공학박사 학위청구논문

Modelling Route Choice Behavior Considering Individual Risk Preferences on Travel Time Reliability

통행시간 신뢰도의 개별 리스크 선호도를 고려한

경로선택행태 모형

2018년 2월

서울대학교 대학원

공과대학 건설환경공학부

조 신 형

Modelling Route Choice Behavior Considering Individual Risk Preferences on Travel Time Reliability

지도교수 고 승 영

이 논문을 공학박사 학위논문으로 제출함.
2017년 11월

서울대학교 대학원
공과대학 건설환경공학부
조 신 형

조신형의 박사학위논문을 인준함.
2017년 12월

위 원 장 이 청 원 (인)

부위원장 고 승 영 (인)

위 원 김 동 규 (인)

위 원 추 상 호 (인)

위 원 이 장 호 (인)

Abstract

The route choice problem is an important factor in traffic operation and transportation planning. There have been many studies to analyze the route choice behavior using travel data. There was a limit in constructing a route choice model by generating an appropriate choice set due to the limitations of the data.

In this study, we construct a choice set generation model and a stochastic route choice model using the observed data. This study estimates the parameters incorporating traveler’s heterogeneity according to the choice set from the choice set generation model and the route choice model.

We define the individual confidence level according to perceived travel time distribution to reflect traveler’s heterogeneity on choice set generation model. In addition, the parameters were estimated using the mixed path-size correction logit model (MPSCL) considering the path overlapping and individual risk preference in the route choice model.

We compare the experienced paths and the derived choice set to construct choice set generation model. In addition, it is possible to estimate better parameters incorporating traveler’s heterogeneity for choice set generation model and route choice model. We compare the choice set from the developed model with that of the conventional choice set generation model. This study shows the superior prediction result in route choice model

reflecting the individual behaviors of the route choice in the urban area on the transportation demand forecasting and traffic operation.

Keywords: Choice set generation model, Route choice model, Traveler heterogeneity, Travel time reliability, Travel time budget, Risk preference

Student Number: 2011-21003

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Backgrounds

Travelers tend to travel for faster, less costly, and more comfortable to the destination. They acquire information about the route through various media and chooses a route to travel through the perception of the individual. Travelers carry out the process of comparing and evaluating the alternatives considered in the choice. There are various factors considered in the route choice process, such as travel time, network type, and information media availability. Transportation Planners attempted to formulate these route choice behaviors and build the model more realistically. Various models have been devised to find mathematical solutions to the route choice problem (Bruynoughe, 1968; Dial, 1971; Fsk, 1980; Chen and Alfa, 1991).

Currently, the development of Intelligent Transportation System (ITS) makes it possible to collect and process various information from the perspective of traffic management. Transportation operators employed the information accumulated over many years as new information. Travelers travel with more information about the routes from the various types of travel information. Moreover, travelers tend to change their routes using the various information. The accumulated enormous data makes transportation operators analyze the travel behaviors apart from the mathematical solution. A large amount of data has the various characteristics of travel pattern.

However, travel data was employed to improve the state of traffic flow in the traffic management. There was an attempt to interpret the distribution of travel time in traffic flow (Zhan et al., 2013). The following figure describes the distribution of travel time used in traffic management and uses it as a traffic measure. In this study, we defined the travel time budget (TTB) applied in transportation planning, of which is the same meaning of planning time used in traffic management.

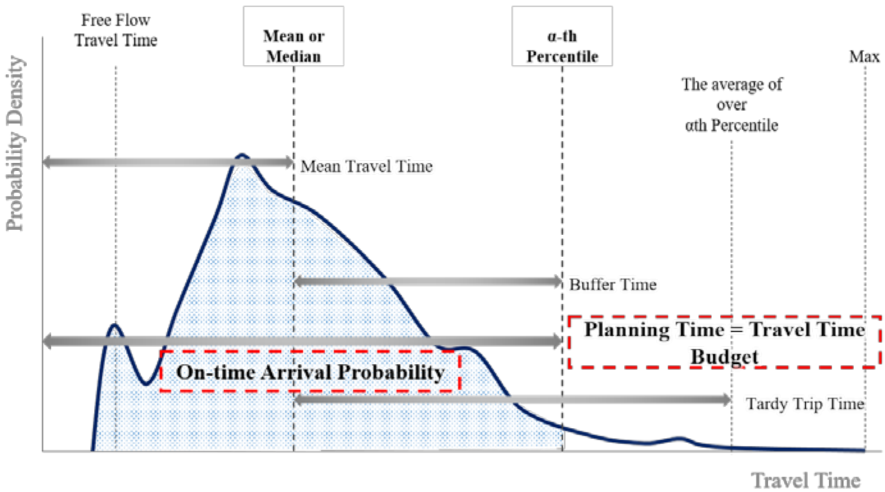


Figure 1.1 Concept of travel time distribution

The route choice model consists of two stages. First, there is a step of constructing a choice set reflecting an individual's travel behavior. This step is modeling of the individual's perceived travel time. The model derives all the possible paths (Universal Set) for a specific OD. We derive a consideration set using the number of observed and experienced paths from the universal set. This set of steps leads to consist of a set of paths. The path set generation process is a model of the path cognitive process for travelers, and the correct number of paths must be selected to increase the accuracy of the choice model estimation process. This process has the challenge of getting the perceived travel time for individuals. There are the different considered set of paths for individual travelers because the travelers have the different experience and attributes for choosing paths. Therefore, it needs to derive a consideration set for individuals to construct the route choice model.

Next, we constructed a stochastic route choice model to derive the utility function by estimating the choice probability from the determined choice set. This process makes it possible to calculate the choice probability for each path from the model. In addition, after estimated the parameters based on the utility function, we validated the developed models using test of model significance. However, there are various alternatives according to the individual experiences and the stochastic travel time in route choice model. Even though it is an effective route in the network, travelers may not choose the route due to their experiences. This problem comes out from the limited information for the alternative routes. Thus, the previous models based on the rationality are somewhat limited to explain this phenomenon. Recently, many studies have carried out to reflect these personal characteristics (i.e., prospect theory, bounded rationality, choice inertia, Etc.)

1.2 Research Purpose

The purpose of this study is to construct a stochastic route choice model using the perceived travel time distribution reflecting individual experiences and characteristics. We defined the perceived travel time for individual travelers as a cumulative distribution function resulting from experienced travel time. The perception of the individual is an important part of the choice set generation model. Recently, there are studies based on the uncertainty of the travel time (i.e., Travel time budget (TTB) model, Percentile user equilibrium (PUE) model, Mean-excess traffic equilibrium (METE) models).

In this study, we estimate the distribution type of accumulated travel time from repeated travel for travelers. It makes possible to define the travel time reliability in stochastic travel time. In addition, we define an individual risk preference and travel time budget (TTB). We construct the route choice model using the determined choice set considering the individual travel behavior.

This study proposes a combined model of the route choice set generation model and the route choice model considering the travel time reliability. We derive the travel time budget (TTB) and the individual confidence level using the cumulative travel time distribution. This study constructs a choice set generation model through Heterogeneous K-α-Reliable Shortest Path Searching method using the individual confidence level derived by individual travel time reliability. We construct a stochastic route choice model incorporating traveler’s heterogeneity for individuals.

1.3 Main contents

There are some important issues to construct a route choice model based on choice set generation incorporating traveler's heterogeneity. It is important about how to set the distribution of travel time and how to distinguish different characteristics for individual travelers. In the previous studies, researchers defined travel time distribution in various types of distribution.

In addition, it is necessary to construct a model reflecting the individual's perceived characteristics on the travel time to reflect the traveler's heterogeneity. We employ the concept of risk preference to reflect these characteristics. It is necessary to generate the route choice set using the travel time reliability and the risk preference for individuals. Finally, it is important to construct a combined model between the generated path set and the route choice model.

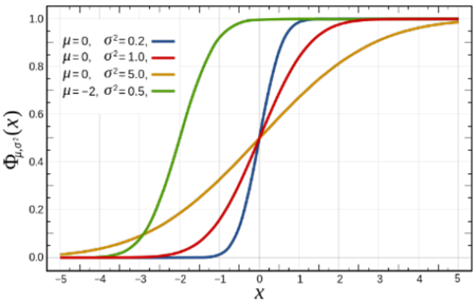
This study mainly deals with the following points.

- Definition of the cumulative distribution function for route travel time reliability → Lognormal distribution.
- Establishment of traveler heterogeneity through identification of traveler's risk preference
- Heterogeneous K-α-Reliable Shortest Path Searching (HKαRSP) based on cumulative distribution function
- Combined model of choice set generation model and stochastic route choice model reflecting travel time reliability

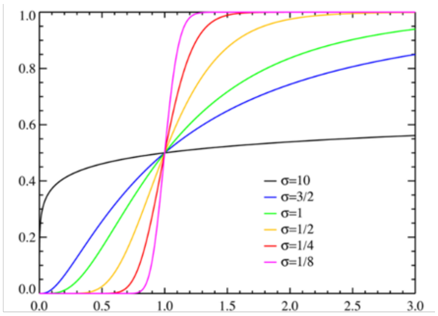
The contents of the previous studies and the parts to be reflected in this study will be summarized.

- **The distribution function of route travel time**

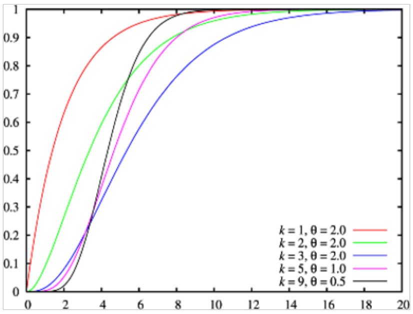
Previous studies have conducted to estimate the distribution of travel time using observed travel time data (Susilawati, 2013). Researchers define the variation of travel time using the cumulative distribution function of travel time and try to solve the problem of travel time estimation using the on-time arrival probability (Chen et al., 2013; Wu and Nie, 2011).



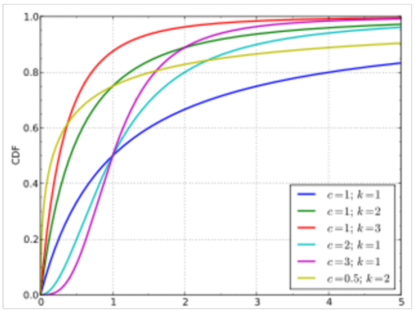
<Normal Dist.>



<Log-Normal Dist.>



<Gamma Dist.>



<Weibull Dist.>

Figure 1.2 Kinds of distribution for travel time

Source: Wikipedia, https://en.wikipedia.org/, accessed on 01.Nov.2017

- **Establishment of traveler group for traveler's risk preference**

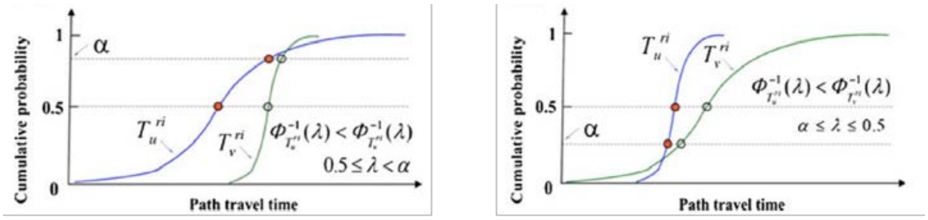
Classifying different characteristics incorporating the traveler's heterogeneity is necessary. We define the risk preferences of travelers according to the distribution of perceived travel time for individuals. The concept of individual confidence level is necessary to classify the travelers. Because of the differences in experienced travel time for individuals, travelers perceive the different travel time. This characteristic makes possible to incorporate the traveler's heterogeneity in route choice model. The on-time arrival probability (α) has the same meaning of the individual confidence level α_l, which means that the individual wants to reach the destination with the level of confidence.

- $\alpha > 0.5$, Risk Averse(avoiding) for on-time arrival
- $\alpha = 0.5$, Risk Neutral for on-time arrival
- $\alpha < 0.5$, Risk Seeking(loving) for on-time arrival

The individual confidence level α_l can be decided by the traveler's perceived travel time for a specific OD pair

- **Heterogeneous K-α-Reliable Shortest Path Searching (HKαRSP) based on risk preference**

The different risk preferences should be defined to generate the different choice set according to the individual confidence level α_l. Researchers tried to generate a path set using the relationship between travel time distribution for each path and risk preference for individual travelers (Chen and Lam, 2013; Chen et al., 2013).



<Risk averse property> <Risk seeking property>

Figure 1.3 The relationship between path travel time and risk preference

Source: Chen et al. (2013), p. 131, Fig. 3

- **Combined model of choice set generation model and stochastic route choice model**

There are some studies to construct route choice model reflecting the traveler's heterogeneity. In addition, some researchers are using revealed preference (RP) data. They analyze the route choice behavior through the questionnaire survey. In addition, studies performed the simulation analysis to generate probabilistic paths due to the lack of actual travel data. Studies may not be possible to reflect the individual characteristics in route choice behavior. In this study, we construct a choice set generation model and stochastic route choice model using observed path travel time data.

1.4 Research Scope

In section 2, we focus on review of choice set generation model and the route choice model and compare the studies analyzing the travel time reliability. We define the terms to establish the models in section 3. In addition, we establish the methodology of choice set generation model and route choice model. Section 4 introduces the revealed preference data used in this study and presents a methodology for processing the route travel data. In section 5, we employ the data to estimate and verify the choice set generation model and the route choice model. This study presents the results by analyzing the choice set generation model and the route choice model empirically by reflecting the travel time reliability. In section 6, we conclude this study with conclusions and discuss the future research.

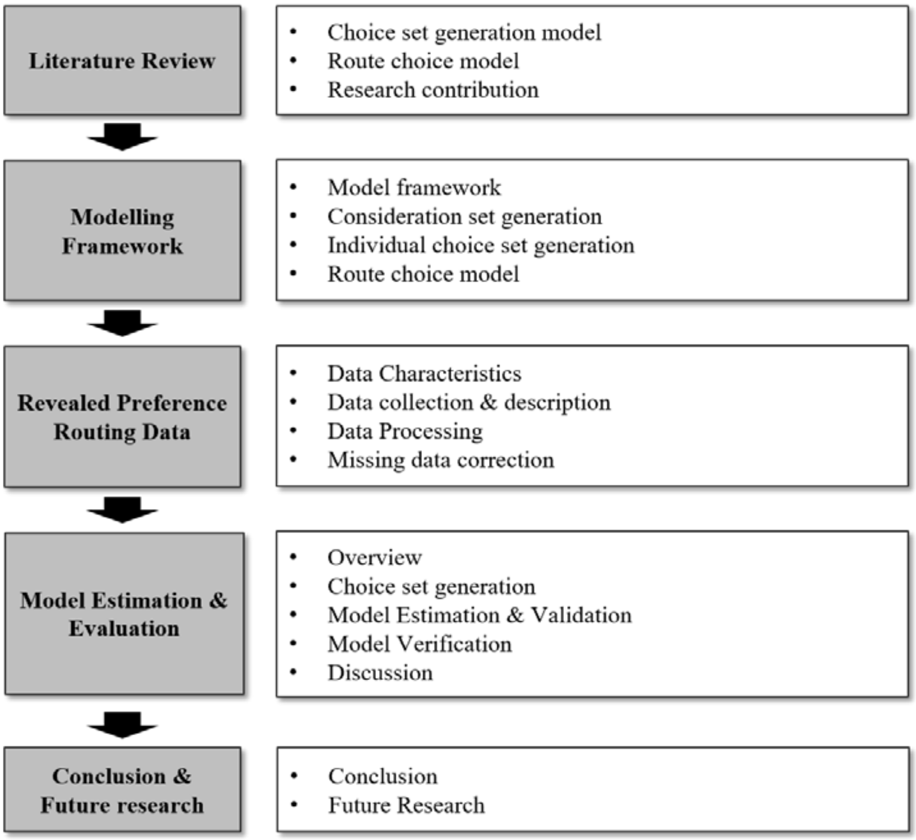


Figure 1.4 Research Procedure

Chapter 2. Literature Review

It is necessary to consider the various developed models to construct route choice model. The existing models mainly deal with the route choice behavior assuming rational behavior without reflecting the cognitive behavior characteristics of travelers. Recent researches suggest the improved models considering traveler's behavioral and network characteristics.

In this study, we set up a model suitable for the analysis through comparison between the models and compare the changes of the route choice probability using the proposed model to derive implications.

The path set generation model is a problem before solving the route choice model and the path assignment model. Researchers have devised the various methods for solving the path set generation problem, and many studies have been carried out for generating the more efficient and highly accurate path set. In addition, studies have been conducted to estimate the route choice probability for travelers using questionnaire survey or revealed travel data. Various types of route choice model have been proposed to improve the existing multinomial logit model (MNL). Researchers studied the improved algorithm for the efficient path-based assignment.

In this study, we construct a path set generation model reflecting the perception of paths for travelers and the travel time reliability. We compared the differences between the existing path set generation model and the proposed model.

In addition, we present a model incorporating the traveler heterogeneity apart from the rationality in the econometric decision. In this regard, we examine the various studies for path set generation model and compare the characteristics of the risk preferences. This study shows the differences for the individual path set generation by comparing the cumulative distribution function reflecting the travel time reliability and the risk preference.

We compare the route choice models for transportation planning. This study analyzes the models with various structural characteristics to verify the change of route choice probability by the individual perception and to apply the generated path set in various route choice model. We use the revealed preference data to show the difference in the actual route choice probability.

2.1 Choice Set Generation

The route choice model has different characteristics from the choice model in transportation analysis (i.e., destination, mode, facility, Etc.). The existing choice model has a characteristic that the set of paths is limited and classified according to the limited budget and purpose. On the other hand, the route choice problem exponentially increases the alternatives according to the number of nodes and links, and the correlation is high.

When travelers consider their route alternatives, there are two approaches in travel choice model. The first is to determine all possible alternatives between OD pairs, and the other is to choose geographically considerable routes among the perceived alternatives. Researchers have tried to devise the method of determining the choice set not only in the transportation planning but also in the transportation operation. Many researchers suggest the methods for choosing the efficient paths (Dial, 1970; Richardson, 1982; Ben-Akiva et al., 1984; De La Barra et al., 1993; Horowitz and Louviere, 1995; Bliemer and Taale, 2006).

The algorithm for searching the paths can be classified into two categories. First, there is a method for finding a mathematical solution for a deterministic method using an efficient algorithm. This method employs the optimization algorithm to find the shortest path most efficiently using deterministic variables (Pollack, 1961; Yen, 1971; Shier, 1979; Martins, 1984; Azevedo et al., 1993). There is a Dijkstra's algorithm for solving the optimization problem, which is a representative algorithm for finding the shortest path using label-setting and label-correcting methods.

Next, there is the heuristic method divided into three categories. First, it is a process of searching the shortest path repeatedly by removing the shortest path after searching. Second, it is a method to search for other shortest paths by giving a redundant penalty on the shortest path. Third, it is a method of branching out from a specific node using the detected shortest path. Recently, researchers have a concern for the development of K-reliable shortest path algorithm considering stochastic traffic flow condition (Ji, 2005; Peer et al., 2007; Hutson and Shier, 2009; Nie and Wu, 2009; Chen et al., 2010; Chen et al., 2016; Zhang et al., 2017).

2.1.1 Deterministic Methods

2.1.1.1 K-Shortest Paths Method

The K-shortest path algorithm uses the shortest path algorithm to generate several shortest paths repeatedly. This approach is similar to the process of finding the shortest path for travelers. It is also categorized into the path searching algorithm allowing loops and not allowing loops.

Lawler's algorithm defines all the paths of the network and repeats the process of finding the shortest paths in all the sets. It is a process searching for the existence of the second shortest path in the remaining subset except for the shortest path.

2.1.1.2 Constrained K-Shortest Paths Method

It is a method to search K efficient paths by generating a large number of alternative paths using the existing K-shortest path algorithm and applying certain constraints. Researchers propose this algorithm for improving the Lawler's algorithm. It generates mutually exclusive subsets in the domain of possible solution, develops them in the form of a tree shape, and find K-shortest paths according to certain constraints using brute force (Van Der Zijpp, N.J. and Fiorenzo-Catalano, S., 2005).

2.1.1.3 Constrained Enumeration Method

The constrained enumeration method is a process of generating a set of paths through the constraints of the traveler's cognitive and behavioral perspectives. This method includes the generating process with a specific criterion and maintains an appropriate set of paths. The model calculates the function by setting the common factors among alternatives, such as distances, time and volume. The parameters of the resistance function utilize the information of the alternatives collected through the investigation. The accuracy of estimated paths is approximately 60% in this model from the observed paths (Pillat et al., 2011).

2.1.1.4 Link Elimination Method

The link elimination method is an algorithm of iteratively searching for the shortest path and performing a process of searching for the next shortest path while removing all links or some links. One method is to find the next shortest path by removing all links used in the shortest path searching (Azvedo et al., 1993). The link elimination method may not generate the paths between ODs by disconnecting the centroid connector or the main link. To improve such a method, researchers proposed the method of eliminating one link, but the computation time is increased due to the increasing number of alternatives rapidly.

2.1.1.5 Link Penalty Method

The link penalty method is the process of finding the shortest path by increasing the resistance on the link instead of removing the link. This method solves the problem of the loss of the previous shortest path caused by breaking the link. An analysis performed the repeated process of searching for the shortest path by increasing the rate of resistance (De la Barra et al., 1993). Other studies attempted to eliminate the various paths that result from minor changes by not increasing the resistance near the origin and destination instead of increasing the rate of resistance (Park and Rilett, 1997). Scott et al. (1997) studied to solve the method of measuring the magnitude of the link resistance by the number of given links by optimization method.

2.1.1.6 Labelling Method

Ben-akiva et al. (1984) defined the path choice set for consisting of "labeled" paths and constructed the path set generation model using the reference function. The different travelers make various choice sets from each point of view. Researchers have constructed a model that divides the paths into ten types and generates the paths using various path attributes. It is composed of various traffic factors such as travel time, distance, safety, and traffic signal. They defined the characteristics of each type of path through the label resistance function. They demonstrated that the set of various paths have the characteristics similar to the set of paths considered by the driver.

The following table summarizes the label and resistance functions presented in the study.

Table 2.1 Path labels and impedances

Criterion	Attribute	Link Impedance
Min. Time	Travel time on a link	Time
Min. Distance	Link length	Distance
Max. Scenic	% of link length through scenic areas	Time(1+β_1% non-scenic)
Min. Traffic lights	Number of traffic lights on link	Time + β_2# of lights
Min. Congestion	Volume/Capacity ratio	Time(1+β_3 high V/C dummy)
Max. Highways	% of link length on highway	Time(1+β_4% non-highway)
Max. Capacity	% of link length of high capacity road	Time(1+β_5 low capacity dummy)
Max. Commercial	% of link length through commercial areas	Time(1+β_6 % low commercial)
Max. Road quality	Dummy variable for quality of pavement	Time(1+β_7 low-quality dummy)
Hierarchical travel pattern	Link “level” in network (3 levels considered)	Time(1+β_{81}(level 1 dummy) + β_{82}(level 2 dummy))

Source: Ben-Akiva et al.(1984) Table 2. Labels and Impedance

Ramming (2002) included the additional six attributes to the existing ten labels, the total for 16 labels. The researcher derived the 188 observations from 91 OD pairs. This study used the internet-based survey data to construct the route choice model in Boston area.

Bekhor and Prato (2009) employed the data from Boston area and Turin area to classify those paths into five labels. The labeling method reflects the perception of the paths for travelers. They tried to grasp the degree of path generation in overlapping.

2.1.1.7 Branch and bound method

Researchers have developed an algorithm for generating the path set using the branch and bound method, which is one of the algorithms for searching the optimal paths in an urban network. The branch and bound method is a technique of finding paths by forming a tree as if it extends to the end node from the origin node. Prato and Bekhor (2006) constructed a process of finding shortest path by defining a link, looking for a segment of a path, and arranging a tree in an array form. They analyzed the process of finding the paths using the tree depth. The model explored the optimal path using directional, temporal, loop, similarity, and movement constraints. They also verified the accuracy of derived paths in comparison of labeling, link elimination, link penalty, and two simulation approaches.

2.1.2 Heuristic Methods

2.1.2.1 K-Dissimilar Paths Method

The k-dissimilar paths method is described in Akgűn et al. (2000) based on the k-shortest path algorithm. It is a process of extracting an optimal path by comparing K paths with a certain minimum value criterion. The method searches the optimal path repeatedly using the shortest path algorithm. The method can flexibly reflect changes of the network and extract a set of spatially constrained paths using criterion. The path does not derive any loop shape. Since the model generates the paths using the dissimilarity of the path, there is little possibility for deriving a similar path. On the other hand, when it needs to generate a specific alternative path, the model cannot derive the corresponding path.

2.1.2.2 Gateway Method

Akgün et al. (2000) devised a gateway method to improve the disadvantage of the dissimilar method. This method is the process of finding the gateway shortest path for a specific OD pair using the spatial variation of the shortest path. This method is possible to generate a path going through a specific node or link so that the model can derive the desired path.

This method set up a gateway node in the middle, and it makes a path from the origin node to the gateway node and a path from the gateway node to the destination node. The model combined the generated paths for one path in OD pair. Moreover, this method removes irrational paths using constraints.

Since the set of paths is repeatedly generated based on the gateway node, it is possible to generate various sets of paths. On the other hand, this method may generate a loop path and cannot distinguish several similar paths.

2.1.2.3 Essentially Least-cost Paths Method

Hunt and Kornhauser (1996) developed an applicable algorithm to generate the set of paths in the regional area. They constructed the shortest path algorithm using cost threshold and derived the different k essentially least-cost paths.

The model derives the minimum path set using a combination of paths reflecting the constraints. It generates the total set of paths according to given constraints and costs. This method generates the reasonable set of paths and derives a path applicable to a certain level of the cost threshold. It applies the method of eliminating similar paths through comparison between overlapping and spatial variating paths. On the other hand, it is difficult to apply to a large network, and it is not possible to generate a specific alternative path.

2.1.2.4 Monte Carlo Method

Sheffi, Y. and Powell, W. B. (1982) constructed a model to generate heuristic paths using Monte Carlo simulation. The Monte Carlo simulation is a method for finding a path set using a shortest path searching method by repeatedly generating link properties. The solution of the path set is generated on the maximum iterations. This method is applied for solving the traffic assignment problem using the polynomial probit model. The model generates the actual link travel time using the distribution through repetitive implementation. The method of traffic assignment is an all-or-nothing assignment. Finally, the derived value is the average value of the traffic volume obtained from the iterative generating process. It is a reasonable method for building choice set generation model. The travelers compare a sufficient number of paths to choose an acceptable path set.

Ramming, M.S. (2002) derived the utility function of the path using Monte Carlo simulation. The accuracy of generated paths is approximately 80% from the observed paths, and it takes the computation time for 20 hours to find feasible path set. This method makes the result not only single OD pair but also multiple OD pairs.

2.1.2.5 Monte Carlo Labelling Combination Method

This method randomly estimates travel time using the Monte Carlo method, which is a method of repeatedly generating parameters. The model selects the different parameters for each simulation and includes them in the algorithm for finding another alternate path. This method shows excellent characteristics in the application of the probit model. The parameter of the link cost function occurs randomly, and it makes the decision variable fixed by the parameters. The extracted value is not largely deviated from the actual value. In this way, the model searches for the several shortest paths repeatedly (Fiorenzo-Catalano and Van der Zijpp, 2001). Compared to the Monte Carlo simulation, this method employs a criterion for selecting different sets of paths, and thus constructs various path sets. They employed the label criteria such as minimum time, distance, cost, and maximum scenery, etc. It is a method for generating the path sets with various objective functions to reflect the different characteristics of individual travelers.

2.1.2.6 K-Reliable Shortest Paths Method

A reliable shortest path problem is a method of generating a path using the stochastic characteristics of travel time. Chen and Ji (2005) developed the reliable shortest path searching method using the distribution of travel time. It is a method for generating K path sets using the stochastic network condition. Here, the sum of the link travel time is same as the path travel time. Travel time budget is required to arrive at the destination with α probability.

$$T_u = \sum_{i=1}^{l-1} T_u^i$$

The estimated travel time is compared with the path travel time for travelers. This characteristic appears as a risk preference and leads to different values depending on the traveler's purpose of travel or socioeconomic indicators. The model generates the reliable shortest path ranked by the value using the mean and standard deviation of the path travel time in distribution function.

$$\Phi_{T^u}^{-1}(\alpha) = T_u + z_\alpha \sigma_u$$

Chen et al. (2016) analyzed the path travel time using the concept of travel time reliability. They constructed path set generation model to model the characteristics of travel time reliability improving the classical K shortest path algorithm (Yen’s algorithm; Yen, 1971). The study employed the floating car data in Wuhan network to derive the K-α-reliable paths.

Table 2.2 Review of choice set generation

STUDY	DATA	SAMPLE	MODEL	CHOICE SET	METHODOLOGY
Ben-Akiva (1984)	-	-	-	Time and distance are most effective Signals fails to be a significant factor	Labeling method (10 labels) Time, distance, scenic, signals, capacity, hierarchical travel pattern, quality of pavement, commercial dev., highway, congestion
Ramming (2002)	Survey	188	-	Median of 30 routes Maximum up-to 51 routes 160 routes met 80% overlap criteria	Labeling (16 labels) Link Elimination (2-49 unique paths) Link Penalty (3% for origins that are close to MIT, 5% for most distant ones and 4% for remaining) Simulation (48 draws – Gaussian distribution with mean and standard deviation)
Bierlaire and Frejinger (2005)	GPS	1282 observed routes with 927 OD pairs	-	Average of 9.3 routes Maximum 22 and min 2	Simulation method Truncated normal distribution with 20 draws Mean and variance based on the observations The observed route is inserted to the choice set

Van Der Zijpp and Fiorenzo-Catalano (2005)	Sample Network	-	Lawler's algorithm	K shortest paths in the sample network	Constrained K-Shortest Paths Method
Prato and Bekhor (2007)	Web-based survey	236 routes, 339 possible alternatives and 182 different ODs	-	Median size 17 and maximum of 44 routes Merged choice set with median size of 32 and maximum 55 routes	Branch and Bound Compared with other three: Labeling (4 labels) Link Elimination (10 iterations) Link Penalty (15 penalties) Simulation (25 and 35 draws)
Bekhor and Prato (2009)	observations	228 in Turin, Italy and 181 in Boston	-	Same route shortest path(SRSP) Same not route shortest path(SNRSP) Not same route(NSR)	Labeling Link elimination Link penalty Simulation Branch and bound
Frejinger et al.(2009)	Synthetic Data	A network with 38 nodes and 64 links	-	-	Probabilistic Method with Random Walk Algorithm
Menghini et al. (2010)	GPS	3387 bike stages	-	-	Breadth-first search algorithm

Papinski and Scott (2011)	GPS	Home-based Work Trips In all 237 trips	-	k-shortest path generates only 52 out of 237 chosen routes	Shortest path using Potential path area (PPA) concept k-shortest path algorithm
Pillat et al.(2011)	GPS / Personal interview	total 18,300 trips with 61 trips per person/300 participants	-	Replicates 60% of the actually chosen routes Maximum allowed factor of 0.90 commonly	Path enumeration Parameters were estimated by using known routes from the questionnaire
Quattrone and Vieteatta (2011)	Survey / GPS	280 routes/52 routes	-	30 routes Overlap of 75% chosen routes	k-shortest path Combination of 5 criteria
Spissu et al. (2011)	GPS data from smartphones	393 observed routes	Min-cost algorithm Cagliari model	Observed route Last 10 min-cost paths from the simulation	Min-cost algorithm through existing Cagliari model with cost function of time and distance
Schussler et al. (2012)	On-person GPS	36,000 car trips by 2,434 persons	-	Choice set size: 20-100 routes	Breadth-first search link elimination Compared with stochastic

Chen et al. (2013)	Hong Kong Real-time Travel Info. System (RTIS)	1,367nodes 3,655links 5min interval 1,052,640 observations	F-TDRSP-LC F-TDRSP-A B-TDRSP	3 reliable path	Time-dependent Reliable Shortest path (TD-RSP)
Chen et al. (2016)	Wuhan Network, China	19,306nodes 46,757links Floating car data	Various algorithm	9 reliable path	K-Reliable Shortest path(KRSP)

Table 2.3 Summary of choice set generation

CONTENTS		CNC	CPC	CRC	CC	OD PAIRS	STUDY
Deterministic	k-shortest paths	O		O		Single	Akg¨un et al. (2000); Fiorenzo-Catalano and Van Der Zijpp (2001); Cascetta et al. (2002); Ramming (2002)
	Constrained k-shortest paths	O		O	O	Single	Fiorenzo-Catalano and Van Der Zijpp (2001); Van Der Zijpp and Fiorenzo-Catalano (2005)
	Constrained enumeration				O	Single	Prato and Bekhor (2006), Pillat et al.(2011)
	Link elimination	O				Single	Martins (1984); Ramming (2002), Bekhor and Prato (2009)
	Link penalty	O				Single	Akg¨un et al. (2000); Ramming (2002)
Heuristic	k-dissimilar paths	O			O	Single	Akg¨un et al. (2000);
	Gateway method			O		Single	Akg¨un et al. (2000);
	Essentially least-cost paths				O	Single	Hunt and Kornhauser (1996)
	Labelling		O			Multiple	BenAkiva et al. (1984); Dial (2000), Bekhor and Prato

							(2009)
	Monte Carlo (MC)	O				Multiple	Sheffi and Powell (1982); Ramming (2002)
	Accelerated MC	O				Multiple	Bliemer et al. (2004)
	MC Labelling combination	O	O			Multiple	Fiorenzo-Catalano and Van Der Zijpp (2001)
	K-Reliable Shortest Path	O		O		Single	Nie and Wu (2009), Chen et al. (2013), Chen et al. (2016)

Remarks: CNA(Change Network Attribute), CPC(Change Path Criteria), CRC(Change Restriction Criteria), CC(Check Constraints)

2.2 Route Choice Model

2.2.1 Multinomial logit (MNL) based Model

2.2.1.1 Multinomial logit model

McFadden (1973) developed the Multinomial Logit Model (MNL) of which is a general model form of choice model. The model is developed from the perspective of econometrics, and it is the basis of the various choice model. In the transportation analysis, Researchers has employed the MNL model in the mode choice model before applying for the route choice model.

Studies have proposed the probit and multinomial logit model to use in stochastic or probabilistic assignment models (Ben-Akiva and Lerman, 1985; Sheffi, 1985; Cascetta, 1990; Otuzar and Williamsen, 1990). MNL model supposed homogeneous error term of traveler's perception to construct the utility function based on the type extreme value distribution. The study supposed the Gumbel distribution, which is the type of log-transformed Weibull distribution.

Researchers employed the MNL model in route choice model to analyze the stochastic user equilibrium (SUE). It supposes the perception of the route for travelers as a type of i.i.d Gumbel distribution. MNL model is developed for calculating the choice probability in between distinct distinguished alternatives. The expression of the function is in below.

$$\mathrm{P(i)} = \frac{e^{-\beta c_i}}{\sum_{j \in A_n} e^{-\beta c_j}}$$

Where, P(i) = the probability value of a traveler using path *i* for homogeneous travelers

A_n = path alternatives considered by individual n

C_i, C_j = cost function reflecting path attribute of paths i and j

β = utility coefficient

There are various contents of attribute reflecting path C_i in previous researches. Many researchers have employed the explanatory variables to make the model more feasible (Dial, 1971; Fisk, 1980; Bekhor et al., 2006; Prato and Bekhor, 2006; Frejinger and Bierlaire, 2007; Bovy et al., 2008).

Dial (1971) developed the traffic assignment model using MNL model with setting the possible choice set generation. The researcher generated the efficient and feasible paths using STOCH algorithm, which is generation method of a path set using a link far from the origin and close to the destination. Fisk (1980) analyzed the SUE problem to minimize the difference of error.

Route choice model is different from the mode choice model by the correlation between alternatives. MNL model does not show the characteristics of overlapping, and the model does not derive the feasible result due to the assuming the independently and identically distributed (i.i.d) random utility. Recently, there are several models of C-logit Model(C-Logit), Path-size Logit (PSL) Model, Cross-Nested Logit (CNL) Model, Implicit Availability/Perception (IAP) Logit Model, Logit Kernel (LK) Model, Etc. to improve the shortcomings of MNL model, which is represented to the overlapping route problem. Figure in below shows the example to interpret the overlapping route problem.

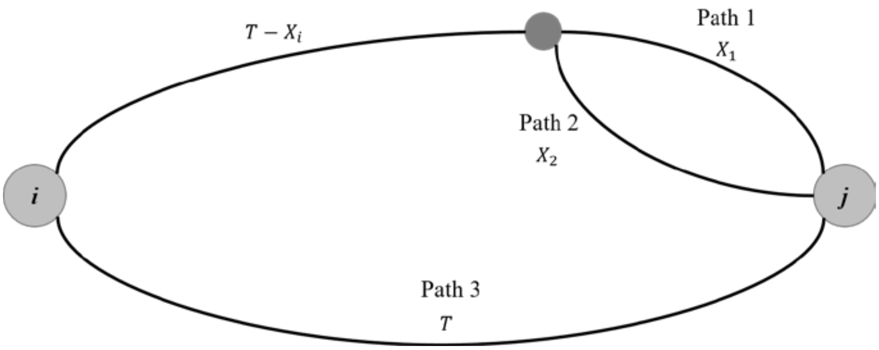


Figure 2.1 The Overlapping Route Problem

2.2.1.2 C-Logit Model(C-Logit)

Logit model can efficiently calibrate the choice behavior from disaggregated data. Route choice model establishes the specific alternatives of choice set to choose a route from the given choice set. The model estimates the more accurate probability using the considered routes. The relationship between routes has the various types of overlapping links. Researchers have proposed the improved model to overcome the defect of MNL model. Independence from Irrelevant Alternatives (IIA) property of MNL model has been derived the unrealistic route choice probability, which has observed by many previous researches (Floian and Fox, 1976; Sheffi, 1985; Daganzo and Sheffi, 1977).

Cascetta et al. (1996) developed the C-logit model to reflect the overlapping between routes. C-Logit model employs the commonality factor (CF) to improve the model structure from the existing MNL model. From the above figure, the probability in C-logit model has the difference by the size of X. When the size of X_i is getting smaller (that is, $X_i \rightarrow T$), so there is no difference between path 1 and path2 that the overlapping does not occur. As a result, because the path 1 and path 2 are the same route in the model, it derives the same result from MNL model. On the other hand, when the X_i is getting bigger, the probability has an effect from the X_i than $\mathrm{T} - X_i$. It is necessary to develop the model reflecting the impact of the size of X_i.

$$\mathrm{P}(i|A_n) = \frac{e^{V_{in}+CF_{in}}}{\sum_{j \in A_n} e^{V_{jn}+CF_{jn}}}$$

Where, $\mathrm{P}(i|A_n)$ = the probability value of travelers using path *i* among the route alternatives A_n for homogeneous travelers

A_n = path alternatives considered by individual *n*

V_{in} V_{jn} = cost function of paths *i* and *j* by individual *n*

CF_{in}, CF_{jn} = commonality factor of paths *i* and *j* by individual *n*

As the length of the overlapping link becomes smaller, the difference between the paths becomes smaller, so CF always has a negative value. CF is in the form of a log transform. The study proposed the equations for commonality factor correlation.

$$CF_{in} = -\beta_0 ln \sum_{j \in C_n} \left(\frac{L_{ij}}{\sqrt{L_i L_j}} \right)^r,$$

$$CF_{in} = -\beta_0 ln \sum_{a \in \Gamma_i} \frac{l_a}{L_i} N_{an},$$

$$CF_{in} = -\beta_0 \sum_{a \in \Gamma_i} \frac{l_a}{L_i} lnN_{an}, \text{ and}$$

$$CF_{in} = -\beta_0 ln\left[1 + \sum_{\substack{j\in C_n \\ j\neq i}} \left(\frac{L_{ij}}{\sqrt{L_i L_j}}\right)\left(\frac{L_i - L_{ij}}{L_j - L_{ij}}\right)\right]$$

Where, β_0, γ = coefficients to estimate the model

L_{ij} = the length paths i and j have in common

Γ_i = the set of arcs in path *i*

l_a = the length of link *a*

$N_{am} = \sum_{j\in C_n} \delta_{aj}$ = the number of shared route for link *a*

δ_{aj} = the link-path incidence dummy, 1 if path j uses link a and 0 otherwise

C-Logit model proposed to improve the IIA property of MNL by eliminating the counter-intuitive results relative to overlapping routes. The model has an easily computable closed form of conventional MNL model with explicit route enumeration. The basic idea of C-Logit model is adding the additional attribute, named communality factor, in the utility function of MNL model to consider the covariance of random residuals of perceived overlapping links. The C-Logit model is mathematically equivalent to the Path-size Logit (PSL) model in the previous study (Ramming, 2002).

2.2.1.3 Path-size Logit Model (PSL)

Researchers developed the Path-size Logit Model (PSL), which is modified MNL model considering the degree of overlapping of the routes to construct the model considering the correlation between routes in the MNL model. (Ben-Akiva and Ramming, 1998) The PSL model is consistent with the adjustment process of the C-Logit model regarding behavioral theory (Ramming, 2002). C-Logit Model and PSL Model include the correction term to interpret the link overlapping on the utility function of alternative routes.

$$P(i|A_n) = \frac{e^{V_i + ln\,PS_{in}}}{\sum_{j \in A_n} e^{V_j + ln\,PS_{jn}}} = \frac{PS_{in} e^{V_i}}{\sum_{j \in A_n} PS_{jn} e^{V_j}}$$

Where, $\mathrm{P}(i|A_n)$ = the probability value of travelers using path *i* among the route alternatives A_n for homogeneous travelers using path-size model

A_n = path alternatives considered by individual n

V_i V_j = cost function of paths *i* and *j*

PS_{in}, PS_{jn} = path-size factor of paths *i* and *j*

It is a method to analyze the utility except for the overlapped links between routes. In the previous figure, when the link length of X_i becomes 0, the problem of overlapping path does not occur. PSL model removes the factor of overlapping links from the alternative path in the j alternatives.

Ben-akiva and Ramming (1998) presented the equation for the path size resistance in below.

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{N_{an}} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in A_n} \left(\frac{\bar{L}}{L_i}\right)^{\gamma} \delta_{aj}}$$

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in A_n} \frac{L_i}{L_j} \delta_{aj}}$$

Where, (l_a/L_i) = a weight corresponding to the path resistance of a particular overlapping link

N_{an} = the number of routes that share a nested link

δ_{aj} = link incidence dummy, if route *j* uses link a and otherwise 0

From the above figure, if X_1 of path 1 is very large and X_2 is small, the choice probability of path 1 will be relatively small. The $\mathrm{T} - X_i$ values are mutually shared values in path 1 and path 2, and it will have the same effect on both paths, so the choice probability is determined by the value of X_1. Considering the choice probability of all paths, path 2 and path 3 have the same path length, but the path 2 is less than path 3 because of the difference in path size. Path 1 yields a relatively small probability due to the size of X_1.

Frejinger et al. (2009) devised an expanded path size logit model that includes additional factors compared to the path-size logit model. The model employs an additional factor to compensate for the sample.

$$EPS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{M_{an}^{EPS}}$$

$$M_{an}^{EPS} = \sum_{j \in A_n} \delta_{aj}\, \Phi_{jn}$$

Where, Φ_{jn} = expansion factor defined by

$$\Phi_{jn}\begin{cases}1, & if\ \ \delta_{jc}=1\ \ or\ \ q(j)R_n \geq 1 \\ \frac{1}{q(j)R_n}, otherwise. & \end{cases}$$

2.2.1.4 Path size correction logit (PSCL)

Since there is no satisfactory derivation based on theoretical arguments, it is necessary to employ the correction terms for the specification (Bovy et al., 2008). They suggested the detailed and systematic derivations of assumption for correcting. Path-size correction factor makes the model result in the better model by the notion of aggregate alternatives. This model is simplified the nested logit model.

Because routes share one or more common links, it requires the following additional assumptions.

- All elemental routes of C_n that share links have the same systematic utility V_{in}
- All error terms ε_{in} are i.i.d. (independently and identically distributed) and thus μ_a is equal to μ.

The equation of path-size correction term is presented in below.

$$P(i|A_n) = \frac{e^{V_{in}+\beta_{PSC}*PSC_{in}}}{\sum_{j\in A_n} e^{V_{jn}+\beta_{PSC}*PSC_{jn}}}$$

Where, $P(i|A_n)$ = the probability value of travelers using path i among the route alternatives A_n for homogeneous travelers using path-size correction logit model

A_n = path alternatives considered by individual n

V_i V_j = cost function of paths i and j

PSC_{in}, PSC_{jn} = path-size correction factor of paths i and j

The path-size correction logit model includes the path-size correction term (PSC_{in}) from the path-size model. The model applies the modified variables to the overlapping of paths. The calculation method of the path-size correction term is as follows.

$$PSC_i = -\sum_{a \in \Gamma_i} \left(\left(\frac{l_a}{L_i} \right) ln \sum_{j \in A_n} \delta_{aj} \right)$$

Where, l_a = the length of link *a*

L_i = the length of route *i*

(l_a/L_i) = a weight corresponding to the path resistance of a particular overlapping link

δ_{aj} = link path incidence dummy, if route *j* uses link *a* 1, and otherwise 0

PSC logit model considers the stochastic correlation of error term to interpret the closed form of logit model as the type of generalized extreme value model.

2.2.1.5 Implicit Availability/Perception (IAP)

Cascetta and Papola (1998) develop the implicit availability/perception (IAP) logit model by aggregating the path generation model and route choice model using the traveler's perception of routes. C-logit model merely includes the correction parameter in MNL model to improve the accuracy of the model. On the other hand, IAP logit model constructs the model by extracting the non-recognized and unavailable routes. Furthermore, it includes the process of finding the paths avoiding the overlapping links. IAP logit model is a process for deriving the probability of traveler n choosing the route i.

$$P_n(i) = \frac{e^{V_i + ln\mu_n(i)}}{\sum_{j \in M} e^{V_j + ln\mu_n(j)}}$$

$$P_n(i) = \frac{\mu_n(i)e^{V_i}}{\sum_{j \in M} \mu_n(j)e^{V_j}}$$

Where, M = master choice set (the set of all possible routes)

$\mu_n(i)$ = 1, if path i is available, while 0 implies either the path is unaware of travelers or unavailable

If all of the routes are non-recognized or impossible routes, the value of $\mu_n(i)$ becomes 0, so the log-transform of $\mu_n(i)$ is negative infinity. Thus, the prerequisite must have at least one possible or perceived route. There is a case for the unknown $\mu_n(i)$. In this case, the expected value ($\overline{\mu_n}(i)$) can be derived by using a random variable. The model also assumes Taylor series expansion to estimate the maximum variance of $\mu_n(i)$ by the Bernoulli distribution.

$$P_n(i) = \frac{exp\left[V_i + ln\overline{\mu_n}(i) - \frac{1-\overline{\mu_n}(i)}{2\overline{\mu_n}(i)}\right]}{\sum_{j \in M} exp\left[V_j + ln\overline{\mu_n}(j) - \frac{1-\overline{\mu_n}(j)}{2\overline{\mu_n}(j)}\right]}$$

The model derives $\overline{\mu_n}(i)$ by assuming the form of the model.

$$\overline{\mu_n}(i) = \frac{1}{1 + exp(\sum_{k=1}^{K} \gamma_k Y_{ink})}$$

Where, Y_{ink} = k[th] variable relating to the availability or perception of alternative *i* for individual *n*

γ_k = coefficients

This model is not developed for a route choice model. But the model constructed for mode choice model with four modes. Researchers constructed a model based on the possibility of recognizing the travelers, rather than the availability.

Cascetta and Papola (2002) conducted a model improving the existing model for route choice. It is important to construct a set of perceived paths in a route choice model with a relatively small size of alternatives. They constructed a route perception model by comparing with the existing route choice set generation model.

They employed $PC_n(j)$ to represent the perceived path. The equation for obtaining the route choice probability is given as follows.

$$P_n(i) = \frac{exp\left[V_{in} + \alpha lnC_n(i) - \alpha \frac{1 - C_n(i)}{2C_n(i)}\right]}{\sum_{j \in M} exp\left[V_{jn} + \alpha lnC_n(j) - \alpha \frac{1 - C_n(j)}{2C_n(j)}\right]}$$

$$PC_n(i) = \frac{1}{1 + exp(-\boldsymbol{\gamma Y_{in}})}$$

Where, $\boldsymbol{Y_{in}}$ = vector of socio-economic characteristics of the decision maker and attribute relative to the perception of each alternative route

$\boldsymbol{\gamma}$ = vector of coefficients

It is possible to confirm the choice set more efficiently and to have a higher fitness index compared with the C-logit model. Furthermore, it derives choice set for six paths to analyze the route choice behavior using the coverage factor.

2.2.2 Generalized extreme value (GEV) based model

2.2.2.1 Paired-Combinational Logit (PCL)

Researchers improved the PCL model in route choice model using the attributes of the route. (Koppelman and Wen, 2002; Prashker and Bekhor, 1998) According to the generalized extreme value model, proposed by McFadden (1978), they employed a method of comparing alternatives by using the correlation between two shared links in the choice set. The probability calculation method for PCL is as follows.

$$P_i = \sum_{\substack{j \in A_n \\ j \neq i}} P(i|ij)P(ij|A_n)$$

$$P(i|ij) = \frac{exp\left(\frac{\mu V_i}{1-\sigma_{ij}}\right)}{exp\left(\frac{\mu V_i}{1-\sigma_{ij}}\right) + exp\left(\frac{\mu V_j}{1-\sigma_{ij}}\right)}$$

$$P(ij|A_n) = \frac{\left[exp\left(\frac{\mu V_i}{1-\sigma_{ij}}\right) + exp\left(\frac{\mu V_j}{1-\sigma_{ij}}\right)\right]^{1-\sigma_{ij}}}{\sum_{k=1}^{A_n-1}\sum_{m=k+1}^{A_n}(1-\sigma_{km})\left[exp\left(\frac{\mu V_k}{1-\sigma_{km}}\right) + exp\left(\frac{\mu V_m}{1-\sigma_{km}}\right)\right]^{1-\sigma_{km}}}$$

$$\sigma_{ij} = \frac{L_{ij}}{\sqrt{L_i L_j}}$$

Where, $P(i|ij)$ = the conditional probability of selecting route *i* provided that the binary pair (i, j) is chosen

$P(ij|A_n)$ = the unobserved probability for the pair (i, j)

σ_{ij} = similarity coefficient (i, j)

L_{ij} = the common length for the route *i* and *j*

L_i, L_j = the length of the route *i*, *j*

The PCL model is superior to other models in calculating the elasticity of the model. This characteristic is more flexible than the MNL model.

2.2.2.2 Cross-Nested Logit (CNL)

Vovsha and Bekhor (1998) proposed the link-nested logit model as cross-nested logit (CNL) model in route choice model. The CNL model is formulated differently from the structure of the existing nested logit model. This model is a method of deriving the model using the parameter applying for the weight value of nest n of alternative *i*. The parameter α_{mi} has a value between 0 and 1, and the sum of α_{mi} is applied to the generalized value of the lowest nest.

$$P\big((i|C_n)\big) = \sum_{m=1}^{M} P(C_{mn}|C_n)P_n(i|C_{mn})$$

$$P\big((i|C_{mn})\big) = \frac{\alpha_{mi}e^{V_{in}}}{\sum_{j\in C_{mn}} \alpha_{mj}e^{V_{jn}}}, P\big((C_{mn}|C_n)\big) = \frac{e^{V_{C_{mn}}+\mu_m I_{C_{mn}}}}{\sum_{l=1}^{M} e^{V_{C_{ln}}+\mu_m I_{C_{ln}}}}$$

$$I_{C_{mn}} = ln \sum_{j\in C_{mn}} \left(\alpha_{mj}e^{V_{jn}}\right)^{1/\mu_m}$$

$$P_n\big((i|C_n)\big) = \frac{\sum_{m=1}^{M}(\alpha_{mi}e^{V_i})\left(\sum_{j\in C_{mn}} \alpha_{mj}e^{V_i}\right)^{\mu_m-1}}{\sum_{m=1}^{M}\left(\sum_{j\in C_{mn}} \alpha_{mj}e^{V_i}\right)}$$

$$= \frac{\sum_{m=1}^{M}(\alpha_{mi}e^{V_i})}{\sum_{j\in C_{mn}}\sum_{m=1}^{M}\left(\alpha_{mj}e^{V_i}\right)} = \frac{e^{V_i}\sum_{m=1}^{M}(\alpha_{mi})}{\sum_{j\in C_{mn}} e^{V_i}\sum_{m=1}^{M}\left(\alpha_{mj}\right)}$$

Where, $P(i|C_n)$ = the marginal probability of choosing nest C_n

$P(C_{mn}|C_n)$ = the marginal probability of choosing nest m

$P_n(i|C_{mn})$ = the conditional probability of choosing route *i* if the route includes the nest C_{mn} for individual *n*

$P_n\big((i|C_n)\big)$ = the conditional probability of choosing route *i* if the route includes the nest C_n for individual *n*

α_{mi}, α_{mj} = inclusion coefficients

μ_m = parameter, if to reduce MNL model, the value are 1

V_i, V_j = cost function of routes *i* and *j*

The model has the same structure as the basic MNL structure with the sum of α_{mi}, which is applied as one constant value. The link-nested logit model analyzed the model by applying the following weights.

$$\alpha_{mi} = \frac{l_m}{L_i}\delta_{mi}$$

Where, l_m = the length of link m

L_i = the length of route i

δ_{mi} = the link-route incidence dummy, δ_{mi}=1 if route includes the link m, and 0 otherwise.

Bekhor et al. (2006) estimated the CNL with the choice sets consisting more than one route. From the research, CNL and PSL model have come out with the best model fit in the route choice model. Moreover, Bierlaire (2001) estimated a CNL mode choice model using inter-city data. He employed the combined revealed preference and stated preference data.

2.2.2.3 Generalized Nested Logit (GNL)

GNL model is generalized the CNL model by adopting the different nesting parameter for each nest in the model Bekhor (2001) proposed the formulation for nesting coefficient is in below:

$$\mu_m = \left(1 - \frac{\sum_{l \in C_n} \alpha_{mi}}{\sum_{l \in C_n} \delta_{mi}}\right)^{\gamma}$$

$$P\big((i|C_{mn})\big) = \frac{(\alpha_{mi} e^{V_{in}})^{\mu_m}}{\sum_{j \in C_{mn}} \left(\alpha_{mj} e^{V_{jn}}\right)^{\mu_m}}$$

Where, α_{mi} = the incursion parameter

δ_{mi} = the link-route incidence dummy, δ_{mi}=1 if route includes the link a, and 0 otherwise.

γ = parameter

The GNL model is introduced to explain the similarity of the probabilistic part of the utility function. Swait (2001) proposed the Choice Set Generation Logit (GenL) model to combine the choice set generation model and route choice model. The model presented the result from inter-city route choice model involving the mode choice among four modes. GNL model is referred to Wen and Koppelman (2001) with new specific parameter μ as "Generalized Nested Logit model."

The error term is divided into two parts in the model. One part represents correlation and heterogeneity, and the other part describes i.i.d extreme value.

2.2.3 Multinomial weibit (MNW) based model

2.2.3.1 Multinomial weibit model (MNW)

Castillo et al. (2008) developed a multinomial weibit model (MNW) based on the Weibull distribution to relieve the independently and identically distributed assumption based on Gumbel distribution for MNL model. This model represents the equation for Weibull distribution of travel time, mean travel cost, and perception variance in below.

$$\text{CDF } F_{G_k^{ij}}(t) = 1 - exp\left\{-\left[\frac{1-\zeta_k^{ij}}{\varphi_k^{ij}}\right]^{\beta_k^{ij}}\right\}, \text{t} \in \ (0,\infty)$$

$$\text{Mean travel cost } g_k^{ij} = \ \zeta_k^{ij} + \ \varphi_k^{ij}\Gamma\left(1+\frac{1}{\beta_k^{ij}}\right)$$

Route perception variance $\left(\sigma_k^{ij}\right)^2$

$$= \left(\varphi_k^{ij}\right)^2 \Gamma\left(1 + \frac{2}{\beta_k^{ij}}\right) - \left(g_k^{ij} - \zeta_k^{ij}\right)^2$$

It is possible to model the route choice probability by the equation according to the above distribution. The probability of choosing route k can be expressed as follows.

$$U_k^{ij} = \left(g_k^{ij} - \zeta^{ij}\right)^{\beta^{ij}} \varepsilon_k^{ij}, \forall k \in R_{ij}, ij \in IJ$$

$$P_k^{ij} = \frac{\left(g_k^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}{\sum_{n \in R_{ij}} \left(g_n^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}, \forall k \in R_{ij}, ij \in IJ$$

Where, g_k^{ij} = the mean of travel cost perceived by travelers for route $k \in R_{ij}$ between origin *i* and destination *j*

ζ_k^{ij} = location parameter

β_k^{ij} = shape parameter

φ_k^{ij} = scale parameter

Γ = Gamma distribution

σ_k^{ij} = perception variance of the Weibull distribution

U_k^{ij} = utility function for route k between origin *i* and destination *j*

ε_k^{ij} = independently Weibull distributed random error term for route k between origin *i* and destination *j*

MNW model can reflect the effect of length of the route between OD pairs and make the difference of choice probability in the model. The longer routes share, the more similar probability than the shorter routes with the same difference between alternatives in route choice model. The model can extract the more feasible choice probability (Kitthamkesorn and Chen, 2013).

2.2.3.2 Path-size Weibit model(PSW)

MNW model does not consider the overlapping links in the route choice model. It needs to include the additional factor to relax the independently identically distributed assumption. The path-size factor was proposed to the MNW model. The precious path-size factor accounts for the elimination of effect on shared links between routes (Ben-akiva and Bierlaire, 1999).

$$\omega_k^{ij} = \sum_{a \in \gamma_k} \frac{l_a}{L_k^{ij}} \frac{1}{\sum_{n \in R_{ij}} \delta_{an}^{ij}}, \forall r \in R_{ij}, ij \in IJ$$

Where, ω_k^{ij} : path-size factor for route $k \in R_{ij}$ between origin i and destination j

l_a: the length of link a

L_k^{ij}: the length of route k between origin i and destination j

γ_k: the set of all links in route k between origin i and destination j

δ_{an}^{ij}: the link-route incidence dummy, δ_{an}^{ij}=1 if route includes the link a, and 0 otherwise.

If routes have the more overlapping links, ω_k^{ij} will have the smaller value. PSW model includes the path-size factor in the utility function and route choice model.

$$U_k^{ij} = \frac{\left(g_k^{ij} - \zeta^{ij}\right)^{\beta^{ij}}}{\omega_k^{ij}} \varepsilon_k^{ij}, \forall k \in R_{ij}, ij \in IJ$$

$$P_k^{ij} = \frac{\omega_k^{ij}\left(g_k^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}{\sum_{n\in R_{ij}} \omega_n^{ij}\left(g_n^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}, \forall k \in R_{ij}, ij \in IJ$$

Where, U_k^{ij} = utility function for route k between origin *i* and destination *j*

g_k^{ij} = the mean of travel cost perceived by travelers for route $k \in R_{ij}$ between origin *i* and destination *j*

ζ_k^{ij} = location parameter

β_k^{ij} = shape parameter

ω_k^{ij} = path-size factor

ε_k^{ij} = independently Weibull distributed random error term for route *k* between origin *i* and destination *j*

PSW model, as well as PSL model, have the same route choice probability for handling the overlapping problem of the routes using the path-size factor. The path-size factor can reflect the correlation of the routes and make the difference of choice probability in the route choice model.

2.2.4 Mixed logit based model

Researchers devised the mixed multinomial logit model to improve the limitations of logit based model. Since the logit based model has the IIA property, it is difficult to reflect the effect of overlapping links between alternatives. They employed the probit model for route choice to relax the correlation on the error term between alternatives. McFadden and Train (1998) proposed the choice model combined with the Gaussian and Type 1 Extreme value error term, which is called mixed logit model. Walker (2000) present the general form of the mixed logit model in below:

$$\mathbf{U} = \mathbf{X\beta} + \boldsymbol{\varepsilon} = \mathbf{X\beta} + \mathbf{FT\zeta} + \boldsymbol{\nu}$$

$$P_k = \Gamma(k|\zeta) = \frac{e^{\mu(\boldsymbol{X_k\beta}+\boldsymbol{F_k T\zeta})}}{\sum_{l \in C_n} e^{\mu(\boldsymbol{X_l\beta}+\boldsymbol{F_l T\zeta})}}$$

Where, $\mathbf{U}$ = the C_n by 1 vector of utility function

$\mathbf{X}$ = the C_n by *K* matrix of variables

$\boldsymbol{\beta}$ = the column vector of *K* unknown parameters for variables

$\boldsymbol{\varepsilon}$ = the C_n by 1 vector of error terms

$\mathbf{F}$ = the C_n by *M* factor loading matrix

$\mathbf{T}$ = the *M* by *M* lower triangular matrix of unknown parameters

$\boldsymbol{\zeta}$ = the *M* by 1 vector of i.i.d standard normal variables as unobservable factors

$\boldsymbol{\nu}$ = the *M* by *M* lower triangular matrix of unknown parameters

$\Gamma(k|\zeta)$ = the probability of chosen route *k* with given ζ.

The $\mathbf{F}$ and $\mathbf{T}$ elements are estimated by the observed data in the choice model. In contrast, since the $\boldsymbol{\zeta}$ element is unobservable factor, the model can express the heterogeneous characteristics of the choice. Lee et al. (2004)

developed the mode choice model using the mixed logit model. They estimated the heterogeneous choice behavior in mode choice, and found that the random coefficient model improves the log-likelihood estimation. In addition, Ramming (2002) estimated the route choice behavior in Boston area. He developed the mixed logit model with the given network knowledge and choice set in the route choice model.

Table 2.4 Review of route choice models

STUDY	DATA	SAMPLE	MODEL	CHOICE SET	EXPLANATORY VARIABLES
Bekhor et al.(2002)	Survey	159	Logit Kernel (LK) model Compared with: MNL, and PSL	-	Distance Free-flow time Dummy variables for road sections Time spend on government numbered routes Delays for different income categories
Bekhor et al.(2006)	Survey	159	MNL/CNL PSL	Choice sets consisting more than 1 routes	Distance Free-flow time Dummy variables for road sections Time spend on government numbered routes Delays for different income categories Dummy for least distance and time paths
Prato and Bekhor (2006)	survey in Turin	-	MNL/GNL/CNL/LNL C-logit PSL	Home-to-work (236 chosen routes, 339 possible alternatives, 182 different ODs)	Level-of-service(distance, free-flow time, and travel time) Landmark dummy variables (1 or 0) Behavioral variables (habit, spatial ability, and familiarity)

Frejinger and Bierlaire (2007)	GPS	2978	Error Component (EC) model Compared five different specifications of EC model with: MNL and PSL	2244 unique observed routes 2179 OD pairs	Path size Estimated travel time Number of speed bumps Number of left turns Avg. link length
Prato and Bekhor (2007)	Survey	216	MNL/GNL/CNL C-logit PSL LK with a factor analytic	At-least 5 alternatives except the chosen route	Level-of-service(distance, free-flow time, and travel time for experience and non-experienced drivers) Landmark dummy variables (1 or 0) Behavioral variables (habit, spatial ability, and familiarity)
Bierlaire, M. and E. Frejinger (2008)	GPS / Reported trips	780	Path Size Logit Model Subnetwork Model	Freeway free-flow CN free-flow Main free-flow Small free-flow	The Comparison of error between True Traveling Route & Route using GPS data (DDR) Calibrating the Parameters for Route Choice Behavior using GPS data

Bliemer and Bovy(2008)	Forecasting probabilities	-	C-logit PSL, PSCL, PCL, CNL	A simple hypothetical network of a single OD pair with 12 available routes	-
Bovy et al.(2008)	Turin and Boston datasets	228/181	MNL PSL Path Size Component Logit (PSCL) model	-	Total length Travel time % of travel time on major roads Dummy for the path with the maximum avg. speed % delay with respect to the free-flow time
Frejinger et al.(2009)	Synthetic dataset	-	Expanded PS Compared with the original PSL model	-	Length Number of speed bumps
Hess and Rose(2009)	Survey	205 persons / 3,280 observations	MNL models allowing for inter-respondent and intra-respondent taste heterogeneity	16 choice situations for 2 alternatives	Free flow travel condition Slowed down time travel condition Travel time variability Running costs Toll costs

Schussler and Axhausen (2010)	GPS	On-person GPS logger 1500 observations	C-logit PSL model	-	Time-of-day dependent travel times on each road types (motorway(MW), extra-urban main(EUM), urban main(UM), and local road(LR)) Travel time proportions on each road type Road type specific path sizes
Kitthamkesorn and Chen (2013)	Winnipeg network	-	PSL MNW, PSW	3~6 from specific OD pairs	Route cost (travel time) Traffic volume
Li et al. (2016)	GPS / On-board unit	95 drivers / 2,182 trips	C-logit PSl, PSCL PCL, CNL	50 unique paths using random walk method	Route specific (Free travel time, Number of intersections) Traveler specific (Age, Gender, Car displacement) OD specific (Distance, Familiarity)

Table 2.5 Summary of route choice model

CONTENTS		MODEL	DESCRIPTION	STUDY
	MNL (Multinomial Logit)	$P(r\|Q_d) = \frac{e^{V_r}}{\sum_{r' \in Q_d} e^{V_{r'}}}$	-	Dial (1971), Fisk (1980) Bekhor et al. (2006), Prato and Bekhor (2006)
	C-Logit	$P(r\|Q_d) = \frac{e^{V_r - \alpha CF_r}}{\sum_{r' \in Q_d} e^{V_{r'} - \alpha CF_{r'}}}$	Subtracting Commonality Factor(CF)	Cascetta et al. (1996), Schussler and Axhausen (2010), Zhou et al. (2012)
MNL	IAP (Implicit Availability / Perception)	$P(r\|Q_d) = \frac{e^{V_r + \ln(q_r)}}{\sum_{r' \in Q_d} e^{V_{r'} + \ln(q_{r'})}}$	Adding Path Awareness (IAP)	Cascetta et al. (2002)
	PSL (Path-size Logit)	$P(r\|Q_d) = \frac{e^{V_r + \beta_r \ln(PS_r)}}{\sum_{r' \in Q_d} e^{V_{r'} + \beta_r \ln(PS_{r'})}}$	Adding ln(size) (Path Size)	Frejinger et al. (2009), Schussler and Axhausen (2010), Li et al. (2016)
	PSCL (Path-size Correction Logit)	$P(r\|Q_d) = \frac{e^{V_r + \beta_r PSC_r}}{\sum_{r' \in Q_d} e^{(V_{r'} + PSC_{r'})}}$	Adding PSC (Path Size Correction)	Bovy et al. (2008)

GEV	PCL (Paired-Combinational Logit)	$P(r\|Q_d) = \frac{e^{\left(\frac{\mu V_r}{1-\sigma_r}\right)}}{\sum_{r' \in Q_d} e^{\left(\frac{\mu V_{r'}}{1-\sigma_{r'}}\right)}}$	Multiplications of Unobserved Probability(P_{ij})	Bliemer and Bovy (2008),
	CNL (Cross-nested Logit)	$P(r\|Q_d) = \frac{\kappa_{mr} e^{\mu V_r}}{\sum_{r' \in Q_d} \kappa_{mr'} e^{\mu V_{r'}}}$	Multiplications of Marginal(nested) Probability	Prato and Bekhor (2006), Bliemer and Bovy (2008)
	GNL (Generalized Nested Logit)	$P(r\|Q_d) = \frac{\alpha_{mr} e^{\mu V_r}}{\sum_{r' \in Q_d} \alpha_{mr'} e^{e^{\mu V_{r'}}}}$	Including the Allocation Parameter($m,\ \alpha_{nm}$)	Prato and Bekhor (2006), Wen and Koppelman (2001)
MNW	MNW (Multinomial Weibit)	$P_k^{ij} = \frac{\left(g_k^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}{\sum_{n \in R_{ij}} \left(g_n^{ij} - \zeta^{ij}\right)^{-\beta^{ij}}}$	Weibull Distribution Based Model (Open Form)	Kitthamkesorn and Chen (2013) Kitthamkesorn and Chen (2014) Sharifi et al. (2015)
	PSW (Path-size Weibit)	$P_{ik} = \frac{\left(c_k^i - \xi_i^0\right)^{-\beta_i}}{\sum_{s \in K_i} \left(c_k^i - \xi_i^0\right)^{-\beta_i}}$		Castillo et al. (2008), Kitthamkesorn and Chen (2013)

Table 2.6 Summary of variable choice

STUDY	VARIABLES for level of service					VARIABLES for network attribute									SAMPLE	MODEL
	FTT	MTT	SDTT	TD	TC	Toll	LM	#BR	#BU	#LT	#INT	%MR	%UI	%D		
Bekhor et al.(2006)	○			○								○	○	○	159	PSL/CNL
Prato and Bekhor(2006)	○	○		○			○								228	GNL/CNL/PSL
Frejinger and Bierlaire(2007)		○		○					○	○					2,978	MMNL/PSL
Prato and Bekhor(2007)	○	○		○			○								216	GNL/CNL/MPSL
Bierlaire and Frejinger(2008)	○														780	PSL
Bliemer and Bovy(2008)		○													Simulation	PSL/PSCL/PCL
Bovy et al.(2008)		○		○								○		○	228	MNL/PSL/PSCL
Frejinger et al.(2009)				○					○						Simulation	PSL
Hess and Rose(2009)	○	○	○		○	○									3,280	MMNL
Schussler and Axhausen(2009)		○										○			1,500	C-logit/PSL
Kitthamkesorn and Chen(2013)		○		○											Simulation	MNW/PSW
Li et al.(2016)	○			○							○				2,182	PCL/CNL
This Study(2017)	○	○	○	○	○	○		○			○		○		**40,000**	**PSCL/MPSCL**

Annotation: FFT(Free Travel Time), MTT(Mean of Travel Time), SDTT(Standard Deviation of Travel Time), TD(Travel Distance), TC(Travel Cost), Toll(Toll fare), LM(LandMark dummy), #BR(Number of Bridge), #LT(Number of Left Turn), #INT(Number of INTersection), %MR(Percentage of Major Road), %UI(Percentage of UnInterrupted flow), %D(Percentage of Delay)

2.3 Review result and limitation

This section summarizes the results of previous studies and presents the limitations of choice set generation models and route choice models. The limitations of the study are as follows.

2.3.1 Limitations for choice set generation

- Simulation-based choice set generation model *(Prato and Bekhor, 2007; Bliemer and Bovy, 2008; Frejinger et al., 2009; Bovy and Fiorenzo-Catalano, 2009; Chen et al., 2016)*

Researchers constructed models to calculate the travel time distribution through simulating the path travel time using a certain amount of observation travel time due to the shortage of observed paths. These studies attempted to reflect the characteristics of the travel time without reflecting actual traveler's behavior. Researchers constructed the choice set generation model and verified the model by applying to the real network.

- Lack of actual travel data *(Van Der Zijpp, N.J. and Fiorenzo-Catalano, S., 2005; Bekhor and Prato, 2009; Pillat et al., 2011; Schussler et al., 2012)*

There is a difficulty in analyzing a more accurate travel behavior due to the lack of observation for travel time between the specific OD pairs. In particular, previous researchers have carried out to analyze the travel patterns of travelers using GPS data. However, there is a limit to describe the traveled paths using a large amount of GPS data. They employed the GPS data to analyze travel patterns. Moreover, there is a model that analyzes travel behavior using stated preference (SP) data.

- Difficulties of identifying the consideration set *(Frejinger and Bierlaire, 2007; Schussler and Axhausen, 2010)*

Constructing a considering choice set for individual travelers is difficult. It should be determined the choice set according to the travel experience of the travelers. In previous researches, researchers employed the GPS data to observe the paths for individuals and to derive a route choice model using the questionnaire survey. There are limitations in constructing a route choice model using observed data without constructing various types of path sets.

- Impossibility for investigating the risk preference for individual travelers *(Chen et al., 2016)*

It is difficult to compare the perceived travel time distribution of the travelers and the choice situations in the actual route choice. Therefore, they defined the risk preference using assumption and analyzed the change of choice behavior. They derived the confidence level from the probability of arriving at a certain time using the travel time distribution. They analyzed the route choice probability by defining the individual risk preference according to the confidence level.

2.3.2 Limitations for route choice model

- Determinations of adequate choice set *(Bekhor et al., 2006; Frejinger and Bierlaire, 2007; Bekhor and Prato, 2009; Schussler and Axhausen, 2010)*

There is a limitation in determining an appropriate choice set to analyze the traveler's route choice behavior. Researchers employed the travel data generated by surveys or GPS observations, and they compared the results of applying for the route choice model. However, it was difficult to create an appropriate number of choice set for individual travelers and to apply the behavioral characteristics.

- Difficulties of considering the changes in choice set for individual travelers *(Schussler and Axhausen, 2010)*

There is a limit on the impossibility for generating the choice set reflecting the travelers' repeated experience. They analyzed the change of choice behavior according to the repeated questions in the experimental questionnaire surveys. There is a problem of biasedness for application of limited responses from the experiment of controlled repetitive questions. It is necessary to observe the actual route choice situation for individual travelers and the route choice behaviors.

- Estimation of the parameters by the result of simulation or questionnaire survey *(Prato and Bekhor, 2006; Hess and Rose, 2009; Frejinger et al., 2009)*

In previous studies, researchers estimated the parameters of the route choice model using simulation or questionnaire survey. This method estimates the parameters after analyzing the traveler's choice behavior with certain criteria. Moreover, the model draws a relatively high fitness index by limiting the behavioral parameters. It is difficult to estimate the perceived choice set and travel time in the previous researches.

2.4 Research Contributions

This research presents the contribution for improving the existing choice set generation and route choice model through reviewing the limitations of previous studies. The contribution of this study can be summarized as the following five categories.

- **Determinations for the size of consideration set & individual choice set using experienced travel data for individual traveler**

We define the relationship between a traveler's experienced path set and a master set. The model generates the choice set for each travelers according to the experienced paths and the considered path set. We employ a large amount of data to determine the choice set considering alternatives for individual traveler. We perform the process of generating a choice set reflecting different heterogeneities for individual travelers.

- **Applications of the different risk preferences for individual travelers in choice set generation model**

Researchers devised a method to derive the path travel time using accumulated travel data (Nassir et al., 2014; Srinivasan, 2014; Chen et al., 2016). However, they performed the estimation process using the integrated data or the non-refined data in the previous studies. Furthermore, there are studies on the relationship between risk preference and traffic assignment (Chen et al., 2010; Chen et al., 2013; Chen et al., 2014; Xu et al., 2014; Chen et al., 2016; Ji et al., 2017). These studies analyzed the sensitivities for traffic assignments according to the assumed risk preferences for travelers.

This research analyzes the risk preference comparing the travel time reliability of the network and individual travelers using the cumulative data of route travel time. Moreover, we established the choice set generation model using K-α-Reliable Shortest Path Searching algorithm (PRPSA-K**α**RSP) incorporating traveler heterogeneity.

- **Choice set generation model using revealed preference data for individual travelers**

It is necessary to generate the appropriate choice set in route choice model (Bliemer and Bovy, 2008). Researchers constructed a model to generate a choice set by tracking the observed paths for individual GPS data (Fosgerau et el., 2013; Nassir et al., 2014; Knapen et al., 2016; Chen et al., 2017).

In this study, we employed the Dedicated Short Range Communication (DSRC) data to identify individual routes. The data is possible to generate the route for individuals by storing the observed travel time. It is possible to identify the individual travel route by the personal identification number, and we can form the distribution of travel time by deriving the route travel time using the accumulated data for about 14 months. We derive the feasible choice set K and individual confidence level α_l using the revealed preference data in choice set generation. This study compares of the chosen route and choice set in the route choice model for individual travelers.

- **Development of stochastic route choice model incorporating traveler heterogeneity**

Researchers generated an appropriate choice set in the route choice model and analyzed the route choice behaviors in the previous research (Bekhor et al., 2006; Frejinger and Bierlaire, 2007; Schussler and Axhausen, 2010). They analyzed the route choice behavior according to the choice set generation model.

On the other hands, since the route choice model is constructed using SP data, it does not reflect actual travel behavior. There was a difficulty in constructing a choice model for repeated travel behaviors, and the model was constructed using the result of repeated SP questionnaire survey (Hess and Rose, 2009; Li et al., 2016). The model has difficulties of considering the changes in choice set for individual travelers (Schussler and Axhausen, 2010)

In this research, we construct the route choice model using K-α-Reliable Shortest Path Searching. We model for the route choice behavior using massive route traveled data for individual travelers. It estimates the precise route choice model by incorporating the travelers' heterogeneity in choice set generation model and route choice model. We establish the combined route choice model using risk preference and heterogeneity of travelers.

- **Possible to estimate for the more accurate parameter in route choice model**

The developed model reflects the characteristics of the network for the traveler’s perceived routes (the number of route experiences, the perceived travel time, the number of uninterrupted flow, Etc.). This model examines for the difference of route choice probability by traveler’s experience. The model made it possible to reflect the travel time budget (TTB) using the data for actual route travel.

In this study, we include a model validation process to examine the appropriateness of the model. We improve the accuracy of estimation according to the difference of choice probability.

Chapter 3. Modelling Framework

This section presents the framework of choice set generation model and the route choice model developed in this study. We summarize the process of choice set generation as presented in the previous research. Furthermore, it is necessary to define various important terms in this research to support the interpretation of this study. This study describes the choice set generation model reflecting the risk preference and develops the route choice model using the generated choice set.

3.1 Overview

We develop the following structure by dividing the process of choice set generation and route choice model. We construct a process to determine the size of consideration set and individual choice set. Consideration choice set is derived by the number of paths including 90% and the set of experienced paths using the observed data from the universal set occurring in the network for a specific OD pair.

In this study, we construct a model generating individual choice sets based on the distribution of perceived travel time using big data. We model a different choice set for individuals using travel time budget and risk preference in the process of individual choice set generation. We analyze the route choice model using the individual choice set derived from the choice set generation model. The individual choice set is a set of paths for incorporating traveler's heterogeneity. It is important to determine the choice set by the different travel behavior for individuals. Finally, we estimate and verify the model using the developed route choice model.

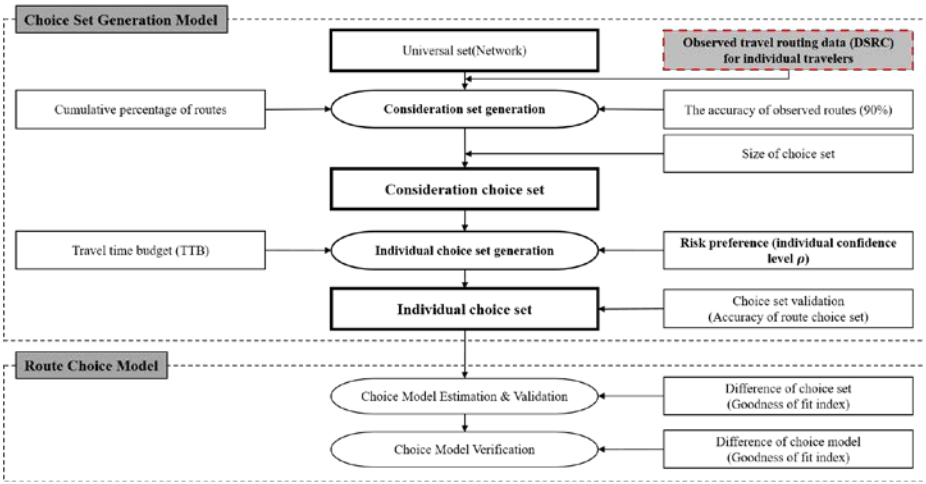


Figure 3.1 Process of route choice modeling

3.2 Terminology

The various terms used in this study need to be defined to meet the purpose of the study. We present various definitions used in previous studies for the terms used in the research. Furthermore, we employ the definitions of terms appropriate for this study. Travel time reliability, travel time variability, perceived travel time, travel time budget, risk preference, and other terms are defined in this chapter to understand the process of this study.

3.2.1 Travel Time Reliability

It is a term used in the degree of reliable for travel time in the network from traveler's perspectives. Travel time reliability is a term similar to travel time variability and is used in some studies to interpret the same meaning.

Asakura and Kashiwadani (1991) mentioned; "travel time reliability pertains to the probability that a trip between a given OD pair can be completed within a specified time" to define the probability of occurrence in travel time.

Chen et al. (1999, 2002) said that "the network can accommodate a certain demand at a given service level" to define the level of service for travel time in view of capacity. This meaning is useful for finding the reliable paths in choice set generation model.

Sumalee (2006) defined as "the probability of the total travel time in Reliable Network Design Problem (RNDP)." to understand the probability of occurrence in travel time.

Bates et al. (2001) defined the travel time reliability for “investigated day-to-day variations in travel time as the random variation in travel time.” They defined as a random change of travel time occurred in a day. It is used in a similar meaning to the travel time variability by focusing on temporal variability.

In this study, we tried to utilize the concept reflecting the variable characteristics of experienced travel time. Therefore, travel time reliability defined as “reasonably captures certain empirically observed features of user behavior such as the aversion to late arrival and the use of a time budget to maximize chances of on-time arrival.” (Recker et al., 2005; Srinivasan, 2014). This definition was used to emphasize the aspect of using travel time to avoid late arrivals from the traveler’s perspectives.

3.2.2 Travel Time Variability

Travel time variability is a term often used in the field of traffic management and is used as a term to describe the variable characteristics of travel time by day/hour occurring on links or route.

Bate et al. (2011) mentioned “An additional component of time, which represents the randomness in travel times over repeated trips.” This is defined as the concept of buffer time by defining it as an additional time to the average travel time occurring in the travel time.

Li et al. (2010) defined as “A feature of transport systems, which adds additional costs and uncertainty to travelers.” It is defined as the additional costs that passengers have to pay for the traffic system.

This study focuses on daily variability and therefore takes into account the distributional characteristics of travel time. Travel time variability is defined as “At varying levels from different perspectives, namely, day-to-day variability, within-day or period-to-period variability and vehicle-to-vehicle variability” (Noland and Polak, 2002; Yildirimoglu et al., 2013) This is defined as the variation of the changing travel time in the situation. This study defines travel time variability as the various observed travel time in various situations to explain the difference in observed travel time.

3.2.3 Perceived Travel Time

Individual perception has a very important meaning in a study based on the theoretical background reflecting individual characteristics. The meaning of "perceived" can be interpreted in various meanings, and each research interprets it in various ways according to how it reflects the perception of the individual. Various studies have conducted to understand the cognitive behavior of individuals.

Yang and Jiang (2014) defined the perceived travel time as “Average travel time at different times is commonly utilized to express perceived travel time considering that travelers predict the travel time of the path based on previous experiences.” The average travel time from the previous experience is expressed as a representative value. They tried to identify the route choice behavior of individual perceived travel time through prospect theory.

Tang et al. (2017) mentioned; “The only attribute that evolves over time and other attributes are assumed fixed from day to day.” Travelers set for the perceived travel time as a constant attribute, and they analyzed the route choice model through the change of other variables.

Vovsha (2011) defined as “A measure has been used to quantify travel time variability in practice including congestion multi-attribute utility functions including early and late schedule delays, trip time variability, and temporal utility profiles for activity participation.” It is an indicator value that quantifies the variation of travel time according to various influence factors. It is explained by the index value according to the characteristics of the network rather than the cognitive characteristic of the individual.

Lo and Tung (2003), Lo et al. (2006), Chen et al. (2011) suggested that the meaning of perceived travel time is “Travelers are assumed to have the ability to learn the travel time variability through their past experiences. Then, they incorporate this knowledge into their daily route choice decisions to reach a habitual equilibrium.” They defined the perceived travel time as the value used to acquire the travel time variation from experience. This definition is employed to study the network equilibrium state.

The perceived travel time used in this study is calculated to refer to the traveler's perception of the accumulated travel time. Therefore, we define the perceived travel time as “A general equilibrium condition is derived for a network whose travel times arise from arbitrary probability distributions, with travelers' choosing their route based on subjective perceptions of the route travel time distributions" (Connors and Sumalee, 2009). The purpose of this study is to analyze travel behavior by constructing the subjective perception of the route in the form of route travel time distribution.

3.2.4 Travel Time Budget (TTB)

Travel time budget represents the characteristics of travel time reflecting the travel time reliability for stochastic deviations apart from the various shortest path searching algorithms using average travel time. It is used as the required time to arrive at the destination to the on-time arrival probability.

Hall (1983) defined as "The route choice problem has to consider the summation of the mean route travel time and the route-specific safety margin, which is referred to as the effective travel time." The researcher defined the travel time budget for the average travel time and the additional travel time as the efficient travel time.

Lo et al. (2006) introduced for definition of travel time budget as "A function of several factors, including (a) the stochastic nature of the network link capacity variations or degradations, (b) the probability desired or required for within budget-time arrivals, referred to as the within budget time reliability (WBTR), and (c) other travelers' route choices." They defined the travel time budget as a function value determined by various variables.

Ji et al. (2017) mentioned that "In order to ensure the on-time arrival, travelers will add some buffer time in the TTB model, which can be formulated as the following chance constraint programming". They derived the SUE using the route choice model by programming the process to add the additional time for on-time arrival.

The definition of travel time budget is "The minimum total travel time threshold that satisfies a chance constraint reliability requirement, that the percentile of total travel time distribution concerning the reliability requirement specified by decision-makers using the confidence level α." (Xu et al., 2014). They defined the travel time budget using the previously defined confidence level α as the value derived from the distribution of total travel time. On the other hand, the travel time budget is derived from the distribution, which is determined by the individual confidence level α_l. Therefore, we include the distributional characteristics and the individual confidence level α_l.

3.2.5 Risk preference

The risk preferences are related to the degree of enduring the risk of late arrival. It is an individual behavioral characteristics occurring in route choice decision.

Erev and Barron (2003), Avineri and Prashker(2011) suggest that the risk preference is "risk-taking behavior of decision makers is characterized by risk aversion in the gain domain and risk seeking in the loss. Such a risk-taking pattern is reversed in iterated tasks (risk seeking in the gain domain and risk aversion in the loss)." They defined the risk taking property using the concept of gain and loss in the function used in the prospect theory.

Avineri (2006) mentioned that “Equilibrium under the condition that no user can increase his/her route prospect value by unilaterally switching routes.” The definition is employed to define the equilibrium state describing the route choice behavior for individuals from prospect theory.

In this study, the risk preference is the risk taking property for travel failure or delay due to travel time reliability. Therefore, we used the definition to show the characteristics of travelers on arrival at scheduled time (travel time budget; TTB). “On-time arrival probability (i.e., α parameter) is also referred as travel time reliability (Bell and Iida, 1997). $\alpha \in (0, 1)$ reflects travelers’ risk taking attitude towards being late, where $\alpha > 0.5$, $\alpha = 0.5$, and $\alpha < 0.5$ represent risk-averse, risk-neutral, and risk-seeking attitudes, respectively.” (Chen et al., 2016). The risk preference is determined by the travel time reliability having the different characteristics for individuals.

3.2.6 Other Terms

3.2.6.1 Path & Route

The terms of path and route need to define for this research. The term of “Path” is used for Heterogeneous K-α-Reliable Shortest path algorithm to search k paths in the model. On the other hand, the term of “Route” is employed for route choice model to estimate the probability from k paths and use of choice set in the model. The term of “Path” is a term used to find k paths, and the term of “Route” is used to derive the choice probability among the selected choice set. In this study, the terms of path and route have the same meaning to establish the combined model for choice set generation and route choice behavior.

3.2.6.2 Confidence level & Individual confidence level

Confidence Level is the value of on-time arrival probability for cumulative distribution from node i to node j in the network by the result of repeated travels. On the other hand, when travel time budget (TTB) of individual on confidence level α has the same TTB of the network from node i to node j, the TTB represent the individual confidence level α_l on the cumulative distribution of network.

Confidence level α is used for on-time arrival with cumulative travel time distribution in the network, otherwise individual confidence level α_l represent the traveler's risk preference in the network from node i to node j by his/her perceived travel time.

3.3 Choice set generation model

In this study, the distribution of individual travel time is set up based on the observed travel time to derive the individual route choice set. We consider the set of different paths according to individual perception for route. We generate the individual choice sets using the travel time budget and risk preference. We define the travel characteristics for individuals using the on-time arrival probability due to the heterogeneous difference between travel time budget for individual and the travel time provided by the network.

3.3.1 Consideration set generation

Determining the choice set in the route choice model is an important process for changing the accuracy of the prediction (Bliemer and Bovy, 2008). It is important to know what routes should be chosen in the network. We construct the set of considered paths for individual travelers and a set of paths from the researcher's perspective. In addition, we derive the size of the consideration set and individual choice set for individual travelers using actual travel data.

3.3.1.1 Choice set notions for individual traveler's perspective

There are four categories of choice set for individual perspectives divided into master choice set, consideration paths, experienced paths, and chosen path. Travelers identify the choice set for their travel by traveling the known routes and finding the alternative routes. We are able to identify the master choice set for a specific OD pair. Since it may be the millions of alternatives in the network, it is very time-consuming task to analyze the choice set. In addition, travelers does not consider the enormous size of choice set for travel. Therefore, we do not have to consider the master choice set in choice set generation model.

We have to generate the appropriate size of choice set to estimate the choice probability. There are experienced paths for determining the proper size of choice set, which is possible to know the exact paths for each traveler. Moreover, the single alternative chosen by traveler belongs to the experienced paths.

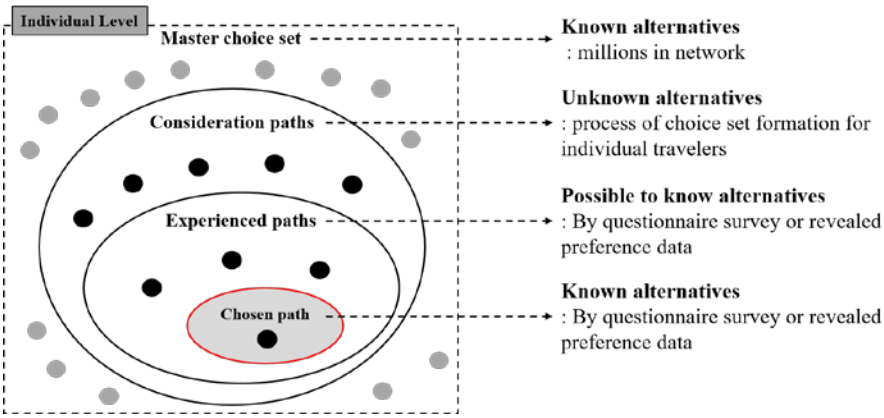


Figure 3.2 Structure of choice set for traveler’s perspectives

3.3.1.2 Choice set notions for researcher’s perspective

Fiorenzo-Catalano (2007) mentioned that it is important to determine the choice set considering researcher’s perspectives. There are differences between traveler’s perspectives and researcher’s perspectives. Since researcher does not know the choice set for individuals, it needs some of the assumptions in choice set generation model. We employ the two kinds of sets, which are the consideration set and individual choice set. Consideration set includes the possible to choose the paths for the most of travelers. In addition, individual choice set has the proper size of the set for individual travelers in route choice behavior.

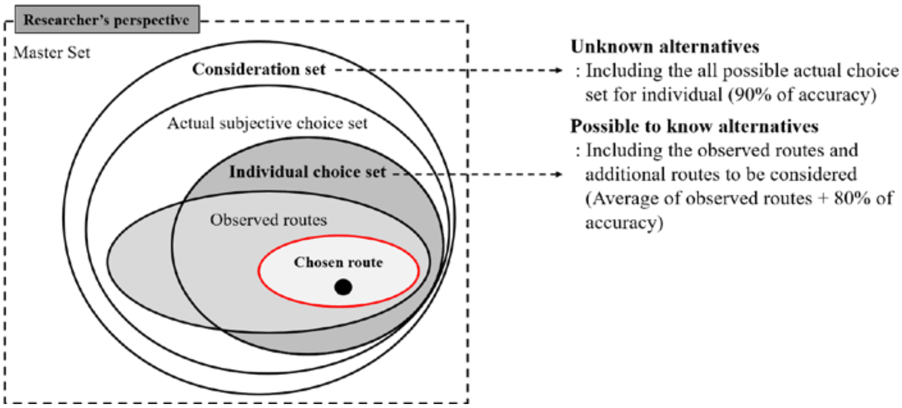


Figure 3.3 Structure of choice set for researcher's perspectives

We formulate the consideration set by including the all of possible actual choice set for most travelers. We set up the 90% of accuracy in chosen route. There are discrepancies in observed routes and individual choice set. We formulate the size of individual choice set using the actual number of observed routes and the 80% of accuracy.

3.3.2 Mathematical definition of travel time budget

It is necessary to define the distribution type of travel time in order to define the travel time budget (TTB). Xu et al. (2014) define the TTB as a concept including a margin of travel time determined by the assumed confidence level α. The value of TTB depends on the type of distribution of travel time and confidence level α. Researchers define the distribution of travel time in various types. Chen and Lam (2013) analyzed the distribution of travel time as the normal distribution and the difference of the traffic assignment according to TTB. Furthermore, Polus (1979) analyzed the distribution of travel time as the gamma distribution, and Nie and Wu (2009) analyzed the traffic assignment model using the travel time distribution. Weibull distribution model (Al-Deek and Emam, 2006) and burr distribution model (Susilawati, 2013) have been applied to analyze the TTB. In this study, the distribution of travel time is assumed to log-normal distribution, and various studies have assumed the log-normal distribution of travel time (Herman and Lam, 1974; Richardson and Taylor, 1978, Srinivasan, 2014). We analyze the distribution fitness of the actual route travel time to verify the distribution of travel time. We evaluate the fitness of the travel time distribution using Kolmogorov-Smirnov test (KS test). The following figure is an analysis of the type of travel time distribution using observed travel data. The KS test shows the log-normal distribution having the appropriate model fit.

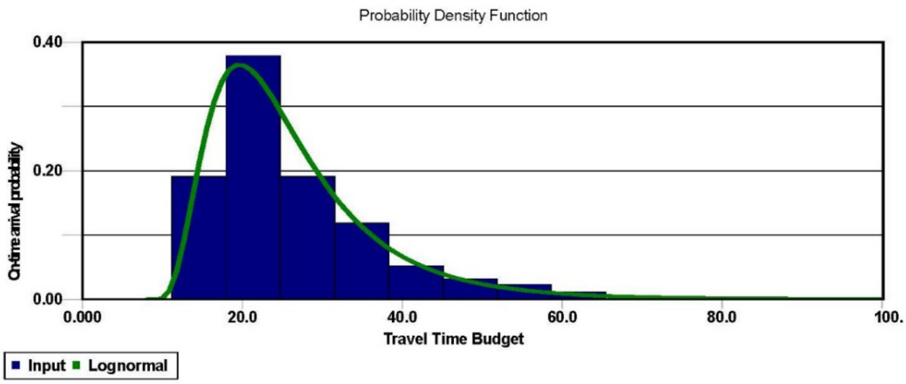


Figure 3.4 Distribution for travel time

Table 3.1 Descriptive statistics

CONTENTS	VALUE	CONTENTS	VALUE
Sample	346	Mean	26.3
Minimum	11.1	Median	23.1
Maximum	85.6	Mode	17.6
Skewness	1.6	Standard Deviation	10.7
Kurtosis	3.6	Coefficient of Variation	40.8

Table 3.2 Goodness of fit test

CONTENTS	KS TEST		RESULT
	STATISTICS	P-VALUE	
Log-normal Dist.	0.038	0.668	Do not Reject
Log-logistics Dist.	0.047	0.400	Do not Reject
Gamma Dist.	0.058	0.188	Do not Reject
Weibull Dist.	0.067	0.087	Do not Reject
Normal Dist.	0.136	0.000	Reject

※ KS Test: Kolmogorov-Smirnov Test Standard: 0.0725

The TTB is defined as the structure as shown in the figure below, in which the TTB is determined according to the confidence level α. TTB has a structure combined with the predictable travel time in the travel time distribution. Therefore, the model structure excluded for the unpredictable increase in travel time. There are two definitions of travel time budget (TTB) in probability density function (PDF) and cumulative distribution function (CDF) in below.

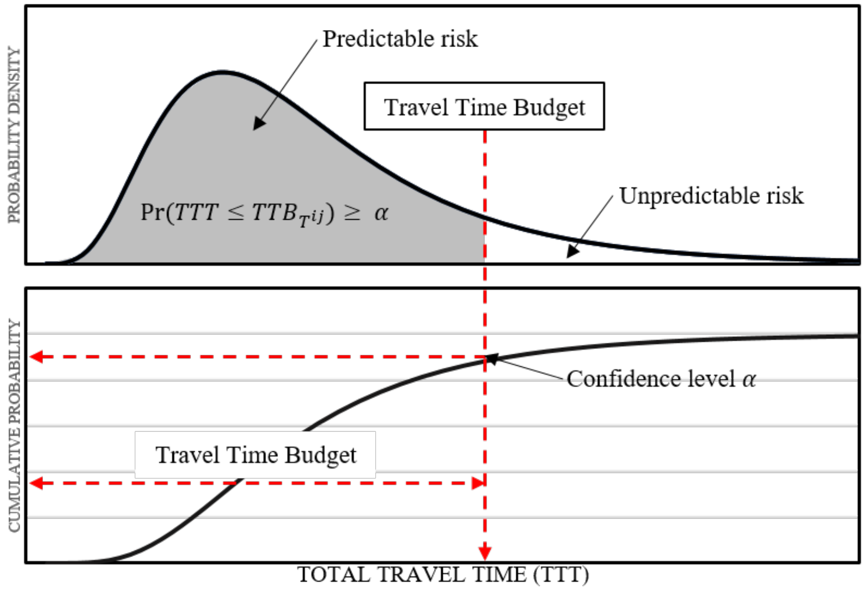


Figure 3.5 Definition of Travel Time Budget

As described above, we analyze the travel time distribution of the route in the form of log-normal distribution. The fitness of distribution is performed using the mean and standard deviation of the path travel time as follows:

Route travel time (Mean: μ, Standard Deviation: σ)

→ *Log-normal distribution:* $ln\,X \sim \mathcal{N}(\mu,\,\sigma)$

The probability density function using log-normal distribution is defined as follows.

$$\textit{Probability Density Function (PDF):}\;\; PDF = \frac{1}{\pi\sqrt{2\pi\sigma}} exp\left(-\frac{(ln\,x-\mu)^2}{2\sigma^2}\right)$$

The CDF can be constructed as a cumulative PDF to derive travel time budget and confidence level α.

Cumulative Density Function (CDF):

$$CDF = \frac{1}{2}\left[1 + erf\left(\frac{ln\,x-\mu}{\sqrt{2}\sigma}\right)\right] = \Phi\left(\frac{ln\,x-\mu}{\sigma}\right) = \alpha$$

$$\textit{Error function:}\;\; erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt \;\rightarrow \textit{Tailor Series:}$$

$$erf(x) = \frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{(2n+1)n!} = \frac{2}{\sqrt{\pi}}\left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \cdots\right)$$

Here, we use the inverse function of CDF to obtain TTB. We employ the inverse of the normal distribution to derive the inverse function of log-normal distribution. The inverse of the normal distribution is defined as:

$$\textit{Normal distribution's CDF:}\;\; \Phi(x) = \frac{1}{2}\left[1 + erf\left(\frac{x}{\sqrt{2}}\right)\right]$$

$$\Phi\left(\frac{ln\,x-\mu}{\sigma}\right) = \alpha \;\rightarrow\; \Phi^{-1}(\alpha) = \frac{ln\,x-\mu}{\sigma}$$

The TTB according to the log-normal distribution can be defined as follows.

$$\textit{Travel Time Budget (TTB):}\;\; x = exp(\mu + \Phi^{-1}(\alpha) \times \sigma)$$

The concept of TTB used in this study can be divided into three types.

- Travel time budget(TTB) required achieving α confidence level in the network from the origin node i to the destination node j $TTB_{T^{ij}}(\alpha)$

$$\mathrm{TTB}_{\mathrm{T}^{\mathrm{ij}}}(\alpha) = \exp\left(\mu^{\mathrm{ij}} + \Phi^{-1}_{\mathrm{T}^{\mathrm{ij}}}(\alpha) \times \sigma^{\mathrm{ij}}\right)$$

- Travel time budget(TTB) of k^th path required achieving α confidence level from the origin node i to the destination node j, $\mathrm{TTB}_{\mathrm{T}^{\mathrm{ij}}_{\mathrm{k}}}(\alpha)$

$$\mathrm{TTB}_{\mathrm{T}^{\mathrm{ij}}_{\mathrm{k}}}(\alpha) = \exp\left(\mu^{\mathrm{ij}}_{\mathrm{k}} + \Phi^{-1}_{\mathrm{T}^{\mathrm{ij}}_{\mathrm{k}}}(\alpha) \times \sigma^{\mathrm{ij}}_{\mathrm{k}}\right)$$

- Travel time budget(TTB) required achieving α confidence level for individual l from the origin node i to the destination node j, $TTB_{T^{ijl}}(\alpha)$

$$\mathrm{TTB}_{\mathrm{T}^{\mathrm{ijl}}}(\alpha) = \exp\left(\mu^{\mathrm{ijl}} + \Phi^{-1}_{\mathrm{T}^{\mathrm{ijl}}}(\alpha) \times \sigma^{\mathrm{ijl}}\right)$$

Where, i = the origin node

j = the destination node

k = the order of α-reliable path or predetermined the number of route choice set

l = the individual traveler

α = Confidence level α (i.e., On-time arrival probability)

μ^{ij} = the mean of travel time distribution from the origin node i to the destination node j

μ^{ij}_{k} = the mean of travel time distribution of kth α-reliable path from the origin node i to the destination node j

μ^{ijl} = the perceived mean of travel time distribution for individual l from the origin node i to the destination node j

σ^{ij} = the standard deviation of travel time distribution from the origin node i to the destination node j

σ_k^{ij} = the standard deviation of travel time distribution of kth α-reliable path from the origin node i to the destination node j

σ^{ijl} = the perceived standard deviation of travel time distribution for individual l from the origin node i to the destination node j

3.3.3 Distributional characteristics for risk preferences

Individual risk preference plays an important role in the choice set generation model. The process of choice set generation modeling is formulated using the individual confidence level α_l, which is referred to as risk preference. Risk preferences have three forms according to the definitions in the previous studies. There are three types of risk preference, risk seeking (α<0.5), risk neutral (α=0.5), and risk averse (α>0.5) concerning the individual confidence level. In this study, we set up a route choice set which is determined according to different risk preferences. The individual confidence level is determined according to the time difference between the perceived travel time of the individual and the travel time provided by the network, and thus a difference occurs in generating the choice set.

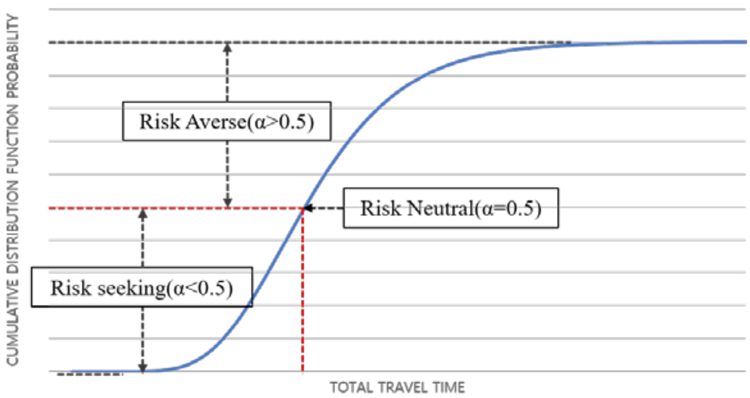


Figure 3.6 Risk preferences

Table 3.3 Comparison of the confidence level and individual confidence level

DIVISION	**Confidence Level α**	**Individual Confidence Level α_I**
Expression	Probability (fractional representation)	Probability (fractional representation)
Value	$0<\alpha<1$	$0<\alpha_I<1$
Viewpoint	Operator (Network)	Traveler (Individual)
Decision	Prior determination	Estimation
Process	Determined by operator or researcher	Estimated by the distribution of experienced travel time for individual traveler
Characteristics	Fixed value Effect on the estimation for individual confidence level	Different value for individual travelers Effect on the determination of route choice set Risk preference
Researches	Chen et al. (2013) / Lo et al. (2006) / Chen and Ji (2005) / Shao et al. (2006), Siu and Lo (2006)	This study (2017)

We construct a route choice model reflecting the behavior of travelers according to risk seeking and risk averse. We define risk-seeking travelers having the less travel time budget than that of travel time provided in the network. When the travel time budget for the individual is derived from the confidence level α according to the average and standard deviation in the travel time distribution, we compare the travel time budget for individuals and network travel time distribution with the confidence level α. In other words, when the experienced travel time for individual is less than the travel time in the network, the travelers have a concern of late arrival by the perceived travel time. This case is defined as the risk-seeking characteristic.

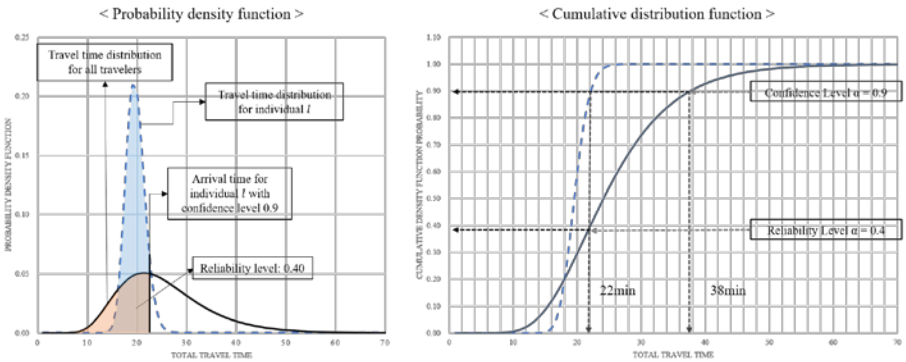


Figure 3.7 Distributional characteristics for risk seeking travelers

The traveler has the mean of perceived travel time less than the network's travel time, and if he/she has the same TTB of perceived travel time in travel, he/she might be afraid of late on time arrival. According to the above figure, the traveler perceived the travel time budget to 22 min with confidence level 0.9. On the other hand, the distribution of travel time in the network (observed travel time for all travelers) provide the individual confidence level 0.40 to traveler A. $TTB_{Tijl}(0.9) = TTB_{Tij}(0.4)$ represent the travel time to 22min in the distribution, the individual A's individual confidence level α is determined to 0.4 with risk seeking property.

The risk averse travelers have the characteristics that occur when the distribution of the individual's travel time is larger than the distribution of the travel time provided by the network according to confidence level α. The distribution of individual travel times is shifted to the right than the distribution of network travel time. The experiences of individual traveler have the more travel time than the time provided by the network because he/she has experienced the more travel time. Risk preference makes traveler calculate the travel time budget to arrive on time.

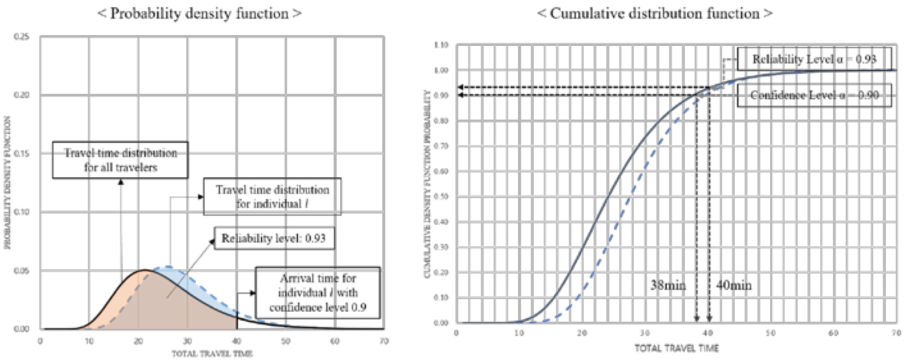


Figure 3.8 Distributional characteristics for risk averse travelers

The traveler has the mean of perceived travel time more than the network's travel time, and if he/she has the same TTB of perceived travel time in travel, he/she might need not to be late on time arrival. From the above figure, the traveler recognizes the travel time to the destination for 40min from the distribution of perceived travel time with confidence level 0.9. Otherwise, the distribution of travel time in the network (observed travel time for all travelers) offer the individual confidence level 0.93 to traveler B. $TTB_{T^{ij}}(0.9) = TTB_{T^{ijl}}(0.93)$ have the travel time to 40min in the distribution, the traveler B's individual confidence level α is set for 0.93 as risk averse characteristic.

Table 3.4 Comparison the confidence level and individual confidence level

DIVISION	Chen et al. (2013)	Xu et al. (2011)	This study (2017)
Travel time reliability	Confidence level α (travel time reliability)	Travel time reliability ρ	Confidence level α Individual confidence level $\boldsymbol{\alpha_l}$
Viewpoint	Researcher	Traveler	Traveler
Attribute	Risk preference (Risk seeking / neutral / averse)	Risk preference (Risk seeking / neutral / averse)	Risk preference (Risk seeking / neutral / averse)
Estimating methodology	Assumption	Estimation	Estimation
Characteristics	K-reliable shortest path searching based path generation	Prospect theory based travel time reliability	Determination for individual route choice set generation by risk preference
Result	Various three routes different from the risk preference	90% of travel time reliability to avoid risk for late	Different route choice set for individual travelers by risk preference

3.3.4 K-α-Reliable Shortest Path Searching algorithm

We construct an analysis algorithm to generate a choice set for individuals using the TTB and risk preference. The individual confidence level α_l derives the choice set for individual travelers. This algorithm generates a set of considered paths using the cumulative distribution of travel time. A set of individual paths is determined according to the number of paths set specified in advance.

It is necessary to prove K paths for each individual mathematically. The TTB using the individual confidence level α should always be smaller than the unchosen paths for the individual alternative paths. These properties can be expressed as follows.

$\textit{Travel Time Budget (TTB) of route K for OD in network}:\ TTB_{T_k^{ij}}(\alpha) =$

$\boldsymbol{exp\left(\mu_k^{ij} + \Phi_{T_k^{ij}}^{-1}(\alpha) \times \sigma_k^{ij}\right)}$

$\textit{Given integer } K \geq 1\textit{, travel time individual confidence level } \alpha \in (0,1)$

- $TTB_{T_k^{ij}}(\alpha) \leq TTB_{T_{k+1}^{ij}}(\alpha)$, $\forall k \in (1, \cdots, K-1)$: *other path among kth paths is smaller than or equal to kth path*
- $TTB_{T_K^{ij}}(\alpha) \leq TTB_{T_U^{ij}}(\alpha)$, $\forall U \in (K+1, \cdots, U)$: *the other paths without k paths is bigger than or equal to kth path*

The above two assumptions must be satisfied to generate the path set. One assumption is that K paths must have a smaller travel time budget than the next K+1th travel time budget. The second assumption is that the Kth path must be less than or equal to the travel time budget for the path from K + 1th to the Uth path.

The K-a-reliable Shortest Path Searching algorithm for generating individual choice set consists of eight steps as shown in the figure below. First, the observed travel time is extracted from the database for a specific OD pair. Next, we set the confidence level α for the travel time reliability to achieve the network performance. It needs to calculate the travel time distribution in the network and the TTB for deriving the individual confidence level α_l. Next, the travel time distribution is formed for each route. Then, we calculate the individual confidence level α according to the individual travel time distribution. At last, we derive the choice set by calculating TTB according to individual confidence level α_l. In this paper, we propose the path choice set generation procedure as K-α-Reliable Shortest Paths (PRPSA-KαRSP).

Step 1.	• Choosing the origin and destination to observe the path (i, j).
Step 2.	• Setting the confidence level, α_{ij} for the satisfaction in the level of service, i.e. α=0.9
Step 3.	• Building the distribution of travel time (T^{ij}) for the whole of path travel from origin *i* to destination *j*, and calculating the travel time budget $(TTB_{T^{ij}}(\alpha))$, with respect to confidence level, α.
Step 4.	• Building the distribution of travel time for kth path (T_k^{ij}), $\forall k \in (1, \cdots, U)$, for the each of observed travel from origin *i* to destination *j*.
Step 5.	• Building the distribution of travel time for individual l (T^{ijl}), for the each of travelers from origin *i* to destination *j*.
Step 6.	• Evaluating the reliability level α^{ijl} for each traveler l, $\alpha^{ijl} = \Phi_{T^{ijl}}\left(\frac{ln(TTB_{T^{ij}}(\alpha_{ij}) - \mu^{ijl}}{\sigma^{ijl}}\right)$, normal distribution $\Phi(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$, erf=error function.
Step 7.	• Calculating the travel time budget for each path $(TTB_{T_k^{ij}}(\alpha^{ijl}))$ with respect to the reliability level, α^{ijl}.
Step 8.	• Choosing the K-α-reliable shortest paths for individual *l*.

Figure 3.9 PRPSA-KαRSP

3.4 Route choice model

We provide the route choice model with the proper assumption in the model. As mentioned before, there are several problems of overlapping links in route choice model. It is necessary to propose the appropriate model form. Moreover, we suggest the model incorporating traveler heterogeneity in route choice behavior. There are four kinds of models compared in this study.

3.4.1 General models and overlapping links

Multinomial Logit Model(MNL) is the general model form of choice modeling. Researchers developed a model for the use of analysis on choice behavior. MNL model has been employed in the route choice model to analyze the disaggregated choice behavior. The expression of MNL function is in below.

$$\mathrm{P(i)} = \frac{e^{-\beta C_i}}{\sum_{j \in A_n} e^{-\beta C_j}}$$

MNL model is sensitive only to the relative scale of an attribute for the routes. Since MNL model assumed the i.i.d Gumbel distribution in perception of route travel time, the model does not consider the correlations between alternative routes.

Researchers tried to develop the improved model form in overlapping problem. They developed the Path-size Logit Model (PSL) for the improved MNL model considering the degree of overlapping links. (Ben-Akiva and Ramming, 1998)

$$P(i|A_n) = \frac{e^{V_i + \ln PS_{in}}}{\sum_{j \in A_n} e^{V_j + \ln PS_{jn}}} = \frac{PS_{in} e^{V_i}}{\sum_{j \in A_n} PS_{jn} e^{V_j}}$$

Ben-akiva and Ramming (1998) presented the equation for the path size resistance in below.

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{N_{an}} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in A_n} \left(\frac{\bar{L}}{L_i}\right)^{\gamma} \delta_{aj}}$$

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in A_n} \frac{L_i}{L_j} \delta_{aj}}$$

The path-size factor is a value of calculation for the overlapping links considering the specific attribute. The overlapping links have the size of one attribute, which is allocated among the routes using the overlapping links. The route size is the sum of link sizes weighted by the scale of the overlapping links contributing to the overall routes. Certain links affect the overall size of the routes including the nested links. The size of the total routes is also affected by the size of the non-overlapping links of the other routes. The more overlapping links makes the less size of the path-size factor.

Bovy et al. (2008) proposed the improved path-size logit model. Since there is no satisfactory derivation based on theoretical arguments, it is necessary to employ the correction terms for the specification. The model considers the impact of choice set in the route choice model.

$$P(i|A_n) = \frac{e^{V_{in} + \beta_{PSC} * PSC_{in}}}{\sum_{j \in A_n} e^{V_{jn} + \beta_{PSC} * PSC_{jn}}}$$

$$PSC_i = \ -\sum_{a \in \Gamma_i} \left(\left(\frac{l_a}{L_i} \right) ln \sum_{j \in A_n} \delta_{aj} \right)$$

Researchers devised the GNL model to generalize the CNL model by employing the nesting parameters. Bekhor (2001) introduced the nesting coefficient in below:

$$\mu_m = \ \left(1 - \frac{\sum_{l \in C_n} \alpha_{mi}}{\sum_{l \in C_n} \delta_{mi}} \right)^{\gamma}$$

Wen and Koppelman (2001) mentioned the model with new specific parameter μ as “Generalized Nested Logit model.”

$$P\big((i|C_{mn})\big) = \frac{(\alpha_{mi} e^{V_{in}})^{\mu_m}}{\sum_{j \in C_{mn}} \left(\alpha_{mj} e^{V_{jn}} \right)^{\mu_m}}$$

Since the improved models are dependent on the generated choice set, the choice set generation model is important to estimate the actual choice probability. There are four route choice models considered in this study.

Table 3.5 Comparison of the logit models

DIVISION	Multinomial Logit Model	Path-size Logit Model	Path-size correction logit model	Generalized nested logit model
Description	Basic Route Choice Model	Route Choice Model Considering Overlapping links		
Equation	$P(i\|A_n) = \frac{e^{V_{in}}}{\sum_{j\in A_n} e^{V_{jn}}}$	$P(i\|A_n) = \frac{e^{V_i + ln\, PS_{in}}}{\sum_{j\in A_n} e^{V_j + ln\, PS_{jn}}} = \frac{PS_{in}e^{V_i}}{\sum_{j\in A_n} PS_{jn}e^{V_j}}$ $PS_i = \sum_{a\in\Gamma_i}\left(\frac{l_a}{L_i}\right)\frac{1}{\sum_{j\in A_n}\delta_{aj}}$	$P(i\|A_n) = \frac{e^{V_{in}+\beta_{PSC}*PSC_{in}}}{\sum_{j\in A_n} e^{V_{jn}+\beta_{PSC}*PSC_{jn}}}$ $\mathrm{PSC_i} = -\sum_{a\in\Gamma_i}\left(\left(\frac{l_a}{L_i}\right)\ln\sum_{j\in A_n}\delta_{aj}\right)$	$P((i\|C_{mn})) = \frac{(\alpha_{mi}e^{V_{in}})^{\mu_m}}{\sum_{j\in C_{mn}}(\alpha_{mj}e^{V_{jn}})^{\mu_m}}$ $\mu_m = \left(1 - \frac{\sum_{l\in C_n}\alpha_{mi}}{\sum_{l\in C_n}\delta_{mi}}\right)^{\gamma}$
Characteristics	- Assume independence between alternatives - Difficulty to apply the route choice model that overlaps many links	- Modification of MNL model considering overlapping links - The probability of choice changes depending on the size of the overlapping link	- Modification of PSL model to reflect the consideration set of paths	- Adopting the different nesting parameter for each nest

For example, there are four routes with overlapping links in below figure. The routes have overlapping links with each other for specific OD pairs, and the degrees of overlap between the different paths are different from each other. In this example, we consider the length of the links to calculate the overlapping factors in the model. We employ the attribute of travel cost for link length to calculate the route choice probability.

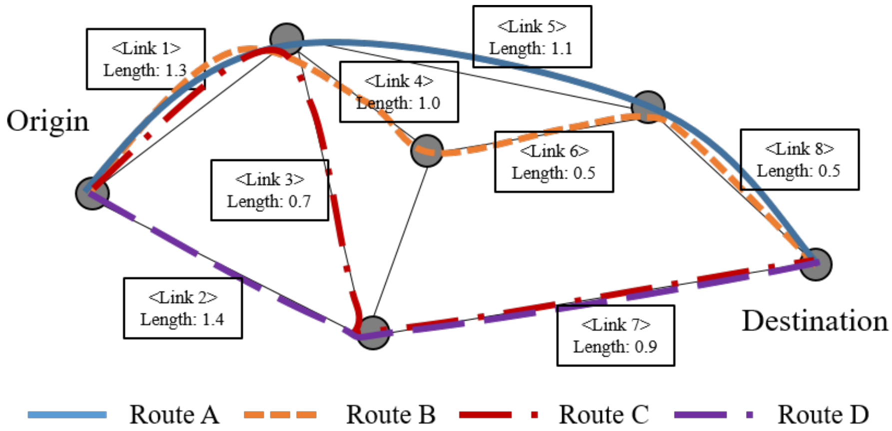


Figure 3.10 Example network

There is the estimated value for factors considering the overlapping links shown in Table 3.6. We present the process of calculation for the route 1 (A).

- $PS_{in} = \frac{1.3}{2.9}\left(\frac{1}{\frac{3.3}{2.9}+\frac{2.9}{2.9}}\right) + \frac{0.5}{2.9}\left(\frac{1}{\frac{3.3}{2.9}}\right)$ = 0.359
- $PSC_{in} = -\left(\frac{1.3}{2.9}ln(2) + \frac{0.5}{2.9}ln(1)\right)$ = -0.311
- $Nesting\ coefficient_{in} = 1 - \left(\frac{\frac{1.3}{2.9}+\frac{0.5}{2.9}}{3}\right)$ = 0.792

Table 3.6 Calculation of the coefficients for overlapping

Division	Route		length	MNL	PSL	PSCL	GNL
Route 1	**A**	**1-5-8**	**2.9**	-	0.359	-0.311	0.792
Route 2	**B**	**1-4-6-8**	**3.3**	-	0.397	-0.271	0.819
Route 3	**C**	**1-3-7**	**2.9**	-	0.611	-0.304	0.747
Route 4	**D**	**2-7**	**2.3**	-	0.320	0.000	0.598

Table 3.7 shows the choice probabilities for the four routes using the improved models considering the overlapping links. The parameters of route length (β) is set for the -0.5 in this example.

Table 3.7 Choice probability for the four models (β = -0.5)

Division	Route		length	MNL	PSL	PSCL	GNL
Route 1	**A**	**1-5-8**	**2.9**	0.240	0.208	0.253	0.232
Route 2	**B**	**1-4-6-8**	**3.3**	0.196	0.188	0.203	0.187
Route 3	**C**	**1-3-7**	**2.9**	0.240	0.354	0.252	0.237
Route 4	**D**	**2-7**	**2.3**	0.324	0.250	0.292	0.344

Route 1 and 3 have the coincidence probability in MNL model due to having the same route cost attributes (link). Even though there are some of overlapping links, the choice probability comes out for the same result. There are three models considering the overlapping links. The four models have the different results of probability due to the factors considering route overlapping properties. We determined to use of path-size correction logit model in overlapping problem because the model is the improved logit based model.

3.4.2 Mixed logit based route choice models

The error term is divided into two parts in the model. One part represents correlation and heterogeneity, and the other part describes i.i.d extreme value. Researchers proposed mixed logit model to overcome the defect of logit model by adding normal error term in the equation to account for the correlation between routes. Because traveler's perceived routes are correlated, the error term is added to illustrate the relationship based on the topology of paths. The equations for MMNL and MPSCL are presented in the below

$$\boldsymbol{Var(\varepsilon) = FTT^T F^T} + \left({}^{\boldsymbol{g}}\!/_{\boldsymbol{\mu}^2}\right)\boldsymbol{I_{J_n}}$$

$$\sum_{\boldsymbol{n}} = \boldsymbol{Var(FT\xi)} = \boldsymbol{\sigma^2}\begin{bmatrix} \boldsymbol{L_{11}} & \cdots & \boldsymbol{L_{1J_n}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{L_{J_n 1}} & \cdots & \boldsymbol{L_{J_n J_n}} \end{bmatrix}$$

MMNL: $\boldsymbol{P_n(i)} = \boldsymbol{\Lambda(i|\xi)} = \frac{\boldsymbol{exp(\mu(X_{in}\beta + F_{in}T\xi))}}{\sum_{\boldsymbol{J\in C_n}} \boldsymbol{exp(\mu(X_{jn}\beta + F_{jn}T\xi))}}$

MPSCL: $\boldsymbol{P_n(i)} = \boldsymbol{\Lambda(i|\xi)} = \frac{\boldsymbol{exp(\mu(X_{in}\beta + F_{in}T\xi) + ln(PSC_{in}))}}{\sum_{\boldsymbol{J\in C_n}} \boldsymbol{exp(\mu(X_{jn}\beta + F_{jn}T\xi) + ln(PSC_{jn}))}}$

Where, **U** = the C_n by 1 vector of utility function

X = the C_n by *K* matrix of variables

β = the column vector of *K* unknown parameters for variables

ε = the C_n by 1 vector of error terms

F = the C_n by *M* factor loading matrix

T = the *M* by *M* lower triangular matrix of unknown parameters

ζ = the *M* by 1 vector of i.i.d standard normal variables as unobservable factors

$\boldsymbol{\nu}$ = the M by M lower triangular matrix of unknown parameters

$\Gamma(k|\zeta)$ = the probability of chosen route k with given ζ.

In this study, we compare the multinomial logit model, path-size correction logit model, mixed logit model, and mixed path-size correction logit model based on the three kinds of choice set generation model to verify the developed model.

Chapter 4. Revealed Preference Routing Data

4.1 Data Characteristics

There are requirements for data to model the distribution of perceived travel time according to the methodology presented in the previous section and to generate the individual choice set. This section presents these requirements and identify whether the data meets the conditions with a description of the data used.

4.1.1 Data required conditions

In this study, we present the three requirements of Availability, Identifiability, and Comparability. These requirements are applied to the data used in this study, and we analyze the data considering these requirements for storing and processing.

- **Availability of actual travel time data**

The data should be able to utilize actual route travel time data for analysis. It should be able to identify the individual travelers and process data collected in real time. Moreover, it should be possible to confirm the route for individual travelers between the specific OD pairs using actual travel time data. In this study, we need a sufficient amount of data to form the travel time distribution between OD pairs because it generates the route choice set using the distribution of travel time. It is necessary to distinguish the path because it is necessary to determine whether to include the path in the choice set using the travel time distribution of each path.

- **Identifiability of routing travel for the travelers**

It is necessary to construct a model reflecting the traveler's behavior. We need to process the data for individual experienced travel time to form of perceived travel time distribution. The data is possible to identify the individual travelers and accumulate the observed data for specific OD pair for a long period. The identification of individual travelers is the most important requirement to trace the route choice behavior. Moreover, it is necessary to be able to analyze the daily variation of the route for a specific OD pair.

- **Comparability for actual route choices**

The selected route and the alternative route should be distinguished in order to trace the chosen route in the network. In other words, the paths must be distinguishable from one another to be able to distinguish between different paths and compare with each other. There are hundreds of possible paths on the network, but the observed paths are small number of paths. To construct a consideration set, each path requires a sufficient amount of observation data to generate a travel time distribution.

4.1.2 DSRC RSE Data

In this study, we employ the Dedicated Short Range Communication data (DSRC data) to construct a choice set generation model for travelers. Dedicated Short Range Communication - Road Side Equipment. (DSRC RSE) observes the vehicle using the on-board unit (OBU) device when the vehicle travels within a certain range. The range of observations is about 3 to 30 meters, and the frequency operates at 5.8 GHz.

Figure 4.1 OBU equipment

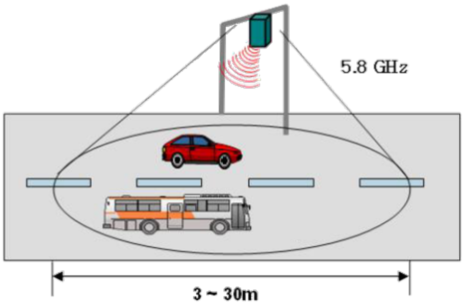
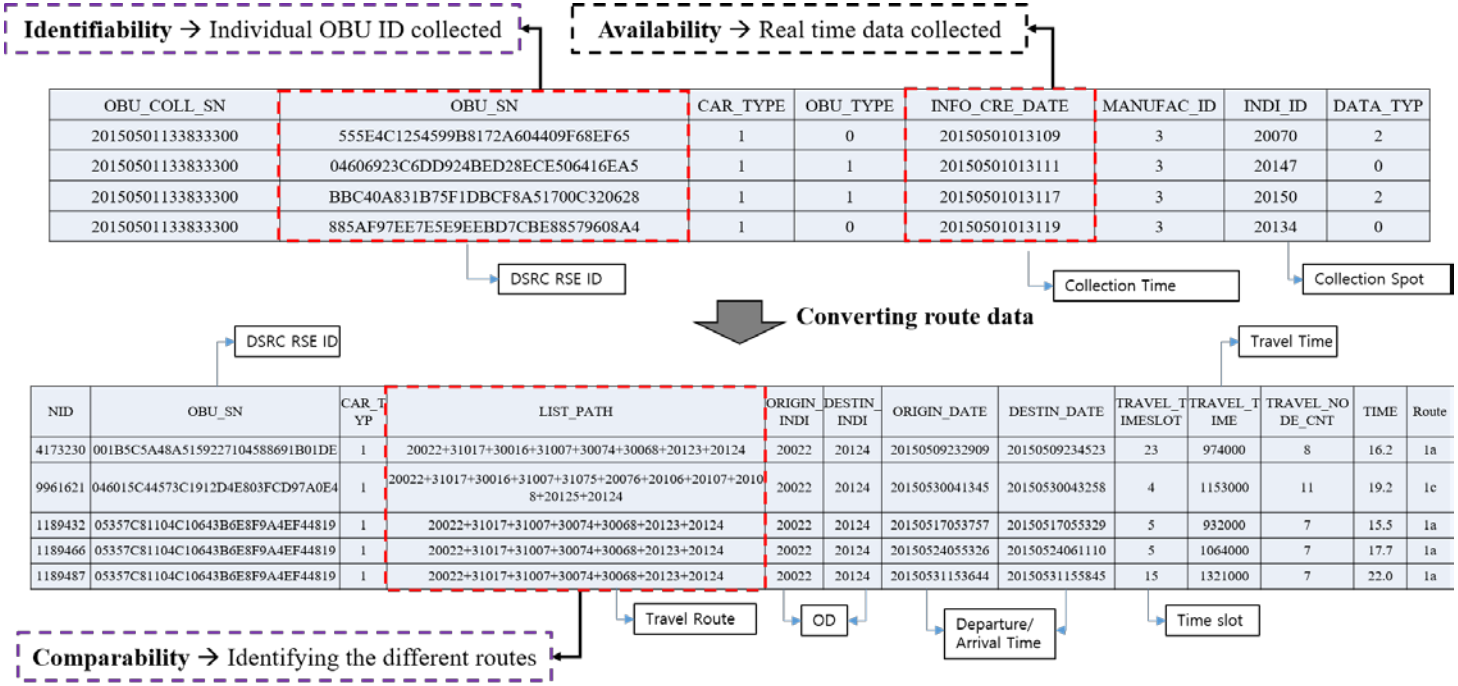


Figure 4.2 DSRC communication range

The DSRC data is the data observed at the point, and the observation time of the vehicle is accumulated in the form of data when the vehicle passes through the point. The data includes various information such as OBU ID, car type, and observed time. OBU_ID makes identify the individual travelers. It is possible to identify the individual because different IDs are assigned to each devices. INFO_CRE_DATE is a record of the observation time of the vehicle. The INDI_ID can be used as an observation point to confirm whether or not the vehicle has passed through the point. OBU_ID, INFO_CRE_DATE, and INDI_ID are mainly used to generate path travel.

Table 4.1 Data table specifications

COLUMN	TYPE	NULL	COMMENTS
OBU_COLL_SN	numeric(20,0)	N	OBU collected serial number
OBU_SN	varchar(50)	N	OBU serial number
CAR_TYP	numeric(1,0)	N	Vehicle type 1: small-sized vehicle 2: medium-sized vehicle 3: heavy vehicle 4: lorry 5: special equipment vehicle 6: compact vehicle
OBU_TYPE	numeric(1,0)	N	OBU type
INFO_CRE_DATE	numeric(14,0)	N	Observed time
MANUFAC_ID	numeric(1,0)	N	Manufacturing ID
INDI_ID	numeric(5,0)	N	Observed spot ID
DATA_TYPE	numeric(1,0)	Y	Data type NULL: before processing 0: normality 1: the same observed time 2: no connection 3: out of the range 5: checking the start point



OBU_COLL_SN	OBU_SN	CAR_TYPE	OBU_TYPE	INFO_CRE_DATE	MANUFAC_ID	INDI_ID	DATA_TYP
20150501133833300	555E4C1254599B8172A604409F68EF65	1	0	20150501013109	3	20070	2
20150501133833300	04606923C6DD924BED28ECE506416EA5	1	1	20150501013111	3	20147	0
20150501133833300	BBC40A831B75F1DBCF8A51700C320628	1	1	20150501013117	3	20150	2
20150501133833300	885AF97EE7E5E9EEBD7CBE88579608A4	1	0	20150501013119	3	20134	0

NID	OBU_SN	CAR_TYP	LIST_PATH	ORIGIN_INDI	DESTIN_INDI	ORIGIN_DATE	DESTIN_DATE	TRAVEL_TIMESLOT	TRAVEL_TIME	TRAVEL_NODE_CNT	TIME	Route
4173230	001B5C5A48A5159227104588691B01DE	1	20022+31017+30016+31007+30074+30068+20123+20124	20022	20124	20150509232909	20150509234523	23	974000	8	16.2	1a
9961621	046015C44573C1912D4E803FCD97A0E4	1	20022+31017+30016+31007+31075+20076+20106+20107+20108+20125+20124	20022	20124	20150530041345	20150530043258	4	1153000	11	19.2	1c
1189432	05357C81104C10643B6E8F9A4EF44819	1	20022+31017+31007+30074+30068+20123+20124	20022	20124	20150517053757	20150517055329	5	932000	7	15.5	1a
1189466	05357C81104C10643B6E8F9A4EF44819	1	20022+31017+31007+30074+30068+20123+20124	20022	20124	20150524055326	20150524061110	5	1064000	7	17.7	1a
1189487	05357C81104C10643B6E8F9A4EF44819	1	20022+31017+31007+30074+30068+20123+20124	20022	20124	20150531153644	20150531155845	15	1321000	7	22.0	1a

Figure 4.3 Data converting & conditions for data

The DSRC RSE data to be analyzed in this study satisfies three conditions. Identifiability is satisfied because the OBU_SN makes it possible to identify the individual travelers. The OBU_SN is an identification number given to an individual as encrypted code for an individual DSRC device. This characteristic allows the researcher to track the long-term travel experiences. Availability is satisfied by INFO_CRE_DATE. Since the real-time data is recorded for the exact time, it enables the vehicle to grasp at the point passing through DSRC RSE. It has an advantage over the data collected for a long period in forming the route travel data. Because it makes the DSRC data to process from point data to route data, the comparability can be satisfied between routes. Distinguishing the travel nodes of each path is possible. Therefore, we intend to construct a choice set generation model and route choice model using DSRC data.

4.2 Data Collection & Description

DSRC data used in this study is in Daegu Metropolitan area. The data was obtained by observing a vehicle passing DSRC RSE equipment on the intersection of the main arterial road in Daegu Metropolitan area. There are a total of 177 points where the equipment is installed, and it is installed not only on the arterial road but also on the uninterrupted flow road. The employed data is analyzed using approximately 1.4 billion data accumulated between November 2014 and December 2015 for 14 months. The DSRC observation node in Daegu area is shown in the following figure.

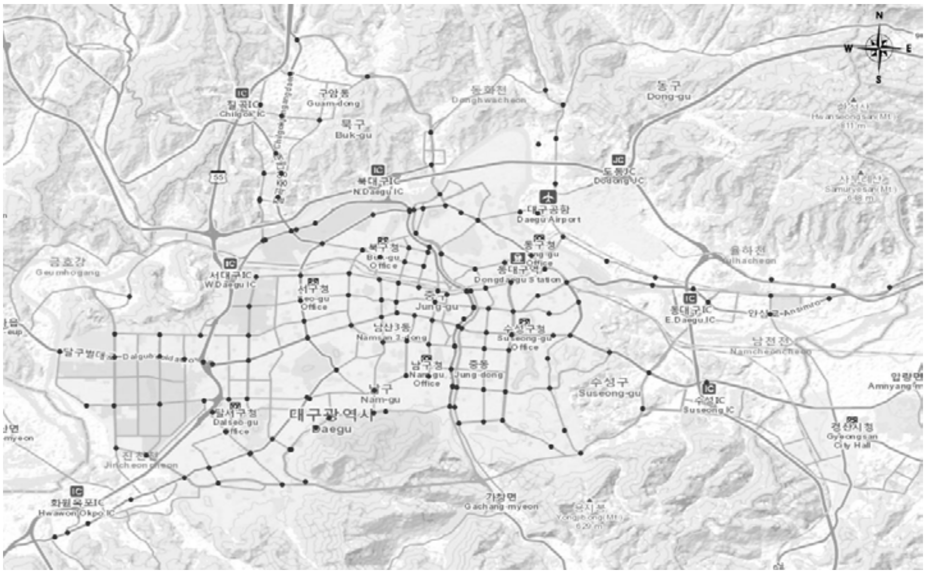


Figure 4.4 Positions of DSRC installation in Daegu City Area – 177nodes

As mentioned above, the analysis of the network is summarized to analyze the accumulated data for 14 months.

Table 4.2 Data descriptions

DIVISION	VALUE	UNIT	REMARKS
Daegu metropolitan area	883.48	㎢	
Population	2.48 million	Persons	
The number of observed travelers	Approximately 0.6 million per month	Persons	-
The Number of OD pair	31,152	Pair	177× 176
The Number of Traveled Route	Approximately 30 million per month	Trips	-
The Average of link distance	1.25	Km	-
The Average number of OD trips	6,015	Trips	-

From the brief data analysis, the number of observed travelers is approximately 0.6 million persons per month. Because the number of nodes is 177, the number of OD pairs come out to 31,152 OD pairs (177× 176). The number of traveled routes are roughly 30 million trips per month. The travelers have traveled for 50 times per month as an average. The average number of OD trips are analyzed 6,015 trips. From the network analysis, moreover, the average of link distance is calculated for 1.25 km.

We constructed a database for collecting, producing, and processing data to analyze a large amount of big data. We modified the data for the analysis using the database. We employ the Java programming language to process the route data and to correct the missing data. Produced path data and corrected data are stored in database so that it is possible to analyze for OD pairs. The data processing is shown in the following figure.

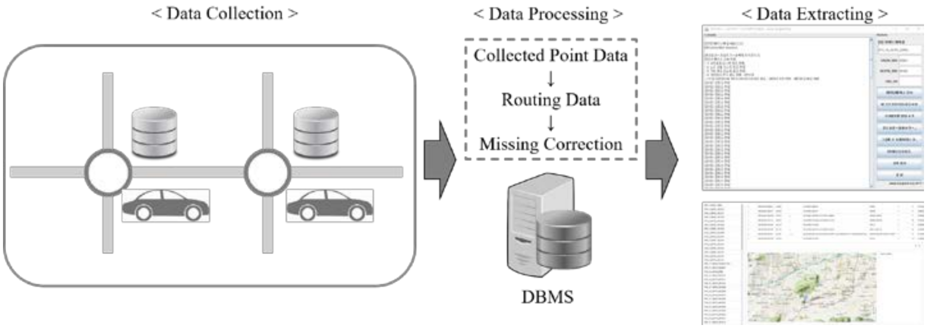


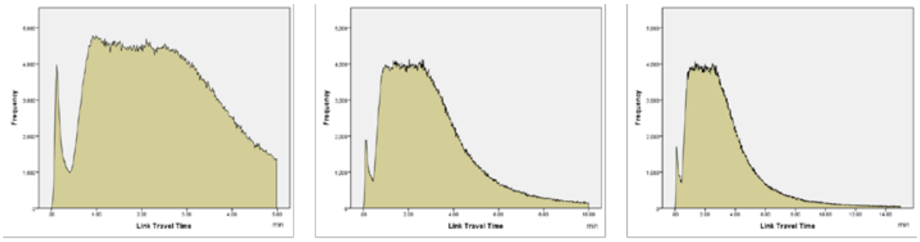
Figure 4.5 Data processing

4.3 Data Processing

Since DSRC data is a point data, it is necessary to convert DSRC data into path data. We need to process for tracking the travelers with the same vehicle ID to identify the route. When the OBU of an individual vehicle arrives at the point of DSRC RSE, the process of routing data needs to classify the data by OBU_SN. Because it is hard to process a large amount of data, we make an index on OBU_SN. It makes possible to produce the path data more rapidly. In this study, we construct a method of classifying data and generating path data based on individual vehicles.

It is necessary to identify the individual vehicle for changing from the point data to path data. We use the INFO_CRE_DATE variable to construct this process. If the observed time for individuals on DSRC RSE is arranged in order, it is possible to produce the path data for individuals. The link time is calculated while moving from the node to the node, and it includes checking whether the path is configured using the link travel time. If the link's travel time is excessive, it should be divided into a different path. Since the whole data has a large amount of data, so it may take a long time for the traveler reaching another node in producing process. These travel patterns are not correct for travelers, and it needs to be classified into different travels. In previous studies, Zhan et al. (2013) estimated the average link travel time. They estimated the travel speed of about 8 miles/hour when the distance of the link is about 80 ~ 300m using GPS data. Converting this to a distance of

1.25 km, which is the average distance of this study, can result in link travel time not exceeding 10 minutes maximum. Moreover, we estimate the distribution of frequency in link travel time using the link travel time data for 1 million of each graph by setting the travel separating time as 5 / 10 / 15 min from the frequency analysis. The distribution of link travel time for 10 min is included in almost 97.7% of travel from the frequency analysis for 15min separating time (ref. 84.8% in 5min, and 94.1% of travel in 10 min for 60min separating time). Therefore, in this study, the path was produced by setting the link travel time to 10 minutes.



- link travel time < 5min - - link travel time < 10min - - link travel time < 15min -

Figure 4.6 Distribution of link travel time

Path data is generated by point data as shown in the figure below. Each link travel time does not exceed 10 minutes, and the route travel time is calculated using the link travel time.

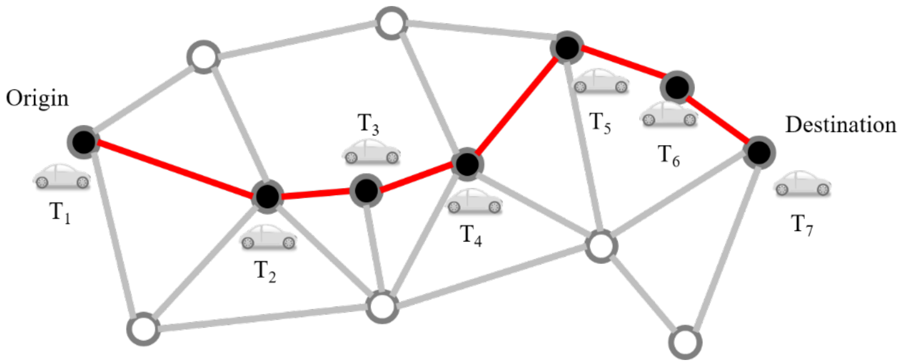


Figure 4.7 Route processing diagram

The process of generating data is performed through the following steps.

< Data processing >

Step 1. Listing the observed points for individual travelers in chronological order

Step 2. Calculating the number of observations of each traveler

Step 3. Comparing the current observation with all observational data for the traveler

Step 4. Checking and calculating the travel time whether the elapsed time between links exceeds a certain level

Step 5. Repeated route generation for all observations of each traveler

Step 6. Constructing route data in all OD pairs for travelers using all observed data

Step 7. Extracting and storing data of each OD pairs

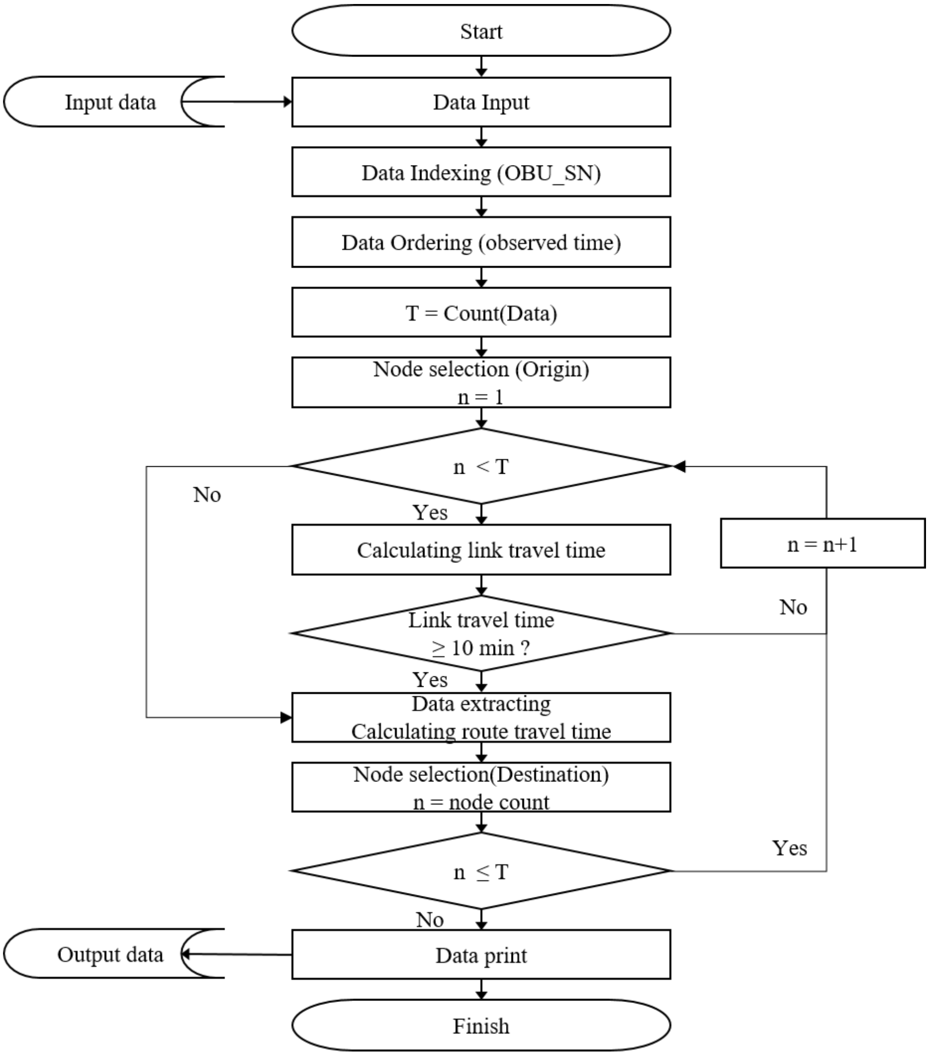


Figure 4.8 Steps for data processing

4.4 Missing Data Correction

We generate the path data by sequentially listing the data observed at the point and generating a path. However, it is impossible to confirm whether the link between the two nodes is a link. We calculate the path travel time using the sum of link travel time. There are the following three types of missing:

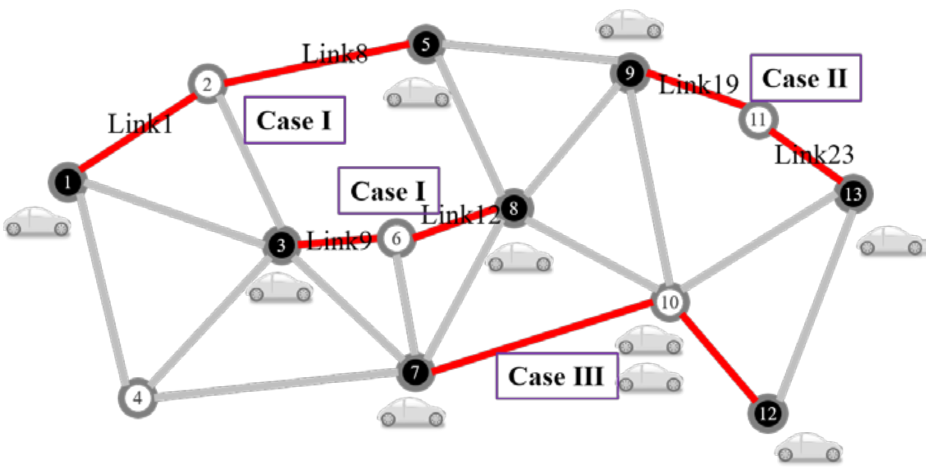


Figure 4.9 The cases of missing data

1) Case 1: Missing data between nodes in arterial road by straightway

(ex. 1 – 5 ➔ 1 – 2 – 5, 3 – 8 ➔ 3 – 6 – 8)

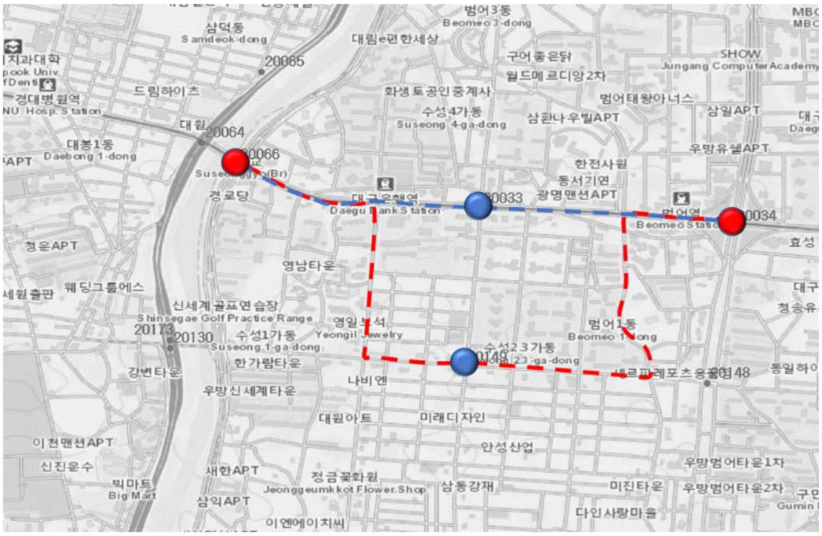


Figure 4.10 Actual missing correction for Case 1

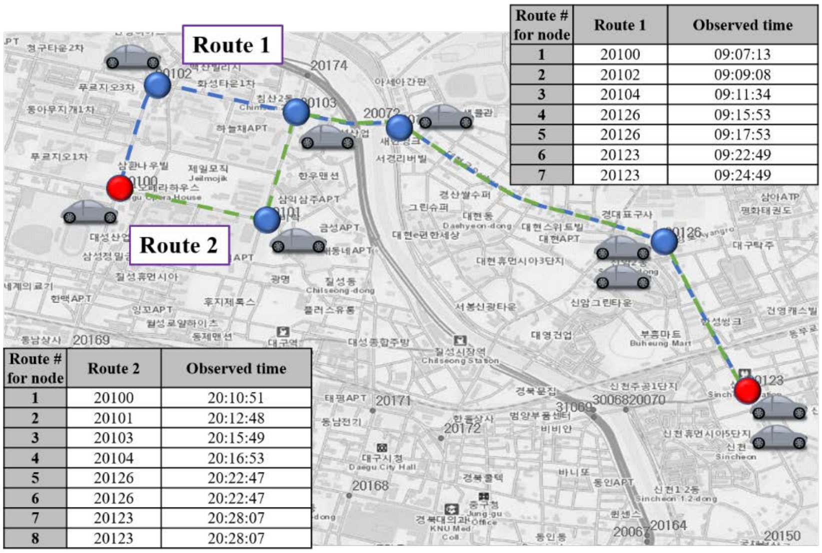
2) Case 2: Missing data between nodes for road type of uninterrupted flow

(ex. 9 – 13 ➔ 9 – 11 – 13)

Figure 4.11 Actual missing correction for Case 2

1) Case 3: Observation for twice at one node

(ex. 7 – 10 – 10 – 12 ➔ 7 – 10 – 12)



Route # for node	Route 1	Observed time
1	20100	09:07:13
2	20102	09:09:08
3	20104	09:11:34
4	20126	09:15:53
5	20126	09:17:53
6	20123	09:22:49
7	20123	09:24:49

Route # for node	Route 2	Observed time
1	20100	20:10:51
2	20101	20:12:48
3	20103	20:15:49
4	20104	20:16:53
5	20126	20:22:47
6	20126	20:22:47
7	20123	20:28:07
8	20123	20:28:07

Figure 4.12 Actual missing correction for Case 3

It needs to define links between the nodes to ensure whether it is connected links. This is similar to the link definition process on the network problem. The process of identifying a link is a process of confirming whether two nodes are connected to each other. Since the used data is point data, there is no definition of path producing process. Therefore, it is necessary to identify the link for the produced path data.

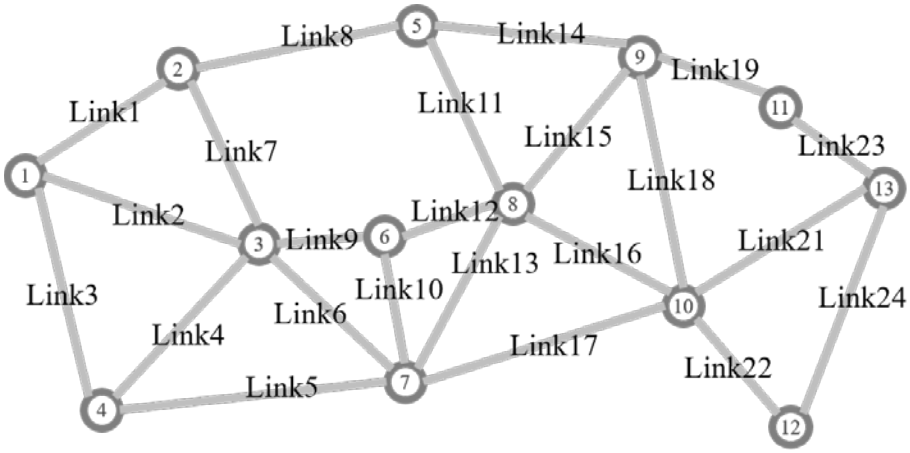


Figure 4.13 Link definition

There is an example to explain the process of missing correction. The following path data was generated. The produced path data is the 1 → 3 → 8 path. This path composes of the missing between node 3 to node 8. It is possible to identify the unconnected link between node 3 and node 8. We perform the missing correction using the defined link. The corrected path data is devised from 1 → 3 → 8 to 1 → 3 → 6 → 8. This process is performed using all the missing data. The developed program performed the various missing correction procedures according to the case presented above.

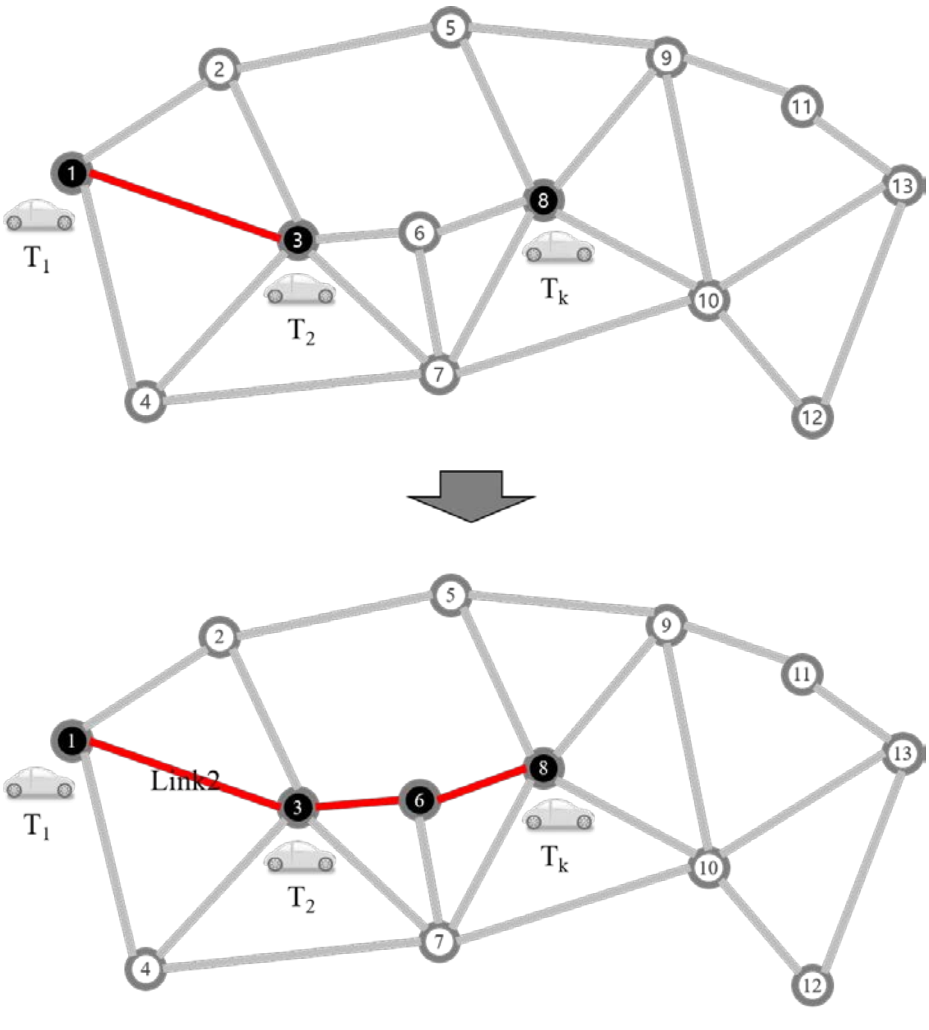


Figure 4.14 Example of missing correction

< Data processing >

Step 1. Calculating the number of nodes for the route

Step 2. Missing data correction for the number of nodes 2 or more

Step 3. Link definition

Step 4. Check whether the link between observed nodes is connected

Step 5. Generating the shortest path between nodes

Step 6. Replace the shortest path for the missing cases

Step 7. Check whether the current n is equal to the entire node

Step 8. Repeated check whether inter-node links are connected links

Step 9. If the current n is the same as the entire number of node T, repeat the missing link correction for the next route

Step 10. Missing correction route extraction and storage

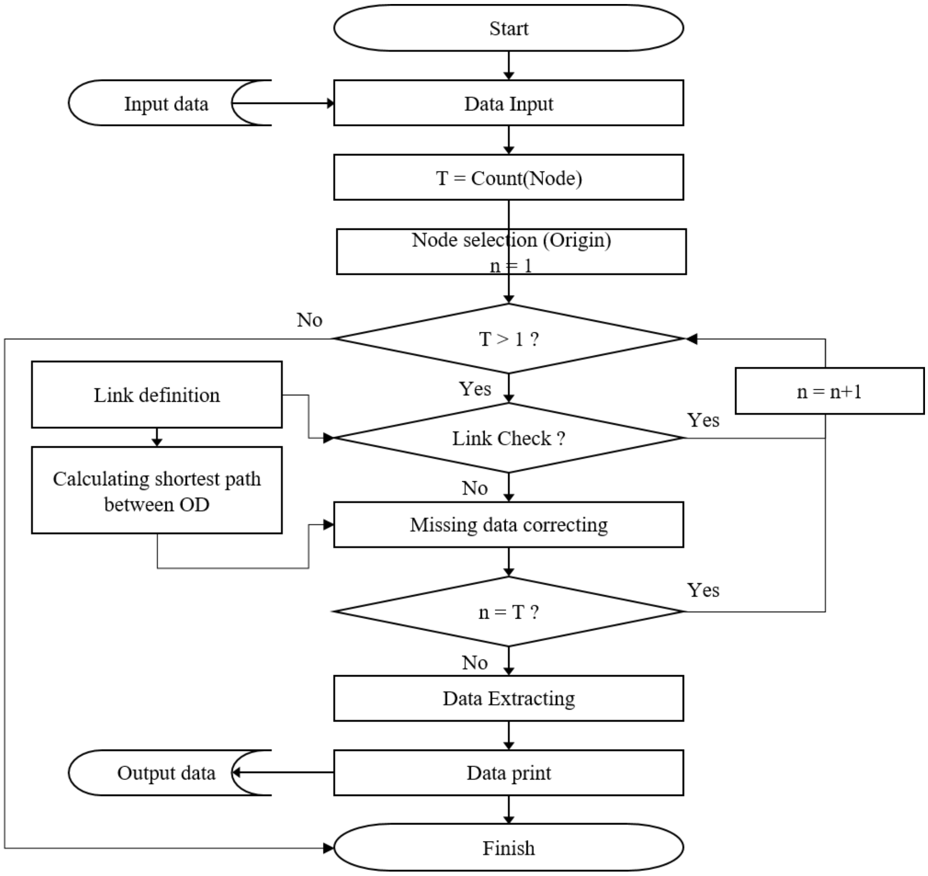


Figure 4.15 Steps for missing data correction

Chapter 5. Model Estimation & Validation

5.1 Overview

We employ the generated data to estimate the parameters of the model and validate this model by comparison with types of choice set in route choice model. Also, the goodness of fit of each model should be evaluated to get the better model having moderate variables. We evaluate the fitness of the estimated model through the process shown in the figure below. We compare the derived models using model verification.

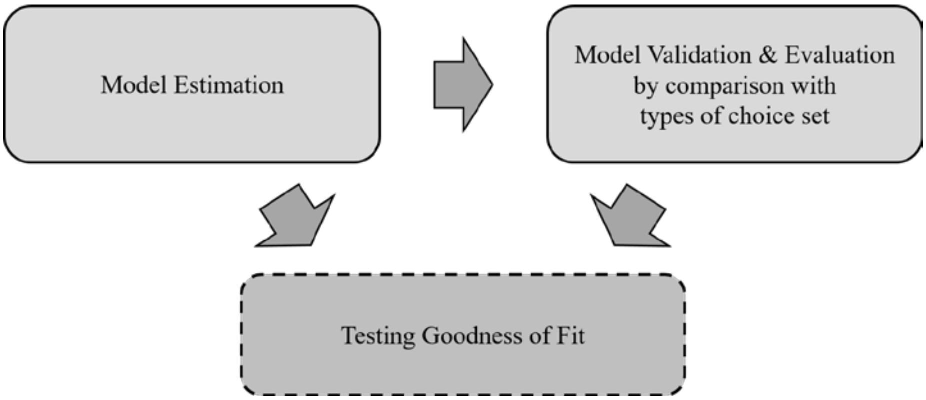


Figure 5.1 Methodology of model estimation and validation

5.2 Choice set generation

We estimate the size of the consideration set using travel data for 76 OD pairs. As mentioned before, it is important to determine the appropriate consideration set by the millions of alternatives. We establish the methodology for choosing the consideration choice set using actual travel data. The consideration choice sets have to include the all of possible choice set for most of the travelers. We set up the adjustment for 90% of coverage probability from the paths for each OD pairs. The size of consideration set comes out to 16 paths for each OD pair.

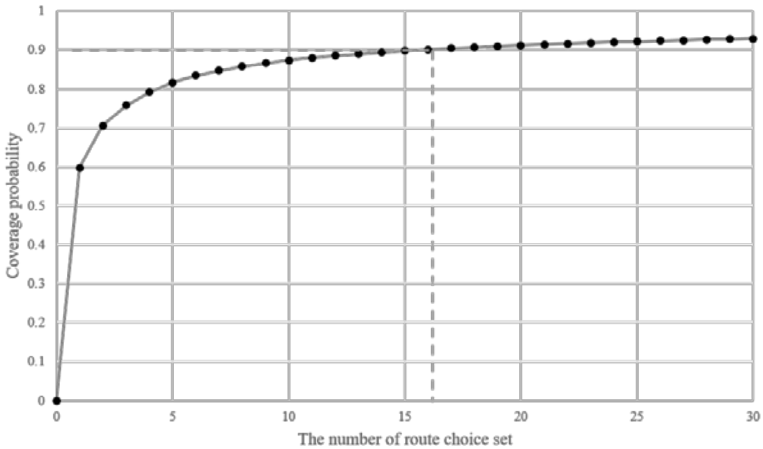


Figure 5.2 Size of consideration choice set

Moreover, the size of individual consideration set is estimated with appropriate for individual travelers considered in route choice. Determination for the proper individual consideration set is estimated from the average of experienced alternatives for individual travelers and 80% of coverage probability. The size of individual consideration set is four paths from the analysis.

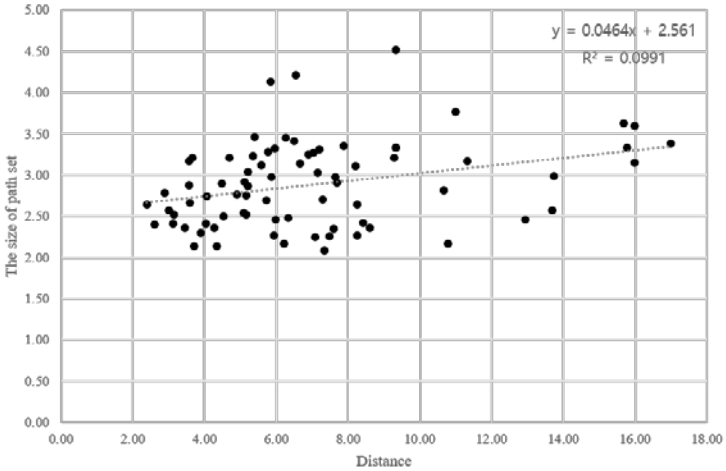


Figure 5.3 Relationship between size of set and distance

We compare the size of individual consideration set concerning the length of the route. However, there is not much correlation between length and the size of a set. There are many Studies on Choice set generation to identify the actual consideration set for individuals. The generated choice set size of 16 and 4 are appropriate from the comparison of the previous researches.

Table 5.1 Comparison of choice set size

Study	Viewpoint	Methodology	Sample	Consideration set	Individual choice set	Application	Remarks
Bovy and Stern (1990)	Traveler	Questionnaire survey	50	15	4	Estimation	Netherlands
			1300	-	2~3		Newcastle
Ramming(2002)	Traveler	Heuristic(Link penalty, elimination), Simulation	160	30~51	-	Prediction	-
Bierlaire and Frejinger(2005)	Traveler	Simulation	1,461	-	9.3	Estimation	-
Hoogendoorn-Lanser (2005)	Traveler	Heuristic(Probabilistic choice set generation)	511	63	2	Estimation	-
Bekhor and Prato(2009)	Researcher	Heuristic(Branch and bound)	181~409	44~61	17~31	Estimation	-
Schussler et al. (2012)	Researcher	Heuristic(Breadth first search)	-	20~100	-	Estimation	-
Kitthamkesorn and Chen(2013)	Researcher	Simulation	-	-	3~6	Demand analysis	-
Li et al.(2016)	Researcher	Heuristic(Link penalty), Simulation	-	50	-	Estimation	-
Chen et al.(2016)	Researcher	K-α-Reliable Shortest Path	-	3	-	Prediction	Wuhan
This Study(2017)	**Traveler**	**K-α-Reliable Shortest Path**	**Millions**	**16**	**4**	**Estimation**	**Daegu**

5.3 Model estimation & Validation

The travel data for 40 specific OD pairs were selected to model the K-α-Reliable Shortest Path Searching based route choice model in Daegu metropolitan area. We establish the heterogeneous K-$\boldsymbol{\alpha}$-reliable shortest path based route choice model to improve the accuracy of estimation.

We propose the three kinds of models to evaluate the superior from the other models. We also formulate the route choice model based on K-$\boldsymbol{\alpha}$-reliable shortest path and observed K shortest path to compare the results for the accuracy of estimated models.

Table 5.2 Model comparison

Division		**Route Choice Model based on Heterogeneous KαRSP (HKαRSP)**	**Route Choice Model based on KαRSP (KαRSP)**	**Route Choice Model based on Observed KSP (KSP)**
Descriptions		Different choice set by risk preference among travelers	Equivalent choice set with K-α-reliable shortest path	Equivalent choice set with observed shortest path
Choice set	Consideration set	16	16	16
	Individual choice set	4	4	4
	Objective choice set	5~7	4	4
Data type		Aggregated panel data		
Route choice model		MNL / MMNL / PSCL / MPSCL model comparison		
Goodness of fit		ρ^2		

We analyze the route choice model to estimate the parameters using the actual travel data. We employ NLOGIT 6.0 program in this study. MNL model set for estimating the model with maximum likelihood estimation.

The equation for the route choice model is established to estimate the more accurate choice model. There are several explanatory variables to consider for the model estimation. There is a process of choosing feasible variables in the route choice model. There are several variables, which is possible to analyze using the DSRC data. We compare the explanatory variables whether the variables improve the goodness of fit or there is multi-collinearity. It is necessary to identify the correlations between variables due to the model estimation. The analysis of Pearson correlation is employed to choose the appropriate variables.

As a result, we choose the variables to improve the model fit, such as the mean of travel time, the standard deviation of travel time, buffer time, travel distance, the ratio of uninterrupted flow, toll fare, and the number of the bridge. Using these variables, we formulate the route choice model.

Table 5.3 Choice of the variables

VARIABLES		CHOICE	EXPLANATION	REASON
Level Of Service (LOS) Variables	Free flow Travel Time (FTT)	X	Observed minimum travel time of the route	Significant decreasing of network attribute variables
	Mean Travel Time (MTT; $\boldsymbol{\mu_k^{ij}}$)	○	Mean travel time of the route	Increasing model significance
	Standard Deviation of Travel Time (SDTT; $\boldsymbol{\sigma_k^{ij}}$)	X	Standard deviation of the route	Multi-collinearity with buffer time
	Buffer Time (BT)	○	90% Travel time budget – 50% travel time for each route by observations	Increasing model significance
	Travel Distance (D)	○	Traveled distance (Sum of euclidian distance)	Increasing model significance
	Travel Cost (TC)	X	The function of travel time and travel distance (in-vehicle operation cost)	Multi-collinearity with mean travel time and travel distance
Network Attribute Variables	% of Uninterrupted flow (RU)	○	The rate of uninterrupted flow on the route	Increasing model significance
	Toll fare (TOLL) (100won)	○	Toll fare on the route including toll road (100won)	Increasing model significance
	# of Intersection (INT)	X	The number of intersection on the route	Decreasing model significance
	# of Bridge (BR)	○	The number of bridge on the route	Increasing model significance

$$L = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}}$$

$$V_k = \beta_1 \mu_k^{ij} + \beta_2 BT_k^{ij} + \beta_3 D_k^{ij} + \beta_4 RU_k^{ij} + \beta_5 TOLL_k^{ij} + \beta_6 BR_k^{ij}$$

Where, V_k: Utility function for alternative k

β_i: The parameters

μ_k^{ij}: The mean travel time for alternative k from node *i* to node *j*

BT_k^{ij}: The buffer travel time for alternative k from node *i* to node *j*

D_k^{ij}: The traveled distance for alternative k from node *i* to node *j*

RU_k^{ij}: The rate of uninterrupted flow for alternative k from node *i* to node *j*

$TOLL_k^{ij}$: The toll fare for alternative k from node *i* to node *j*

BR_k^{ij}: The number of bridges for alternative k from node *i* to node *j*

We analyze the relative comparison of variables between alternatives in the model without alternative specific constants (ASCs). It is necessary to retain the dummy variables to avoid the unbiasedness (Ramming, 2002). There are some dummy variables to interpret the alternatives in the model (i.e., ratio of uninterrupted flow, toll road, and bridge). Since travelers tend to consider the more travel attribute than an immanent attribute of alternatives in route choice, the model needs the additional variables.

We conduct the model validation to confirm the appropriacy of the proposed model. The goodness of fit index (ρ^2, $\bar{\rho}^2$) has been employed to validate the estimated model.

$$\rho^2 = 1 - \left(\frac{L(\beta)}{L(0)}\right), \bar{\rho}^2 = 1 - \left(\frac{L(\beta) - K}{L(0)}\right)$$

Where, $L(\beta)$: the origin node

$L(0)$: the destination node

K: degrees of freedom

The result of route choice behavior computes the different model fit by the methodology of choice set generation. Chi-squared value is estimated to test the significance level at 0.00%, and the result comes out significant model fit for those of models.

In this study, we compare the results of the route choice model based on the different choice set generation model using the MNL, PSCL, MMNL, and MPSCL models. The estimation for MNL model is conducted to derive the basic choice behaviors. Furthermore, we employ the PSCL model reflecting the overlapping links and analyze the model considering the traveler's heterogeneity using the MMNL model. Finally, this study derives the MPSCL model combining the PSCL model and the MMNL model.

As a result of estimating the parameters using the MNL model, the fitness of the HK$\boldsymbol{\alpha}$RSP model was analyzed to have the highest value, and all of the service level attributes are the negative values. On the other hand, the parameter estimation results of the KSP - based route choice model is analyzed to have the best model fit given t-statistics for parameters.

Table 5.4 Result of Multinomial logit model; MNL

EXPLANATORY VARIABLES	HKαRSP		KαRSP		KSP	
	coefficient	t-statistics	coefficient	t-statistics	coefficient	t-statistics
Level of Service (LOS) Attribute Variable						
Mean of travel time (μ_k^{ij})	-0.1385***	-44.40	-0.0641***	-33.94	-0.0068***	-3.33
Buffer Time (BT_k^{ij})	-0.0404***	-5.83	-	-	-0.3527***	-66.22
Travel distance ($DIST_k^{ij}$)	-0.6519***	-52.15	-0.9733***	-82.61	-0.8501***	-55.40
Network Attribute Variable						
Ratio of Uninterrupted flow ($\boldsymbol{UNINT_k^{ij}}$)	-	-	0.4054***	9.44	0.4451***	12.89
Toll fare ($\boldsymbol{TOLL_k^{ij}}$) (100won)	-	-	-	-	-0.0731***	-15.11
The number of Bridge ($\boldsymbol{BRIDGE_k^{ij}}$)	-0.7081***	-33.48	-0.4636***	-24.92	-0.1105***	-6.35
Goodness of Fit						
Observations	40,000		40,000		40,000	
$\boldsymbol{LL(0)}$	-55,451.8		-55,451.8		-55,451.8	
$\boldsymbol{LL(\beta)}$	-48,612.5		-48,967.6		-48,661.53	
$\boldsymbol{\rho^2}$	0.1233		0.1169		0.1225	
$\boldsymbol{\bar{\rho}^2}$	0.1232		0.1168		0.1223	

The analysis of the model fits presented in the <Table 5.4> shows the limitation to reflect the actual behavior regarding fitness index of the models. In conclusion, it can be concluded that the parameter estimation using the MNL has a limitation in accurately considering the actual behaviors so that applying the model reflecting the overlap between the routes will give better results.

Table 5.5 Result of Path-size Correction multinomial logit model; PSCL

EXPLANATORY VARIABLES	HKαRSP		KαRSP		KSP	
	coefficient	t-statistics	coefficient	t-statistics	coefficient	t-statistics
Level of Service (LOS) Attribute Variable						
Mean of travel time (μ_k^{ij})	-0.0465***	-14.66	-0.0381***	-16.98	0.0077***	3.78
Buffer Time (BT_k^{ij})	-0.3029***	-34.17	-0.2291***	-23.75	-0.3388***	-51.74
Travel distance ($DIST_k^{ij}$)	-0.6107***	-38.93	-1.1093***	-66.69	-0.8638***	-60.85
Network Attribute Variable						
Ratio of Uninterrupted flow ($\boldsymbol{UNINT_k^{ij}}$)	4.7544***	63.43	2.4218***	48.75	-0.1099***	-2.80
Toll fare ($\boldsymbol{TOLL_k^{ij}}$) (100won)	-0.0666***	-10.99	-	-	-0.0342***	-6.97
The number of Bridge ($\boldsymbol{BRIDGE_k^{ij}}$)	-0.9497***	-39.55	-0.8996***	-41.93	-0.0683***	-3.92
Path-size correction (PSC)	-3.8680***	-110.61	-3.5059***	-115.21	-1.6276***	-75.46
Goodness of Fit						
Observations	40,000		40,000		40,000	
$\boldsymbol{LL(0)}$	-55,451.8		-55,451.8		-55,451.8	
$\boldsymbol{LL(\beta)}$	-36,905.1		-36,552.9		-44,954.9	
$\boldsymbol{\rho^2}$	0.3345		0.3408		0.1893	
$\boldsymbol{\bar{\rho}^2}$	0.3343		0.3406		0.1891	

The model reflecting the overlapping links is found to have a better fit than that of MNL model. Service level attribute variables are mostly derived as significant. The route choice model based on KαRSP was found to yield the best results. This is because the improvement effect of the model is greatest due to the influence of the path overlapping.

Table 5.6 Result of Mixed multinomial logit model; MMNL with truncated normal distribution

EXPLANATORY VARIABLES		HKαRSP		KαRSP		KSP	
		coefficient	t-statistics	coefficient	t-statistics	coefficient	t-statistics
Level of Service (LOS) Attribute Variable							
Mean of travel time (μ_k^{ij})	Constant	-0.6837***	-60.57	-0.1144***	-31.27	-0.0676***	-18.99
	SD	1.3979***	71.53	0.2416***	24.80	0.3424***	44.19
Buffer Time (BT_k^{ij})	Constant	-	-	-	-	-0.4362***	-56.20
	SD	-	-	-	-	-	-
Travel distance ($DIST_k^{ij}$)	Constant	-0.5697***	-32.37	-1.0139***	-80.13	-0.9190***	-55.95
	SD	-	-	-	-	0.4816***	9.73
Network Attribute Variable							
Ratio of Uninterrupted flow ($\boldsymbol{UNINT_k^{ij}}$)		-	-	0.3218***	7.13	0.4448***	11.51
Toll fare ($\boldsymbol{TOLL_k^{ij}}$) (100won)		-	-	-	-	-0.0867***	-14.57
The number of Bridge ($\boldsymbol{BRIDGE_k^{ij}}$)		-1.2124***	-36.61	-0.4979***	-24.82	-0.0863***	-4.56
Goodness of Fit							
Observations		40,000		40,000		40,000	
$\boldsymbol{LL(0)}$		-55,451.8		-55,451.8		-55,451.8	
$\boldsymbol{LL(\beta)}$		-43,077.4		-48,887.4		-48,940.1	
$\boldsymbol{\rho^2}$		0.2217		0.1184		0.1342	
$\boldsymbol{\bar{\rho}^2}$		0.2215		0.1182		0.1340	

The HKαRSP model reflecting individual heterogeneity was analyzed to show the best fit index. On the other hand, the results of parameter estimation are not significant given t-statistics. Therefore, we considered that a better model can be derived by examining the model reflecting link overlapping and individual heterogeneity in route choice model.

Table 5.7 Result of Mixed Path-size correction multinomial logit model; MPSCL with truncated normal distribution

EXPLANATORY VARIABLES		**HKαRSP**		**KαRSP**		**KSP**	
		coefficient	**t-statistics**	**coefficient**	**t-statistics**	**coefficient**	**t-statistics**
Level of Service (LOS) Attribute Variable							
Mean of travel time (μ_k^{ij})	Constant	-0.3121***	-61.22	-0.0404***	-32.36	-0.0343***	-10.35
	SD	1.3852***	68.83	0.3073***	29.63	0.2865***	38.23
Buffer Time (BT_k^{ij})	Constant	-0.3129***	-26.30	-0.2300***	-24.04	-0.4009***	-52.23
	SD	-	-	-	-	-	-
Travel distance ($DIST_k^{ij}$)	Constant	-0.5840***	-29.42	-1.1117***	-67.68	-0.9177***	-59.22
	SD	-	-	-	-	-	-
Network Attribute Variable							
Ratio of Uninterrupted flow ($\boldsymbol{UNINT_k^{ij}}$)		4.5800***	53.88	2.4270***	49.65	-	-
Toll fare ($\boldsymbol{TOLL_k^{ij}}$) (100won)		-0.0959***	-11.70	-	-	-0.0454***	-8.21
The number of Bridge ($\boldsymbol{BRIDGE_k^{ij}}$)		-1.1636***	-38.75	-0.9013***	-43.07	-0.0508***	-2.71
Path-size correction (PSC)		-4.0237***	-99.25	-3.5107***	-117.47	-1.7035***	-73.76
Goodness of Fit							
Observations		40,000		40,000		40,000	
$\boldsymbol{LL(0)}$		-55,451.8		-55,451.8		-55,451.8	
$\boldsymbol{LL(\beta)}$		-35,558.1		-36,551.9		-44,550.1	
$\boldsymbol{\rho^2}$		0.3588		0.3408		0.1966	
$\boldsymbol{\bar{\rho}^2}$		0.3586		0.3407		0.1964	

The parameters of the estimated model have the appropriate value in the model and draw the significance at 1% level for the all of models. PSCL model has the scale of the parameter for travel distance because PSCL model reflects the path-size correction factor by the link length of which shared in the alternative routes.

The KαRSP incorporating traveler heterogeneity in route choice model has the best goodness of fit index in those of route choice model. The proposed model results in the most precise model to predict the route choice probability by adopting the choice set generation with traveler heterogeneity.

PSCL model makes much-improved ρ^2 than MNL model; it means that the effect of considering overlapping links have the significant impact on the model estimation. In the model comparison, MPSCL model represents the most significant model fit from the analyzed models

Table 5.8 Comparison of goodness of fit

DIVISION	HKαRSP	KαRSP	KSP
MNL	0.1233	0.1169	0.1225
PSCL	0.3345	0.3408	0.1893
MMNL	0.2215	0.1340	0.1174
MPSCL	0.3588	0.3408	0.1966

The parameters of the estimated model have the appropriate value in the model and draw the significance at 1% level for the most of models. Mixed logit based models with traveler heterogeneity have the higher goodness of fit indexes for all of the models due to the employment of variated parameters. The models have the parameters for the standard deviations of the variables, of which are mean of travel time, the standard deviation of travel time, and travel distance. Those of parameters represent the variations of the variables, in other words, the variables have a different effect on the choice for individual travelers, of which is modeled by the random parameter (mixed logit model).

The parameters for the mean and standard deviation of travel time have the greater effect on the model in MPSCL model than those in PSCL model, which means that the travel time makes the differences in choice probability with traveler heterogeneity. The random parameters have a positive effect on the log likelihood estimation comparing basic models to improve the accuracy of parameters.

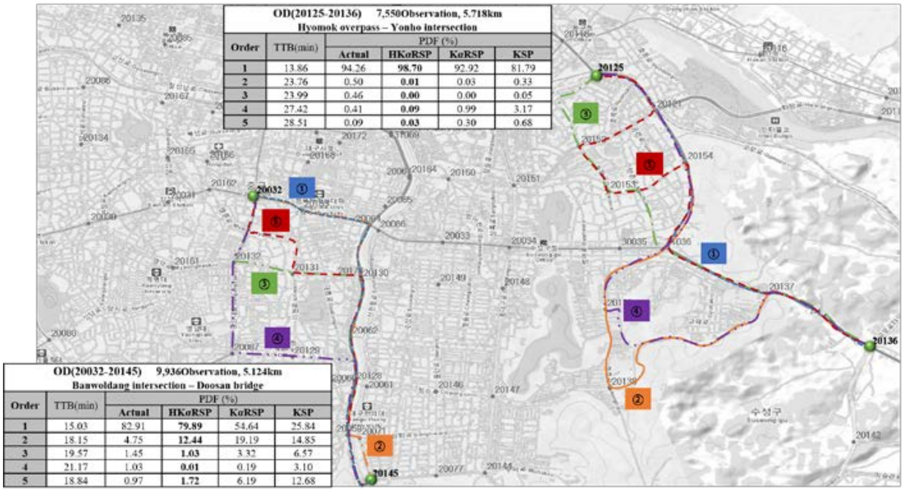
5.4 Model verification

We verify the model by the prediction for the route choice probability using the estimated parameters. The category of the verification sets for three in prediction for the route choice probability based on the distance (Short / Medium / Long distance). There may be differences according to the distance between OD pairs.

Table 5.9 The OD pairs for verifications

Division	Origin		Destination		Land Use	Observation	Distance	Observed choice set
	ID	Name	ID	Name				
Short	20032	Banwoldang intersection	20145	Doosan bridge	CBD	10,391	4.940	2.915
	20125	Hyomok overpass	20136	Yonho intersection	Residential	9,999	5.718	2.690
Medium	20021	Seodaegu industrial intersection	20107	Gongsansuwonji intersection	Industrial	19,555	7.487	2.259
	20098	Dogok intersection	20150	Cheongu intersection	Residential	14,369	7.874	3.357
Long	20044	Sangin intersection	20141	Gwangye three-way intersection	Residential	4,977	10.793	2.167
	20124	Dongdaegu station intersection	20024	Gyemyung university intersection	Commercial (Station)	4,459	16.792	2.859

The result of short distance (0~5km) shows the preference for the shortest distance and arterial road. Travelers tend to choose one route mainly, and the proposed model shows the most similar model estimation results. The routes may not have the distinct differences in the choice property.

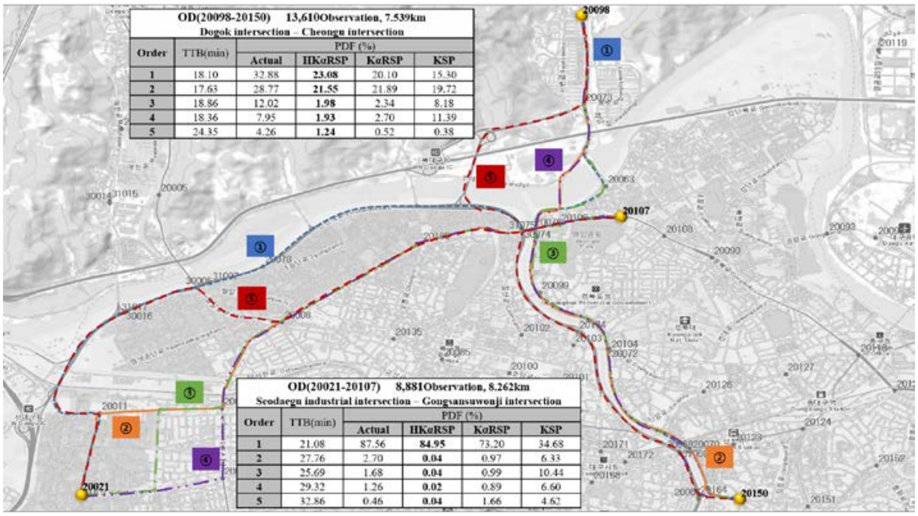


OD(20125-20136) 7,550Observation, 5.718km Hyomok overpass – Youho intersection					
Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	13.86	94.26	**98.70**	92.92	81.79
2	23.76	0.50	**0.01**	0.03	0.33
3	23.99	0.46	**0.00**	0.00	0.05
4	27.42	0.41	**0.09**	0.99	3.17
5	28.51	0.09	**0.03**	0.30	0.68

OD(20032-20145) 9,936Observation, 5.124km Banwoldang intersection – Doosan bridge					
Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	15.03	82.91	**79.89**	54.64	25.84
2	18.15	4.75	**12.44**	19.19	14.85
3	19.57	1.45	**1.03**	3.32	6.57
4	21.17	1.03	**0.01**	0.19	3.10
5	18.84	0.97	**1.72**	6.19	12.68

Figure 5.4 Model verification for short distance

The result of medium distance (5~10km) reveals the preference for the use of uninterrupted flow road. Since the travels are generated from the relatively long distance in the urban area, travelers tend to choose the uninterrupted flow.

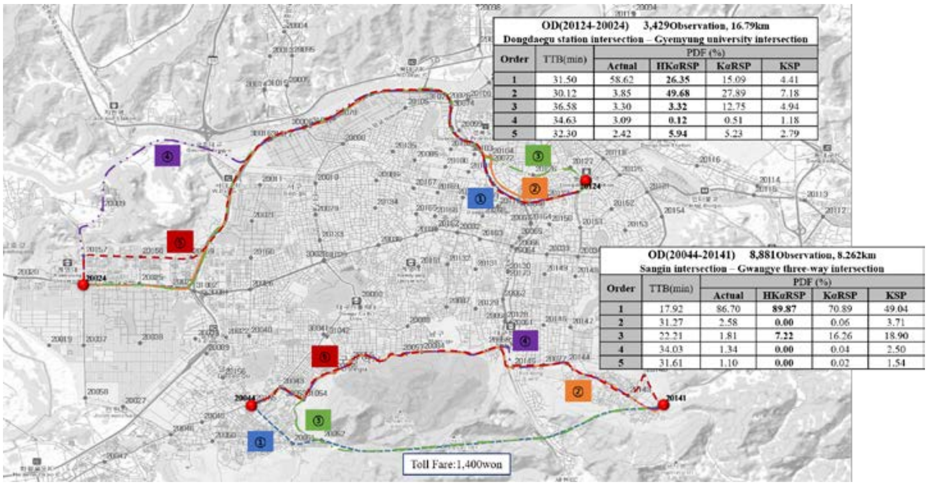


OD(20098-20150) 13,610Observation, 7.539km Dogok intersection – Cheongu intersection					
Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	18.10	32.88	**23.08**	20.10	15.30
2	17.63	28.77	**21.55**	21.89	19.72
3	18.86	12.02	**1.98**	2.34	8.18
4	18.36	7.95	**1.93**	2.70	11.39
5	24.35	4.26	**1.24**	0.52	0.38

OD(20021-20107) 8,881Observation, 8.262km Seodaegu industrial intersection – Gongsansuwonji intersection					
Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	21.08	87.56	**84.95**	73.20	34.68
2	27.76	2.70	**0.04**	0.97	6.33
3	25.69	1.68	**0.04**	0.99	10.44
4	29.32	1.26	**0.02**	0.89	6.60
5	32.86	0.46	**0.04**	1.66	4.62

Figure 5.5 Model verification for medium distance

The result of long distance (10km~) derives the tendency of preference for the use of uninterrupted flow and faster toll road. Even though there is toll road in the route, travelers tend to choose the faster toll road. Moreover, the travelers have many alternative routes for the travel due to the network. The more size of the network makes the more possible routes for specific OD pairs.



OD(20124-20024) 3,429Observation, 16.79km
Dongdaegu station intersection – Gyemyung university intersection

Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	31.50	58.62	**26.35**	15.09	4.41
2	30.12	3.85	**49.68**	27.89	7.18
3	36.58	3.30	**3.32**	12.75	4.94
4	34.63	3.09	**0.12**	0.51	1.18
5	32.30	2.42	**5.94**	5.23	2.79

OD(20044-20141) 8,881Observation, 8.262km
Sangin intersection – Gwangye three-way intersection

Order	TTB(min)	PDF (%)			
		Actual	HKαRSP	KαRSP	KSP
1	17.92	86.70	**89.87**	70.89	49.04
2	31.27	2.58	**0.00**	0.06	3.71
3	22.21	1.81	**7.22**	16.26	18.90
4	34.03	1.34	**0.00**	0.04	2.50
5	31.61	1.10	**0.00**	0.02	1.54

Figure 5.6 Model verification for long distance

The proposed model is superior to the other models. Moreover, the estimation results for different OD pairs can be derived according to the distance of the route.

5.5 Discussion

The use of buffer time is deriving the better goodness of fit than the standard deviation of travel time for travel time reliability. There is an impact of level of service attribute variables in the route choice model. The model has the more sensitive to the travel time reliability (buffer time; BT) than average travel time (μ_k^{ij}). Less distance makes the better model fit than the mean of travel time in route choice.

Furthermore, there is an impact of additional network attribute variables in the route choice model. The more ratio of uninterrupted flow reveals the more choice probability in the model. There is a tendency for less preference to use of toll road for travel in the urban area. Travelers tend to avoid to cross the bridge in the model due to the traffic congestion.

We derive the choice set generation using the observed chosen routes. The observed consideration choice set is estimated for 16 paths with 90% of accuracy. The four individual choice set for travelers is estimated by the average of the observed number of chosen routes. The model has the better goodness of fit in the mixed logit model. Reflecting the traveler heterogeneity in the mixed logit model makes the accuracy of estimations. That is due to the consistency of the structure for choice set generation model and route choice model.

We evaluate the best fit for PSCL/MPSCL model. Due to the coincidence of consideration set generation and path size correction term, the model has the better model fitness indexes. Consideration of identified choice set for travelers is adopted in PSCL/MPSCL route choice model.

There is a process of verification for the probability of route choice. The model has the better accuracy of prediction for route choice probability in HKαRSP model than KαRSP and KSP model. We derive the better prediction according to the different travel distance.

Chapter 6. Conclusion

6.1 Conclusion

We formulate the route choice model using the travel time budget (TTB) and risk preferences incorporating traveler's heterogeneity. This study suggests the definition of the distribution type of travel time and test the fitness of distribution. We establish the concept of travel time budget (TTB) and derive the individual TTB by using individual confidence level and confidence level in the network. There is a process of construction for traveler's choice set generation model using risk preferences.

The individual TTBs are used to analyze individual risk preference, and different sets of routes can be generated for the individual confidence level. In addition to eliminating the unreasonable route choice set, the route choice set can more realistically handle the choice set generation model considered by the actual travelers. We employed the K-α-reliable shortest path searching methodology to construct a choice set generation model, of which is called probabilistic reliable path search algorithm for finding the K-α-reliable shortest paths (PRPSA-KαRSP).

This study introduces the route choice model based on the K-α-reliable shortest path generation model. The result of route choice model using RP data has the feasible goodness of fit similar to that of SP data-based route choice model. Most of the variables have negative value without ‘uninterrupted flow’ variable in the models. The most significant model comes out MPSL model with respect to the goodness of fit. In order to verify the model derived from this study, we applied it to the route choice model for specific OD pairs using the parameter values. The model developed in this study shows that the estimated value has more appropriate choice probability than the other models.

There are some limitations to this study. It needs the process of the determination for feasible consideration choice set from hundreds of observed route choice data. Because of the urban route choice modeling, the difference between alternatives does not make much change of choice probability.

Because there are many unreasonable choices from the choice set, the goodness of fit index comes out lower value than the analysis using stated preference data. Due to the missing data, it needs to the process for missing correction to identify each route in the network. It needs to install the more observing spot in the network for the clear route processing.

To interpret the route choice behavior, it is necessary getting the socio-economic data for individual travelers. The other variables of network attribute can make the model more feasible than the suggested model. It needs to compare the other route choice model using this RP data (i.e., multinomial weibit model, multinomial probit model, etc.)

6.2 Further research

6.2.1 Route choice model with risk preferences

It needs to construct the variated route choice model by risk preferences. There is no clearance for the characteristics of the risk neutral, risk seeking, and risk averse (i.e., risk neutral = 0.5 ?). In other word, the definition of risk preference has the problem of risk neutral case. It is too limiting to state that the traveler is risk neutral only when s/he has exactly $\alpha = 0.5$. This property may occur that the risk-neutral segment of the population is asymptotically zero.

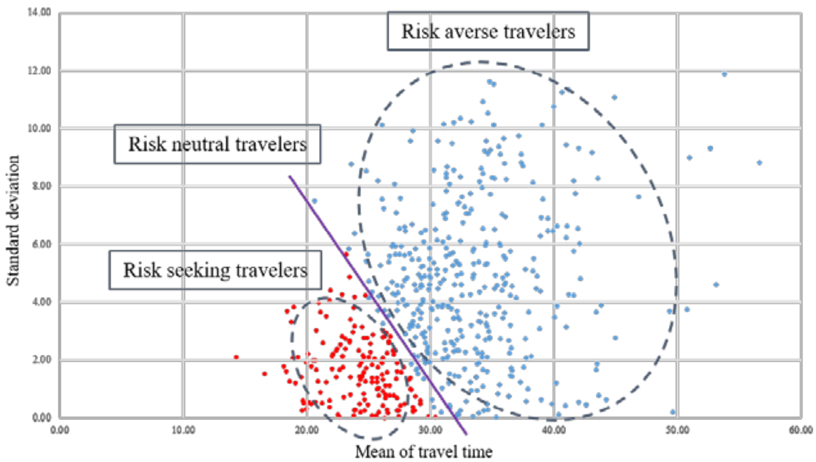


Figure 6.1 Risk preference in previous researches

It is possible to suggest the improved risk preference by comparison of the route choice characteristics toward risk preference for the traveler's behavior in route choice behavior.

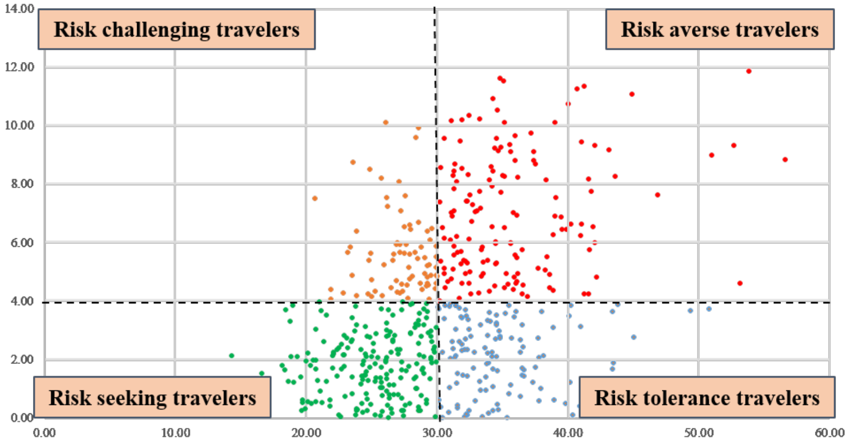
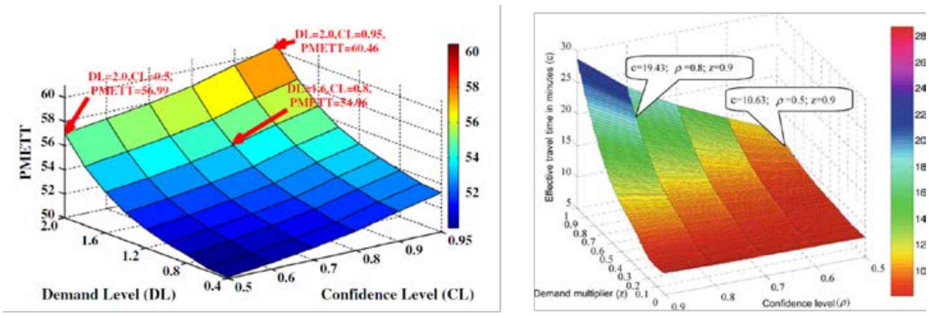


Figure 6.2 Improved risk preference

6.2.2 Travel time reliability model in SUE (Path-based Assignment)

It is possible to model construction combining demand fluctuation and travel time reliability model in the network. There are several models for SUE using sample network (i.e., Dupuis network, Sioux Falls network, and Etc.). After constructing the model, the actual route travel data can be used to estimate and verify the parameters of the model (i.e., Daegu DSRC data).

The developed model can be verified by the error between the actual route choice probability and the prediction probability using the observed path data. Moreover, sensitivity analysis improves the accuracy of the estimated parameters.



Source: Cited Chen et al.(2011) Fig.14 Source: Cited Shao et al.(2006) Fig.4

Figure 6.3 SUE model in previous researches

REFERENCES

1. Akgün, Vedat, Erhan Erkut, and Rajan Batta. 2000. "On finding dissimilar paths." *European Journal of Operational Research* 121 (2):232-246.
2. Avineri, Erel. 2006. "The effect of reference point on stochastic network equilibrium." *Transportation Science* 40 (4):409-420.
3. Avineri, Erel, and Joseph N Prashker. 2005. "Sensitivity to travel time variability: travelers' learning perspective." *Transportation Research Part C: Emerging Technologies* 13 (2):157-183.
4. Bekhor, Shlomo, Moshe E Ben-Akiva, and M Scott Ramming. 2006. "Evaluation of choice set generation algorithms for route choice models." *Annals of Operations Research* 144 (1):235-247.
5. Bekhor, Shlomo, Tomer Toledo, and Joseph N Prashker. 2008. "Effects of choice set size and route choice models on path-based traffic assignment." *Transportmetrica* 4 (2):117-133.
6. Ben-Akiva, Moshe, MJ Bergman, Andrew J Daly, and Rohit Ramaswamy. 1984. "Modeling inter-urban route choice behaviour." *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*.
7. Ben-Elia, Eran, Ido Erev, and Yoram Shiftan. 2008. "The combined effect of information and experience on drivers' route-choice behavior." *Transportation* 35 (2):165-177.
8. Ben-Elia, Eran, Robert Ishaq, and Yoram Shiftan. 2013. "“If only I had taken the other road...”: Regret, risk and reinforced learning in informed route-choice." *Transportation* 40 (2):269-293.
9. Bierlaire, Michel. 2016. *PythonBiogeme: a short introduction.* Report TRANSP-OR.
10. Bovy, Piet HL, and Stella Fiorenzo-Catalano. 2007. "Stochastic route choice set generation: behavioral and probabilistic foundations."

Transportmetrica 3 (3):173-189.

11. Bovy, Piet, Shlomo Bekhor, and Carlo Prato. 2008. "The factor of revisited path size: Alternative derivation." *Transportation Research Record: Journal of the Transportation Research Board* (2076):132-140.
12. Bovy, Piet HL. 2009. "On modelling route choice sets in transportation networks: a synthesis." *Transport reviews* 29 (1):43-68.
13. Cao, Zhiguang, Hongliang Guo, Jie Zhang, Dusit Niyato, and Ulrich Fastenrath. 2016. "Finding the shortest path in stochastic vehicle routing: A cardinality minimization approach." *IEEE Transactions on Intelligent Transportation Systems* 17 (6):1688-1702.
14. Cascetta, Ennio, Francesco Russo, Francesco A Viola, and Antonino Vitetta. 2002. "A model of route perception in urban road networks." *Transportation Research Part B: Methodological* 36 (7):577-592.
15. Chen, Anthony, and Zhaowang Ji. 2005. "Path finding under uncertainty." *Journal of advanced transportation* 39 (1):19-37.
16. Chen, Anthony, Zhaowang Ji, and Will Recker. 2002. "Travel time reliability with risk-sensitive travelers." *Transportation Research Record: Journal of the Transportation Research Board* (1783):27-33.
17. Chen, Anthony, and Zhong Zhou. 2010. "The α-reliable mean-excess traffic equilibrium model with stochastic travel times." *Transportation Research Part B: Methodological* 44 (4):493-513.
18. Chen, Anthony, Zhong Zhou, and William HK Lam. 2011. "Modeling stochastic perception error in the mean-excess traffic equilibrium model." *Transportation Research Part B: Methodological* 45 (10):1619-1640.
19. Chen, Bi Yu, William HK Lam, and Qingquan Li. 2016. "Efficient solution algorithm for finding spatially dependent reliable shortest path in road networks." *Journal of Advanced Transportation* 50 (7):1413-1431.
20. Chen, Bi Yu, William HK Lam, Agachai Sumalee, Qingquan Li, Hu Shao, and Zhixiang Fang. 2013. "Finding reliable shortest paths in road networks under uncertainty." *Networks and spatial economics* 13 (2):123-

148.

21. Chen, Bi Yu, William HK Lam, Agachai Sumalee, Qingquan Li, and Mei Lam Tam. 2014. "Reliable shortest path problems in stochastic time-dependent networks." *Journal of Intelligent Transportation Systems* 18 (2):177-189.
22. Chen, Bi Yu, William HK Lam, Agachai Sumalee, and Zhi-lin Li. 2012. "Reliable shortest path finding in stochastic networks with spatial correlated link travel times." *International Journal of Geographical Information Science* 26 (2):365-386.
23. Chen, Bi Yu, Qingquan Li, and William HK Lam. 2016. "Finding the k reliable shortest paths under travel time uncertainty." *Transportation Research Part B: Methodological* 94:189-203.
24. Chen, Bi Yu, Hui Yuan, Qingquan Li, Shih-Lung Shaw, William HK Lam, and Xiaoling Chen. 2016. "Spatiotemporal data model for network time geographic analysis in the era of big data." *International Journal of Geographical Information Science* 30 (6):1041-1071.
25. Chien, Steven I-Jy, and Chandra Mouly Kuchipudi. 2003. "Dynamic travel time prediction with real-time and historic data." *Journal of transportation engineering* 129 (6):608-616.
26. Chootinan, Piya, and Anthony Chen. 2011. "Confidence interval estimation for path flow estimator." *Transportation Research Part B: Methodological* 45 (10):1680-1698.
27. Chorus, Caspar G. 2014. "Risk aversion, regret aversion and travel choice inertia: an experimental study." *Transportation Planning and Technology* 37 (4):321-332.
28. Chorus, Caspar G, and Benedict GC Dellaert. 2012. "Travel choice inertia: the joint role of risk aversion and learning." *Journal of Transport Economics and Policy (JTEP)* 46 (1):139-155.
29. Connors, Richard D, and Agachai Sumalee. 2009. "A network equilibrium model with travellers' perception of stochastic travel times."

Transportation Research Part B: Methodological 43 (6):614-624.

30. de Palma, Andre, and Nathalie Picard. 2005. "Route choice decision under travel time uncertainty." *Transportation Research Part A: Policy and Practice* 39 (4):295-324.
31. Catalano, SF, and N Van der Zijpp. 2001. "A forecasting model for inland navigation based on route enumeration." *Proceedings of the AET European Transport Conference*, Held 10-12 September, 2001 Homerton College, Cambridge, UK-CD-ROM.
32. Fiorenzo-Catalano, Stella, Rob Van Nes, and Piet HL Bovy. 2004. "Choice set generation for multi-modal travel analysis." *European Journal of Transport and Infrastructure Research* 4 (2):195-209.
33. Fonzone, Achille, Jan-Dirk Schmöcker, Jiangshan Ma, and Daisuke Fukuda. 2012. "Link-Based Route Choice Considering Risk Aversion, Disappointment, and Regret." *Transportation Research Record: Journal of the Transportation Research Board* (2322):119-128.
34. Fosgerau, Mogens, and Daisuke Fukuda. 2012. "Valuing travel time variability: Characteristics of the travel time distribution on an urban road." *Transportation Research Part C: Emerging Technologies* 24:83-101.
35. Frejinger, Emma, Michel Bierlaire, and Moshe Ben-Akiva. 2009. "Sampling of alternatives for route choice modeling." *Transportation Research Part B: Methodological* 43 (10):984-994.
36. Gan, Hongcheng, and Yang Bai. 2014. "The effect of travel time variability on route choice decision: a generalized linear mixed model based analysis." *Transportation* 41 (2):339-350.
37. Gao, Song, Emma Frejinger, and Moshe Ben-Akiva. 2010. "Adaptive route choices in risky traffic networks: A prospect theory approach." *Transportation research part C: emerging technologies* 18 (5):727-740.
38. Gayah, Vikash V, Vinayak V Dixit, and S Ilgin Guler. 2014. "Relationship between mean and day-to-day variation in travel time in urban networks."

EURO Journal on Transportation and Logistics 3 (3-4):227-243.

39. Givon, Moshe. 1984. "Variety seeking through brand switching." *Marketing Science* 3 (1):1-22.
40. Guerriero, Francesca, and Patrizia Beraldi. 2008. "The α-reliable shortest path problem." *Algorithmic Operations Research* 3 (1).
41. Haghani, Milad, Zahra Shahhoseini, and Majid Sarvi. 2016. "Path sets size, model specification, or model estimation: Which one matters most in predicting stochastic user equilibrium traffic flow?" *Journal of Traffic and Transportation Engineering (English Edition)* 3 (3):181-191.
42. Hess, Stephane, and John M Rose. 2009. "Allowing for intra-respondent variations in coefficients estimated on repeated choice data." *Transportation Research Part B: Methodological* 43 (6):708-719.
43. Hess, Stephane, and David A Hensher. 2010. "Using conditioning on observed choices to retrieve individual-specific attribute processing strategies." *Transportation Research Part B: Methodological* 44 (6):781-790.
44. Hess, Stephane, and John M Rose. 2012. "Can scale and coefficient heterogeneity be separated in random coefficients models?" *Transportation* 39 (6):1225-1239.
45. Hess, Stephane, and Kenneth Train. 2017. "Correlation and scale in mixed logit models." *Journal of Choice Modelling* 23:1-8.
46. Hess, Stephane, and Kenneth E Train. 2011. "Recovery of inter-and intra-personal heterogeneity using mixed logit models." *Transportation Research Part B: Methodological* 45 (7):973-990.
47. Hoogendoorn-Lanser, Sascha, Rob van Nes, and Piet Bovy. 2005. "Path size modeling in multimodal route choice analysis." *Transportation Research Record: Journal of the Transportation Research Board* (1921):27-34.
48. Hunt, David, and Alain Kornhauser. 1996. "Assigning traffic over essentially-least-cost paths." *Transportation Research Record: Journal of*

the Transportation Research Board (1556):1-7.

49. Hutson, Kevin R, and Douglas R Shier. 2009. "Extended dominance and a stochastic shortest path problem." *Computers & Operations Research* 36 (2):584-596.

50. Jha, Mithilesh, Samer Madanat, and Srinivas Peeta. 1998. "Perception updating and day-to-day travel choice dynamics in traffic networks with information provision." *Transportation Research Part C: Emerging Technologies* 6 (3):189-212.

51. Ji, Xiaoyu. 2005. "Models and algorithm for stochastic shortest path problem." *Applied Mathematics and Computation* 170 (1):503-514.

52. Ji, Xiangfeng, Xuegang Jeff Ban, Mengtian Li, Jian Zhang, and Bin Ran. 2017. “Non-expected route choice model under risk on stochastic traffic networks.” *Networks and Spatial Economics*, 17(3), 777-807.

53. Jou, Rong-Chang, and Ke-Hong Chen. 2013. "An application of cumulative prospect theory to freeway drivers’ route choice behaviours." *Transportation Research Part A: Policy and Practice* 49:123-131.

54. Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the econometric society*, 47, 278.

55. Kitthamkesorn, Songyot, and Anthony Chen. 2013. "A path-size weibit stochastic user equilibrium model." *Transportation Research Part B: Methodological* 57:378-397.

56. Koppelman, Frank S, and Chieh-Hua Wen. 2000. "The paired combinatorial logit model: properties, estimation and application." *Transportation Research Part B: Methodological* 34 (2):75-89.

57. Lam, William HK, SC Wong, and Hong K Lo. 2009. *Transportation and Traffic Theory 2009: Golden Jubilee*: Springer.

58. Lee, Jang-Ho, Kyung-Soo Chon, and ChangHo Park. 2004. "Accommodating heterogeneity and heteroscedasticity in intercity travel mode choice model: Formulation and application to HoNam, South Korea,

high-speed rail demand analysis." *Transportation Research Record: Journal of the Transportation Research Board* (1898):69-78.

59. Lee, Jang-Ho, Shin-Hyung Cho, Dong-Kyu Kim, and Chungwon Lee. 2016. "Valuation of Travel Time Reliability Accommodating Heterogeneity of Route Choice Behaviors." *Transportation Research Record: Journal of the Transportation Research Board* (2565):86-93.

60. Li, Caixia, Sreenatha Gopalarao, and Tapabrata Ray. 2016. "A path-based flow formulation for the traffic assignment problem." *Transportation Planning and Technology* 39 (6):597-611.

61. Li, Dawei, Tomio Miwa, Takayuki Morikawa, and Pan Liu. 2016. "Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets." *Transportation Research Part C: Emerging Technologies* 67:31-46.

62. Li, Hao, Huizhao Tu, and David A Hensher. 2016. "Integrating the mean–variance and scheduling approaches to allow for schedule delay and trip time variability under uncertainty." *Transportation Research Part A: Policy and Practice* 89:151-163.

63. Li, Meng, and Hai-Jun Huang. 2017. "A regret theory-based route choice model." *Transportmetrica A: Transport Science* 13 (3):250-272.

64. Lo, Hong K, XW Luo, and Barbara WY Siu. 2006. "Degradable transport network: travel time budget of travelers with heterogeneous risk aversion." *Transportation Research Part B: Methodological* 40 (9):792-806.

65. Menghini, Gianluca, Nelson Carrasco, Nadine Schüssler, and Kay W Axhausen. 2010. "Route choice of cyclists in Zurich." *Transportation research part A: policy and practice* 44 (9):754-765.

66. Nie, Yu Marco, and Xing Wu. 2009. "Reliable a priori shortest path problem with limited spatial and temporal dependencies." *In Transportation and traffic theory 2009: golden jubilee*, 169-195. Springer.

67. Nie, Yu Marco, and Xing Wu. 2009. "Shortest path problem considering

on-time arrival probability." *Transportation Research Part B: Methodological* 43 (6):597-613.

68. Pan, Yiyong, Lu Sun, and Minli Ge. 2013. "Finding reliable shortest path in stochastic time-dependent network." *Procedia-Social and Behavioral Sciences* 96:451-460.
69. Peer, SK, and Dinesh K Sharma. 2007. "Finding the shortest path in stochastic networks." *Computers & Mathematics with Applications* 53 (5):729-740.
70. Pillat, Juliane, Eileen Mandir, and Markus Friedrich. 2011. "Dynamic Choice Set Generation Based on Global Positioning System Trajectories and Stated Preference Data." *Transportation Research Record: Journal of the Transportation Research Board* (2231):18-26.
71. Prashker, Joseph N, and Shlomo Bekhor. 2004. "Route choice models used in the stochastic user equilibrium problem: a review." *Transport reviews* 24 (4):437-463.
72. Prato, Carlo, and Shlomo Bekhor. 2006. "Applying branch-and-bound technique to route choice set generation." *Transportation Research Record: Journal of the Transportation Research Board* (1985):19-28.
73. Prato, Carlo, Thomas Rasmussen, and Otto Nielsen. 2014. "Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers." *Transportation Research Record: Journal of the Transportation Research Board* (2412):20-27.
74. Prato, Carlo Giacomo. 2009. "Route choice modeling: past, present and future research directions." *Journal of Choice Modelling* 2 (1):65-100.
75. Ramezani, Mohsen, and Nikolas Geroliminis. 2012. "On the estimation of arterial route travel time distribution with Markov chains." *Transportation Research Part B: Methodological* 46 (10):1576-1590.
76. Ramming, Michael Scott. 2001. *Network knowledge and route choice.* Massachusetts Institute of Technology.
77. Ramos, Giselle Moraes, Winnie Daamen, and Serge Hoogendoorn. 2014.

"A state-of-the-art review: developments in utility theory, prospect theory and regret theory to investigate travellers' behaviour in situations involving travel time uncertainty." *Transport Reviews* 34 (1):46-67.

78. Rasmussen, Thomas Kjær, David Paul Watling, Carlo Giacomo Prato, and Otto Anker Nielsen. 2015. "Stochastic user equilibrium with equilibrated choice sets: Part ii–solving the restricted sue for the logit family." *Transportation Research Part B: Methodological* 77:146-165.
79. Rasouli, Soora, and Harry Timmermans. 2014. "Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research." *Travel Behaviour and Society* 1 (3):79-90.
80. Razo, Michael, and Song Gao. 2013. "A rank-dependent expected utility model for strategic route choice with stated preference data." *Transportation Research Part C: Emerging Technologies* 27:117-130.
81. Revelt, David, and Kenneth Train. 1998. "Mixed logit with repeated choices: households' choices of appliance efficiency level." *Review of economics and statistics* 80 (4):647-657.
82. Ryu, Seungkyu, Anthony Chen, Xiangdong Xu, and Keechoo Choi. 2014. "Modeling demand elasticity and route overlapping in stochastic user equilibrium through paired combinatorial logit model." *Transportation Research Record: Journal of the Transportation Research Board* (2429):8-19.
83. Shao, Hu, William HK Lam, and Mei Lam Tam. 2006. "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand." *Networks and Spatial Economics* 6 (3):173-204.
84. Sharifi, Mohammad Sadra, Anthony Chen, Songyot Kitthamkesorn, and Ziqi Song. 2015. "Link-Based Stochastic Loading Methods for Weibit Route Choice Model." *Transportation Research Record: Journal of the*

Transportation Research Board (2497):84-94.

85. Sheffi, Yosef, and Warren B Powell. 1982. "An algorithm for the equilibrium assignment problem with random link times." *Networks* 12 (2):191-207.
86. Sheng, Yuhong, and Yuan Gao. 2016. "Shortest path problem of uncertain random network." *Computers & Industrial Engineering* 99:97-105.
87. Siu, Barbara WY, and Hong K Lo. 2008. "Doubly uncertain transportation network: degradable capacity and stochastic demand." *European Journal of Operational Research* 191 (1):166-181.
88. Srinivasan, Karthik K, AA Prakash, and Ravi Seshadri. 2014. "Finding most reliable paths on networks with correlated and shifted log–normal travel times." *Transportation Research Part B: Methodological* 66:110-128.
89. Sumalee, Agachai, David Watling, and Shoichiro Nakayama. 2006. "Reliable network design problem: case with uncertain demand and total travel time reliability." *Transportation Research Record: Journal of the Transportation Research Board* (1964):81-90.
90. Sun, Chao, Lin Cheng, Senlai Zhu, Fei Han, and Zhaoming Chu. 2017. "Multi-criteria user equilibrium model considering travel time, travel time reliability and distance." *Transportation Research Part D: Transport and Environment*. Online published.
91. Susilawati, Susilawati, Michael AP Taylor, and Sekhar VC Somenahalli. 2013. "Distributions of travel time variability on urban roads." *Journal of Advanced Transportation* 47 (8):720-736.
92. Tang, Yue, Song Gao, and Eran Ben-Elia. 2017. "An exploratory study of instance-based learning for route choice with random travel times." *Journal of Choice Modelling*. 24:22-35.
93. Train, Kenneth. 1978. "A validation test of a disaggregate mode choice model." *Transportation Research* 12 (3):167-174.
94. Train, Kenneth. 2016. "Mixed logit with a flexible mixing distribution."

Journal of choice modelling 19:40-53.

95. Tversky, Amos, and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty* 5 (4):297-323.
96. Uchida, Kenetsu. 2014. "Estimating the value of travel time and of travel time reliability in road networks." *Transportation research part B: methodological* 66:129-147.
97. Van der Zij, N.J., and S. Fiorenzo-Catalano. “Path Enumeration By Finding The Constrained K-Shortest Paths.” *Transportation Research Part B: Methodological*, 39(6):545-563.
98. Vacca, Alessandro, Carlo Giacomo Prato, and Italo Meloni. 2015. "Estimating Route Choice Models from Stochastically Generated Choice Sets on Large-Scale Networks: Correcting for Unequal Sampling Probability." *Transportation Research Record: Journal of the Transportation Research Board* (2493):11-18.
99. Vovsha, Peter, and Shlomo Bekhor. 1998. "Link-nested logit model of route choice: overcoming route overlapping problem." *Transportation Research Record: Journal of the Transportation Research Board* (1645):133-142.
100. Vreeswijk, Jaap, Tom Thomas, Eric van Berkum, and Bart van Arem. 2013. "Drivers' perception of route alternatives as indicator for the indifference band." *Transportation Research Record: Journal of the Transportation Research Board* (2383):10-17.
101. Wang, Guangchao, Shoufeng Ma, and Ning Jia. 2013. "A combined framework for modeling the evolution of traveler route choice under risk." *Transportation Research Part C: Emerging Technologies* 35:156-179.
102. Wang, Judith YT, Matthias Ehrgott, and Anthony Chen. 2014. "A bi-objective user equilibrium model of travel time reliability in a road

network." *Transportation Research Part B: Methodological* 66:4-15.

103. Watling, David. 2006. "User equilibrium traffic network assignment with stochastic travel times and late arrival penalty." *European journal of operational research* 175 (3):1539-1556.

104. Watling, David Paul, Thomas Kjær Rasmussen, Carlo Giacomo Prato, and Otto Anker Nielsen. 2015. "Stochastic user equilibrium with equilibrated choice sets: Part I–Model formulations under alternative distributions and restrictions." *Transportation Research Part B: Methodological* 77:166-181.

105. Wu, Xing, and Yu Nie. 2009. "Implementation issues for the reliable a priori shortest path problem." *Transportation Research Record: Journal of the Transportation Research Board* (2091):51-60.

106. Wu, Xing, and Yu Marco Nie. 2011. "Modeling heterogeneous risk-taking behavior in route choice: A stochastic dominance approach." *Transportation research part A: policy and practice* 45 (9):896-915.

107. Xie, Chi, and Zugang Liu. 2014. "On the stochastic network equilibrium with heterogeneous choice inertia." *Transportation Research Part B: Methodological* 66:90-109.

108. Xu, Hongli, Yingyan Lou, Yafeng Yin, and Jing Zhou. 2011. "A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing." *Transportation Research Part B: Methodological* 45 (2):311-328.

109. Xu, Hongli, Jing Zhou, and Wei Xu. 2011. "A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory." *Transportation Research Part C: Emerging Technologies* 19 (2):218-228.

110. Xu, Xiangdong, Anthony Chen, and Lin Cheng. 2013. "Assessing the effects of stochastic perception error under travel time variability." *Transportation* 40 (3):525-548.

111. Xu, Xiangdong, Anthony Chen, Lin Cheng, and Hong K Lo. 2014.

"Modeling distribution tail in network performance assessment: A mean-excess total travel time risk measure and analytical estimation method." *Transportation Research Part B: Methodological* 66:32-49.

112. Xu, Xiangdong, Anthony Chen, Zhong Zhou, and Shlomo Bekhor. 2012. "Path-based algorithms to solve c-logit stochastic user equilibrium assignment problem." *Transportation Research Record: Journal of the Transportation Research Board* (2279):21-30.
113. Yai, Tetsuo, Seiji Iwakura, and Shigeru Morichi. 1997. "Multinomial probit with structured covariance for route choice behavior." *Transportation Research Part B: Methodological* 31 (3):195-207.
114. Zhan, Xianyuan, Samiul Hasan, Satish V Ukkusuri, and Camille Kamga. 2013. "Urban link travel time estimation using large-scale taxi data with partial information." *Transportation Research Part C: Emerging Technologies* 33:37-49.
115. Zhang, Junlin, and Hai Yang. 2015. "Modeling route choice inertia in network equilibrium with heterogeneous prevailing choice sets." *Transportation Research Part C: Emerging Technologies* 57:42-54.
116. Zhang, Yuli, Zuo-Jun Max Shen, and Shiji Song. 2017. "Lagrangian relaxation for the reliable shortest path problem with correlated link travel times." *Transportation Research Part B: Methodological*. 204:501-521.
117. Zhou, Zhong, and Anthony Chen. 2008. "Comparative analysis of three user equilibrium models under stochastic demand." *Journal of Advanced Transportation* 42 (3):239-263.
118. Zockaie, Ali, Yu Nie, Xing Wu, and Hani Mahmassani. 2013. "Impacts of Correlations on Reliable Shortest Path Finding: A Simulation-Based Study." *Transportation Research Record: Journal of the Transportation Research Board* (2334):1-9.

APPENDIX

Appendix 1. Notations

SIGN	NOTATION
i	The origin node
j	The destination node
k	The order of α-reliable path or predetermined the number of route choice set
l	The individual traveler
p^{ij}	The observed paths from the origin node i to the destination node j
p_k^{ij}	The k^{th} α-reliable path from the origin node i to the destination node j
T^{ij}	The travel time distribution from the origin node i to the destination node j
T_k^{ij}	The travel time distribution of k^{th} α-reliable path from the origin node i to the destination node j
T^{ijl}	The perceived travel time distribution for individual l from the origin node i to the destination node j
U^{ij}	Total route choice set from the origin node i to the destination node j
μ^{ij}	The mean of travel time distribution from the origin node i to the destination node j
μ_k^{ij}	The mean of travel time distribution of k^{th} α-reliable path from the origin node i to the destination node j
μ^{ijl}	The perceived mean of travel time distribution for individual l from the origin node i to the destination node j
σ^{ij}	The standard deviation of travel time distribution from the origin node i to the destination node j
σ_k^{ij}	The standard deviation of travel time distribution of k^{th} α-reliable path from the origin node i to the destination node j
σ^{ijl}	The perceived standard deviation of travel time distribution for individual l from the origin node i to the destination node j
α	Confidence level α (i.e., On-time arrival probability)

α_l	Individual confidence level α_l for individual l from the origin node i to the destination node j
Z_α	The inverse of cumulative distribution function of the standard normal distribution at confidence level α
$\Phi^{-1}_{T^{ij}}(\alpha)$	The value of normalized distribution function to achieve α confidence level from the origin node i to the destination node j
$\Phi^{-1}_{T^{ij}_k}(\alpha)$	The value of normalized distribution function to achieve α confidence level for k[th] α-reliable path from the origin node i to the destination node j
$\Phi^{-1}_{T^{ijl}}(\alpha_l)$	The value of normalized distribution function to achieve α individual confidence level for individual l from the origin node i to the destination node j
$TTB_{T^{ij}}(\alpha)$	Travel time budget(TTB) required achieving α confidence level from the origin node i to the destination node j
$TTB_{T^{ij}_k}(\alpha)$	Travel time budget(TTB) of k[th] path required achieving α confidence level from the origin node i to the destination node j
$TTB_{T^{ijl}}(\alpha_l)$	Travel time budget(TTB) required achieving α individual confidence level for individual l from the origin node i to the destination node j

Appendix 2. Data list for OD pairs (40 OD pairs)

Origin		Destination		Land Use (Origin)	Observation	Corrected observation	Distance	Observed choice set	Attribute
ID	Name	ID	Name						
20008	Manpyeong intersection	20025	Sungseo intersection	Industrial	5,664	5,568	7.282	2.704	SS
20021	Seosaegu industry intersection	20107	Gongsansuwonji intersection	Industrial	9,939	8,881	8.262	2.642	SU
20024	Sindang intersection	20107	Gongsansuwonji intersection	Industrial	13,132	8,498	13.685	2.572	SS
20030	Bangogae intersection	20149	Suseong market intersection	Residential	7,399	7,283	4.274	2.365	UU
20031	Sinnam intersection	20168	Gongpyeong intersection	CBD	17,989	16,327	5.170	2.752	UU
20032	Banwoldang intersection	20145	Doosan bridge	CBD	10,391	9,936	5.124	2.915	US
20034	Beomeo intersection	20174	Docheong bridge (south)	CBD	4,213	4,091	5.170	2.522	UU
20043	Wolchon intersection	20142	Beoman intersection	Residential	3,582	3,276	12.948	2.462	SS
20044	Sangin intersection	20141	Gwangye intersection	Residential	4,977	4,917	10.793	2.167	SS
20050	Jincheonnam intersection	20104	Gyungdae bridge	Residential	6,642	5,536	15.670	3.628	SU
20051	Wolgok intersection	20032	Banwoldang intersection	Residential	7,391	7,267	9.331	3.329	SU
20063	Sangyeok bridge intersection	20125	Hyomok overpass	Industrial	16,208	15,858	4.333	2.133	SU
20071	Sangdong bridge	20051	Wolgok intersection	Residential	15,211	14,600	7.598	2.349	SS
20073	Seobyun bridge	20130	Daebong bridge	Residential	9,784	7,829	7.647	2.982	SU

20085	Gosung intersection	20079	Sinpyeong intersection	Residential	10,278	10,266	3.564	3.171	US
20087	Youngnam univ. hospital intersection	20031	Sinnam intersection	University	4,988	4,962	2.896	2.786	SU
20090	Bokhyun intersection	20170	Daegu station intersection	Residential	19,702	19,249	3.460	2.362	SU
20098	Dogok intersection	20150	Cheonggu intersection	Residential	14,369	13,610	7.874	3.357	SU
20101	Chilsung homeplus intersection	20144	Doosan five-way intersection	CBD	6,027	4,345	9.295	3.215	US
20102	Chimsan intersection	20034	Beomeo intersection	Residential	16,812	12,543	5.950	3.324	UU
20102	Chimsan intersection	20144	Doosan five-way intersection	Residential	9,641	7,135	9.335	3.337	US
20107	Gongsan suwonji intersection	20115	Hyundai eunha mansion intersection	Industrial	4,069	3,941	7.331	2.091	SS
20107	Gongsan suwonji intersection	20134	Bisan intersection	Industrial	10,059	9,914	5.385	3.462	SU
20111	Banyawol intersection	20152	Hyosin intersection	Residential	19,555	19,525	7.487	2.259	SU
20124	Dongdaegu station intersection	20171	Gyodong intersection	Commercial	16,725	14,241	2.391	2.643	UU
20125	Hyomok overpass	20136	Yunho intersection	Residential	9,999	7,550	5.718	2.690	SS
20129	Namgu office intersection	20151	MBC intersection	Residential	4,747	4,304	5.214	3.034	SU
20132	Myungduk intersection	20100	Namchimsan intersection	Residential	6,332	6,300	3.004	2.572	UU
20132	Myungduk intersection	20144	Doosan five-way intersection	Residential	11,300	10,821	5.773	3.278	US
20137	Damtigogae overpass	20032	Banwoldang intersection	Residential	10,539	10,447	6.217	2.171	SU
20142	Beoman intersection	20163	Bongsan six-way intersection	Residential	10,219	10,040	8.426	2.416	SU

20144	Doosan five-way intersection	20139	Sunsportplaza three-way intersection	Residential	13,566	13,519	3.590	2.661	SS
20144	Doosan five-way intersection	20132	Myungduk intersection	Residential	12,433	12,002	5.833	4.135	SU
20144	Doosan five-way intersection	20102	Chimsan intersection	Residential	9,725	8,784	9.335	4.517	SU
20146	Deulangil intersection	20139	Sunsportplaza three-way intersection	Residential	20,866	20,846	3.156	2.526	SS
20150	Cheonggu intersection	20050	Jincheonnam intersection	Residential	7,091	6,806	13.729	2.990	US
20151	MBC intersection	20166	Seosung intersection	Commercial	8,475	8,387	4.039	2.414	UU
20152	Hyosin intersection	20130	Daebong bridge	Residential	5,776	5,682	4.071	2.749	UU
20161	Gyemyung intersection	20151	MBC intersection	University	10,283	10,045	5.340	3.232	SU
20162	Gyesan intersection	20025	Sungseo intersection	CBD	16,275	16,184	8.262	2.274	US
20164	Dongsin bridge	20051	Wolgok intersection	CBD	9,651	9,147	11.328	3.170	US
20166	Seosung intersection	20130	Daebong bridge	CBD	2,978	2,821	3.111	2.415	UU

Annotation: UU(Travel for urban to urban), US(Travel for urban to suburb), SS(Suburb to suburb), SU(Suburb to suburb)

초 록

통행시간 신뢰도의 개별 리스크 선호도를 고려한 경로선택행태 모형

서울대학교 대학원
공과대학 건설환경공학부
조 신 형

경로의 선택문제는 교통운영 및 교통계획 분야에서 중요한 요소이다. Revealed preference 데이터와 Stated preference 데이터를 이용하여 통행자의 경로선택 특성을 분석하고자 하는 연구가 많이 수행되었다. 데이터의 한계로 인하여 적정 수준의 경로집합을 생성하여 선택모형을 구축하는 데에 한계가 있었다.

본 연구에서는 관측된 데이터를 활용하여 통행자들의 경로집합생성 모형을 구축하고 확률적 경로선택모형을 구축하는 연구를 수행한다. 통행자들의 서로 다른 경로에 대한 선호다양성을 경로생성모형과 경로선택모형에 반영하여 최적의 모형을 선정하고 모수를 추정하였다.

경로집합생성에서 통행자 개인별 선호다양성을 반영하기 위해 경험한 통행시간 분포에 따른 개인별 신뢰수준을 정의하고

경로집합을 도출하였다. 또한, 경로선택모형에서 경로의 중첩과 선택에 대한 개인별 선호다양성을 반영한 모형을 이용하여 모수를 추정하였다.

본 연구를 통하여 통행자 개인의 경로집합과 분석적 측면의 경로집합생성을 비교하고 모형화하였다. 또한, 경로집합생성모형과 경로선택모형에 대해 개인별 선호다양성을 반영함으로써 보다 우수한 모수를 추정할 수 있게 되었다. 기존의 경로집합생성모형에 따른 결과와의 비교를 통하여 본 연구의 우수성을 제고할 수 있다. 본 연구는 교통수요예측이나 교통운영적인 관점에서 도시내 경로선택에 대한 개인별 특성을 반영하여 현실적인 모사가 보다 우수한 것을 확인할 수 있다.

주요어: 경로집합생성모형, 경로선택모형, 선호다양성, 통행시간 신뢰도, 통행시간예산, 위험선호

학 번: 2011-21003