DO TAXI DRIVERS CHOOSE THE SHORTEST ROUTES?

A Large-Scale Analysis of Route Choice Behavior of Taxi Drivers Using Floating Car Data in Shanghai

Junyan Li

Superviors:
Juliane Cron M.Sc. (Technische Universität München)
Univ.Prof. Mag.rer.nat. Dr.rer.nat Georg Gartner (Technische Universität Wien)
Dr. Haosheng Huang (Universität Zürich)

CONTENTS

Part 1

INTRODUCTION

1. Introduction

Context:

•**Route choice** is one of the fundamental features of **human wayfinding** (Golledge,1995)

•Understanding route choice behavior of drivers:

- urban planning
- traffic management
- infrastructure construction

1. Introduction

Context:

•**Following the shortest route** is traditionally believed to be one of the major features of human route choice behavior. (Golledge, 1995).

•Compared to **ordinary drivers**, **taxi drivers** are more experienced in wayfinding (Yao et al., 2013).

1. Introduction

Motivation:

- **Several questions need to be addressed:**
  - humans decision making process : rational + irrational components (de Carvalho et al., 2016).
    - makes accurate prediction of route selection difficult.
  - While theories predict that multiple equilibriums that result in choosing different routes (Smith, 1979, 1983, 1984),
    - hard to justify the theory prediction without large scale observation data
  - observation data : GPS-based point data.
    - needs pre-processing and integration with road network (Guo and Huang,2016)
- **availability of large scale Floating Car Data (FCD)**

Introduction

Research questions:

- To what extent do real routes overlap with the shortest routes, and how much longer are real routes compared with the shortest routes?

- Are other factors *(e.g. speed limitation, road type, number of left turns, number of intersections, road usage frequency)* impacting the deviation of real routes from the shortest routes?

Introduction

Importance:

•most of the existing studies focus on the behavior of ordinary drivers

•to provide cross-validation for previous studies with large scale empirical data

•To help better understanding of the patterns of route choice.

Part 2

THEORETICAL & TECHNICAL BACKGROUND

Theoretical & technical background

Route choice behavior

- Rationality and bounded rationality
  - Rationality is the fundamental assumption in many models of drivers' route choice behavior (Miller-Hooks and Mahmassani, 2000; Razo and Gao, 2013; Smith, 1979, 1983, 1984)
  - Irrationality component of drivers' evaluation also adds to the complexity of the problem. (de Carvalhoet al., 2016; Guo and Huang, 2016).
- Travel distance (Wardrop, 1952; Tang et al., 2016; Ciscal-Terry et al., 2016; Manley et al., 2015).
- Travel time and congestion (Prato et al.,2014; Prato and Rasmussen, 2016).
- Road type (Manley et al., 2015).

Theoretical & technical background

Floating Car Data:

- **GPS-based FCD** is one of the reliable traffic data sources (Paulin and Bessler,2013)
- It provides detailed trip information, such as:
    - longitude, latitude
    - time
    - speed
    - direction
- **Advantages**:
    - FCD reveals much more information, compared to survey data
    - stores information of massive number of taxi drivers with relatively cheap costs.
    - allows to systematically examine the route choice behavior of taxi drivers.

Theoretical & technical background

Floating Car Data:

•At the **macro** level, FCD is used to construct real time traffic information system as well as inference of mobilities and congestions.(e.g. Schafer et al., 2002; Kerner et al., 2005; Huber et al., 1999 and Gühnemann et al., 2004)

•At the **micro** level, FCD is used as a way to predict various information from individual drivers. (e.g. Simroth and Zähle, 2011; Tang et al., 2016; Yao et al., 2013; Manley et al., 2015)

Theoretical & technical background

Map matching:

Map matching is the process of integrating the GPS-based point with spatial road network data to identify the location of a vehicle on the road network.

Quddus et al. (2007) provide an overview of algorithms.

Theoretical & technical background

Map matching:

•geometric-based algorithms

- point-to-point matching (Bentley and Maurer, 1980; Bernstein and Kornhauser, 1998)
- point-to-curve matching (White et al., 2000)
- curve-to-curve matching (Phuyal, 2002; White et al., 2000)
- Limitations:
  - based on the geometric distances between the GPS points and the road network.
  - sensitive to road network density and are unstable with presence of outliers

Theoretical & technical background

Map matching:

•topological analysis

- take into account of the topological features of road networks
  - geometry
  - connectivity
  - contiguity
- commonly adopt a weighted scheme that utilizing available information including *road turn*, *travel speed*, *travel direction*, and *similarity* between the vehicle trajectory and the road network (Greenfeld, 2002; Quddus et al., 2003; Yu, 2006).

Map matching:

- More advanced : e.g. ***fuzzy logic and Bayesian analysis*** (El Najjar and Bonnifait, 2005; Krakiwsky et al., 1988; Kim,1998; Newson and Krumm, 2009)

- regards the GPS points as a serial revealing of **latent positions** on the road network that transfer between different states from previous to current time (Newson and Krumm, 2009).

- **Hidden Markov Model :** The probability of a set of GPS points located on certain combinations of road segments.
  - nearest early position
  - transferring probability matrix

- won't be impacted too much by outliers.

Theoretical & technical background

Algorithms for finding the shortest route:

•Shortest path:

- The problem to find the shortest path by constructing a tree with minimum total length between n node (Dijkstra,1959)

•K shortest path:

- Given the need to compare a set of routes between two end nodes, the k-shortest path algorithm was developed to find the specified number of shortest routes from the road network in an order from 1 to k. (Hoffman and Pavley,1959; Yen,1971)
- Wilson(1996) introduces the algorithm with additional weighting scheme.

Theoretical & technical background

State of the art and research gaps

- **Yao et al. (2013) – FCD Data in Beijing**
  - Influencing factors: faster travel speed, less left turns and more proportion of highways.
  - a very small subsample of taxi trips (in total 221 trips)

Theoretical & technical background

State of the art and research gaps

•Yao et al. (2013) – FCD Data in Beijing

- Influencing factors: faster travel speed, less left turns and more proportion of highways.
- a very small subsample of taxi trips (in total 221 trips)

•Sun et al. (2014) – FCD Data in Shenzhen

- Influencing factors: travel distance, travel time and road preference
- total number of trips is around 4000

Theoretical & technical background

State of the art and research gaps

•Yao et al. (2013) – FCD Data in Beijing

- Influencing factors: faster travel speed, less left turns and more proportion of highways.
- a very small subsample of taxi trips (in total 221 trips)

•Sun et al. (2014) – FCD Data in Shenzhen

- Influencing factors: travel distance, travel time and road preference
- total number of trips is around 4000

•Manley et al. (2015) – FCD Data in London

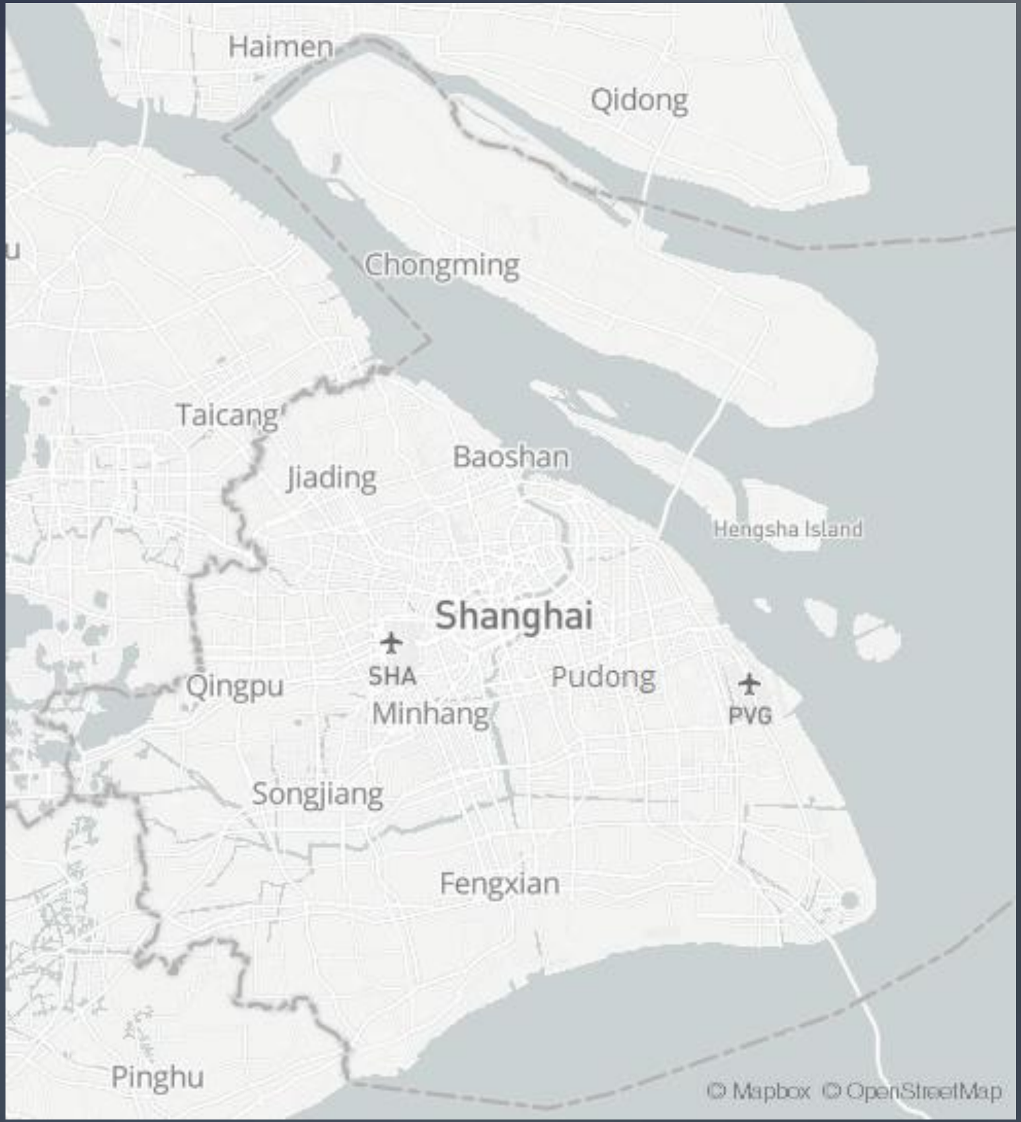
- Influencing factors: travel time, turns and road category
- total number of trips is 700,000 with about 3000 drivers,
- limitations:
  - instead of analyzing the whole London network, the authors only select cases in smaller regions, which may not be representing the whole city.
  - the number of trips in their study is still small than real situations.

Part 3

DATA & STUDY AREA

Data & Study Area

Area of Study:



Data & Study Area

Data Structure:

Column	Attributes	Example
Date	Date of FCD Data	*20100615*
Taxi id	taxi id, consists 5 letters	*10003*
Lon	longitude of current location	*121.529558*
Lat	latitude of current location	*31.23496*
Velocity	instant speed	*45.5*
Direction	driving orientation (valid number from 0 to 359)	*155*
Passenger	if the taxi is with passenger or not (No - 0, Yes – 1)	*1*
Time	exact time of current location	*2010-06-15 09:10:05*

Data & Study Area

Valid number of points and routes for each day included in the analysis:

Around **3,6 million** valid trips, and **650 million** FCD Points

Date	FCD Points	Routes
20100615	55,083,889	275,958
20100616	56,300,003	268,583
20100617	51,978,838	272,594
20100618	47,390,204	278,074
20100619	50,444,522	290,734
20100620	51,259,864	258,392
20100622	47,272,895	271,188
20100623	48,146,691	280,074
20100624	48,267,510	283,901
20100625	48,598,111	295,160
SUM	648,428,867	3,617,821

Part 4

PRE-PROCESSING

Pre-processing

- **Identification of OD pairs**
- **Filtering of irrelevant points**
- **Map Matching**

Pre-processing

- **Identification of OD pairs**
- Filtering of irrelevant points
- Map Matching

Pre-processing

Identification of OD pairs

Taxi driver ID	Time	Passenger	Lon	Lat
...	...	...	...	...
1010	00:00	0	X	X
1010	00:10	1	X	X
1010	00:20	1	X	X
1010	00:30	1	X	X
1010	00:40	1	X	X
1010	00:50	0	X	X
1010	00:60	0	X	X
1010	01:10	0	X	X
1010	01:20	1	X	X
1010	01:30	1	X	X
...	...	...	...	...

Pick-up (origin)

Drop-off (destination)

Pick-up (origin)

Pre-processing

- Identification of OD pairs
- **Filtering of irrelevant points**
- Map Matching

Pre-processing

Filtering of irrelevant points:

- ***Why?***

Pre-processing

Filtering of irrelevant points:

- ***Why?***
  - taxi driver may forget to turn off the passenger-on-board indicator when the taxi is off-service and not moving.

Pre-processing

Filtering of irrelevant points:

- ***Why?***
  - taxi driver may forget to turn off the passenger-on-board indicator when the taxi is off-service and not moving.
  - taxi drivers may mistakenly turn off or turn on the passenger-on-board indicator in a very short time

Pre-processing

- Identification of OD pairs
- Filtering of irrelevant points
- **Map Matching**

Pre-processing

Map Matching:

Hidden Markov Model (HMM) (Newson and Krumm, 2009)

Consider:
- recorded GPS locations
- timestamp of recorded points
- connectivity of the road network

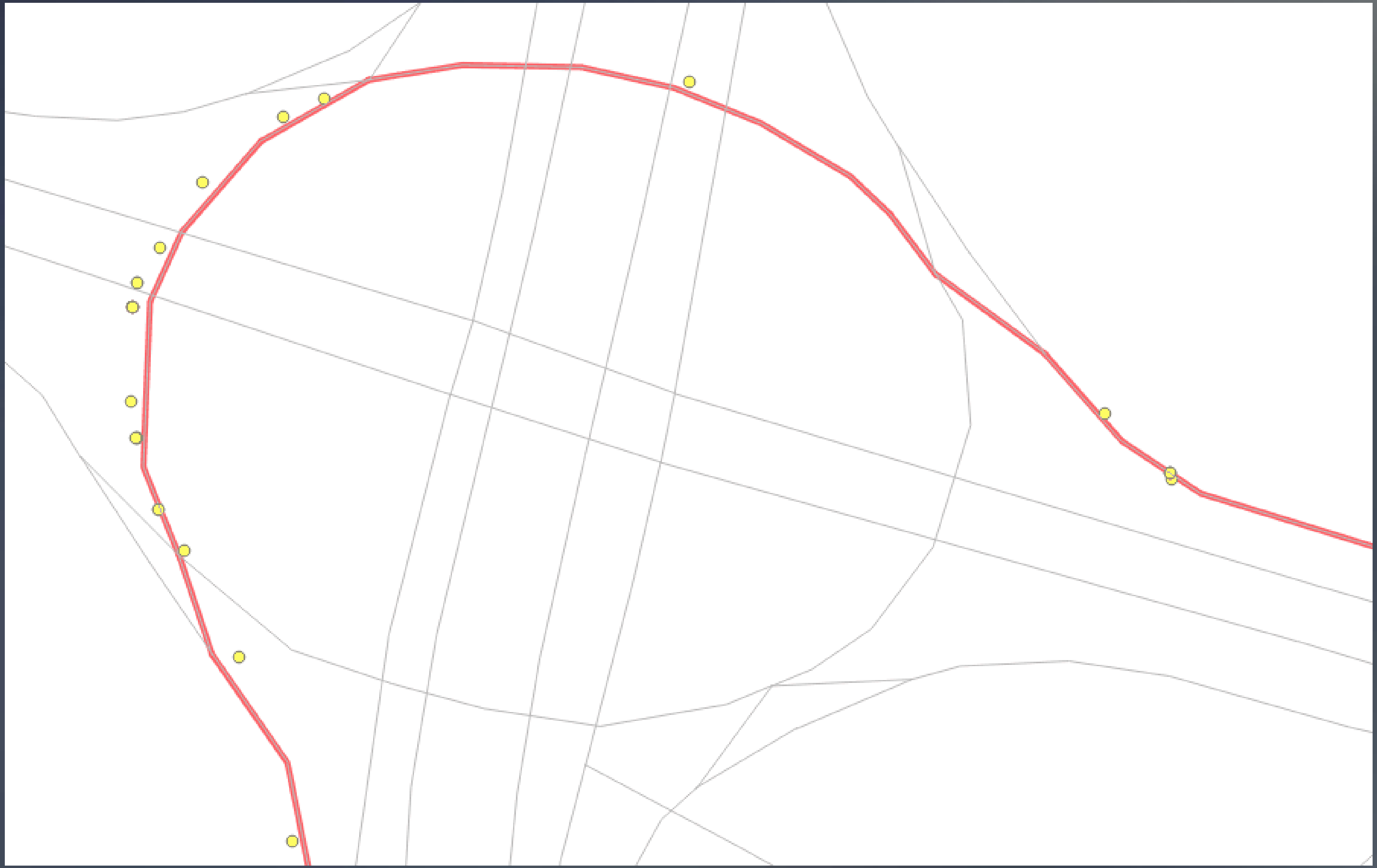
Calculating the path with the highest likelihood

Pre-processing

Map Matching:

Pre-processing

Map Matching:



Part 5

OVERALL PATTERNS

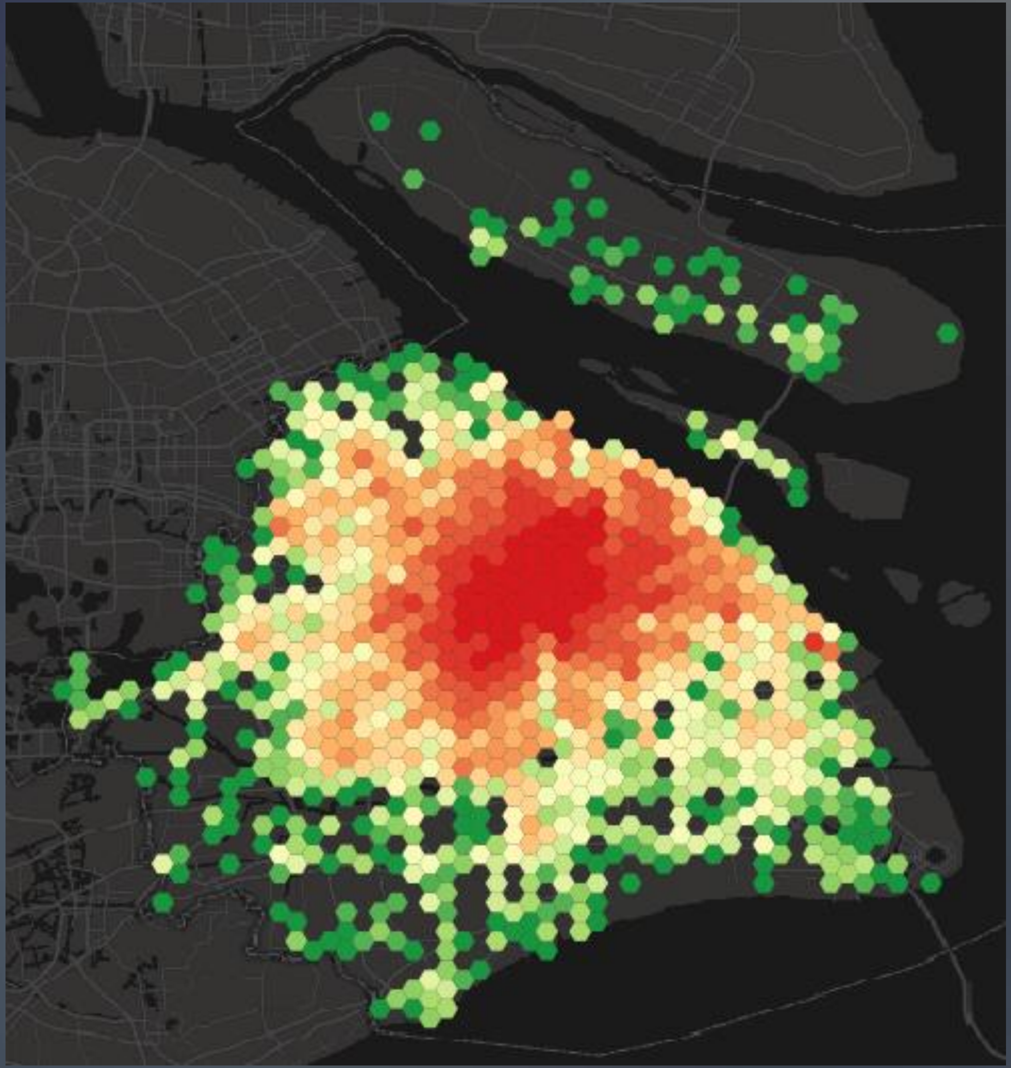
Overall Patterns

- **Distribution of service areas**
- Taxi service activities
- Real routes

Overall Patterns

Distribution of service areas:

centered around city center

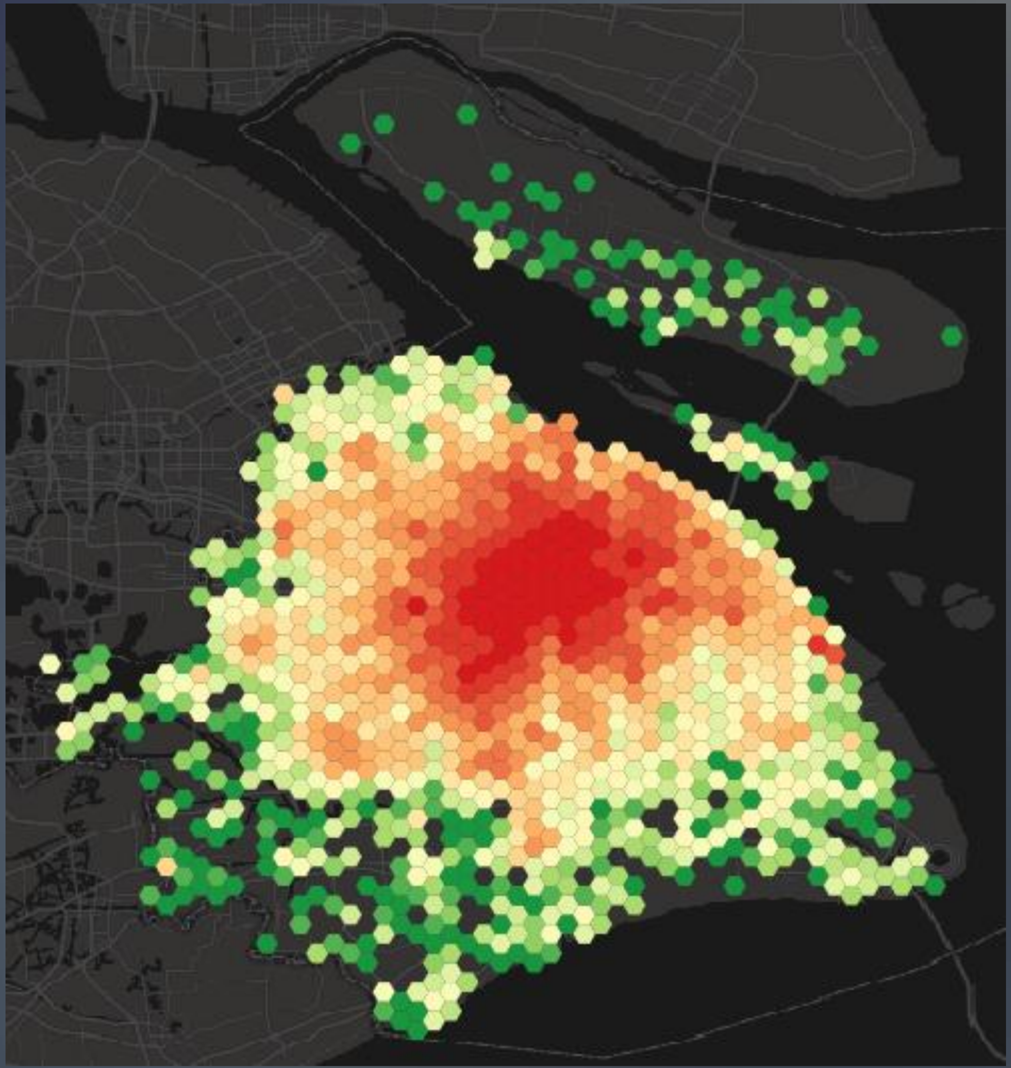


Density of the pick-up locations

Overall Patterns

Distribution of service areas:

centered around city center



Density of the drop-off locations

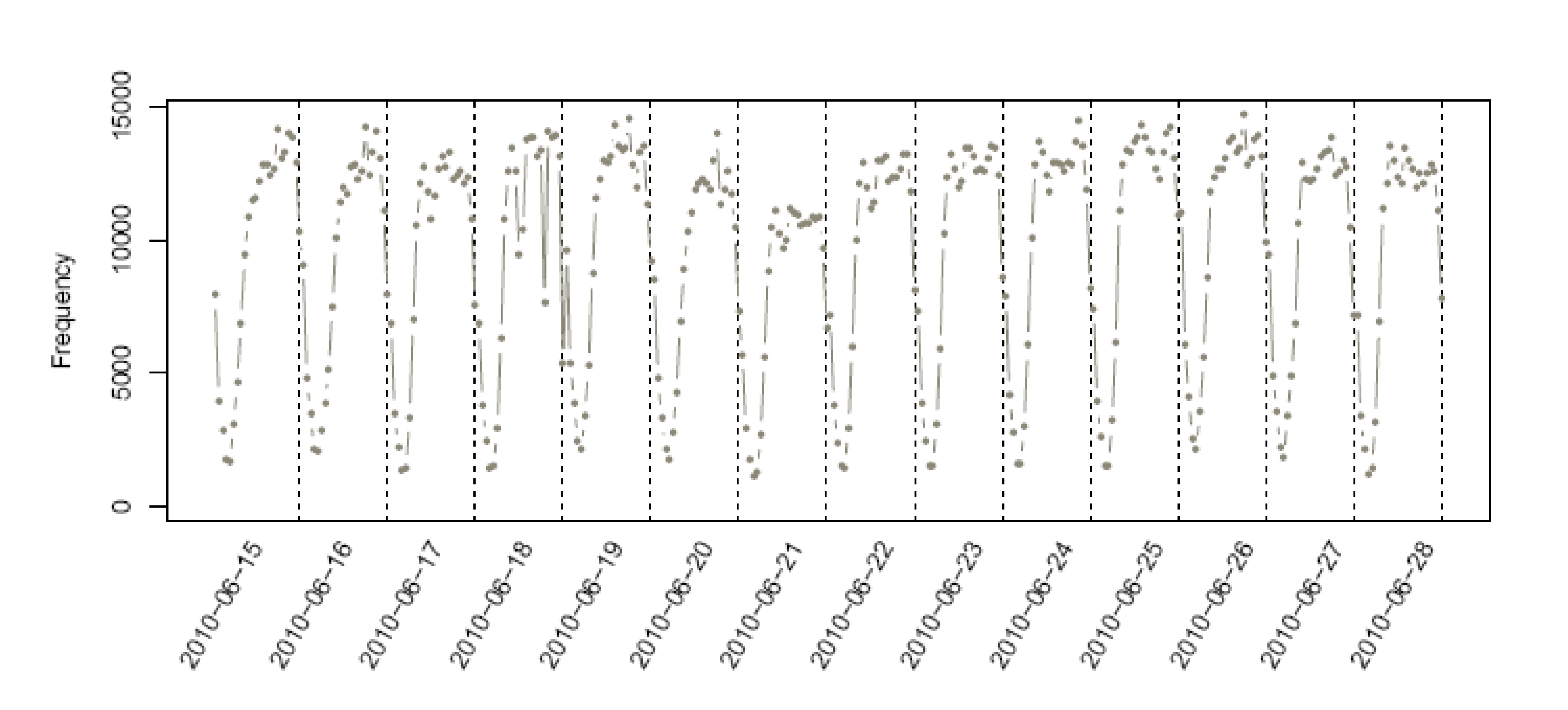
Overall Patterns

- Distribution of service areas
- **Taxi service activities**
- Real routes

Overall Patterns

Taxi service activities:

- Regularity of taxi services

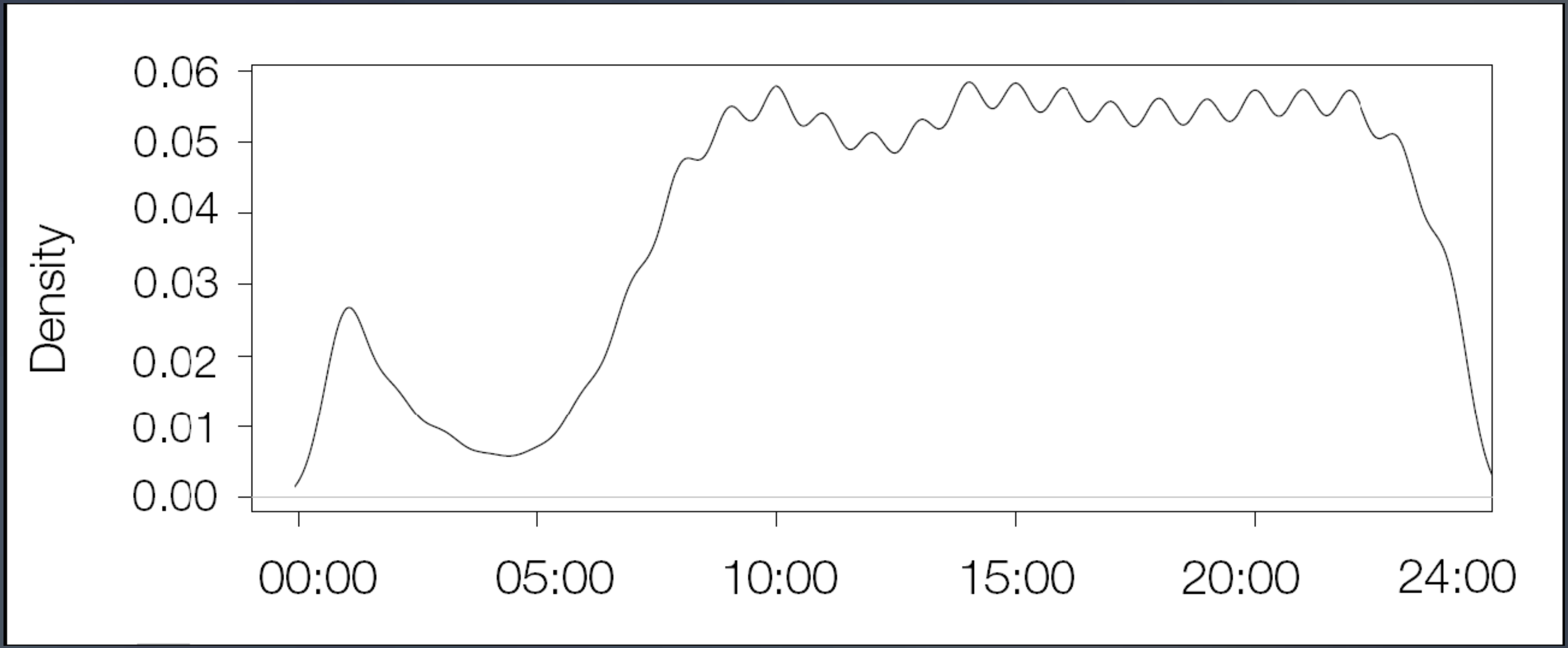


Event plot of pick-up activities across days

Overall Patterns

Taxi service activities:

- Regularity of taxi services



Distribution of pick-up time around the clock

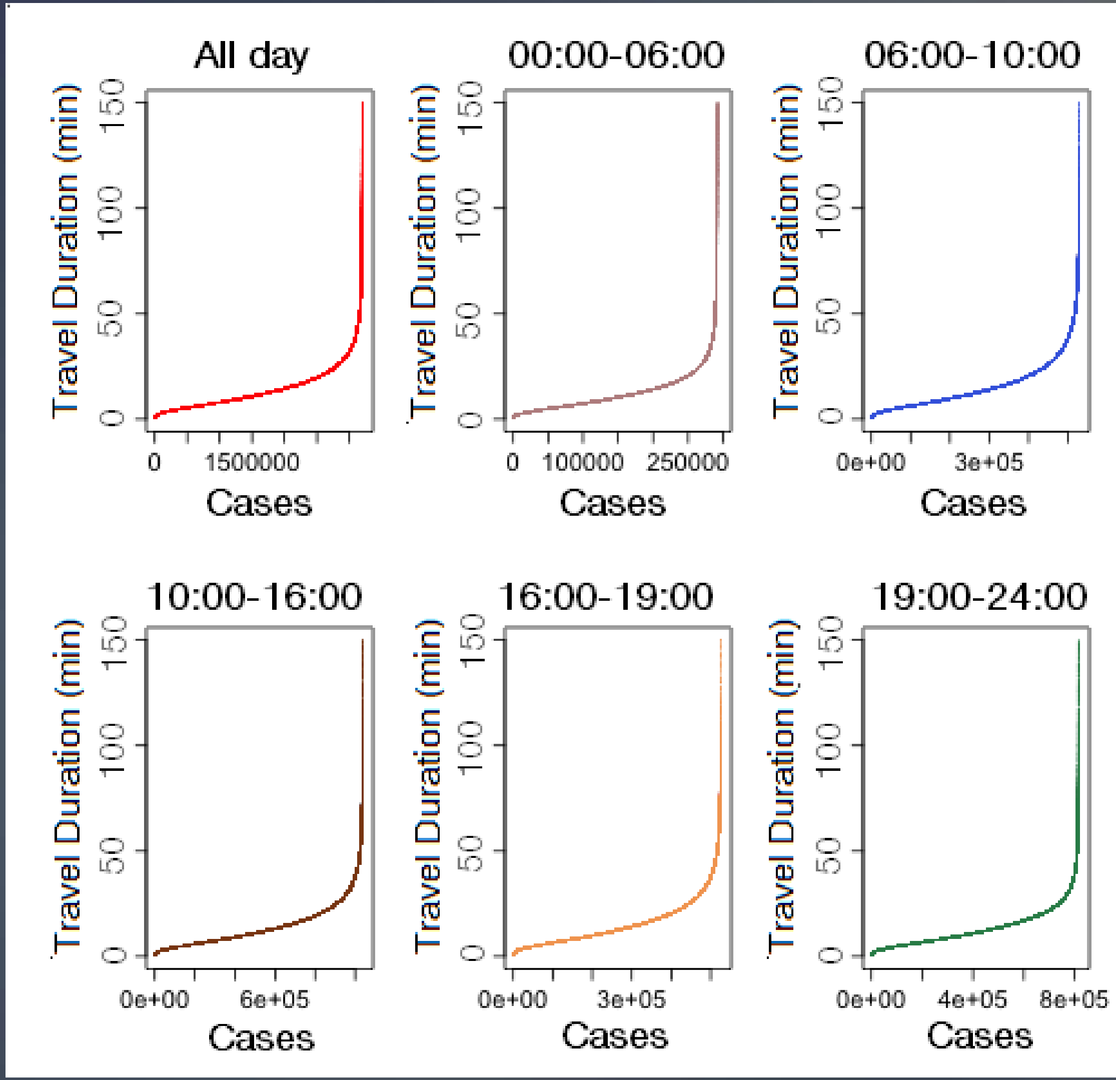
Overall Patterns

Taxi service activities:

- Service duration

  Distribution of travel duration

  Mean: 14.54 min



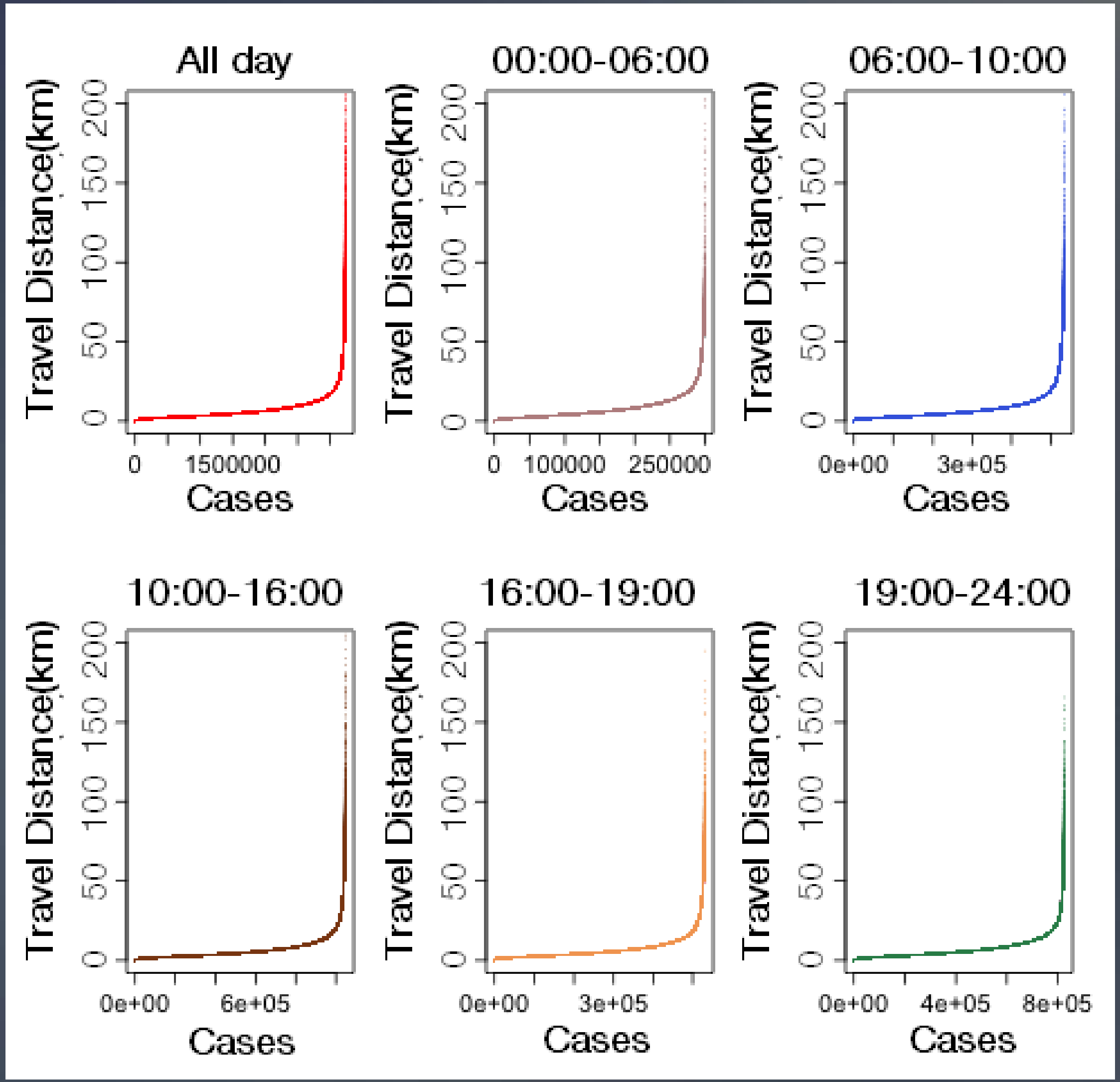
Overall Patterns

Taxi service activities:

- Service distance

Distribution of travel distance of on-service taxi drivers

Mean: 7.27km



Overall Patterns

- Distribution of service areas
- Taxi service activities
- **Real routes**

Overall Patterns

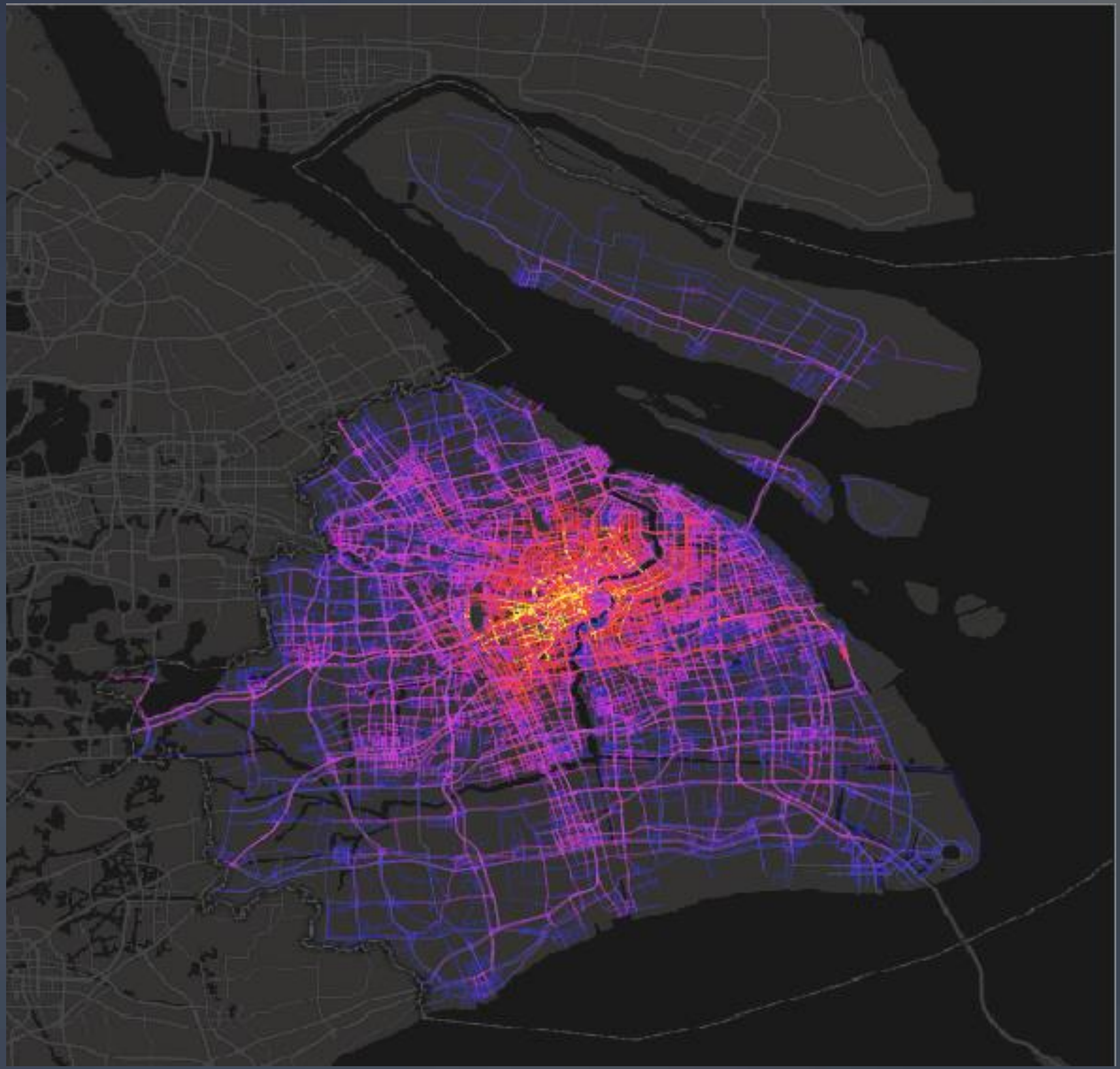
Real routes:

a) Map-matched real routes

Overall Patterns

Real routes:

b) Density of the routes



Part 6

COMPARISON ANALYSIS

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

Percentage of shared length (PSL) :

- To investigate the extent of similarity between the real and the shortest routes.
- The PSL is calculated by first identifying the overlapped segments of real routes and the shortest routes and then dividing the overlapped length by real total length.

$$\text{PSL} = \frac{\sum_{i \in \Phi} r_i}{\sum_{j \in \Theta} r_j}$$

where Φ and Θ denote the set of overlapped segments and the segments contained in real routes, i denotes individual segments, and r_i denotes the segment length.

Comparison Analysis

Percentage of shared length (PSL) :

- Overall result

PSL	Percentage (%)	Cumulative percentage(%)
1	16.66727	16.66727
0.8 -1	5.54906	22.21633
0.5 - 0.8	19.64954	41.86587
0.2 - 0.5	26.71682	68.58269
0 -0.2	31.41731	100
Average:	0.4565	

Table 4. Distribution of PSL

Comparison Analysis

Percentage of shared length (PSL) :

- Variance between different time periods

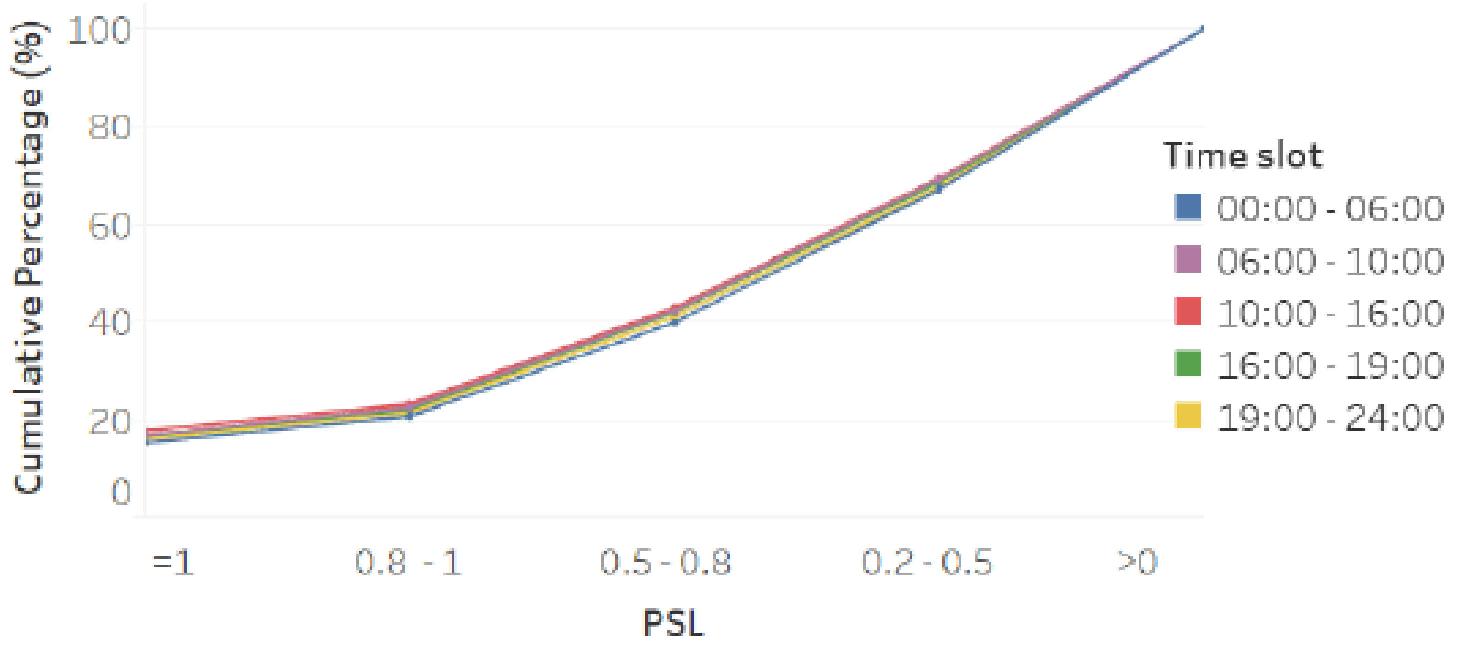


Figure 14. PSL grouped by time slots

Comparison Analysis

Percentage of shared length (PSL) :

- Variance between different time periods

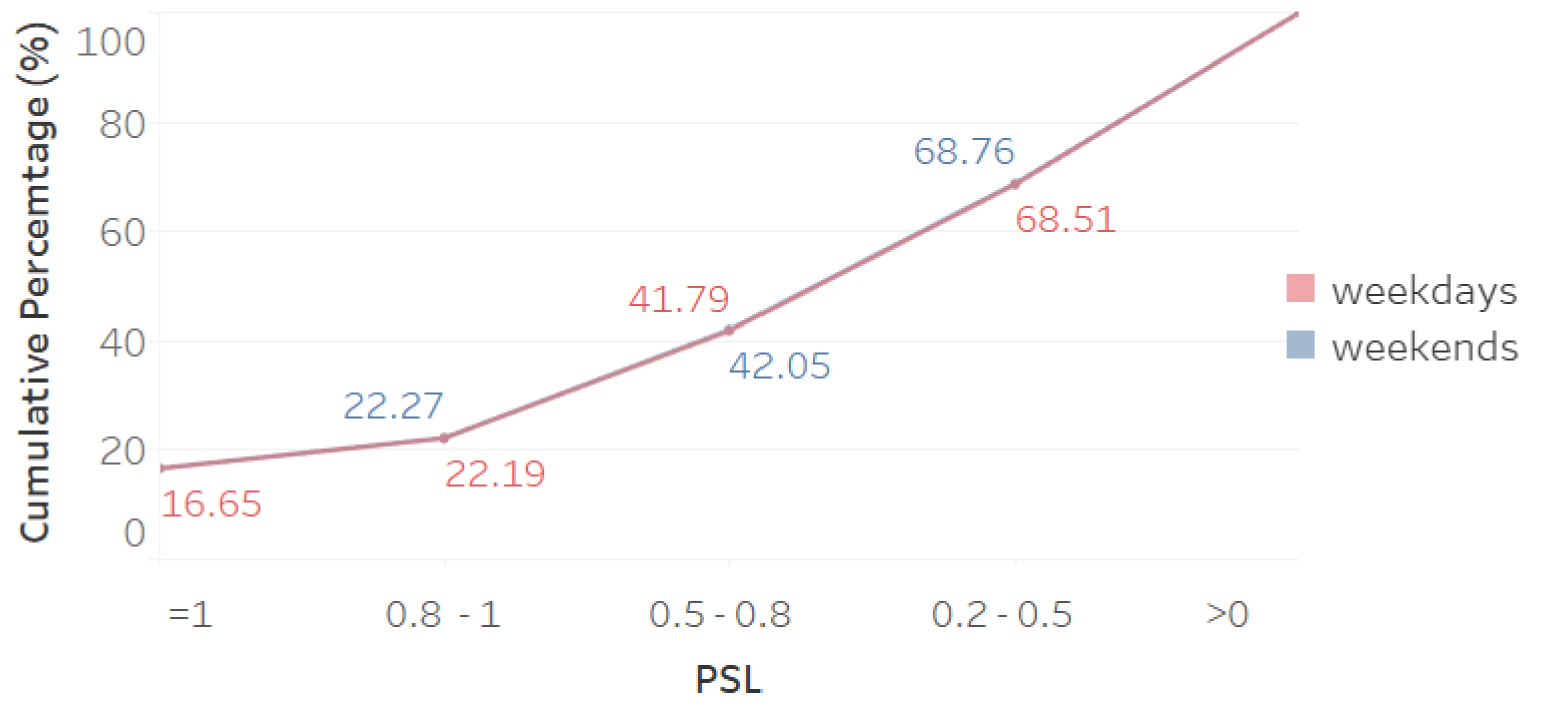


Figure 15. PSL grouped by weekdays and weekends

Comparison Analysis

Percentage of shared length (PSL) :

- Variance between different travel distances

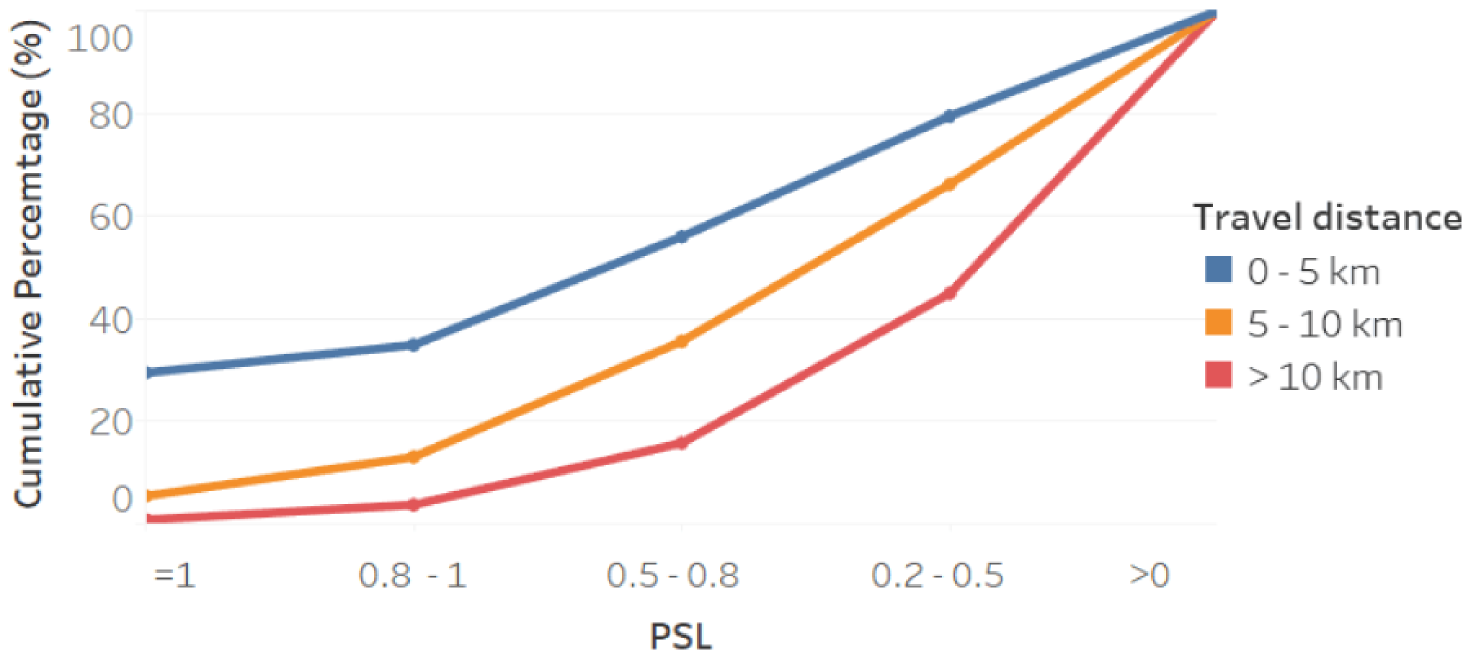


Figure 16. PSL grouped by lengths of real routes

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

Percentage of length difference (PLD)

•PLD compares the total length between the real and the shortest routes to show how much longer is the actual route to the shortest routes.

$$\mathrm{PLD} = \frac{\sum_{i\in\Gamma} r_i - \sum_{j\in\Delta} r_j}{\sum_{i\in\Gamma} r_i}$$

where Γ and Δ denote respectively the set of real route segments and the shortest route segments, i denotes an individual segment, and r_i and r_j denotes the length of the segment. Or in short: $\mathrm{PLD} = \frac{L1-L2}{L1}$, which means the value is calculated by subtracting the length of the shortest route ($L2$) from the real route ($L1$) and then divided by the length of the real route.

Comparison Analysis

Percentage of length difference (PLD)

- Overall result

PLD	Percentage (%)	Cumulative percentage(%)
1	0	0
0.8 -1	0.6421462	0.6421462
0.5 - 0.8	4.2816988	4.923845
0.2 - 0.5	26.540425	31.46427
0 -0.2	68.53573	100
Average:	0.1609	

Table 5. Distribution of PLD

Comparison Analysis

Percentage of length difference (PLD)

- Variance between different time periods

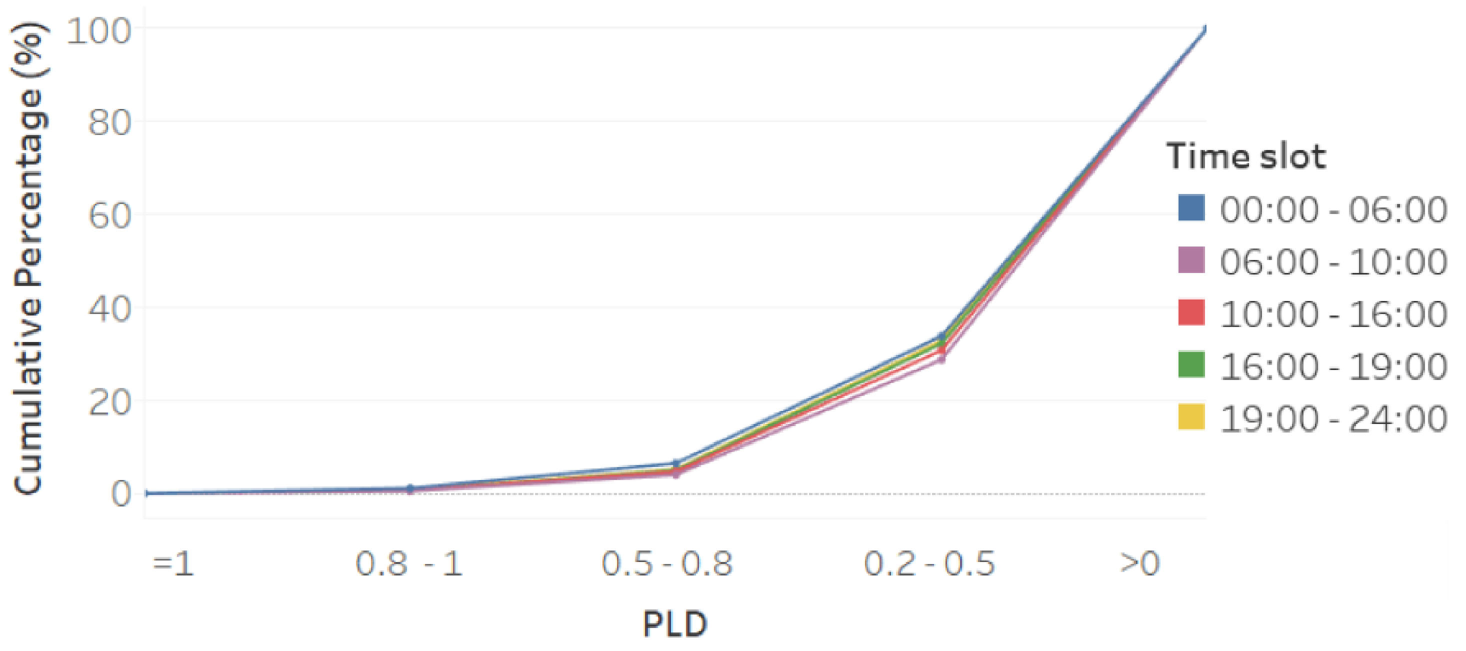


Figure 17. PLD grouped by time slots

Comparison Analysis

Percentage of length difference (PLD)

- Variance between different time periods

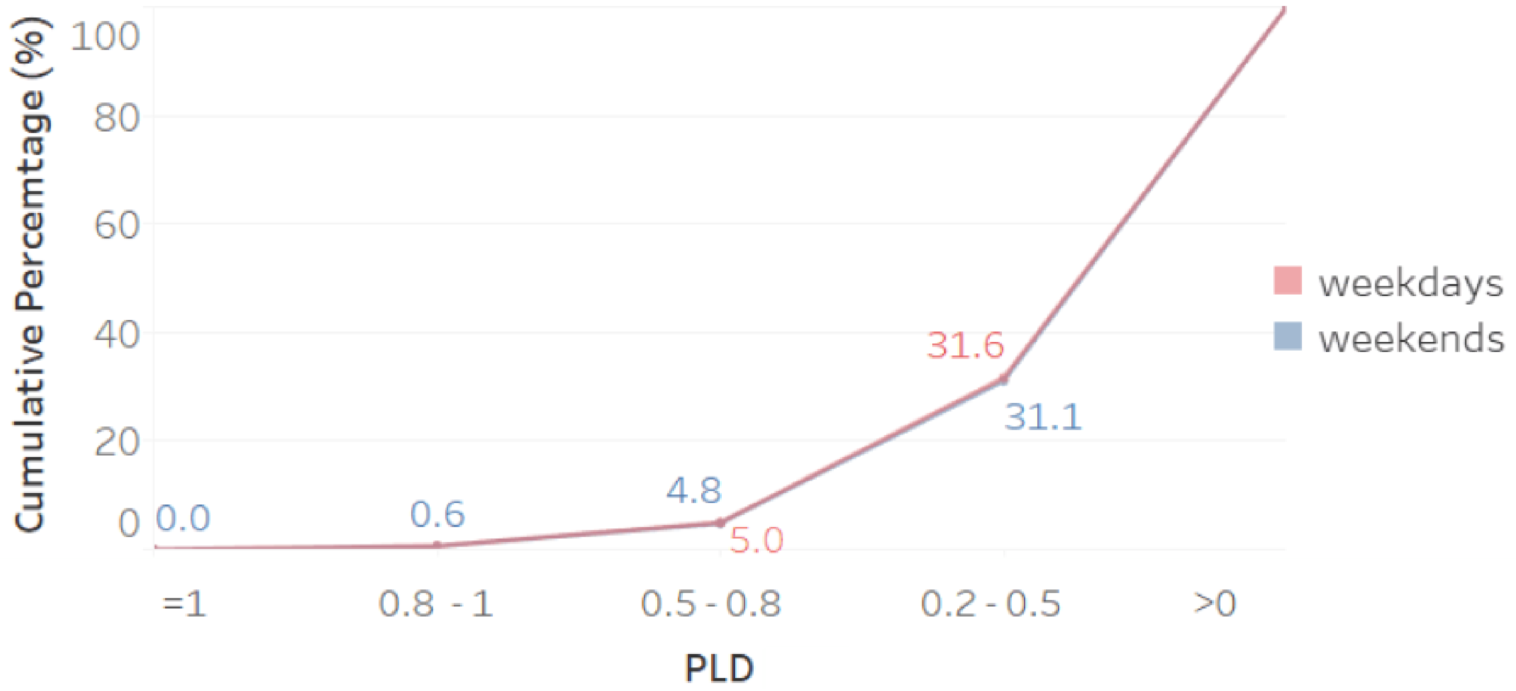


Figure 18. PLD grouped by weekdays and weekends

Comparison Analysis

Percentage of length difference (PLD)

•Variance between different travel lengths

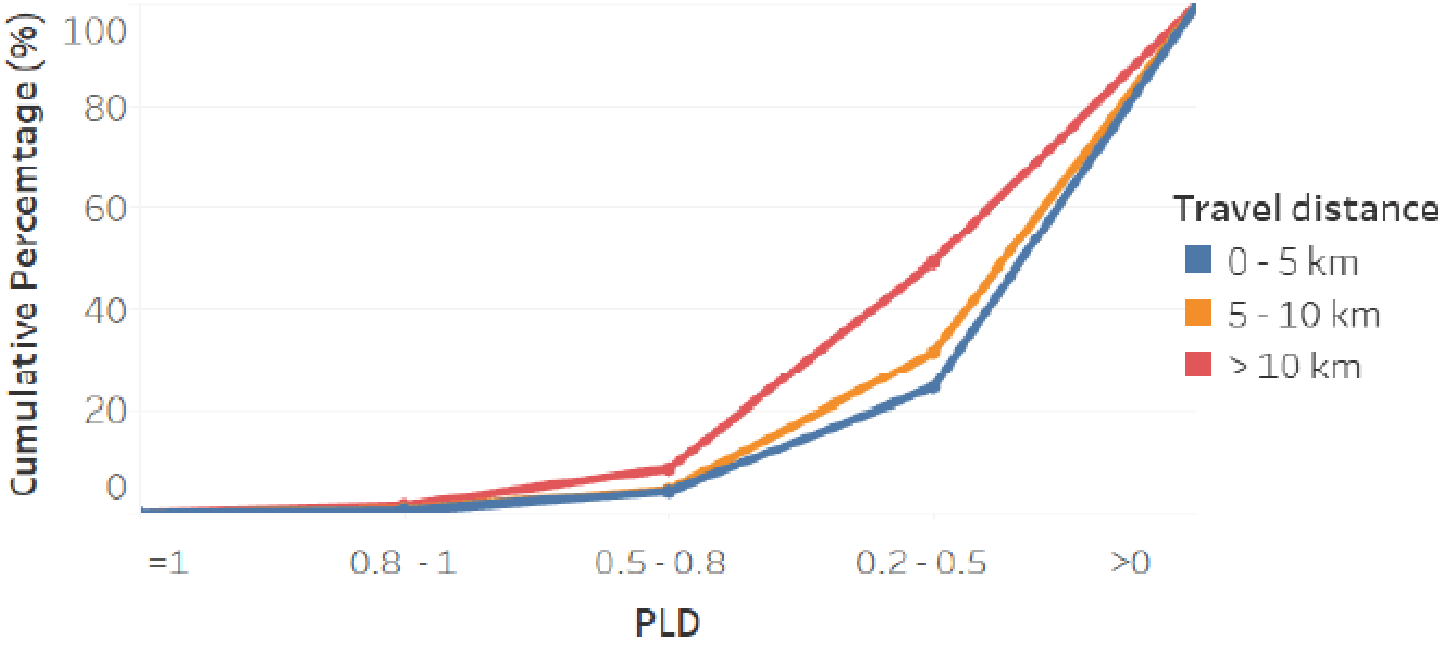


Figure 19. PLD grouped by lengths of real routes

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

Route Circuity

For each OD pair a, b, the route circuity is calculated as follows,

$$rcr = \frac{l_{ab} - d_{ab}}{l_{ab}}$$

, where l_{ab} refers to the route distance, and d_{ab} refers to the geometric distance between point a and point b, and

$$d_{ab} = \arccos(\sin\theta_a * \sin\theta_b + \cos\theta_a * \cos\theta_b * \cos\Delta\lambda) * R$$

where θ refers to latitude, λ refers to longitude, R refers to the radius of the earth.

Comparison Analysis

Route Circuity

- Overall result

Route Circuity	Percentage in real routes (%)	Cumulative percentage(%)	Percentage in shortest routes (%)	Cumulative percentage(%)
1	0	0	0	0
0.8 -1	1.1759	1.1759	0.09560	0.0956
0.5 - 0.8	12.1467	13.3225	1.3021	1.3977
0.2 - 0.5	62.7011	76.0236	37.6061	39.0037
0 -0.2	23.9764	100	60.9963	100
Average	Real routes:	0.3154		
	Shortest routes:	0.1862		

Table 6. Distribution of Route Circuity of real routes and shortest routes

Comparison Analysis

Route Circuity

• Variance between different time periods

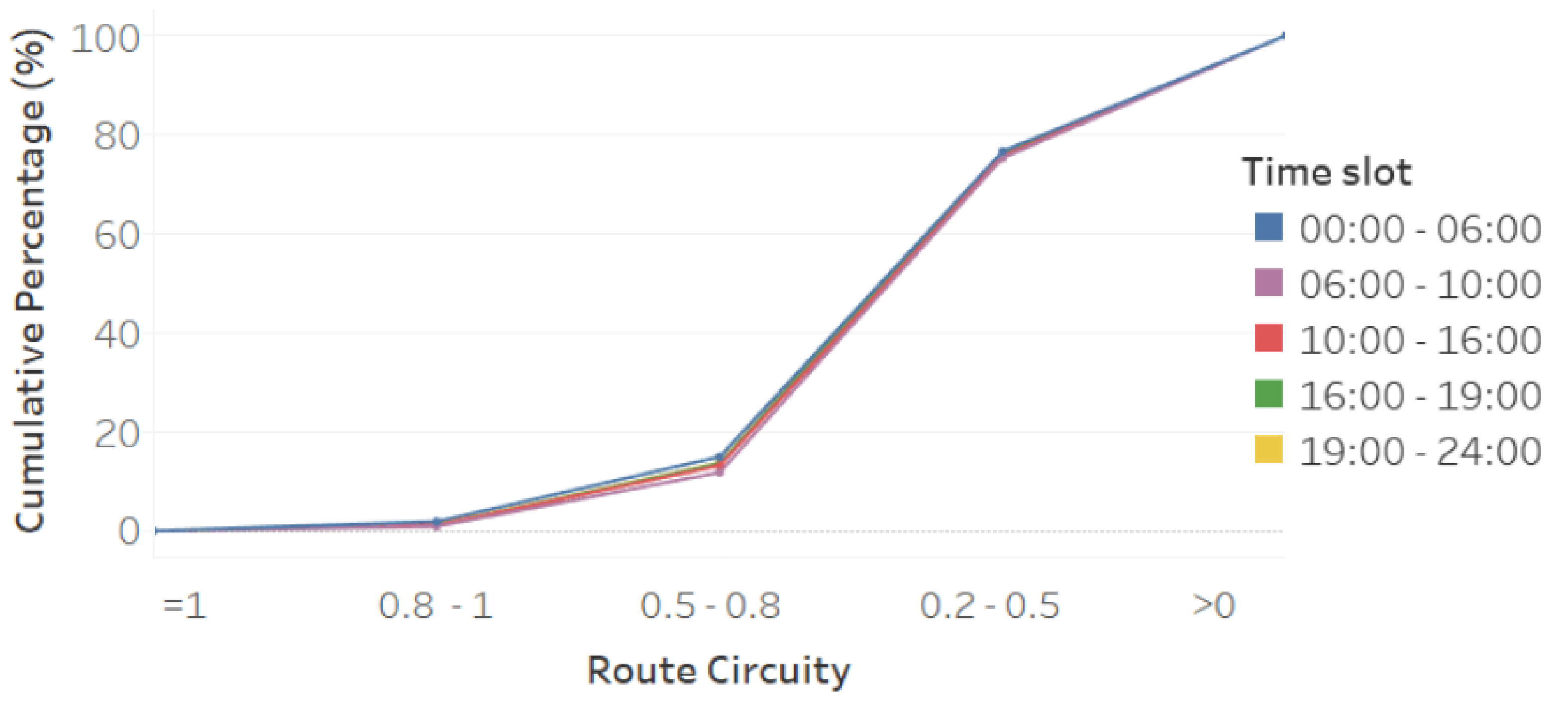


Figure 20. Route circuity grouped by time slots

Comparison Analysis

Route Circuity

- Variance between different time periods

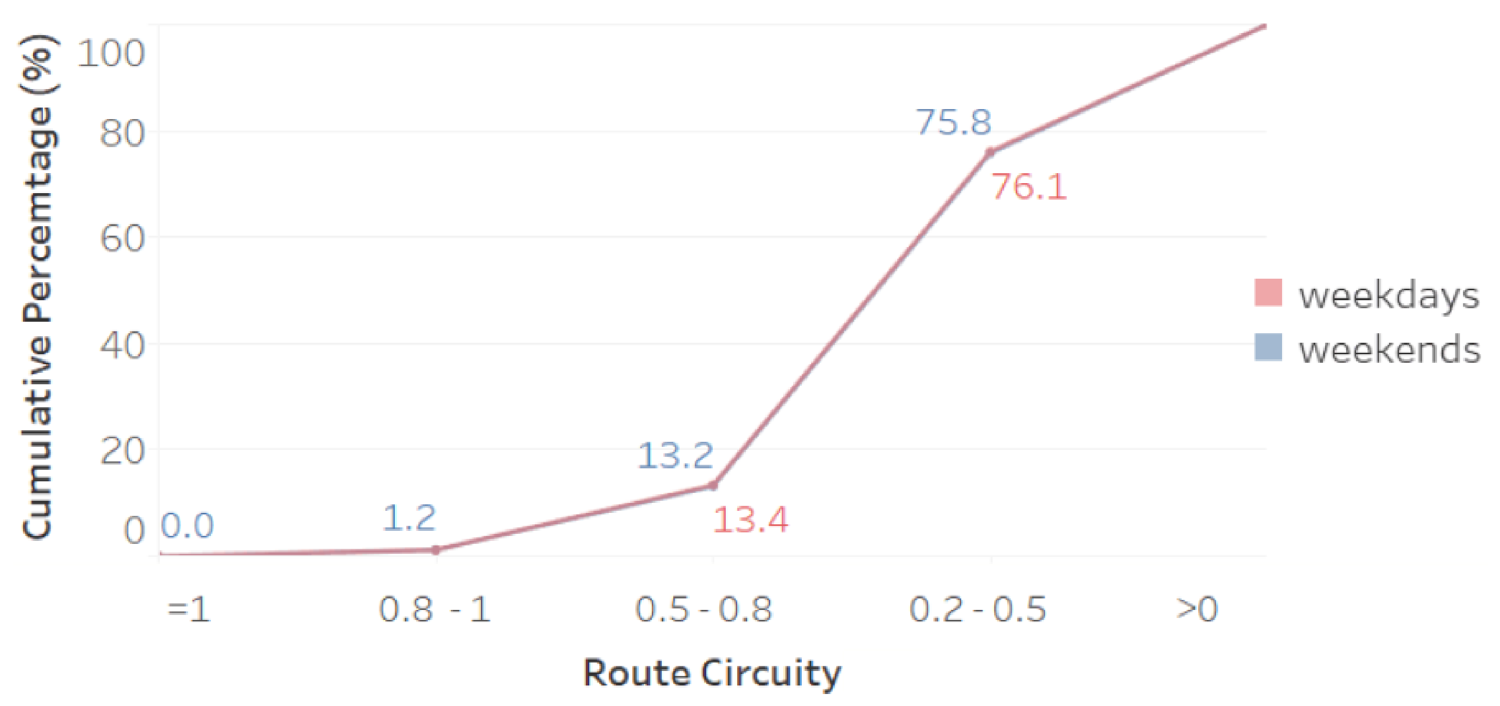


Figure 21. Route circuity grouped by weekdays and weekends

Comparison Analysis

Route Circuity

- Variance between different travel distance

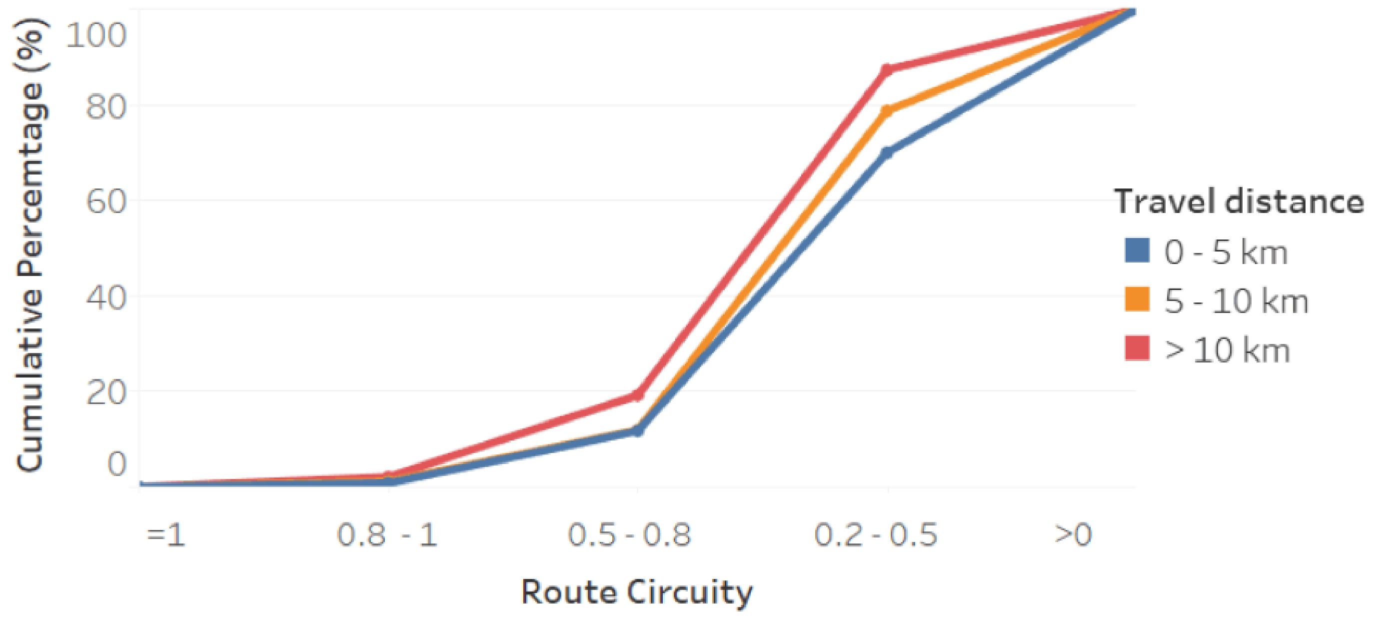


Figure 22. Route circuity grouped by length of real routes

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

Road type

As traditionally categorized, five types of road based on Open Street Map data are analyzed: motorway, primary, secondary, tertiary and residential roads.

- Motorway:120 km/h
- Primary road: 70km/h
- Secondary road: 60 km/h
- Tertiary road: 40km/h
- Residential road: 40km/h

Comparison Analysis

Road type

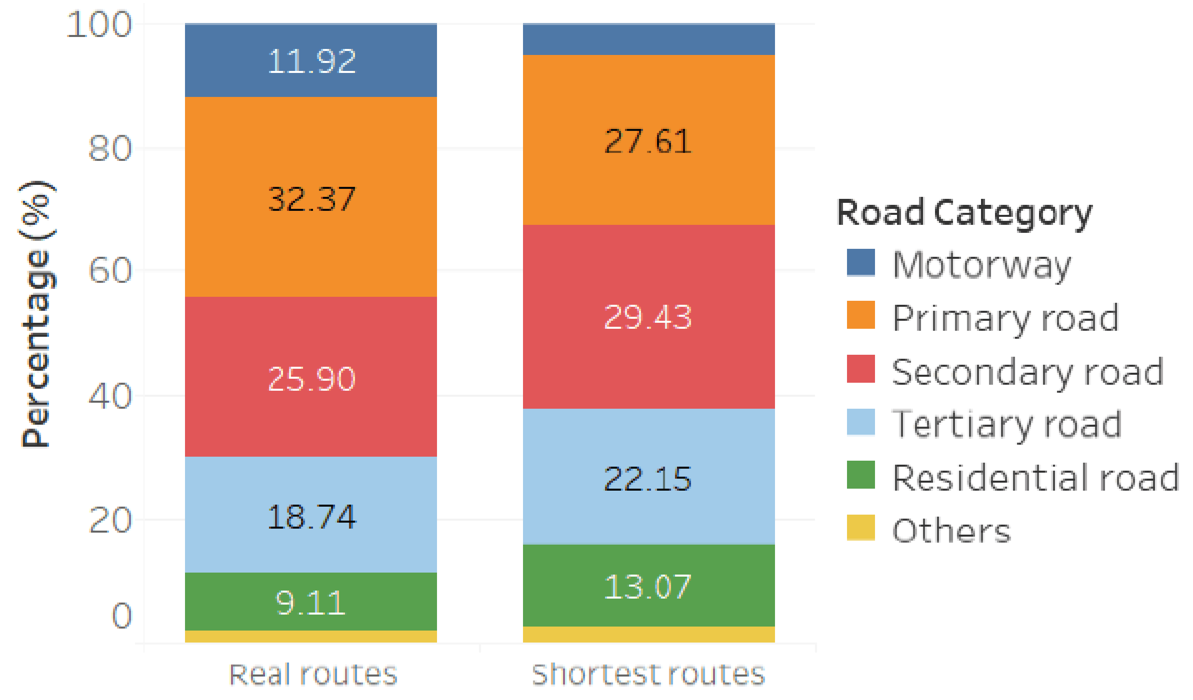


Figure 23. Share of different types of roads in the real and shortest routes

Comparison Analysis

Road type

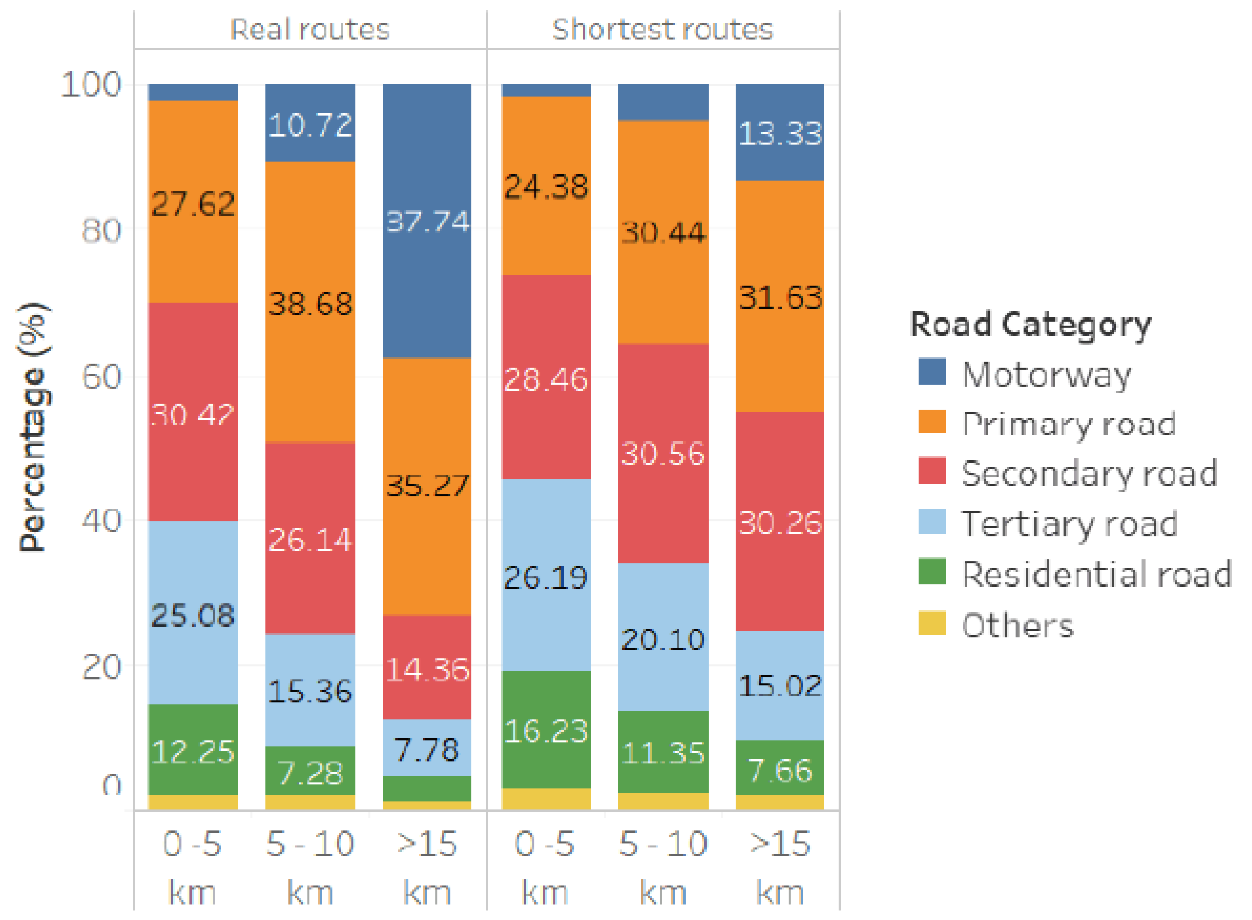


Figure 24. Share of different types of roads across travel distances

Comparison Analysis

Road type

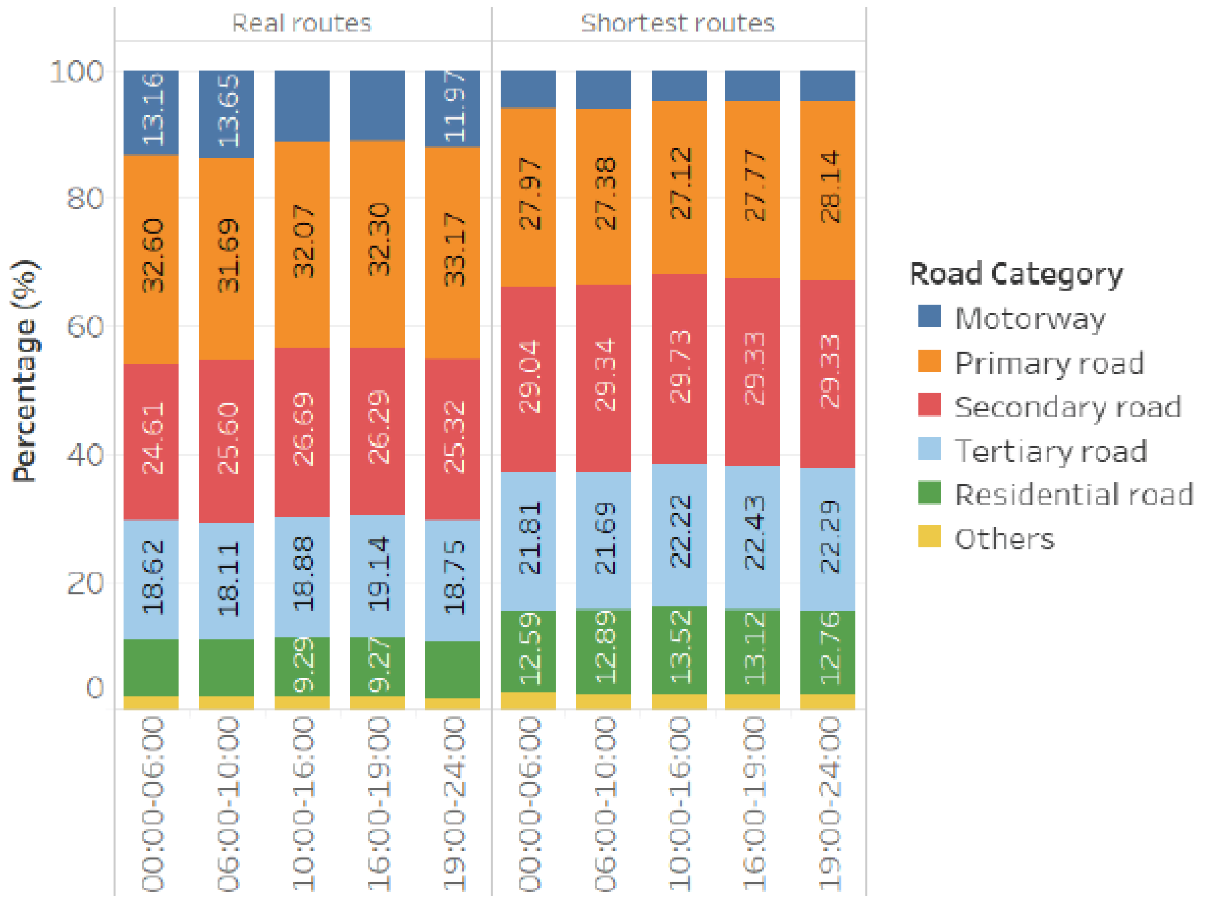


Figure 25. Share of different types of roads across time slots

Comparison Analysis

Road type

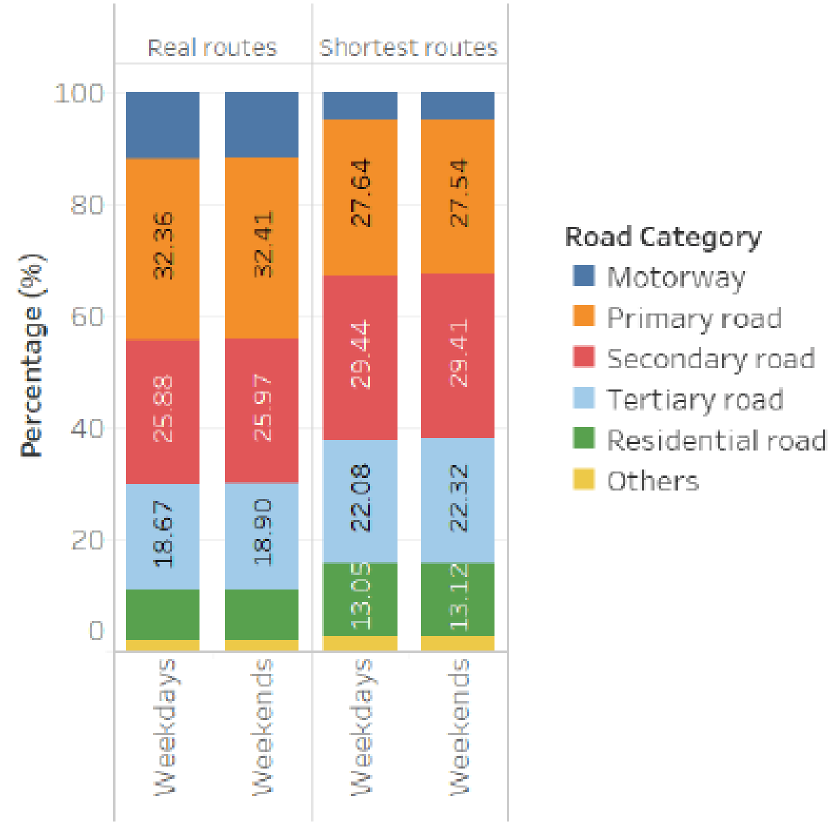


Figure 26. Share of different types of roads in weekdays and weekends

Comparison Analysis

- **Route-based Analysis**
  - Percentage of shared length (PSL)
  - Percentage of length difference (PLD)
  - Route Circuity
- **Route segment-based analysis**
  - Road type
  - Intersections and number of turns

Comparison Analysis

Intersections and number of turns

- Right-hand traffic in China
- left turns (210˚ – 330˚)
- right turns (30˚ - 150˚)
- The straight ways are the ways whose nodes are nether the left- nor the right-turns, or the ways that only connect with two nodes.

Comparison Analysis

Intersections and number of turns

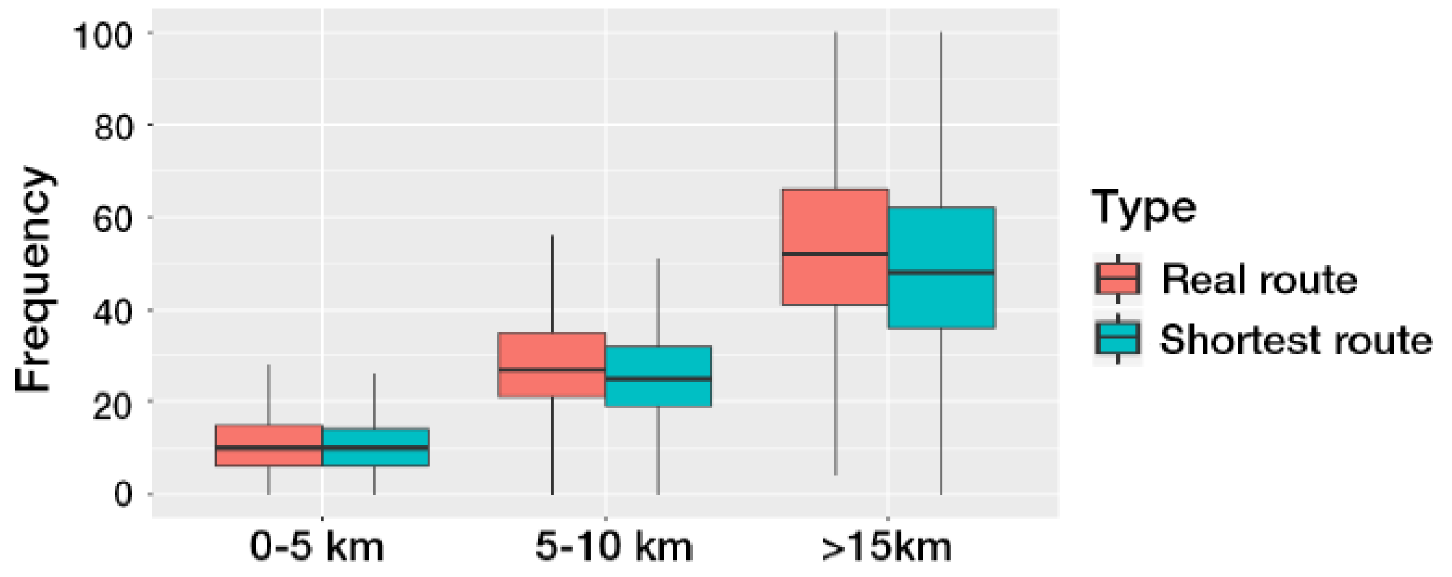


Figure 27. Comparison of numbers of straight ways in the real and the shortest routes

Comparison Analysis

Intersections and number of turns

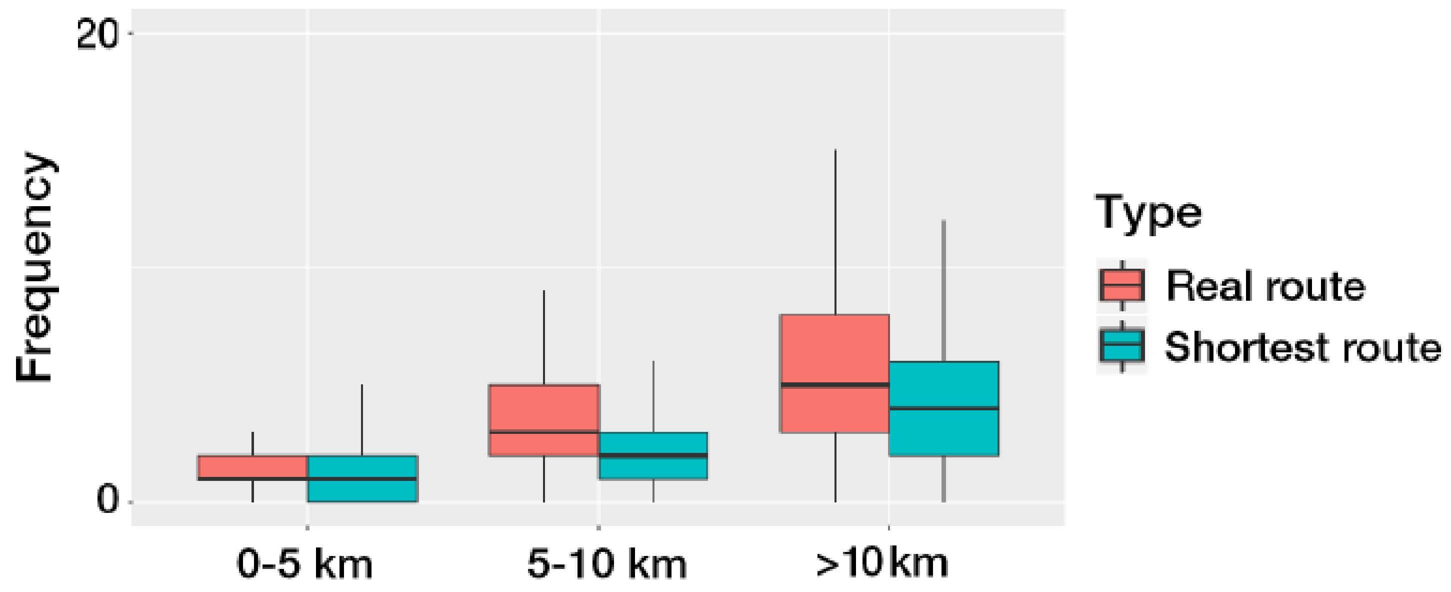


Figure 28. Comparison of numbers of left turns in the real and the shortest routes

Comparison Analysis

Intersections and number of turns

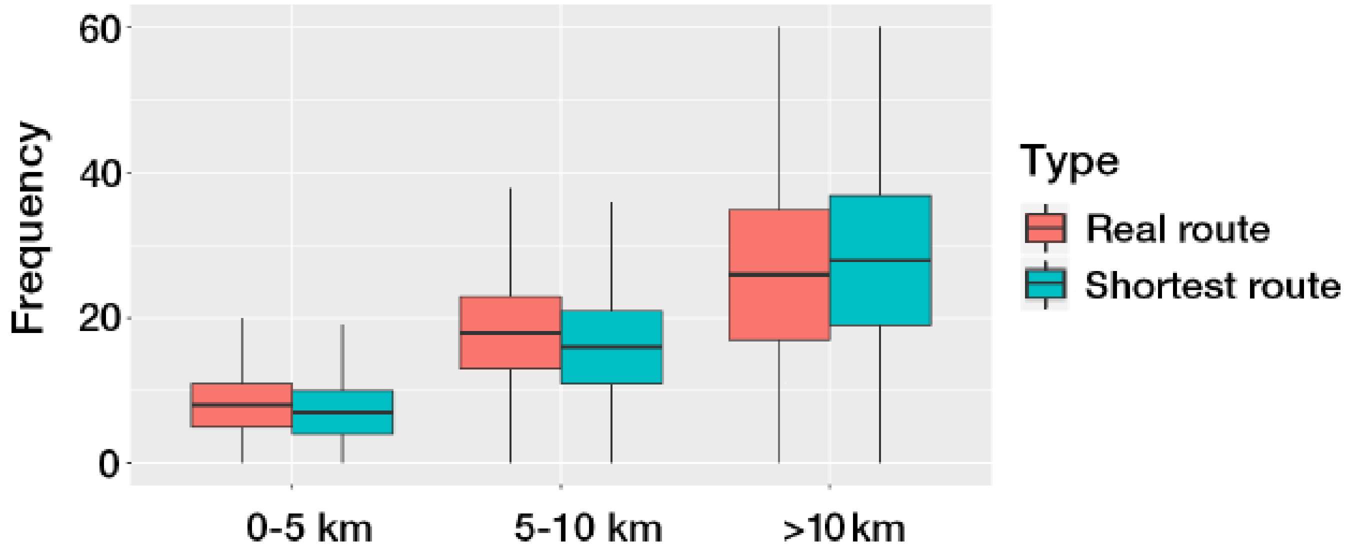


Figure 30. Comparison of numbers of intersections in the real and the shortest routes

Part 7

PREFERENCE ANALYSIS

Preference Analysis

- **Scenarios and their application to general routes**
- Scenarios applied to airport-city center routes

Preference Analysis

Scenarios and their application to general routes

Scenario 0 : Minimal distance

Preference Analysis

Scenarios and their application to general routes

Scenario 0 : Minimal distance
Scenario 1 : Least free flow travel time

Preference Analysis

Scenarios and their application to general routes

Scenario 0 :	Minimal distance
Scenario 1 :	Least free flow travel time
Scenario 2 :	Least observed travel time

Preference Analysis

Scenarios and their application to general routes

Scenario 0 :	Minimal distance
Scenario 1 :	Least free flow travel time
Scenario 2 :	Least observed travel time
Scenario 3 :	Minimal (road category * distance)
Scenario 4 :	Minimal (road category * time)

Preference Analysis

Scenarios and their application to general routes

Scenario 0 :	Minimal distance
Scenario 1 :	Least free flow travel time
Scenario 2 :	Least observed travel time
Scenario 3 :	Minimal (road category * distance)
Scenario 4 :	Minimal (road category * time)
Scenario 5 :	Minimal (road usage frequency * distance)
Scenario 6 :	Minimal (road usage frequency * time)

Preference Analysis

Scenarios and their application to general routes

Scenario 0 :	Minimal distance
Scenario 1 :	Least free flow travel time
Scenario 2 :	Least observed travel time
Scenario 3 :	Minimal (road category * distance)
Scenario 4 :	Minimal (road category * time)
Scenario 5 :	Minimal (road usage frequency * distance)
Scenario 6 :	Minimal (road usage frequency * time)
Scenario 7 :	Minimal (left turns * distance)
Scenario 8 :	Minimal (left turns * time)

Preference Analysis

Scenarios and their application to general routes

Scenario 0 :	Minimal distance
Scenario 1 :	Least free flow travel time
Scenario 2 :	Least observed travel time
Scenario 3 :	Minimal (road category * distance)
Scenario 4 :	Minimal (road category * time)
Scenario 5 :	Minimal (road usage frequency * distance)
Scenario 6 :	Minimal (road usage frequency * time)
Scenario 7 :	Minimal (left turns * distance)
Scenario 8 :	Minimal (left turns * time)
Scenario 9 :	Minimal (turns * distance)
Scenario 10 :	Minimal (turns * time)

Preference Analysis

Scenarios and their application to general routes

	Scenarios										
Factors	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Length	■			■		■		■		■	
Free flow		■			■		■		■		■
Observed travel time			■								
Road type				■	■						
Road use frequency						■	■				
Left turns								■	■		
Turns										■	■

Table 7. Scenarios: the gray cells indicate the inclusion of the factor into the corresponding scenario.

Preference Analysis

Scenarios and their application to general routes

Scenario	Weighted Criteria	mean(%)	Over 75% match
0	Minimal distance	46.61	25.69
1	Least freeflow travel time	51.30	30.77
2	Least observed travel time	51.75	30.92
3	Minimal (road category * distance)	48.79	29.39
4	Minimal (road category * time)	45.80	26.72
5	Minimal (road usage frequency * distance)	53.56	32.52
6	Minimal (road usage frequency * time)	52.48	31.94
7	Minimal (left turns * distance)	46.61	25.60
8	Minimal (left turns * time)	51.06	30.50
9	Minimal (turns * distance)	36.23	14.52
10	Minimal (turns * time)	38.46	16.53

Table 8. Prediction accuracy of different scenarios (cells show the cumulative percentage of cases within specified accuracy category)

Preference Analysis

- Scenarios and their application to general routes
- **Scenarios applied to airport-city center routes**

Preference Analysis

Scenarios applied to airport-city center routes

- DBSCAN
  - *Density-Based Spatial Clustering of Applications with Noise*
  - to identify the clusters of points (Ester et al., 1996)
  - two parameters:
    - the minimum number of points (minpts)
    - the range of clustering (ε)

Preference Analysis

Scenarios applied to airport-city center routes

- DBSCAN

minpts	N	ε 100	200	300	400	500	600	700	800	900	1000
	100%	4227	1822	558	147	53	36	27	18	12	12
	200%	3204	1291	400	99	44	36	27	19	15	12
	300%	2585	929	296	77	36	20	18	15	10	9
	400%	2107	735	236	59	27	18	12	10	10	8
	500%	1746	594	206	70	24	11	10	8	7	6
	600%	1455	487	180	68	24	15	10	7	6	5
	700%	1227	409	154	66	28	15	7	9	6	5
	800%	1050	345	145	66	22	18	11	7	5	4
	900%	889	302	125	61	28	13	7	5	3	4
	1000%	774	267	103	59	26	15	10	6	3	3
	1100%	693	228	93	51	29	16	10	8	4	3
	1200%	606	203	86	51	24	15	9	8	5	3
	1300%	518	186	87	43	22	11	11	7	5	4
	1400%	453	158	77	41	29	13	11	9	7	6
	1500%	391	145	81	34	29	12	11	10	5	7
	1600%	352	130	79	35	25	11	11	9	7	5
	1700%	322	124	76	38	19	15	9	5	4	5
	1800%	282	104	68	29	19	15	8	8	4	2
	1900%	259	88	61	26	18	15	7	7	5	1
	2000%	229	83	58	30	16	14	8	5	4	2

Table 9. Number of cluster with different combinations of parameters (minpts in the first column and ϵ in the first row)

ε = 500
minpts = 3000

Preference Analysis

Scenarios applied to airport-city center routes

- Results

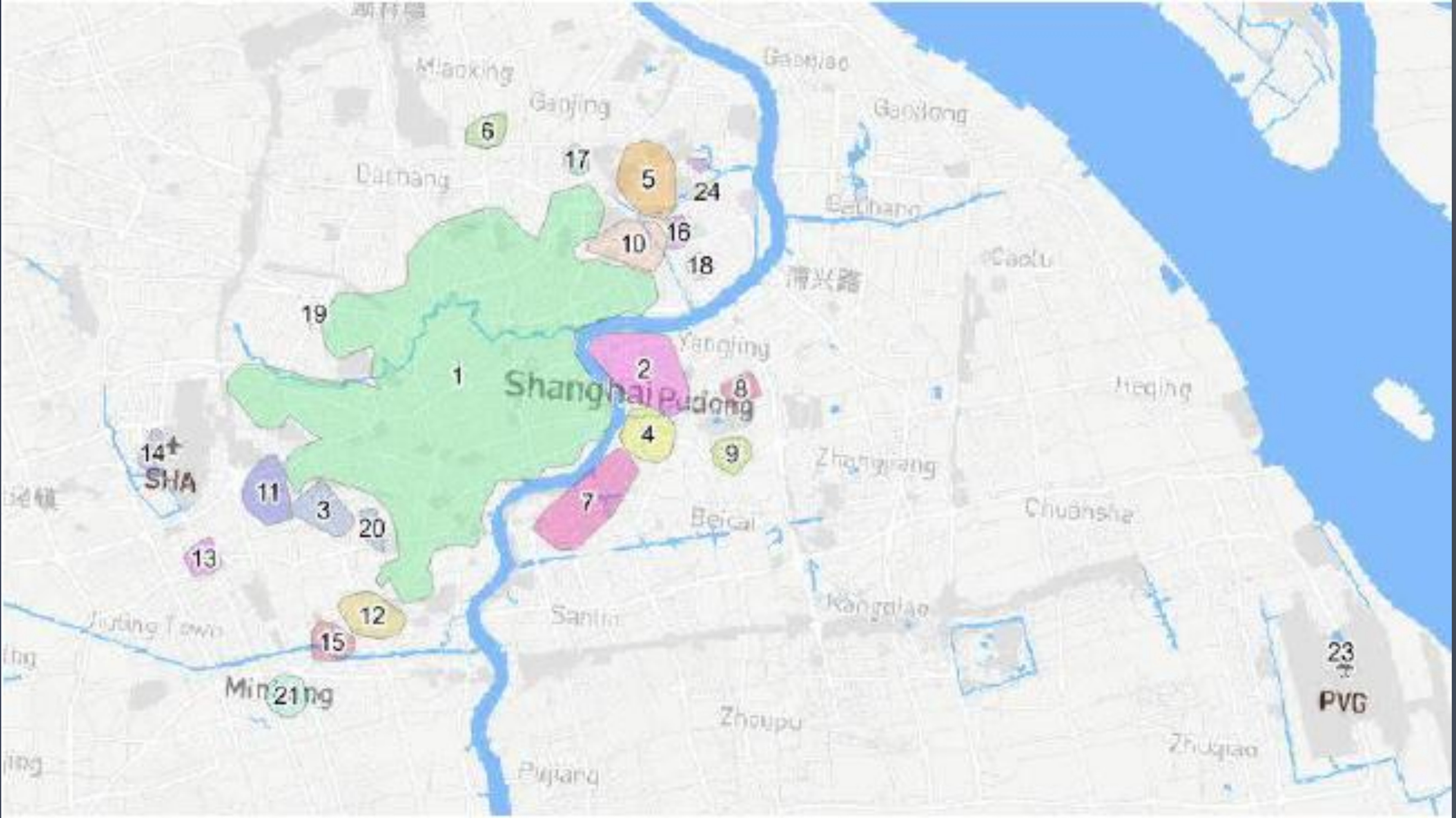


Figure 31. Clusters of most frequently visited places

Preference Analysis

Scenarios applied to airport-city center routes

- Results

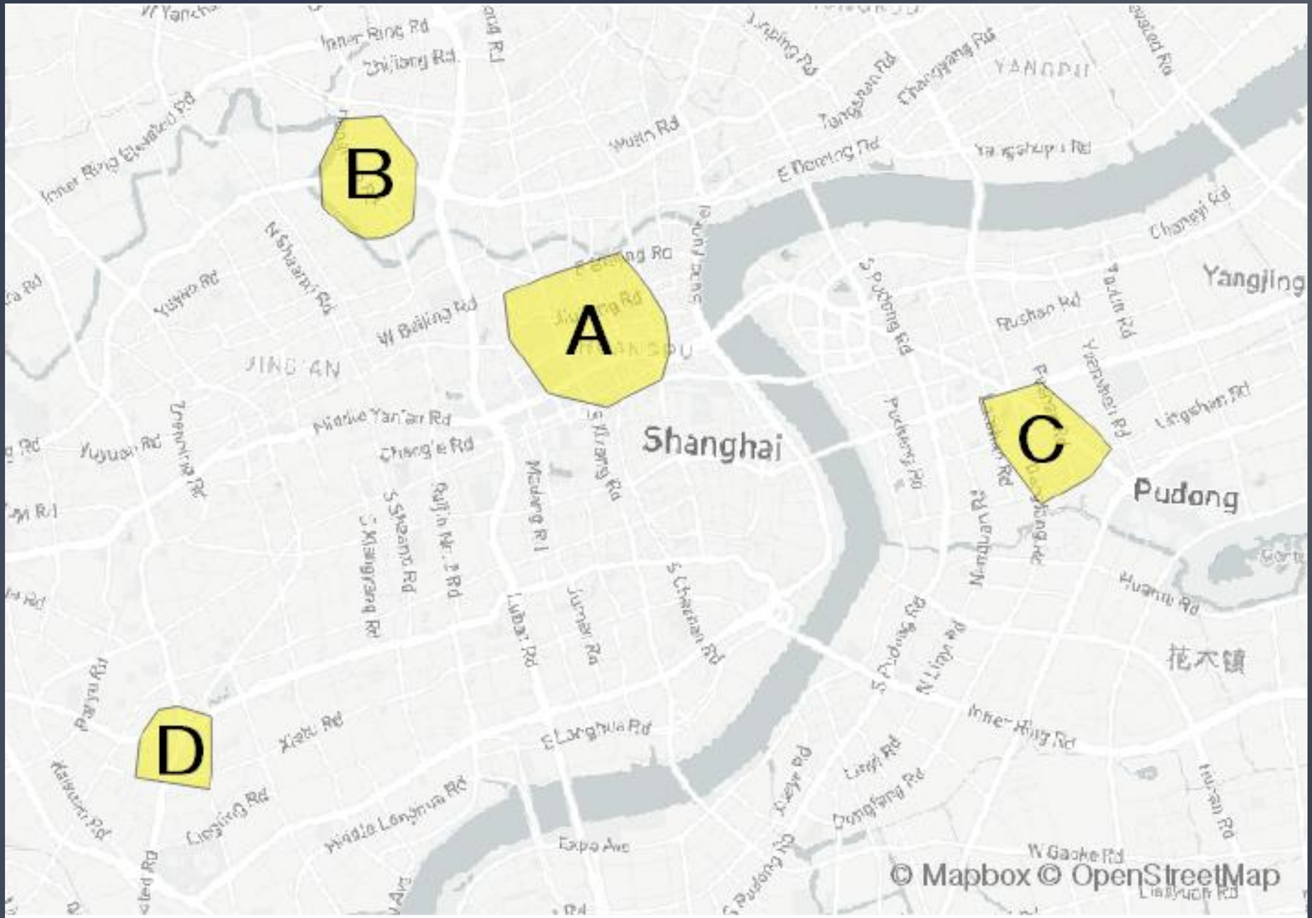


Figure 32. Clusters of most frequently visited places by routes from the Pudong International Airport

Preference Analysis

Scenarios applied to airport-city center routes

- Results

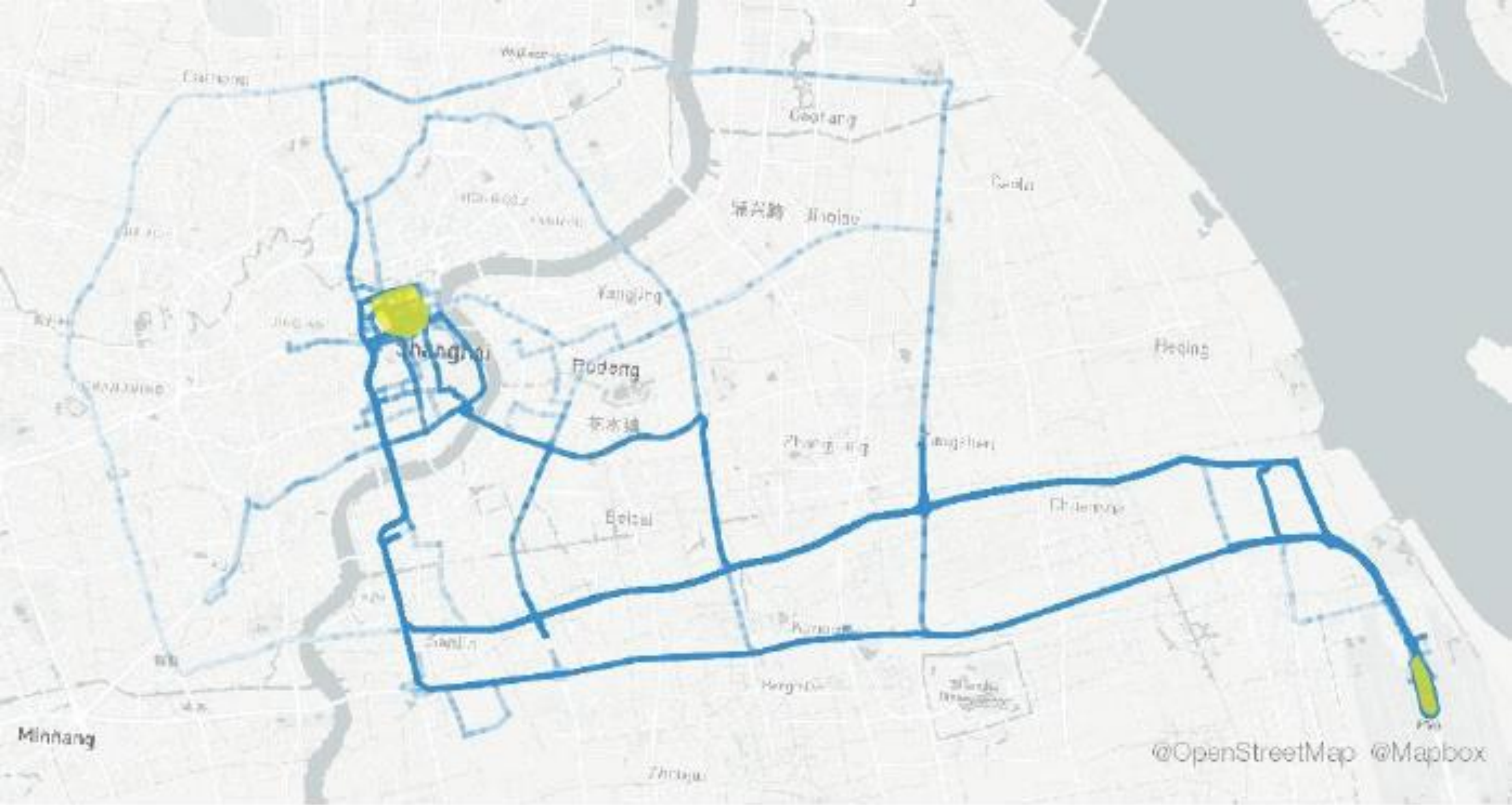


Figure 33. Routes from the Pudong International Airport to city center

Preference Analysis

Scenarios applied to airport-city center routes

- Results

Scenario	Weighted Criteria	Mean (%)	Over 75% match
0	Minimal distance	13.25	0.00
1	Least freeflow travel time	24.22	4.67
2	Least observed travel time	48.00	12.62
3	Minimal (road category * distance)	51.79	17.76
4	Minimal (road category * time)	30.78	9.35
5	Minimal (road usage frequency * distance)	32.34	0.00
6	Minimal (road usage frequency * time)	25.84	5.14
7	Minimal (turns * distance)	10.68	0.00
8	Minimal (turns * time)	20.52	4.67
9	Minimal (left turns * distance)	13.26	0.00
10	Minimal (left turns * time)	24.20	4.66

Table 10. Prediction accuracy of different scenarios to the routes from Pudong International Airport to the city center

Part 8
CONCLUSION, LIMITATION & OUTLOOK

Conclusion, Limitation & Outlook

Conclusion:

Real routes to some extent **overlap** with the shortest routes with an average value of ***46%***.

Conclusion, Limitation & Outlook

Conclusion:

Real routes to some extent **overlap** with the shortest routes with an average value of ***46%***.

Other factors, especially the **road type**, **road usage frequency**, **speed limit** and **observed travel speed**, have impacts on the route choices.

Conclusion, Limitation & Outlook

Conclusion:

In the **general** situation, the best scenario in combination with **travel distance** and **road usage frequency** predicts ***53.56%*** of real routes compare to ***46.61%*** in the shortest routes.

Conclusion, Limitation & Outlook

Conclusion:

In the **general** situation, the best scenario in combination with **travel distance** and **road usage frequency** predicts ***53.56%*** of real routes compare to ***46.61%*** in the shortest routes.

In **long distant** trips, the scenario with **road type** and **travel distance** predicts ***51.79%*** of real routes compared to ***13.25%*** in the shortest routes.

Conclusion, Limitation & Outlook

Limitation & Outlook:

•due to computational constraints, only the top 10 routes to predict the best matches in Scenario 7, 8, 9 and 10 could be calculated for each real routes.

Conclusion, Limitation & Outlook

Limitation & Outlook:

•due to computational constraints, only the top 10 routes to predict the best matches in Scenario 7, 8, 9 and 10 could be calculated for each real routes.

- to optimize the efficiency of the k-shortest-path algorithm so that it is computable with large-scale road networks.

Conclusion, Limitation & Outlook

Limitation & Outlook:

•due to computational constraints, only the top 10 routes to predict the best matches in Scenario 7, 8, 9 and 10 could be calculated for each real routes.

- to optimize the efficiency of the k-shortest-path algorithm so that it is computable with large-scale road networks.

•period of study

- extend the period of study and examine the impact of other external events on the taxi drivers' behavior.

Conclusion, Limitation & Outlook

Limitation & Outlook:

- due to computational constraints, only the top 10 routes to predict the best matches in Scenario 7, 8, 9 and 10 could be calculated for each real routes.
  - to optimize the efficiency of the k-shortest-path algorithm so that it is computable with large-scale road networks.

- period of study
  - extend the period of study and examine the impact of other external events on the taxi drivers' behavior.

- context aware analysis
  - extend the analysis to investigate the impact on context

Reference

Arnott, R., A. de Palma, and R. Lindsey (1992). Route Choice with Heterogeneous Drivers and Group-speciffc Congestion Costs. *Regional Science and Urban Economics 22 (1),* 71-102.

Ciscal-Terry, W., M. Dell'Amico, N. S. Hadjidimitriou, and M. Iori (2016). An Analysis of Drivers Route Choice Behaviour Using GPS Data and Optimal Alternatives. *Journal of Transport Geography* 51, 119-129.

de Carvalho, E. S., S. Rasouli, and H. Timmermans (2016). Use of Concepts of Regret and Disappointment to Model Dynamic Route Choice Behavior under Uncertainty. *Transportation Research Board 95th Annual Meeting* 2567, 131-138.

Dijkstra, E. W. (1959). A Note on Two Problems in Connection with Graphs. *Numerische Mathematik,* 269-271.

Eppstein, D. (1998). Finding the k Shortest Paths. *SIAM Journal of Computing* 28 (2), 652{673.

Reference

Manley, E. J., J. D. Addison, and T. Cheng (2015). Shortest Path or Anchor-based Route Choice: A Large-scale Empirical Analysis of Minicab Routing in London. *Journal of Transport Geography* 43, 123-139.

Newson, P. and J. Krumm (2009). Hidden Markov Map Matching through Noise and Sparseness. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* - GIS '09 , 336.

Papinski, D. and D. M. Scott (2011). A GIS-based Toolkit for Route Choice Analysis. *Journal of Transport Geography* 19 (3), 434-442.

Quddus, M. A., W. Y. Ochieng, and R. B. Noland (2007). Current Map-matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions. *Transportation Research Part C* 15, 312-328.

Sun, D. J., C. Zhang, L. Zhang, F. Chen, and Z.-R. Peng (2014). Urban Travel Behavior Analyses and Route Prediction Based on Floating Car Data. *Transportation Letters* 6 (3), 118{125.

THANKS!