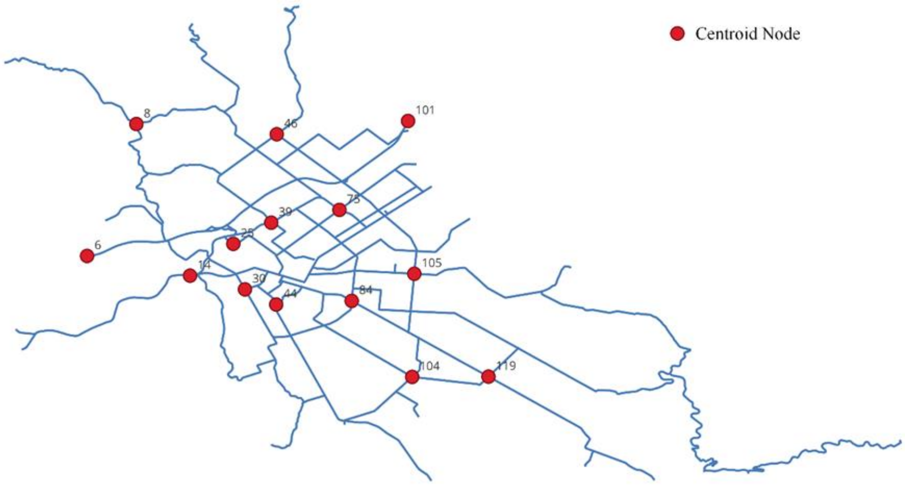


Fig. 3.5 Asahikawa network.

3.6.2 Observation data preparation (ETC2.0)

Electronic Toll Collection 2.0 (ETC2.0) is an advanced technology used for electronic toll collection on highways and toll roads. ETC2.0 builds upon the original ETC system and incorporates additional features and capabilities, such as facilitating the collection of comprehensive road traffic data. By integrating sensors and detectors into the ETC2.0 infrastructure, it can capture real-time traffic information such as vehicle volume, speed, and link travel time. This data can be used for traffic management, congestion monitoring, and transportation planning purposes.

The link delay times collected by ETC2.0 in October 2021 were used in the calculation. The data was collected for 31 days from 9:00 am to 10:00 am from October 1st to 31st including weekend days. Only the links with more than 25 observed link delay times are chosen.

Table 3.3 shows an example of observation data where '-' indicates no observation data is available.

Table 3.3. Observed link delay times for the Asahikawa network.

Link No. / Day	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 3.3 that there are some outliers, such as a delay time of 6808.4 [sec], for which data cleansing is required before the estimation of traffic states.

In this study, an outlier is defined as a value greater than three times the median absolute deviation (MAD). When the number of observations is n, a set of observed values is B that is defined as:

$$\mathrm{B} = \{b_1 \quad \cdots \quad b_n\} \tag{3.62}$$

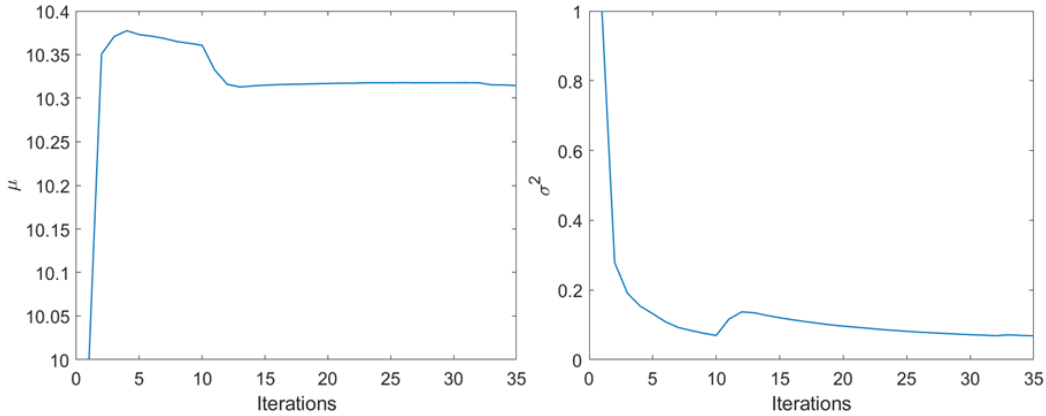
where b_i is the ith observed value and some of them may be $-\infty$ (unobserved). Let B^* denote the subset of B comprised of $b_i > -\infty$ denoted by b_i^*. The MAD is defined as

$$MAD = median(|b_i^* - median(\mathrm{B}^*)| \; \forall b_i^* \in \mathrm{B}^*) \tag{3.63}$$

Outliers can be identified by using the above procedure and incorporated into subsequent numerical calculations after being set to zero.

3.6.3 Results of Asahikawa Network

The estimated μ and σ^2 for the Asahikawa network are 10.31446 and 0.0684 and their convergence process is illustrated in Fig. 3.6.



(a) (b)

Fig. 3.6 Convergence process of μ (a) and σ^2 (b) for the Asahikawa network.

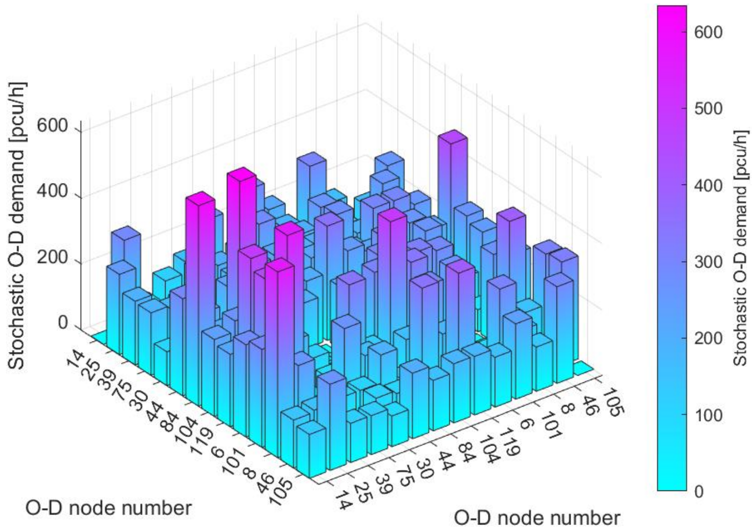


Fig. 3.7 Stochastic demand for each O-D pair for the Asahikawa network.

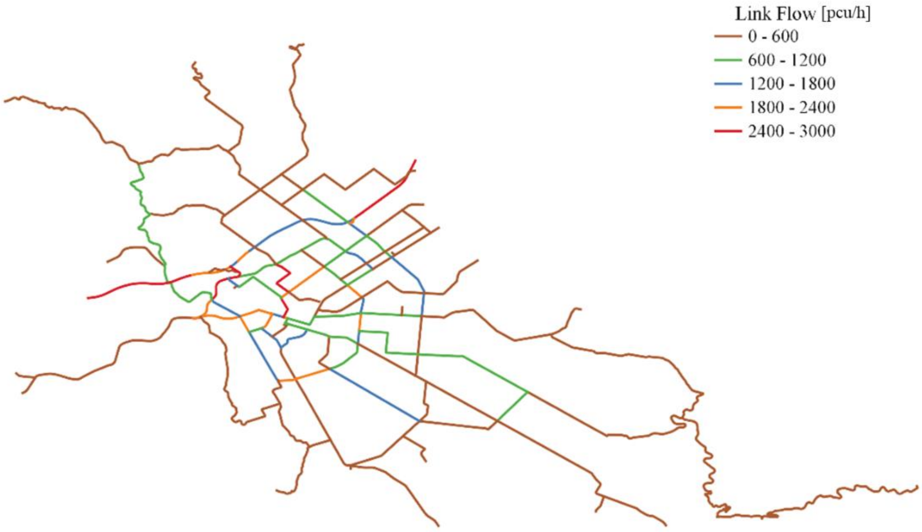


Fig. 3.8 Estimated mean link flow.

Fig. 3.7 shows the estimated mean demand for each O-D pair (there are a total of 182 O-D pairs). The total mean O-D demand estimated is 3.2119×10^4[pcu/h]. The estimated mean O-D demands for O-D pairs 130-88, 130-61, 130-5, 106-61, 58-15 and 84-28 are significantly larger than those for other O-D pairs. The reason for this is that centroid node 84 is located near a primary school, centroid nodes 48 and 28 are located near residential areas, and centroid node 58 is close to a high school. There are major routes from the outskirts to the center of Asahikawa City in the vicinity of centroid nodes 106 and 130.

In this study, a road section in the network is comprised of opposing two directed links. Flows of opposing two directed links are almost the same. Since the number of such links is large, we choose one of two opposing links for a road section in the network. Fig. 3.8 shows the estimated mean link flow in the same direction selected. The mean link flow in the central part is from 600 [pcu/h] to 1,200 [pcu/h], and the links with red color have link flows between 2,400 [pcu/h] and 3,000 [pcu/h]. The reason for this is that the links are located near the centroid nodes.

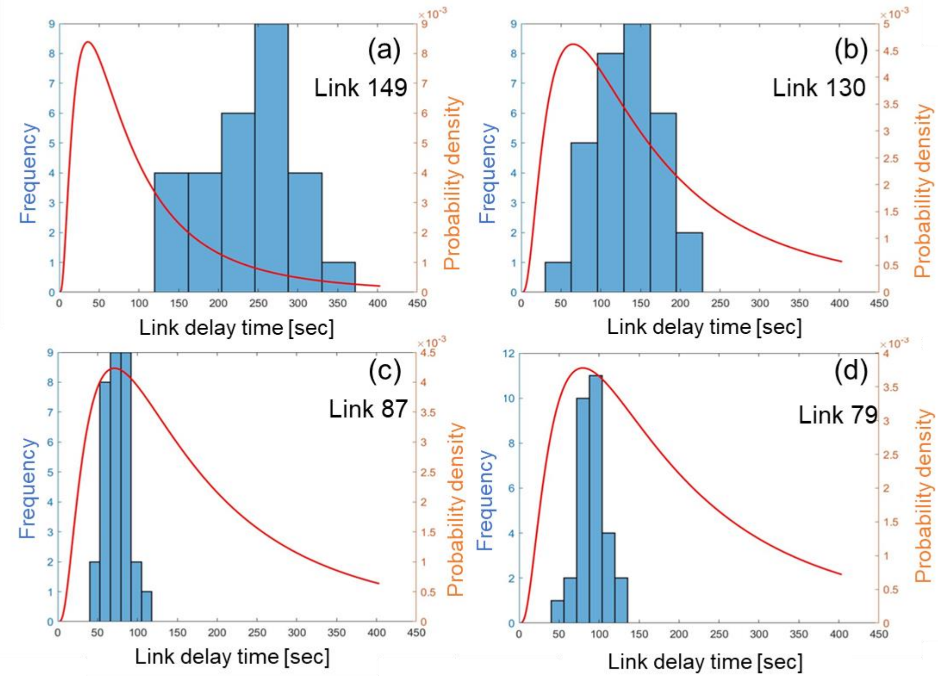


Fig. 3.9 Histogram of link delay times and estimated PDF of log-normal distribution.

Fig. 3.9 shows the relationship between observed link delay times shown by the histogram and the probability density function (PDF) of the estimated lognormal distribution of a link in the network (all results are summarized in Chapter 3.9). The histograms represent observed link delay times. The estimated link delay times are reasonably close to observed link delay times such as subplots (c) and (d). However, a few results are notably different from the observation data such as subplots (a) and (b) due to the locations of centroids predefined before the calculation.

3.7 KL divergence

The Kullback-Leibler (KL) divergence is used for quantifying how much the estimated distribution differs from the true distribution of observed delay time as shown in Fig.3.9.

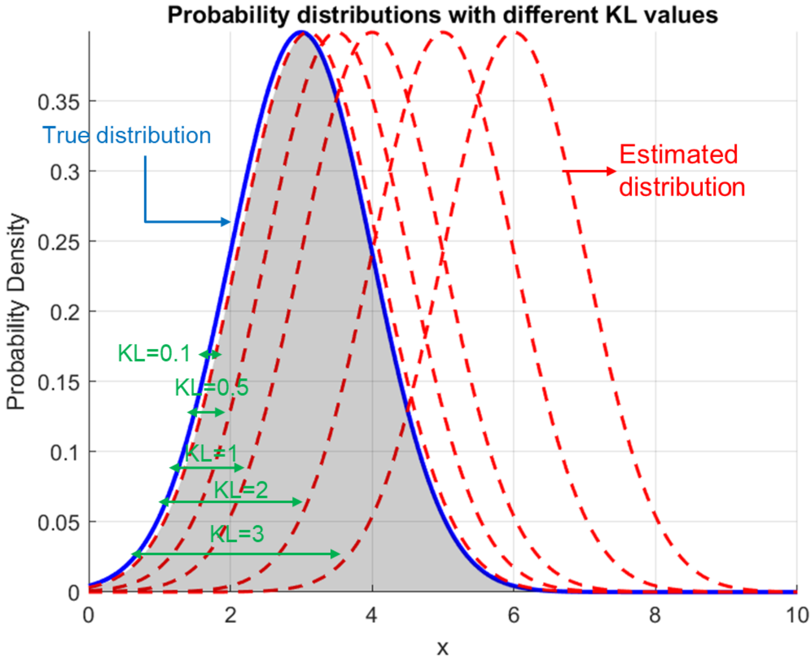


Fig. 3.10 Probability distribution with different KL values.

The blue curve represents the true distribution, while the red dashed curves represent estimated distributions with varying KL values in Fig.3.10. Small KL values (e.g., KL=0.1) indicate a closer match between the estimated and true distributions, while large KL values (e.g., KL=3) show significant deviations.

The KL divergence in this study is calculated as:

$$KL(f(x)\|g(x)) = \int_{ln(-\infty)}^{n(\infty)} f(x) \cdot ln\frac{f(x)}{g(x)}dx \tag{3.64}$$

where $f(x)$ and $g(x)$ are the probability density function of the link delay time estimated by the proposed model, and that obtained by fitting a lognormal distribution to the observed link delay times, respectively.

The link delay times collected by ETC2.0 in October 2021 for the period of 31 days were used in the calculation. Fig. 3.11 shows the corresponding KL divergence results.

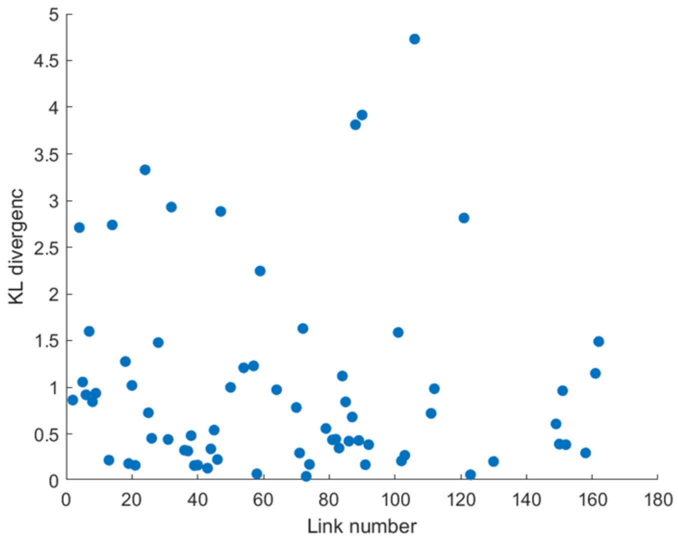


Fig. 3.11 Results of KL divergence.

From Fig.3.11, it is shown that 68% of the KL divergence values fall within the range of 0 and 1, and all of the KL divergence values fall within the range of 0 and 5. The results of the KL divergence analysis show the estimated link delay times by the proposed method are reasonably close to the real observed link delay times in the Asahikawa network.

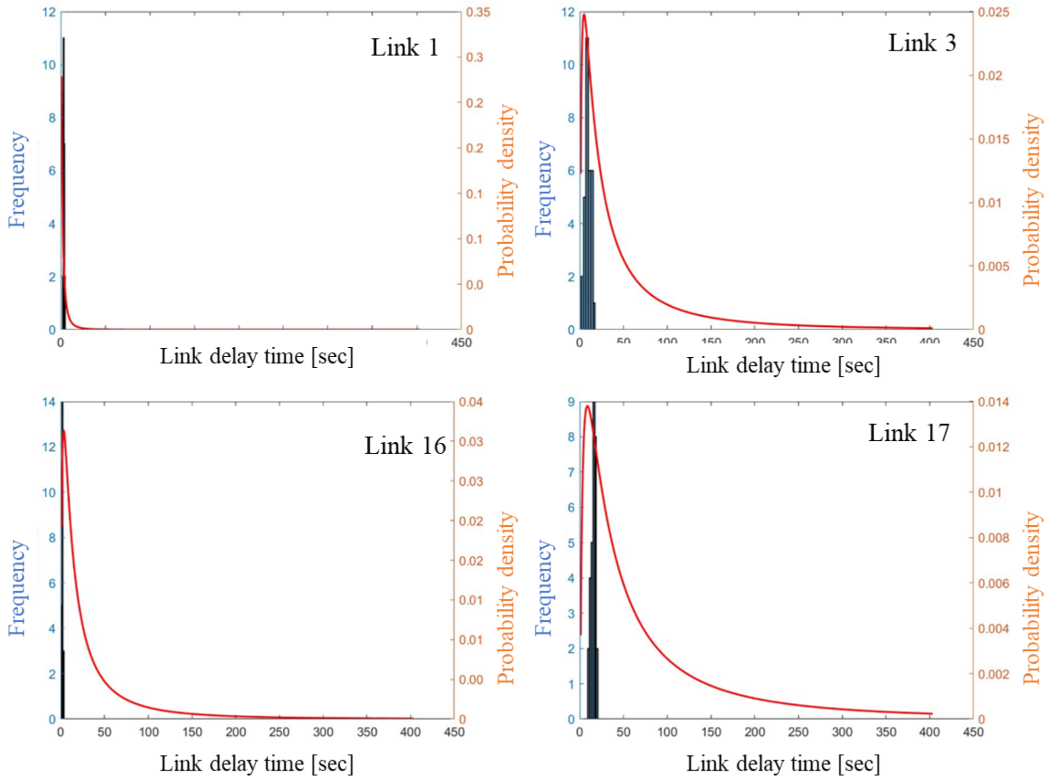
3.8 Conclusions of Chapter 3

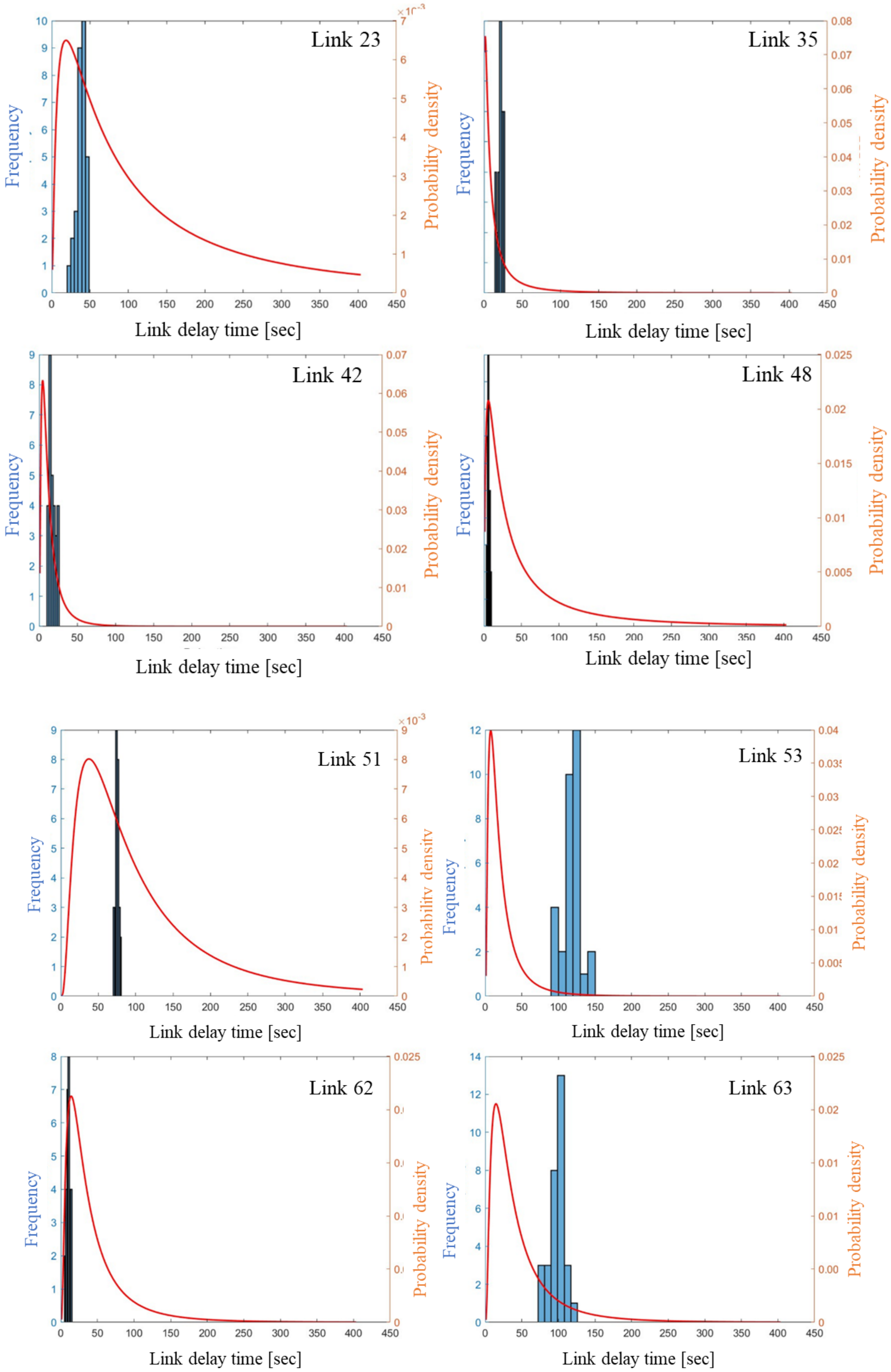
In this chapter, a traffic state estimation problem is formulated as a FIMLE problem under UE constraints. In addition, a solution algorithm, which is based on the sensitivity analysis for UE, for the traffic state estimation problem is developed. Since the market penetration rate of cars from which floating car data can be obtained is still low in Japan, therefore complete observation of network-level data is impossible. The problem formulated in this study can address such incomplete traffic data.

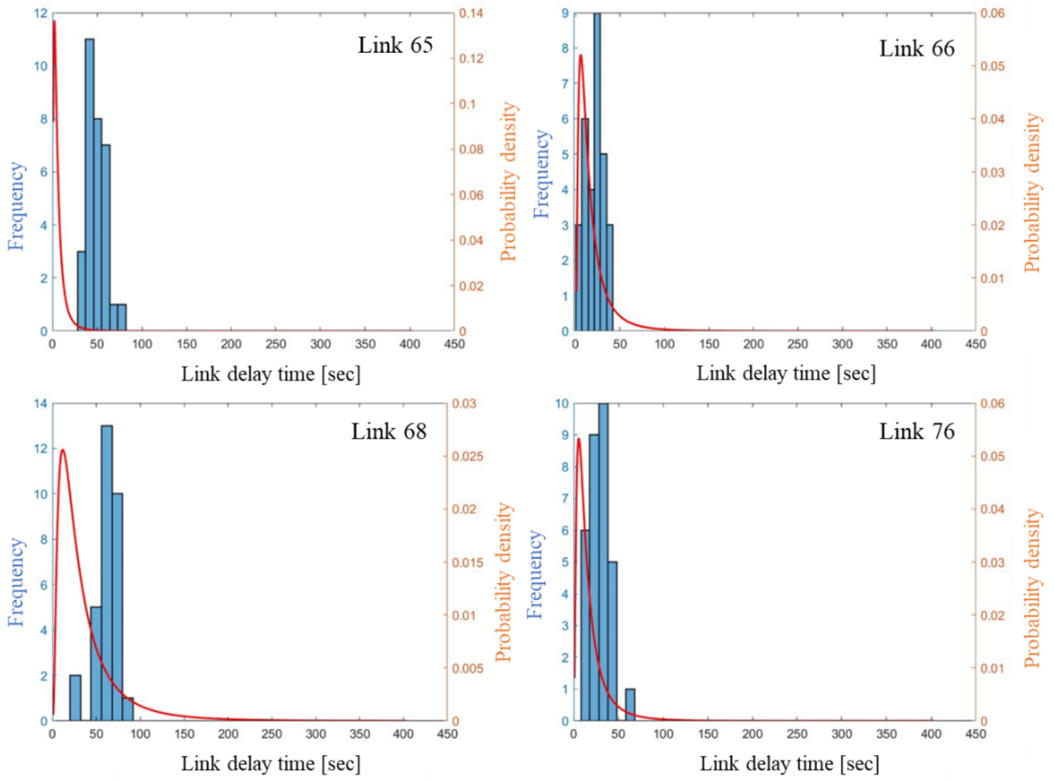
Numerical calculations were carried out in a test network and the real Asahikawa network using real observed link delay times. Especially, the algorithm developed in this study estimated relatively reasonable link delay times in the Asahikawa network. The results of KL divergence show that the proposed method performs well in accurately estimating the link delay times for large networks. However, the accuracy of the estimated link delay times can heavily depend on the locations of the centroids and the number of predefined O-D pairs. We will address such issues in future studies.

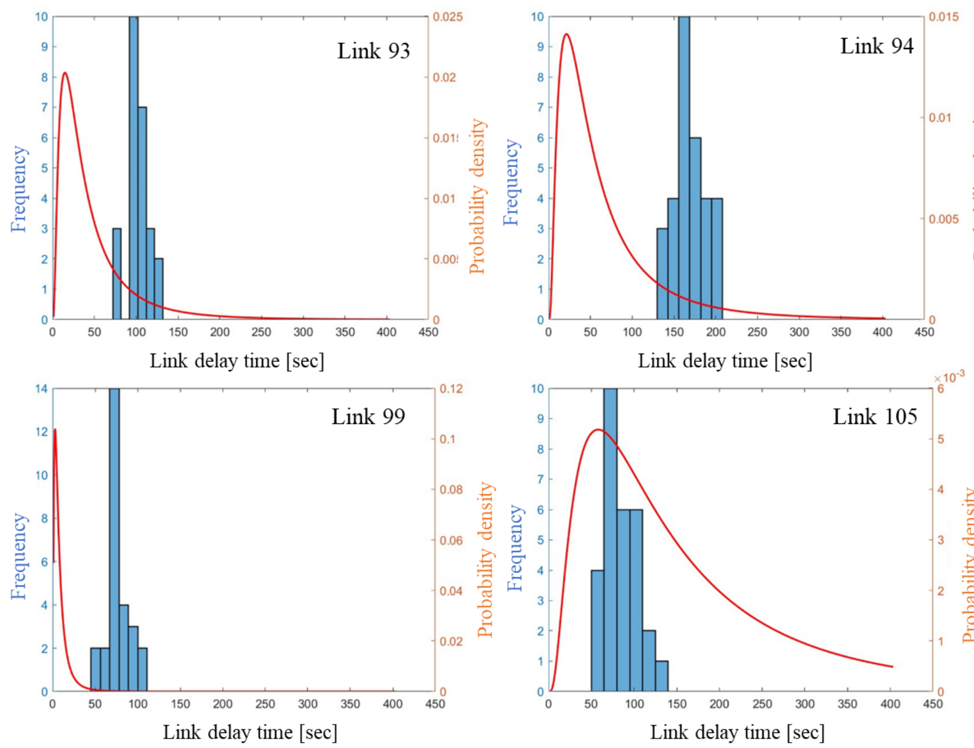
3.9 Appendix 3

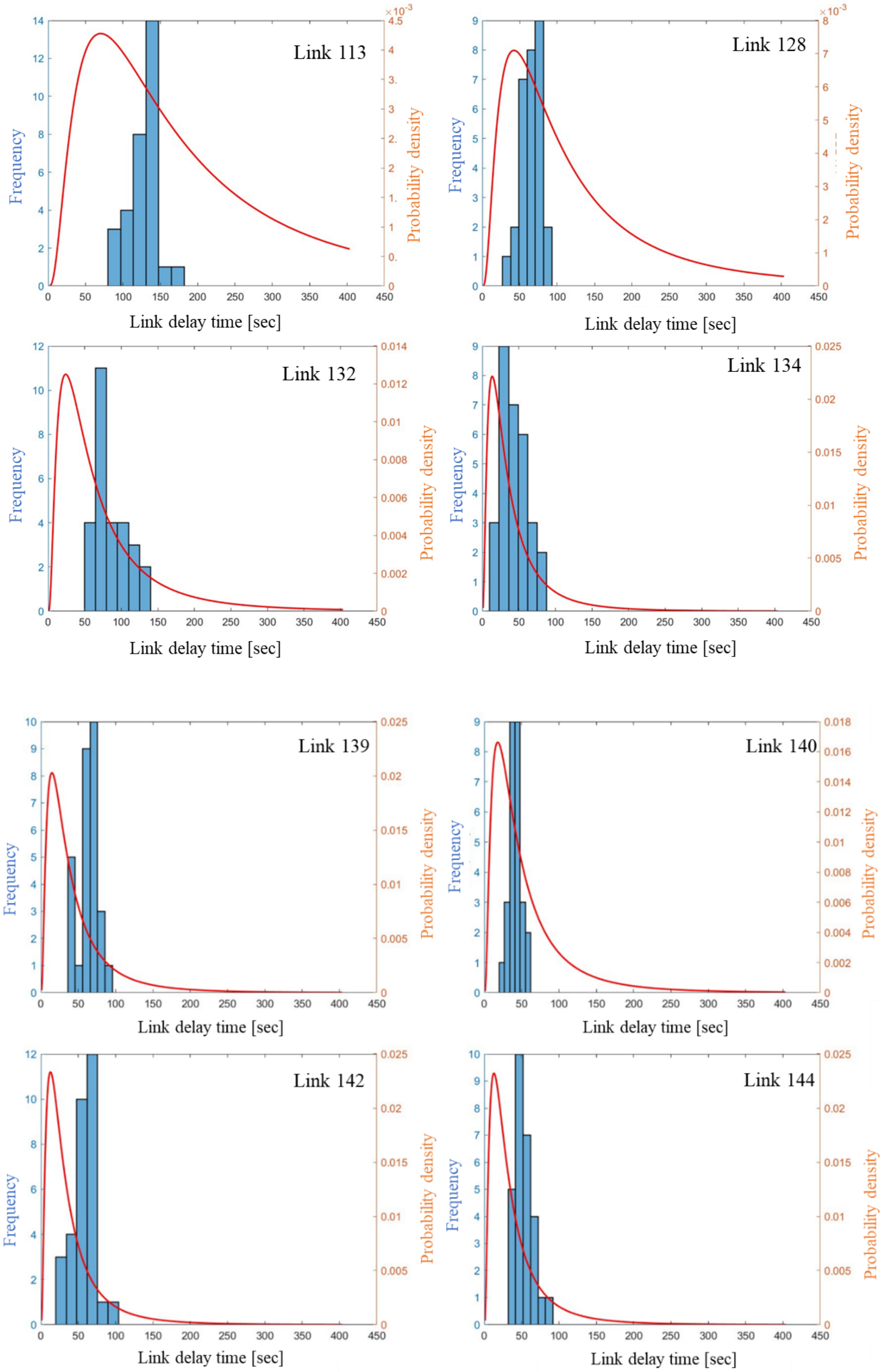
All of the results of the relationship between observed link delay times and PDF of log-normal distribution for each link estimated by the method proposed in this study are summarized here.

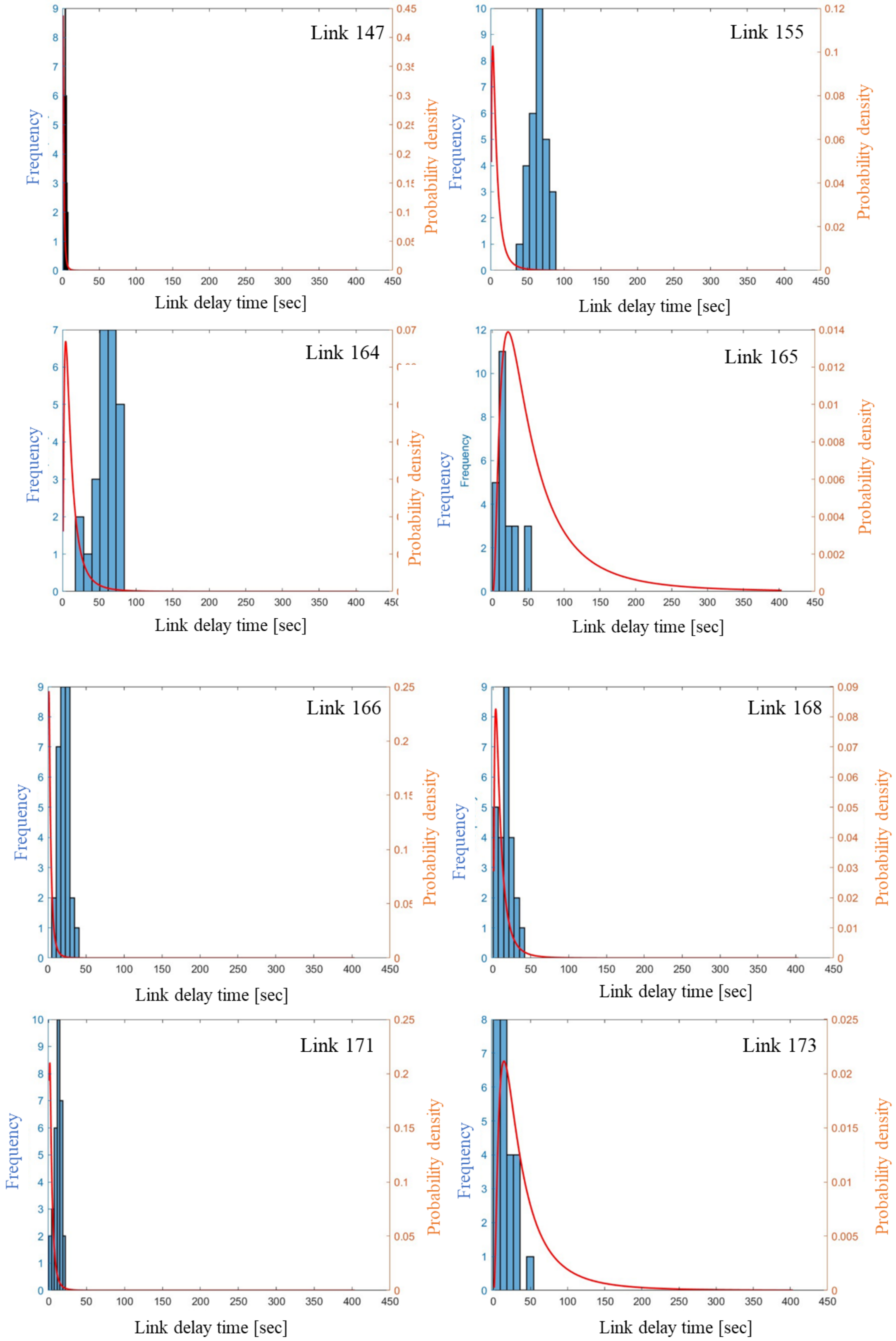


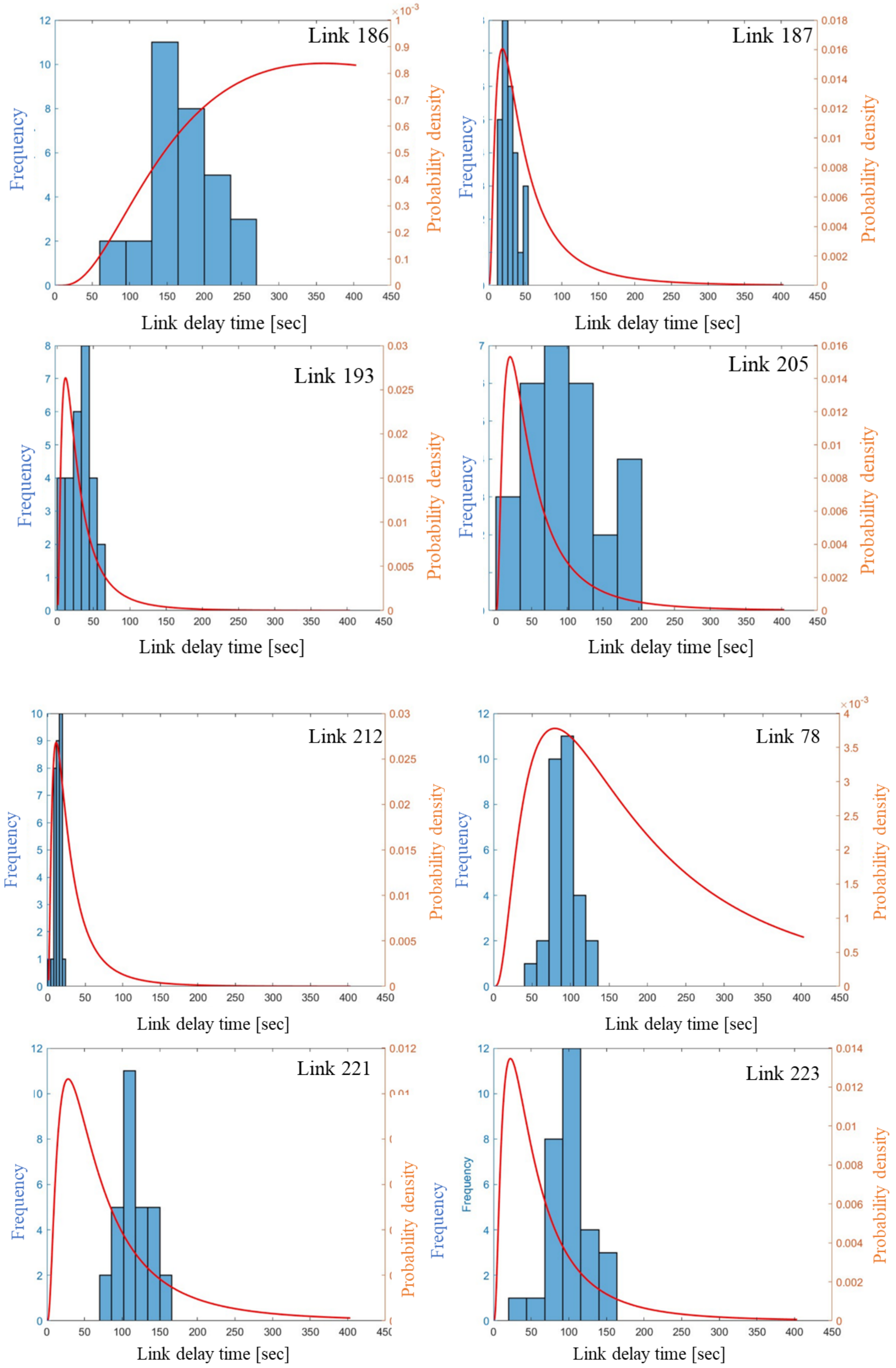

Link 23
Link 35
Link 42
Link 48
Link 51
Link 53
Link 62
Link 63
Frequency
Probability density
Link delay time [sec]

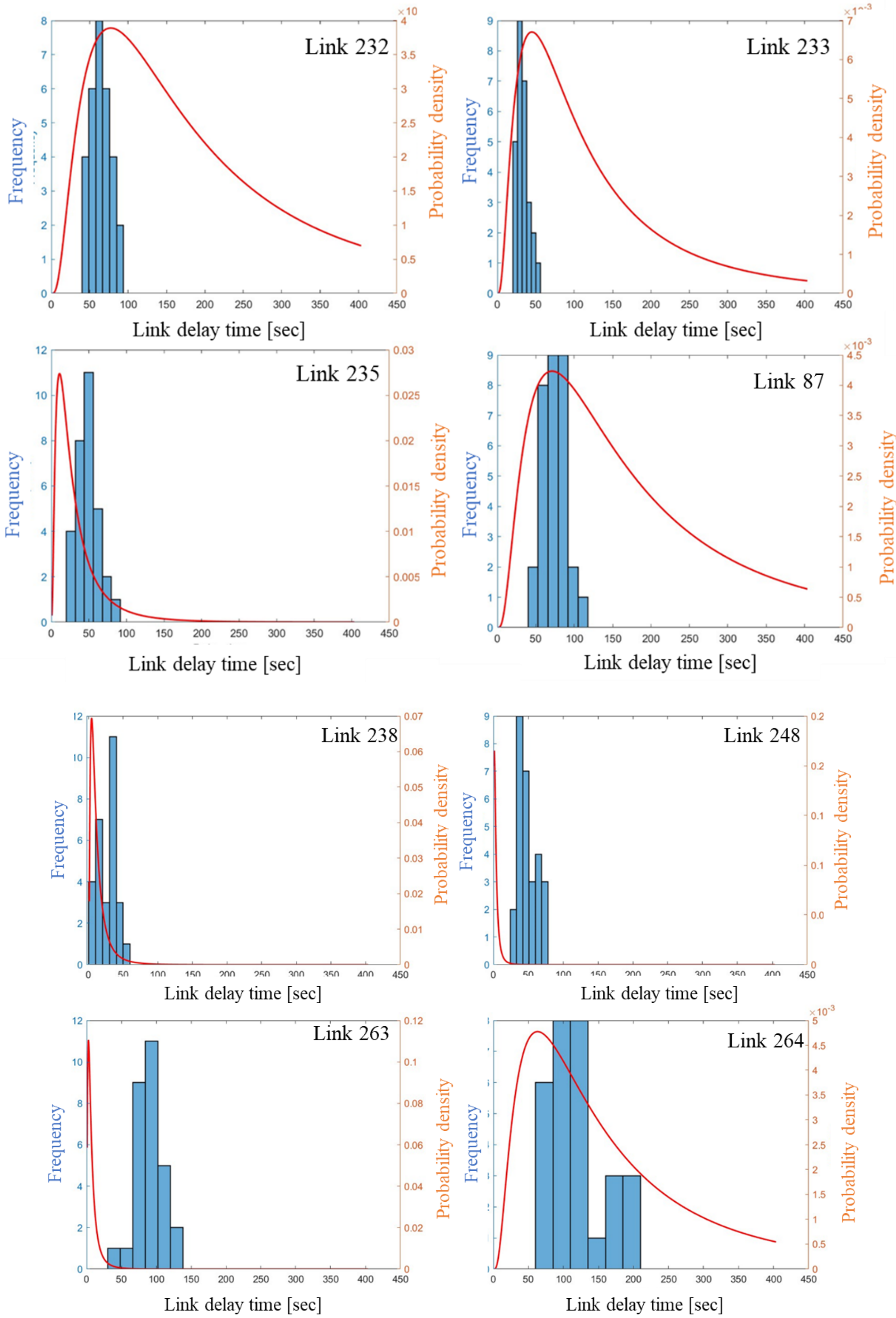

Link 65
Link 66
Link 68
Link 76
Frequency
Probability density
Link delay time [sec]

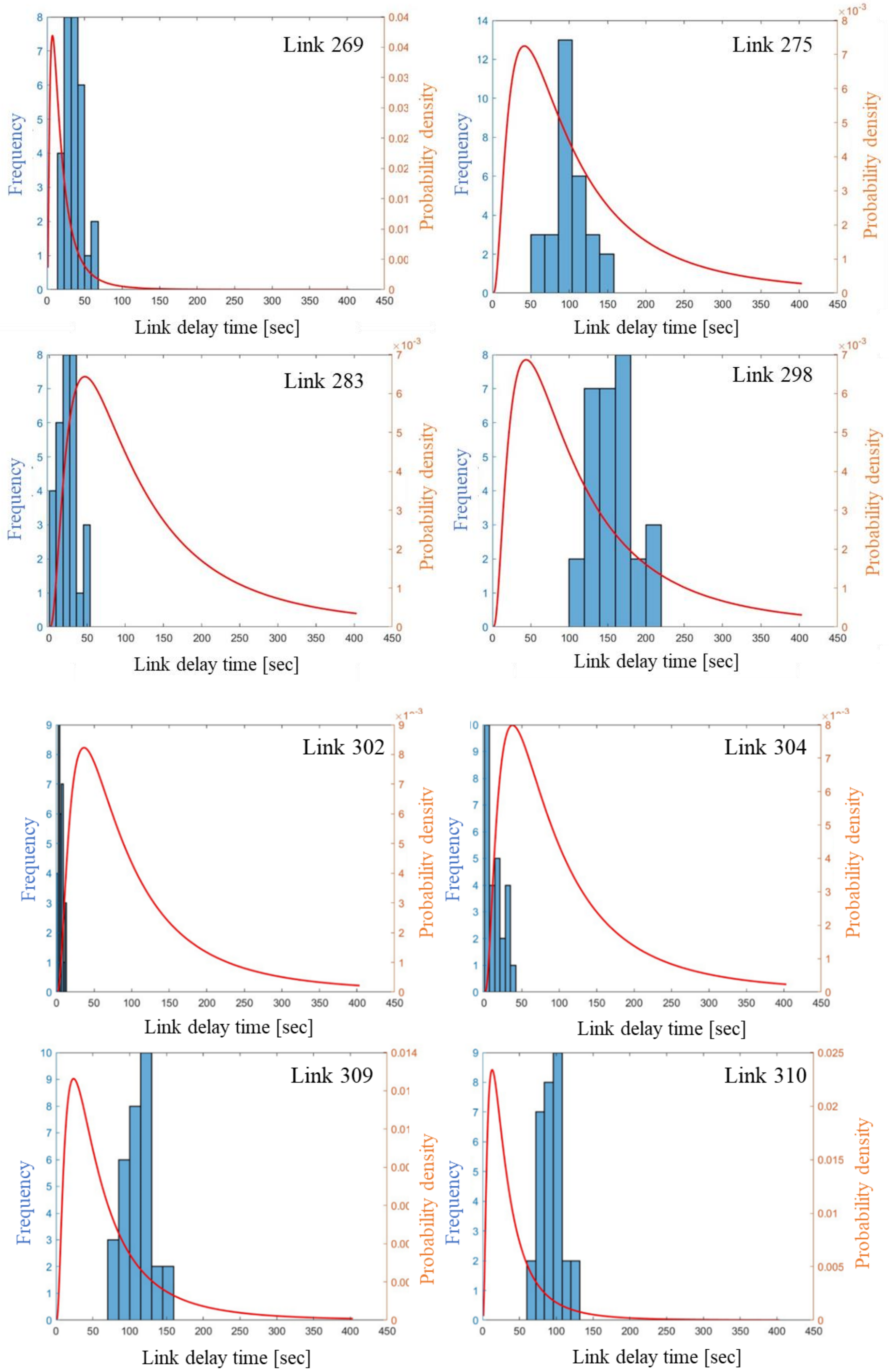

Link 93
Link 94
Link 99
Link 105
Frequency
Probability density
Link delay time [sec]

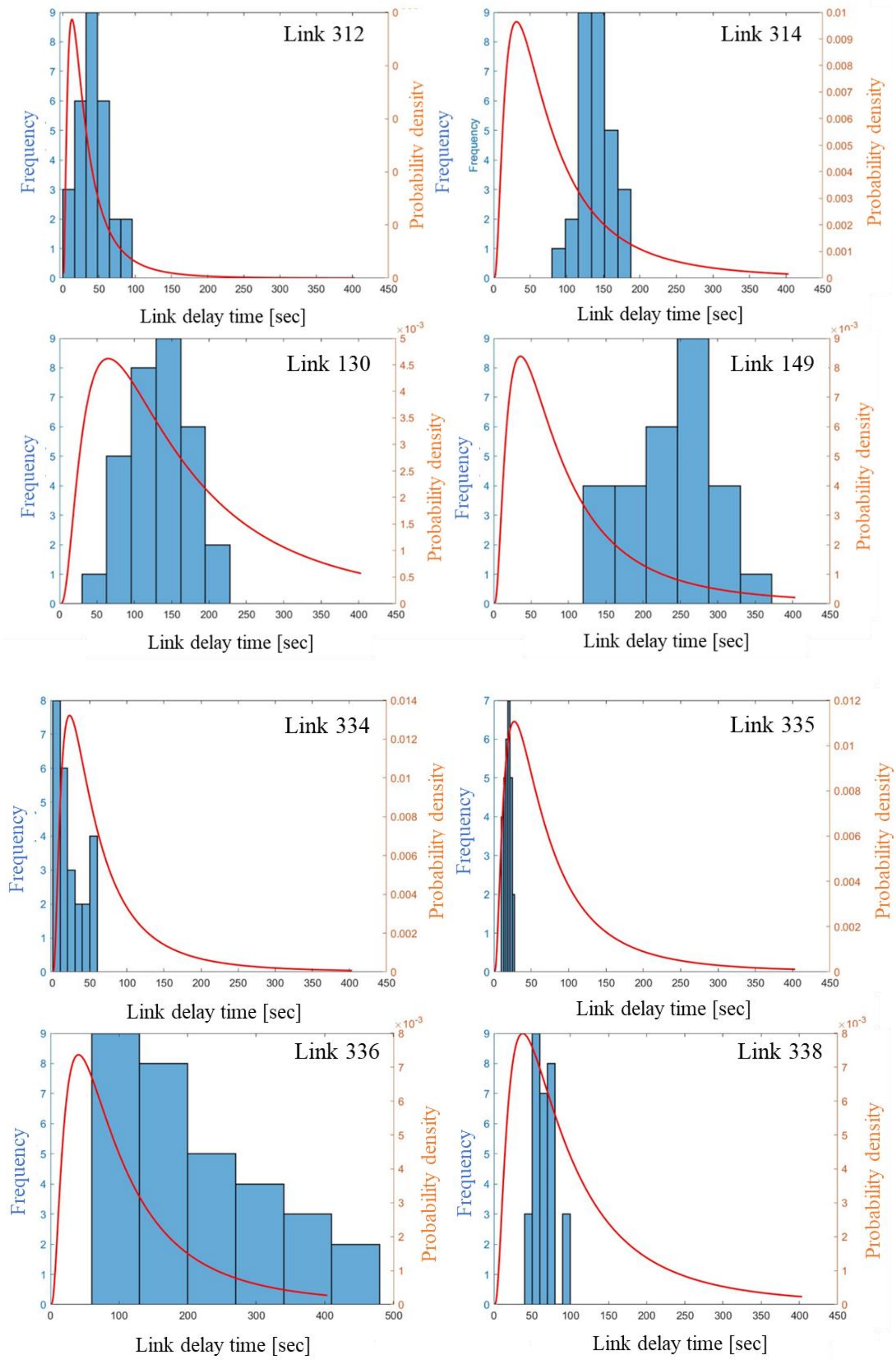

Link 113
Link 128
Link 132
Link 134
Link 139
Link 140
Link 142
Link 144
Frequency
Probability density
Link delay time [sec]


Link 147
Frequency
Probability density
Link delay time [sec]
Link 155
Frequency
Probability density
Link delay time [sec]
Link 164
Frequency
Probability density
Link delay time [sec]
Link 165
Frequency
Probability density
Link delay time [sec]
Link 166
Frequency
Probability density
Link delay time [sec]
Link 168
Frequency
Probability density
Link delay time [sec]
Link 171
Frequency
Probability density
Link delay time [sec]
Link 173
Frequency
Probability density
Link delay time [sec]


Link 186
Link 187
Link 193
Link 205
Link 212
Link 78
Link 221
Link 223
Frequency
Probability density
Link delay time [sec]
×10^-3


Link 232
Link 233
Link 235
Link 87
Link 238
Link 248
Link 263
Link 264
Frequency
Probability density
Link delay time [sec]


Link 269
Link 275
Link 283
Link 298
Link 302
Link 304
Link 309
Link 310
Frequency
Probability density
Link delay time [sec]


Link 312
Link 314
Link 130
Link 149
Link 334
Link 335
Link 336
Link 338
Frequency
Probability density
Link delay time [sec]

4. Development of sensitivity analysis-based algorithm for a bus network design problem

4.1 Introduction

With the advancement of CABs technology and the increasing urban transportation demand, designing efficient transit networks and optimizing service frequencies have become particularly crucial. However, research addressing both optimal bus route network design and service frequencies is lacking. This paper proposes a bi-level programming model (BIPM) to simultaneously optimize bus route network and service frequencies, considering two travel modes: CABs and Human Driving Vehicles (HDVs). The upper-level problem aims to minimize the total network cost, including travel and operational costs. This problem is known as a discrete optimization problem. A continuous relaxation approach is employed for this discrete problem, allowing the application of a solution algorithm for nonlinear optimization to solve the relaxed problem. Furthermore, this optimization problem must be solved under an equilibrium constraint, formulated as a UE traffic assignment problem at the lower-level. To solve the bi-level problem, a sensitivity analysis-based algorithm is developed. Numerical experiments are used to validate the efficacy of the proposed method. The results demonstrate the capability of the model and algorithm to concurrently optimize both bus route networks and service frequencies, providing a valuable reference for future bus network design.

CABs are poised to play a critical role in the future of urban mobility, offering significant potential for public transportation systems globally. Additionally, CABs offer multiple advantages over Human Driving Buses (HDBs), which are increasingly recognized in the field of transportation. Compared to HDBs, computer-controlled CABs can significantly reduce operational costs by eliminating the need for drivers. Furthermore, CABs can enhance operational efficiency by gathering information about the surrounding vehicles and environment. Moreover, when humans drive vehicles, variations in their perception of danger lead to different reaction times and, consequently, different headways, thereby affecting the efficiency of road capacity usage. In contrast, CABs exhibit significantly shorter reaction times and shorter headways, thereby optimizing road capacity utilization.

Although CABs present considerable advantages over HDVs, fully realizing the potential of CABs to optimize the efficiency of transportation networks necessitates targeted optimizations that consider the interdependence among various components of the network. Factors influencing the effectiveness of CABs within a transportation network include: (a) route design, (b) frequency setting, (c) timetabling, (d) vehicle scheduling, and (e) crew scheduling and rostering. Among these, route design and frequency setting are especially critical. Route design refers to the process of determining the specific paths or lines that CABs will follow within a network. This process directly affects the accessibility and coverage of the transit system, which in turn influences the efficiency of the network by determining how well the service meets demand and how effectively it reduces travel times and congestion. Similarly, frequency setting involves determining how often CABs will serve each route. This factor directly impacts the level of service, influencing passenger waiting times, vehicle load factors, and the overall reliability and attractiveness of the transit system, thereby affecting the overall operational efficiency of the network. More importantly, the interaction between route design and frequency setting significantly influences the overall performance of the network. Therefore, optimizing only one of these factors is usually insufficient to fully harness the potential of CABs. By jointly optimizing both route design and frequency setting, considering their interaction, it is possible to significantly improve the operational efficiency of transportation networks employing CABs. This, in turn, will enable CABs to better fulfill their potential and contribute to the enhancement of the transportation networks in which they are deployed.

1) Optimization research on bus networks involving HDBs and CABs

Current research directions can be broadly categorized into bus route design, frequency setting, and timetabling. In the literature, Chen et al. (2018) optimized the bus network by integrating the local route service and the short-turn approach respectively. Pternea et al. (2015) tackled a sustainability-focused variant of the bus route network design problem, incorporating environmental costs. Jiang (2022) addressed the concept of direct traveler density and extended it to the bus network design problem by incorporating demand density, which relates to both direct demand and transfers, as well as route lengths. Zhang et al. (2019) proposed a method for the bus network design problem, which was to build the CABs dedicated line to improve the overall system performance.

Conversely, Ceder (1984) examined and analyzed various data collection methods for bus operators to efficiently determine bus frequencies and headways. Afanasiev and Liberman (1983) designed a network of urban bus rapid transit services in multiple medium-sized cities to address commuter demand. With the development of autonomous driving technology. Sadrani et al. (2022) optimized service frequencies and vehicle size for autonomous bus systems, considering both user and operator costs. Tian et al. (2022) discuss that the “one-off” replacement of all service lines with autonomous buses is not available because of the low acceptance of autonomous buses. Hatzenbühler et al. (2022) examined how network design and frequency setting might change when Autonomous Vehicle (AV) systems are implemented on fixed-route networks, compared to conventional public transportation systems.

Furthermore, some researchers aimed to solve the problem of timetabling. Gkiotsalitis and Alesiani (2019) improved the bus timetable incorporating travel time and passenger demand. Chu et al. (2019)

simultaneously optimized the bus timetables and passenger travel path choices.

This is even though there are many current research directions and each of them is well established. For the study of optimal bus route network design and service frequencies, however, most of the research mentioned above primarily focused on route design and optimal frequency setting as separate issues. However, integrating bus route network design with frequency setting ensures that routes are strategically designed to cover key demand areas and are complemented by frequencies that align with the varying demand levels across the network.

2) Mathematical models for bus network design problems

Transit network designs have traditionally functioned as a decision support tool for policymakers, enabling them to evaluate the potential benefits and effects of various transportation projects. Up to now, some researchers and policymakers have solved transit network design problems using a single-level model. Ceder (1984) derived bus frequency using a single-level model. Ceder and Wilson (1986) proposed an algorithm for designing new bus routes that consider both passenger needs and operator interests. Bi-Level Programming Model (BIPM), in contrast, offers a more sophisticated approach by explicitly modeling these interactions, allowing for a more comprehensive understanding of transit network dynamics. Many researchers have recognized the potential of BIPM in addressing the complexities of transit network design. For instance, Jiang (2022) developed a multi-objective BIPM in which the lower-level model focuses on reliability-based transit assignment, while the upper-level model aims to determine optimal bus frequencies to simultaneously minimize the total travel cost of the entire network and maximize network fairness. Similarly, Uchida et al. (2005, 2007) and Gao et al. (2004) utilized BIPM to optimize service frequencies, the total cost of the entire network including travel cost and operation cost in the objective function of the upper-level model and the equilibrium constraint is used in the lower-level model. Yang and Bell (1997) developed a BIPM to determine the optimal toll charges.

In contrast to the BIPMs discussed earlier, which focused on optimizing the frequencies or bus routes separately, this paper addresses an upper-level model that emphasizes reducing the total travel cost of the entire network. Since the total travel cost is influenced by both bus routes and frequency settings, these two aspects are considered simultaneously. The problem addressed in this paper focuses on the optimization of both bus routes and service frequencies and this is known as a mixed integer nonlinear problem. The problem is then relaxed into continuous form and a nonlinear optimization algorithm is applied to solve this relaxed problem.

3) Optimal routes and frequencies problem

Addressing the optimal routes and frequencies problem for buses is crucial for transportation systems. By optimizing bus routes and service frequencies, cities can improve accessibility, reduce operational costs, and better match supply with passenger demand. For instance, Ruano-Daza et al. (2018) provide an effective way to handle the complexity of bus rapid transit systems design, including route selection and frequency optimization, under various constraints. Szeto and Wu (2011) adopt a bi-level optimization framework and highlight the significance of balancing operational performance with passenger convenience, particularly in suburban contexts where transit demands can vary largely. Moghaddam et al. (2019) present a robust approach utilizing a weighted multi-objective genetic algorithm (GA) to address the problem of simultaneous optimization of bus route network design and frequency setting. While earlier papers considered UE within traditional transit contexts, this study emphasizes how service frequency adjustments in a multimodal network (including HDVs and CABs) influence both modal and route choices. The model proposed in this study captures more realistic travel behavior under evolving technological and operational conditions. One of the innovations of this study is the integration of CABs into bus network design and the dynamic treatment of link capacities in a mixed flow of CABs and HDVs.

4.2 Objectives

Travelers' behavior in a multimodal network, including modal choice and route choice, is influenced by the service frequencies assigned to each bus route within the bus network. Changes in service frequencies can directly affect the attractiveness and utility of different routes, thereby influencing overall traffic patterns. This study introduces a comprehensive optimization problem that simultaneously addresses: 1) the bus route network design problem and 2) determining optimal frequencies.

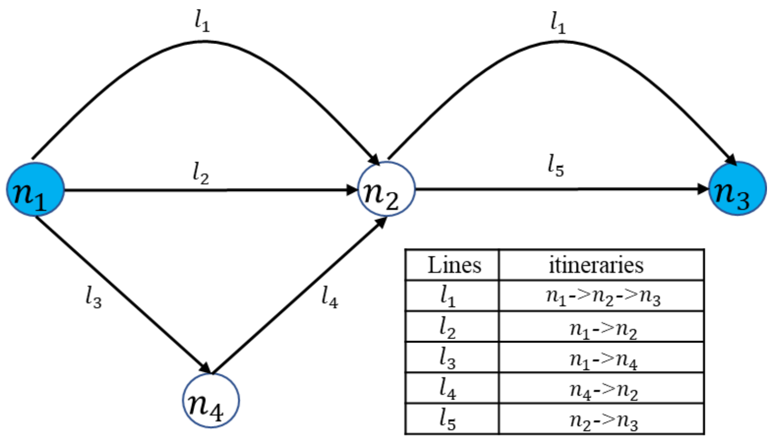
The main contributions of this study are as follows:

1. A BIPM is formulated for the optimization of bus route networks and service frequencies. CABs are introduced to BIPM as a public transportation system. The BIPM offers advantages in addressing the bus network design problem. The upper-level problem typically focuses on optimizing network design, which includes route design and service frequencies. While the lower-level problem captures the user behavior and traffic flow dynamics, modeled through a UE traffic assignment.
2. This paper considers mixed flows of CABs and HDVs, therefore traffic capacity of each link in a road network is represented by a function of the penetration rate of CABs on the link.
3. The upper-level model is known as a discrete optimization problem, which is formulated as a

continuous form in the BIPM. A nonlinear optimization algorithm is applied to solve the BIPM. This optimization problem needs to be solved subject to equilibrium constraints, represented as a UE traffic assignment problem at the lower-level problem. A sensitivity-based algorithm is developed for solving BIPM.

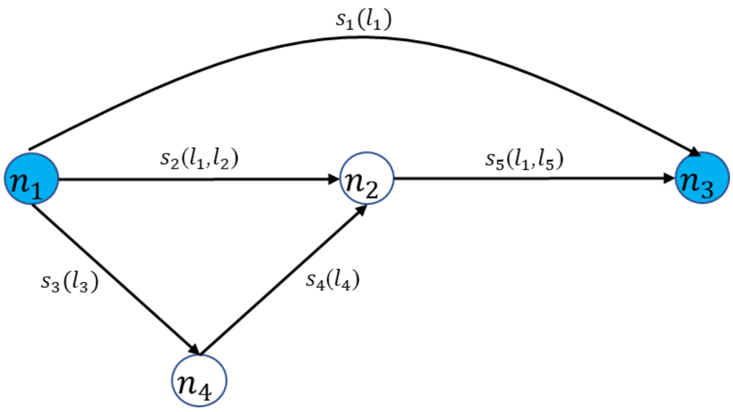
4.3 Network representation

Consider a general bus network as shown in Fig. 4.1(a), which shows a simple example of the primitive network taken from De Cea and Fernández (1993). There are five lines in this network (l_1-l_5). A line (or transit line) is a group of buses or trains that runs back and forth between two transit stops on a transit network. The primitive network is transformed into a modified network shown in Fig. 4.1(b), i.e., $G(N, S)$ with node set N representing bus stops and link set S representing route sections. The route section is a portion of a route between two consecutive transfer nodes. In Fig. 4.1(a), there are six paths between one O-D pair. There are three combined paths between the same O-D pair in Fig. 4.1(b): s_1, s_2-s_5 and s_3-s_4-s_5.



Lines	itineraries
l_1	n_1->n_2->n_3
l_2	n_1->n_2
l_3	n_1->n_4
l_4	n_4->n_2
l_5	n_2->n_3

(a) Primitive network and its itinerary



(b) Bus network represented by route section

Fig. 4.1 Network representation for the bus network

The bus lines selected by bus passengers on route section s are called "attractive lines" for them instead of all lines on route section s. It is necessary to identify the set of attractive lines for bus passengers in bus planning. The set of attractive lines on route section s, i.e., $\psi(s)$, can be determined by solving the following problem (De Cea and Fernández, 1993):

$$\min_{\{x_s^l\}} Z_s = \frac{\alpha + \sum_{l\in\psi(s)} \hat{t}_s^l \cdot f_l \cdot x_s^l}{f^s} \tag{4.1}$$

s.t.

$$x_s^l = \{0,1\}\ \forall l \in \psi(s), f^s = \sum_{l\in\psi(s)} f_l \cdot x_s^l \tag{4.2}$$

where, $\psi(s)$, f_l and $\hat{t}_s^l$ denote a set of all bus lines passing through link, the frequency of bus line l and the constant in-vehicle travel time of route section s using bus line l, respectively. The objective function defined by Equation (4.1) expresses the expected total travel time.

The multi-modal network in this study is expressed as a hyper-network, i.e., the road network including the walking network plus the modified network representing the bus network as described above.

4.4 Disutility of link

In this paper, we assume that travelers evaluate the inconvenience or disutility of traveling on a link based on four main factors: travel time, in-vehicle congestion effects, expected waiting time, and fares. As we will discuss in the next chapter, the in-vehicle congestion effect and travel time can be merged into a single measure called perceived travel time. Thus, in general, the total disutility of using link s can be defined as:

$$d_s = \pi_t \cdot \hat{t}_s + \rho_w \cdot w_s + \tau_p \cdot p_s \tag{4.3}$$

where π_t, ρ_w and τ_p denote coefficients on the perceived travel time ($\hat{t}_s$), waiting time (w_s) and fare (p_s), respectively.

Then, the utility function of a route can be defined as the sum of the associated link utilities, plus mode-specific constant(s).

$$\pi_{i,j} = \sum_{m \in \{W,H,B\}} \lambda^m \cdot \xi_{i,j,m} + \sum_{s \in S} d_s \cdot \delta_{i,j,s} \tag{4.4}$$

where λ^m denotes the MSC for mode m. For the jth intermodal route between O-D pair, $\xi_{i,j,m} = 1$ if the route includes at least one link with transport mode is m and $\delta_{i,j,s} = 1$ if it is associated with link s. Otherwise, $\xi_{i,j,m}$ and $\delta_{i,j,s}$ equal zero. The monetary costs of HDVs derive mainly from car ownership expenses, fuel costs, and tolls. The cost associated with vehicle ownership is represented in the MSC for HDVs.

4.4.1 Notations

The notation adopted in this paper is shown below:

Notation	Representation
S	Set of links comprising the multi-modal network
$M = \{W, H, B\}$	Set of transport modes where W, H and B represent walking, passenger car and bus, respectively
S_m	Set of links for transport mode $m \in M$, i.e., $S = \bigcup_{m \in M} S_m$
I	Set of origin-destination (O-D) pairs
J_i	Set of paths serving between O-D pair $i \in I$
L_i	Set of bus lines serving between O-D pair $i \in I$
L	Set of bus lines in the multi-modal network, i.e., $L = \bigcup_{i \in I} L_i$
$f_{i,l}$ (f_l)	Service frequency of bus line $l \in L_i$ serving between O-D pair $i \in I$ (Service frequency of bus line $l \in L$)

$\psi(s)$	Set of bus lines that travel on link $s \in S_H$
χ_s	Traffic capacity of link $s \in S_H$
v_s	Flow on link $s \in S$
v_s^l	Flow on link $s \in S$ using line $l \in L$
σ	Free-flow speed
w	Backwards wave speed
π_t	Coefficient for travel time
ρ_w	Coefficient for waiting time
τ_p	1/(value of time) which is used to covert the money to time
$\mathcal{i}$	Cell $\mathcal{i} \in \mathcal{I}$
$\mathcal{I}$	Set of cells
$\mathcal{N}$	Maximum number of vehicles that can fit in cell $\mathcal{i} \in \mathcal{I}$
$\mathcal{Q}_i(t)$	Maximum flow in cell $\mathcal{i} \in \mathcal{I}$ at time t
$k_m(x,t)$	The density of vehicles of class $m = \{\text{HDVs}, \text{CABs}\}$ at space-time point (x,t)
k^{jam}	Jam density
E_b	Passenger car equivalent for bus
α_s, β_s	Calibration parameters of link s
t_s	Free flow travel time on the link s
ϑ_m	Vehicle length of transport mode $m \in \{H, B\}$
θ	Coefficient converting fare caused by frequency setting
$\mathbf{v}$	Vector of link flows
$\mathbf{v}^*$	Vector of equilibrium link flows
$\mathbf{f}$	Vector of frequencies
$\mathbf{h}$	Vector of path flows
$\mathbf{q}$	Vector of O-D demands
$\Omega_{\mathbf{v}}$	A feasible set of link flows
$\Omega_{\mathbf{h}}$	A feasible set of path flows
$\mathbf{\Delta}, \mathbf{\Lambda}$	Link/path, O-D pair/path incidence matrix

4.4.2 Link capacity

Autonomous driving technology is rapidly advancing, with testing on public roads becoming increasingly common. As autonomous vehicles become available to the public, the precision of their onboard computers and advanced communication systems may enable new traffic management strategies that could significantly improve network capacity. For example, Dresner and Stone (2004) introduced the Tile-Based Reservation (TBR) intersection policy, which has been shown to reduce delays beyond what can be achieved with optimized traffic signal systems (Fajardo et al., 2011). This policy exemplifies how autonomous vehicles can create more efficient intersection behaviors. Additionally, autonomous vehicles have the potential to increase link capacity by reducing reaction times, which in turn allows for shorter following distances. Furthermore, autonomous vehicles are likely to be less susceptible to certain adverse road conditions compared to HDVs, thereby improving road performance under challenging conditions.

However, these capacity improvements are not straightforward because, for many years to come, autonomous vehicles will have to share the road with HDVs due to the slow adoption process and limited affordability of autonomous technology for all travelers. This mixed-flow environment adds complexity to understanding how capacity improvements can be achieved. Importantly, the TBR policy is designed to function effectively even on shared roads, as demonstrated by Dresner and Stone (2007). Similarly, capacity improvements at the link level can also be safely implemented in a mixed-traffic setting. Nevertheless, developing robust models that accurately capture these link and intersection capacity improvements under shared-road conditions remains an open research challenge.

At present, most existing models that simulate autonomous vehicle behavior rely on micro-simulation techniques. While these models provide detailed insights, they are computationally expensive and thus impractical for large-scale traffic assignment problems, which are commonly used to model route choice and network-wide traffic flows. Recognizing this limitation, Levin and Boyles (2015a) adopted static link performance functions to predict capacity improvements as a function of the proportion of autonomous vehicles on a given link. Their approach was based on Greenshields' (1935) capacity model, which provides a fundamental relationship between traffic flow, density, and speed. However, in reality, the proportion of AVs on each link varies over time due to changing traffic patterns and vehicle distributions.

Dynamic traffic assignment (DTA) models offer a more accurate representation of traffic flow and can account for time-varying changes in link capacities. Unlike static models, DTA allows for the integration of temporal effects, which are particularly important when modeling capacity changes, influenced by varying proportions of AVs. For instance, Kesting et al. (2010) predicted theoretical link capacities for adaptive cruise control systems and extrapolated these findings using linear regression to estimate the effects of different proportions of connected vehicles (CVs) and non-connected vehicles (non-CVs).

Levin and Boyles (2016) developed a multiclass cell transmission model (CTM) that admits variations in capacity and backward wave speed in response to class proportions within each cell. The CTM model is shown to be consistent with hydrodynamic theory. And then they developed a car following model incorporating driver reaction time to predict capacity and backward wave speed for mixed flow conditions.

If there is mixed flow on the link, the capacity is not fixed. Levin and Boyles (2016) devised a method to calculate the link capacity under the mixed flow of two different types of vehicles with the same vehicle length. Firstly, the multiclass extension of CTM is introduced. Here are some assumptions based on Levin and Boyles (2016).

1. All vehicles are assumed to travel at the same speed. While actual vehicle speeds vary in real-world scenarios, DTA models commonly adopt the assumption that all vehicles exhibit identical speed behavior. This simplification is considered reasonable even when multiple vehicle classes are present, as CABs may adjust their speeds to match surrounding traffic, potentially even exceeding the speed limit if necessary (Miller, 2014). Although Tuerprasert and Aswakul (2010) incorporated differing vehicle speeds in their CTM, the focus in this chapter on HDVs and CABs assumes that most speed variations arise from differences in HDVs behavior, which are typically not accounted for in standard DTA models
2. The density of each vehicle class within a cell is assumed to be uniformly distributed. In single-class CTM, it is generally assumed that the density within a cell is uniform. We extend this assumption to include class-specific densities for multiple vehicle classes.
3. An arbitrary number of vehicle classes is considered. Different types of CABs may be certified with varying reaction times, leading to differences in their car-following behavior.
4. The backward wave speed is assumed to be less than or equal to the free-flow speed. This assumption is essential for determining the cell length based on the free flow speed.

Let $\mathcal{k}_m(x, t)$ be the density of vehicles of class m at space-time point (x, t) with total density denoted by:

$$\mathcal{k}(x, t) = \sum_{m \in M} \mathcal{k}_m(x, t) \tag{4.5}$$

Similarly, let

$$q_m(x,t) = u\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right) \cdot k_m(x,t) \tag{4.6}$$

be the class-specific flow, with the total flow given by

$$q(x,t) = \sum_{m \in M} \chi_m(x,t) \tag{4.7}$$

and let the function $u\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right)$ denote the possible speed with class proportions of $\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}$.

Speed is limited by free-flow speed, capacity, and backward wave propagation:

$$u(k_1, \dots, k_{|M|}) = \min\left\{\sigma, \frac{q^{\max}\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right)}{k}, w\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right) \cdot \frac{k^{\text{jam}} - k}{k}\right\} \tag{4.8}$$

where σ is free-flow speed, $w\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right)$ is the backwards wave speed, $\frac{q^{\max}\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right)}{k}$ is the capacity when the proportions of density in each class are $\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right)$, and k^{jam} is jam density. For consistency, conservation of flow must be satisfied, i.e. $\frac{\partial q_m(x,t)}{\partial x} = \frac{\partial k_m(x,t)}{\partial t}$ for all $m \in M$.

As with Daganzo (1994), to form the CTM, time is discretized into timesteps of dt. Links are then discretized into cells labeled by $i = 1, \dots, \mathcal{I}$ such that vehicles traveling at free-flow speed will travel exactly the distance of one cell per timestep. Let $n_i^m(t)$ be vehicles of class m in cell i at time t, where

$$n_i(t) = \sum_{m \in M} n_i^m(t) \tag{4.9}$$

Let $y_i^m(t)$ be vehicle of class m entering cell i from cell $i-1$ at time t. Then cell occupancy is defined by:

$$n_i^m(t+1) = n_i^m(t) + y_i^m(t) - y_{i+1}^m(t) \tag{4.10}$$

with total transition flows given by:

$$\begin{aligned} y_i(t) &= \sum_{m \in M} y_i^m(t) \\ &= \min\left\{\sum_{m \in M} n_{i-1}^m(t), Q_i(t), \frac{w_i(t)}{\sigma}\left(\mathcal{N} - \sum_{m \in M} n_i^m(t)\right)\right\} \end{aligned} \tag{4.11}$$

where $\mathcal{N}$ is the maximum number of vehicles that can fit in cell i and $Q_i(t)$ is the maximum flow.

Equation (4.11) defines the total transition flows, which will now be defined specific to vehicle class. To avoid dividing by zero, assume

$$n_{i-1}(t) > 0 \tag{4.12}$$

if $n_{i-1}(t) = 0$, there is no flow to propagate. As stated in Assumption 2, class-specific density is assumed to be uniformly distributed throughout the cell. Then class-specific transition flows are proportional to $\frac{n_{i-1}^m(t)}{n_{i-1}(t)}$.

$$y_i^m(t) = \frac{n_{i-1}^m(t)}{n_{i-1}(t)} \cdot \min\left\{\sum_{m\in M} n_{i-1}^m(t), Q_i(t), \frac{w_i(t)}{\sigma} \cdot \left(N - \sum_{m\in M} n_i^m(t)\right)\right\} \tag{4.13}$$

Equation (4.13) may be simplified to

$$y_i^m(t) = \min\left\{n_{i-1}^m(t), \frac{n_{i-1}^m(t)}{n_{i-1}(t)} \cdot Q_i(t), \frac{n_{i-1}^m(t)}{n_{i-1}(t)} \cdot \frac{w_i(t)}{\sigma} \cdot \left(N - \sum_{m\in M} n_i^m(t)\right)\right\} \tag{4.14}$$

which shows that the flow of class m is restricted by three factors:

1) Class-specific cell occupancy.
2) Proportional share of the capacity.
3) Proportional share of congested flow.

Assume class-specific flow is proportional to density, i.e. $\frac{k_m}{k}$ and all classes travel at the same speed. Also assume that $k > 0$, because if $k = 0$ then flow is also 0. Then:

$$q_m(x,t) = \frac{k_m}{k} \cdot \min\left\{\sigma \cdot k, q^{\max}\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right), w\left(\frac{k_1}{k}, \dots, \frac{k_{|M|}}{k}\right) \cdot \left(k^{\text{jam}} - k\right)\right\} \tag{4.15}$$

Let dt be the timestep and choose cell length such that $\sigma \cdot dt = 1$. Then cell length is 1, $\sigma = 1$, $x = i$, $k^{\text{jam}} = N$, $q^{\max}(t) = Q_i(t)$, and $k(x,t) = n_i(t)$. Cell length is chosen so that flow may traverse at most one cell per timestep to satisfy the Courant–Friedrichs–Lewy condition (Courant et al., 1928). Then:

$$q_m(x,t) = \frac{n_i^m(t)}{n_i(t)} \cdot \min\left\{n_i(t), q_i^{\max}(t), \frac{w_i(t)}{\sigma} \cdot \left(k^{\text{jam}} - k\right)\right\} = y_{i+1}^m(t) \tag{4.16}$$

except for the subindex of n the last term, which should be $i + 1$. As with Daganzo (1994) this difference is disregarded. (See Daganzo, 1995b for more discussion on this issue.) Therefore

$$\frac{\partial q_m(x,t)}{\partial x} = y_{i+1}^m(t) - y_i^m(t) \tag{4.17}$$

Since

$$\frac{\partial k_m(x,t)}{\partial t} = n_i^m(t+1) - n_i^m(t) \tag{4.18}$$

is the rate of change in cell occupancy with respect to time, the conservation of flow equation:

$$\frac{\partial q_m(x,t)}{\partial x} = -\frac{\partial k_m(x,t)}{\partial t} \tag{4.19}$$

is satisfied by the cell propagation function of Equation (4.10).

Levin and Boyles (2016) presented a car-following model based on kinematics to predict the speed-density relationship as a function of the reaction times of multiple classes. Car following models can be divided into several types as described by Brackstone and McDonald (1999) and Gartner et al. (2005). For instance, some predict fluctuations in the acceleration behavior of an individual driver in response to the vehicle ahead. However, for DTA a simpler model is more appropriate to predict the speed of traffic at a macroscopic level. Newell (2002) greatly simplified car following to be consistent with the hydrodynamic theory, but the model does not include the effects of reaction time. Instead, the car following model used here builds from the collision avoidance theory of Kometani and Sasaki (1959) to predict the allowed headway for a given speed, which varies with driver reaction time. The inverse relationship predicts speed as a function of the headway, which is determined by density. This car-following model results in the triangular fundamental diagram used by Newell (1993) and Yperman et al. (2005).

Although this car-following model is useful in predicting the effects of a heterogeneous vehicle composition on capacity and wave speed, other effects such as roadway conditions are not included. Furthermore, CTM assumes a trapezoidal fundamental diagram that admits a lower restriction on capacity. Therefore, the effect of reaction times on capacity and backward wave speed are used to appropriately scale link characteristics for realistic city network models.

Suppose that vehicle 2 follows vehicle 1 at speed u with vehicle lengths ϑ. Vehicle 1 decelerates at a to a full stop starting at time $t = 0$, and vehicle 2 follows suit after a reaction time of ω. The safe following distance, $\mathcal{L}$, is determined by kinematics.

The position of vehicle 1 is given by:

$$x_1(t) = \begin{cases} u \cdot t - \frac{1}{2} \cdot a \cdot t^2, & t \leq \frac{u}{a} \\ \frac{u^2}{2 \cdot a}, & t > \frac{u}{a} \end{cases} \tag{4.20}$$

where $\frac{u}{a}$ is the time required to reach a full stop. For $t > \frac{u}{a}$, the position of vehicle 1 is constant after its full stop. The position of vehicle 2, including the following distance of $\mathcal{L}$, is:

$$x_2(t) = \begin{cases} u \cdot t - \mathcal{L}, & t \leq \omega \\ u \cdot t - \frac{1}{2} \cdot a \cdot (t - \omega)^2, & t > \omega \end{cases} \tag{4.21}$$

The difference is

$$x_1(t) - x_2(t) = \begin{cases} u \cdot t - \frac{1}{2} \cdot a \cdot t^2 + \mathcal{L}, & t \le \omega \\ -a \cdot t \cdot \omega + \frac{1}{2} \cdot a \cdot (t - \omega)^2 + \mathcal{L}, & \omega < t \le \frac{u}{a} \\ \frac{u^2}{2 \cdot a} - u \cdot t + \frac{1}{2} \cdot a \cdot (t - \omega)^2 + \mathcal{L}, & t > \frac{u}{a} \end{cases} \tag{4.22}$$

and the minimum distance occurs when both vehicles are stopped, at $\frac{u}{a} + \omega$. To avoid a collision,

$$\mathcal{L} \ge -\frac{u^2}{2 \cdot a} + u \cdot \left(\frac{u}{a} + \omega\right) - \frac{1}{2} \cdot a \cdot \left(\frac{u}{a}\right)^2 + \vartheta = u \cdot \omega + \vartheta \tag{4.23}$$

Equivalently, Equation (4.23) may be expressed as:

$$u \le \frac{\mathcal{L} - \vartheta}{\omega} \tag{4.24}$$

which restricts speed based on following distance (from density). Flow may be determined from the relationship:

$$q = \left(\frac{\mathcal{L} - \vartheta}{\omega}\right) \cdot k \tag{4.25}$$

with $\mathcal{L} = \frac{1}{k}$, which is linear with respect to density.

The maximum density at which a speed of u is possible is $\frac{1}{u \cdot \omega + \vartheta}$ from Equation (4.24), and therefore capacity for free-flow speed of σ is:

$$q^{\max} = \sigma \cdot \frac{1}{\sigma \cdot \omega + \vartheta} \tag{4.26}$$

Backwards wave speed is:

$$w = -\frac{\frac{\sigma}{\sigma \cdot \omega + \vartheta}}{\frac{1}{\sigma \cdot \omega + \vartheta} - \frac{1}{\vartheta}} = \frac{\vartheta}{\omega} \tag{4.27}$$

which increases as reaction time decreases. Note that if $\omega < \frac{\vartheta}{\sigma}$, which may be possible for computer reaction times, then backward wave speed exceeds free-flow speed. If $w > \sigma$ for CTM, then the cell lengths would need to be derived from the backward wave speed, not the forward. That would complicate the cell transition flows. To avoid this issue, assume that $w \le \sigma$.

The car following model is designed to estimate the capacity and backward wave speed when the reaction time varies but is uniform across all vehicles. This section expands the model for heterogeneous flow with different vehicles having different reaction times. Let the density be disaggregated into k_m for each vehicle class m. Consider the case where speed is limited by density. Assuming that all vehicles travel at the same speed, for all vehicle classes:

$$u = \frac{\mathcal{L}_m - \vartheta}{\omega_m} \tag{4.28}$$

where $\mathcal{L}_m$ is the headway allotted and ω_m is the reaction time for vehicles of class m. Also, with appropriate units,

$$\sum_{m \in M} k_m \cdot \mathcal{L}_m = 1 \tag{4.29}$$

is the total distance occupied by the vehicles. Thus

$$\sum_{m \in M} k_m \cdot (\mathcal{L}_m - \vartheta) = 1 - k \cdot \vartheta \tag{4.30}$$

By Equation (4.28)

$$\sum_{m \in M} k_m \cdot u \cdot \omega_m = 1 - k \cdot \vartheta \tag{4.31}$$

and

$$u = \frac{1 - k \cdot \vartheta}{\sum_{m \in M} k_m \cdot \omega_m} \tag{4.32}$$

Equation (4.32) may be rewritten as

$$u \cdot \sum_{m \in M} k_m \cdot \omega_m = 1 - k \cdot \vartheta \tag{4.33}$$

Dividing both sides by k yields

$$u \cdot \sum_{m \in M} \frac{k_m}{k} \cdot \omega_m + \vartheta = \frac{1}{k} \tag{4.34}$$

Assuming that vehicle class proportions $\frac{k_m}{k}$ remain constant because all vehicles travel at the same speed, the maximum density for which a speed of σ is possible is

$$k = \frac{1}{\sigma \cdot \sum_{m \in M} \frac{k_m}{k} \cdot \omega_m + \vartheta} \tag{4.35}$$

which follows by taking the reciprocal of Equation (4.34). Capacity is

$$\chi^{\max} = \sigma \cdot \frac{1}{\sigma \cdot \sum_{m \in M} \frac{k_m}{k} \cdot \omega_m + \vartheta} \tag{4.36}$$

The vehicle lengths of CABs and HDVs are different in this study. Therefore, by expanding the method proposed by Levin and Boyles (2016), the capacity in this study is calculated by assuming two different vehicle lengths. The capacity under the mixed flows is calculated as:

$$\chi_s(v_s, \mathbf{f}) = \frac{\sigma}{\sigma \cdot \left(\frac{v_s}{v_s + v_{s_B}} \cdot \omega_H + \frac{v_{s_B}}{v_s + v_{s_B}} \cdot \omega_B\right) + \left(\frac{v_s}{v_s + v_{s_B}} \cdot \vartheta_H + \frac{v_s}{v_s + v_{s_B}} \cdot \vartheta_B\right)} \quad \forall s \in S_H \tag{4.37}$$

where

$$v_{s_B} = \sum_{l \in \psi(s)} f_l \cdot E_b \tag{4.38}$$

$$\omega_m = \frac{1}{\chi_m} - \frac{\vartheta_m}{\sigma} \ \forall\, m \in \{H, B\} \tag{4.39}$$

σ is the free-flow speed of vehicles. χ_m is the capacity when both the reaction time and the vehicle length are same among vehicles with the same transport mode of $m \in \{H, B\}$. ω_m is the reaction time of vehicle of which transport mode is $m \in \{H, B\}$ (Wang et al., 2019). $\omega_s \triangleq \left(\frac{v_s}{v_s + v_{s_B}} \cdot \omega_H + \frac{v_{s_B}}{v_s + v_{s_B}} \cdot \omega_B\right)$ in Equation (4.37) is the average reaction time of the mixed flow on s. $\vartheta_s \triangleq \left(\frac{v_s}{v_s + v_{s_B}} \cdot \vartheta_H + \frac{v_s}{v_s + v_{s_B}} \cdot \vartheta_B\right)$ is the average vehicle length of the mixed flow on s. From Equation (4.39), the critical headway time of the mixed flow, which is the reciprocal of the corresponding capacity, is calculated as:

$$\frac{1}{\chi_s} = \omega_s + \frac{\vartheta_s}{\sigma} \quad \forall s \in S_H \tag{4.40}$$

Computer-controlled CABs have shorter reaction times than both HDVs and HDBs, which leads to differences in road capacity.

For the road network and associated bus services, the actual travel time is presumed to adhere to a standard bureau of public roads (BPR) function, incorporating an interaction term between CABs and HDVs flow. Thus, the actual travel time, $\bar{t}_s$, for link $s \in S_H$ within the network can be expressed as:

$$\bar{t}_s(v_s, \mathbf{f}) = t_s \cdot \left(1 + \alpha_s \cdot \left(\frac{v_s + v_{s_B}}{\chi_s}\right)^{\beta_s}\right) \quad \forall\, s \in S_H \tag{4.41}$$

Since this study assumes that each HDV carries one person, the passenger flow v_s represents the flow of HDVs on link $s \in S_H$.

4.4.3 Cost functions

A volume-delay function can be used to characterize the link congestion caused by flow on it. The perceived in-vehicle travel time depends on the two terms: (i) the passenger flow $\tilde{v}_s$ on the competing route sections and (ii) the passenger flow v_s to board on route section s.

i. The passenger flow $\tilde{v}_s$ on competing route sections.

   It is computed by $\tilde{v}_s = \sum_{s \in S_{n(s)}^{l+}} v_s^l + \sum_{s \in S_{n(s)}^{l-}} v_s^l$, where $n(s)$ indicates that n is the origin node of route section s and $S_{n(s)}^{l+}$ represents the set of the route sections that originate from node $n(s)$, excluding route section s, contains line $l \in \psi(s)$ as an attractive line, meanwhile, $S_{n(s)}^{l-}$ refers to the set of route sections containing line $l \in \psi(s)$ as an attractive line, with the node preceding $n(s)$ and the node after $n(s)$.

ii. The passenger flow v_s to board on route section s.

Referring to Uchida et al. (2005, 2007), the perceived in-vehicle travel time on line l over road link s can be modeled using an in-vehicle discomfort function of BPR function type:

$$\hat{t}_s(\mathbf{v}, \mathbf{f}) = \frac{\sum_{l \in \psi(s)} f_l \cdot \hat{t}_s^l}{\sum_{l \in \psi(s)} f_l} \quad \forall\, s \in S_B \tag{4.42}$$

where

$$\hat{t}_s^l = \tilde{t}_s \cdot \left(1 + \beta_1 \cdot \left(\frac{v_s + \tilde{v}_s}{f_l \cdot k_l}\right)^{\gamma_1}\right) \quad \forall\, s \in S_B \tag{4.43}$$

$\tilde{t}_s$ represents the same travel time as that of a link on which bus line l over road link s travels, and k_l means practical capacity of line $l \in \psi(s)$. Then, $\tilde{t}_s$ can be written as follows,

$$\tilde{t}_s = \sum_{s \in \kappa(s,l)} \bar{t}_s \tag{4.44}$$

where $\kappa(s, l)$ denotes a set of direct links with respect to line l that are included in the set of attractive lines $l \in \psi(s)$ over link s.

The calculations for the waiting times for CABs are based on the notion of in-vehicle perceived travel time, which considers both the number of passengers already on board and those waiting to board at link s. The waiting time for passengers at link s is influenced by the number of passengers boarding before link s and those remaining in the vehicle after link s, as well as the passengers waiting to board at link s itself. The first term is $\tilde{v}_s$ and the second term is simply v_s as defined earlier.

$$w_s(\mathbf{v}, \mathbf{f}) = \frac{\alpha}{\sum_{l \in \psi(s)} f_l} + \beta_2 \cdot \left(\frac{v_s + \tilde{v}_s}{\sum_{l \in \psi(s)} f_l \cdot k_l}\right)^{\gamma_2} \quad \forall\, s \in S_B \tag{4.45}$$

where β_2 and γ_2 indicate calibration parameters and α is assumed to be 1 in this study, i.e., arrival headway time of CABs/HDBs follows an exponential distribution.

Giving the example for calculating the link cost for CABs based on Fig. 4.1, for the link s_2 between node n_1 and n_2, the following relationship holds:

$$\psi(s_2) = \{l_1, l_2\}$$

$$v_{s_2} = v_{s_2}^{l_1} + v_{s_2}^{l_2}$$

$$\tilde{v}_{s_2}^{l_1} = v_{s_1} \; and \; \tilde{v}_{s_2}^{l_2} = 0$$

$$\tilde{v}_{s_2} = v_{s_1}$$

Then it is assumed that there is a road link $s \in S_H$ between node N_1 and N_2, the actual travel time for link s based on Equation (4.41) is given by:

$$\bar{t}_s = t_s \cdot \left(1 + \alpha_s \cdot \left(\frac{v_s + v_{s_B}}{\chi_s}\right)^{\beta_s}\right)$$

Therefore, according to Equation (4.44), $\tilde{t}_{s_2}$ is:

$$\tilde{t}_{s_2} = \bar{t}_s$$

In-vehicle discomfort $\hat{t}_{s_2}$ for link s_2 based on Equation (4.42) is:

$$\hat{t}_{s_2} = \frac{f_{l_1} \cdot \hat{t}_{s_2}^{l_1} + f_{l_2} \cdot \hat{t}_{s_2}^{l_2}}{f_{l_1} + f_{l_2}}$$

where

$$\hat{t}_{s_2}^{l_1} = \tilde{t}_{s_2} \cdot \left(1 + \beta_1 \cdot \left(\frac{v_{s_2}^{l_1} + v_{s_1}}{f_{l_1} \cdot k_{l_1}}\right)^{\gamma_1}\right)$$

Similarly,

$$\hat{t}_{s_2}^{l_2} = \tilde{t}_{s_2} \cdot \left(1 + \beta_1 \cdot \left(\frac{v_{s_2}^{l_2}}{f_{l_2} \cdot k_{l_2}}\right)^{\gamma_1}\right)$$

The waiting time w_{s_2} of link s_2 based on Equation (4.45) can be calculated as

$$\begin{aligned} w_{s_2} &= \frac{\alpha}{f_{l_1} + f_{l_2}} + \beta_2 \cdot \left(\frac{v_{s_2} + \tilde{v}_{s_2}}{f_{l_1} \cdot k_{l_1} + f_{l_2} \cdot k_{l_2}}\right)^{\gamma_2} \\ &= \frac{\alpha}{f_{l_1} + f_{l_2}} + \beta_2 \cdot \left(\frac{v_{s_2} + v_{s_1}}{f_{l_1} \cdot k_{l_1} + f_{l_2} \cdot k_{l_2}}\right)^{\gamma_2} \end{aligned}$$

4.5 Formulation of UE assignment and sensitivity analysis for multiclass UE assignment

4.5.1 Single-class and multi-class UE assignment

According to Equation (4.3), the disutility of the link $s \in S_B$ for CABs is:

$$d_s = \pi_t \cdot \hat{t}_s + \rho_s \cdot w_s + \tau_p \cdot p_s \;\; \forall\, s \in S_B \tag{4.46}$$

since the travelers who use HDVs experience no waiting time, disutility of road link $s \in S_H \cup S_W$ are:

$$d_s = \pi_t \cdot \bar{t}_s \;\; \forall\, s \in S_H \cup S_W \tag{4.47}$$

then the general path cost function between O-D pair $i \in I$ based on Equation (4.4) is given by:

$$\pi_{i,j} = \sum_{m \in M} \lambda_m \cdot \xi_{i,j,m} + \sum_{s \in S} d_s \cdot \delta_{i,j,s} \;\; \forall\, i \in I, \forall\, j \in J_i \tag{4.48}$$

Let $h_{i,j}$ denote the passenger flow on path j between O-D pair $i \in I$, and q_i is the travel demand between O-D pair $i \in I$. Then we have:

$$\sum_{j \in J_i} h_{i,j} = q_i \quad \forall i \in I \tag{4.49}$$

$$v_s = \sum_{i \in I} \sum_{j \in J_i} h_{i,j} \cdot \delta_{i,j,s} \quad \forall s \in S \tag{4.50}$$

Equation (4.49) is the flow conservation condition and Equation (4.50) defines the relationship between link flow and path flows.

Each link flow in the network is calculated by the Wardrop's first principle. In other words, travelers using the same O-D pair, cannot reduce their travel costs by individually altering their paths within their available set of choices.

All paths used by travelers in the network for an O-D pair $i \in I$ have equal and minimal travel times. When $h_{i,j}$ satisfies the Wardropian first principle, the following condition hold:

$$\pi_{i,j} = \begin{cases} = \mu_i, & h_{i,j} \geq 0 \\ \geq \mu_i, & h_{i,j} = 0 \end{cases} \quad \forall i \in I, \forall j \in J_i \tag{4.51}$$

Equation (4.51) says that paths with travel costs greater than the equilibrium cost carry no traffic flow. The path-based single-class UE (SCUE) has been demonstrated to be equivalent to identifying a solution $\mathbf{h}^*$ for the following variational inequality (VI) problem:

$$\boldsymbol{\pi}(\mathbf{h}^*)^{\mathrm{T}} \cdot (\mathbf{h} - \mathbf{h}^*) \geq 0 \ \mathbf{h} \in \Omega_{\mathbf{h}} \tag{4.52}$$

$$\Omega_{\mathbf{h}} = \{\mathbf{h} | \boldsymbol{\Lambda} \cdot \mathbf{h} = \mathbf{q}, \mathbf{h} \geq \mathbf{0}\} \tag{4.53}$$

For the proof of the VI formulation mentioned above, one can see Florian and Spiess (1983) and De Cea and Fernández (1993).

When we address a problem under the fixed flow of HDV and CABs/HDBs shown by Equation $(4.49a)$, the VI problem shown by Equation (4.52) needs to be modified so that it can address a multi-class UE (MCUE) as discussed below.

$$\sum_{j \in J_{i,t}} h_{i,j,t} = q_{i,t} \quad \forall i \in I, \forall t \in \{p, b\} \tag{4.49a}$$

$q_{i,t}$ is the O-D demand traveling between O-D pair $i \in I$ using HDVs when $t = p$ and that using HDBs/CABs when $t = b$. The corresponding path flows are denoted by $h_{i,j,t}$. The set of paths for HDVs users between O-D pair $\forall i \in I$ is $J_{i,p}$, and that for CABs/HDBs users is $J_{i,b}$. It is assumed $J_i = J_{i,p} \cup J_{i,b}$. Every traveler in the multi-modal network is either HDVs user or CABs/HDBs user, and travels on the real road network since HDVs and CAVs/CABs share the same road space. However, the path utility of HDVs users is different from CABs/HDBs users even though they travel on the same road links as shown by Equation (4.48). Path disutilities of two classes of users, i.e., HDVs users and CABs/HDBs users, are as follows:

$$\pi_{i,j,p} = \lambda_H + \lambda_W + \sum_{s \in S} d_s \cdot \delta_{i,j,s} \;\; \forall\, i \in I, \forall\, j \in J_{i,p} \tag{4.48a}$$

$$\pi_{i,j,b} = \lambda_W + \lambda_B + \sum_{s \in S} d_s \cdot \delta_{i,j,s} \;\; \forall\, i \in I, \forall\, j \in J_{i,b} \tag{4.48b}$$

where $\pi_{i,j,p}$ and $\pi_{i,j,b}$ are disutilities of paths $j \in J_{i,p}$ and $j \in J_{i,b}$, respectively. The corresponding path flows are denoted by $h_{i,j,p}$ and $h_{i,j,b}$, respectively. The MCUE conditions are the following:

$$\pi_{i,j,t} = \begin{cases} = \mu_{i,t}, & h_{i,j,t} \geq 0 \\ \geq \mu_{i,t}, & h_{i,j,t} = 0 \end{cases} \;\; \forall i \in I, \forall\, j \in J_{i,t}, \forall t \in \{p, b\} \tag{4.51a}$$

vector of $\pi_{i,j,p}$ or $\pi_{i,j,b}$ is denoted as $\boldsymbol{\pi}_p$ or $\boldsymbol{\pi}_b$. Similarly, vector of $h_{i,j,p}$ or $h_{i,j,b}$ is denoted as $\mathbf{h}_p$ or $\mathbf{h}_b$. The VI problem equivalent to Equation $(4.19a)$ is shown below:

$$\boldsymbol{\pi}_p(\mathbf{h}_p^*)^{\mathrm{T}} \cdot (\mathbf{h}_p - \mathbf{h}_p^*) + \boldsymbol{\pi}_b(\mathbf{h}_b^*)^{\mathrm{T}} \cdot (\mathbf{h}_b - \mathbf{h}_b^*) \geq 0 \;\; \mathbf{h}_p \in \Omega_{\mathbf{h}_p}, \mathbf{h}_b \in \Omega_{\mathbf{h}_b} \tag{4.52a}$$

where $\Omega_{\mathbf{h}_p}$ and $\Omega_{\mathbf{h}_b}$ are feasible set of path flows for HDVs and that for CABs/HDBs, respectively. $\Omega_{\mathbf{h}_p}$ and $\Omega_{\mathbf{h}_b}$ are respectively defined as $\Omega_{\mathbf{h}_p} = \{\mathbf{h}_p | \boldsymbol{\Lambda}_p \cdot \mathbf{h}_p = \mathbf{q}_p, \mathbf{h}_p \geq \mathbf{0}\}$ and $\Omega_{\mathbf{h}_b} = \{\mathbf{h}_b | \boldsymbol{\Lambda}_b \cdot \mathbf{h}_b = \mathbf{q}_b, \mathbf{h}_b \geq \mathbf{0}\}$, where $\boldsymbol{\Lambda}_t$ is O-D pair/path incidence matrix and $\mathbf{q}_t$ is vector of O-D flows, for $t \in \{p, b\}$.

Nagurney (2000) established conditions under which a solution exists for a VI problem with the same structure as Equation (4.52). Specifically, it was demonstrated that when the path disutility $\pi_{i,j}$ is a continuous function defined in a closed and convex feasible region $\mathbf{h} \in \Omega_{\mathbf{h}}$, there exists at least one solution to Equation (4.52). This result is supported by the properties described in Equations (4.49), (4.50), and (4.53). In more straightforward cases where the link disutility is separable with respect to flow on the link, the proof of solution uniqueness typically hinges on the strict monotonicity of the link disutility function. However, the link disutility in this study is non-separable. As a result, the uniqueness of solutions to Equation (4.52) cannot be assured under the given conditions.

4.5.2 Sensitivity analysis for multi-class UE assignment

This chapter presents the calculation of the local linear approximation of UE link flow for both CABs and HDVs with respect to the perturbed parameters $\boldsymbol{\varepsilon}$ in the MCUE problem shown in Equation $(4.52a)$. The local approximation of UE link flow will be used to develop an algorithm for solving the bus network design problem to be formulated in Chapter 4.6. Let $\mathbf{h}^{**} = [\mathbf{h}_b^{**} \;\; \mathbf{h}_p^{**}]$ represent equilibrium path flow solution to the VI problem in Equation $(4.52a)$ at $\boldsymbol{\varepsilon} = \mathbf{0}$. The KKT conditions at $\mathbf{h}^{**}$ are:

$$\begin{bmatrix}\boldsymbol{\pi}_b(\mathbf{h}^{**},\mathbf{0})\\\boldsymbol{\pi}_p(\mathbf{h}^{**},\mathbf{0})\end{bmatrix}-\begin{bmatrix}\boldsymbol{\varrho}_b\\\boldsymbol{\varrho}_p\end{bmatrix}-\begin{bmatrix}\boldsymbol{\Lambda}_b^{\mathrm{T}}\cdot\boldsymbol{\mu}_b\\\boldsymbol{\Lambda}_p^{\mathrm{T}}\cdot\boldsymbol{\mu}_p\end{bmatrix}=\mathbf{0} \tag{4.54}$$

$$\begin{bmatrix}\boldsymbol{\varrho}_b^{\mathrm{T}} & \boldsymbol{\varrho}_p^{\mathrm{T}}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{h}_b^{**}\\\mathbf{h}_p^{**}\end{bmatrix}=\mathbf{0} \tag{4.55}$$

$$\begin{bmatrix}\boldsymbol{\Lambda}_b & \mathbf{0}\\\mathbf{0} & \boldsymbol{\Lambda}_h\end{bmatrix}\begin{bmatrix}\mathbf{h}_b^{**}\\\mathbf{h}_p^{**}\end{bmatrix}-\begin{bmatrix}\mathbf{q}_b\\\mathbf{q}_p\end{bmatrix}=\mathbf{0} \tag{4.56}$$

$$\begin{bmatrix}\boldsymbol{\varrho}_b\\\boldsymbol{\varrho}_p\end{bmatrix}\geq\mathbf{0} \tag{4.57}$$

$$\begin{bmatrix}\mathbf{h}_b^{**}\\\mathbf{h}_p^{**}\end{bmatrix}\geq\mathbf{0} \tag{4.58}$$

where $\boldsymbol{\pi}_t$ is the vector of path disutilities, $\mathbf{h}_t^{**}$ is the vector of equilibrium path flows, $\mathbf{q}_t$ is the vector of O-D flows and $\boldsymbol{\mu}_t$ and $\boldsymbol{\varrho}_t$ are vectors of Lagrange multipliers, respectively, for $t\in\{b,p\}$.

Referring to Tobin and Friesz (1988), the Equations (4.54)-(4.58) can be simplified as:

$$\begin{bmatrix}\boldsymbol{\pi}_b^0(\mathbf{h}^{**},\mathbf{0})\\\boldsymbol{\pi}_p^0(\mathbf{h}^{**},\mathbf{0})\end{bmatrix}-\begin{bmatrix}\boldsymbol{\Lambda}_b^{0\mathrm{T}}\cdot\boldsymbol{\mu}_b\\\boldsymbol{\Lambda}_p^{0\mathrm{T}}\cdot\boldsymbol{\mu}_p\end{bmatrix}=\mathbf{0} \tag{4.59}$$

$$\begin{bmatrix}\boldsymbol{\Lambda}_b^0 & \mathbf{0}\\\mathbf{0} & \boldsymbol{\Lambda}_p^0\end{bmatrix}\begin{bmatrix}\mathbf{h}_b^0\\\mathbf{h}_p^0\end{bmatrix}-\begin{bmatrix}\mathbf{q}_b\\\mathbf{q}_p\end{bmatrix}=\mathbf{0} \tag{4.60}$$

where the superscript 0 represents the corresponding reduced vectors and matrixes.

Then, differentiating both sides of Equations (4.59) and (4.60) with respect to the perturbations $\boldsymbol{\varepsilon}$ ($\boldsymbol{\varepsilon}=\mathbf{0}$) yields:

$$\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_b^0\\\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_p^0\\\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_b\\\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_p\end{bmatrix}=\boldsymbol{J}_{\mathbf{h}^{**}}(\mathbf{0})^{-1}\begin{bmatrix}-\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\pi}_b^0(\mathbf{h}^{**},\mathbf{0})\\-\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\pi}_p^0(\mathbf{h}^{**},\mathbf{0})\\\mathbf{0}\\\mathbf{0}\end{bmatrix} \tag{4.61}$$

where

$$\boldsymbol{J}_{\mathbf{h}^{**}}(\mathbf{0})=\begin{bmatrix}\nabla_{\mathbf{h}_b}\boldsymbol{\pi}_b^0(\mathbf{h}^{**},\mathbf{0}) & \nabla_{\mathbf{h}_p}\boldsymbol{\pi}_b^0(\mathbf{h}^{**},\mathbf{0}) & -\boldsymbol{\Lambda}_b^{0\mathrm{T}} & \mathbf{0}\\\nabla_{\mathbf{h}_b}\boldsymbol{\pi}_p^0(\mathbf{h}^{**},\mathbf{0}) & \nabla_{\mathbf{h}_p}\boldsymbol{\pi}_p^0(\mathbf{h}^{**},\mathbf{0}) & \mathbf{0} & -\boldsymbol{\Lambda}_p^{0\mathrm{T}}\\\boldsymbol{\Lambda}_b^0 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\\mathbf{0} & \boldsymbol{\Lambda}_p^0 & \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{4.62}$$

For the calculation of the inverse matrix $\boldsymbol{J}_{\mathbf{h}^{**}}(\mathbf{0})^{-1}$ of Equation (4.62), the reader is referred to calculation methods shown in Yang and Bell (2007). Thereby, the first derivative of path flows $\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_b^0\\\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_p^0\end{bmatrix}$ with respect to $\mathbf{f}$ at $\boldsymbol{\varepsilon}=\mathbf{0}$ can be calculated by (4.61), and the following relationship holds:

$$\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_b^0\\\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_p^0\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Delta}_b^{0\mathrm{T}} & \boldsymbol{\Delta}_p^{0\mathrm{T}}\end{bmatrix}\cdot\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_b^0\\\nabla_{\boldsymbol{\varepsilon}}\mathbf{h}_p^0\end{bmatrix} \tag{4.63}$$

considering an initial solution $\mathbf{f}^*$ and the corresponding equilibrium link flow $\mathbf{v}_t^*(\mathbf{f}^*)$ $(t\in\{b,p\})$, then,

$$\begin{bmatrix} \tilde{\mathbf{v}}_b^*(\mathbf{f}+\boldsymbol{\varepsilon}) \\ \tilde{\mathbf{v}}_p^*(\mathbf{f}+\boldsymbol{\varepsilon}) \end{bmatrix} \approx \begin{bmatrix} \mathbf{v}_b^*(\mathbf{f}^*) + \nabla_{\boldsymbol{\varepsilon}} \mathbf{v}_b^0 \cdot \boldsymbol{\varepsilon} \\ \mathbf{v}_p^*(\mathbf{f}^*) + \nabla_{\boldsymbol{\varepsilon}} \mathbf{v}_p^0 \cdot \boldsymbol{\varepsilon} \end{bmatrix} \tag{4.64}$$

the local linear approximation of UE link flows for SCUE is:

$$\tilde{\mathbf{v}}^*(\mathbf{f}+\boldsymbol{\varepsilon}) \approx \mathbf{v}^*(\mathbf{f}) + \nabla_{\boldsymbol{\varepsilon}} \mathbf{v}^0 \cdot \boldsymbol{\varepsilon} \tag{4.65}$$

the detailed calculation process can be found in Tobin and Friesz (1988), and Gao et al. (2004).

4.6 Bus route network and service frequencies problem

The travel behavior of travelers within a multi-modal network where bus service is available, including their choices of transport mode and path, is influenced by the service frequencies assigned to each bus route (or bus line). Changes in service frequency can directly affect the attractiveness and utility of different paths, thereby influencing overall travel patterns.

This study introduces an optimization problem that simultaneously addresses: 1) bus route network design problem and 2) optimal frequencies for bus routes. The bus route network design problem addressed in this study is to choose one from several bus route candidates which are predefined for each bus route O-D pair in a road network. Furthermore, the frequency of each bus route selected needs to be optimized. The bus network design problem in this study chooses one from several bus route candidates and determines its service frequency simultaneously, so that the total disutility (or the total cost) across the network is minimized.

The example network is used to explain the optimization problem of bus route network design and service frequencies of each selected bus route. The example network has 4 O-D pairs, and for O-D pair of O_1-D_1 which is served by bus route candidates $l_{1,1}$ and $l_{1,2}$. Table.4.1 shows the information on bus lines and stop nodes for each bus line.

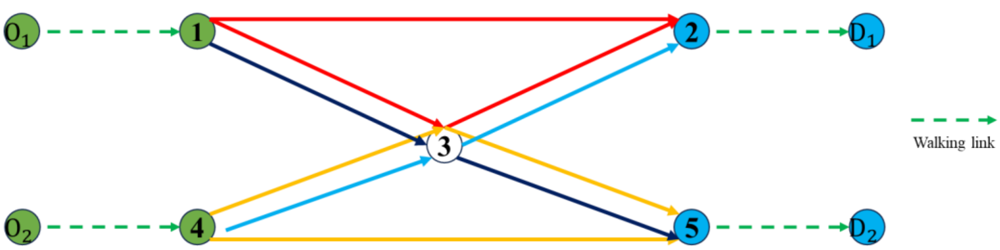


Fig.4.2 Example network

Table.4.1 Bus lines and stop nodes

O-D pair	Line	Stop nodes
	$l_{1,1}$	1-2
$O_1 - D_1$	$l_{1,2}$	1-3-2
$O_1 - D_2$	$l_{2,1}$	4-3-2
$O_2 - D_1$	$l_{3,1}$	1-3-5
	$l_{4,1}$	4-5
$O_2 - D_2$	$l_{4,2}$	4-3-5

The nonlinear complementary conditions equivalent to the conditions for choosing one from three bus line candidates are shown below:

$$f_{1,1} \cdot f_{1,2} = 0 \tag{4.66}$$

$$\zeta \leq f_{1,1}, f_{1,2} \leq f_{max} \tag{4.67}$$

where $\zeta > 0$ is a predetermined, positive and small value, and f_{max} represents the maximum bus service frequency.

The possible combinations are:

$$\begin{cases} f_{1,1} \\ f_{1,2} \\ \text{both are not chosen} \end{cases} \tag{4.68}$$

Consider a road network with $\mathcal{k}$ O-D pairs, where each O-D pair is served by $\mathcal{b}$ bus route candidates. The maximum number of combinations to choose at most one from $\mathcal{b}$ bus route candidates in the network is $(\mathcal{b}+1)^{\mathcal{k}}$, and that increases exponentially as the number of O-D pairs increases. Furthermore, the frequency of each bus route selected needs to be optimized.

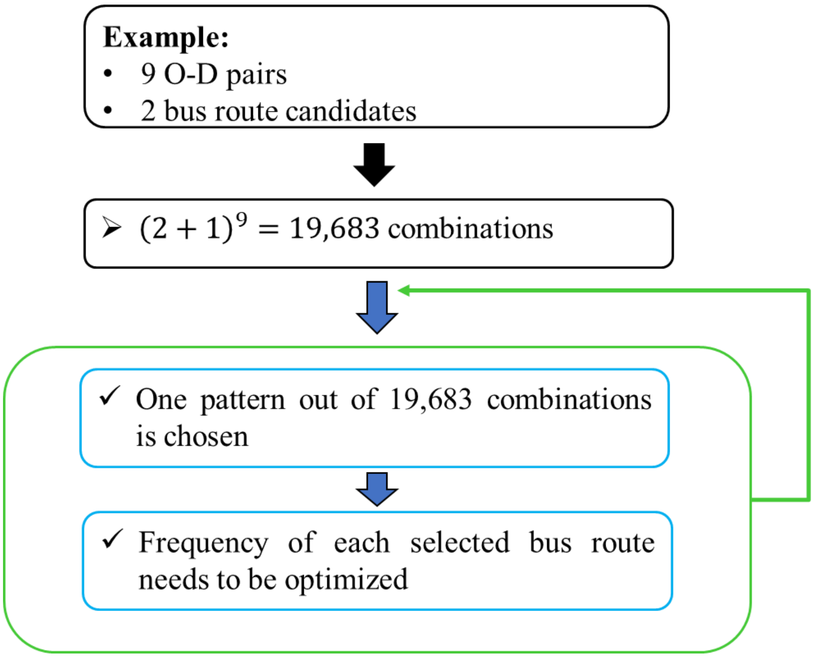


Fig.4.3 Process of solving the bus network work design problem

The nonlinear complementary conditions shown in Equations (4.66) and (4.67) imply that no more than one bus route is chosen per O-D pair $i \in I$. These conditions can be expressed as:

$$\sum_{l_1,l_2(\neq l_1)\in L_i}\left(-(f_{i,l_1}+f_{i,l_2})+\sqrt{{f_{i,l_1}}^2+{f_{i,l_2}}^2}\right)^2<\phi \tag{4.69}$$

where $\phi > 0$ is a small constant number.

Therefore, the bus network design problem in this study is formulated as follows:

$$\min_{\mathbf{f}} Z(\mathbf{v}^*(\mathbf{f}),\mathbf{f}) = \mathbf{d}(\mathbf{v}^*(\mathbf{f}),\mathbf{f})^T\cdot\mathbf{v}^*(\mathbf{f})+\theta\cdot\mathbf{F}^T\cdot\mathbf{f} \tag{4.70}$$

s.t.

$$\boldsymbol{\zeta}\leq\mathbf{f}\leq\mathbf{f}_{max} \tag{4.71}$$

$$\sum_{i\in I}\sum_{l_1,l_2(\neq l_1)\in L_i}\left(-(f_{i,l_1}+f_{i,l_2})+\sqrt{{f_{i,l_1}}^2+{f_{i,l_2}}^2}\right)^2<\phi \tag{4.72}$$

where $\mathbf{v}^*(\mathbf{f})$ is the vector of equilibrium link flows shown in Equation (4.70). θ is a coefficient converting fare caused by frequency setting to travel cost, $\boldsymbol{\zeta}>\mathbf{0}$ is vector of a predetermined, positive and small values, $\mathbf{f}_{max}$ is vector of maximum bus service frequencies and $\mathbf{F}$ is the vector of operational costs per frequency increase on the bus lines. The objective function is total disutility in the multimodal network.

Depending on O-D demands in the multimodal network, there may be a situation where setting more than one bus route for each O-D pair is more efficient for reducing total cost. Such examinations are possible to replace the constraint (4.69) with $(4.\mathrm{A}5)$ shown in Chapter 4.12.

4.7 Sensitivity analysis-based solution algorithm for multimodal NDPs

The objective function shown by Equations (4.70)-(4.72) is a nonlinear optimization problem with nonlinear constraints. We first solve the UE problem under predefined service frequencies to obtain new equilibrium link flows and then calculate the local linear approximation of the UE link flows $\tilde{\mathbf{v}}^*(\mathbf{f})$ using the sensitivity analysis. Optimal frequencies and bus routes which minimize Equation (4.70) can be obtained. Optimal service frequencies obtained here are inputs to the UE problem. By repeating this process, we can iteratively derive new optimal frequency values and corresponding optimal bus routes. The solution algorithm developed in this study can be outlined as follows:

Step 0: Set an initial feasible solution $\mathbf{f}_0$ and initialize the outer iteration counter $k = 0$.

Step 1: Calculate the vector of UE link flows $\mathbf{v}^*(\mathbf{f}_k)$ as the response from passengers to the current vector of frequencies $\mathbf{f}_k$. This can be calculated by applying the Method of Successive Averages (MSA; Sheffi, 1985).

Step 2: Solve the following sub-problem for the bus network design problem:

$$\min_{\boldsymbol{\varepsilon}} Z\left(\tilde{\mathbf{v}}^*\left(\hat{\mathbf{f}}_k\right), \hat{\mathbf{f}}_k\right) = \mathbf{d}\left(\tilde{\mathbf{v}}^*\left(\hat{\mathbf{f}}_k\right), \hat{\mathbf{f}}_k\right)^T \cdot \tilde{\mathbf{v}}^*\left(\hat{\mathbf{f}}_k\right) + \theta \cdot \mathbf{F}^T \cdot \hat{\mathbf{f}}_k \tag{4.73}$$

s.t.

$$\boldsymbol{\zeta} \le \hat{\mathbf{f}}_k \le \mathbf{f}_{max} \tag{4.74}$$

$$\sum_{i \in I} \sum_{l_1, l_2 (\neq l_1) \in L_i} \left(-\left(\hat{f}_{k,i,l_1} + \hat{f}_{k,i,l_2}\right) + \sqrt{\hat{f}_{k,i,l_1}{}^2 + \hat{f}_{k,i,l_2}{}^2} \right)^2 < \phi \tag{4.75}$$

where $\hat{f}_{k,i,l} = f_{k,i,l} + \varepsilon_{i,l}$ and $\hat{\mathbf{f}}_k = \mathbf{f}_k + \boldsymbol{\varepsilon}$. $\tilde{\mathbf{v}}^*\left(\hat{\mathbf{f}}_k\right)$ is the vector of linearly approximated equilibrium link flows given by Equation (4.65). The solution obtained here is denoted as $\boldsymbol{\varepsilon}^*$.

Step 3: If $max|\boldsymbol{\varepsilon}^*| < \varsigma$ then calculation stops, otherwise update $\mathbf{f}_{k+1} = (1 - 1/k) \cdot \mathbf{f}_k + (1/k) \cdot \hat{\mathbf{f}}_k$, set

$k = k + 1$ and return to Step 1, where ς is a predetermined, positive, small value.

4.8. Numerical experiments

4.8.1 Effects of CABs on link capacity

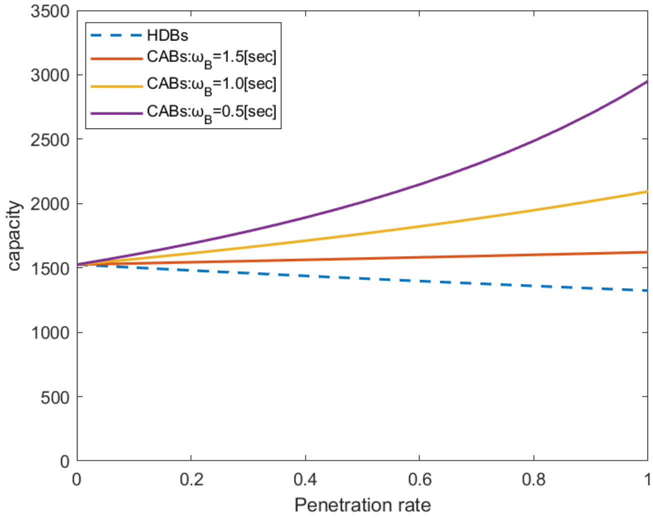


Fig. 4.4 Effect of HDBs or CABs penetration on capacity.

Fig. 4.4 illustrates the relationship between the penetration rate of CABs and capacity, with various scenarios based on different reaction times. The initial capacity when only HDVs are present is 1,525 [pcu/h], which is calculated by using σ =50 [km/h], ξ_H =5 [m] and ω_H =2.0 [sec], from Equation (4.40). As the penetration rate of HDBs increases, the link capacity gradually decreases from the baseline value of 1,525 [pcu/h] to 1,323 [pcu/h]. Conversely, the capacity improves across all scenarios where CABs are implemented. Even at low penetration rates, CABs with shorter reaction times begin to outperform HDBs in terms of capacity. At higher penetration rates, the benefits of CABs become increasingly apparent, the capacity reaches 1,621 [pcu/h] when ω_B =1.5 [sec]. While the capacity increases to 2,093 [pcu/h] when ω_B =1.0 [sec] and reaches nearly double that of HDBs in scenarios where the reaction time of CABs is ω_B =0.5 [sec]. Fig. 4.4 shows that the implementation of CABs with shorter reaction times can increase capacity. A higher penetration rate of CABs and shorter reaction times can increase capacity.

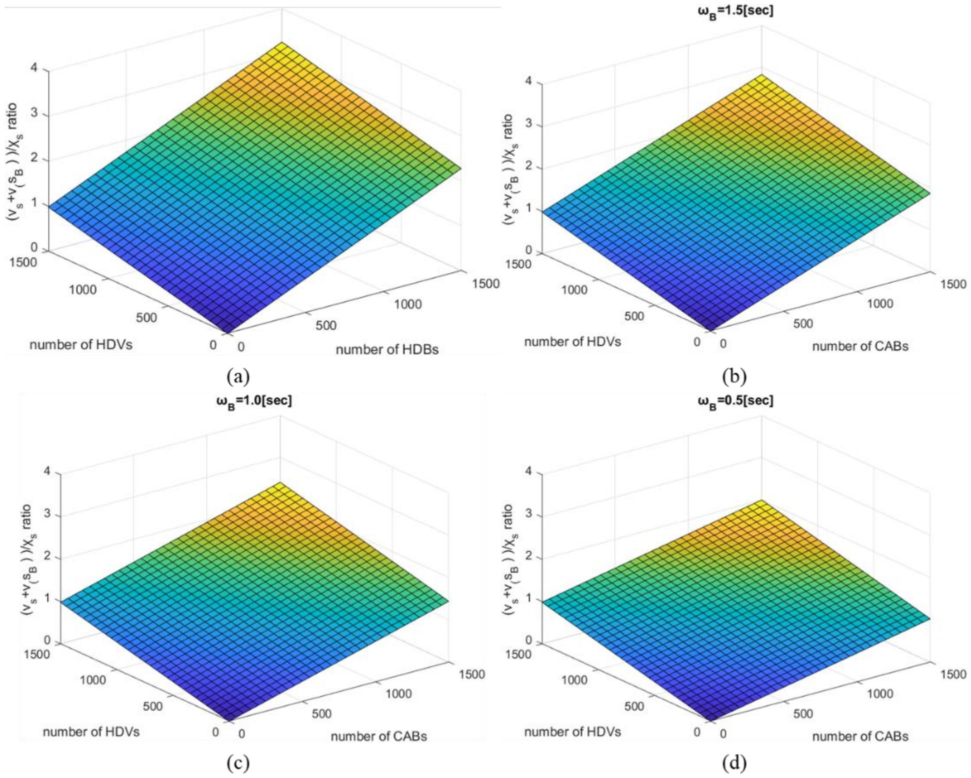


Fig. 4.5 Results of $\left(v_s + v_{s_B}\right)/\chi_s$ ratios in mixed flow of: (a) HDVs and HDBs, (b) HDVs and CABs (ω_B =1.5 [sec]), (c) HDVs and CABs (ω_B =1.0 [sec]), and (d) HDVs and CABs (ω_B =0.5 [sec]).

Fig. 4.5 depicts the volume-to-capacity ratio $\left(v_s + v_{s_B}\right)/\chi_s$ as defined in Equation (4.41), in relation to the increasing number of vehicles across different scenarios. Note that the calculations for the number of HDBs and CABs must incorporate the passenger car equivalent factor for buses, as shown in Equation (4.38). In Fig. 4.5(a), (b), (c) and (d), when the network consists only of HDVs (i.e., with zero HDBs or CABs), the changes in the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio do not change. In Fig. 4.5(a), when the network is full entirely of HDBs, the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio is higher compared to a scenario with only HDVs. As the number of HDBs increases, the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio increases more quickly, indicating that HDBs alone can contribute to congestion. When 1,500 HDBs are traveling on the link, the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio is 2.26.

In contrast, in the scenarios where the network is served exclusively by CABs under different reaction times, the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio is 1.85, which is lower than in the scenario with only HDBs, as shown in Fig. 4.3(b). The ratio decreases to 1.43 when the $\omega_B = 1.0$ [sec]. The $\left(v_s + v_{s_B}\right)/\chi_s$ ratio further decreases to 1.02 when $\omega_B = 0.5$ [sec], as shown in Fig. 4.5(d).

When comparing Fig. 4.5(a), (b), (c), and (d) with each other, particularly focusing on the scenarios with 1500 HDBs (Fig. 4.5(a)) or CABs (Fig. 4.5(b), (c) and (d)) mixed with 1500 HDVs on a link. The $\left(v_s + v_{s_B}\right)/\chi_s$ ratio peaks at 3.25 with the maximum number of vehicles (both HDBs and HDVs) as shown in Fig. 4.5(a). For the other scenarios, which involve the mixed flow of CABs and HDVs under different reaction times as shown in Fig. 4.5(b), (c) and (d), the $\left(v_s + v_{s_B}\right)/\chi_s$ ratio are 2.83, 2.42 and 2.00, respectively.

4.9 Numerical calculations of small network

4.9.1 Definitions of small network

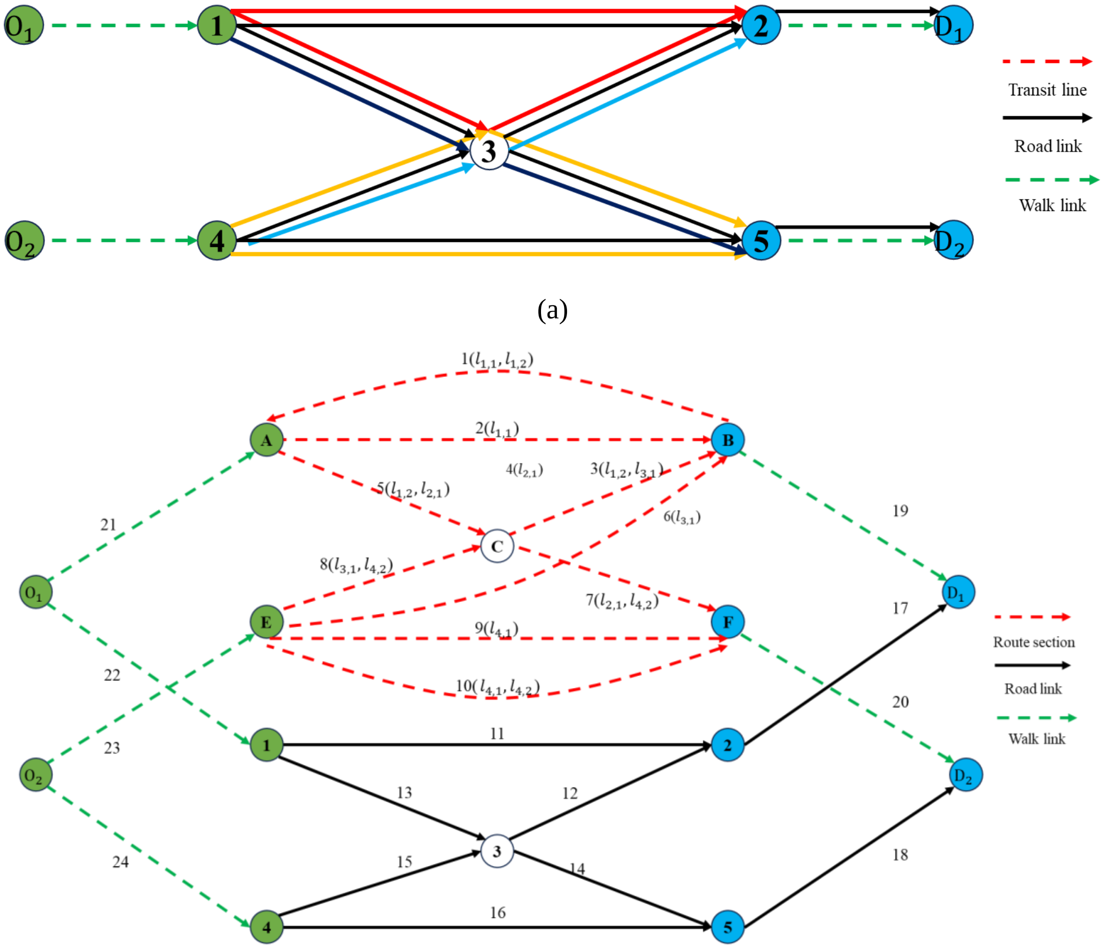
Fig. 4.6 illustrates the primitive network, and its corresponding hyper-network used in the test computation. The primitive network depicted in Fig. 4.6(a) includes 6 bus lines, 8 road links and 4 walking links. The hyper-network in Fig. 4.6(b) consists of 10 route sections for buses, 8 road links for HDBs and 6 walking links. For simplicity, all lines are considered as attractive lines.

The network includes four O-D movements: from origin O_1 to destination D_1, O_1 to D_2, O_2 to D_1, and O_2 to D_2. The passenger demands for O-D pairs $\mathrm{O}_1\mathrm{D}_1$ and $\mathrm{O}_2\mathrm{D}_2$ are set at 2,000 [passengers/h] [passengers/h], while for O-D pairs $\mathrm{O}_1 - \mathrm{D}_2$ and $\mathrm{O}_2 - \mathrm{D}_1$, each demand is 5,000 [passengers/h]. The parameter θ is set to $0.025[min/yen]$, and the fare per frequency increase (F_l) for each bus line l is set as $800{,}000\ [\text{yen/unit period}]$. The vehicle lengths of $\vartheta_H = 5\ [\text{m}]$ and $\vartheta_B = 10\ [\text{m}]$ are assumed to HDVs and CABs, respectively. It is assumed that the reaction time for HDBs, is 2.0 [sec], the same as that for HDVs (McGehee et al., 2000). For CABs, the reaction time is assumed to be 1.5 [sec], considering human intervention (Favarò et al., 2018; Dixit et al., 2016).

Passengers can choose between both transport modes for any O-D pair. Tables 4.2 and 4.3 provide the bus line information and basic link data. The parameters for calculating the link disabilities and line disutility coefficients for each transport mode are listed in Tables 4.3 and 4.5.

It is assumed that passengers choosing CABs need to walk to the bus stop from their origin, and both CABs and HDVs users must walk to their final destination. CABs users must walk from the bus stop to their destination, while HDVs users need to park their vehicles in the parking lot before walking to their destination.

Passengers choosing CABs can transfer to any bus line available at their stop. For example, if a passenger is travelling from O_1 to D_2, the direct bus line is Line $l_{2,1}$. However, passengers can also take bus Line $l_{1,2}$ or Line $l_{2,1}$ first to reach node 3, and then transfer to Line $l_{4,2}$ to reach D_2.



(b)

Fig. 4.6 Network representations of test network, (a) primitive network; (b) hyper-network.

Table 4.2 Candidates for bus lines in the test network.

O-D	Line	Node sequences
	$l_{1,1}$	O_1->1->2-> D_1
O_1-D_1	$l_{1,2}$	O_1->1->3->2-> D_1
O_1-D_2	$l_{2,1}$	O_1->1->3->5-> D_2
O_2-D_1	$l_{3,1}$	O_2->4->3->2-> D_1
	$l_{4,1}$	O_2->4->5-> D_2
O_2-D_2	$l_{4,2}$	O_2->4->3->5-> D_2

Table 4.3 Basic link data information of test network.

Link number	Link type	t_s [min]	p_s [yen]	ω_H [sec]	ω_B [sec]	α_s, β_s
1	Route section		200			
2	Route section		150			
3	Route section		100			
4	Route section		150			
5	Route section		100			
6	Route section		150			
7	Route section		100			
8	Route section		100			
9	Route section		150			
10	Route section		150			
11	Road link	3		2.0	1.5	2, 3
12	Road link	3		2.0	1.5	2, 3
13	Road link	3		2.0	1.5	2, 3
14	Road link	3		2.0	1.5	2, 3
15	Road link	3		2.0	1.5	2, 3
16	Road link	3		2.0	1.5	2, 3
17	Road link	3		2.0	1.5	2, 3
18	Road link	3		2.0	1.5	2, 3
19	Walk link	5				
20	Walk link	5				
21	Walk link	5				
22	Walk link	5				
23	Walk link	5				
24	Walk link	5				

Table 4.4 Value of parameters.

Line	Mode	k_l [passengers/service]	(φ_1, γ_1)	(φ_2, γ_2)
1	CABs/HDBs	50	2, 3	2, 3
2	HDVs	∞		
3	Walking	∞		

Table 4.5 Coefficients of disutility.

Line	Mode	λ_m	π	ρ	τ
1	CABs/HDBs	30	1	1	0.02
2	HDVs	50	1	1	0.02
3	Walking	5	1	1	0.02

4.9.2 Results of small network

Fig. 4.7 shows the results of changes in frequency, and it can be easily seen that the optimal computation is done before 35 iterations.

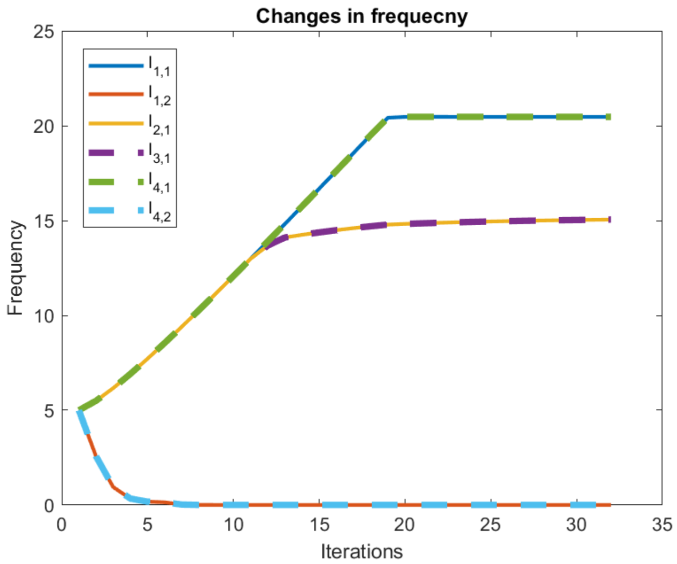


Fig. 4.7 Changes in frequency of test network

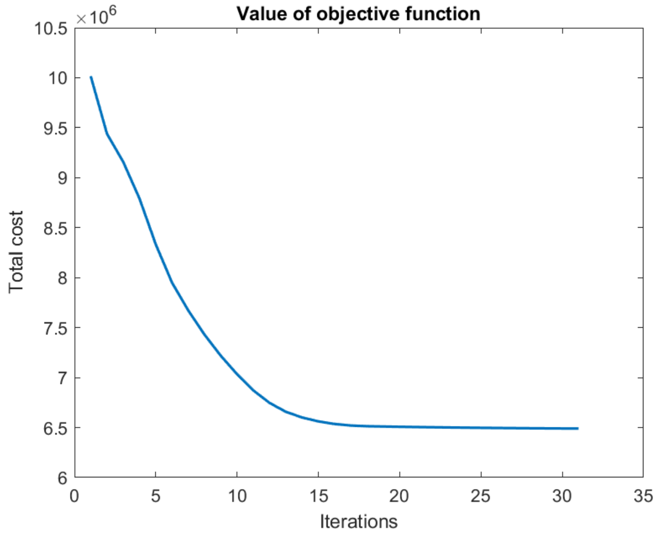


Fig. 4.8 Value of objective function

When the frequencies of all six bus lines are optimized, the resulting solutions are $(f_{1,1}, f_{1,2}, f_{2,1}, f_{3,1}, f_{4,1}, f_{4,2}) = (20.5, 0.0, 15.1, 15.1, 20.5, 0.0)$. The optimal bus route network is comprised of Lines $l_{1,1}$, $l_{2,1}$, $l_{3,1}$ and $l_{4,1}$ as shown in Fig. 4.5, with the remaining frequencies approaching zero. Furthermore, the objective function value is 6.4892×10^6, as shown in Fig. 4.8.

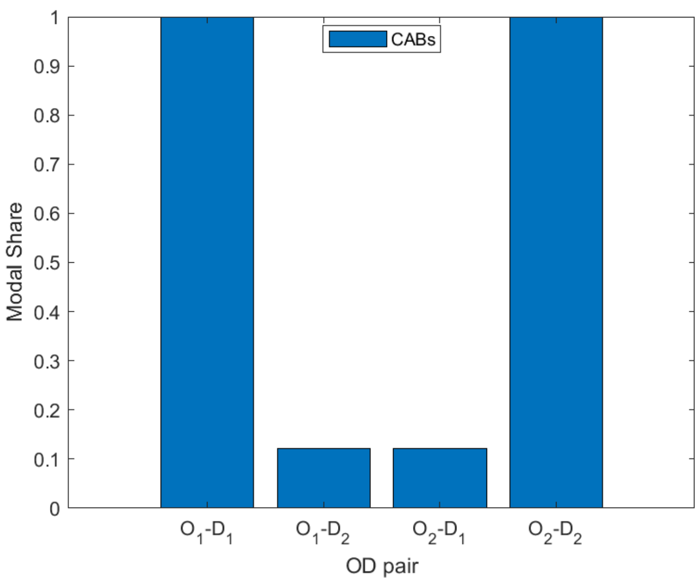


Fig. 4.9 Modal share of CABs for each O-D pair

Fig. 4.9 presents the modal share of CABs for each O-D pair. The modal shares of CABs for the O-D pairs O_1-D_1 and O_2-D_2 are close to 100%, indicating that almost all passengers in these O-D pairs choose CABs. This is due to the direct lines and lack of transfer lines available (e.g., from O_1 to D_1, passengers can take Line $l_{1,1}$ directly to the destination without transfer), making CABs a more attractive option compared to HDVs. Notably, the capacity of a given link is higher when all vehicles are CABs, as opposed to a scenario where all vehicles are HDVs. This is evidenced by Fig. 4.4. This increased capacity results in a reduction in travel disutility for CABs relative to HDVs. In contrast, the O-D pairs O_1-D_2 and O_2-D_1 exhibit significantly lower modal shares for CABs, around 10%, due to the presence of node 3 as a transfer point, reflecting the time cost associated with transferring at this node, as shown in Fig. 4.6(a). Transfers often result in additional waiting time and potential discomfort, this results in very few users willing to choose CABs.

The higher modal shares for CABs in certain O-D pairs suggest that when the transfers are minimal, CABs become a preferred mode of transport. However, if multiple transfers are required, travelers may prefer HDVs, resulting in lower CABs modal share. These findings underscore the importance of optimizing route design, transfer coordination, and service frequency to enhance the attractiveness of CABs in urban transportation networks.

Table 4.6 The results of different initial frequency values.

Initial values	$l_{1,1}$	$l_{1,2}$	$l_{2,1}$	$l_{3,1}$	$l_{4,1}$	$l_{4,2}$
(5, 5, 5, 5, 5, 5)	20.5	0.0	15.2	15.2	20.5	0.0
(6, 7, 8, 9, 4, 3)	20.5	0.0	15.2	15.2	20.5	0.0
(4, 9, 12, 7, 6, 8)	20.5	0.0	15.2	15.2	20.5	0.0

To assess algorithm convergence, various initial values were employed to solve the bus network design problem, as detailed in Table 4.6. The results indicate that different initial values consistently converge to the same value, providing evidence of algorithm convergence. This consistent convergence across diverse initial values suggests that the obtained solution is likely close to the global optimum.

4.10 Numerical calculations of Sioux-falls network

4.10.1 Definitions of Sioux-falls network

The model formulation and solution algorithm were also applied to a bus network design problem in the Sioux-falls network. In our modified Sioux-falls network, 36 paths for HDVs and 18 bus lines are created arbitrarily. The link sequences and stop nodes are detailed in Tables 4.7 and 4.8, respectively. In

Table 4.8, node numbers in parentheses indicate nodes that the bus line passes through without stopping. The initial service frequency for each bus line is assumed to be 5 [services/h]. The O-D demand is set to 3,000 [passengers/h] for each pair, with a fare per frequency increase (F_l) for each bus line l is $1{,}000{,}000$ [yen/unit period] for both HDBs and CABs. Other related settings remain consistent with those in the previous calculation example (Tables 4.4 and 4.5). For numerical calculations, 253 hyper-paths were enumerated using the Yen (1970) method.

The Sioux-falls network uses the same setup as the test network. This setup includes the hypothetical origins where CABs passengers must walk to the bus stop, and both HDVs and CABs users must walk to their destinations.

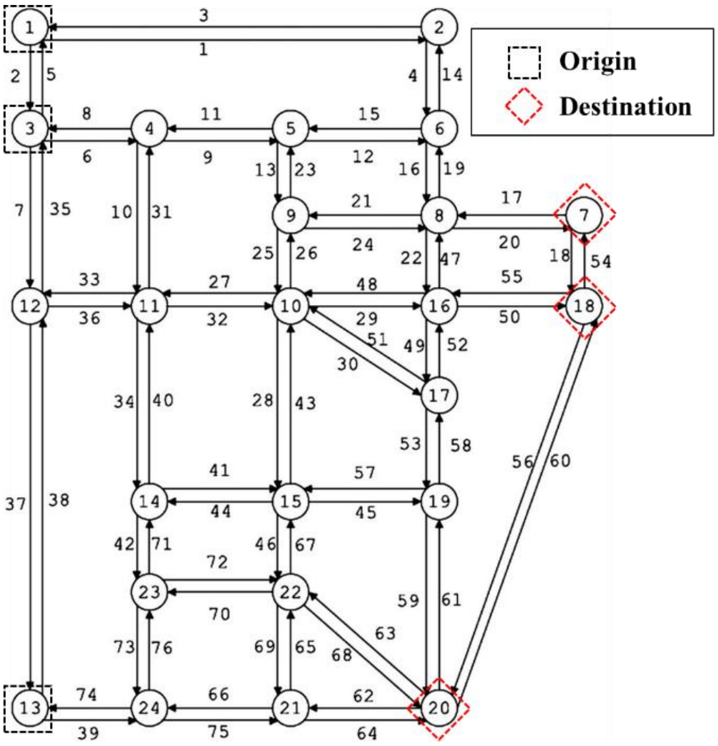


Fig. 4.10 Sioux-falls network.

Table 4.7 Road paths in the Sioux-falls network.

O-D pair	Path	Link sequences	O-D pair	Path	Link sequences
O-D1 (1-7)	1	1-4-16-20	O-D6 (3-20)	21	6-9-12-16-22-49-53 -59
	2	2-6-9-12-16-20		22	6-10-32-30-53-39
	3	2-6-9-13-24-20		23	7-37-39-75-64
	4	1-4-15-13-24-20		24	7-36-34-42-70-68
O-D2 (1-18)	5	1-4-16-20-18	O-D7 (13-7)	25	39-75-64-60-54
	6	1-4-16-22-50		26	38-36-32-29-50-54
	7	2-7-36-32-29-50		27	38-36-32-26-24-20
	8	2-6-9-12-16-22-50		28	39-75-64-61-58-52-47-20
O-D3 (1-20)	9	1-4-16-22-49-53-59	O-D8 (13-18)	29	39-75-64-60
	10	2-7-36-32-29-49-53-59		30	38-36-32-29-50
	11	2-7-37-39-75-64		31	39-75-65-67-43-29-50
	12	2-6-10-32-29-49-53-59		32	39-76-71-40-32-29-50
O-D4 (3-7)	13	6-9-12-16-20	O-D9 (13-20)	33	39-76-72-69-64
	14	6-9-13-24-20		34	39-76-72-68
	15	7-36-32-29-50-54		35	39-75-64
	16	7-36-32-29-47-20		36	39-76-71-41-46-68
O-D5 (3-18)	17	6-9-12-16-20-18			
	18	6-10-32-29-50			
	19	7-36-32-29-50			
	20	6-9-13-25-29-50			

Table 4.8 Candidates for bus lines in the Sioux-falls network.

O-D pair	Line	Stop nodes	O-D pair	Line	Stop nodes
O-D1 (1-7)	1	1-(2)-6-7	O-D6 (3-20)	11	3-12-11-(10)-17-(19)-20
	2	1-(2)-6-5-(9)-(8)-7		12	3-(4)-(5)-6-(8)-16-17-(19)-20
O-D2 (1-18)	3	1-(2)-6-5-(9)-(8)-7-18	O-D7 (13-7)	13	13-12-(11)-10-(9)-(8)-7
	4	1-(3)-(4)-(5)-6-(8)-16-18		14	13-12-(11)-10-16-(18)-7
O-D3 (1-20)	5	1-(2)-(6)-(8)-16-(17)-(19)-20	O-D8 (13-18)	15	13-12-11-10-(16)-18
	6	1-(3)-12-(11)-(10)-16-(17)-(19)-20		16	13-24-(23)-(14)-11-(10)-(16)-18
O-D4 (3-7)	7	3-(4)-5-(6)-(8)-7	O-D9 (13-20)	17	13-24-21-20
	8	3-(4)-5-(9)-8-7		18	13-24-23-22-20
O-D5 (3-18)	9	3-(4)-5-(6)-(8)-7-18			
	10	3-(4)-5-(9)-10-(16)-18			

4.10.2 Results of Sioux-falls network under single class UE

Two scenarios were computed in the Sioux-falls network: 1) a mixed flow of HDBs and HDVs and 2) a mixed flow of CABs and HDVs. The optimal bus route network for each scenario is comprised of Lines 6, 11, and 14, as displayed in Fig. 4.11.

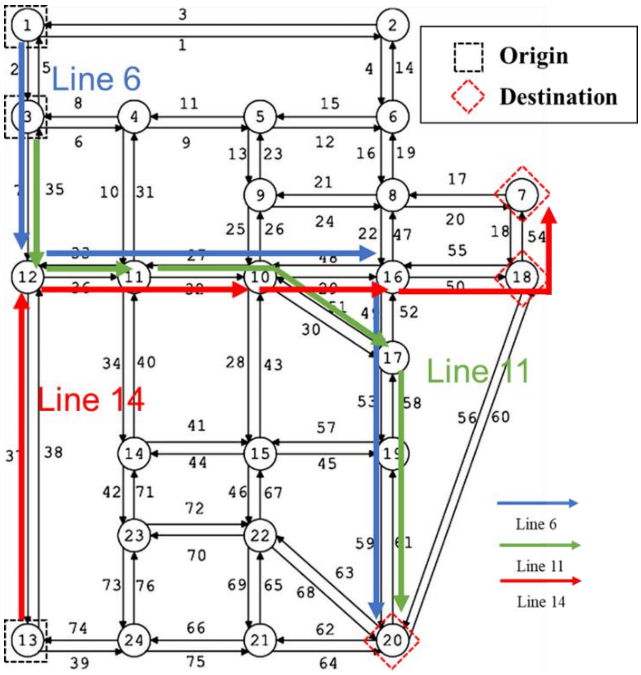


Fig. 4.11 Results of optimal bus route of Sioux-falls network.

Passengers can transfer at any node served by these bus lines. For example, passengers traveling from origin 1 to destination 7 can take Line 6 to node 12. At node 12, they can transfer to either Line 6 or Line 14 to reach node 16 and then continue on Line 14 to arrive at destination 7.

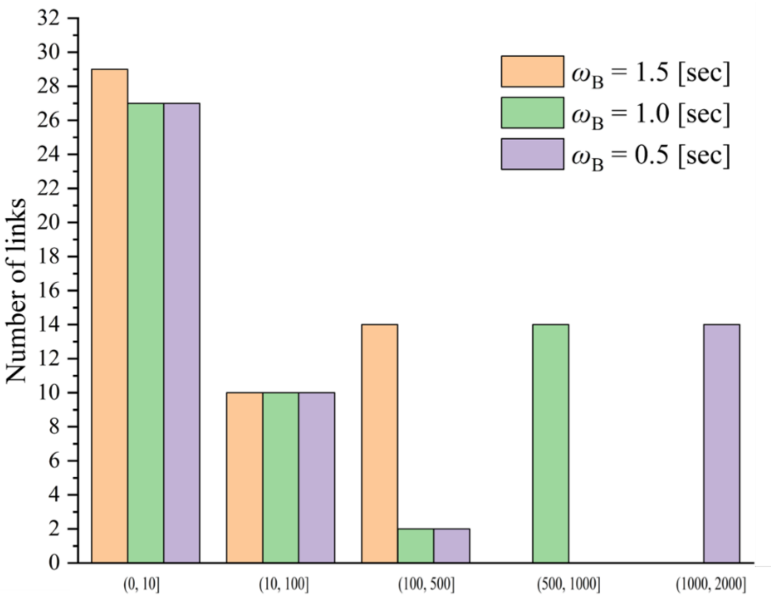


Fig. 4.12 Results of difference in link capacity

Fig. 4.12 presents a comparative analysis of the differences in link capacity between scenarios where CABs are mixed with HDVs and where HDBs are mixed with HDVs. The horizontal axis illustrates the improvement in capacity, i.e., capacity with a mixed flow of CABs and HDVs and minus the one of HDBs and HDVs. Vertical axis values near zero indicate there is no increase in capacity since the link is exclusively occupied by HDVs. As the reaction times of the CABs decrease (from $\omega_B = 1.5$ [sec] to $\omega_B = 0.5$ [sec]), improvement in capacity on certain links become more pronounced.

When the buses are HDBs with the reaction time of $\omega_B = 2.0$ [sec], the total network cost is calculated as 8.6223×10^6. In contrast, the total cost with CABs is 8.5431×10^6 when $\omega_B = 1.5$ [sec]. The total cost further decreases to 8.4673×10^6 when $\omega_B = 1.0$ [sec]. For the scenario where CABs have the shortest reaction time ($\omega_B = 0.5$ [sec]), the total cost is 8.3695×10^6.

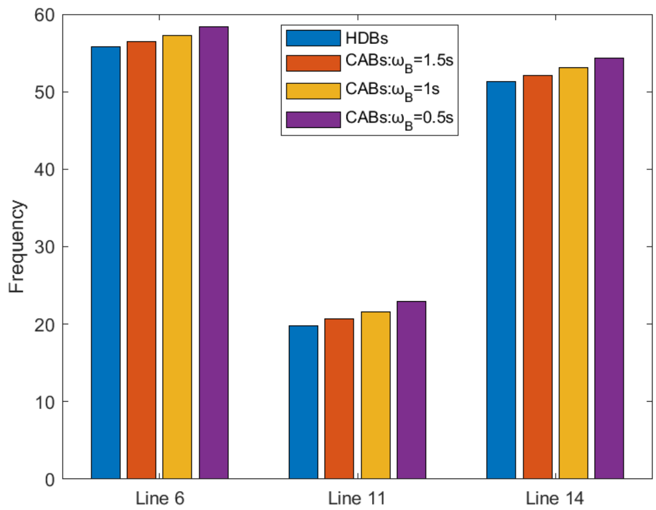


Fig. 4.13 Changes in frequency of Sioux-falls network

Fig. 4.13 shows the frequencies of three bus lines (Lines 6, 11, and 14) under different scenarios: one where HDBs exist in the network and another where replacing HDBs with CABs with different reaction times ($\omega_B = 1.5$ [sec], 1.0 [sec], 0.5 [sec]). The frequency for Line 6 remains consistently high across all scenarios, regardless of whether the network is served by HDBs or CABs with different reaction times. Line 11 shows a lower frequency relative to Lines 6 and 14 across all scenarios. This is because passengers travelling on O-D pair 3-20 can transfer at node 12 and use Line 6 to reach their final destination directly. However, the frequency increases as the reaction time of CABs decreases (ω_B decreases from 1.5 [sec] to 0.5 [sec]) for all lines.

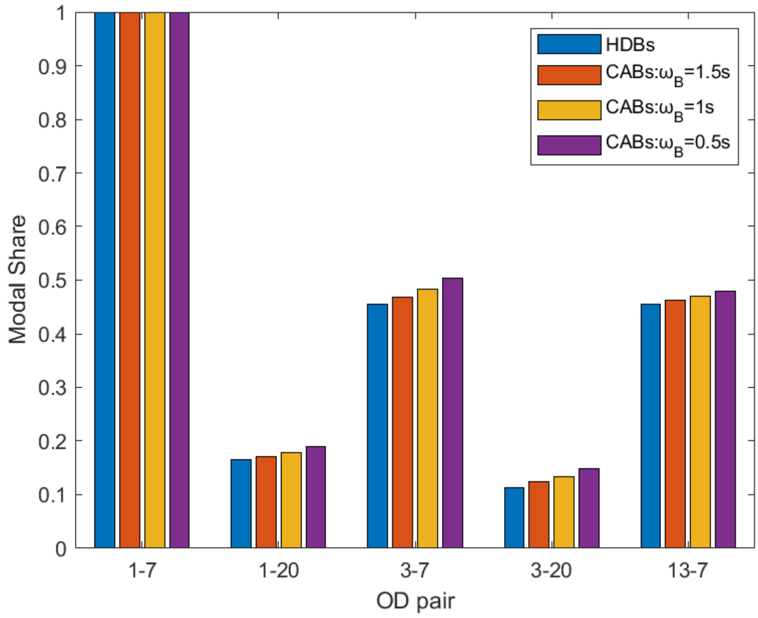


Fig. 4.14 Comparison of modal share for each O-D pair.

Fig. 4.14 shows modal shares across several O-D pairs for different scenarios. For the O-D pair (1-7), available bus routes are comprised of two direct lines which are either Line 6 or Line 11. For all O-D pairs except for O-D pair (1-7), the results show the increases in the modal share for CABs with shorter reaction times. For the O-D pairs (1-18), (3-18), (13-18) and (13-20), the bus was not chosen. Bus routes for these O-D pairs do not provide direct routes or time-efficient travel options. These routes involve detours, excessive stops, or inefficient connections, leading to increasing travel time. As a result, passengers are more likely to choose HDVs for a more direct and quicker travel. This analysis highlights the importance of designing bus networks that minimize transfers across most O-D pairs. CABs have a slightly higher modal share than HDBs, indicating a preference for CAB.

4.10.3 Results of Sioux-falls network under multi-class UE

This chapter presents the numerical analysis of the Sioux-falls network using the MCUE assignment. The setup is the same as the one in Chapter 4.5.1, with modifications only to the demand in the Sioux-falls network. As shown in Fig. 4.15, the entire demand for the O-D pair 1-7 is set as 3,000 [pcu/h]. Assuming that O-D pair 1-7 has been served by CABs only. For the remaining O-D pairs, the demands of CABs and HDVs for each are set as 1,500 [pcu/h].

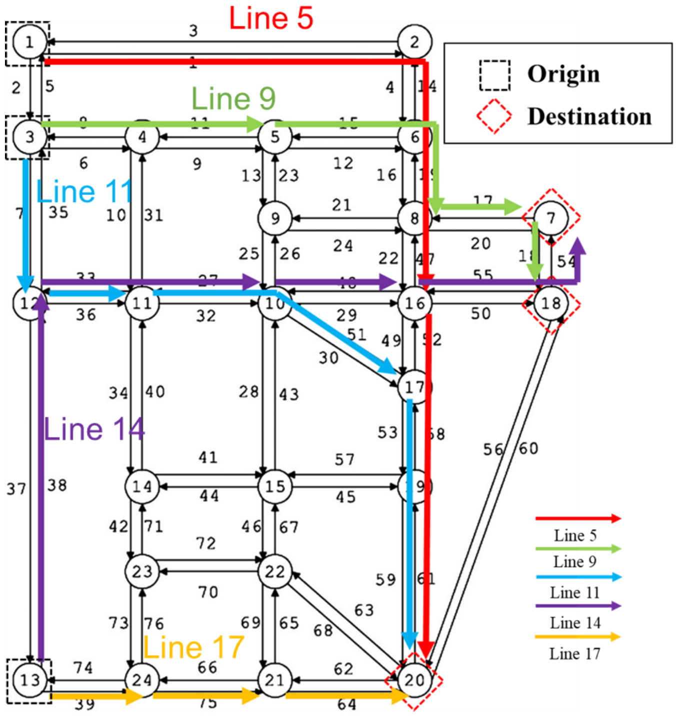
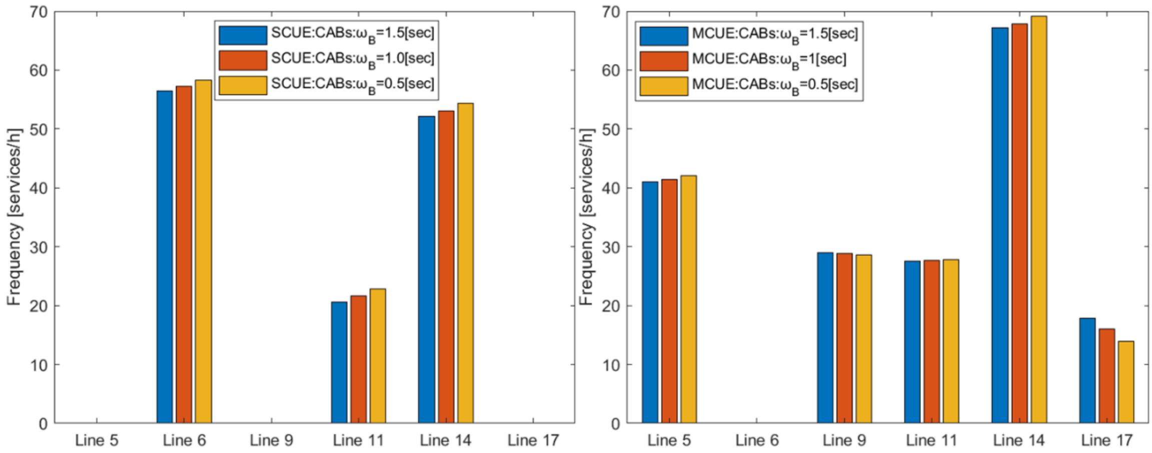


Fig. 4.15 Results of Sioux-falls network by MCUE assignment.

Fig. 4.15 displays the results for the Sioux-falls network calculated by using MCUE assignment. The optimal frequencies for the optimal bus route network is comprised of bus lines $(f_5, f_9, f_{11}, f_{14}, f_{17})$, which remains consistent across scenarios with varying CABs reaction times ($\omega_B = 1.5$ [sec], 1 [sec], 0.5 [sec]). Since several O-D pairs lack direct bus lines, passengers need to use bus lines serving other O-D pairs. For instance, passengers travelling from O-D pair 1-18 start from origin 1 by taking bus Line 5 to reach node 16, transfer to bus Line 11 to reach node 7, and get on bus Line 9 to their final destination which is node 18.

The total cost with CABs under MCUE is 7.8207×10^6, showing an 8.3% reduction compared to the total cost under SCUE at $\omega_B = 1.5$ [sec]. At $\omega_B = 1.0$ [sec], the total cost decreases to 7.7491×10^6, which is an 8.5% reduction compared to the SCUE. Furthermore, at $\omega_B = 0.5$ [sec], the total cost further drops to 7.6649×10^6, indicating an 8.4% reduction compared to the SCUE.



(a) (b)

Fig. 4.16 Changes in frequency under: (a) SCUE assignment; (b) MCUE assignment.

Fig. 4.16 displays results of service frequencies under (a) SCUE assignment and (b) MCUE assignment, considering different CABs reaction times. Lines 5, 9, and 17 are selected but Line 6 was not selected in the scenarios under MCUE . Compared to the SCUE scenario, Lines 5, 9, and 17 are included in the MCUE results, since the MCUE model captures variations in preferences and behaviors across user classes and considers the travel patterns of non-driving groups so that determine the optimal bus route network which can connect all origin and destination nodes.

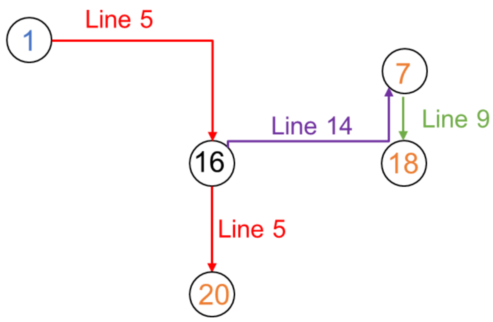


Fig. 4.17 The bus route network for O-D pairs 1-7/18/20

For example, Fig.4.17 shows that the passengers who departure from origin node 1 to destination node 7 can take Line 5 to arrive at node 16 and then transfer to Line 14 to arrive at destination node 7, furthermore, taking Line 9 at node 7 can reach the destination node 18. Passengers can use Line 5 to travel between origin node 1 and destination node 20.

For Line 5, we assume that the O-D pair 1-7 is served exclusively by CABs, this line is necessary to meet the demand. Line 9 provides the possibility for users who travel between O-D pair 3-18 directly.

As the reaction times of CABs decrease (with lower ω_B), the frequencies increase on all bus lines except Line 17.

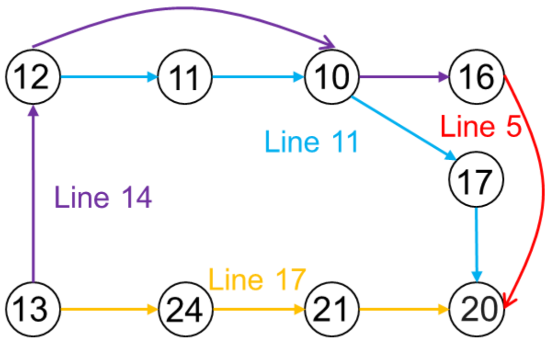


Fig. 4.18 Line 17 serves O-D pair 13-20

This is because Line 17, which serves the O-D pair 13-20, can be effectively replaced by Lines 5 and 14 as shown in Fig.4.18. Consequently, the frequency of Line 17 decreases as the reaction time decreases.

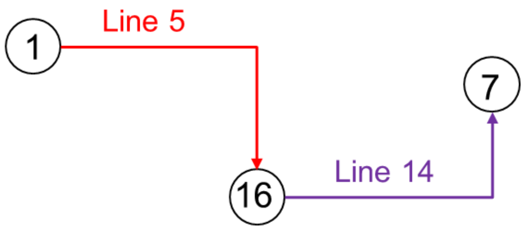


Fig. 4.19 Lines 5 and 14 serve O-D pair 1-7

In the MCUE scenario, the frequencies of bus Line 14 are higher than Lines 5, 9, 11 and 17 with different reaction times. Firstly, O-D pair 1-7 has been served by CABs only, and passengers must use Line 17 to arrive at destination node 7.

4.10.4 Calculation of circular bus routes for Sioux-falls network

A circular bus route, often referred to as a loop or circular route, forms a closed loop, beginning and ending at the same location. The bus runs along circular or elliptical routes, typically serving a sequence of stops without backtracking. Circular bus routes are typically designed to connect key points within a specific area, such as a downtown district or a suburban neighborhood, enabling passengers to travel between these points without requiring transfers. The spanning tree method, as employed in Kim et al. (2017) to address the bus route network design problem, ensures network connectivity while explicitly avoiding the formation of cycles. This feature of the spanning tree method inherently limits its applicability to the design of circular bus routes. In contrast, the method proposed in this study is specifically designed to directly address the circular bus route network problem if necessary, providing a more suitable solution for scenarios requiring circular bus routes.

Next, we will examine the potential of the proposed method to generate circular bus routes. One O-D demand is added to the Sioux-falls network, as shown in Table 4.9. O-D demand for O-D pair 7-13 is set to 5,000 [passengers/h], while the demands for the remaining O-D pairs are set at 3,000 [passengers/h]. The fare per frequency increase (F_l) for each bus line l is $2{,}000{,}000$ [yen/unit period]. The method proposed by Yen (1970) enumerates 4,088 hyper-paths.

Table 4.9 Road routes and candidates for bus lines added to Sioux-falls network.

O-D pair	path	Link sequences	O-D pair	Line	Stop nodes
O-D10 (7-13)	37	18-56-62-66-74	O-D10(7-13)	19	7-(18)-20-(21)-(24)-13
	38	18-55-48-27-33-37			
	39	18-55-48-28-46-69-66-74		20	7-(16)-(10)-(11)-12-13
	40	18-55-49-53-59-62-66-74			

As shown in Fig. 4.20, the calculation results for circular bus routes indicate that the optimal bus routes consist of Lines 4, 5, 9, 11, 14, 18 and 19. Each of these bus lines is strategically designed to cover key demand corridors, crucial for ensuring efficient passenger travel with minimal transfers. Lines 14 and 19, which connect the O-D pair 7-13, form a circular route that significantly enhances overall network connectivity.

The selection of Line 19 rather than Line 20 is thanks to the ability of Line 19 to balance passenger demand across the network, particularly as node 20 serves as both a crucial destination and a major transfer node. Line 19 provides a direct connection between nodes 7 and 20, transforming node 7 into both an origin and destination point and an essential transfer node as well. This allows passengers traveling to node 20 to transfer at node 7. Circular routes are especially effective in urban bus systems because they offer continuous service in densely populated areas, minimizing the need for backtracking and reducing travel times for passengers.

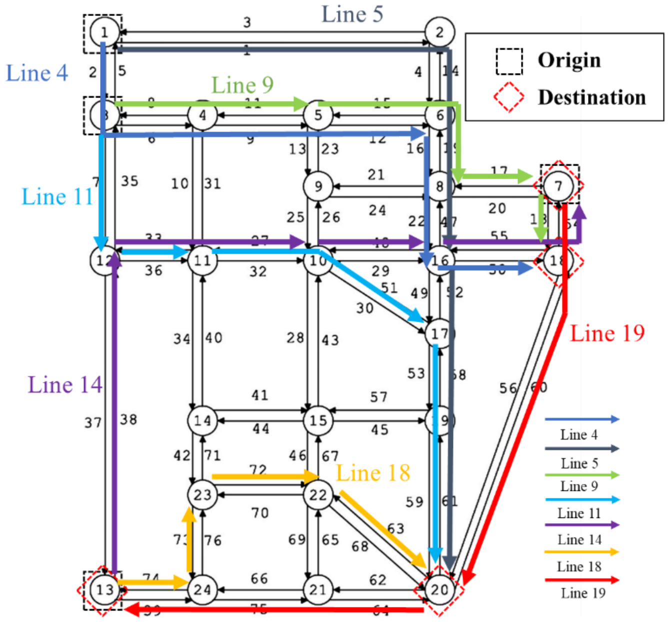


Fig. 4.20 Result of circular bus routes.

The optimal frequencies for the optimal bus route network is comprised of bus lines are $(f_4, f_5, f_9, f_{11}, f_{14}, f_{18}, f_{19}) = (24.6, 39.7, 39.0, 29.5, 58.3, 20.1, 27.9)$, with a total cost of 1.6528×10^7.

4.11 Conclusions of Chapter 4

This study presents a BIPM designed to simultaneously optimize bus route network and service frequencies, accounting for the mixed flow environments of CABs and HDVs. The upper-level problem aims to minimize the total cost of the entire network, encompassing total travel time and bus operation costs, while the lower-level problem addresses the UE problem. The results derived from the proposed model can provide policymakers with guidance for determining the optimal bus routes and their corresponding service frequencies.

Firstly, the impact of penetration rate and reaction times of CABs on capacity is analyzed. A higher penetration rate of CABs with shorter reaction times can increase capacity. Secondly, when there are only HDBs in network, the volume-to-capacity ratio $(v_s + v_{s_B})/\chi_s$ reaches 2.26, indicating severe congestion. In contrast, replacing HDBs with CABs, the ratio decreases to 1.85, 1.43, and 1.02 as reaction times decreases from $\omega_B = 1.5$ [sec] to $\omega_B = 0.5$ [sec]. The numerical analysis confirms that replacing CABs with HDBs can increase capacity and reduce total network costs. For instance, when reaction times decrease from $\omega_B = 2.0$ [sec] for HDBs to $\omega_B = 0.5$ [sec] for CABs, the total cost decreases from 8.6223×10^6 to 8.3695×10^6.

Our findings also underscore the essential role of transfer coordination in maximizing CABs utilization. In cases where bus routes involve multiple transfers, passengers may choose alternative modes, resulting in lower CABs use. These results not only underscore the advantages of CABs but also offer valuable insights for both policymakers and urban planners in designing efficient and user-friendly bus networks.

Future research will focus on developing strategies to minimize passenger transfers and further examine the influence of disutility function coefficients on the uniqueness and stability of solutions. These efforts aim to refine our current model and algorithm, ensuring the established model can support the design of bus networks in the future.

4.12 Appendix

In the computational case presented in this paper, we provide results based on a maximum of one bus line per O-D pair as shown in Equation (4.69). However, in large cities, a single bus line may be insufficient to accommodate the increasing traffic demand. This limitation highlights that having only one bus line per O-D pair may not be adequate to meet the actual demand. Therefore, we extend the method employed in this study to identify up to k bus lines for each O-D pair, ensuring better coverage and capacity to meet the growing demand in key urban areas.

Let $L_{i,k}$ denote a subset of L_i ($L_{i,k} \subset L_i$) with the number of elements of $|L_{i,k}| = k$ where we assume $|L_i| > |L_{i,k}|$. Then we define the following two variables

$$\tilde{f}_{L_{i,k}} = \sum_{l \in L_{i,k}} f_l \tag{4.A1}$$

$$\omega(\tilde{f}_{L_{i,k}}, f_l) = \left(-(\tilde{f}_{L_{i,k}} + f_l) + \sqrt{(\tilde{f}_{L_{i,k}})^2 + f_l^{\,2}}\right)^2 \quad \forall l \in L_i \setminus L_{i,k} \tag{4.A2}$$

We developed a condition that ensures that at most two bus lines can be selected from four candidates available between O-D pair $i \in I$, i.e., $L_i = \{1,2,3,4\}$ and $L_{i,2}$. This condition is written as:

$$\prod_{L_{i,2} \subset L_i} \sum_{l \in L_i \setminus L_{i,2}} \omega(\tilde{f}_{L_{i,2}}, f_l) < \phi \tag{4.A3}$$

Proof:

Left-hand side of above equation is

$$\prod_{L_{i,2} \subset L_i} \sum_{l \in L_i \setminus L_{i,2}} \omega(\tilde{f}_{L_{i,2}}, f_l) = \left(\omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) + \omega(\tilde{f}_{i,\{1,2\}}, f_{i,4})\right) \cdot \left(\omega(\tilde{f}_{i,\{1,3\}}, f_{i,2}) + \omega(\tilde{f}_{i,\{1,3\}}, f_{i,4})\right) \cdot$$
$$\left(\omega(\tilde{f}_{i,\{1,4\}}, f_{i,2}) + \omega(\tilde{f}_{i,\{1,4\}}, f_{i,3})\right) \cdot \left(\omega(\tilde{f}_{i,\{2,3\}}, f_{i,1}) + \omega(\tilde{f}_{i,\{2,3\}}, f_{i,4})\right) \cdot$$
$$\left(\omega(\tilde{f}_{i,\{2,4\}}, f_{i,1}) + \omega(\tilde{f}_{i,\{2,4\}}, f_{i,3})\right) \cdot \left(\omega(\tilde{f}_{i,\{3,4\}}, f_{i,1}) + \omega(\tilde{f}_{i,\{3,4\}}, f_{i,2})\right) < \phi \tag{4.A4}$$

When $f_l = 0$ $(l = 1, \dots, 4)$, (4.A4) equals zero. When $f_1 > 0$ and $f_l = 0$ $(l = 2, \dots, 4)$, (4.A4) equals zero since $\omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) = \omega(\tilde{f}_{i,\{1,2\}}, f_{i,4}) = 0 \Leftrightarrow \omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) + \omega(\tilde{f}_{i,\{1,2\}}, f_{i,4}) = 0$. When $f_l > 0$ $(l = 1,2)$ and $f_k > 0$ $(k = 3,4)$, (4.A4) equals zero since $\omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) = \omega(\tilde{f}_{i,\{1,2\}}, f_{i,4}) = 0 \Leftrightarrow \omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) + \omega(\tilde{f}_{i,\{1,2\}}, f_{i,4}) = 0$. When $f_l > 0$ $(l = 1, \dots, 3)$ and $f_4 = 0$, (4.A4) is greater than zero since $\omega(\tilde{f}_{i,\{1,2\}}, f_{i,3}) > 0$, $\omega(\tilde{f}_{i,\{1,3\}}, f_{i,2}) > 0$, $\omega(\tilde{f}_{i,\{1,4\}}, f_{i,n}) > 0$ $(n = 2,3)$, $\omega(\tilde{f}_{i,\{2,3\}}, f_{i,1}) > 0$, $\omega(\tilde{f}_{i,\{2,4\}}, f_{i,n}) > 0$ $(n = 1,3)$, and $\omega(\tilde{f}_{i,\{3,4\}}, f_{i,n}) > 0$ $(n = 1,2)$. In a similar way, it is shown that when $f_l > 0$ $(l = 1,2,4)$ and $f_3 = 0$, (4.A4) are greater than zero. It is shown that when more than two elements in L_i take positive values, (4.A4) is always greater than zero. This completes the proof.

This constraint can be generalized so that it ensures at most k bus lines can be selected for each O-D pair as

$$\sum_{i \in I} \prod_{L_{i,k} \subset L_i} \sum_{l \in L_i \setminus L_{i,k}} \omega\left(\tilde{f}_{L_{i,k}}, f_l\right) < \phi \qquad (4.\mathrm{A}5)$$

5. Conclusions

The inclusion of UE constraints in NDPs reflects the real-world dynamics of travelers' behavior in transport networks, where travelers adjust their path choices in response to changes in path cost in the network, seeking to minimize their travel costs. The incorporation of UE constraints makes NDPs computationally intensive to solve, as each network adjustment requires reevaluation of the equilibrium state. The NDPs addressed in this thesis focus on two key aspects: traffic state estimation and optimization of both bus route network design and service frequencies. This thesis proposes sensitivity analysis-based algorithms for solving large-scale NDPs while maintaining solution accuracy, since sensitivity analysis can approximate the UE link (or path) flows so that reduce the computational burden.

The first method of this thesis is to develop an efficient algorithm based on the sensitivity analysis for UE flows to solve the traffic states estimation problem by using floating car data. Since the market penetration rate of cars from which floating car data can be obtained is still low in Japan, therefore complete observation data of the entire network is impossible. FIMLE can deal with incomplete data by using all available information, making it robust in scenarios where traditional MLE cannot handle it. Therefore, the traffic states in a road network are obtained by solving a FIMLE problem under UE constraints.

Using the sensitivity analysis-based algorithm to estimate traffic states in the large-scale Asahikawa network in Hokkaido, Japan, the results show that the method proposed in this study estimated relatively reasonable link delay times in the Asahikawa network as summarized in Chapter 3.6. The results of KL divergence show that the proposed algorithm performs well in accurately estimating the link delay times for large-scale networks. However, the accuracy of the estimated link delay times can heavily depend on the locations of the centroids and the number of predefined O-D pairs. We will address such issues in future studies.

The second method is to propose a BIPM to solve an optimization problem of bus route network design and service frequencies, and CABs are introduced as a public transport system. The upper-level problem typically focuses on minimizing the total cost (or disutility), while the lower-level problem captures the user behavior and traffic flow dynamics, modeled through a UE traffic assignment. This optimization problem is subject to UE constraints. An algorithm based on the sensitivity analysis method is developed for solving the optimization problem. This thesis considers mixed flows of CABs and HDVs, therefore traffic capacity of each link in a road network is represented by a function of the penetration rate of CABs on the link.

We analysis the impact of penetration rate of CABs with different reaction times on capacity firstly.

A higher penetration rate of CABs with shorter reaction times can increase capacity as discussed. Secondly, replacing CABs with HDBs can reduce total network costs. For instance, when reaction times decrease from $\omega_B = 2.0$ [sec] for HDBs to $\omega_B = 0.5$ [sec] for CABs, the total cost decreases from 8.6223×10^6 to 8.3695×10^6. The proposed method is specifically designed to directly address the circular bus route network problem if necessary, providing a more suitable solution for scenarios requiring circular bus routes, which can be seen Chapter 4.10.4. Our findings also underscore the essential role of transfer coordination in maximizing CABs utilization. In cases where bus routes involve multiple transfers, passengers may choose alternative modes, resulting in lower CABs use. Future research will focus on developing strategies to minimize passenger transfers and further examine the influence of disutility function coefficients on the uniqueness and stability of solutions. These efforts aim to refine our current model and algorithm, ensuring the established model can support the design of bus networks in future.

As discussed above, this thesis develops efficient solution algorithms based on sensitivity analysis for NDPs with UE constraints. Numerical experiments are conducted on large-scale networks including real-world network (Asahikawa network in Japan) to validate the proposed algorithms. The results confirm the feasibility and effectiveness of the proposed sensitivity analysis-based algorithms in efficiently solving large-scale NDPs under UE constraints.

References

Abdulaal, M. and LeBlanc, L.J. (1979). Continuous equilibrium network design models. Transportation Research Part B 13 (10), 19–32

Afanasiev, L. L. and Liberman, S. Y. (1983). Principles for organizing express bus services. Transportation Research Part A-Policy and Practice, 17, 343-346.

Ban, J.X., Liu, H.X. Ferris M.C. and Bin, R. (2006). A general MPCC model and its solution algorithm for continuous network design problem. Mathematical and Computer Modelling 43 (5–6), 493–505

Bento, L.C., Parafita, R., Santos, S. and Nunes, U. (2013). Intelligent traffic management at intersections: legacy mode for vehicles not equipped with V2V and V2I communications. In: Communications, vol. 2, p. V2I.

Brackstone, M. and McDonald, M. (1999). Car-following: a historical review. Transp. Res. F: Traffic Psychol. Behav. 2 (4), 181–196.

Ceder, A. and Wilson, N. H. M. (1986). Bus network design. Transportation Research Part B-Methodological, 20, 331-344.

Ceder, A. (1984). Bus frequency determination using passenger count data. Transportation Research Part A-Policy and Practice, 18, 439-453.

Ceylan, H. and Bell, MGH.(2004). Sensitivity analysis on stochastic equilibrium transportation networks using genetic algorithm. Journal of Advanced Transportation, 38,291-321.

Chen, J. X., Liu, Z. Y., Wang, S. A. and Chen, X. W. (2018). Continuum approximation modeling of transit network design considering local route service and short-turn strategy. Transportation Research Part E-Logistics and Transportation Review, 119, 165-188.

Chiou, S.W. (2005). Bilevel programming for the continuous transport network design problem. Transportation Research Part B 39 (4), 361–383. Chiou, S.W., 2009. A subgradient optimization model for continuous road network design problem. Applied Mathematical Modelling 33 (3), 1386–1396.

Cho, HJ., Smith, TE. and Friesz, TL. (2000). A reduction method for local sensitivity analysis of network equilibrium arc flows. Transportation Research, 34B, 31-51.

Chu, J. C., Korsesthakarn, K., Hsu, Y. T. and Wu, H. Y. (2019). Models and a solution algorithm for planning transfer synchronization of bus timetables. Transportation Research Part E-Logistics and Transportation Review, 131, 247-266.

Clark, SD. and Watlting DP. (2000). Probit-based sensitivity analysis for general traffic networks.

Transportation Research Record, 1733, 88-95.

Clark, SD. and Watlting DP. (2002). Sensitivity analysis of the probit-based stochastic user equilibrium assignment model. Transportation Research, 36B, 617-635.

Dafermos, SC. and Nagurney, A. (1983). Sensitivity analysis for the asymmetric network equilibrium problem. Mathematical Programming, 28, 174-184.

Daganzo, C.F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp. Res. B:

Meth. 28 (4), 269–287.

Daganzo, C.F. (1995). The cell transmission model, part II: Network traffic. Transp. Res. B: Meth. 29 (2), 79–93.

Daganzo, C.F. (1995). A finite difference approximation of the kinematic wave model of traffic flow. Transp. Res. B: Meth. 29 (4), 261–276.

Davis, GA.(1994). Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. Transportation Research, 28B, 61-75.

De Cea, J. and Fernández, E. (1993). Transit assignment for congested public transport systems: An equilibrium model. Transportation Science, 27, 133-147.

Dresner, K. and Stone, P. (2004). Multiagent traffic management: a reservation-based intersection control mechanism. Proceedings of 3rd International Joint Conference on Autonomous Agents & Multiagent Systems, vol. 2. IEEE Computer Society, pp. 530–537.

Dresner, K.M. and Stone, P. (2007). Sharing the Road: Autonomous Vehicles Meet Human Drivers. In: IJCAI, vol. 7, pp. 1263–1268.

Dixit, V. V., Chand, S. and Nair, D. J. (2016). Autonomous vehicles: Disengagements, accidents and reaction times. PLoS One, 11.

Fajardo, D., Au, T.C., Waller, S.T., Stone, P. and Yang, D. (2011). Automated intersection control. Transp. Res. Rec.: J. Transp. Res. Board 2259 (1), 223–232.

Favarò, F., Eurich, S. and Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. Accident Analysis & Prevention, 110, 136-148.

Fiacco, AV. (1983). Intraduction to sensitivity and stability analysis in nonlinear programming. Academic Press, New York.

Friesz,TL., Tobin, RL., Cho, HJ. and Mehta, NJ. (1990). Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. Mathematical Programming, 48, 265-284.

Florian, M. and Spiess, H. (1983). On binary mode choice/assignment models. Transportation Science, 17, 32-47.

Gao, Z., Sun, H. and Zhang, H. (2007). A globally convergent algorithm for transportation continuous network design problem. Optimization and Engineering 8 (3), 241–257.

Gao, Z. Y., Sun, H. J. and Shan, L. L. (2004). A continuous equilibrium network design model and algorithm for transit systems. Transportation Research Part B-Methodological, 38, 235-250.

Gartner, N.H., Messer, C.J. and Rathi, A.K. (2005). Revised monograph on traffic flow theory: a state-of-the-art report. Special Report by the Transportation Research Board of the National Research Council.

Gkiotsalitis, K. and Alesiani, F. (2019). Robust timetable optimization for bus lines subject to resource and regulatory constraints. Transportation Research Part E-Logistics and Transportation Review, 128, 30-51.

Greenshields, B.D., Channing, W. and Miller, H. (1935). A study of traffic capacity. Highway Research Board Proceedings, vol. 1935. National Research Council (USA), Highway Research Board.

Hall, MA. (1978). Properties of the equilibrium state in transportation networks. Transportation Science, 48,265-284.

Hara, Y., Suzuki, J., and Kuwahara, M. (2018). Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. *Transportation Research Part C*. 86, 622-638.

Hazelton, M.L. (2010). Bayesian inference for network-based models with a linear inverse structure. Transportation Research Part B. 44(5), 674–685

Hatzenbühler, J., Cats, O. and Jenelius, E. (2022). Network design for line-based autonomous bus services. Transportation, 49, 467-502.

Jenelius, E. and Koutsopoulos, H.N. (2013). Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B. 53, 64–81

Jiang, Y. (2022). Reliability-based equitable transit frequency design. Transportmetrica A-Transport Science, 18, 879-909.

Kesting, A., Treiber, M. and Helbing, D. (2010). Enhanced intelligent driver model to assess the impact of driving strategies on traffic capacity. Philos. Trans. Roy. Soc. A: Math. Phys. Eng. Sci. 368 (1928), 4585–4605.

Kim, H.-C., Kim, I. and Hahn, J.-S. (2017). Schematic transit network design using minimum spanning tree algorithm. Journal of the Eastern Asia Society for Transportation Studies, 12, 1299-1310.

Kometani, E. and Sasaki, T. (1959). Dynamic behaviour of traffic with a nonlinear spacing-speed relationship. In: Proceedings of the Symposium on Theory of Traffic Flow, Research Laboratories, General Motors. Elsevier, New York, pp. 105–119.

Kyparisis, J. (1988). Perturbed solutions of variational inequality problem over polyhedral sets. Journal of Optimization Theory and Applications, 57, 295-305.

Kyparisis, J. (1990). Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. Mathematics of Operations Research, 15, 295-305.

Leurent, F. (1988). Sensitivity and error analysis of the dual criteria traffic assignment model. Transportation Research, 32B, 189-204.

Levin, M. W. and Boyles, S. D. (2016). A multiclass cell transmission model for shared human and autonomous vehicle roads. Transportation Research Part C-Emerging Technologies, 62, 103-116.

Levin, Michael W. and Boyles, Stephen D. (2015). Effects of autonomous vehicle ownership on trip, mode, and route choice. In: Publication Pending in Transportation Research Record.

Liu, H.X. and Ma, W. (2009). A virtual vehicle probe model for time dependent travel time estimation on signalized arterials. Transportation Research Part C. 17(1), 11–26

Marcotte, P. (1983). Network optimization with continuous control parameters. Transportation Science 17 (2), 181–197.

Marcotte, P. and Marquis, G. (1992). Efficient implementation of heuristics for the continuous network design problem. Annals of Operations Research 34 (1), 163–176.

Meng, Q., Yang, H. and Bell, M.G.H. (2001). An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. Transportation Research Part B 35 (1), 83–105.

McGehee, D. V., Mazzae, E. N. and Baldwin, G. H. S. (2000). Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track. Proc. Hum. Factors Ergon. Soc. Annu. Meet., 44, 3-320-3-323.

Miller, Joe. (2014). Google Cars ’designed to Speed’ BBC News. BBC, n.d. Web. 25 Aug. 2014.

Moghaddam, S. M. H. M., Rao, K. R., Tiwari, G. and Biyani, P. (2019). Simultaneous bus transit route network and frequency setting search algorithm. J Transp Eng a-Syst, 145.

Nagurney, A. (2000). A multiclass, multicriteria traffic network equilibrium model. Math. Comput. Modell., 32, 393-411.

Newell, G.F. (1993). A simplified theory of kinematic waves in highway traffic, part I: General theory. Transp. Res. B: Meth. 27 (4), 281–287.

Newell, G.F. (2002). A simplified car-following theory: a lower order model. Transp. Res. B: Meth. 36 (3), 195–205.

Qui,Y. and Magnanti, TL.(1989). Sensitivity analysis for variational inequalities defined on polyhedral set. Mathematics of Operations Research, 14, 410-432.

Parry, K. and Hazelton, M.L. (2013). Bayesian inference for day-to-day dynamic traffic models. Transportation Research Part B. 50, 104–115

Patriksson, M. (2004). Sensitivity analysis of traffic equilibria. Transportation Science, 38, 258-281.

Patriksson, M. and Rockafellar, RT. (2003). Sensitivity analysis of aggregated variational inequality problems, with application to traffic equilibria. Transportation Science, 37, 56-68.

Pternea, M., Kepaptsoglou, K. and Karlaftis, M. G. (2015). Sustainable urban transit network design. Transportation Research Part A-Policy and Practice, 77, 276-291.

Ramezani, M. and Geroliminis, N. (2012). On the estimation of arterial route travel time distribution with Markov chains. Transportation Research Part B. 46(10),1576–1590

Ruano-Daza, E., Cobos, C., Torres-Jimenez, J., Mendoza, M. and Paz, A. (2018). A multiobjective bilevel approach based on global-best harmony search for defining optimal routes and frequencies for bus rapid transit systems. Appl. Soft Comput., 67, 567-583.

Sadrani, M., Tirachini, A. and Antoniou, C. (2022). Optimization of service frequency and vehicle size for automated bus systems with crowding externalities and travel time stochasticity. Transportation Research Part C: Emerging Technologies, 143, 103793.

Sheffi, Y. (1984). Urban transportation networks: Equilibrium analysis with mathematical programming methods, Englewood Cliffs, N.J, Prentice-Hall.

Smith, M.J. (1979). The existence, uniqueness and stability of traffic equilibrium. Transportation Research, 13B, 295-304

Suwansirikul, C., Friesz, T.L. and Tobin, R.L. (1987). Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problems. Transportation Science 21 (4), 254–263.

Szeto, W. Y. and Wu, Y. Z. (2011). A simultaneous bus route design and frequency setting problem for tin shui wai, Hong Kong. Eur. J. Oper. Res., 209, 141-155.

Tan, H.N., Gershwin, S.B. and Athans, M. (1979). Hybrid Optimization in Urban Traffic Networks. Report No. DOT-TSCRSPA-79-7, Laboratory for Information and Decision System. MIT Press, Cambridge, MA.

Tian, Q., Wang, D. Z. W. and Lin, Y. H. (2022). Optimal deployment of autonomous buses into a transit service network. Transportation Research Part E: Logistics and Transportation Review, 165, 102865.

Tani, R. and Uchida, K. (2021). An Estimation Method for Spatiotemporal Traffic States Based on Incomplete Traffic Observation. The 8th International Symposium on Dynamic Traffic Assignment.

Tani, R., Owada, T. and Uchida, K. (2018). Path travel time estimating method by incomplete traffic data. International Journal of Intelligent Transportation Systems Research.

Tobin, R. L. & Friesz, T. L. (1988). Sensitivity analysis for equilibrium network flow. Transportation Science, 22, 242-250.

Tuerprasert, K. and Aswakul, C. (2010). Multiclass cell transmission model for heterogeneous mobility in general topology of road network. J. Intell. Transp. Syst. 14 (2), 68–82.

Uchida, K., Sumalee, A., Watling, D. and Connors, R. (2005). Study on optimal frequency design problem for multimodal network using probit-based user equilibrium assignment. Transp. Res. Rec., 1923, 236-245.

Uchida, K., Sumalee, A., Watling, D. and Connors, R. (2007). A study on network design problems for multi-modal networks by probit-based stochastic user equilibrium. Networks Spatial Econ., 7, 213-240.

Wang, J., Peeta, S. and He, X. (2019). Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transportation Research Part B: Methodological, 126, 139-168.

Wei, C. and Asakura, Y. (2013). A Bayesian approach to traffic estimation in stochastic user equilibrium networks. Transportation Research Part C. 36, 446–459

Yang, H., Zhang, X. and Meng, Q. (2004). Modeling private highways in networks with entry-exit based toll charges. Transportation Research Part B 38 (3), 191– 213.

Yang, H. and Michael, GH. Bell. (2009). 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.

Yang, H. and Bell, M.G.H. (1998). Models and algorithms for road network design: a review and some new developments. Transport Reviews 18 (3), 257–278.

Yang, H. and Bell, M.G.H. (2001). Special issue: transportation bilevel programming problems: recent methodological advances. Transportation Research Part B 35 (1), 1–4.

Yang, H. and Bell, M. G. (2007). Sensitivity analysis of network traffic equilibrium revisited: The corrected approach. 4th IMA International Conference on Mathematics in TransportInstitute of Mathematics and its Applications.

Yang, H. (1997). Sensitivity analysis for the elastic demand network equilibrium problem with applications. Transportation Research, 31B,55-70.

Yang, H. and Bell., M. G. H. (1997). Traffic restraint, road pricing and network equilibrium. Transportation Research Part B-Methodological, 31, 303-314.

Yang, H. and Yagar, S. (1994). Traffic assignment and traffic control in general freeway-arterial corridor systems. Transportation Research Part B 28 (6), 463–486.

Yang, H. and Yagar, S., (1995). Traffic assignment and signal control in saturated road networks. Transportation Research Part A 29 (2), 125–139.

Yen, J. Y. (1970). An algorithm for finding shortest routes from all source nodes to a given destination in general networks. Q. Appl. Math., 27, 526-530.

Ying, JQ. (2005). Sensitivity analysis based method for optimal road network pricing. Annals of Operations Research,133, 303-317.

Ying, JQ. and Miyagi, T. (2001). Sensitivity analysis for stochastic user equilibrium network flows- A dual approach. Transportation Science, 35,124-133.

Yperman, I., Logghe, S. and Immers, B. (2005). The link transmission model: an efficient implementation of the kinematic wave theory in traffic networks. In: Proceedings of the 10th EWGT Meeting, Poznan, Poland.

Zhang, W., Jenelius, E. and Badia, H. (2019). Efficiency of semi-autonomous and fully autonomous bus services in trunk-and-branches networks. J. Adv. Transp.

Zheng, F. and Van Zuylen, H. (2013). Urban link travel time estimation based on sparse probe vehicle data. Transportation Research Part C. 31, 145–157

Acknowledgments

Completing this PhD has been a significant challenge and achievement in my academic journey. This process has only been made possible through the encouragement, support, and guidance of numerous individuals. I would like to express my heartfelt gratitude to everyone who has contributed to my academic, personal, and professional development during these past years.

First and foremost, I would like to thank my advisor, Uchida Kenetsu , for his unwavering support, patience, and intellectual guidance throughout my PhD journey. Your keen insights, constructive criticism, and constant encouragement have been invaluable to me. The countless discussions we had, whether in person or over email, have profoundly shaped my thinking and understanding of the field. You pushed me to ask deeper questions and helped me grow as both a researcher and a scholar. For this, I am truly grateful.

I would like to extend my heartfelt thanks to Ryuichi Tani, whose guidance and support have been invaluable throughout my academic journey. Your patience in answering my numerous questions, providing feedback on my work, and offering insightful suggestions has significantly improved the quality of my research. The time and effort you dedicated to reviewing my drafts, clarifying complex concepts, and sharing your expertise have made a profound impact on my understanding of the subject. I am deeply appreciative of your dedication and the encouragement you provided when I needed it most. Thank you for being an essential part of my academic success.

Special thanks to my fellow lab members and colleagues, who have made this journey more enjoyable and less isolating. The camaraderie, support, and intellectual exchange that we have shared in our lab have not only helped me refine my ideas but also given me the emotional strength to push through difficult times. Thank you for the constructive criticism during presentations, the late-night brainstorming sessions, and, perhaps most importantly, the laughter that reminded me to enjoy the process. In particular, I would like to acknowledge (Mr. Yi-fan Wang, Mr. Rong-ze Yu and Mr. Zheng Fan) for their collaboration and friendship over the years.

To my family, I owe the greatest debt of gratitude. My parents have always been my greatest source of strength. From my early days in school to now, your unconditional love and belief in my abilities have driven me forward. You have been my biggest cheerleader, reminding me to stay true to my goals and passions. The sacrifices you made for my education and well-being are immeasurable, and I will forever be grateful.

To my brothers, Mr. Shi-chang Guo and Mr. Shao-zu Sun , thank you for being my source of laughter and sanity throughout this long and winding journey. You've always been there when I needed a break from the academic world, and your lighthearted support has meant so much to me. You've celebrated every milestone with me, no matter how small, and I am incredibly fortunate to have you in my corner.

I would also like to extend my deepest gratitude to my friends who have supported me during this time. (Mr. Zhi-yuan Wang, Mr. Chao Zhang, Mr. Wen-biao Yin, Mr. Yang-bo Zhao and Ms. Lu-meng Jia), thank you for your understanding when I couldn't attend gatherings, for the words of encouragement, and for offering me a space to unwind and recharge. Your support has been a constant reminder that the PhD journey does not have to be taken alone. You were there for me not only during the high points but also when I felt like giving up. Your presence in my life has been a true blessing.

Finally, I would like to acknowledge the financial support I received during my studies. The Rotary Yoneyama Memorial Foundation provided me with the resources necessary to carry out my research, and without their funding, this thesis would not have been possible. I would also like to thank the administrative staff at Hokkaido University (Ms. Azusa Kimura) for her invaluable assistance throughout the years. Your work behind the scenes has enabled me to focus on my research, and I am deeply appreciative of your efforts.

In addition to all the individual people I have mentioned, I am profoundly grateful for the wider academic community in which I have been able to learn, grow, and share my work. Attending conferences, workshops, and seminars has enriched my understanding of the field and allowed me to engage with the brilliant minds who are shaping it. I have benefited immensely from the collective knowledge and passion of this community.

This thesis is the culmination of many years of hard work, sacrifice, and dedication, but it is also the result of the support of so many people who believed in me along the way. I dedicate this work to all of you. Your contributions, both large and small, have made a lasting impact on my life, and I am deeply thankful for your role in helping me reach this milestone.