

Het verkeerskundig  
laboratorium  
voor studenten

ITS EDU LAB

# Dynamic OD matrix estimation using floating car data

Rúna Ásmundsdóttir

March 2008



Rijkswaterstaat

 TU Delft



# Dynamic OD matrix estimation using floating car data

Floating car data used for a-priori estimation and  
route choice analysis

MSc thesis – Civil Engineering  
Delft University of Technology

---

## Colophon

**Published by:** ITS Edulab

Telephone: 010 – 282 5881

E-mail: [henk.taale@rws.nl](mailto:henk.taale@rws.nl)

**Written by:** Rúna Ásmundsdóttir

**Graduation** Prof. Dr. H.J. van Zuylen

**committee:** Ir. H. Taale

Dr. Y.S. Chen

Dr. Ir. J.W.C. Lint

Ir. M. van den Berg

Ir. P.B.L. Wiggeraad

ITS Edulab is a project of Rijkswaterstaat – Centre for Transport and Navigation and the Delft University of Technology



**Date:** March 20, 2008

---

# Abstract

---

Travel demand in terms of an *Origin-Destination* (OD) matrix is an essential input for network traffic assignment and traffic simulation models. Several different methods that estimate OD matrices exist. For most of them, a-priori information (in the form of matrices) is necessary. Until this day, that information has been acquired from non-current data sources, such as household surveys and road interviews. Due to possible changes in the network or the traffic demand there is no guarantee that these data are still accurate.

The quality of OD matrix estimations depends largely on the quality of the input data. This thesis is aimed at improving the input data and hence improving the estimations themselves.

Detailed traffic data has been collected in the city centre of Chengdu in China. The data comprise traffic counts, video camera data and *Floating Car Data* (FCD) that come from taxis. In this thesis the FCD are used to derive a-priori matrices and analyze route choices and *Trip Length Distributions* (TLD).

From the results it can be concluded that FCD can indeed be used for estimating a-priori matrices and analysing the route choices. However, due to the fact that FCD are only sample data, they lack some information. This lack of information can be compensated with traditional a-priori matrices and route choice analyses. In that way, additional, current information is added to the traditional input data, which increases their reliability.

The TLD analyses made with the FCD are consistent between days, but due to the limited size of the study area, short trips are overestimated. With a larger study area the TLD could be used to scale OD matrices so that they match the total traffic. That eliminates estimation bias caused by the data coming from taxis.

Since the real OD matrix estimations are not known, the reliability of the developed methods cannot be estimated. It is however clear that using FCD for the examined purposes is very feasible. For further research, it is recommended that the real OD matrices should be found. Other recommendations are for instance to enlarge the study area and add information to the FCD regarding the taxi's occupancy.

---

---

---

# Preface

---

This report is the final product of my education towards becoming an engineer. I feel indebted to my supervisors, Henk van Zuylen for giving me the opportunity to undertake this challenging work and for his support during its completion, Yusen Chen for his endless patience and invaluable assistance, Henk Taale for great advice and comments on my work, Hans van Lint for helping me structure this thesis in a good way, and Monique van den Berg and Paul Wiggeraad for their constructive criticism and advice.

Furthermore, I would like to thank Rijkswaterstaat. I feel privileged for the facilities that I was offered at the Edulab and for the financial support that I received. Special thanks go as well to my fellow "Edulabbers", without them these months would not have been nearly as fun.

Finally, I would like to express my gratitude to my family for standing by me through everything I have undertaken in my life, to my friends for their moral support and simply for being great people, and to Adriaan who has supported me more than he will ever know.

Rúna Ásmundsdóttir

---

---



---

# Index

---

<b>1.</b>	<b>Introduction</b>	<b>19</b>
1.1	OD matrix estimation	19
1.2	Problem formulation	20
1.3	Research goals	21
1.4	Thesis outline	21
<b>2.</b>	<b>Floating Car Data</b>	<b>23</b>
2.1	Functional concepts of FCD	23
2.2	Previous FCD research	25
2.3	Conclusions	26
<b>3.</b>	<b>Taxonomy of OD matrix estimation methods</b>	<b>27</b>
3.1	Time dimension	27
3.2	Mapping of OD flows into the network	30
3.3	Levels in estimation procedure	32
3.4	Operational applications	33
3.5	The type of network used	34
3.6	The solution approach used	34
3.7	Input data used	42
3.8	Summary and conclusions	44
<b>4.</b>	<b>Methodology of the developed estimation processes</b>	<b>47</b>
4.1	Methodology	48
4.1.1.	Rules for determining origins and destinations within FCD	48
4.1.2.	A-priori matrix estimation with FCD	50
4.1.3.	Route choice analysis with FCD	52
4.1.4.	Trip length distribution analysis with FCD	53
4.1.5.	The application of NEST	54
4.2	Classification of the developed OD matrix estimation method	55
4.2.1.	Time dimension	55
4.2.2.	Mapping of OD flows into the network	55
4.2.3.	Levels in estimation procedure	55
4.2.4.	Operational application	55
4.2.5.	The network used	56
4.2.6.	The solution approach used	56
4.2.7.	The input data used	57
4.3	Summary	58
<b>5.</b>	<b>Data preparation &amp; experimental setup</b>	<b>61</b>
5.1	Data mapping	61
5.2	Data cleaning	63
5.3	Experimental setup	63
5.3.1.	Data used for FCD processing	63
5.3.2.	Scenarios of OD matrix estimation	64
5.4	Summary	65

---

<b>6.</b>	<b>Results</b>	<b>67</b>
6.1	Considerations about the rules	68
6.2	Sensitivity analysis	69
6.2.1.	Conclusions	75
6.3	The driving behaviour of taxis	75
6.4	Estimations made with the FCD	77
6.4.1.	A-priori matrix estimation	77
6.4.2.	Route choice analysis	81
6.4.3.	Trip length distribution	83
6.4.4.	Conclusions	85
6.5	The complete OD matrix estimation method	86
6.6	Comparison with another method	87
6.7	Summary and answers to questions	89
<b>7.</b>	<b>Conclusions and suggestions</b>	<b>91</b>
7.1	Research findings	91
7.1.1.	Rules for determining origins and destinations within FCD	91
7.1.2.	A-priori matrices estimated from FCD	92
7.1.3.	Route choice analysis with FCD	92
7.1.4.	Trip length distribution analysis with FCD	92
7.1.5.	The complete OD matrix estimation method	92
7.2	Conclusions	93
7.3	Suggestions for further research	93
<b>8.</b>	<b>References</b>	<b>95</b>
<b>9.</b>	<b>Appendices</b>	<b>99</b>
9.1	Appendix 1: The recurrence model of platoon dispersion	99
9.2	Appendix 2: Space-time extended networks	100
9.3	Appendix 3: REMDOE	101
9.4	Appendix 4: Building an a priori OD matrix from junction turning fractions	102
9.5	Appendix 5: The MATLAB codes	103
9.5.1.	Data cleaning	103
9.5.2.	ODs assigned, OD matrices derived, paths computed and path lengths calculated.	103
9.5.3.	Functions	111
9.5.4.	Trip length distribution	113
9.5.5.	Input for NEST written	115
9.6	Appendix 6: Comparison of the structure of OD matrices estimated with different a-priori matrices	122
9.7	Appendix 7: Comparison of the structure of OD matrices estimated with the developed method and with REMDOE	142

---

## LIST OF FIGURES

Figure 1.1: The steps generally taken when estimating OD matrices.....	19
Figure 2.1: FCD with probe-vehicles.....	24
Figure 2.2: FCD system setup .....	24
Figure 3.1: Dynamic assignment framework (Van Zuylen et al., 2006) .....	31
Figure 3.2: Three of the possible routes between A and D .....	32
Figure 3.3: The two levels in a bi-level approach.....	33
Figure 3.4: A simple network with traffic counts and three possible OD matrices.....	35
Figure 3.5: The taxonomy of OD matrix estimation methods .....	45
Figure 4.1: Classification of the developed estimation method, based on the taxonomy defined in Chapter 3 .....	47
Figure 4.2: A suggested process for deriving CFCD .....	51
Figure 4.3: The shortest path is inserted to complete the path.....	52
Figure 4.4: The inputs, output and iteration process of NEST .....	54
Figure 4.5: The study area – Chengdu’s city centre and the network.....	56
Figure 5.1: A plot of the mapped data locations with a few node numbers .....	62
Figure 6.1: : Fly-over and tunnels are located on the edges of the network.....	68
Figure 6.2: The trip distribution of May 9, with different values for the parameter <i>real stop</i> .....	69
Figure 6.3: The trip distribution of May 10, with different values for the parameter <i>real stop</i> .....	69
Figure 6.4: The change in number of total trips on May 9 when the parameter <i>real stop</i> is varied.....	70
Figure 6.5: The change in number of total trips on May 10 when the parameter <i>real stop</i> is varied.....	70
Figure 6.6: The trip distribution of May 9, with different values for the parameter <i>break</i> .....	72
Figure 6.7: The trip distribution of May 10, with different values for the parameter <i>break</i> .....	72
Figure 6.8: The change in number of total trips on May 9 when the parameter <i>break</i> is varied .....	73
Figure 6.9: The change in number of total trips on May 10 when the parameter <i>break</i> is varied .....	73
Figure 6.10: A “typical” trip distribution in the Netherlands.....	76
Figure 6.11: Traffic pattern from the FCD of 4 different days.....	76
Figure 6.12: The ratio between the number of taxis on links and traffic counts on links for 10 minutes time slices in the morning of May 9, 2007.....	77
Figure 6.13: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39 .....	80
Figure 6.14: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39 .....	80

---

Figure 6.15: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39 .....	80
Figure 6.16: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39 .....	81
Figure 6.17: The trip length distribution for all the trips of May 9 and 10 and the trip length distribution given by the OVG from 1996 .....	83
Figure 6.18: The trip length distribution, of all trips shorter than 50 km, for all the detected trips of May 9 May 10 and the trip length distribution given by the OVG from 1996 .....	84
Figure 6.19: The trip length distributions of each time slice calculated from the estimated OD matrices of May 9.....	85
Figure 6.20: The number of estimated trips from the two different methods shown graphically .....	88
Figure 6.21: Comparison between OD matrix estimated in this report and OD matrix estimated with REMDOE. Time period 8:30-8:39 .....	89
Figure 9.1: Space-time extended network.....	100
Figure 9.2: The classification of REMDOE.....	101
Figure 9.3: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:30-7:39 .....	122
Figure 9.4: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:40-7:49 .....	122
Figure 9.5: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:50-7:59 .....	122
Figure 9.6: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:00-8:09 .....	123
Figure 9.7: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:10-8:19 .....	123
Figure 9.8: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:20-8:29 .....	123
Figure 9.9: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39 .....	124
Figure 9.10: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:40-8:49 .....	124
Figure 9.11: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. The whole time period 7:30-8:49 .....	124
Figure 9.12: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:30-7:39 .....	125

---

---

Figure 9.13: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:40-7:49 .....	125
Figure 9.14: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:50-7:59 .....	125
Figure 9.15: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:00-8:09 .....	126
Figure 9.16: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:10-8:19 .....	126
Figure 9.17: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:20-8:29 .....	126
Figure 9.18: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39 .....	127
Figure 9.19: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:40-8:49 .....	127
Figure 9.20: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Total time period.....	127
Figure 9.21: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39 .....	128
Figure 9.22: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. The whole time period 7:30-8:49 .....	128
Figure 9.23: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39 .....	128
Figure 9.24: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Total time period.....	129
Figure 9.25: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:30-7:39 .....	130
Figure 9.26: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:40-7:49 .....	130
Figure 9.27: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:50-7:59 .....	130
Figure 9.28: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:00-8:09 .....	131
Figure 9.29: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:10-8:19 .....	131

---

---

Figure 9.30: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:20-8:29 .....	131
Figure 9.31: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39 .....	132
Figure 9.32: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:40-8:49 .....	132
Figure 9.33: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. The whole time period 7:30-8:49 .....	132
Figure 9.34: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:30-7:39 .....	133
Figure 9.35: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:40-7:49 .....	133
Figure 9.36: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:50-7:59 .....	133
Figure 9.37: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:00-8:09 .....	134
Figure 9.38: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:10-8:19 .....	134
Figure 9.39: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:20-8:29 .....	134
Figure 9.40: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39 .....	135
Figure 9.41: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:40-8:49 .....	135
Figure 9.42: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Total time period.....	135
Figure 9.43: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:30-7:39 .....	136
Figure 9.44: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:40-7:49 .....	136
Figure 9.45: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:50-7:59 .....	136
Figure 9.46: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:00-8:09 .....	137

---

---

Figure 9.47: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:10-8:19 .....	137
Figure 9.48: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:20-8:29 .....	137
Figure 9.49: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39 .....	138
Figure 9.50: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:40-8:49 .....	138
Figure 9.51: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Total time period 7:30-8:49 .....	138
Figure 9.52: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:30-7:39 .....	139
Figure 9.53: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:40-7:49 .....	139
Figure 9.54: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:50-7:59 .....	139
Figure 9.55: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:00-8:09 .....	140
Figure 9.56: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:10-8:09 .....	140
Figure 9.57: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:20-8:29 .....	140
Figure 9.58: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39 .....	141
Figure 9.59: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:40-8:49 .....	141
Figure 9.60: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Total time period.....	141
Figure 9.61: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:30-7:39 .....	142
Figure 9.62: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:40-7:49 .....	142
Figure 9.63: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:50-7:59 .....	142
Figure 9.64: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:00-8:09 .....	143
Figure 9.65: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:10-8:19 .....	143

---

---

Figure 9.66: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:20-8:29 .....	143
Figure 9.67: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:30-8:39 .....	144
Figure 9.68: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:40-8:49 .....	144
Figure 9.69: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:30-7:39 ..	144
Figure 9.70: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:40-7:49 ..	145
Figure 9.71: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:50-7:59 ..	145
Figure 9.72: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:00-8:09 ..	145
Figure 9.73: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:10-8:19 ..	146
Figure 9.74: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:20-8:29 ..	146
Figure 9.75: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:30-8:39 ..	146
Figure 9.76: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:40-8:49 ..	147



---

## LIST OF TABLES

Table 3.1: Combinations of OD matrix estimation methods and TA..	31
Table 3.2: Classification of several OD matrix estimation methods ...	46
Table 4.1: The tree possible situations and corresponding measures for the route choice analysis .....	53
Table 4.2: The Chengdu FCD properties .....	57
Table 5.1: The data used for different parts of the FCD processing...	63
Table 5.2: Properties of the used FCD .....	64
Table 6.1: The change in the number and percentage of trips in peak periods on May 9, caused by variation of the parameter <i>real stop</i> .....	71
Table 6.2: The change in the number and percentage of trips in peak periods on May 10, caused by variation of the parameter <i>real stop</i> .....	71
Table 6.3: The change in the number and percentage of trips in peak periods on May 9, caused by variation of the parameter <i>break</i> .....	74
Table 6.4 The change in the number and percentage of trips in peak periods on May 10, caused by variation of the parameter <i>break</i> .....	74
Table 6.5: The number of iterations necessary for the OD matrix estimation of each time slice with different a-priori matrices.....	78
Table 6.6: The number of trips in both the a-priori matrices and the estimated OD matrices.....	78
Table 6.7: The values from equation (6.1) for different proportional differences .....	79
Table 6.8: The number of detected trips in the morning of 4 days....	81
Table 6.9: The percentage of OD pairs with detected routes per time slice .....	82
Table 6.10: The proportion of OD pairs with a certain amount of detected trips on May 9.....	82
Table 6.11: The trip length distribution of May 9, May 10 and given by the OVG from 1996.....	84
Table 6.12: The trip length distributions of each time slice calculated from the estimated OD matrices of May 9.....	85
Table 6.13: The number of summed and total trips for the whole period.....	86
Table 6.14: The calculations of the chance that the trips detected within the FCD exceed the time frame of 10 minutes.....	87
Table 6.15: The number of estimated trips per time slice from the two different methods .....	88

---

## LIST OF ABBREVIATIONS (in order of appearance)

TA:	Traffic Assignment
OD:	Origin-Destination
FCD:	Floating Car Data
TLD:	Trip Length Distribution
GSM:	Global System for Mobile communication
GPS:	Global Positioning System
MAD:	Median Absolute Deviation
ITS:	Intelligent Transport Systems
ATIS:	Advanced Traveller Information Systems
STEN:	Space-Time Extended Network
STA:	Static Traffic Assignment
DTA:	Dynamic Traffic Assignment
DNL:	Dynamic Network Loading
ML:	Maximum Likelihood
MVN:	MultiVariate Normal
GLS:	Generalized Least Squares
AVI:	Automated Vehicle Identification
O:	Origin
D:	Destination
OFCD:	Original FCD matrix
PFCD:	OFCD where missing measurements are filled with 1
CFCD:	OFCD scaled and filled with a complete matrix
UOD:	Unit OD matrix
TC:	OD matrix based on traffic counts and turning fractions

## LIST OF MOST IMPORTANT SYMBOLS

$ij$	Index for origin-destination pairs, $ij=1,\dots,N_{OD}$
$a$	Index for links with traffic flow measurements, $a=1,\dots,N_a$
$T$	Estimated OD matrix
$t$	A-priori matrix.
$V$	Estimated flows
$v$	Counted flows
$p_{ij}^a$	The fraction of trips from $i$ to $j$ that travel via link $a$
$h'$	Departure time
$h$	Observation time
$F_1$	Objective function that minimizes the distance between the estimated OD matrix and the a-priori matrix
$F_2$	Objective function that minimizes the distance between the estimated flows and the observed flows
$\gamma_1$	A weight parameter set on $F_1$
$\gamma_2$	A weight parameter set on $F_2$
$\alpha$	Sampling fraction, sampled trips divided by total trips

---

# 1.Introduction

---

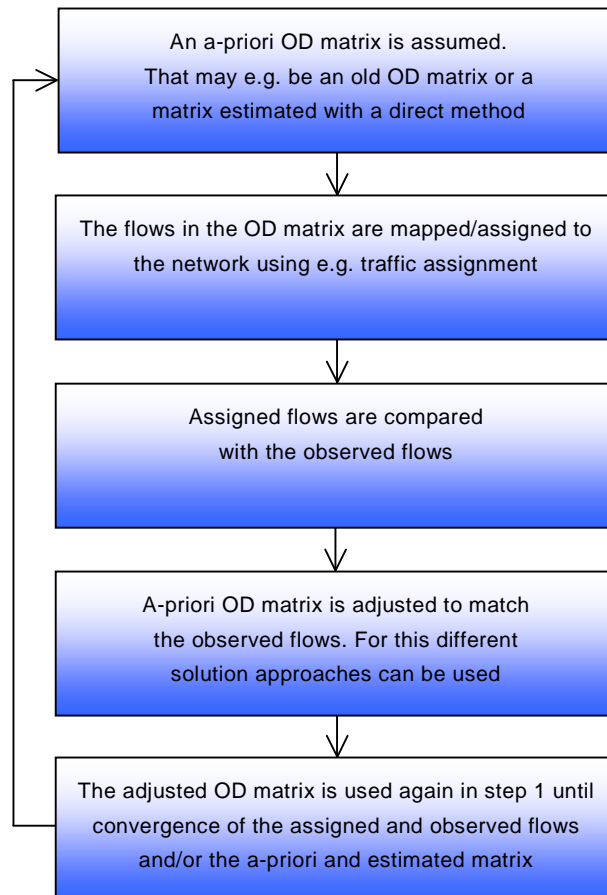
In the world today, congestion is a widespread and serious problem. Not only does it increase travel times, but it has negative effects on the environment due to an increase in emissions as well. In order to solve this problem of congestion, there are two main solutions: to extend the existing infrastructure or to improve its utilization. In both cases it is necessary to know beforehand the effect that the taken measure will have. To predict that effect, traffic simulation models and *Traffic Assignment* (TA) models are used. An essential input for both, is the travel demand in terms of an Origin-Destination (OD) matrix.

## 1.1 OD matrix estimation

There are many different ways to estimate OD matrices. Direct methods that estimate an OD matrix from e.g. household surveys and road interviews are expensive, time consuming and labour intensive. Thus, OD matrix estimation methods that use readily available but indirect data like traffic counts are used instead. This subject has been thoroughly studied in the past two to three decades and several different methods have been developed. For most of these methods, the steps that are taken in the estimation are basically the same. Those steps are shown in Figure 1.1.

---

Figure 1.1: The steps generally taken when estimating OD matrices



---

In the first step, where a-priori information is found and an a-priori matrix is constructed, traditionally, either direct methods or simply an old OD matrix from the network are used. Thus, the a-priori matrix is not built on current data, i.e. data from that current time. Due to possible changes in the network or the traffic demand there is no guarantee that these data are still accurate. The accuracy of the estimated OD matrix depends largely on the quality of the input data. Hence, it can be assumed that by using current data for the a-priori matrix, the accuracy of the estimation can be improved.

The abovementioned assumption is the drive behind the idea of this thesis work. The main focus is on improving the input for OD matrix estimation methods and hence improving the estimation itself. A new method is developed where current data is used to build an a-priori matrix. In addition, a method to analyze route choices and trip lengths is developed and with this information, a full OD matrix estimation method is provided.

## 1.2 Problem formulation

To the author's knowledge, there exist no OD matrix estimation methods that make use of current data as a-priori information. Usually, the only available current data source is traffic counts that are used in the later steps of the estimation. However, when other observation data sources exist the possibilities for estimating OD matrices increase.

The dataset used in this thesis work is vast. It includes very detailed traffic counts from loops, *Floating Car Data* (FCD) from taxis and video camera data from various points in the network. The idea is that the FCD can be used to estimate an a-priori matrix and examine the route choice and the *Trip Length Distribution* (TLD).

The FCD are a very rich source of information but still there are two serious questions that are worth to be raised regarding their reliability.

First of all, the FCD comes from just a sample of the whole traffic. Thus, they are only sample data where information might be missing and the real traffic volumes are unknown. This has to be taken into account when constructing the a-priori matrix and examining the route choice and the TLD. Hence the first question:

*Do FCD comprise enough information to build a good a-priori matrix and do they give sufficient information about the route choice and the TLD?*

The second question deals with the fact that the data come from taxis. Taxi drivers might behave different in the traffic than other drivers. This leads to the second question:

---

*Are data coming from taxis representative for the whole traffic, and if not, can the bias be estimated and adjusted?*

In this thesis, answers to these two questions will be given.

### **1.3 Research goals**

The main goal of this research is to develop a new method that estimates an a-priori matrix from current FCD. The purpose of estimating this matrix is to increase the quality of the input for OD matrix estimations and hence the estimations themselves. In addition, the FCD are used to analyze route choices and trip lengths.

The a-priori matrix and the route choice are then used together with traffic counts to estimate OD matrices. The final product of this thesis will consequently be a complete OD matrix estimation method.

### **1.4 Thesis outline**

The structure of this thesis will be as follows:

*Chapter 2* introduces the functional concepts of FCD and some of their advantages and disadvantages are discussed. Furthermore, a few papers that have been written about the different usage of FCD are reviewed.

*Chapter 3* discusses several different methods that exist for OD matrix estimation. A few well-known and important developments are mentioned in that context and a general taxonomy for OD matrix estimation methods is introduced.

In *Chapter 4* the methodologies of the developed methods are discussed and the complete OD matrix estimation is classified.

*Chapter 5* discusses how the data were prepared before they were used for the analyses and the experimental setup is described.

In *Chapter 6* results from the thesis work are presented and discussed. Furthermore, the questions raised in the beginning of this thesis are answered.

*Chapter 7* contains conclusions and suggestions for possible future steps of research.

---

---

## 2. Floating Car Data

---

FCD are a relatively new source of data. The information included in FCD is much richer than from traditional traffic data such as traffic counts, but it also lacks some information, for instance concerning traffic volumes. Still, the possibilities for utilizing FCD in traffic management are considered to be vast. Several investigations and experiments have already been made with FCD but, to the author's knowledge, none of them are in the field of OD matrix estimation.

In this chapter the functional concepts of FCD are described and their advantages and disadvantages discussed. Furthermore, a few papers that have been written about different usage of FCD are listed and reviewed.

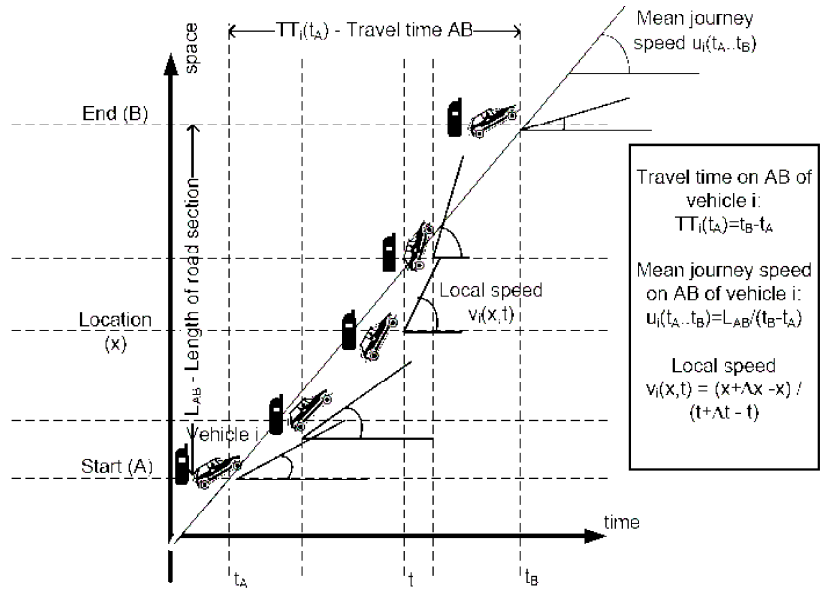
### 2.1 Functional concepts of FCD

FCD come from so-called *probe-vehicles*, i.e. vehicles that are equipped with the necessary devices to transmit data to a traffic centre at regular time intervals. The data comprise information on the status of the vehicle, for instance its location and speed (Coëmet et al., 1999). The difference between these data and data from local traffic sensors<sup>1</sup> is that the FCD sensors actually measure traffic quantities over road sections while the local quantities can only be generalized over space, at the price of assuming that vehicle operations are both homogeneous and stationary during the observation period and over the considered road section. However, an important limitation of FCD is that, until this time, only a fraction of the traffic has been equipped as probe-vehicles. Therefore, the FCD do not give any information about the complete volumes of the traffic. Furthermore, there are limitations associated with the equipment: the signal might include errors and in some locations the reception is interrupted. Figure 2.1 shows how different vehicle information can be obtained from FCD (Van Zuylen et al., 2006).

---

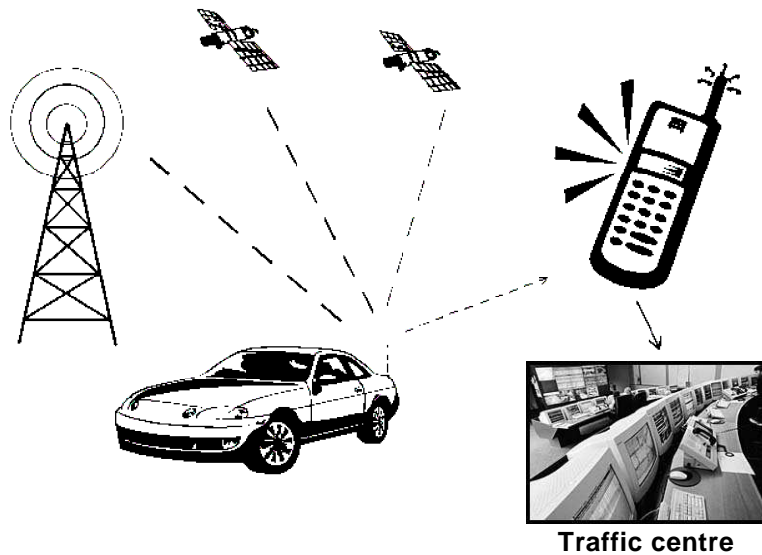
<sup>1</sup> Data from local traffic sensors are for instance vehicle counts, flows, time or harmonic mean speeds, "local" densities, proportions of vehicle types, vehicle lengths, etc.

Figure 2.1: FCD with probe-vehicles  
(Van Zuylen et al., 2006)



The equipment in the probe-vehicles is typically a *Global System for Mobile communication (GSM)* sending out a *Global Positioning Signal (GPS)*. In the traffic centre the data are processed in order to make them useable. Figure 2.2 illustrates the setup of a FCD system. The accuracy of the data depends on the frequency of the positioning and broadcasting of the data, the accuracy of the GPS and the number of probe-vehicles (Coëmet et al., 1999).

Figure 2.2: FCD system setup  
(Coëmet et al., 1999)





---

## 2.2 Previous FCD research

As mentioned earlier, several papers about the utilization of FCD exist, of which none is about OD matrix estimation. In this section, a few of them have been selected for discussion.

### **FCD – Part of “Roads of the Future” research program**

In The Netherlands, the Ministry of Transport, Public Works and Water Management (Ministerie van Verkeer en Waterstaat) has carried out an experiment with FCD. The purpose of this experiment was to investigate the usefulness of FCD and to get an understanding of the possibilities and problems with FCD. The experiment was part of a large innovation research program called “Roads to the Future”.

Approximately 60 vehicles in the city of Rotterdam were equipped with GPS and GSM devices and the data were used to estimate travel times. The results of the experiment were satisfactory. After the data had been filtered, about 75% of all the measurements could be used to estimate the travel times. The accuracy of the estimated travel times lies within 1% of the actual travel times for relatively larger road sections. In the report, it is mentioned that the FCD can as well be used for deriving OD and route-choice information (Coëmet et al., 1999 and Taale et al., 2000).

### **Deriving road networks from FCD**

For all applications of FCD it is essential to know on which roads the vehicles are travelling. For that purpose a digital network is used in most applications. The production and maintenance of these networks requires a lot of work and resources. Furthermore, the current digital networks have an inherent static nature while the real road networks are dynamic by nature – new roads are built and old ones reconstructed. Temporary changes such as road works and accidents also influence the network. In order to overcome this problem Hamerslag and Taale (2001) suggest an algorithm that derives road networks from FCD. The idea behind this is: “where there are vehicles, there must be a road”.

### **FCD for traffic monitoring**

Torp and Lahrmann (2005) proposed a complete prototype system that uses FCD for both automatic and manual detection of queues in traffic. The system consists of small hardware units placed in mobile traffic report units (taxis were used) and backstage databases that collect all the data from the report units. The automatic detection was based on analyzing GPS data from the taxis. The manual detection was based on taxi drivers reporting traffic queues by using the equipment in the taxis. A one-month field test, where 10 taxis were used, showed that the system is operational and that the communication costs are very low. The field test also provoked new questions, such as how many taxis are needed to do real-time queue detection, how to combine automatic and manual queue detection, and how to integrate the FCD with existing queue detection systems.

---

### **Local MAD method for probe vehicle data processing**

Ban et al. (2007) presented a local *Median Absolute Deviation* (MAD)<sup>2</sup> method that processes travel times from raw probe-vehicles data. Travel times generated by probe-vehicles may contain a significant amount of outliers that must be filtered. For this filtering the local MAD method is applied locally to each time window (band) with a fixed duration. A sensitivity analysis showed that for data with more than 2000 data samples per day, a bandwidth of 15 – 30 min should be used.

### **Real time route analysis based on FCD technology**

Zajicek and Reinthaler (2007) proposed a system that uses FCD to calculate detailed routes and travel times for hazardous goods transport in the Austrian road network. Furthermore the FCD are used to calculate historical time series and actual travel times.

During the 14<sup>th</sup> World Congress on Intelligent Transport Systems in Beijing, China (October 2007), discussion about FCD research was quite prominent. The above-mentioned papers by Ban et al. (2007) and Zajicek and Reinthaler (2007) were among those. Other papers on the topic are e.g. *The application of floating car system in Beijing* by Wen and Chen (2007) and *Validating travel times calculated on the basis of taxi floating car data with test drivers* by Brockfeld et al. (2007).

## **2.3 Conclusions**

The information included in FCD is much richer than that of traditional traffic data. However, it also has some limitations due to the fact that until now only a fraction of the traffic is equipped as probe-vehicles. Thus, it contains for instance no information about the traffic volumes. Despite this, it is believed that there are many possibilities in the field of FCD, of which some have already been researched. It however appears that no research exists on the utilization of FCD for OD matrix estimation.

One of the drawbacks of today's OD matrix estimation methods is that they normally use a-priori matrices build on non-current data that might no longer be valid. By using FCD, from the same period as the OD matrix estimation is being made, to estimate an a-priori matrix this drawback can possibly be eliminated. The same applies for the mapping of the OD flows. The flows detected within the FCD are real flows and thus they must give better outcomes than estimated flows.

---

<sup>2</sup> A statistical measure for capturing the variation of a given set of data points.

---

## 3. Taxonomy of OD matrix estimation methods

---

Existing OD matrix estimation methods can be classified according to several different classification methods. In this chapter, a selection of those methods is discussed. Based on this discussion, taxonomy rules for OD matrix estimation methods are built. Below, the selected classification methods are listed in the same order as they are discussed in the chapter.

**The time dimension of the estimation can differ between methods**

**There are different ways that can be used to map the traffic to the links in the network**

**The mapping of the traffic and the OD matrix estimation can be done either separately or simultaneously**

**The operational application can differ between methods**

**The type of network used in the estimations can differ**

**There are several different solution approaches that can be used in the estimation**

**Different methods might use different input data**

For all the classification methods, important developments and methods are mentioned.

The last section of the chapter provides a summary of all the classification methods that were discussed and the corresponding taxonomy rules. Furthermore, a few well-known OD matrix estimation methods are listed and classified.

### 3.1 Time dimension

The classical *static* OD matrix estimation problem is concerned with estimating OD flows given a set of link flows for a certain period of time, e.g. morning or afternoon peak-periods. This problem has been thoroughly studied through the years and various methods have been developed in order to solve it. An overview of static OD estimation methods using traffic counts can be found in Abrahamsson (1998).

---

Several methods for static OD matrix estimation using traffic counts have been introduced since 1998 and are thus not mentioned in Abrahamsson's overview. A few of them are hereafter mentioned:

**Cascetta and Postorino (2001)**

The problem is formulated for general congested network as a fixed-point<sup>3</sup> problem of an implicit function, which results from the solution of a mathematical programming problem. In other words, the solution of the problem is an OD matrix that, once assigned to the network, reproduces flows and costs that are consistent with the values used to compute the assignment fractions. Several fixed-point heuristic algorithms are proposed and their performances are compared on a small test network.

**Lo and Chan (2003)**

A bi-level (see Section 3.3) OD matrix and link choice proportions estimation. This method performs statistical estimation and TA alternately until convergence, in order to obtain the best estimators for both the OD matrix and the link choice proportions.

**Sherali et al. (2003)**

An approach based only on a partial set of link volume information. This introduces nonlinearities in the model's cost function because of the dependence of link travel costs on link volumes, and requires a determination of a fixed-point (rather than an optimal) solution to the proposed model. The fixed-point is determined heuristically by iteratively approximating the nonlinear model using a sequence of linear programs.

The disadvantage of static models is that they cannot handle the time-dynamic characteristics of traffic flow. Furthermore, they assume that the observed link flows represent a steady-state situation that remains stable over a period of time. *Dynamic* OD matrix estimation methods however specifically account for the time-dependent traffic flows. These studies are becoming increasingly important due to more demanding needs for *Intelligent Transportation Systems* (ITS) and *Advanced Traveller Information Systems* (ATIS). One of the biggest challenges regarding dynamic OD matrix estimation concerns the difference between the departure time and the observation time. Observations are made with a certain time interval. Hence, while a vehicle is travelling through traffic network it also travels between different time intervals. This can be troublesome since the travel time on a certain link in the network is not necessarily equal to the time interval. Several methods have been proposed to solve the problem of dynamic OD matrix estimation. Following are a few examples of the most important ones:

---

<sup>3</sup> A fixed-point of a function is a point that is mapped to itself by the function. That is,  $x$  is a fixed point of the function  $f$  if and only if  $f(x) = x$ . In this case it is the OD matrix that is determined by the OD matrix itself.

---

### **Cremer and Keller (1981, 1984 and 1987)**

Cremer and Keller introduced one of the first publications on the subject of dynamic OD matrix estimation in 1981. Further development of their work was published in 1984 and 1987. In their approach from 1987 the basic idea is that the traffic flow through a facility is treated as a dynamic process in which the sequences of short-time exit flow counts depend, by causal relationships, upon the time-variable sequences of entrance flow volumes. In this way it is assumed that enough information can be obtained from the counts at the entrances and the exits to obtain unique and bias-free estimates for the unknown OD matrix without further a-priori information.

### **Willumsen (1984):**

Van Zuylen and Willumsen proposed in 1980 a static entropy maximization model for OD matrix estimation. In this paper, from 1984, that model is extended to a dynamic version. This method however eliminates errors and hence requires completely consistent data. This means that in order to get a feasible solution, the flows into links must always equal the flows out of them.

### **Bell (1991b)**

Two methods are proposed for dynamic OD matrix estimation for junctions or small networks. Previous methods assumed that the time taken by vehicles to traverse the junction or network is either small in relation to the chosen time interval or equal to some fixed number of time intervals. In reality there is however a distribution of travel times that may span a number of time intervals. Bell's methods take that into account. The first method, which is appropriate if travel times are approximately geometrically distributed, makes use of the recurrence model of platoon dispersion (see Appendix 1 – Section 9.1). The second one makes no assumptions about the form of the distribution of travel times, but requires an estimation of substantially more parameters.

### **Van Der Zijpp (1996, 1997)**

In the Ph.D. thesis of Van Der Zijpp (1996), a dynamic OD matrix estimation method is proposed using a *Space-Time Extended Network* (STEN) (see Appendix 2 – Section 9.2). This method is valid for motorway networks where traffic is counted on all entries and exits (i.e. corridor network (see Section 3.5)). Van Der Zijpp proposed a further development of this work in 1997, where automated vehicle identification data are used as well.

### **Sherali and Park (2001)**

In this paper a least squares optimization approach for dynamic OD matrix estimation for a general road network is proposed. The algorithm uses a decomposition scheme that employs a restricted master program, along with dynamic shortest path sub-problems. That is done in order to generate additional path information as needed to solve the problem. The sub-problem is a dynamic shortest path problem on an expanded time-space network, while the master program is solved by using a so-called projected conjugate gradient method.

---

### **Tsekeris and Stathopoulos (2005)**

This paper proposes a demand-orientated methodological approach for estimating the operational performance of extended traffic networks, particularly under the effect of routing information provision. This information is based on the dynamic estimation of the most recent OD matrix using time-series of traffic counts. The estimation of the most recent OD matrix provides a better representation of the prevailing traffic conditions than those produced by the loading of the historical a-priori matrix onto the network. The results of these applications demonstrated a significant impact of traffic information provision on the network performance, in terms of the reduction of travel times and the increase of travel speeds.

### **Zhou and Mahmassani (2007)**

Zhou and Mahmassani have published several reports on the subject of dynamic OD matrix estimation. In this recent report, a structural state space model to systematically incorporate regular demand pattern information, structural deviations and random fluctuations is presented. Furthermore, a polynomial trend filter is developed to capture possible structural deviations in real-time demand.

## **3.2 Mapping of OD flows into the network**

An OD matrix has information about how many vehicles are travelling between the zones in the network, but it does not include information about the route choice of these vehicles. In order to be able to compare the assigned flows with the actual flows, one has to map the OD flows on the links in the network. The three most common ways to do this mapping are: *Traffic assignment* (TA) – either *static* (STA) or *dynamic* (DTA) – and direct estimation of the path-flows.

In the case of STA, only one time period is considered and the travel demand is assumed to be constant. DTA models are however able to capture the true dynamic nature of the traffic. Therefore, in several ways the DTA model provides richer and more reliable information than the STA model. The outcomes from STA can be very unrealistic. Following are examples of that:

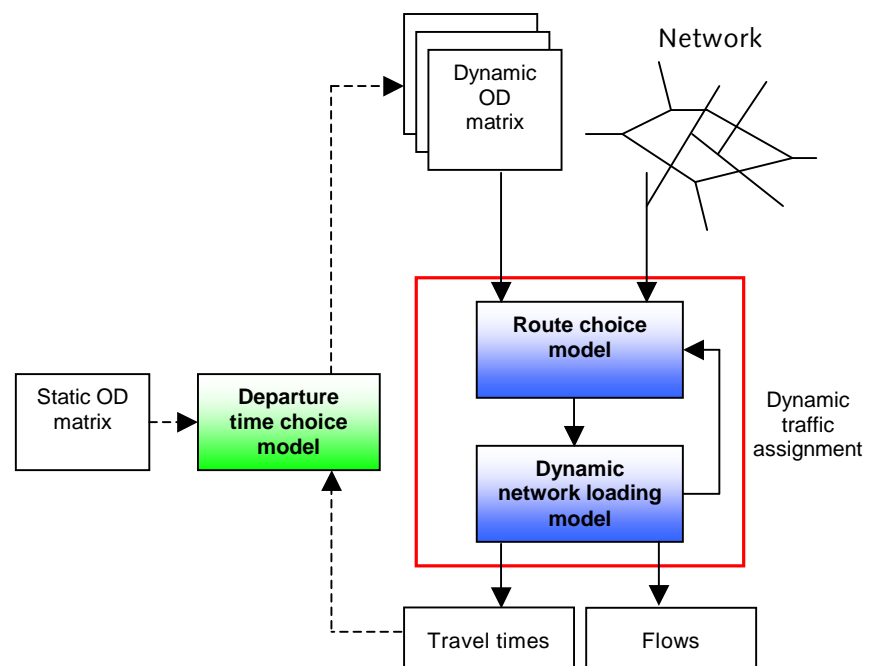
- The incorrect assumption is made that all vehicles can complete their trips within a certain period of time.
- The true congestion due to peak demand is underestimated.
- The true congestion due to dislocation of bottlenecks is overestimated.

Further explanation of these examples can be found in Van Zuylen et al. (2006).

The framework for DTA can be seen in Figure 3.1. A model set normally consists of a *route choice model* and a *Dynamic Network Loading* (DNL) model. The route choice model distributes the trips in the dynamic OD matrix over the available routes for each departure

time and each OD pair. The route flows are transferred to the DNL model that simulates the route flows over the network and computes the dynamic link travel times and dynamic link flows. Then, route travel times are transferred back into the route choice model and users may adapt their chosen route according to the new traffic conditions. The DTA is thus an iterative procedure, converging to dynamic traffic equilibrium. In addition of choosing which route to take, travellers can also decide to change their departure time. The *departure time choice model* uses the OD travel times for each departure time to determine the optimal departure time for all travellers and produces *departure time rates*. When these departure time rates are multiplied with the static OD matrix, a new dynamic OD matrix is created and a new DTA is performed. The departure time choice model is an optional component in DTA models (Van Zuylen et al., 2006).

Figure 3.1: Dynamic assignment framework (Van Zuylen et al., 2006)



Existing OD matrix estimation methods can thus be classified according to whether they use TA or not and, if they do, whether they use STA or DTA. For dynamic OD matrix estimation methods, all three possibilities exist. For static methods, there are only two possibilities since DTA would not be used. These five combinations are listed in Table 3.1.

Table 3.1: Combinations of OD matrix estimation methods and TA

		Traffic assignment		
		Static	Dynamic	None
OD matrix	Static	Yes	No	Yes
	Dynamic	Yes	Yes	Yes

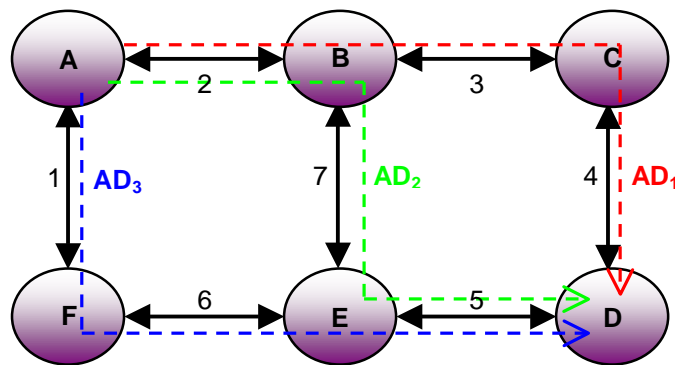
Dynamic OD estimation methods that use DTA include Cascetta et al. (1993), Ashok and Ben-Akiva (1993), Sherali and Park (2001) and Zhou and Mahmassani (2007). Examples of dynamic OD matrix estimation methods that do not use DTA are Cremer and Keller (1981, 1984 and

1987), Willumsen (1984), Nihan and Davis (1987 and 1989), Bell (1991b), Wu (1997) and Lin and Chang (2007).

The third approach is to estimate the path flows directly. The OD relations are split into different paths between the OD pairs after which the flow on each path is estimated. When this approach is used, the problem's dimension surely increases considerably but the route choice parameters are directly estimated instead of calculated from an assignment.

To further explain these path flow approaches the simple network in Figure 3.2 can be used. It shows 3 out of 4 possible routes between A and D. For these 3 routes, 3 different OD pairs have to be made;  $AD_1$  – using links 2-3-4;  $AD_2$  – using links 2-7-5; and  $AD_3$  – using links 1-6-5.

Figure 3.2: Three of the possible routes between A and D



An important drawback of these path flow approaches is that the same set of measurements is used for both the route choice and the OD matrix estimation. Therefore, both estimates are based on the same timescale. The equilibrium based route choice model requires that the measurement set contains information about completed trips. Thus, a lower bound is imposed on the admissible aggregation period. That lower bound is the time it takes for most vehicles to reach their destination. In large networks this causes long aggregation periods that are ill-suited for traffic control (Vukovic, 2007).

### 3.3 Levels in estimation procedure

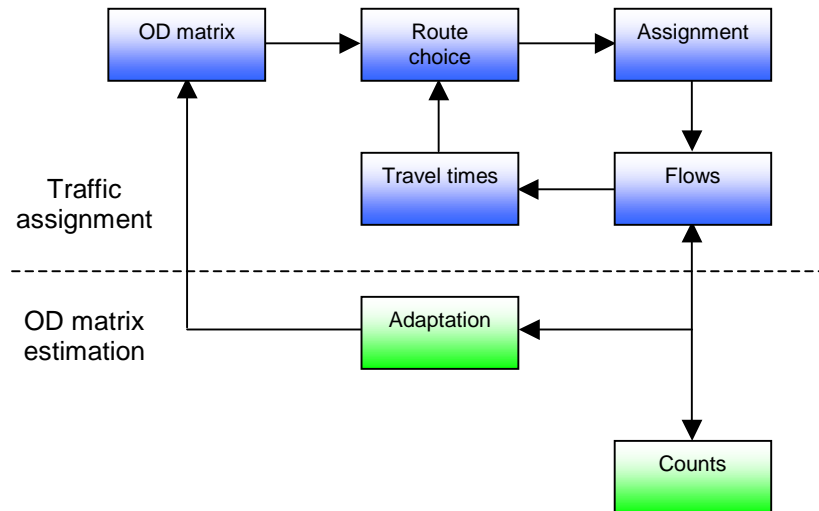
Link choice proportions in a network change with traffic conditions. In other words, when a certain link gets congested, travellers might deviate from their usual route and use other links. Thus, some OD matrix estimation methods use a *bi-level* approach where one level is for the TA and the other for the OD matrix estimation (Figure 3.3). These two levels work alternately in an iterative process until convergence.

An example of a bi-level approach is the dynamic OD matrix estimation method proposed by Zhou et al. (2003). In that method the upper-level problem is a constrained ordinary least-squares problem, which estimates the dynamic OD demand based on given link-flow



proportions. At the lower level, the link-flow proportions are generated from a dynamic traffic network-loading problem, which is solved by a DTA simulation program. Other examples are the static OD matrix estimation method proposed by Lo and Chan (2003) and a method proposed by Fisk (1988), where the entropy maximization method of Van Zuylen and Willumsen (1980) is combined with TA.

Figure 3.3: The two levels in a bi-level approach



For the *one-level* approaches the TA process is either done separately (which can lead to inconsistencies in the link choice proportions, especially in a congested network) or the path flows are estimated directly.

Many methods that use the one-level approach exist such as the static OD matrix estimation method proposed by Van Zuylen and Willumsen (1980).

### 3.4 Operational applications

Another basic difference between existing OD matrix estimation methods is their operational application, i.e. whether they operate on-line or off-line.

In the on-line methods, the new information for the evolution of the OD matrix arrives piecemeal. These methods are needed for ITS and ATIS. The models should have real-time processing capabilities, including the facility to be frequently updated with changes in the traffic pattern (Sherali and Park, 2001).

Examples of methods that work on-line are Cremer and Keller (1987) and Sherali and Park (2001).

In the off-line approaches all the data for the OD matrix are available at once. These models provide more accurate estimates based on historical data, but are computationally more intensive. The off-line models are well suited for planning purposes (Sherali and Park, 2001).

---

Examples of methods that work off-line are Van Der Zijpp (1996 and 1997) and Chen (1993).

On-line methods are more complex than off-line methods since they have to predict the destination of each started trip.

### 3.5 The type of network used

There are two types of networks that can be used for OD matrix estimation, a *corridor network* and a *general network*. When corridor networks are used there can be no route choice. Furthermore, it is essential that all entries and exits in the network are continuously monitored. For the general network, that requirement is relaxed and route choice can be used. General networks are important when handling real-world situations, in which corridor networks hardly exist. Corridor networks are mostly used for single motorways.

Several researchers, such as Cremer and Keller (1987) and Van Der Zijpp (1996 and 1997), have used corridor networks for their approaches. Methods using general networks have been proposed by e.g. Van Zuylen (1981), Cascetta et al. (1993) and Sherali and Park (2001).

### 3.6 The solution approach used

Volume counts in a traffic network impose a set of linear constraints on the OD matrix  $T$ . Thus, when estimating the OD matrix, the basic equation is the following:

$$\sum_{ij} T_{ij} \cdot p_{ij}^a = V_a, \quad a \in A \quad (3.1)$$

Where  $V_a$  is the volume on link  $a$ ,  $T_{ij}$  is the number of trips from  $i$  to  $j$ ,  $p_{ij}^a$  is the fraction of these trips that travel via link  $a$ , and  $A$  is the set of links in the transportation network. This equation is only valid when both the time dimension (see Section 3.1) and the mapping (see Section 3.2) are static. For the time dimension to be dynamic the departure time ( $h'$ ) has to be included in  $T_{ij}$  and the observation time ( $h$ ) has to be included in  $V_a$ . Then (3.1) becomes:

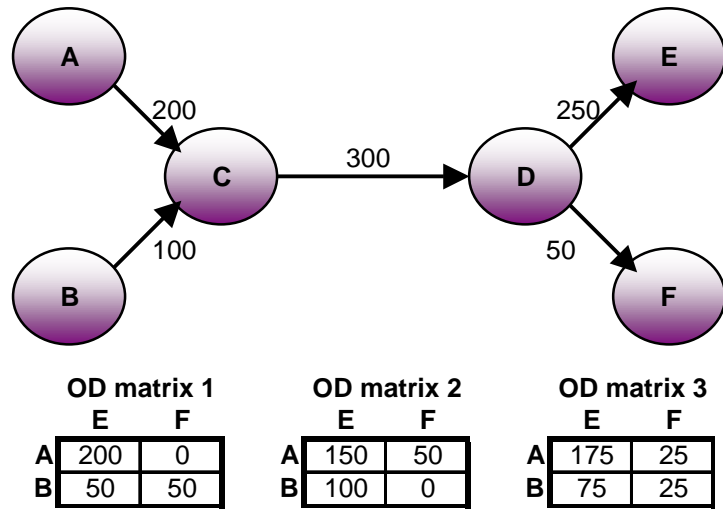
$$\sum_{ij} \sum_{h'=0}^h T_{ij}(h') p_{ij}^a = V_a(h), \quad a \in A \quad (3.2)$$

For the mapping to be dynamic, the departure time and observation time have to be included in  $p_{ij}^a$ . Then (3.2) becomes:

$$\sum_{ij} \sum_{h'=0}^h T_{ij}(h') p_{ij}^a(h, h') = V_a(h), \quad a \in A \quad (3.3)$$

The information contained in equations (3.1) – (3.3) is by itself not sufficient to find a unique OD matrix. The problem is under-determined, i.e. there is not only one single OD matrix that fits to the traffic counts. This can be explained with the simple (static) example shown in Figure 3.4.

Figure 3.4: A simple network with traffic counts and three possible OD matrices



Even in this case of a very simple and fully measured network there is more than one possible solution for the OD matrix, the three matrices shown are just an example of many possible matrices. For a more realistic network, with multiple route choice alternatives and OD pairs sharing common links, this problem becomes even larger.

To overcome this under-determination of the problem, several methods have been developed where certain constraints are used to find the most likely OD matrix. The major part of all OD matrix estimation methods use a specific *objective function* in their solution approach and most of them an a-priori matrix as well. Other constraints could be e.g. the distribution of the OD matrix and any other data that contain information about the structure of the OD matrix. However, those constraints are not commonly used. In the case where an objective function and an a-priori matrix are used, two functions have to be minimized: one for the distance between the estimated OD matrix,  $T$ , and the a-priori matrix,  $t$ , and another one for the difference between the estimated flows,  $V$ , and the counted flows,  $v$ . Hence the related optimization problems can be expressed in the following general form:

$$\min F(T, V) = \gamma_1 F_1(T, t) + \gamma_2 F_2(V, v) \quad T, V \geq 0 \quad (3.4)$$

where  $F_1$  and  $F_2$  are objective functions. If the a-priori matrix is very reliable and accurate,  $\gamma_1$  should be large compared to  $\gamma_2$ , which should result in an estimated OD matrix close to the a-priori matrix. In this case larger deviations between the estimated flows and the observed flows can be accepted. On the other hand, if the observed traffic counts are reliable compared to the information in the a-priori matrix then  $\gamma_2$  should be large compared to  $\gamma_1$ . The latter part of

---

equation (3.4) would then guide the estimation and lead to estimated flows that are close to the observed values. In this case larger deviations between the estimated OD matrix and the a-priori matrix can be accepted (Abrahamsson, 1998).

Following is a description of some of the most common solution approaches that have been used in OD matrix estimation methods. In the description it is assumed that an a-priori matrix is used for the estimation. The notations used are the same as above and can be summarized in the following manner:

- $T$ : the estimated OD matrix
- $t$ : the a-priori matrix
- $V$ : the estimated flows
- $v$ : the observed flows
- $F_1$ : the objective function that minimizes the distance between the estimated OD matrix and the a-priori matrix
- $F_2$ : the objective function that minimizes that distance between the estimated flows and the observed flows
- $p_{ij}^a$ : the fraction of trips from  $i$  to  $j$  that travel via link  $a$

### Maximum Likelihood

The *Maximum Likelihood* (ML) approach is a statistical method that maximizes the likelihood of observing the a-priori OD matrix and the observed traffic counts, conditional on the true (estimated) OD matrix. It is assumed that the elements of the a-priori OD matrix are obtained as observations of a set of random variables. The observed traffic counts constitute another source of information about the OD matrix to be estimated. The errors of the observed traffic counts and the a-priori matrix are usually considered to be statistically independent. Due to this independence, the likelihood of observing both the a-priori OD matrix and the traffic counts is equal to the product of the two likelihoods:

$$L(t, v|T) = L(t|T) \cdot L(v|T) \quad (3.5)$$

When the ML principle is applied for this problem it amounts to seeking the OD matrix that maximizes this likelihood. With the convention that  $0 \cdot \ln(0) = 0$  it is also possible to maximize the logarithm of the product.

If a simple random sampling in a region with a stable travel pattern is used to obtain the a-priori OD matrix, it may be assumed that the a-priori OD matrix follows a multinomial distribution. This is dependent on small sampling fractions  $\alpha_i$ . If  $t_i$  trips are sampled out of a total of  $T_i$  trips at the origin  $i$  then  $\alpha_i = t_i / T_i$ . The logarithm of the probability  $L(t|T)$  then results in:

$$\ln L(t|T) = \sum_{ij} t_{ij} \ln(\alpha_i T_{ij}) + const. \quad (3.6)$$

This equation corresponds to  $F_1$  in equation (3.4). If the sampling fractions are large enough, Poisson probability distribution can be assumed for the a-priori OD matrix. Then we get for  $F_1$ :

$$\ln L(t|T) = \sum_{ij} (-\alpha_i T_{ij} + t_{ij} \ln(\alpha_i T_{ij})) + const. \quad (3.7)$$

If the observed traffic counts are assumed to be generated by a Poisson distribution as well, and independent of the a-priori OD matrix, another similar expression for the probability  $L(v|V(T))$  can be found:

$$\ln L(v|V(T)) = \sum_{a \in A} (v_a \ln(V_a(T)) - V_a(T)) + const. \quad (3.8)$$

where  $v_a(T)$  is the flow volume on link  $a$  resulting from an assignment of  $T$ . This equation corresponds to  $F_2$  in equation (3.4). If one assumes a *MultiVariate Normal* (MVN) distribution for the error terms of the observed traffic counts with zero mean and a variance-covariance matrix  $W$ , we get for  $F_2$ :

$$\ln L(v|V(T)) = -\frac{1}{2} (v - V(T))' W^{-1} (v - V(T)) + const. \quad (3.9)$$

If equations (3.8) and (3.9) and proportional assignment are valid assumptions, the OD matrix estimation problem can be formulated in the following manner:

$$\begin{aligned} \max & \left[ \sum_{ij} (-\alpha_i T_{ij} + t_{ij} \ln(\alpha_i T_{ij})) + \sum_{a \in A} (v_a \ln(V_a(T)) - V_a(T)) \right] \\ \text{s.t.} & \begin{cases} \sum_{ij} p_{ij}^a T_{ij} = V_a, \forall a \in A \\ T_{ij} \geq 0 \end{cases} \end{aligned} \quad (3.10)$$

OD matrix estimation methods using ML have been proposed by e.g. Spiess (1987) (where equation (3.10) is one of the optimization problems considered and solved with an algorithm of the cyclic coordinate ascent type) and Nihan and Davis (1989) (Abrahamsson, 1998).

---

### Generalized least squares

For the *Generalized Least Squares* (GLS) approach it can be assumed that the a-priori OD matrix is obtained from the estimated "true" OD matrix with a probabilistic error term. The traffic counts may in the same way be viewed as obtained from a stochastic equation:

$$\begin{aligned}t &= T + \eta \\v &= V(T) + e\end{aligned}\tag{3.11}$$

where  $\eta$  is the probabilistic error that relates the a-priori matrix with the estimated OD matrix and  $e$  the error that relates the observed flows with the estimated flows. Frequently, both  $\eta$  and  $e$  are assumed to have zero means.

For the GLS estimator that is derived below, no distributional assumptions need to be made for  $\eta$  and  $e$ ; there is only a requirement on the existence of dispersion matrices<sup>4</sup>. When no accurate dispersion matrices are available unity matrices (with diagonal elements equal to 1) have often been used. This independence of distributional assumptions is an important advantage of the GLS approach. Like in the ML approach the a-priori OD matrix and the observed traffic counts are independent from each other. If the dispersion matrix of the traffic counts is  $W$  and the a-priori OD matrix has an error with a variance-covariance matrix  $Z$ , the GLS estimator can be obtained by solving the following:

$$\begin{aligned}\min & \left[ \frac{1}{2} (t - T)' Z^{-1} (t - T) + \frac{1}{2} (v - V(T))' W^{-1} (v - V(T)) \right] \\s.t. & \quad T_{ij} \geq 0\end{aligned}\tag{3.12}$$

An important factor of the GLS approach is that the observed traffic counts and the a-priori OD matrix are readily combined. For example, if either of the dispersion matrices is close to zero (which reflects great confidence in that part of the information), the inverse of the matrix is very large. This means that the weights on the corresponding deviations are large and as a result, the model reproduces this part of the observed information when the minimum is attained. The approximation of the dispersion matrix  $Z$  can be done in different ways, amongst other if an origin-based simple random sampling is adopted, an approximation that becomes sparse may be developed. The dispersion matrix  $W$  is often considered to be diagonal and therefore no covariances between the different traffic counts are assumed (Abrahamsson, 1998).

Several OD matrix estimation methods using the GLS approach exist. Cascetta (1984) derived two estimators for static methods. Nihan and

---

<sup>4</sup> A matrix containing the scattering values of a variable around the mean or median of a distribution

---

Davis (1987, 1989) and Bell (1991a) also proposed static OD matrix estimation methods. In Bell's method the non-negativity constraints on the estimated OD matrix are explicitly considered. One of the first publications on the subject of dynamic OD matrix estimation was by Cremer and Keller (1981) who used the GLS approach for their method. Further development of their work can be found in their 1984 and 1987 publications where they propose an addition to the GLS in order to satisfy certain equality constraints. A recent method for dynamic OD estimation can be found in Sherali and Park (2001), where a bi-level, least squares optimization approach is proposed for a general road network.

### Entropy maximization and minimum information

Because the information provided by the traffic counts on some links is insufficient to determine a unique OD matrix, it is possible to argue that one should choose a "minimum information" OD matrix. That is an OD matrix that adds as little information as possible to the information in the a-priori matrix, while taking into account the equations relating the observed traffic counts to the estimated OD volumes. The estimated, minimum information matrix is obtained from minimizing the function  $I$ :

$$I = \sum_{ij} T_{ij} \ln\left(\frac{T_{ij}}{t_{ij}}\right) \quad (3.13)$$

This equation is the minimum information or entropy maximizing function. An OD matrix minimizing while reproducing the traffic count constrained in the a-priori OD matrix into account may be derived in the following way:

$$T_{ij} = t_{ij} e^{\lambda_1 p_{ij}^1 + \lambda_2 p_{ij}^2 + \dots + \lambda_A p_{ij}^A} \quad (3.14)$$

where each  $\lambda_i$  is a Lagrange multiplier associated with the constraint that relates the link flow with the OD matrix. Equation (3.14) relies on the assumption of a proportional assignment, i.e. constant  $p_{ij}^a$  (Abrahamsson, 1998).

Van Zuylen and Willumsen (1980) proposed two important models of this type, which later were improved by Van Zuylen (1981). Willumsen (1984) extended these models to a dynamic version and Fisk (1988) proposed another extension to a congested case by introducing the user-equilibrium conditions as constraints.

---

### Bayesian inference

The *Bayesian inference* approach considers the a-priori OD matrix as a prior probability function  $\Pr(T)$  of the estimated OD matrix. If the observed traffic counts are considered as another source of information about the estimated OD matrix with a probability  $L(v|T)$  then the Bayes' theorem can provide a method for combining those two sources of information. For the posterior probability  $f(T|v)$  of observing the estimated OD matrix conditional on the observed traffic counts we have the following:

$$f(T|v) \approx L(v|T) \cdot \Pr(T) \quad (3.15)$$

This posterior probability function allows for the determination of a confidence region for the estimated OD matrix. However, due to practical computational complications only point estimators can be obtained. This can take the form of the maximum value of the logarithm of the posterior distribution, the OD matrix that maximizes  $\ln f(T|v)$ . A Poisson probability or a MVN distribution is usually assumed for the observed traffic counts (the first term in equation (3.15)). The logarithm of  $L(v|T)$  will then be expressed by equations (3.8) or (3.9). A multinomial distribution can then be assumed for the probability function  $\Pr(T)$ . Using Stirling's approximation results in:

$$\ln \Pr(T) = -\sum_{ij} T_{ij} \ln\left(\frac{T_{ij}}{t_{ij}}\right) + const. \quad (3.16)$$

That is the minimum information function. A similar function can also be obtained with a Poisson approximation of the multinomial distribution.

Maher (1983) proposed that if one assumes that a MVN distribution holds for  $\Pr(T)$ , with mean  $q$  and dispersion matrix  $Z_q$ , we get:

$$\ln \Pr(T) \approx -\frac{1}{2}(T - q)' Z_q^{-1} (T - q) + const. \quad (3.17)$$

Here it is assumed that the proportional assignment holds. For the observed traffic counts the MVN assumption is made and it is shown that in this case the estimated OD matrix also becomes MVN distributed. Another, more recent, method using the Bayesian inference approach was proposed by Van Der Zijpp (1996, 1997).

The Bayesian inference approach is a statistical inference technique with some properties in common with the previously discussed ML and GLS approaches. However, the roles assumed by the estimated OD matrix in the classical inference approaches (ML and GLS) and the Bayesian inference approach differ. In the former case, the true  $T_{ij}$  are



---

parameters of the likelihood function  $L(t, v|T)$ , while in the latter case the  $T_{ij}$  are random variables with given prior distributions (Abrahamsson, 1998).

### Kalman filtering

The *Kalman filter* is a widely used and efficient, recursive incremental algorithm that is used to solve a GLS problem using the assumptions that  $\eta = N(0, \sigma_\eta^2)$  and  $e = N(0, \sigma_e^2)$ . In OD matrix estimation, Kalman filtering is only used for the dynamic case.

The discrete Kalman filter deals with solving the linear stochastic difference equation

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (3.18)$$

with a measurement  $z \in \mathfrak{R}^m$  that is

$$z_k = Hx_k + v_k \quad (3.19)$$

where the random variables  $w_k$  and  $v_k$  represent the process and measurement noise. These random variables are assumed to be independent of each other, white noise and with normal probability distributions.

The extended Kalman filter deals with the non-linear stochastic difference equation

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \quad (3.20)$$

with a measurement  $z \in \mathfrak{R}^m$  that is

$$z_k = h(x_k, v_k) \quad (3.21)$$

where the random variables  $w_k$  and  $v_k$  represent the process and measurement noise. Here the non-linear function  $f$  in equation (3.20) relates the state at the previous time step  $k - 1$  to the state of the current time step  $k$ . Included, as parameters, are a driving function  $u_{k-1}$  and the zero-mean process noise  $w_k$ . The non-linear function  $h$  relates the state  $x_k$  to the measurement  $z_k$  (Welch and Bishop, 2006).

The Kalman filter provides an efficient computational remedy to estimate the state of a process, in a way that minimizes the mean of the squared error. The Kalman filter is, in many aspects, very powerful. It can support estimations of past, present, and future states, and it can do so even when the precise nature of the modelled system is unknown (Welch and Bishop, 2006). Further advantages of the Kalman filter are the possibility to process measurements that are interdependent, and that it does not only give an estimate for a split matrix, but also for a

---

covariance matrix that gives an indication of the reliability of the estimate.

Various researchers have used the Kalman filter for their model developments, e.g. Cremer and Keller (1987), Nihan and Davis (1989), Ashok and Ben-Akiva (1993), and very recently Zhou and Mahmassani (2007).

A more detailed and rather simple description of both the discrete Kalman filter and the extended Kalman filter can be found in Welch and Bishop (2006).

### **3.7 Input data used**

The accuracy of an estimated OD matrix depends largely on the quality of the input data. As mentioned earlier in Section 3.6 most OD matrix estimation methods use both traffic counts and a-priori information as input. In those cases, the amount of counted links in the network and the similarities between the a-priori information and the current situation are important.

Methods that use both traffic counts and a-priori matrices include Van Zuylen and Willumsen (1980), Nihan and Davis (1987), Fisk (1988) and Sherali and Park (2001). In Cremer and Keller (1987), only traffic counts are used.

The traffic data for the OD matrix estimation could also be obtained with *Automated Vehicle Identification (AVI)* techniques such as automated license plate recognition, or they may originate from FCD, as in this research. In Van Der Zijpp (1997) traffic counts and AVI data are used together for a dynamic OD matrix estimation. In this case, data fusion becomes part of the problem. There are, to the knowledge of the author, no records of using FCD in OD matrix estimation.

---

### **INTERMEZZO – The work of Vukovic**

In Vukovic (2007) four state of the art OD estimators are chosen and their performance in the MiOS simulation suite<sup>5</sup> is determined for a range of scenarios. The result of the research is that augmented Kalman filtering is the best-suited OD matrix estimation method for on-line OD matrix estimation, while the other three methods incorporate simplifications that ensure that operational requirements are met to a lesser degree. Furthermore, it is shown that it is impossible to guarantee successful OD matrix estimations from any of the tested methods. It is recommended that researchers in the field should in the future concentrate on the addition of information to the OD matrix estimation procedure as well as on further development of maximum likelihood methods.

There are a few main differences between the work of Vukovic and the work done for this thesis. Firstly, all the methods examined by Vukovic operate on-line while the method developed here works off-line. Secondly, in this thesis additional information is used for the OD matrix estimation like Vukovic recommended. The results found in this thesis might however be used to pull real-time estimations down to ground truth, and hence increase their reliability.

---

<sup>5</sup> MiOS is a microscopic on-line simulation suite that is currently being developed at the Delft University of Technology,

---

### 3.8 Summary and conclusions

There exist many different approaches to OD matrix estimation and surely some are more developed and accurate than others. However, the needed accuracy of the OD matrix estimation depends heavily on the intended use of the OD matrix. While long term strategic alternatives can be based on rather coarse estimates, real-time ITS and ATIS require more accurate and considerably richer OD matrix information.

In this chapter several of these different approaches have been discussed. The topics addressed in this discussion can be summarized into a few classification rules for OD matrix estimation. Below these rules are listed and in Figure 3.5 the taxonomy of OD matrix estimation methods, built on these rules, is shown.

**OD matrix estimation methods are split into two main groups; static and dynamic.**

**The mapping of the OD flows to the links in the network is most commonly done either with Traffic assignment or path-flow estimation. If Traffic assignment is used it can be either static or dynamic.**

**The mapping can be done either on one level or two levels (bi-level).**

**Methods can work either on-line or off-line.**

**There are two types of networks: general and corridor.**

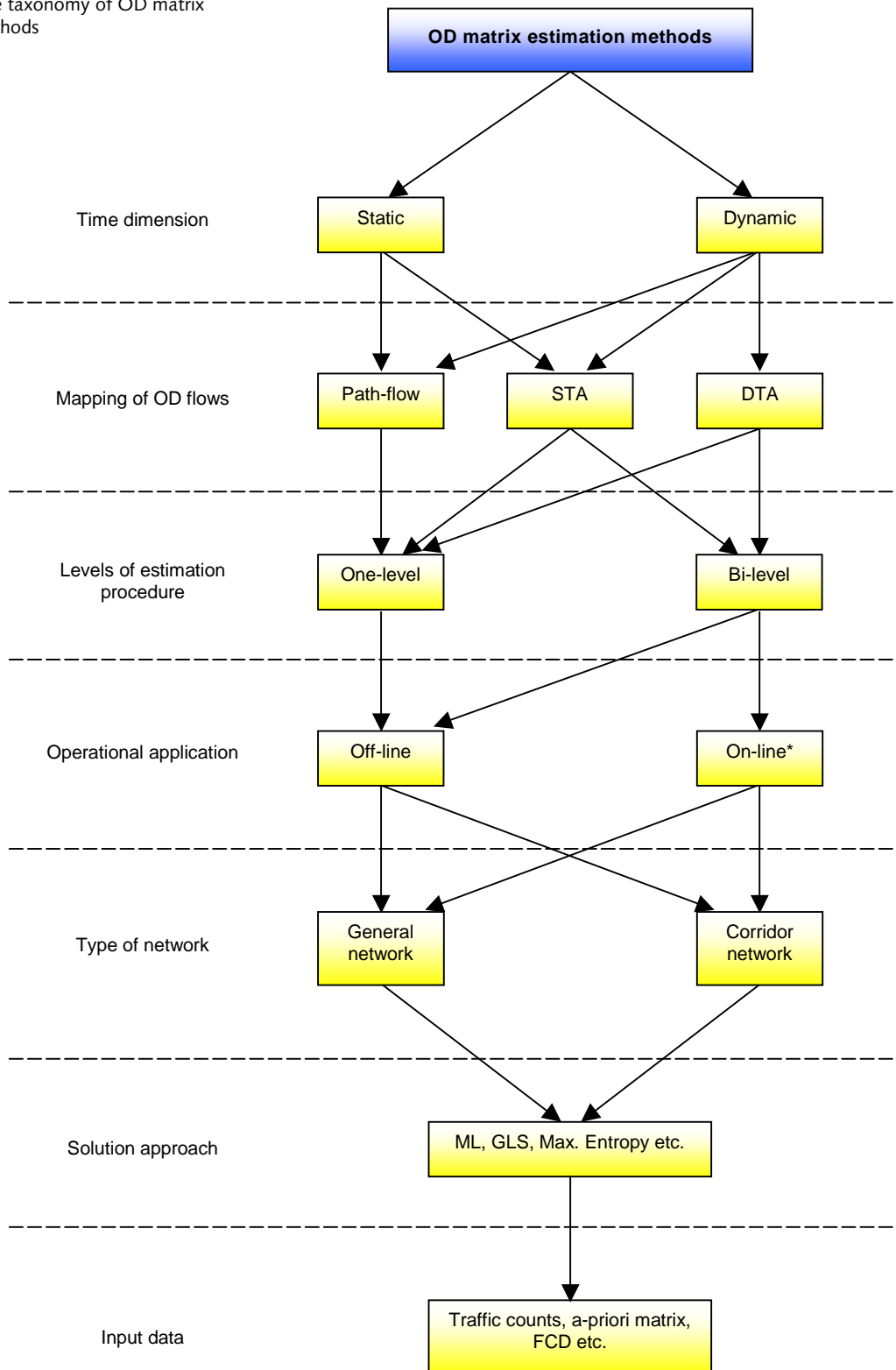
**There are several different solution approaches that can be used, i.e. ML, GLS, maximum entropy and Kalman filtering.**

**There are a few different data sources that can be used, i.e. traffic counts, FCD and video camera data.**

The main focus in this research is on the last topic addressed in this chapter, the input data. Most OD matrix estimation methods that use traffic counts use a-priori information as well. It is clear that the more accurate the a-priori information is, the more accurate the estimation will be. Until now the a-priori information have been based on non-current data, which are not guaranteed to mirror the current situation. In practice, most dynamic OD matrices are derived with existing static OD matrices used in strategic models as a-priori matrix. It can thus be assumed that the OD matrix estimation can be improved by using current data as a-priori information.

The methods that are developed in this thesis are further described and classified in Chapter 4.

Figure 3.5: The taxonomy of OD matrix estimation methods



\*Only dynamic methods can work on-line

In addition, Table 3.2 lists several well-known estimation methods and classifies them according to the rules above.

Table 3.2: Classification of several OD matrix estimation methods

	<b>Time dimension</b>	<b>Mapping</b>	<b>Use of TA</b>	<b>On-line vs. off-line</b>	<b>Type of network used</b>	<b>Solution approach</b>	<b>Input data</b>
<b>Van Zuylen and Willumsen (1980)</b>	Static	One-level	None	Off-line	General	Entropy	Traffic counts and a-priori matrix (optional)
<b>Cremer and Keller (1987)</b>	Dynamic	One-level	None	On-line	Corridor	E.g. GLS and Kalman filtering	Traffic counts
<b>Nihan and Davis (1987)</b>	Dynamic	One-level	None	Off-line	Corridor	Recursive LS and Kalman filtering	Traffic counts and a-priori matrix
<b>Bell (1991a)</b>	Static	One-level	None	Off-line	General	GLS	Survey data and traffic counts
<b>Bell (1991b)</b>	Dynamic	One-level	None	Off-line	Corridor	Recurrence model of platoon dispersion	Traffic counts
<b>Van Der Zijpp (1997)</b>	Dynamic	One-level	None	Off-line	Corridor	Bayesian	Traffic counts, vehicle trajectories and a-priori matrix (optional)
<b>Sherali and Park (2001)</b>	Dynamic	One-level	None	Both on-line and off-line	General	GLS	Traffic counts and a-priori matrix (optional)
<b>Zhou, Qin and Mahmassani (2003)</b>	Dynamic	Bi-level	DTA	On-line	General	GLS	Traffic counts
<b>Tsekeris and Stathopoulos (2005)</b>	Dynamic	Bi-level	DTA	On-line	General	Entropy maximization	Traffic counts

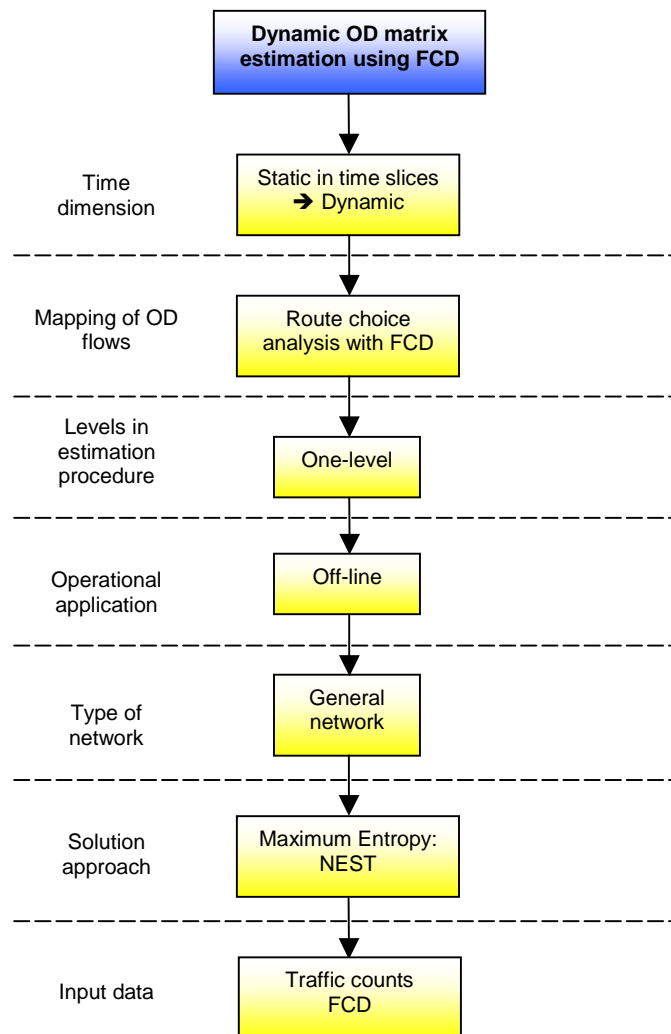
# 4. Methodology of the developed estimation processes

In the upcoming chapter the developed methods, which use FCD for a-priori matrix estimation, route choice analysis and *Trip Length Distribution* (TLD) are discussed and the complete OD matrix estimation that uses these FCD a-priori matrices and route choices is described.

In the first part of the chapter, the methodology of the a-priori matrix estimation, the route choice analysis and the TLD analysis are discussed and the application of a computer program called NEST<sup>6</sup> is described.

In the second part of the chapter, the classification of the complete OD matrix estimation method, which can be seen in Figure 4.1, is introduced. All of the seven classification levels are discussed separately.

Figure 4.1: Classification of the developed estimation method, based on the taxonomy defined in Chapter 3



<sup>6</sup> Networkflow ESTimator. A program developed by Van Zuylen based on Van Zuylen (1980).

---

## 4.1 Methodology

The first problem to be solved when analysing the FCD is how to define a beginning and an end of a trip, i.e. an *Origin* (O) and a *Destination* (D), within the data. That definition affects all information derived from the FCD. Hence, a necessary step prior to the analyses is to define rules that determine when a trip starts and ends.

Some FCD include information about the vehicle's occupancy, but the data used in this thesis does not. Thus the rules have to be based on the time of the measurements and the speed of the vehicles.

When the rules have been defined the FCD can be analyzed further and information like OD matrices, route choices and TLD can be derived.

In the following sections these rules are defined and the methodology of the a-priori matrix estimation, the route choice analysis and the TLD are discussed. Finally, the application of the computer program NEST is described.

### 4.1.1. Rules for determining origins and destinations within FCD

Since the FCD come from taxis, a trip starts at an O when a passenger enters the taxi and ends at a D when he steps out. The FCD do not give any information about whether the taxi is occupied or not. Thus, in order to derive an OD matrix from the data, a time limit has to be defined in order to distinguish between *real stops* (where passengers leave or enter the vehicle) and *intermediate stops* (e.g. on traffic lights and in traffic jams).

Another factor that distinguishes a measurement as an O or a D is when the driver takes a *break*, i.e. stops and turns the vehicle off. The GPS equipment turns on and off with the vehicle, therefore data are only transmitted when the vehicle is on. Data might also get lost due to high-rise buildings or closed facilities such as tunnels or fly-overs. Thus, a time limit needs to be defined in order to distinguish these missing data from each other.

Yet another definition for an O and a D is when a driver leaves or enters the study area. The driver could be leaving the study area only shortly during a trip that starts and ends within the study area or he might have a *break* or *real stop* while located outside the study area. Hence, a time limit needs to be set that defines when a vehicle has actually left the study area.

Finally, the first and last measurements from a vehicle must either be an O or a D.



---

The abovementioned factors can be summarized into four rules that define a start and an end of a trip within the FCD.

### **Rule 1 – Real stop**

Stopping at traffic lights should under normal circumstances not exceed 2 minutes, and when a vehicle is stopped in a traffic jam it can be assumed that its speed is not completely 0km/h for more than 2 minutes. When a trip in a taxi ends, the driver has to print out a receipt and the passenger has to pay for the ride. This process is assumed to take more than 2 minutes. These assumptions lead to the first rule of the FCD processing:

*A stop is considered to be a real stop if the measured speed is 0km/h for 2 minutes<sup>7</sup> or more. Stops that last less than 2 minutes are considered to be intermediate stops. Thus, the last measurement before a real stop is a D and the first measurement after a real stop is an O.*

### **Rule 2 – Break**

When a vehicle is driving, usually the GPS equipment sends out measurements every 1-minute. When the driver takes a *break* and turns off the vehicle, and hence the GPS equipment, it can be assumed that the last measurement before the *break* is a D and the first measurement after the *break* is an O. It is assumed that a disruption of data emissions due to high-rise buildings, tunnels and fly-overs should not, under normal conditions, exceed 2 minutes. This assumption leads to the second rule of the FCD processing:

*When the time between two measurements exceeds 2 minutes<sup>8</sup> it can be assumed that the driver has taken a break. Thus, the last measurement before the break is a D and the first measurement after the break is an O.*

### **Rule 3 – Vehicles entering/leaving study area**

The first and last measurements with speed larger than 0 km/h that are detected from a vehicle before/after leaving/entering the study area are an O or a D. The driver could however also be leaving the study area for a short time within a trip that begins and ends inside the study area. In this thesis it is assumed that when a driver leaves the study area for 2 minutes or longer he has actually left the study area. This assumption leads to the third rule of the FCD processing:

---

<sup>7</sup> This variable was tested in the sensitivity analysis in Section 6.1

<sup>8</sup> This variable was tested in the sensitivity analysis in Section 6.1

---

*A vehicle is defined to have left the study area if it dwells outside it for 2 minutes<sup>9</sup> or longer. The first measurement with speed larger than 0 km/h that is detected from a vehicle after it enters the study area is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle before it leaves the study area is a D.*

#### **Rule 4 – First and last measurements from a vehicle**

The first and last measurements with speed larger than 0 km/h that are detected from a vehicle are an O or a D. This leads to the fourth rule of the FCD processing:

*The first measurement with speed larger than 0 km/h that is detected from a vehicle is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle is a D.*

Surely, more than one rule can apply for one measurement, for instance there can be a *real stop* both before and after a measurement, a *break* right after the first measurement from a vehicle or just before the last measurement from a vehicle. In those cases these measurements are according to the rules both an O and a D. That is however not logical, thus they are neither assigned an O nor a D.

#### **4.1.2. A-priori matrix estimation with FCD**

After the beginnings and the ends of trips within the FCD have been determined, OD matrices can be derived directly from the dataset (these matrices are hereafter referred to as OFCD matrices). However, one of the disadvantages of the FCD is that they are only from a sample of the whole traffic. Hence, these OFCD matrices are also only a sample and therefore the detected trips are only a fraction of the total traffic volume. When the time slices are small it is thus highly likely that the OFCD matrices have no measured trips for OD pairs with low values. Due to this, the OFCD matrices probably do not include enough information to serve as good a-priori matrices. In order to do so, their missing values need to be replaced and they should be scaled up to match the real traffic volumes. But how can that be done in a good way?

A rough gesture would be to replace all the missing values with a low value, like 1 for instance (these matrices are hereafter referred to as PFCD). But since the values in the OFCD matrices are already very low, that would make the structure of the PFCD matrices rather uniform and thus probably different from the actual OD matrices, plus that it does not solve the problem of the traffic volumes.

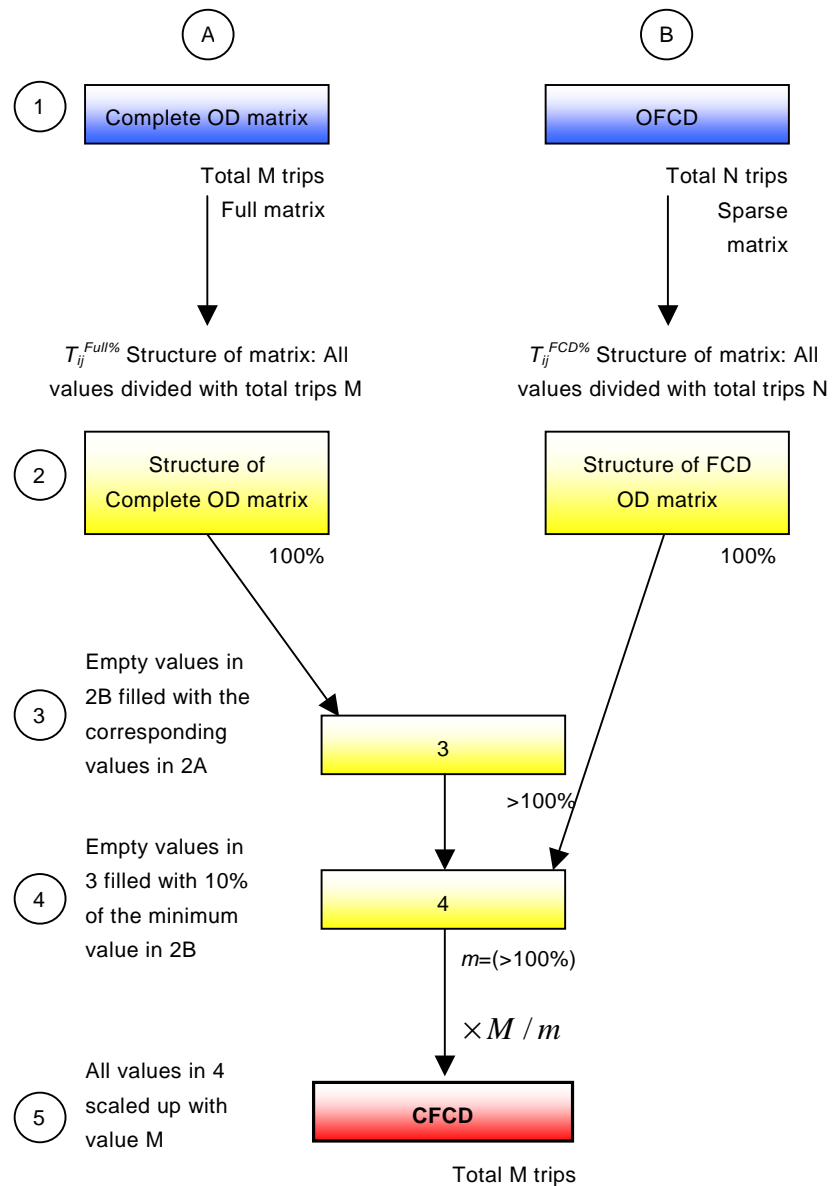
---

<sup>9</sup> This variable was tested in the sensitivity analysis in Section 6.1

One solution is to use a complete OD matrix from the area, both for the scaling and the filling of missing measurements. That matrix can for instance come from a historical database, a traffic survey or another estimation method, and it can be used both for replacing the missing measurements and for the scaling. In this way, additional information is added to the matrices that are traditionally used as a-priori matrices. The resulting matrix will hereafter be referred to as CFCD.

In Figure 4.2 a procedure for this replacement of missing measurements and scaling is suggested.

Figure 4.2: A suggested process for deriving CFCD



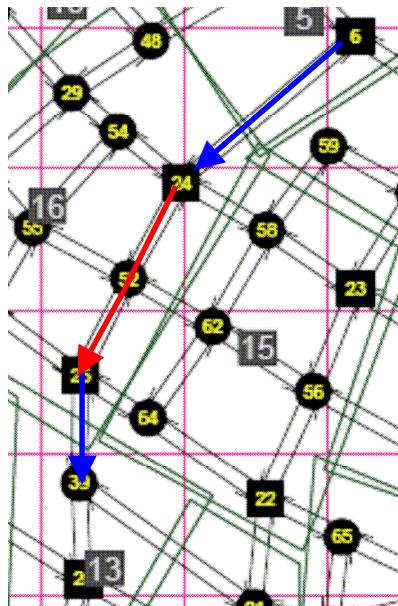
#### 4.1.3. Route choice analysis with FCD

Since the FCD contain information about the link on which the vehicles are driving as well as the driving direction, the paths of the vehicles can be traced through the network. The measured paths however are sometimes not complete. Due to e.g. missing measurements, measurement errors or just the fact that there is normally a whole minute between two measurements, links are often missing to connect two consecutive measurements. Thus, in order to construct a complete route, the computed shortest paths<sup>10</sup> are inserted between the links when needed. An example of this follows.

##### EXAMPLE

When a vehicle is detected on the link between nodes 6 and 24 in one measurements and on the link between nodes 25 and 39 in the next measurement, the shortest path between 24 and 25 (node 52) has to be inserted to complete the path. This is shown in Figure 4.3 where the blue arrows are part of the measured path and the red arrow is the inserted shortest path.

Figure 4.3: The shortest path is inserted to complete the path.



When all the routes have been constructed, the route choice analysis can be made, i.e. the values of  $p_{ij}^a$  can be calculated. The route choice analysis can then be used for the mapping of the OD flows to the network. There are three potential situations that exist for all the OD pairs and for those situations three different measures to analyze the routes. These situations and measures are listed in Table 4.1.

<sup>10</sup> Calculated with DYNASMART

Table 4.1: The tree possible situations and corresponding measures for the route choice analysis

Situation	Measure
There is no detected route between an OD	A route is found to compensate for the missing information, in this thesis the calculated shortest path is used
There is only one detected route between an OD	That particular route is used
There are more than one routes detected between OD	For all the links in those routes the parameter $p_{ij}^a$ needs to be found

The first two situations are simple since  $p_{ij}^a = 100\%$ . In case of the last situation  $p_{ij}^a$  is calculated: the total number of times that each link in the paths appears is divided with the total number of paths. To explain this, the following example is given.

**EXAMPLE**

Lets assume that for the OD pair between zone A and B, i.e.  $T_{AB}$ , the following routes are detected:

1 – 2 – 3 – 4      with one trip  
 1 – 2 – 4          with two trips  
 1 – 4                with three trips

Then we get:

$$p_{AB}^1 = (1+2+3)/6 * 100 = 100\%$$

$$p_{AB}^2 = (1+2)/6 * 100 = 50\%$$

$$p_{AB}^3 = 1/6 * 100 = 16,7\%$$

$$p_{AB}^4 = (1+2+3)/6 * 100 = 100\%$$

#### 4.1.4. Trip length distribution analysis with FCD

In addition to the necessary input for NEST the TLD was examined. The TLD can be calculated in two different ways: directly from the FCD and from the estimated OD matrix. Those are described below.

##### TLD obtained directly from FCD

When the paths within the FCD are completed, the trip lengths can be calculated. The length of all the links in the network is known. In order to calculate the length of the trips, the lengths of the links used for each trip are added up.

##### TLD calculated from estimated OD matrices

Another approach to estimate the TLD is to use the estimated OD matrix to calculate the trip lengths. In the OD matrix the number of trips between each OD pair is given, with that information and the computed shortest paths the TLD can be calculated.



---

## 4.2 Classification of the developed OD matrix estimation method

In the following sections the classification of the developed estimation method (see Figure 4.1) is discussed. All the classification levels are addressed separately.

### 4.2.1. Time dimension

The developed method performs the estimation statically in small time slices of 10 minutes. This is a simple way to extend a static method into a dynamic one and the problem's dimension is minimized. However, this method has some drawbacks, since the travel times between the ODs will probably exceed the estimation time frequently. Furthermore, there is no connection between the time slices because the estimations are made separately.

### 4.2.2. Mapping of OD flows into the network

The mapping of the OD flows to the links in the network will not be done in one of the traditional ways that were described in Section 3.2, i.e. there will be no actual TA process used nor will the path-flows be estimated directly.

In this thesis a new method is developed, in which the route choices are measured dynamically with the FCD. In that way, actual flows are used instead of only estimated flows. However, especially when the time slices are small, there is a risk of no routs being detected between some OD pairs. In those cases, the calculated shortest paths are used. This process is described in Section 4.1.3.

### 4.2.3. Levels in estimation procedure

Since the data is already available, the abovementioned route choice estimation will be done separately before the actual OD matrix estimation. This means that the OD matrix estimation is done on one level.

### 4.2.4. Operational application

On-line methods have to predict the destination for all starting trips. Thus, they are quite complex. The data used in this thesis for the complete OD matrix estimation are already available and thus it is possible to do the estimation off-line.

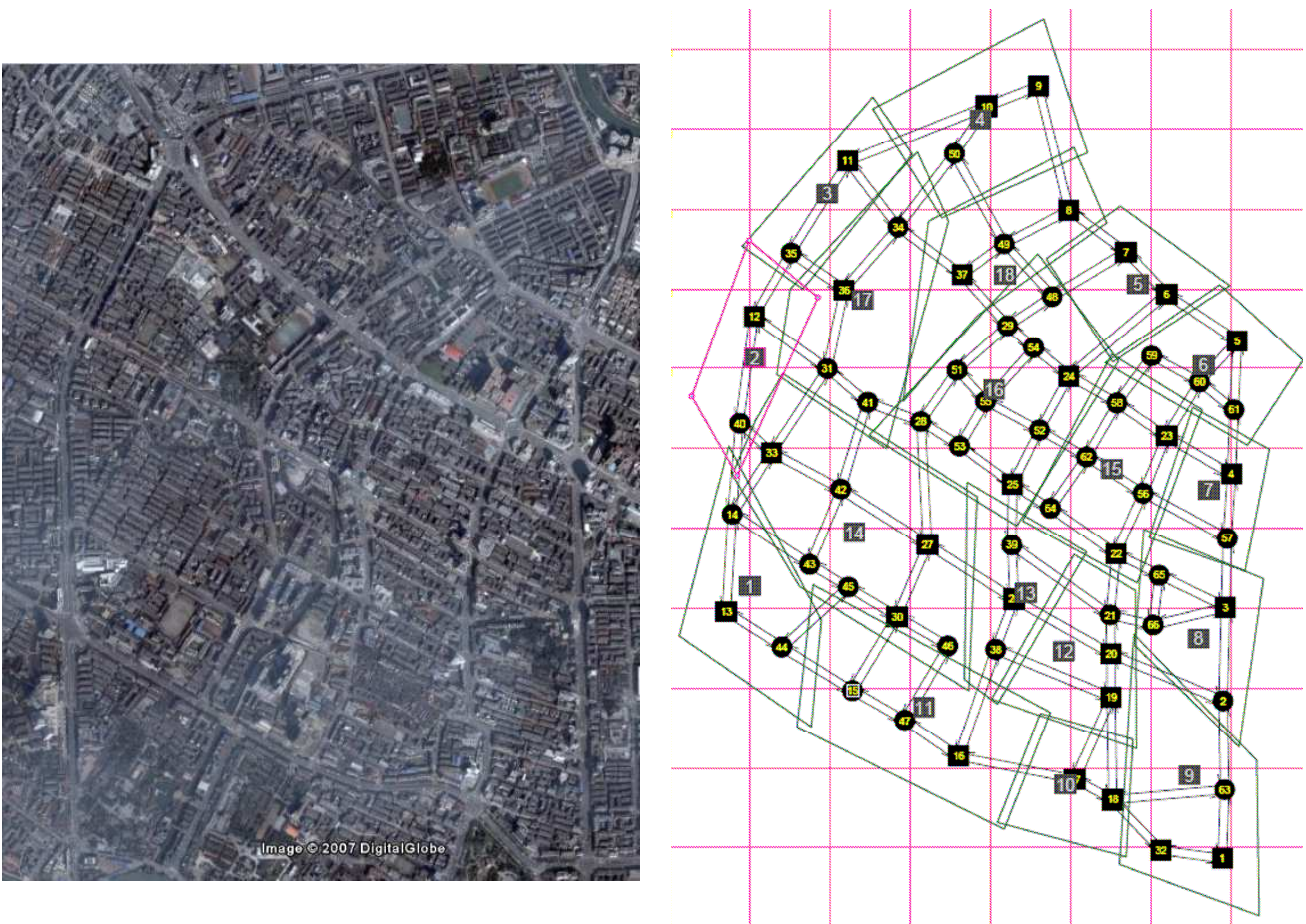
The estimations of the a-priori matrix, the route choices and the TLD are done with the knowledge of the destination of each trip, i.e. they are estimated off-line as well.

#### 4.2.5. The network used

The study area used in this thesis work is the centre of the city Chengdu in southwest China. A satellite image of the area, taken from Google Earth, can be seen in Figure 4.5a. This area is about 2 km wide and 3 km long, i.e. about 6 km<sup>2</sup>, which is rather small.

The network, built on the study area, can be seen in Figure 4.5b. It is a general network that is divided into 18 zones and has 66 nodes and 218 links. The zones in the network are circumscribed with polygons, the zone numbers are in the centre of each zone and the node numbers are on the nodes. For a network with 18 zones, the size of a generated OD matrix will be  $18 \times 18 = 324$  OD pairs.

Figure 4.5: The study area – Chengdu's city centre and the network.



#### 4.2.6. The solution approach used

A computer program called NEST, developed by Van Zuylen based on Van Zuylen (1981), is used as a tool for the OD matrix estimation. This program uses maximum entropy and minimum information to estimate the most likely OD matrix for each of the time slices. NEST performs a static estimation, but by having the time slices relatively small, the estimation is extended in a simple way to a dynamic one. NEST and its application were further discussed in Section 4.1.5.



Many other tools exist that can calculate OD matrices statically but here NEST is chosen due to its availability and short calculation times.

#### 4.2.7. The input data used

There are two<sup>11</sup> types of data used for the complete OD matrix estimation:

- Traffic counts, used in a traditional way to compare the assigned flows with the real flows,
- FCD, used to estimate a-priori matrices, route choices (and additionally, TLD), and

Following is a more detailed description of the input data.

##### Traffic counts

There are traffic counts available from 47 links in the network. The counted periods are in the morning peak, between 7:30 and 8:50, and the time interval between counts is 10 minutes.

##### FCD

The FCD from Chengdu were collected in a period of 2 months. During that period a large amount of taxis were driving around the whole city, but on the days from which data was used for this thesis work, on average 1960 probe-vehicles were detected within the study area. The interval between each measurement is 1 minute, thus the dataset is very large. If necessary, it would be possible to get the data with a smaller time interval but for this work it was considered to be unnecessary. Table 4.2 lists these main properties of the data.

.....  
Table 4.2: The Chengdu FCD properties

<b>Number of probe-vehicles in study area</b>	1960 on average
<b>Time interval between measurements</b>	1 minute
<b>Duration of data collection</b>	2 months
<b>Type of data collection equipment</b>	GPS

The Chengdu FCD comprise a lot of information about the vehicle's status. Some of the information is irrelevant to the work done in this thesis but the information that was used is the following:

- Vehicle plate number and vehicle ID.
- Date and time of the measurement.
- Longitude and latitude coordinates.
- The speed of the vehicle.

The GPS equipment switches on and off with the vehicle. The drivers can also turn it off, but since they are specifically asked not to do so, normally that is not the case. The GPS signal is lost when the vehicle is

---

<sup>11</sup> When the CFCD matrices (see Section 4.1.2) are used as a-priori matrices, additional input data is needed. That is a complete matrix from the study area. These matrices can for instance be derived from historical data or survey data.

---

located within a closed structure, like garages and tunnels and it might get lost in locations where high-rise buildings block the reception.

### 4.3 Summary

The first part of this chapter discusses the methodology of the new methods developed in this thesis. Rules that determine beginnings and ends of trips are defined and thereafter the process of deriving a-priori matrices and analysing the route choices and the TLD are described. The OD matrix estimation tool, NEST, that uses the a-priori matrices and the route choice analysis for its estimation, is as well described in the chapter. A recap of this part of the chapter follows below.

#### Rules for determining origins and destinations

Four rules were defined that determine Os and Ds within the FCD.

1. A stop is considered to be a *real stop* if the measured speed is 0km/h for 2 minutes or more. Stops that last less than 2 minutes are considered to be intermediate stops. Thus, the last measurement before a *real stop* is a D and the first measurement after a *real stop* is an O.
2. When the time between two measurements exceeds 2 minutes it can be assumed that the driver has taken a *break*. Thus, the last measurement before the *break* is a D and the first measurement after the *break* is an O.
3. The first measurement with speed larger than 0 km/h that is detected from a vehicle after it enters the study area is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle before it leaves the study area is a D. A vehicle is defined to be outside the study area if it dwells there for 2 minutes or longer.
4. The first measurement with speed larger than 0 km/h that is detected from a vehicle is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle is a D.

#### A-priori matrix estimation with FCD

Three different OD matrices that can be derived from FCD were discussed:

OFCD is the original OD matrix derived from the FCD. Due to its sparse nature and low traffic volumes, it might not comprise sufficient information to serve as a good a-priori matrix.

PFCD is a matrix where the missing values of OFCD are replaced with a low value, e.g. 1. This way of constructing an a-priori matrix however only solves a part of the problem; missing values are replaced but the

---

traffic volumes are still too low. Furthermore, since the values in the OFCD are most likely rather low, the structure of the PFCD might not resemble the structure of the real OD matrix.

Thus, the third matrix is suggested: CFCD. For constructing CFCD a combination of FCD and historical data or survey data is necessary (i.e. a traditional a-priori matrix). If those data do not exist, a matrix estimated with another method can be used. The way this matrix is constructed can be seen in Figure 4.2.

#### **Route choice analysis with FCD**

The routes of the vehicles can be detected throughout the network. Those routes are often not complete, i.e. sometimes links are missing between two consecutive measurements. In those cases, the computed shortest paths between those two measurements are inserted.

When the routes are complete, there might still be OD pairs with no detected routes. In those cases, the shortest paths between the OD pairs are used. For all the routes, the value of  $p_{ij}^a$  needs to be found for all the links in the route. When shortest paths are used and when only one route is detected for an OD pair this is simple since all  $p_{ij}^a = 100\%$ . However, when more than one route is detected for an OD pair the values of  $p_{ij}^a$  need to be calculated. The calculation procedure is described in the chapter.

#### **Trip length distribution analysis with FCD**

When all the routes have been identified and completed, the TLD can be analysed. There are two ways to derive the TLD: directly from the FCD and from the estimated OD matrix. These are described in the chapter.

#### **The application of NEST**

The chapter describes roughly how NEST uses the a-priori matrix, the route choices and the traffic counts to estimate an OD matrix.

In the latter part of this chapter, the classification of the complete OD matrix estimation method that is developed in this thesis was discussed. This classification can be seen in Figure 4.1.

---

---

## 5.Data preparation & experimental setup

---

The original FCD include only raw information about the location of the vehicles. This information needs to be adjusted and the data points mapped into the network in order to be able to work with the data. Furthermore, within the data there are several measurements that have no assigned link. That information can be cleaned out from the dataset and the measurements with no assigned link number hence treated as missing measurements. In the first part of this chapter this data preparation is described.

In the latter part of this chapter the experimental setup of the thesis work is explained.

### 5.1 Data mapping

The FCD originally include only the coordinates of the vehicles but not their location within the network, i.e. position on links and nodes. Hence, before it is possible to process the data, they must be mapped to the network. Chen (2007) designed a method that maps the data points. This method is described below.

In order to locate the measurement points in the network, the coordinates of the original data are converted to the coordinate system of the network. However, due to errors in the GPS data and the fact that not all roads in the study area are included in the network, the measurement points do not all fit exactly on a specific node. Certain measures have to be taken in order to assign the measurements to zones and links. Below, these measures are described.

#### **Zone determined**

When a measurement is detected outside the study area it is assigned a zone number 0 or  $-1$ . When a measurement is detected within only one of the 18 zones it is assigned that zone number. If a measurement is detected in an overlap of zones its distance from the centre of the zones is used to decide which zone number to use and the nearest zone centre's number is assigned.

#### **Nodes and links determined**

The measurements are assigned to the nearest link. However, the distance from that link cannot be larger than one and half the link length. If it is further away, the measurement is not assigned to any link. A measurement can thus have assigned zone number within the study area, but no assigned link. The link is represented with the numbers of the nodes on its ends; hence two numbers from 0 (for no assigned link) to 66.

---

### Direction and location on link determined

The order of the nodes (i.e. which one is node 1 and node 2) indicates the driving direction of the vehicle. In order to determine the driving direction, the next and the two previous measurements have to be considered.

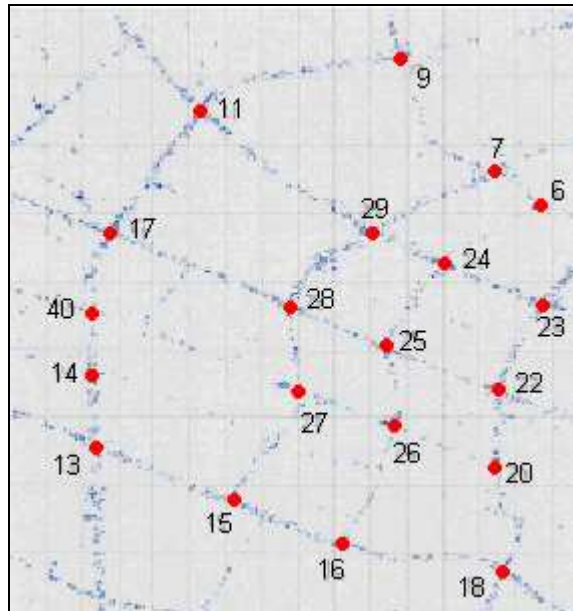
When measurements are assigned to a link, they are assigned to a location on the link as well. This location is indicated by the distance from node 1 in percentage of the total length of the link.

After the mapping, the data comprise the following information:

- The vehicle number (starting with 1 and ending with the total number of vehicles)
- Date (yyyymmdd)
- Time (hhmmss)
- Vehicle speed (km/h)
- Zone number (-1–18)
- Node 1 (0-66) and node 2 (0-66) (indicating the travelling direction)
- Location on link (%), i.e. percentage distance from node 1.

When the locations of the mapped data points are plotted one can see a rather clear picture of the network. Figure 5.1 shows a plot of the mapped data locations and a few node numbers.

.....  
Figure 5.1: A plot of the mapped data locations with a few node numbers



---

## 5.2 Data cleaning

Before the data were used, all the measurements with node number 0 were removed. Most of these measurements are from outside the study area but also from within, where the nearest link is too far away from the measurement. By doing this, measurements with no assigned links can be treated as missing measurements. Thus, rule 2 and 3 are combined.

The data cleaning has two important benefits. The calculations are simplified and the size of the dataset is decreased.

The data cleaning was performed in MATLAB and the code can be seen in Appendix 5, Section 9.5.1.

## 5.3 Experimental setup

A few decisions have to be made before the data experimenting can start. The part of the dataset that will be used for the work needs to be selected and different scenarios for the OD matrix estimation need to be defined. In the following two sections these issues are discussed.

### 5.3.1. Data used for FCD processing

As mentioned before, the FCD were collected over a period of 2 months. It is however unnecessary to use the whole dataset for the experiments. Thus, data from 4 separate dates were selected for usage. Those dates were: May 9, May 10, June 6 and June 7, 2007.

For a sensitivity analysis (see Section 6.2) all the data from 9 May and 10 May were used. All the data from the 4 days were used in order to examine the traffic pattern of the taxis by comparing the trip distribution between days while a part of the morning peak of May 9, 7:30-8:50, was used to examine whether the distribution of taxis over the network is constant over time (see Section 6.3). That same part of the morning period of May 9 was used for both the a-priori matrix estimation (see Section 6.4.1) and the route choice analysis (see Section 6.4.2). For the TLD (see Section 6.4.3) all the data from 9 and 10 May was used. The data usage is listed in Table 5.1.

.....  
Table 5.1: The data used for different parts of the FCD processing

	May 9	May 10	June 6	June 7
<b>Traffic pattern of taxis</b>	All data used	All data used	All data used	All data used
<b>Distribution of taxis over the network</b>	7:30-8:50	–	–	–
<b>Sensitivity analysis</b>	All data used	All data used	–	–
<b>A-priori matrices</b>	7:30-8:50	–	–	–
<b>Route choice analysis</b>	7:30-8:50	–	–	–
<b>Trip length distribution</b>	All data used	All data used	–	–

The properties of all the used FCD can be seen in Table 5.2.

.....  
Table 5.2: Properties of the used FCD

Date	Number of vehicles detected	Number of measurements	Number of measurements after data cleaning
May 9	2.264	215.432	105.455
May 10	2.280	233.519	115.029
June 6	865	139.906	36.222
June 7	2.415	381.449	95.053

Later in this thesis, both when the a-priori matrices are constructed and when the estimated OD matrices are evaluated, OD matrices estimated with a dynamic OD matrix estimation method called REMDOE (Chen, 1993), are used. The classification of REMDOE can be seen in Appendix 3 – Section 9.3. Those matrices were estimated for the morning period of May 9. Therefore, the data from May 9 were chosen for most of the FCD processing. The other dates were chosen randomly.

### 5.3.2. Scenarios of OD matrix estimation

The real OD matrices for the study area are unfortunately not known. Hence, it is impossible to tell whether any OD matrix estimation is good or bad, but the effect that different a-priori matrices have on the estimation outcomes can be examined. In order to do so, five different a-priori matrices were used as an input for NEST and the estimated OD matrices and the estimated traffic volumes compared. Three of these a-priori matrices have already been mentioned (see Section 4.1.2) the other two are explained below. These five a-priori matrices are:

- **UOD:** A unit OD matrix, where all interzonal trips have the value 1 and all intrazonal trips the value 0
- **OFCD:** see Section 4.1.2
- **PFCF:** see Section 4.1.2
- **CFCD:** see Section 4.1.2
- **TC:** A matrix based on traffic counts and turning fractions (see Appendix 4 – Section 9.4)

Hereafter the abovementioned acronyms will be used when the different a-priori matrices are discussed.

Since neither historical data nor survey data exist for the study area, the CFCD used here is a combination of FCD and OD matrices estimated with the previously mentioned method REMDOE (Chen, 1993). The estimations done with REMDOE were made using the traffic counts and the TC matrices as a-priori matrices. REMDOE uses DTA for the mapping of the OD flows to the network.

The other inputs for NEST, i.e. the route choice analysis and the traffic counts are kept the same for all the estimations.



---

## 5.4 Summary

The original FCD includes information about the location of the vehicles in coordinates. This information does not give indication of the location of the vehicles within the network, i.e. in which zone and on which link they are located. Hence, before the data could be used, they had to be mapped into the network.

When the data had been mapped, they include the following information:

- Vehicle number
- Date and time of the measurement
- The speed of the vehicle
- The zone in which the vehicle is located
- The end nodes of the link on which the vehicle is located
- The location on the link

In addition to the data mapping all the measurements with no assigned link were removed from the dataset. By doing so, rules 2 and 3, which were defined in Section 4.1.1, were combined.

The data used for the analyses were from May 9 and 10 and June 6 and 7, 2007. The NEST inputs, i.e. the a-priori matrices and the route choice analysis were only done for a part of the morning period of May 9. This time and day were selected in order to be able to compare the conclusions to estimations done another estimation method called REMDOE.

Since the real OD matrices for the study area are not known, it is not possible to estimate the quality of estimated OD matrices. It is however possible to examine the effects which different a-priori matrices have on the estimated OD matrices. For this purpose, four different a-priori matrices were chosen, those are: OFCD, PFCD, CFCD and TC.

---

---

## 6. Results

---

In the beginning of this thesis two important questions were raised about whether FCD, due to their sparse nature, comprise enough information to build good a-priori matrices and whether they are representative of the whole traffic. In this chapter, these questions are answered.

In the first part of the chapter some considerations about the rules that were defined in Section 4.1.1 are expressed. Following that, a sensitivity analysis is performed for the parameters *real stop* and *break*. Based on the sensitivity analysis, a final decision is made regarding the value of those two parameters.

When the values of the two parameters have been confirmed, the driving behaviour of the taxis is inspected. Firstly, the difference between a "typical" trip distribution and the trip distribution derived from the FCD is investigated. Secondly, the distribution of the taxis over the network is examined.

In Section 6.4 the estimations made with the FCD are discussed, i.e.:

- A-priori matrix estimation: The effects that five different a-priori matrices have on the OD matrix estimation are examined and a final decision is made regarding which matrix is best suited as an a-priori matrix.
- Route choice analysis: The amount of OD pairs that have a detected route during each time period is checked. The number of detected routes per OD pair is also measured.
- Trip length distribution: There are two different ways to estimate the TLD, directly from the FCD and from the estimated OD matrix. These two ways are tested and the applicability of the outcomes discussed.

The last parts of this chapter discuss the complete OD matrix estimation method. In Section 6.5 the estimated OD matrices are discussed and in Section 6.6 they are compared with OD matrices estimated with REMDOE.

All the calculations in the chapter are done either in MATLAB or Excel. The MATLAB codes are displayed in Appendix 5 – Section 9.5.

---

## 6.1 Considerations about the rules

In rules 1, 2 and 3 (see Section 4.1.1), parameters were determined to define a *real stop*, *break* and when a vehicle actually leaves the study area. By removing the measurements with node number 0, rule 2 and 3 were combined. The parameters are based on assumptions only. Hence, one might consider whether they give good results or not. Hereafter considerations about rule 1 and the combination of 2 and 3 are discussed.

### Rule 1 – Real stop

The first rule is based on the assumption that it takes 2 minutes or more for a passenger to pay, get a receipt and exit the vehicle, while stopping on an intersection or in a traffic jam should not exceed 2 minutes. However, according to a contact person in Chengdu, this payment procedure might not take so long.

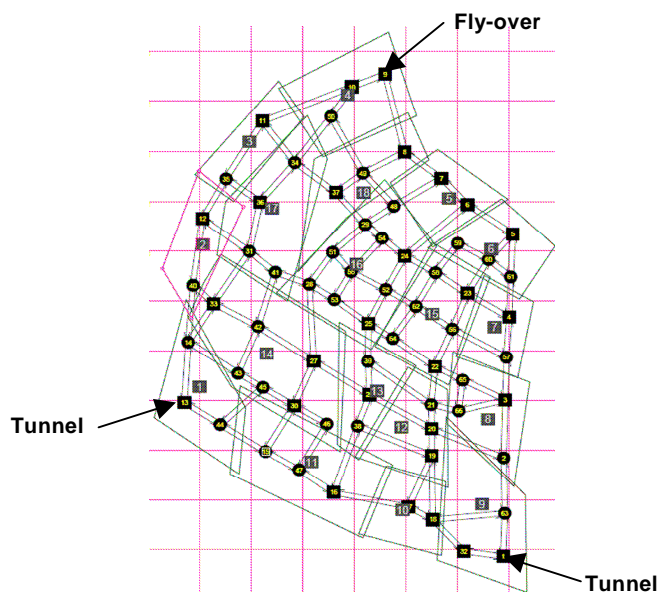
### Rule 2 & 3 – Break & vehicles entering/leaving the study area

The second rule is based on the assumption that passing a high-rise building and driving through tunnels should not take more than 2 minutes. There are two tunnels and one fly-over in the study area but they are all located on the edges of the network as can be seen in Figure 6.1. Hence, missing measurements due to tunnels and fly-overs are, in this study, irrelevant. However, in the case of congestion it might occur that a vehicle dwells for a longer period than 2 minutes below a high-rise building and thus, measurements could get lost for more than 2 minutes.

There are also some considerations about the assumptions made for the third rule. There is no guarantee that when a vehicle dwells for longer than 2 minutes outside the study area it has either a *real stop* or a *break* in the meantime.

In order to test the validity of the parameters determined in rule 1 and the combined rule 2 and 3, a sensitivity analysis was performed. In the following section, this analysis is discussed.

.....  
Figure 6.1: Fly-over and tunnels are located on the edges of the network



## 6.2 Sensitivity analysis

In order to test the validity of the decided values of the parameters *real stop* (rule 1) and *break* (rule 2 and 3 combined) a sensitivity analysis was performed. The parameters were tested separately with data recorded during two days, May 9 and 10, 2007. The trip distribution, i.e. the distribution of the number of trips per hour of the day, was examined as well as the total number of trips. In order to test either parameter, several calculations were done with a different value of each parameter while the other one was fixed to 2 minutes.

### Sensitivity of the parameter *real stop*

The parameter *break* was fixed to 2 minutes while the parameter *real stop* was changed for each calculation. Seventeen calculations were performed with the value of *real stop* ranging from 1 minute to 1.440 minutes (24hours). Figure 6.2 shows a comparison between the trip distributions for different values of the parameter *real stop* for the data collected on May 9. Figure 6.3 shows the same comparison for May 10.

Figure 6.2: The trip distribution of May 9, with different values for the parameter *real stop*

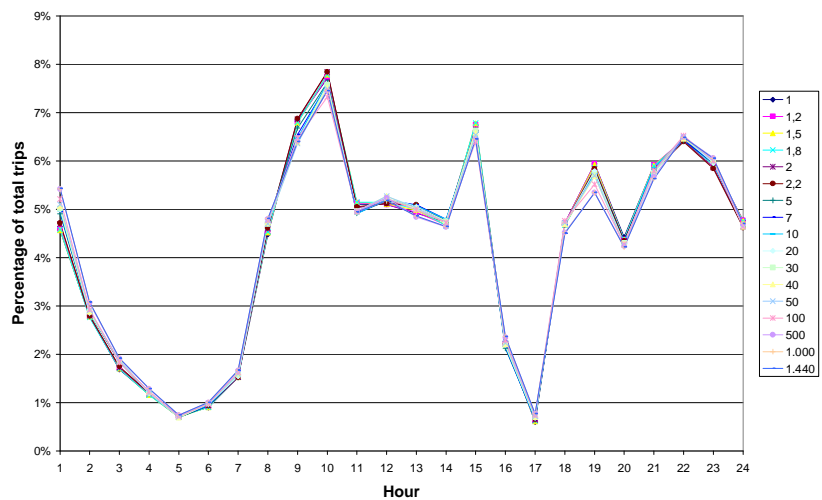


Figure 6.3: The trip distribution of May 10, with different values for the parameter *real stop*

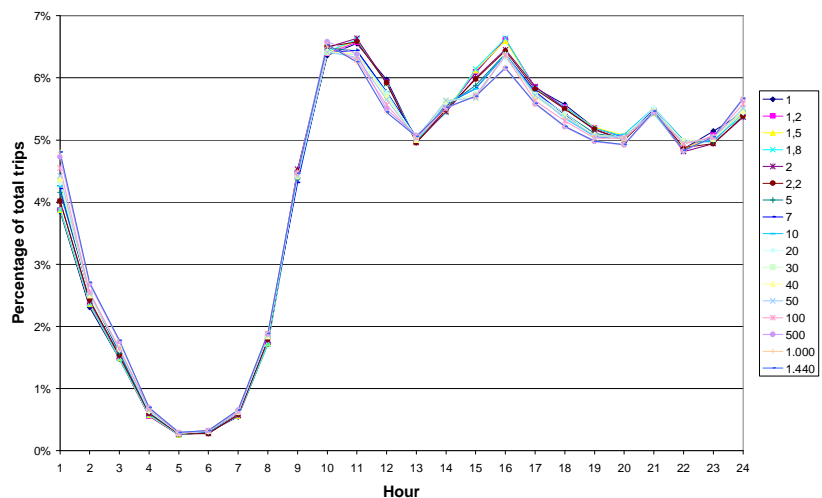


Figure 6.4 shows how the number of detected trips on May 9 changes with different values of the parameter *break* (both in the case when intrazonal trips<sup>12</sup> are included (all trips) and for interzonal trips<sup>13</sup> only). Figure 6.5 shows the same for May 10.

Figure 6.4: The change in number of total trips on May 9 when the parameter *real stop* is varied

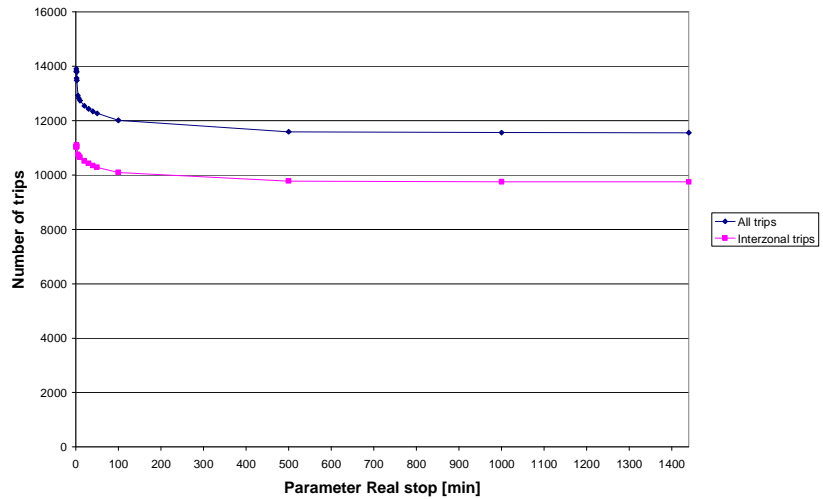
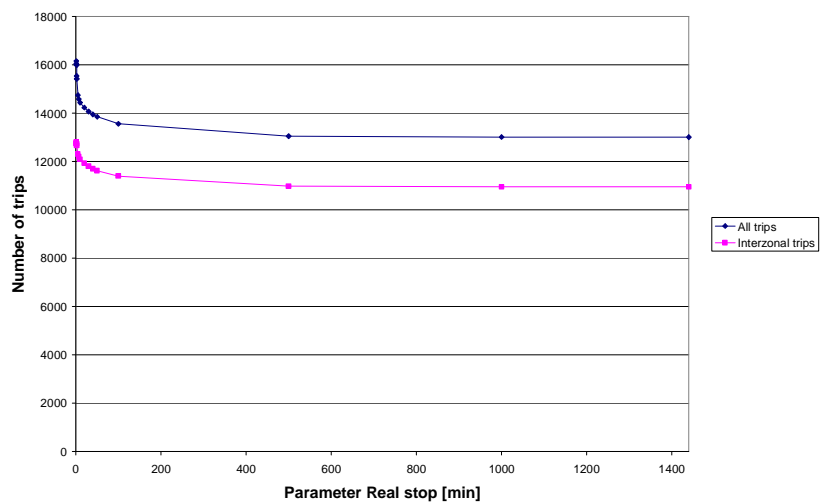


Figure 6.5: The change in number of total trips on May 10 when the parameter *real stop* is varied



From Figures 6.2-6.5 it can be concluded that variation of the parameter *real stop* will change the structure of the trip distribution only slightly, while the change in number of trips is quite substantial. The number of trips decreases when the parameter is increased until it stabilizes. The number becomes more or less stable when the parameter is set as 500 minutes and larger, and it becomes completely stable when the parameter approaches the size of 24 hours.

Table 6.1 and Table 6.2 show the number of trips (both including and excluding intrazonal trips) for the different value of *real stop* as well as

<sup>12</sup> Intrazonal trips are trips that start and end in the same zone.

<sup>13</sup> Interzonal trips are trips that start and end in different zones.

the percentage of the trips that occur in the morning and afternoon peak (7:00-10:00 and 16:00-19:00).

.....  
 Table 6.1: The change in the number and percentage of trips in peak periods on May 9, caused by variation of the parameter *real stop*

Break = 2 minutes – May 9, 2007						
Real stop [min]	All trips			Interzonal trips only		
	Trips	Morning peak	Afternoon peak	Trips	Morning peak	Afternoon peak
1	13912	19%	11%	11051	18%	10%
1,2	13830	19%	11%	11029	18%	10%
1,5	13805	19%	11%	11036	18%	10%
1,8	13788	19%	11%	11038	18%	10%
2	13561	19%	11%	11100	18%	11%
2,2	13482	19%	11%	11054	18%	11%
5	12921	19%	11%	10756	18%	11%
7	12825	19%	11%	10708	18%	11%
10	12738	19%	11%	10655	18%	11%
20	12547	19%	11%	10515	18%	11%
30	12435	19%	11%	10426	18%	11%
40	12343	19%	11%	10346	18%	11%
50	12265	19%	11%	10286	18%	11%
100	12009	19%	11%	10098	18%	11%
500	11591	19%	11%	9775	18%	10%
1.000	11560	19%	11%	9750	18%	10%
1.440	11559	19%	11%	9748	18%	10%

.....  
 Table 6.2: The change in the number and percentage of trips in peak periods on May 10, caused by variation of the parameter *real stop*

Break = 2 minutes – May 10, 2007						
Real stop [min]	All trips			Interzonal trips only		
	Trips	Morning peak	Afternoon peak	Trips	Morning peak	Afternoon peak
1	16157	12%	17%	12809	12%	16%
1,2	16060	13%	17%	12761	12%	16%
1,5	16014	13%	17%	12738	12%	16%
1,8	15989	13%	17%	12740	12%	16%
2	15540	13%	17%	12731	12%	16%
2,2	15423	13%	16%	12664	12%	16%
5	14748	13%	16%	12309	12%	16%
7	14582	13%	16%	12197	12%	16%
10	14443	13%	16%	12094	12%	16%
20	14232	13%	16%	11931	12%	16%
30	14069	13%	16%	11808	12%	16%
40	13952	13%	16%	11704	12%	16%
50	13851	13%	16%	11620	12%	16%
100	13563	13%	16%	11399	13%	16%
500	13052	13%	16%	10980	13%	15%
1.000	13014	13%	16%	10949	13%	15%
1.440	13013	13%	16%	10948	13%	15%

In a traffic pattern given by the OVG<sup>14</sup> from 1996 (Centraal Bureau voor de Statistiek, 1997), 15,2% of the trips are made in the morning peak and 21,6% in the afternoon peak. The fractions derived from the FCD are relatively low compared to this.

### Sensitivity of the parameter break

The parameter *real stop* was fixed to 2 minutes while the parameter *break* was changed for each calculation. Seventeen calculations were performed with the value of *break* ranging from 1 minute to 1.440 minutes (24hours). Figure 6.6 shows a comparison between the trip distributions for different values of the parameter *break* for the data collected on May 9. Figure 6.6 shows the same comparison for May 10.

Figure 6.6: The trip distribution of May 9, with different values for the parameter *break*

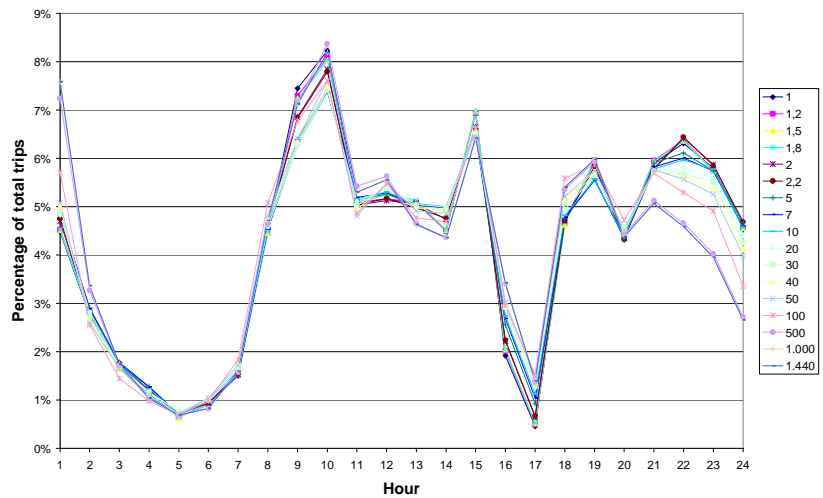


Figure 6.7: The trip distribution of May 10, with different values for the parameter *break*

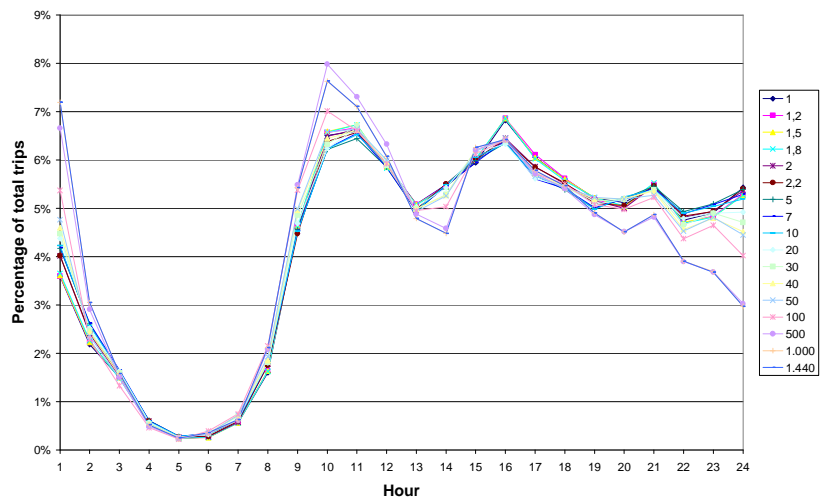


Figure 6.8 shows how the number of trips on May 9 changes with different values of the parameter *break*, both including and excluding intrazonal trips. Figure 6.9 shows the same for May 10.

<sup>14</sup> OVG is the Dutch Travel Diary



Figure 6.8: The change in number of total trips on May 9 when the parameter *break* is varied

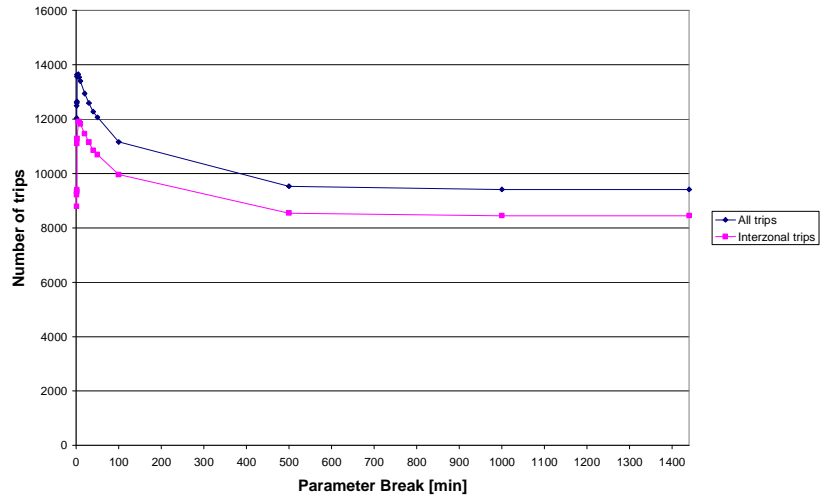
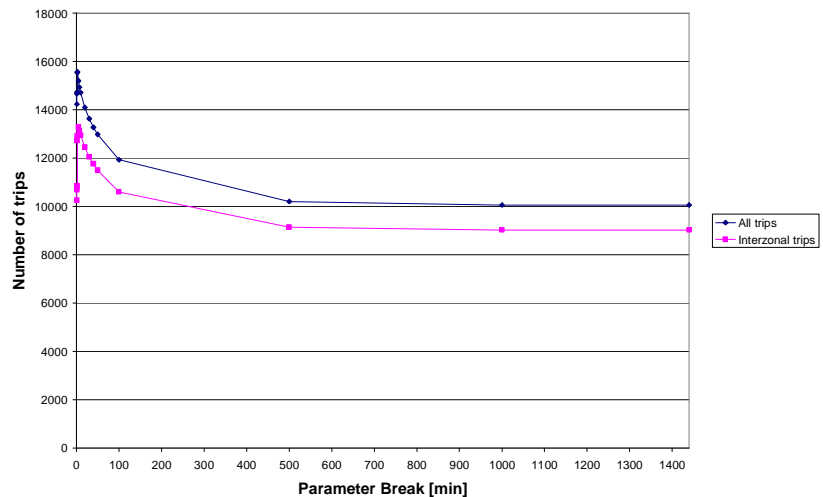


Figure 6.9: The change in number of total trips on May 10 when the parameter *break* is varied



From Figures 6.6-6.9 it can be concluded that variation of the parameter *break* will change the structure of the trip distribution only slightly (though a bit more than when changing the parameter *real stop*) while the change in number of trips is quite substantial. Interestingly, the number of trips increases in the beginning when the parameter *break* is rather low. This is due to the definition that a measurement is neither an O nor a D if there is a *break* or a *real stop* both before and after it. Hence, when the *break* parameter is small, several Os and Ds are cancelled out. The number of trips becomes more or less stable when the parameter is set as 500 minutes and larger, and it becomes completely stable when the parameter approaches the size of 24 hours.

Table 6.3 and Table 6.4 show the number of trips (both when intrazonal trips are included and for interzonal trips only) for the different value of *break* as well as the percentage of the trips that occur in the morning and afternoon peak (7:00-10:00 and 16:00-19:00).

Table 6.3: The change in the number and percentage of trips in peak periods on May 9, caused by variation of the parameter *break*

Real stop = 2 minutes – May 9, 2007						
Break [min]	All trips			Interzonal trips only		
	Trips	Morning peak	Afternoon peak	Trips	Morning peak	Afternoon peak
1	12043	20%	11%	8787	18%	10%
1,2	12497	20%	11%	9219	18%	10%
1,5	12602	20%	11%	9333	18%	10%
1,8	12647	20%	11%	9399	18%	10%
2	13561	19%	11%	11100	18%	11%
2,2	13633	19%	11%	11279	18%	11%
5	13647	18%	11%	11902	18%	11%
7	13545	18%	11%	11911	18%	11%
10	13402	18%	12%	11820	18%	11%
20	12944	18%	12%	11464	18%	11%
30	12594	18%	12%	11149	18%	12%
40	12273	19%	11%	8787	18%	10%
50	12069	19%	11%	9219	18%	10%
100	11159	19%	11%	9333	18%	10%
500	9532	20%	11%	9399	18%	10%
1.000	9414	20%	11%	11100	18%	11%
10.000	9412	20%	11%	11279	18%	11%

Table 6.4 The change in the number and percentage of trips in peak periods on May 10, caused by variation of the parameter *break*

Real stop = 2 minutes May – 10, 2007						
Break [min]	Intrazonal trips included			Interzonal trips only		
	Trips	Morning peak	Afternoon peak	Trips	Morning peak	Afternoon peak
1	14233	13%	17%	10266	12%	16%
1,2	14660	13%	17%	10689	12%	16%
1,5	14700	13%	17%	10756	12%	16%
1,8	14749	13%	17%	10845	12%	16%
2	15540	13%	17%	12731	12%	16%
2,2	15564	13%	16%	12921	12%	16%
5	15198	13%	16%	13282	12%	16%
7	14936	13%	16%	13135	12%	16%
10	14713	13%	16%	12952	12%	16%
20	14098	13%	16%	12452	12%	16%
30	13634	13%	16%	12051	13%	16%
40	13276	13%	16%	11755	13%	16%
50	12982	14%	16%	11498	13%	16%
100	11933	15%	16%	10603	14%	16%
500	10208	16%	16%	9135	15%	16%
1.000	10060	15%	16%	9024	15%	16%
10.000	10059	15%	16%	9023	15%	16%

The fractions in the tables above, derived from the FCD, are in most cases relatively low compared to the previously mentioned fractions given by the OVG from 1996.

---

### 6.2.1. Conclusions

The sensitivity analysis of the parameters *real stop* and *break* shows that the value of both parameters does not have a great effect on the trip distribution. They, however, affect the total number of trips considerably.

When an a-priori matrix is constructed, it can be scaled up to match the total amount of traffic. Hence, the trip distribution is of more importance than the total number of trips. The route choice analysis is based on the location of the vehicles and the values of the parameters do not have any effect on that. The values of these parameters have the largest impact on the TLD.

Since the main topics of this thesis are the a-priori matrix estimation and the route choice analysis, it is concluded that the values of the parameters should be kept as they were in Section 4.1.1, both at 2 minutes.

In order to eliminate all doubts about when trips are really beginning and ending, additional information from the FCD is required, i.e. information about the vehicle's occupancy. That would considerably increase the reliability of the information derived from the FCD.

## 6.3 The driving behaviour of taxis

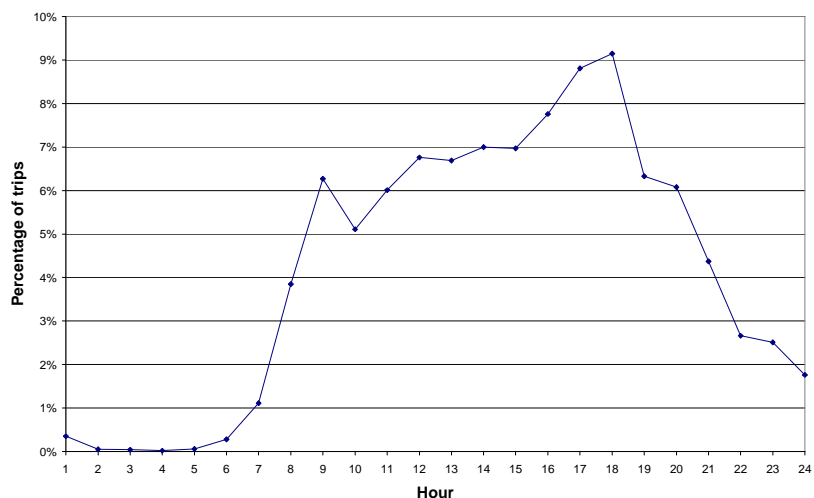
One of the two important questions that were raised in the beginning of this thesis is whether data from taxis are representative for the whole traffic. A large part of the traffic in China consists of taxis, yet it is highly likely that their driving behaviour differs from the driving behaviour of the common driver.

An example of a "typical" trip distribution<sup>15</sup> can be seen in Figure 6.10, while the trip distributions derived from the FCD of 4 different days (May 9 and 10, June 6 and 7) can be seen in Figure 6.11. Clearly there is a considerable difference between the "typical" distribution and the FCD distribution. However, from Figure 6.11 it can be seen that the FCD trip distribution is rather consistent between days. The distribution of May 9 differs from the other distributions in the afternoon period, but apart from that the difference between those days is not large. From this it can be concluded that these distributions are "typical" taxi trip distributions and therefore outcomes from the FCD analyses should be consistent between days.

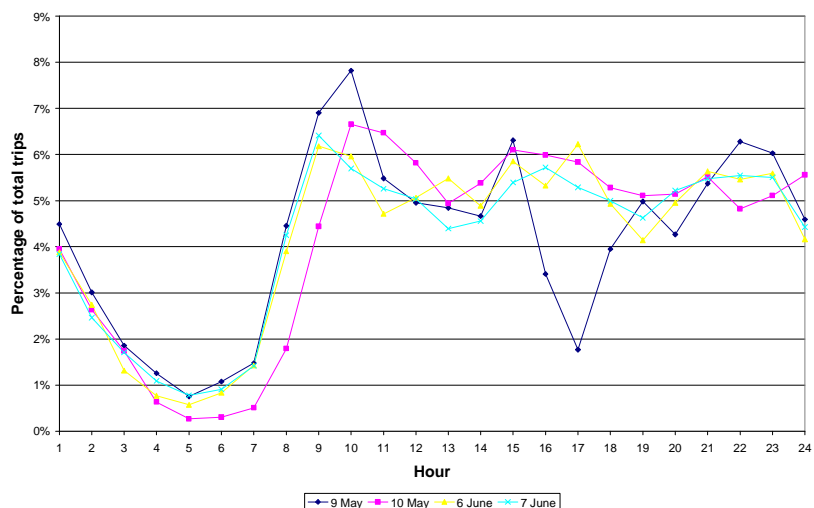
---

<sup>15</sup> Given by the OVG from 1996 (Centraal Bureau voor de Statistiek, 1997)

.....  
 Figure 6.10: A "typical" trip distribution  
 in the Netherlands  
 Given by the OVG of 1996



.....  
 Figure 6.11: Traffic pattern from the  
 FCD of 4 different days



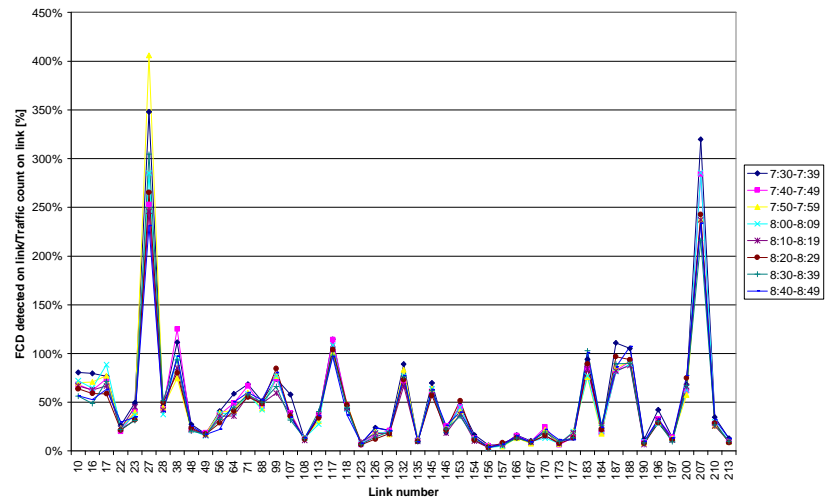
The "typical" trip distribution above comes from a Dutch survey done by the CBS<sup>16</sup>, thus it might indeed differ from the situation in the study area. If a similar kind of survey were done for the study area and a "typical" trip distribution constructed, it would perhaps be possible to use that trip distribution to scale the FCD so they would match the total traffic.

Another thing that can be examined regarding the driving behaviour of taxis is whether their distribution over the network is consistent over time. For the FCD collected on May 9, the taxis on the links were counted and compared with the traffic counts. Figure 6.12 shows this comparison, i.e. the ratio between taxis and traffic counts for 10-minute time slices on the 47 links that have traffic counts. According to this, the distribution of taxis over the network is consistent over time.

<sup>16</sup> Centraal Bureau voor de Statistiek

However, the ratio varies quite considerably, from 3% to 406%. That indicates that taxis are taking different routes than the other vehicles. There are two links where the amount of taxis is fairly larger than the volumes given by the traffic counts. This might be caused by congestion on those links. Under those circumstances the vehicles move slowly over the link (hence, no *break* or *real stop* occurs), which can result in the repetition of the link within the path and thus, an overestimation of the taxis on that link. The important thing is though that over the considered time period (which is a peak period) this is constant.

Figure 6.12: The ratio between the number of taxis on links and traffic counts on links for 10 minutes time slices in the morning of May 9, 2007



## 6.4 Estimations made with the FCD

In this section the three different estimations made with the FCD are discussed, i.e. the a-priori matrix estimation, the route choice analysis and the TLD.

### 6.4.1. A-priori matrix estimation

In Section 4.1.2, three different ways of building an a-priori matrix from the FCD were discussed:

- OFCD: The original FCD OD matrix
- PFCD: The original FCD OD matrix with the value 1 inserted where values are missing
- CFCD: The original FCD OD matrix with scale and missing measurements adjusted by using another matrix from the study area

In the before mentioned section, some doubts were expressed about the first two ways and hence the third one was suggested. Here, the complete matrices, used for the construction of CFCD, are estimated with REMDOE (Chen, 1993). In this section, the effects that these different a-priori matrices have on the OD matrix estimation are

examined. Furthermore, the effects from a unit matrix, UOD, and another a-priori matrix, TC, are tested.

Table 6.5 shows how many iterations NEST performs before convergence is reached between the flows in the estimated OD matrix and the counted flows (with maximum relative error set as 0,25). In the case of the OFCD, the estimation nearly always stops due to non-converging iterations or because the a-priori matrix gives 0 volumes on counted links. Thus, it is clear that the information from the FCD alone is not sufficient to build a good a-priori matrix. The other estimations are all completed.

.....  
Table 6.5: The number of iterations necessary for the OD matrix estimation of each time slice with different a-priori matrices

Time Slice	UOD	OFCD	PFCD	CFCD	TC
7:30-7:39	16	Stopped	16	15	15
7:40-7:49	13	Stopped	9	7	5
7:50-7:59	5	Stopped	5	5	5
8:00-8:09	6	Stopped	7	7	8
8:10-8:19	7	Stopped	7	8	8
8:20-8:29	7	Stopped	7	6	8
8:30-8:39	19	19	14	18	17
8:40-8:49	3	Stopped	3	3	4
<b>Total</b> 7:30-8:49	9	12	9	9	10

The iteration speed does not give an indication about the quality of the a-priori matrix. Hence, the number of trips was also considered. The total number of trips in both the a-priori matrices and the estimated OD matrices of all the time slices can be seen in Table 6.6.

.....  
Table 6.6: The number of trips in both the a-priori matrices and the estimated OD matrices

Time Slice	Number of trips									
	UOD		OFCD		PFCD		CFCD		TC	
	A-priori	OD	A-priori	OD	A-priori	OD	A-priori	OD	A-priori	OD
7:30-7:39	306	5.056	107	X	338	5.172	5.086	5.453	7.192	5.234
7:40-7:49	306	5.192	111	X	341	5.186	6.514	5.491	7.874	5.578
7:50-7:59	306	5.602	119	X	354	5.615	8.328	5.607	7.990	5.518
8:00-8:09	306	5.610	117	X	351	5.679	7.125	5.934	8.181	5.760
8:10-8:19	306	6.252	119	X	349	6.300	8.129	7.001	8.476	6.471
8:20-8:29	306	5.908	128	X	364	5.841	6.889	5.763	8.461	5.996
8:30-8:39	306	5.538	116	5.516	352	5.528	8.135	5.536	8.209	5.183
8:40-8:49	306	5.702	88	X	334	6.045	10.204	6.319	7.727	5.875
<b>Total</b> 7:30-8:49	306	27.764	905	28.184	2.783	28.036	60.411	28.315	64.110	26.953

When the numbers in the table above are examined, it is clear that the different a-priori matrices do not influence the number of estimated trips. That is due to the fact that NEST normalizes the a-priori matrices. Thus, the structures of the 5 different estimated OD matrices were examined. The estimated matrices resulting from using CFCD were used as a base for the comparison, i.e. they were compared with the structure of all the other matrices.

In order to compare the structure of two matrices the following equation is used.

$$abs\left(\frac{T_{ij}^{A\%} - T_{ij}^{B\%}}{T_{ij}^{A\%} + T_{ij}^{B\%}} \cdot 2\right) \quad (6.1)$$

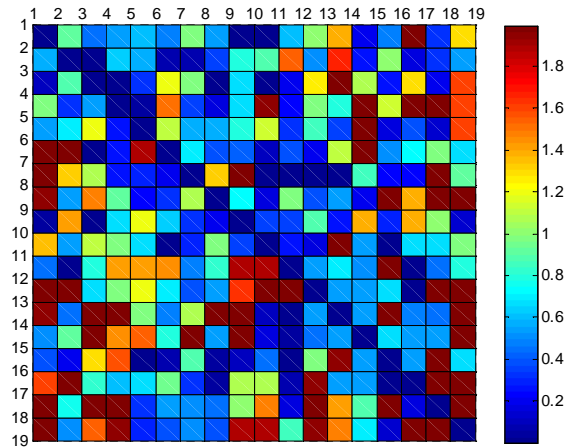
In order to get  $T_{ij}^{A\%}$  and  $T_{ij}^{B\%}$ , each value in both matrices is divided with the matrix's total number of trips, i.e. the two matrices are normalized. Then the difference between each corresponding value within the two matrices is divided with the average of both values. By applying this equation, large differences between two values approach the value 2. If the difference is e.g. 3 fold, the equation gives the value 1. Table 6.7 shows which value equation (6.1) gives for different proportional differences between two corresponding matrix values.

.....  
 Table 6.7: The values from equation (6.1) for different proportional differences

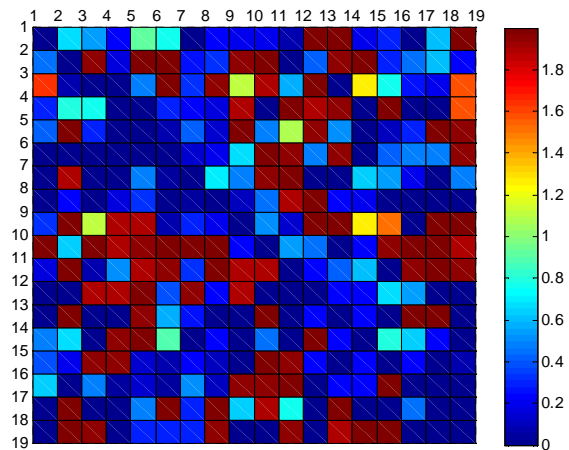
Proportional difference between corresponding cells, if $T_{ij}^{A\%} > T_{ij}^{B\%}$ : $(T_{ij}^{A\%} / T_{ij}^{B\%})$ if $T_{ij}^{B\%} > T_{ij}^{A\%}$ : $(T_{ij}^{B\%} / T_{ij}^{A\%})$	Value from equation (6.1)
1	0,00
2	0,67
3	1,00
4	1,20
5	1,33
10	1,64
15	1,75
20	1,81
30	1,87
40	1,90
50	1,92
60	1,93
70	1,94
80	1,95
90	1,96
100	1,96
300	1,99
500	1,99
1000	2,00

Figures 6.13-6.16 show the comparisons of the estimated OD matrices of the seventh time slice, 8:30-8:39. The squares in the figures represent an OD pair and the colours represent the value calculated with equation (6.1). The rest of the comparison figures as well as comparison plots, where the values of each matrix cells are plotted using two axes for the two matrices, can be seen in Appendix 6 – Section 9.6.

.....  
 ..  
 Figure 6.13: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39



.....  
 ..  
 Figure 6.14: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39



.....  
 ..  
 Figure 6.15: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39

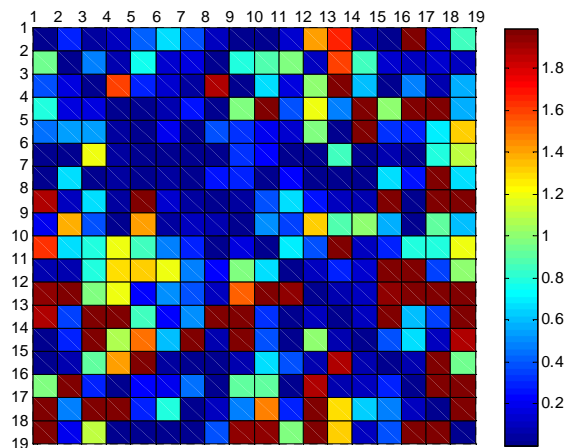
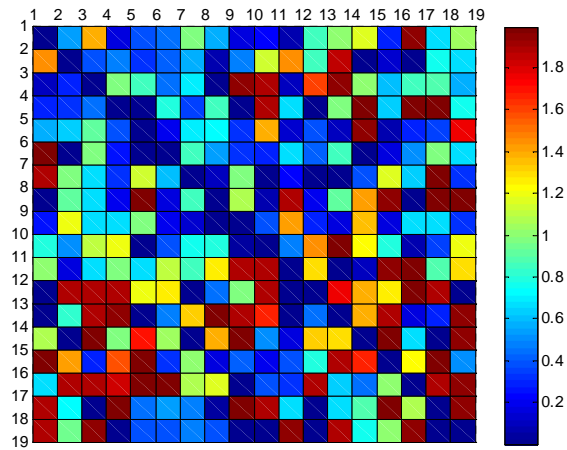




Figure 6.16: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39



In the figures above, blue colour indicates a small structural difference between the cells of the two matrices while red colour indicates a large difference. When the figures are examined it can be seen that the different a-priori matrices do indeed influence the structure of the estimated OD matrix.

From this it can be concluded that when more information is included in the a-priori matrix it must lead to a better OD matrix estimation. Since the FCD alone are not enough to build a useable a-priori matrix, combining it with e.g. historical data or survey data should lead to a better result. Thus, the CFCD are best suited as a-priori matrices.

#### 6.4.2. Route choice analysis

The numbers of trips detected on the 4 days considered are the following:

- May 9: 13.561 trips
- May 10: 15.540 trips
- June 6: 4.427 trips
- June 7: 12.769 trips

Table 6.8 shows the number of trips detected during each time slice for those 4 days, both the total number of trips and the interzonal trips only.

Table 6.8: The number of detected trips in the morning of 4 days

Time slice	May 9		May 10		June 6		June 7	
	Total	Inter-zonal	Total	Inter-zonal	Total	Inter-zonal	Total	Inter-zonal
7:30-7:39	124	107	47	41	39	34	92	77
7:40-7:49	142	111	59	52	44	39	97	83
7:50-7:59	140	119	64	56	33	29	147	120
8:00-8:09	158	117	72	60	45	34	128	103
8:10-8:19	157	119	81	59	58	47	120	88
8:20-8:29	184	128	69	45	42	26	151	98
8:30-8:39	166	116	107	79	43	29	142	104
8:40-8:49	136	88	178	127	46	41	158	119

For all the OD pairs detected in the FCD within each time slice, the route choice was analysed as described in Section 4.1.3. After the shortest paths have been inserted for the missing OD pairs and the value of  $p_{ij}^a$  has been calculated, the routes can be used directly as an input for NEST.

For a proper dynamic mapping of the OD flows into the network, relatively many OD pairs need to have detected routes and preferably more than one. Table 6.9 shows, for the 4 days, the proportion of OD pairs that have detected routes during each time period.

.....  
Table 6.9: The percentage of OD pairs with detected routes per time slice

Time slice	Percentage of OD pairs with detected routes			
	May 9	May 10	June 6	June 7
7:30-7:39	24,5%	10,5%	9,5%	17,0%
7:40-7:49	24,8%	14,4%	9,8%	18,3%
7:50-7:59	23,2%	13,4%	7,5%	21,6%
8:00-8:09	23,5%	15,0%	10,1%	20,9%
8:10-8:19	24,8%	13,4%	9,8%	15,0%
8:20-8:29	22,9%	12,7%	7,8%	19,6%
8:30-8:39	22,9%	19,0%	7,2%	19,6%
8:40-8:49	19,6%	26,1%	8,2%	22,5%
<b>Total 7:30-8:49</b>	<b>66,3%</b>	<b>54,2%</b>	<b>33,7%</b>	<b>52,3%</b>

The highest percentages in the table above are just around 25%, which means that minimum 75% of OD flows are mapped into the network with calculated shortest paths.

Table 6.10 shows which percentages of OD pairs in the morning of May 9 have a certain amount of detected routes. The OD pairs with no detected routes are left out of the calculations. For all time slices, the majority of OD pairs with detected routes has only 1 detected route.

.....  
Table 6.10: The proportion of OD pairs with a certain amount of detected trips on May 9

Time slice	Number of detected routes per OD pair							
	1	2	3	4	5	6	7	8
7:30-7:39	74,7%	17,3%	4,0%	0,0%	2,7%	1,3%	0,0%	0,0%
7:40-7:49	69,7%	14,5%	15,8%	0,0%	0,0%	0,0%	0,0%	0,0%
7:50-7:59	62,0%	21,1%	9,9%	4,2%	0,0%	2,8%	0,0%	0,0%
8:00-8:09	69,4%	13,9%	6,9%	5,6%	2,8%	1,4%	0,0%	0,0%
8:10-8:19	69,7%	14,5%	9,2%	5,3%	0,0%	0,0%	1,3%	0,0%
8:20-8:29	62,9%	15,7%	8,6%	8,6%	1,4%	0,0%	1,4%	1,4%
8:30-8:39	70,0%	12,9%	8,6%	2,9%	2,9%	1,4%	1,4%	0,0%
8:40-8:49	73,3%	18,3%	1,7%	3,3%	1,7%	1,7%	0,0%	0,0%
<b>Total 7:30-8:49</b>	<b>38,2%</b>	<b>18,5%</b>	<b>16,3%</b>	<b>10,7%</b>	<b>3,4%</b>	<b>4,5%</b>	<b>6,7%</b>	<b>1,7%</b>

From this discussion, it can be concluded that the FCD alone do not comprise enough information to analyse the route choice. In this thesis, the calculated shortest paths are used to compensate for the lack of information.

### 6.4.3. Trip length distribution

In Section 4.1.4 two possible ways to get the TLD were mentioned: directly from the detected trips within the FCD and from the estimated OD matrices, using the calculated shortest paths. Following is a discussion concerning these two methods.

#### TLD obtained directly from FCD

Figure 6.17 shows the TLD of the FCD from May 9 and 10 as well as a “typical” TLD given by the OVG from 1996 (Centraal Bureau voor de Statistiek, 1997). Figure 6.18 shows a close-up of the TLD for all trips shorter than 50 km. Table 6.11 shows the numerical values that are plotted in the figures. The shape of the TLD is very consistent between the two days but it differs considerably from the OVG TLD. The proportion of short trips is much higher in the FCD than given by the OVG. This difference can be understood when considering the study area. The size of the study area is only about 6 km<sup>2</sup> and a trip is assumed to begin or end when a vehicle leaves or enters the study area. Furthermore, the longest computed shortest trip is only 3,5 km (between zone 9 and zone 4). It is therefore logical that within this area, short trips are overestimated. In order to get a TLD that resembles the one given by OVG, the study area needs to be enlarged.

.....  
Figure 6.17: The trip length distribution for all the trips of May 9 and 10 and the trip length distribution given by the OVG from 1996

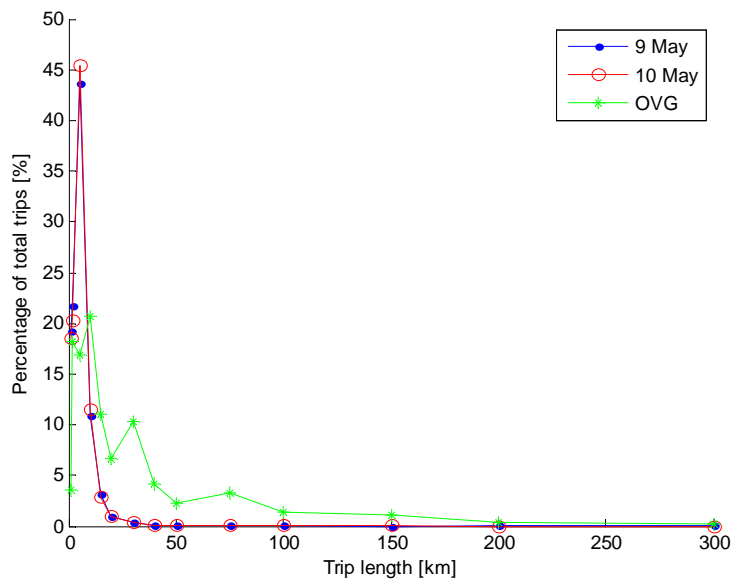


Figure 6.18: The trip length distribution, of all trips shorter than 50 km, for all the detected trips of May 9 May 10 and the trip length distribution given by the OVG from 1996

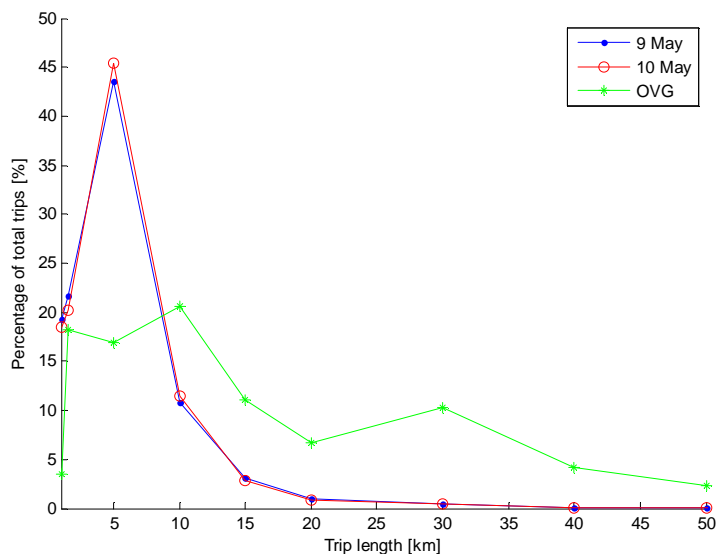


Table 6.11: The trip length distribution of May 9, May 10 and given by the OVG from 1996

Trip length [km]	May 9	May 10	OVG
0-1	19,28%	18,54%	3,55%
1-1,5	21,69%	20,19%	18,18%
1,5-5	43,63%	45,46%	16,89%
5-10	10,87%	11,51%	20,66%
10-15	3,08%	2,86%	11,04%
15-20	0,93%	0,92%	6,65%
20-30	0,43%	0,41%	10,26%
30-40	0,05%	0,04%	4,12%
40-50	0,01%	0,02%	2,31%
50-75	0,01%	0,02%	3,26%
75-100	0,01%	0,01%	1,39%
100-150	0,00%	0,01%	1,14%
150-200	0,01%	0,00%	0,39%
200-300	0,01%	0,00%	0,17%

### TLD calculated from estimated OD matrices

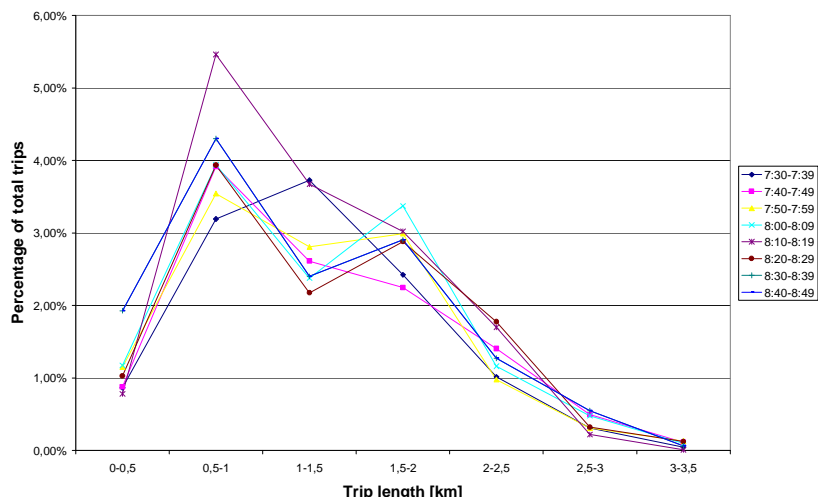
When the TLD is calculated from the estimated OD matrix, the same problem exists as before. Since the calculated shortest paths are used to estimate the TLD, the trips in the TLD can maximum equal the longest calculated shortest path, which here is only 3,5 km. The same applies here as in the former method; if the study area would be bigger, the TLD would most probably resemble the OVG data from 1996 (Centraal Bureau voor de Statistiek, 1997) more closely.

Table 6.12 and Figure 6.19 show the TLD calculated for the estimated OD matrix of May 9. There is some consistency between the patterns of the different time slices but this information is probably not enough to be of any use.

Table 6.12: The trip length distributions of each time slice calculated from the estimated OD matrices of May 9

Trip length [km]	7:30-7:39	7:40-7:49	7:50-7:59	8:00-8:09	8:10-8:19	8:20-8:29	8:30-8:39	8:40-8:49
0-0,5	0,86%	0,88%	1,15%	1,17%	0,78%	1,03%	1,54%	1,93%
0,5-1	3,19%	3,92%	3,54%	3,94%	5,46%	3,94%	3,50%	4,30%
1-1,5	3,73%	2,61%	2,81%	2,38%	3,67%	2,17%	2,20%	2,40%
1,5-2	2,42%	2,25%	2,99%	3,37%	3,02%	2,88%	2,65%	2,90%
2-2,5	1,01%	1,40%	0,98%	1,16%	1,70%	1,77%	1,33%	1,27%
2,5-3	0,31%	0,49%	0,31%	0,48%	0,22%	0,32%	0,49%	0,55%
3-3,5	0,04%	0,11%	0,13%	0,10%	0,01%	0,12%	0,04%	0,06%

Figure 6.19: The trip length distributions of each time slice calculated from the estimated OD matrices of May 9



#### 6.4.4. Conclusions

From Figures 6.12-6.15 it can be seen that different a-priori matrices do indeed influence the structure of the estimated OD matrices. Hence, it can be concluded that the more information the a-priori matrices comprise, the better the estimation will be. Since the FCD alone is not enough to build a useable a-priori matrix, combining it with another source of information like for instance historical data or survey data should give a good result. In this thesis, neither historical data nor survey data exist. Thus, OD matrices estimated for the same time periods with REMDOE are used.

When the route choices are analysed, it appears that when the time periods are 10 minute long, only a small fraction (at best 25%) of the OD pairs have a detected route, and most of them only one. The missing routes thus need to be compensated. This is done by assigning the calculated shortest paths to the OD pairs with no detected routes.

For both methods of estimating the TLD, the short trips are most likely overestimated. That is due to the limited size of the study area. In order to get a better TLD, the study area needs to be enlarged. With a better TLD it would be possible to scale the a-priori matrices or the estimated OD matrices so that they fit a "typical" TLD, and hence the whole traffic.

## 6.5 The complete OD matrix estimation method

OD matrix estimations were made using the program NEST, with traffic counts, FCD route choice and different a-priori matrices as input. Table 6.6 displayed the estimated flows. If this information is examined, one might notice a striking difference between the summed number of estimated trips of each time slice and the number of estimated trips for the whole time period. Table 6.13 shows these numbers. In all cases, the summed number of estimated trips is about 60% larger than the estimated trips of the whole time period. This large difference is caused by the fact that NEST performs static OD matrix estimations. In static estimation methods, it is assumed that all trips that start within a time period also end within that same time period. Thus, when the estimation is done in small consecutive time steps (as in this case) there is a risk of counting the same trips more than once and thus overestimating the number of trips. This is a common problem of all static OD matrix estimation methods. One way to eliminate this problem would be to extend the network to a STEN network. That is however not done here.

.....  
Table 6.13: The number of summed and total trips for the whole period

	UOD	PFCD	CFCD	TC
<b>Summed from all time slices</b>	44.860	45.366	47.104	45.615
<b>Estimated for whole period</b>	27.764	28.036	28.315	26.953

When the trip lengths from the FCD of May 9 are examined it is possible to calculate the chance that a trip exceeds the time frame of 10 minutes. The measured trip lengths from the FCD are rounded to the nearest integer and counted. It is assumed that the departure times are uniformly distributed and using that assumption the chance that a trip of a certain length exceeds the time frame is calculated. The chance that the measured trips exceed the time frame is then the product of the percentage of total trips and the previously mentioned chance. Table 6.14 shows these calculations. The total chance that a trip exceeds the time frame is, according to these calculations, 40%. The reason why this is lower than 60% might be due to the fact that 16% of the measured trips are longer than 20 minutes (two time periods) and those are thus counted more than twice.

Table 6.14: The calculations of the chance that the trips detected within the FCD exceed the time frame of 10 minutes

Trip time [min]	Number of trips	Percentage of total trips	Chance that trip of this length exceeds the time frame	Chance that measured trips exceed the time frame
0	114	1%	0%	0,0%
1	2.911	27%	10%	2,7%
2	2.280	21%	20%	4,2%
3	1.316	12%	30%	3,7%
4	837	8%	40%	3,1%
5	522	5%	50%	2,4%
6	328	3%	60%	1,8%
7	214	2%	70%	1,4%
8	164	2%	80%	1,2%
9	89	1%	90%	0,7%
10	70	1%	100%	0,6%
>10	1.971	18%	100%	18,2%
<b>Total</b>	<b>10.816</b>	<b>100%</b>		<b>40,1%</b>

## 6.6 Comparison with another method

Due to the fact that the “true” OD matrix is not known, it is very hard to say whether estimated OD matrices are good or not. They can however be compared with matrices estimated with another method.

The comparison method is the before mentioned dynamic OD matrix estimation method REMDOE and its input data are traffic counts and the TC a-priori matrices.

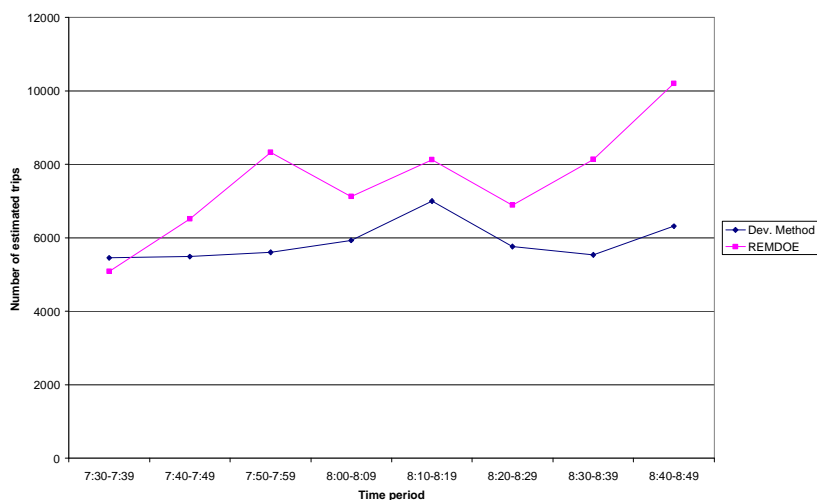
The OD matrices that are compared with the REMDOE OD matrices are estimated with the developed OD matrix estimation method, using the traffic counts, the CFCD a-priori matrices and the FCD route choices combined with calculated shortest paths as input.

Table 6.15 shows the number of trips estimated for all the time slices using the two different methods. The information of Table 6.15 is plotted in Figure 6.20. REMDOE nearly always estimates a higher number of trips but the patterns in Figure 6.20 do share some similarities.

Table 6.15: The number of estimated trips per time slice from the two different methods

Time slice	Number of estimated trips	
	Developed method	REMDOE
7:30-7:39	5.453	5.086
7:40-7:49	5.491	6.514
7:50-7:59	5.607	8.328
8:00-8:09	5.934	7.125
8:10-8:19	7.001	8.129
8:20-8:29	5.763	6.889
8:30-8:39	5.536	8.135
8:40-8:49	6.319	10.204
Total 7:30-8:49	5.453	5.086

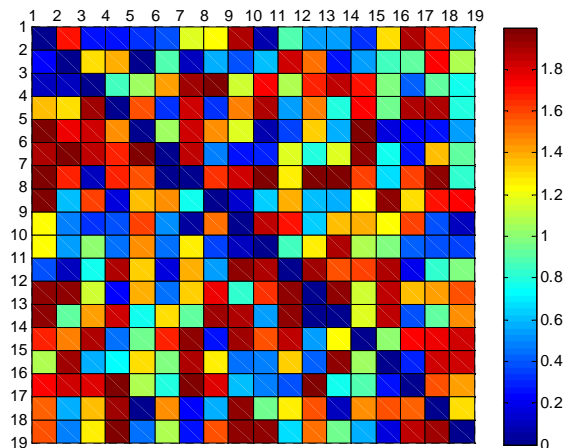
Figure 6.20: The number of estimated trips from the two different methods shown graphically



In order to compare the structure of the estimated OD matrices, equation 6.1 was used in the same way as in Section 6.4.1. Figure 6.21 shows the difference of the two matrices (the figures from the other time slices as well as comparison plots can be found in Appendix 7 – Section 9.7). Judging from those figures, there is not much similarity between the structures of the OD matrices estimated with these two different methods. However, as was mentioned before, since the “true” matrix is not known, it is impossible to tell which method gives better results.



Figure 6.21: Comparison between OD matrix estimated in this report and OD matrix estimated with REMDOE. Time period 8:30-8:39



## 6.7 Summary and answers to questions

In the beginning of this thesis the following two questions were raised:

*Are data coming from taxis representative for the whole traffic, and if not, can the bias be estimated and adjusted?*

*Do FCD comprise enough information to build a good a-priori matrix and do they give sufficient information about the route choice and the TLD?*

In this chapter, those questions are answered. Hereafter the answers are summarized.

When the distribution of the trips detected within the FCD is compared with a "typical" trip distribution, it is evident that they differ quite considerably. The trip distribution of the FCD is however rather consistent between days and thus, outcomes from FCD analysis should be consistent as well. If the "typical" trip distribution for the study area would be known, it might be used to scale the outcomes of the FCD analyses or the estimated OD matrix. Another way to scale these outcomes would be to match the TLDs to a "typical" TLD, but in order to be able to do that, the study area needs to be of a considerable size.

Concluding, the answer to the former question is:

*No, data from taxis is not representative for the whole traffic, but the bias can be adjusted for.*

---

Five different a-priori matrices were tested as an input for NEST. Those matrices were: UOD, OFCD, PFCD, CFCD and TC.

All the estimations resulted in estimated OD matrices except the one that used OFCD, i.e. the original matrices derived from the FCD. Thus, it is clear that the information included in FCD alone is not enough to construct a good a-priori matrix. The structure of the successfully estimated OD matrices differs when different a-priori matrices are used. It is concluded that the more information comprised in the a-priori matrices the better the estimation must be.

When 4 days are examined, at most 25% of the OD pairs have a detected route during a time period of 10 minutes. This means that at least 75% of the OD flows need to be mapped in another way. In this thesis this is done with calculated shortest path.

The TLD was done in two ways. In both cases, the conclusion was that the study area was too small to get a good TLD.

The answer to the latter question is thus:

*No, FCD do neither comprise enough information to build a good a-priori matrix nor analyse the route choice. Other information sources, for instance historical data or survey data for the a-priori matrix estimation and calculated shortest path for the route choice analysis, need to be combined with the FCD. As mentioned before, in order to get a good TLD, the study area needs to be relatively large.*

Apart from answering the two questions above, this chapter included a sensitivity analysis that confirmed that the values of the parameters *real stop* and *break* should be 2 minutes.

Furthermore, the difference in number of estimated trips when estimating in short consecutive time slices and estimating for the whole time period at once was discussed. Since NEST performs a static OD matrix estimation there is a risk of an overestimation of trips when the estimation is done in time slices. In this case, the trips are overestimated by around 60%.

In the last section of this chapter, the OD matrices estimated with the developed method (using the CFCD matrices for a-priori matrices, the combination of FCD routes and calculated shortest paths for the mapping and the traffic counts) were compared with OD matrices estimated with REMDOE (using TC for a-priori matrices, DTA for mapping and the traffic counts). The differences were considerable, but there is unfortunately no way to tell which ones are better, since the "true" matrices are not known.

---

## 7. Conclusions and suggestions

---

The main goal of this thesis work was to develop a new method that uses FCD to estimate OD matrices that can be used as a-priori matrices for OD matrix estimation. The purpose is to increase the quality of the input for OD matrix estimation and hence the estimation itself. Furthermore, methods that analyze route choices and TLD from FCD were developed and a complete OD matrix estimation method that uses the FCD a-priori matrices and route choices was developed.

In this chapter, those goals are returned to in a summary of the research findings and some practical recommendations for further research are suggested.

### 7.1 Research findings

The research findings of this thesis will be summarized in the following sections.

**7.1.1. Rules for determining origins and destinations within FCD**  
Based on a few assumptions, rules were created in order to define the measurements within the FCD that correspond to Os or Ds. Those rules are the following:

**Rule 1**

*A stop is considered to be a real stop if the measured speed is 0km/h for 2 minutes or more. Stops that last less than 2 minutes are considered to be intermediate stops. Thus, the last measurement before a real stop is an O and the first measurement after a real stop is a D.*

**Rule 2**

*When the time between two measurements exceeds 2 minutes it can be assumed that the driver has taken a break. Consequently, the last measurement before the break is a D and the first measurement after the break is an O.*

**Rule 3**

*The first measurement with speed larger than 0 km/h that is detected from a vehicle after it enters the study area is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle before it leaves the study area is a D. A vehicle is defined to be outside the study area if it dwells there for 2 minutes or longer.*

---

#### **Rule 4**

*The first measurement with speed larger than 0 km/h that is detected from a vehicle is an O. The last measurement with speed larger than 0 km/h that is detected from a vehicle is a D.*

By dismissing all measurements with no assigned links, rules 2 and 3 were combined. These rules were confirmed with a sensitivity analysis.

#### **7.1.2. A-priori matrices estimated from FCD**

A-priori matrices that are derived from the FCD for 10 minute time periods are by nature considerably sparse and their volumes are low. That is due to the fact that the data come from taxis, which are only a fraction of the whole traffic.

When those original OD matrices are used as a-priori matrices the estimation procedure nearly always discontinues due to zero values in the a-priori matrix where traffic counts are given or because of non-converging calculations. It is thus clear that the FCD alone do not comprise enough information to serve for good a-priori matrix estimations. Hence, the a-priori matrices are constructed with a combination of FCD and historical data, survey data or matrices estimated with other methods. Traditionally, the a-priori matrices are based on those two sources alone, but by doing this, the information included in the a-priori matrix is optimized.

#### **7.1.3. Route choice analysis with FCD**

For the route choice analysis, the same applies as for the a-priori matrix estimation; the FCD alone do not comprise enough information. When the route choices are examined for 10 minute time period, only a maximum of 25% of the OD pairs have a detected route and the majority of those only one route. The missing routes thus need to be compensated in some way. This was done with computed shortest path.

#### **7.1.4. Trip length distribution analysis with FCD**

The TLD derived from the FCD is very consistent between days but due to the limited size of the study area, short trips are overestimated. The same applies for the TLD derived from estimated OD matrices.

In this thesis the TLD is only examined but not used for practical purposes. With improved TLD it might be possible to scale the estimated OD matrices so that they match the whole traffic.

#### **7.1.5. The complete OD matrix estimation method**

The problem of the suggested complete OD matrix estimation method is that when the static method is extended to a dynamic one (by doing the estimations in small time slices) the number of trips is overestimated by around 60%. This is caused by some trips starting in one time period

---

and ending in another, consequently those are counted more than once.

## 7.2 Conclusions

From this thesis work it can be concluded that it is surely feasible to use FCD for a-priori matrix estimation, route choice analysis and TLD analysis. However, since the FCD are only sample data, they by themselves are mostly not sufficient. The a-priori matrix estimation needs to be made with a combination of e.g. historical data or survey data (traditional a-priori matrices) and FCD. In that way the information in the a-priori matrix is maximized. Furthermore, when the routes are examined, only a fraction of the OD pairs has a detected route while the rest needs to be compensated in some way.

These facts do not mean that using FCD for the abovementioned purposes is pointless. The FCD is current data, and thus, combining it with traditional input must increase its quality and lead to better OD matrix estimations. Due to the lack of knowledge of the real OD matrices it is however not possible to say whether the methods developed in this thesis are reliable.

## 7.3 Suggestions for further research

In the following discussion a few suggestions are made for further research.

### Improving the FCD

Even though the rules for determining origins and destinations within the FCD were confirmed with a sensitivity analysis, it is certain that they do not give completely correct information about the real Os and Ds within the data. Thus, it is suggested that for further research, additional information should be added to the FCD. That information is regarding the occupancy of the taxis. With that information, there will be no doubt when a trip is beginning and ending. This would result in more reliable outcomes from the FCD analyses.

### Involving DTA

In the thesis it was shown that the information included in the FCD alone is neither enough for the a-priori matrix estimation nor the route choice analysis. The only solution to this, is equipping more vehicles as probe-vehicles. Besides, more diverse probe-vehicles would also improve the reliability of the data. In the future, this might become the case, but at the moment, matters such as privacy issues prevent it from happening. Hence, in order to compensate for the lack of information, measures like those described in this thesis need to be taken. In the case of the route choice analysis, calculated shortest paths were used for the mapping of OD flows with no detected routes. Thus, the mapping is done in a mixture of dynamic and static way. Since a large part of the OD pairs do not have a detected route, the major part of

---

the mapping is static. It is thus suggested that the missing routes should be estimated with DTA. Furthermore, when there is only one trip detected between two zones, that trip is used for the mapping, regardless of whether it is logical or not. It would probably give better results to include at least one estimated route as well.

#### **Enlarging the network**

The largest problem discovered in this thesis regarding the TDL, is that the study area is simply too small to get a good estimate. The TDL can possibly be used to scale the OD matrices so that they match the real traffic. It is suggested that, before further research in that field is performed, data from a larger area should be gleaned.

#### **Extending the network**

In order to avoid the overestimation of trips due to the small time slices used in the developed OD matrix estimation, the network can be extended to a STEN network. Making the time slices larger (e.g. 15-20 minutes) can also improve the estimations, but that would make the method less dynamic.

#### **Finding the true OD matrices**

Finally, in order to examine the reliability of the methods developed in this thesis, it is suggested that the real OD matrices should be discovered. One way to do so is to trace the traffic manually through the study area. Doing that might cost a lot of man-hours but that would provide a way to check the reliability of all the developed methods. This information can also be used for further development of the methods.

---

## 8. References

---

- Abrahamsson, T. (1998) *Estimation of Origin-Destination matrices using traffic counts – a literature survey*. Report no. IR-98-021. International Institute for Applied Systems Analysis
- Ashok, K., Ben-Akiva, M.E. (1993) Dynamic origin-destination matrix estimation for real-time traffic management systems. *Proc. 12<sup>th</sup> Int. Symposium on Transportation and Traffic Theory*, Berkeley, California
- Ban, X., Li, Y., Skabardonis, A. (2007) Local MAD method for probe vehicle data processing. *Proc. 14<sup>th</sup> world congress on intelligent transport systems*, Beijing, China
- Bell, M.G.H. (1991a) The estimation of origin-destination matrices by constrained generalised least squares, *Transportation Research B*, Vol. 25, No. 1, pp. 13-22
- Bell, M.G.H. (1991b) The Real Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion, *Transportation Research B*, Vol. 25, No. 2-3, pp. 115-125
- Brockfeld, E., Passfeld, B., Wagner, P. (2007) Validating travel times calculated on the basis of taxi floating car data with test drivers. *Proc. 14<sup>th</sup> world congress on intelligent transport systems*, Beijing, China
- Cascetta, E. (1984) Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator, *Transportation Research B*, Vol. 18, No. 4-5, pp. 289-299
- Cascetta, E., Inaudi, D., Marquis, G. (1993) Dynamic Estimators of Origin-Destination matrices using traffic counts. *Transportation Science* 27, pp. 363-373
- Cascetta, E., Postorino, M.N. (2001) Fixed Point Approaches to the Estimation of O-D Matrices Using Traffic Counts on Congested Networks. *Transportation Science* 35, pp. 134-147
- Centraal Bureau voor de Statistiek (1997) *De mobiliteit van de Nederlandse bevolking in 1996*, Voorburg/Heerlen
- Chen, Y.S. (1993-2007) *REMDOE – a dynamic OD estimator*, Software manual
- Chen, Y.S., Van Zuylen, H.J. et al. (2007) *Chengdu feasibility study for traffic information and route guidance*, Project report

---

Coëmet, M., De Hoog, A., Taale, H., Zijderhand, F. (1999) *Floating car data*, final report for roads to the future. Ministry of Transport, Public Works and Water Management (in Dutch)

Cremer, M., Keller, H. (1981) Dynamic identification of flows from traffic counts at complex intersections, *Proc. 8<sup>th</sup> Int. Symposium on Transportation and Traffic Theory*, University of Toronto Press, Toronto Canada

Cremer, M., Keller, H. (1984) A systems dynamics approach to the estimation of entry and exit O-D flows, *Proc. 9<sup>th</sup> International Symposium on Transportation and Traffic Theory*, Delft University of Technology, Delft The Netherlands

Cremer, M., Keller, H. (1987) A New Class of Dynamic Methods for the Identification of Origin-Destination Flows, *Transportation Research B*, Vol. 21, No. 2, pp. 117-132

Fisk, C.S. (1988) On combining maximum entropy trip matrix estimation with user optimal assignment, *Transportation Research B*, Vol. 40, No. 1, pp. 69-73

Hamerslag, R., Taale, H. (2001) Deriving road networks from floating car data, *9<sup>th</sup> world conference on transport research*, Seoul, Korea

Lin, P.-W., Chang, G.-L. (2007) A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix, *Transportation Research B*, Vol. 41, No. 5, pp. 554-572

Lo, H.-P., Chan, C.-P. (2003) Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts, *Transportation Research A*, Vol. 37, No. 9, pp. 771-788

Maher, M.J. (1983) Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, *Transportation Research B*, Vol. 17, No. 6, pp. 435-447

Nihan, N.L., Davis, G.A. (1987) Recursive Estimation of Origin-Destination Matrices from Input-Output counts, *Transportation Research B*, Vol. 21, No. 2, pp. 149-163

Nihan, N.L., Davis, G.A. (1989) Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts, *Transportation Science*, Vol. 23, No. 2

Sherali, H.D., Park, T. (2001) Estimation of dynamic origin-destination trip tables for a general network, *Transportation Research B*, Vol. 35, No. 3, pp. 217-235

Sherali, H.D., Narayanan, A., Sivanandan, R. (2003) Estimation of origin-destination trip-tables based on a partial set of traffic link volumes, *Transportation Research B*, Vol. 37, No. 9, pp. 815-836



- 
- Spiess, H. (1987) A maximum likelihood model for estimating origin-destination matrices, *Transportation Research B*, Vol. 21, No. 5, pp. 395-412
- Taale, H., De Hoog, A., Smulders, S.A., Tool, O. (2000) The results of a Dutch experiment with floating car data, *9<sup>th</sup> IFAC symposium on control in transportation systems*, Braunschweig, Germany
- Torp, K., Lahrmann, H. (2005) Floating car data for traffic monitoring, *5<sup>th</sup> European congress and exhibition of intelligent transport systems and services*, Hannover, Germany
- Tsekeris, T., Stathopoulos, A. (2005) Demand-oriented approach to estimate the operational performance of urban networks with and without traffic information provision, *EJTIR*, 5, no. 2, pp. 81-96
- Van der Zijpp, N. J. (1996) *Dynamic Origin-Destination Matrix Estimation on Motorway Networks*, PhD Thesis, Delft University of Technology
- Van der Zijpp, N. J. (1997) Dynamic OD-matrix estimation from traffic counts and automated vehicle identification data, *Transportation Research Record 1607*
- Van Zuylen, H.J., Willumsen, L.G. (1980) The most likely trip matrix estimated from traffic counts, *Transportation Research B*, Vol. 14, No. 3, pp. 281-293
- Van Zuylen, H.J. (1981) Some improvements in the estimation of an OD matrix from traffic counts, Proc. 8th Int. Symposium on Transportation and Traffic Theory, University of Toronto Press, Toronto Canada
- Van Zuylen, H.J., Van Lint, J.W.C., Hoogendoorn, S.P., Bliemer, M., Muller, T.H.J., Chen, Y.S., Taale, H., Zuurbier, F. (2006) *ITS and dynamic traffic management*, lecture notes, code CT5804, Delft University of Technology
- Vukovic, D. (2007) *Improved OD estimations for MiOS*, MSc thesis, Delft University of Technology
- Welch, G., Bishop, G. (2006) An introduction to the Kalman filter, *TR95-041*, University of North Carolina at Chapel Hill, Department of Computer Science
- Wen, H., Chen, F. (2007) The application of floating car system in Beijing. Proc. *14<sup>th</sup> world congress on intelligent transport systems*, Beijing, China

---

Willumsen, L.G. (1984) Estimating time-dependent trip matrices from traffic counts, Proc. 9th International Symposium on Transportation and Traffic Theory, Delft University of Technology, Delft The Netherlands

Wu, J. (1997) A real-time origin-destination matrix updating algorithm for on-line applications, *Transportation Research B*, Vol. 31, No. 5, pp. 381-396

Zajicek, J., Reinthaler, M. (2007) Real time route analysis based on floating car technology. *Proc. 14<sup>th</sup> world congress on intelligent transport systems*, Beijing, China

Zhou, X., Qin, X., Mahmassani, H.S. (2003) Dynamic origin-destination demand estimation using multi-day link traffic counts for planning applications, *Transportation Research Record*, No. 1831, 30-38

Zhou, X., Mahmassani, H.S. (2007) A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework, *Transportation Research B*, Vol. 41, No. 8, pp. 823-840

---

## 9. Appendices

---

### 9.1 Appendix 1: The recurrence model of platoon dispersion

In Bell (1991b) the *recurrence model of platoon dispersion* is described in the following manner:

If we let

- $q_i(t)$  = the traffic flow measured by the detector at entrance  $i$  to the junction or network,
- $y_j(t)$  = the traffic flow measured by the detector at exit  $j$  from the junction or network,
- $b_{ij}$  = the proportion of traffic from entrance  $i$  destined for exit  $j$ ,
- $b_{ijk}$  = the proportion of traffic from entrance  $i$  destined for exit  $j$  with a travel time, when truncated, of  $k$  intervals,
- $\alpha_j$  = the platoon dispersion factor for exit  $j$ , and
- $d$  = a discount factor applied to the sum of squared errors ( $0 \leq d \leq 1$ ).

Then the following vectors exist:

$$\begin{aligned} \mathbf{q}(t) &= [q_1(t), q_2(t), \dots]^T \\ \mathbf{b}_j &= [b_{1j}, b_{2j}, \dots]^T \\ \mathbf{b}_{jk} &= [b_{1jk}, b_{2jk}, \dots]^T \\ \mathbf{0} &= [0, 0, \dots]^T \end{aligned}$$

When it is assumed that the fastest vehicles reach the exit from any entrance within 1 interval and that for each exit there is a single travel time distribution of geometric form, then the following recurrence model of platoon dispersion applies

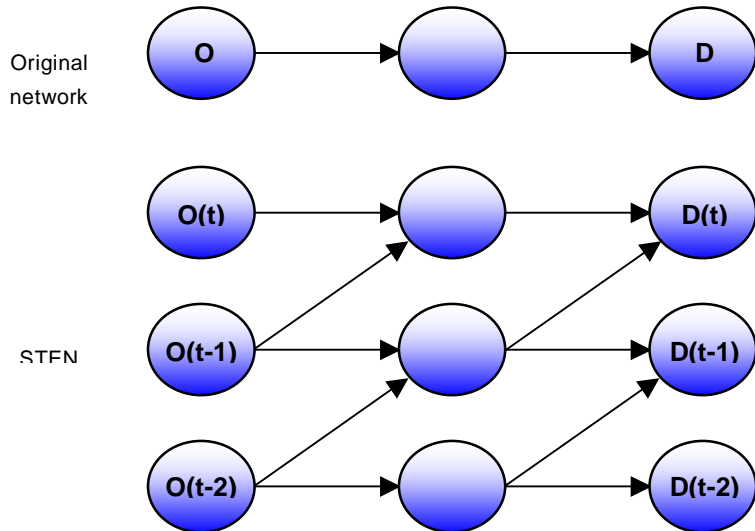
$$y_j = (1 - \alpha_j)y_j(t-1) + \alpha_j \mathbf{b}_j^T \mathbf{q}(t) \text{ for each exit } j. \quad (\text{A1.1})$$

This model has formed an important component of traffic models since it was first introduced in the TRANSYT network optimization program (Bell, 1991b).

## 9.2 Appendix 2: Space-time extended networks

A *Space-Time Extended Network (STEN)* represents time explicitly by having a layer of all nodes of the network for each relevant departure time interval<sup>17</sup>. In Figure 9.1 an example of STEN is showed.

Figure 9.1: Space-time extended network



Each level in STEN represents a departure time interval, thus the amount levels are equal to the amount of times that there are departure time intervals that have influence on the current detector counts. When traffic travels between two nodes within the same time interval, horizontal links are used between the nodes. If the travel time exceeds the time interval, the nodes are connected with a diagonal link. When networks are represented with STEN, dynamic OD matrix estimation problems can be restated as static OD matrix estimation problems on the STEN network. This method allows the extension of almost all the static formulations without any changes in their underlying logic. The only changes are the network that they operate on. This method is therefore rather flexible (Vukovic, 2007).

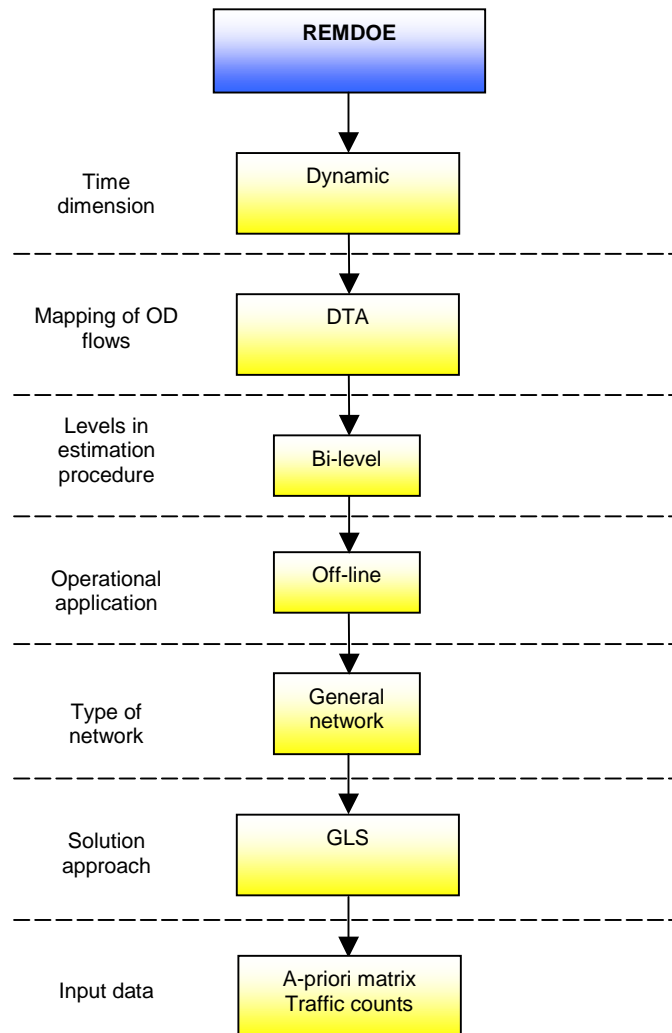
<sup>17</sup> Relevant intervals are the time intervals in which detector counts are generated during the current measurement interval.

---

### 9.3 Appendix 3: REMDOE

REMDOE is a dynamic OD matrix estimator. Its classification can be seen in Figure 9.2.

.....  
Figure 9.2: The classification of  
REMDOE



---

## 9.4 Appendix 4: Building an a priori OD matrix from junction turning fractions

For the 17 observed junctions in the study area, turning fraction per direction is available.

At each junction, each departing turning movement is considered as follows:

- Each departing turning movement is associated to one origin zone
- The distance from this origin to all destinations is calculated
- Calculate proportional from this origin to all destinations, based on the gravity model
- OD trip is obtained by multiplying the turning proportional with the total departing turning flow

This is repeated for all departing turning movements at all junctions. That gives the distribution of observed flows over possible origins and destinations.

Note that normally 12 departing tuning movements exist at one junction. 17 junctions would give possibilities to have all origins considered. All destinations are automatically considered in this case. As these 17 junction observations cover major flows in the study area, the method would allow an initial estimation of a priori OD matrix.

---

## 9.5 Appendix 5: The MATLAB codes

Below are the most important MATLAB codes that were written for this thesis.

### 9.5.1. Data cleaning

The code below shows how the data of June 7 was cleaned. The other three days were cleaned in the same way.

```
s = importdata('ODt_out_path_n_07Jun.txt');
Nodel = s(:,7);

i=1;
A=[];
while i<size(s,1)+1
    if Nodel(i)>0
        A=[A; s(i,:)];
    end
    i=i+1;
end
A=A';

fid = fopen('Input 7 June.txt', 'wt');
fprintf(fid, '%1.0f %8.0f %6.0f %3.0f %1.3f %2.0f %2.0f %2.0f %2.1f\n', A);
fclose(fid)
```

### 9.5.2. ODs assigned, OD matrices derived, paths computed and path lengths calculated.

The following is the main data processing code. In this code the following is done:

- Origins (O) and destinations (D) are assigned to the measurements according to the defined rules
- Trip times are calculated
- Paths and path lengths are calculated and written into a text file
- OD matrices are calculated and written into a text file

The code can as well be used to perform a sensitivity analysis.

```
% Program for analysing FCD. O's and D's are assigned to measurements, OD
% matrices are derived, Paths are computed and Path lengths are calculated

clear
tic;

% variables Break (br) and Real stop (stop) defined, variable values are
% for the sensitivity analysis only
A=[1 1.2 1.5 1.8 2 2.2 5 7 10 20 30 40 50 100 500 1000 1440];
% A=[2];

% This is only for the sensitivity analysis. The name of the file, where
% the different OD pairs and OD matrices are written, is defined.
filename=('SensitivityAnalysis 9 May.xls');
letter=65;
numb=1;
```

---

```

% this loop is only used for the sensitivity analysis
for i=1:size(A,2)

    br = A(i);
    stop = 2;

    % load FCD
    % Vnr: Vehicle number
    % date: The date of the measurement
    % t: time of measurement in h:mm:ss
    % tsec: time of measurement in seconds from midnight
    % v: vehicle speed
    % zone: the zone where the vehicle is detected
    % Node1: the starting node of the link where the vehicle is located
    % Node2: the end node of the link where the vehicle is located
    % PercOfLink: The vehicle's distance from Node1, percentage of total length

    s = importdata('Input 9 May.txt');
    Vnr = s(:,1);
    date = s(:,2);
    t = s(:,3);
    tsec = floor(t/10000)*3600+floor((t-floor(t/10000)*10000)/100)*60+t-
floor(t/100)*100;
    v = s(:,5);
    zone = s(:,6);
    Node1 = s(:,7);
    Node2 = s(:,8);
    PercOfLink = s(:,9);
    fprintf(1, ' dataimport complete');

    % dT is a vector with time between two measurements,
    % Ts is a vector with measurements how long a vehicle is stop
    % Stop1 is a vector where 1 is applied to a stop when it exceeds the
    %parameter stop, i.e. Real stops, 0 to all others.
    dT = zeros(size(s,1),1);
    Ts = zeros(size(s,1),1);
    Stop1 = zeros(size(s,1),1);
    for i = 2:size(s,1)
        if Vnr(i)==Vnr(i-1)
            dT(i,:)=(tsec(i)-tsec(i-1))/60;
        else
            dT(i,:)=0;
        end
        if v(i) == 0
            if v(i-1) == 0
                Ts(i,:) = Ts(i-1)+ dT(i);
            else
                Ts(i,:) = dT(i);
            end
        end
        if v(i) == 0 && Ts(i)>stop
            Stop1(i,:) = 1;
        else
            Stop1(i,:) = 0;
        end
    end
    fprintf(1, ' dT, Ts and Stop1 complete');

```

---



---

```

% Stop2 applies 1 to all measurements that belong to a Real stop, 0 to all
%others.
Stop2=Stop1;
j=1;
while j<50
    for i = 1:size(s,1)-1
        if v(i)==0 && Stop2(i+1)==1
            Stop2(i,:)=1;
        end
    end
    j=j+1;
end
fprintf(1,' Stop2 complete');

% When time between measurements exceeds the parameter br, Bre applies 1 to
%the measurement after the break, 0 to all others.
Bre = zeros(size(s,1),1);
for i = 1:size(s,1)
    if dT(i) > br
        Bre(i,:) = 1;
    else
        Bre(i,:) = 0;
    end
end
fprintf(1,' Bre complete');

% OD applies 1 to all origins and 2 to all destinations.
OD = zeros(size(s,1),1);
for i = 1:size(s,1)

    % first and last measurements
    if i == 1 && v(i)==0
        for m = 1:size(s,1)
            if v(m)~=0, break
            end
        end
        if Bre(i+1)~=1 && Stop2(i+1)~=1
            OD(m,:) = 1;
        end
    elseif i == size(s,1) && v(i)==0
        for n=size(s,1):-1:1
            if v(n)~=0, break
            end
        end
        if Bre(i)~=1 && Stop2(i-1)
            OD(n,:) = 2;
        end
    elseif i==1 && v(i)~=0
        if Bre(i+1)~=1 && Stop2(i+1)~=1
            OD(i,:)=1;
        end
    elseif i==size(s,1) && v(i)~=0
        if Bre(i)~=1 && Stop2(i-1)~=1
            OD(i,:)=2;
        end
    end
end

```

---

---

```

% first and last measurement from a vehicle
elseif Vnr(i)>Vnr(i-1) && v(i)==0
    for m=i:size(s,1)
        if v(m)~=0, break
        end
    end
    end
    if Bre(i+1)~=1 && Stop2(i+1)~=1
    OD(m,:)=1;
    end
elseif Vnr(i)<Vnr(i+1) && v(i)==0
    for n=i:-1:1
        if v(n)~=0, break
        end
    end
    if Bre(i)~=1 && Stop2(i-1)~=1
    OD(n,:)=2;
    end
elseif Vnr(i)>Vnr(i-1)&& v(i)~=0
    if Bre(i+1)~=1 || Stop2(i+1)~=1
    OD(i,:) = 1;
    end
elseif Vnr(i)<Vnr(i+1)&& v(i)~=0
    if Bre(i)~=1 || Stop2(i-1)~=1
    OD(i,:) = 2;
    end
end
end
end

% origins and destinations due to breaks and real stops
for i = 2:size(s,1)-1
    if v(i)~=0
        if Bre(i)==1 || Stop2(i-1)==1
            OD(i,:) = 1;
        elseif Bre(i+1)==1 || Stop2(i+1)==1
            OD(i,:) = 2;
        elseif (Bre(i)==1 && Bre(i+1)==1) || (Stop2(i-1)==1 &&
Stop2(i+1)==1) || (Bre(i)==1 && Stop2(i+1)==1) || (Bre(i+1)==1 && Stop2(i-1)==1)
            OD(i,:) = 0;
        end
    end
end
end
fprintf(1, ' OD complete');

% OD2 replaces the zeros in OD for the value of the previous O/D.
OD2 = zeros(size(s,1),1);
for i = 1:size(s,1)
    if i == 1
        OD2(i,:)=OD(i);
    elseif OD(i)==1 || OD(i)==2
        OD2(i,:)=OD(i);
    else
        OD2(i,:)=OD2(i-1);
    end
end
end
fprintf(1, ' OD2 complete');

% StartTime and StartTimeSec are vectors with the starting times of all

```

---

---

```

% trips
StartTime = [];
StartTimeSec = [];
j=1;
for i=1
    if OD2(i)==1
        StartTime(j,:)=t(i);
        StartTimeSec(j,:)=tsec(i);
        j=j+1;
    end
end
for i=2:size(s,1)
    if OD2(i)==1 && OD2(i-1)~=1
        StartTime(j,:)=t(i);
        StartTimeSec(j,:)=tsec(i);
        j=j+1;
    end
end
fprintf(1,' StartTime complete');

% EndTime and EndTimeSec are vectors with the end times of all trips
EndTime = [];
EndTimeSec = [];
j=1;
for i=2:size(s,1)
    if OD2(i)==2 && OD2(i-1)==1
        EndTime(j,:)=t(i);
        EndTimeSec(j,:)=tsec(i);
        j=j+1;
    end
end
fprintf(1,' EndTime complete');

TripTimeSec=EndTimeSec-StartTimeSec;
fprintf(1,' EndTime complete');

% zone2 is a vector with the zone numbers of O/D only, corresponding to
% OD2
zone2 = zeros(size(s,1),1);
for i=1
    if OD2(i)==0
        zone2(i,:)=zone(i+1);
    else
        zone2(i,:)=zone(i);
    end
end
for i=2:size(s,1)
    if (OD2(i)==2 && OD2(i-1)==1) || (OD2(i)==1 && OD2(i-1)==2)
        zone2(i,:)=zone(i);
    else
        zone2(i,:)=zone2(i-1);
    end
end
fprintf(1,' zone2 complete');

% ODpair is a matrix with all the OD pairs and trip times in seconds
ODpair = [];
x=1;
for i = 1:size(s,1)-1

```

---

---

```

    if OD2(i) == 1 && OD2(i+1) == 2
        pair(i,:)=[StartTimeSec(x) zone2(i) zone2(i+1)];
        x=x+1;
        ODpair=[ODpair;pair(i,:)];
    end
end
fprintf(1,' ODpair complete');

% MeasuredPath is the detected path of the vehicles for each trip
MeasuredPath = ones(size(ODpair,1),700)*999;
x=1;          %measurement
j=1;          %Path
while x<=size(s,1)
    k=1;      %Placement in path
    if OD(x)==1
        MeasuredPath(j,(2*k-1):(2*k))=[Node1(x) Node2(x)];
        x=x+1;
        k=k+1;
        if OD(x)==0 || OD(x)==1
            while OD(x)==0 || OD(x)==1
                MeasuredPath(j,(2*k-1):(2*k))=[Node1(x) Node2(x)];
                x=x+1;
                k=k+1;
            end
        end
    end
    if OD(x)==2
        MeasuredPath(j,(2*k-1):(2*k))=[Node1(x) Node2(x)];
        x=x+1;
    end
    j=j+1;
else
    x=x+1;
end
end
fprintf(1,' MeasuredPath complete');

Paths = ones(size(ODpair,1),800)*999;
i=2087;
while i<=size(MeasuredPath,1)
    Path=[MeasuredPath(i,1)];
    j=2;
    while MeasuredPath(i,j)~=999
        if MeasuredPath(i,j+1)~=999
            CoPath=CompPath(MeasuredPath(i,j),MeasuredPath(i,j+1));
            Path=[Path CoPath];
        else
            Path=[Path MeasuredPath(i,j)];
            Paths(i,1:size(Path,2))=Path;
        end
        j=j+2;
    end
    i=i+1;
end
fprintf(1,' Paths complete');

% PathLength calculates the length of each path
PathLength=zeros(size(Paths,1),1);
i=1;
while i<=size(Paths,1)

```

---

---

```

        k=1;
        while Paths(i,k+1)~=999 && k<=size(Paths,2)

PathLength(i)=PathLength(i)+Linklength(Paths(i,k),Paths(i,k+1))*1000;
        k=k+1;
        end
        i=i+1;
    end
    fprintf(1, ' PathLength complete');

% ODandPath includes the start time of the trip, the OD zones and the
% path. TimePathLength includes the start time of the trip and the path
% length.
ODandPath=[ODpair Paths];
TimePathLength=[StartTimeSec PathLength];

% Outcomes of ODandPath printed in an text file
fid = fopen('ODandPath 9May.txt', 'w+');
for i=1:size(ODandPath,1)
    j=1;
    while ODandPath(i,j+1)~=999
        fprintf(fid, '%5.0f', ODandPath(i,j));
        j=j+1;
    end
    fprintf(fid, '%5.0f\n',ODandPath(i,j));
end
fclose(fid);

% Outcomes of TimePathLength printed in a text file
fid = fopen('PathLength 9May.txt', 'w+');
fprintf(fid, '%5.0f %5.0f\n',TimePathLength);
fclose(fid);

% Below the OD matrices are calculated
% ODmatrixTotal is an OD matrix from all the data.
ODmatrixTotal = [];
for i = 1:max(zone2) %i is origin
    Row = [];
    for j = 1:max(zone2) %j is dest
        k = find(ODpair(:,2)==(i) & ODpair(:,3)==(j));
        number(j,:) = length(k);
        Row = [Row number(j,:)];
    end
    ODmatrixTotal = [ODmatrixTotal;Row];
end
fprintf(1, ' ODmatrix complete');

TimeSlice = [2700:3600:31800];
ODmatrixTimeSlice = [];
for x=1:size(TimeSlice,2)-1
    for i = 1:max(zone2) %i is origin
        RowTimeSlice = [];
        for j = 1:max(zone2) %j is dest
            kTimeSlice = find(ODpair(:,1)>=TimeSlice(x) &
ODpair(:,1)<TimeSlice(x+1) & ODpair(:,2)==(i) & ODpair(:,3)==(j));
            numberTimeSlice(j,:) = length(kTimeSlice);

```

---

---

```

        RowTimeSlice = [RowTimeSlice numberTimeSlice(j,:)];
    end
    ODmatrixTimeSlice = [ODmatrixTimeSlice;RowTimeSlice];
end
end
fprintf(1,' ODmatrixTimeSlice complete');

% ODmatrixAM is an OD matrix for AM peak 7:00-10:00
ODmatrixAM = [];
for i = 1:max(zone2) %i is origin
    RowAM = [];
    for j = 1:max(zone2) %j is dest
        kAM = find(ODpair(:,1)>=25200 & ODpair(:,1)<36000 & ODpair(:,2)==(i)
& ODpair(:,3)==(j));
        numberAM(j,:) = length(kAM);
        RowAM = [RowAM numberAM(j,:)];
    end
    ODmatrixAM = [ODmatrixAM;RowAM];
end
fprintf(1,' ODmatrixAM complete');

% ODmatrixPM is an OD matrix for PM peak 16:00-19:00
ODmatrixPM = [];
for i = 1:max(zone2) %i is origin
    RowPM = [];
    for j = 1:max(zone2) %j is dest
        kPM = find(ODpair(:,1)>=57600 & ODpair(:,1)<68400 & ODpair(:,2)==(i)
& ODpair(:,3)==(j));
        numberPM(j,:) = length(kPM);
        RowPM = [RowPM numberPM(j,:)];
    end
    ODmatrixPM = [ODmatrixPM;RowPM];
end
fprintf(1,' ODmatrixPM complete');

if letter>90 && letter<116
    letter1=letter-26;
    range1 = [sprintf('%s', 65) sprintf('%s', letter1) '1'];
elseif letter>115
    letter1=letter-52;
    range1 = [sprintf('%s', 66) sprintf('%s', letter1) '1'];
else
    range1 = [sprintf('%s', letter) '1'];
end
range2 = ['A' num2str(numb)];

sheet1=['ODpair'];
sheet2=['ODmatrixTotal'];
sheet3=['ODmatrixAM'];
sheet4=['ODmatrixPM'];
sheet5=['ODmatrixTimeSlice'];

warning off MATLAB:xlswrite:AddSheet

xlswrite(filename, ODpair, sheet1,range1);
xlswrite(filename, ODmatrixTotal, sheet2,range2);
xlswrite(filename, ODmatrixAM, sheet3,range2);
xlswrite(filename, ODmatrixPM, sheet4,range2);
xlswrite(filename, ODmatrixTimeSlice, sheet5,range2);

```

---

---

```

        letter=letter+4;
        numb=numb+20;
end
beep;
toc

```

### 9.5.3. Functions

The function below calculates the shortest path between two links.

```

function f = CompPath(N1,N2)

p = xlsread('CompPath.xls');
Nod1 = p(:,1);
Nod2 = p(:,2);
Time = p(:,3);
Dist = p(:,4);
NBNod = p(:,5);
List = ones(size(Nod1,1),max(NBNod)+1)*999;
for i = 1:size(Nod1,1)
    List(i,1:NBNod(i)) = p(i,6:(5+NBNod(i)));
end

for i=1:size(Nod1,1)
    if Nod1(i)==N1 && Nod2(i)==N2
        g=List(i,:);
        x=1;
        while g(x)< 999
            x=x+1;
        end
        f=List(i,1:x-1);
    end
end
end

```

The following function calculates the shortest path between two zones.

```

function f = CompPathZone(Z1,Z2)

p = xlsread('CompPathZone.xls');
Zon1 = p(:,1);
Zon2 = p(:,2);
Time = p(:,3);
Dist = p(:,4);
NBNodZ = p(:,5);
List = ones(size(Zon1,1),max(NBNodZ)+1)*999;
for i = 1:size(Zon1,1)
    List(i,1:NBNodZ(i)) = p(i,6:(5+NBNodZ(i)));
end

for i=1:size(Zon1,1)
    if Zon1(i)==Z1 && Zon2(i)==Z2
        g=List(i,:);
        x=1;
        while g(x)< 999

```

---

```

        x=x+1;
    end
    f=List(i,1:x-1);
end
end

```

The following MATLAB code is the function that gives the length of links.

```

function f = Linklength(N1,N2)

p = importdata('Link_DistTime.txt');
Link = p(:,1);
Nod1 = p(:,2);
Nod2 = p(:,3);
Dist = p(:,4); %km
Time = p(:,6); %min

for i=1:size(Nod1,1)
    if Nod1(i)==N1 && Nod2(i)==N2
        f=Dist(i,:);
    end
end

```

The following function gives the number of nodes in the shortest path between two nodes.

```

function f = NBNodes(N1,N2)

p = xlsread('CompPath.xls');
Nod1 = p(:,1);
Nod2 = p(:,2);
Time = p(:,3);
Dist = p(:,4);
NBN = p(:,5);

for i=1:size(Nod1,1)
    if Nod1(i)==N1 && Nod2(i)==N2
        f=NBN(i);
    end
end

```

The following function gives the number of nodes in the shortest path between two zones.

```

function f = NBNodesZ(Z1,Z2)

p = xlsread('CompPathZone.xls');
Zon1 = p(:,1);
Zon2 = p(:,2);
Time = p(:,3);
Dist = p(:,4);
NBNZ = p(:,5);

for i=1:size(Zon1,1)
    if Zon1(i)==Z1 && Zon2(i)==Z2

```



---

```

        f=NBNZ(i);
    end
end

```

#### 9.5.4. Trip length distribution

The code below calculates and draws the TLD from FCD as well as data given by the OVG (1996).

```

clear
s=importdata('PathLength 9May.txt')
Space=[0 1 1.5 5 10 15 20 30 40 50 75 100 150 200 300];

TL1=[];
j=1;
while j<size(Space,2)
    n=0;
    for i=1:size(s,1)
        if s(i,2)>Space(j)*1000 && s(i,2)<=Space(j+1)*1000
            n=n+1;
        end
    end
    TL1=[TL1;Space(j+1) n];
    j=j+1;
end

sum=0;
for i=1:size(TL1,1)
    sum=sum+TL1(i,2);
end
hold on
figure(1);
plot(TL1(:,1),TL1(:,2)/sum*100, '-b')

r=importdata('PathLength 10May.txt')

TL2=[];
j=1;
while j<size(Space,2)
    n=0;
    for i=1:size(r,1)
        if r(i,2)>Space(j)*1000 && r(i,2)<=Space(j+1)*1000
            n=n+1;
        end
    end
    TL2=[TL2;Space(j+1) n];
    j=j+1;
end

sum=0;
for i=1:size(TL2,1)
    sum=sum+TL2(i,2);
end

hold on
figure(1);
plot(TL2(:,1),TL2(:,2)/sum*100,'o-r')

```

---

---

```

%OVG TLD
OVG=[1 1.5 5 10 15 20 30 40 50 75 100 150 200 300; 3.55 18.18 16.89 20.66 11.04
6.65 10.26 4.12 2.31 3.26 1.39 1.14 0.39 0.17]';
hold on
figure(1);
plot(OVG(:,1),OVG(:,2),'*-g')
h = legend('9 May','10 May','OVG');
xlabel('Trip length [km]');
ylabel('Percentage of total trips [%]');

```

The following code calculates and plots the TLD from an OD matrix

```

clear
S=importdata('ShortestPathLengths.txt');
od=importdata('ODmatrixTOT.txt');

OD=[];
for i=1:18
    for j=1:18
        if i~=j
            OD=[OD;od(i,j)];
        end
    end
end

S=[S OD];

Space=[0 0.5 1 1.5 2 2.5 3 3.5 4];

TL1=[];
j=1;
while j<size(Space,2)
    n=0;
    for i=1:size(S,1)
        if S(i,3)>Space(j) && S(i,3)<=Space(j+1)
            n=n+S(i,4);
        end
    end
    TL1=[TL1;Space(j+1) n];
    j=j+1;
end

sum=0;
for i=1:size(TL1,1)
    sum=sum+TL1(i,2);
end

for i=1:size(TL1,1)
    TL1(i,2)=TL1(i,2)/sum*100;
end
hold on
plot(TL1(:,1),TL1(:,2),'--r')
xlabel('Trip length [km]');
ylabel('Percentage of total trips [%]');
set(gca,'XTick',[0 0.5 1 1.5 2 2.5 3 3.5 4]);

```

---

```
h = legend('7:30-7:39', '7:40-7:49', '7:50-7:59', '8:00-8:09', '8:10-8:19', '8:20-8:29', '8:30-8:39', '8:40-8:49');
```

### 9.5.5. Input for NEST written

In the following MATLAB code, the input for NEST is written into a text file. Furthermore, the parameters  $p_{ij}^a$  are calculated.

```
% Program that writes the input file for NEST
tic;

% Name of NEST inputfile defined - Needs to be changed for each time slice
fid = fopen('CDPIII.txt', 'w+');

% Traffic counts imported - CountsTime needs to be changed for each time
% slice
Counts=importdata('LinkCounts.txt');
% CountsTime=Counts(:,4); %Time00 %I
% CountsTime=Counts(:,5); %Time10 %II
CountsTime=Counts(:,6); %Time20 %III
% CountsTime=Counts(:,7); %Time30 %IV
% CountsTime=Counts(:,8); %Time40 %V
% CountsTime=Counts(:,9); %Time50 %VI
% CountsTime=Counts(:,10); %Time60 %VII
% CountsTime=Counts(:,11); %Time70 %VIII
% CountsTime=Counts(:,12); %Total

% Definition of Links imported - Same for all time slices
L=importdata('Links.txt');
L=[L(:,2) L(:,3)];

% Paths imported - Needs to be changed for each time slice
s=importdata('ODandPath 9May Time20.txt');
StartTime = s(:,1);
Zone1 = s(:,2);
Zone2 = s(:,3);
Path = s(:,4:size(s,2));

% A-priori matrix imported - Needs to be changed for each time slice
Apr=importdata('PFCD20.txt');
Apr=Apr(:,3);

% #1 Title printed - Needs to be changed for each time slice
fprintf(fid, 'Chengdu input for NEST 9 May with PFCD martix Time20 7:50-7:59\n');

% Hereafter nothing needs to be changed for different timeslices!!

% #2 Specifications printed - Same for all time slices
Accuracy=0.25; %Max rel. err allowed between estima. and actual flows
RouteChoice=1; %All-or-nothing=0; fractional distribution=1
MaxIterations=25; %Max iterations the program will perform
Output=1; %Complete output=1; Shortened output=0
Input=0; %Normal input=0; Shortened input=1

Specifications=[Accuracy RouteChoice MaxIterations Output Input]';
```

---

---

```

fprintf(fid, '%1.3f %4.0f %4.0f %4.0f %4.0f\n', Specifications);

% #3 Counted Links printed - Same for all time slices
CountedLinks=Counts(:,2);
i=1;
while i<=size(CountedLinks,1)
    if CountedLinks(i)<10
        fprintf(fid, '%5.0f%1.0f', [0 CountedLinks(i)]);
        fprintf(fid, ' ');
    else
        fprintf(fid, '%6.0f', [CountedLinks(i)]);
        fprintf(fid, ' ');
    end
    if i==10 || i==20 || i==30 || i==40 || i==50
        fprintf(fid, '\n');
    end
    i=i+1;
end
fprintf(fid, '\n');

% Star printed
fprintf(fid, '*\n');

% #4 OD connections printed - Same for all time slices
x=1;
for i=1:18
    for j=1:18
        if i<10 && j<10
            fprintf(fid, '%3.0f%1.0f%1.0f%1.0f', [0 i 0 j]);
            fprintf(fid, ' ');
        elseif i>=10 && j<10
            fprintf(fid, '%4.0f%1.0f%1.0f', [i 0 j]);
            fprintf(fid, ' ');
        elseif i<10 && j>=10
            fprintf(fid, '%3.0f%1.0f%2.0f', [0 i j]);
            fprintf(fid, ' ');
        elseif i>=10 && j>=10
            fprintf(fid, '%4.0f%2.0f', [i j]);
            fprintf(fid, ' ');
        end
        x=x+1;
        if x==11 || x==21 || x==31 || x==41 || x==51 || x==61 || x==71
            fprintf(fid, '\n');
        end
        if x==81 || x==91 || x==101 || x==111 || x==121 || x==131 || x==141
            fprintf(fid, '\n');
        end
        if x==151 || x==161 || x==171 || x==181 || x==191 || x==201
            fprintf(fid, '\n');
        end
        if x==211 || x==221 || x==231 || x==241 || x==251 || x==261
            fprintf(fid, '\n');
        end
        if x==271 || x==281 || x==291 || x==301 || x==311 || x==321
            fprintf(fid, '\n');
        end
    end
end
end
fprintf(fid, '\n');

```

---

---

```

% Star printed
fprintf(fid, '*\n');

% #5 Routes and route choice printed - file is changed above, otherwise
% same for all time slices

% Paths organized by zones and intrazonal trips erased
Q=[];
z1=1;
while z1<=18
    z2=1;
    while z2<=18
        for i=1:size(s,1)
            if Zone1(i)==z1 && Zone2(i)==z2
                if z1~=z2
                    Q=[Q; Zone1(i) Zone2(i) Path(i,:)];
                end
            end
        end
        z2=z2+1;
    end
    z1=z1+1;
end

% Computed shortest paths inserted when OD connection is missing
P=[];
z1=1;
while z1<=18
    z2=1;
    while z2<=18
        m=0;
        for i=1:size(s,1)
            if Zone1(i)==z1 && Zone2(i)==z2
                if z1~=z2
                    m=m+1;
                    P=[P; m z1 z2 Path(i,:)];
                end
            end
        end
        if m==0
            if z1~=z2
                P=[P; m z1 z2 CompPathZone(z1,z2) zeros(1,size(Path,2)-
NBNodesZ(z1,z2))];
            end
        end
        z2=z2+1;
    end
    z1=z1+1;
end

% The paths of each OD pair counted
R=P;
i=1;
while i<size(R,1)
    if R(i,1)==1
        m=0;
        while i<size(R,1) && R(i,1)<R(i+1,1)
            i=i+1;

```

---

---

```

        m=m+1;
    end
    R((i-m):i,1)=m+1;
end
i=i+1;
end

%Node numbers replaced by link numbers
T=[R(:,1:3) zeros(size(R,1),size(R,2)-3)];
i=1;
while i<=size(R,1)
    j=4;
    while R(i,j+1)~=0
        for x=1:size(L,1)
            if L(x,:)=[R(i,j) R(i,j+1)]
                l=x;
            end
        end
        T(i,j)=l;
        j=j+1;
    end
    i=i+1;
end

% Paths where shortest path is used printed
i=1;
while i<=size(T,1)
    if T(i,1)==0
        S1=[T(i,2:size(T,2))];
        j=1;
        S=[];
        while j<=size(S1,2)
            if S1(j)>0
                S=[S S1(j)];
            end
            j=j+1;
        end

        for x=1
            if S(x)<10 && S(x+1)<10
                fprintf(fid, '%3.0f%1.0f%1.0f%1.0f', [0 S(x) 0 S(x+1)]);
                fprintf(fid, ' ');
            elseif S(x)<10 && S(x+1)>=10
                fprintf(fid, '%3.0f%1.0f%2.0f', [0 S(x) S(x+1)]);
                fprintf(fid, ' ');
            elseif S(x)>=10 && S(x+1)<10
                fprintf(fid, '%4.0f%1.0f%1.0f', [S(x) 0 S(x+1)]);
                fprintf(fid, ' ');
            elseif S(x)>=10 && S(x+1)>=10
                fprintf(fid, '%4.0f%2.0f', [S(x) S(x+1)]);
                fprintf(fid, ' ');
            end
        end

        for x=3:size(S,2)
            for m=1:size(CountedLinks,1)
                if S(x)==CountedLinks(m)
                    if S(x)<10

```

---

---

```

        fprintf(fid, '%5.0f%1.0f', [0 S(x)]);
        fprintf(fid, ' ');
    else
        fprintf(fid, '%6.0f',S(x));
        fprintf(fid, ' ');
    end
end
end
end
end
    fprintf(fid, '\n');
end
i=i+1;
end

% Paths with measured paths printed
i=1;
while i<=size(T,1)
    if T(i,1)>=1
        A1=[];
        for x=i:(i+T(i,1)-1)
            A1=[A1 T(x,4:size(T,2))];
        end
        j=1;
        A=[];
        while j<=size(A1,2)
            if A1(j)>0
                A=[A A1(j)];
            end
            j=j+1;
        end
        E=[];
        y=1;
        while y<=max(A)
            t=0;
            for j=1:size(A,2)
                if A(j)==y
                    t=t+1;
                end
            end
            E=[E t/T(i,1)*100];
            y=y+1;
        end
        x=1;
        n=0;
        while x<=size(E,2)

            if n==0 || n==80 || n==160 || n==240
                if n~=0
                    fprintf(fid, '\n');
                end
                if T(i,2)<10 && T(i,3)<10
                    fprintf(fid, '%3.0f%1.0f%1.0f%1.0f', [0 T(i,2) 0 T(i,3)]);
                    fprintf(fid, ' ');
                elseif T(i,2)<10 && T(i,3)>=10
                    fprintf(fid, '%3.0f%1.0f%2.0f', [0 T(i,2) T(i,3)]);
                    fprintf(fid, ' ');
                end
            end
            x=x+1;
        end
    end
    i=i+1;
end

```

---

---

```

elseif T(i,2)>=10 && T(i,3)<10
    fprintf(fid, '%4.0f%1.0f%1.0f', [T(i,2) 0 T(i,3)]);
    fprintf(fid, ' ');
elseif T(i,2)>=10 && T(i,3)>=10
    fprintf(fid, '%4.0f%2.0f', [T(i,2) T(i,3)]);
    fprintf(fid, ' ');
end
n=n+8;
end

if E(x)>0
    for m=1:size(CountedLinks,1)
        if x==CountedLinks(m)
            if E(x)>=100
                if x<10
                    fprintf(fid,'%5.0f%1.0f',[0 x]);
                    fprintf(fid, ' ');
                else
                    fprintf(fid,'%6.0f',x);
                    fprintf(fid, ' ');
                end
                n=n+8;
            elseif E(x)>0 && E(x)<10
                if x<10
                    fprintf(fid,'%5.0f%1.0f%1.0f%1.0f',[0 x 0
E(x)]);

                else
                    fprintf(fid,'%6.0f%1.0f%1.0f',[x 0 E(x)]);
                end
                n=n+8;
            elseif E(x)>=10 && E(x)<100
                if x<10
                    fprintf(fid,'%5.0f%1.0f%2.0f',[0 x E(x)]);
                else
                    fprintf(fid,'%6.0f%2.0f',[x E(x)]);
                end
                n=n+8;
            end
        end
    end
end
end
x=x+1;
end
fprintf(fid, '\n');
i=i+T(i,1)-1;
end
i=i+1;
end

% Star printed
fprintf(fid, '*\n');

% #6 Counts printed - CountsTime is changed above, otherwise same for all
% time slices
for i=1:size(Counts,1)
    if CountsTime(i)==0
        fprintf(fid,'%5.0f',1);
    else
        fprintf(fid,'%5.0f',CountsTime(i));
    end
end

```

---



---

```

    end
    if i==16 || i==32 || i==48
        fprintf(fid, '\n');
    end
end
fprintf(fid, '\n');

% Star printed
fprintf(fid, '*\n');

% #7 A-priori matrix printed - A-priori matrix changed above, otherwise
% same for all time slices
for i=1:size(Apr,1)
    fprintf(fid, '%5.0f', Apr(i));
    if i==16 || i==32 || i==48 || i==64 || i==80 || i==96 || i==112
        fprintf(fid, '\n');
    end
    if i==128 || i==144 || i==160 || i==176 || i==192 || i==208 || i==224
        fprintf(fid, '\n');
    end
    if i==240 || i==256 || i==272 || i==288 || i==304 || i==320 || i==336
        fprintf(fid, '\n');
    end
end

fclose(fid);
toc
beep

```

## 9.6 Appendix 6: Comparison of the structure of OD matrices estimated with different a-priori matrices

### CFCD vs. UOD

Figure 9.3: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:30-7:39

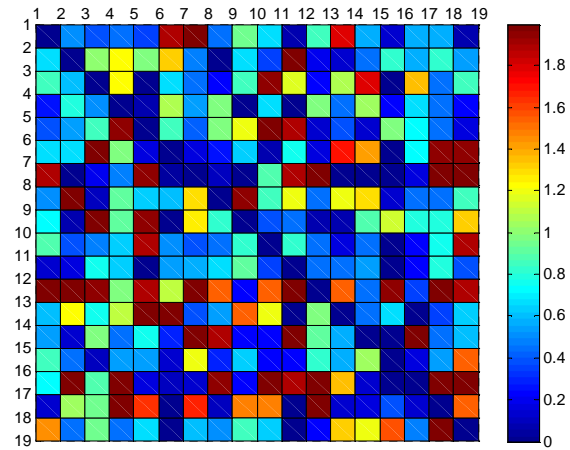


Figure 9.4: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:40-7:49

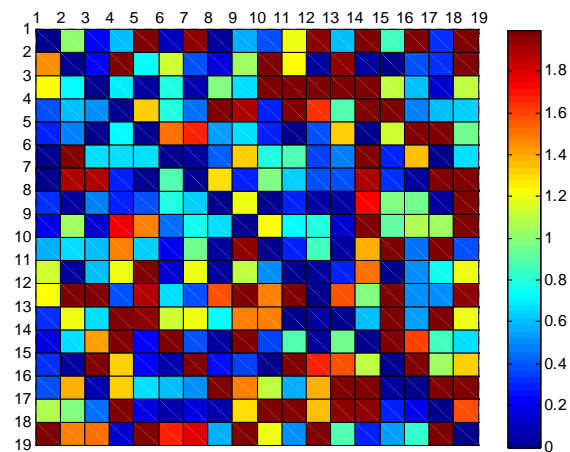
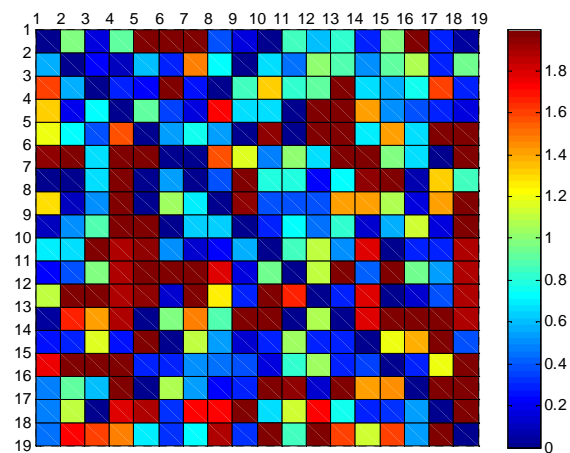
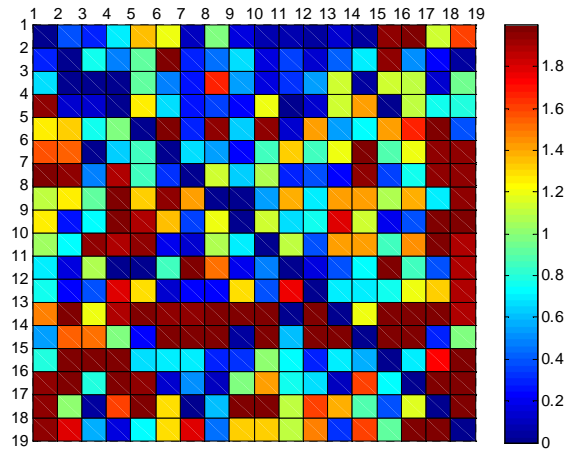


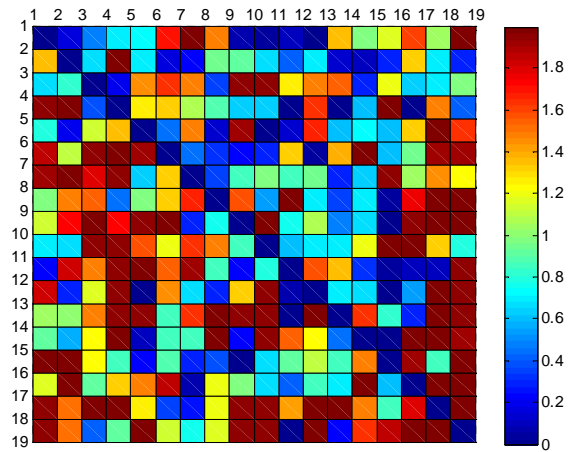
Figure 9.5: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:50-7:59



.....  
 Figure 9.6: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:00-8:09



.....  
 Figure 9.7: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:10-8:19



.....  
 Figure 9.8: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:20-8:29

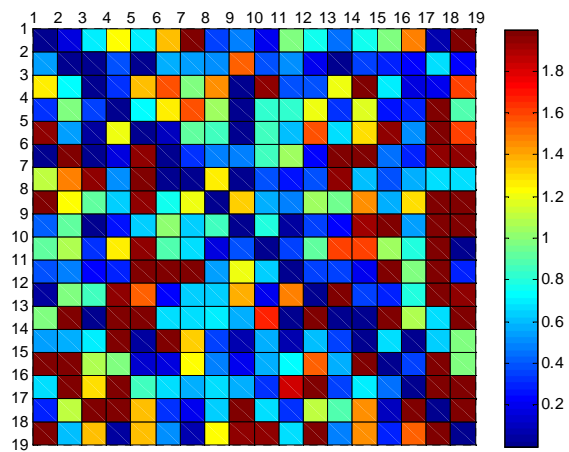


Figure 9.9: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39

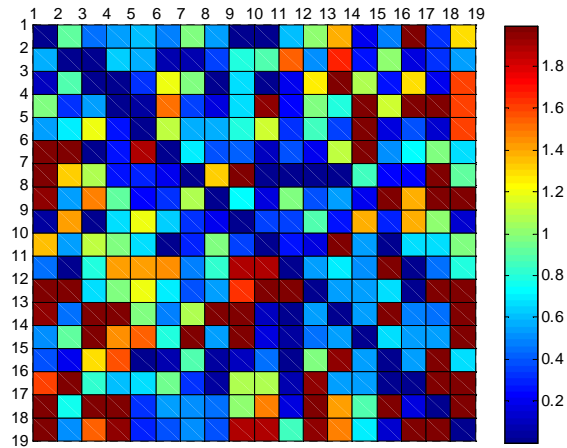


Figure 9.10: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:40-8:49

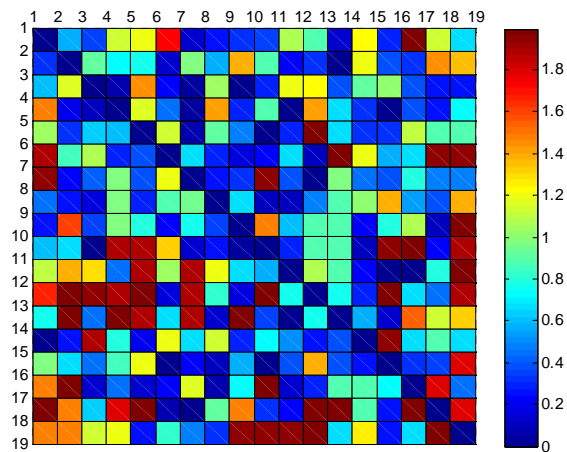


Figure 9.11: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. The whole time period 7:30-8:49

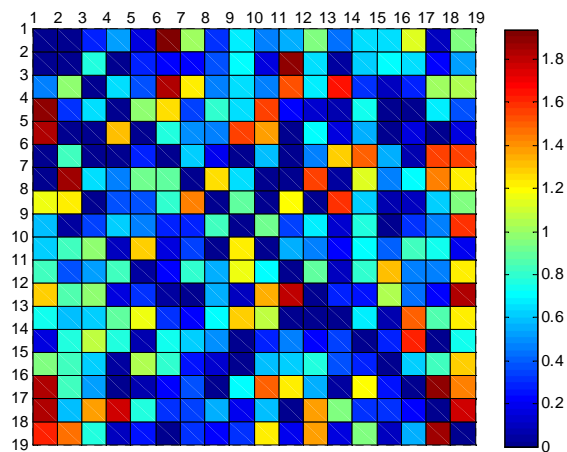


Figure 9.12: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:30-7:39

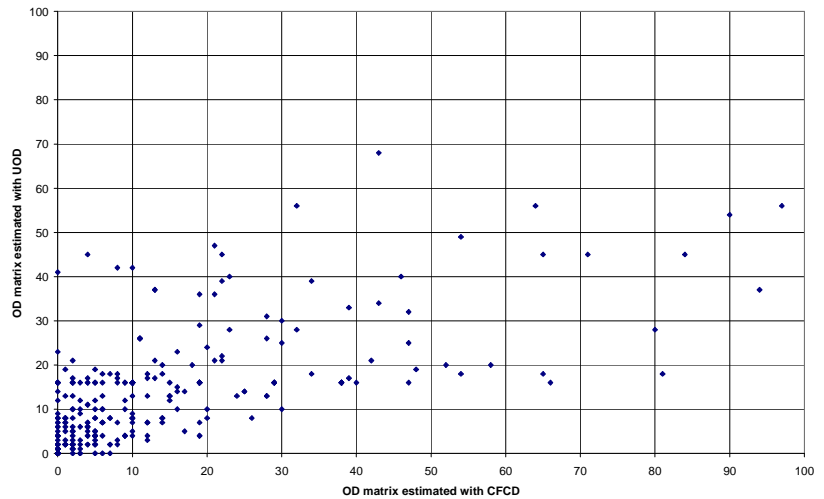


Figure 9.13: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:40-7:49

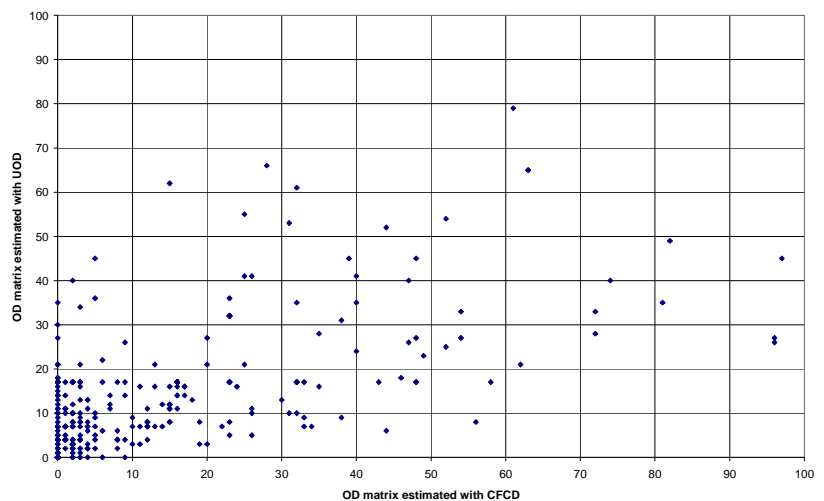


Figure 9.14: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 7:50-7:59

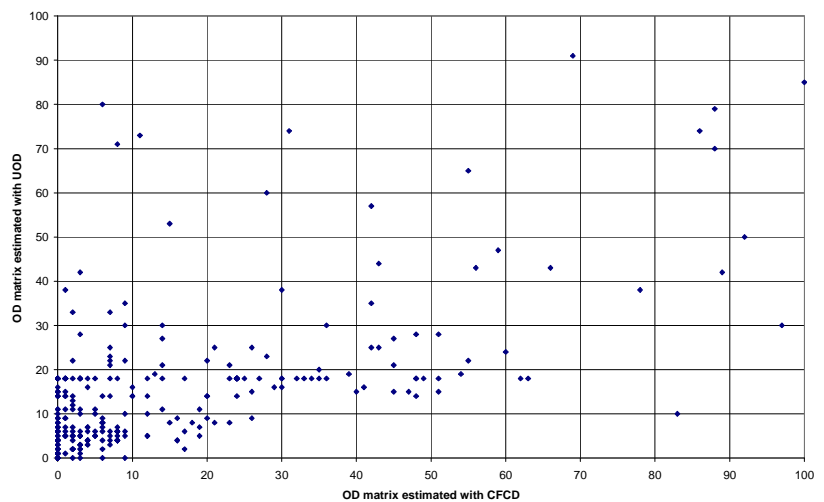


Figure 9.15: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:00-8:09

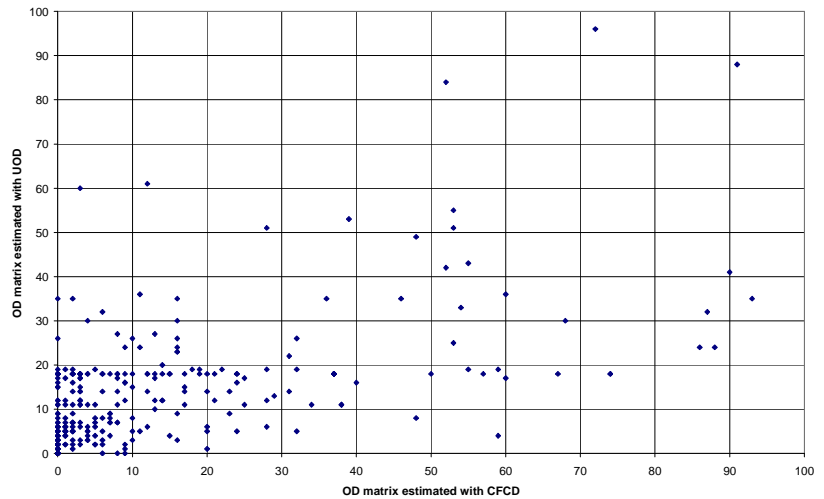


Figure 9.16: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:10-8:19

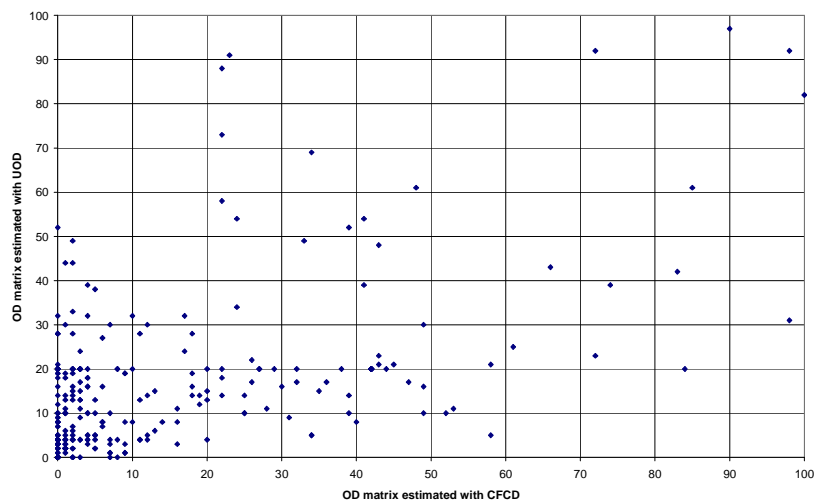


Figure 9.17: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:20-8:29

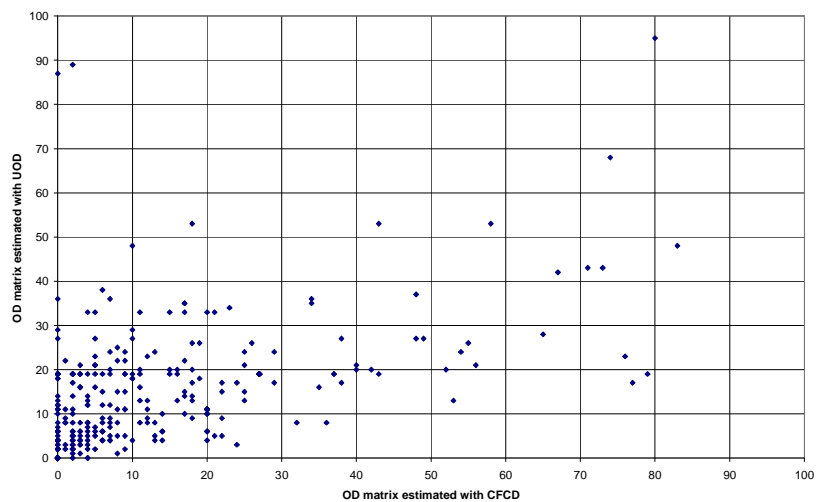


Figure 9.18: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:30-8:39

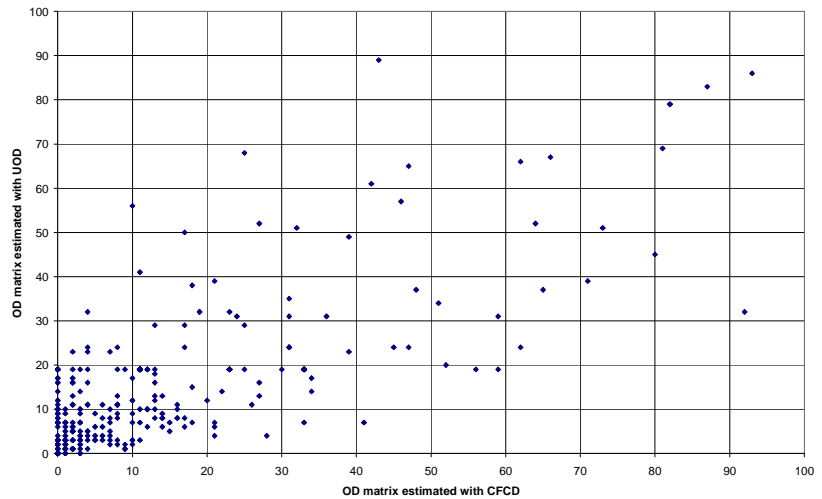


Figure 9.19: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Time period 8:40-8:49

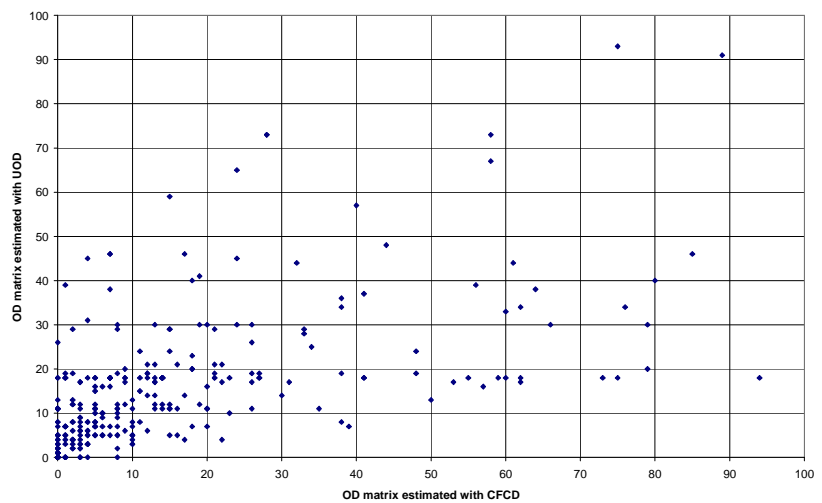
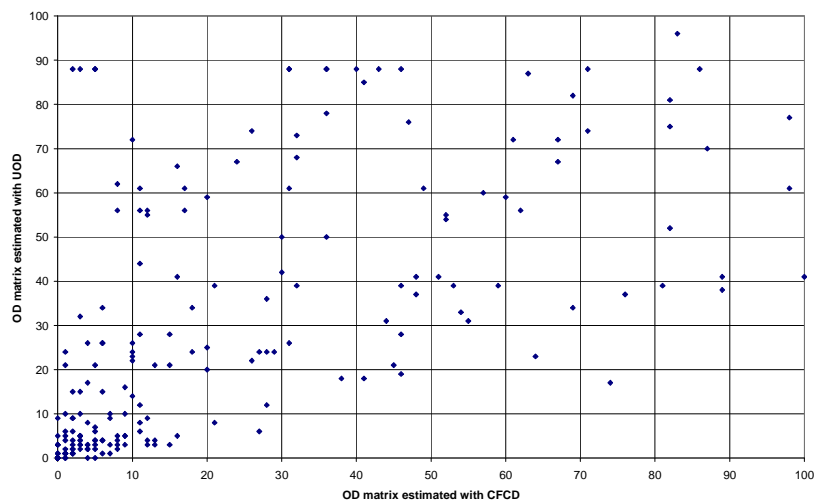


Figure 9.20: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and UOD. Total time period



## CFCD vs. OFCD

Figure 9.21: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39

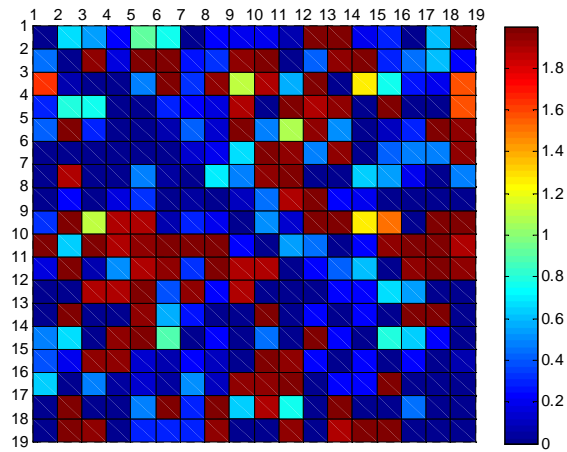


Figure 9.22: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. The whole time period 7:30-8:49

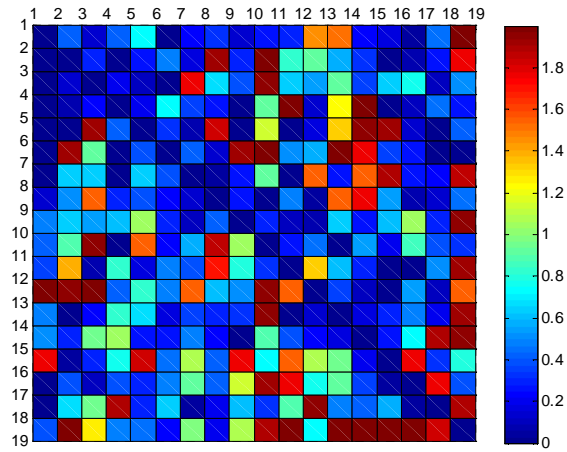
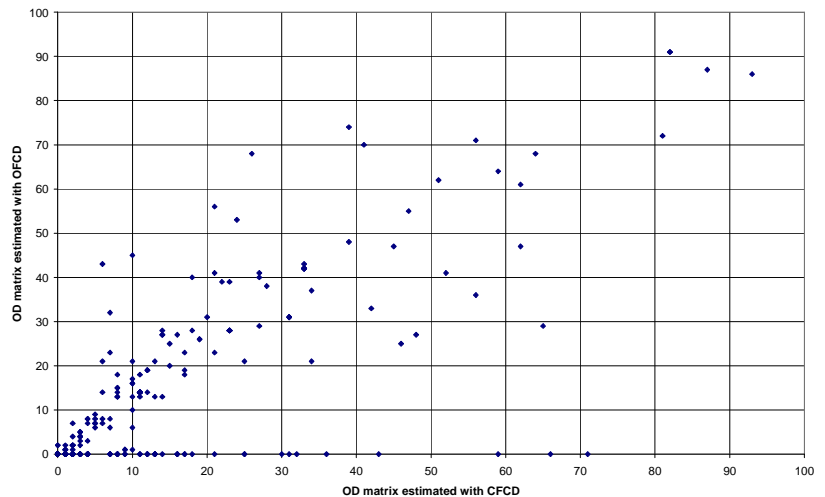
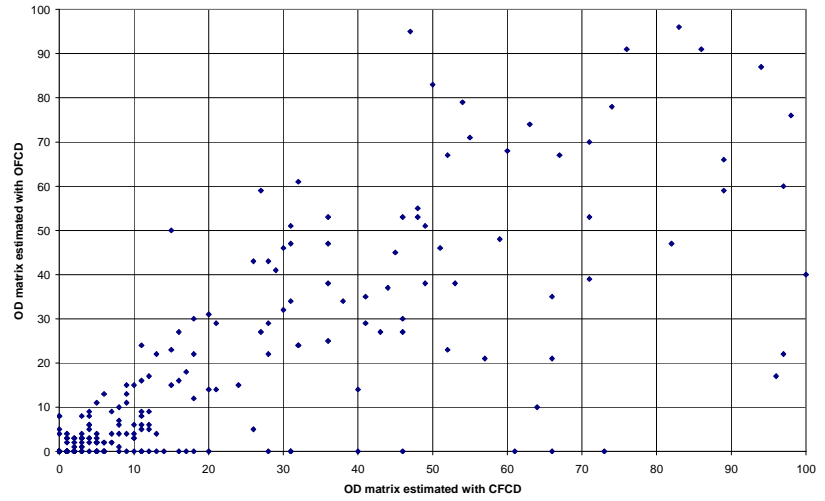


Figure 9.23: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Time period 8:30-8:39





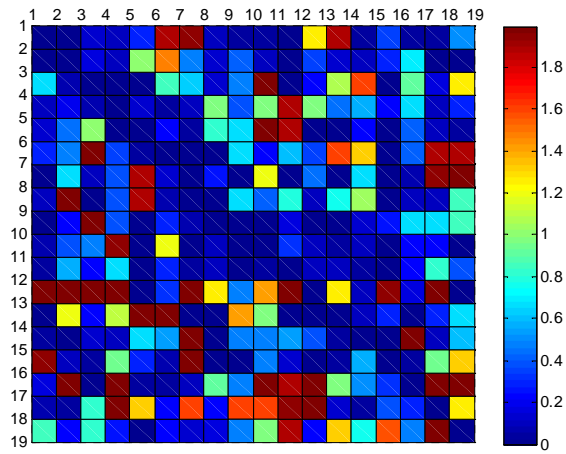
.....  
Figure 9.24: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and OFCD. Total time period



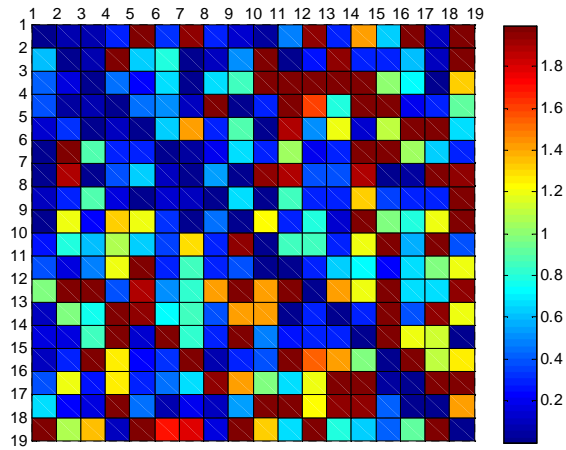
---

## CFCD vs. PFCD

.....  
Figure 9.25: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:30-7:39



.....  
Figure 9.26: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:40-7:49



.....  
Figure 9.27: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:50-7:59

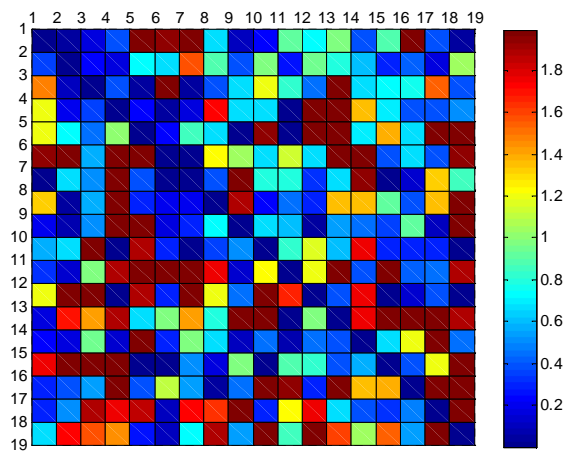


Figure 9.28: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:00-8:09

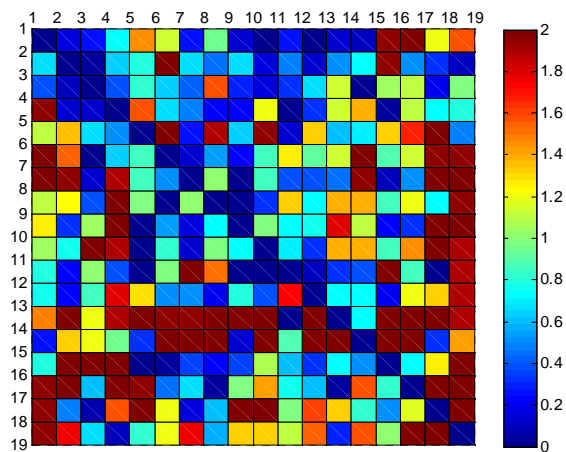


Figure 9.29: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:10-8:19

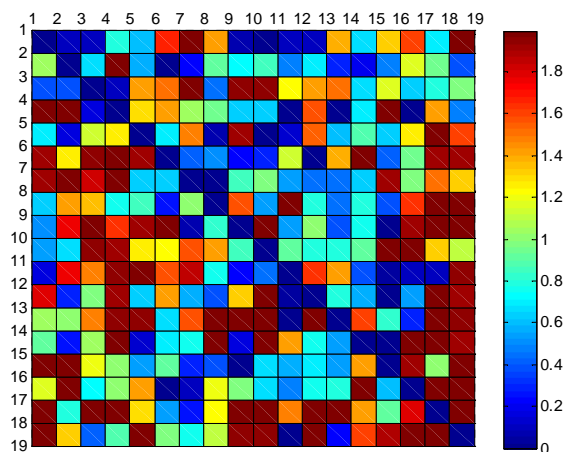


Figure 9.30: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:20-8:29

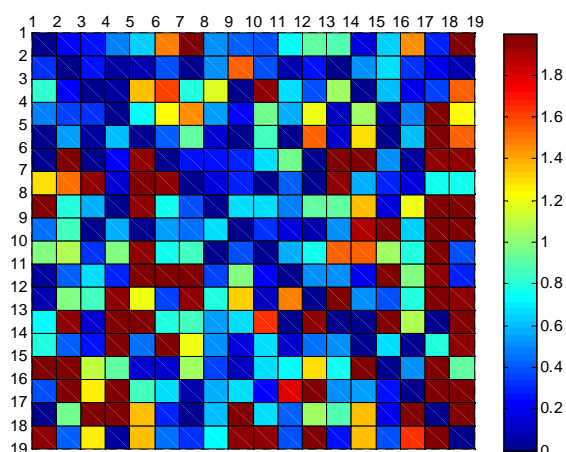


Figure 9.31: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39

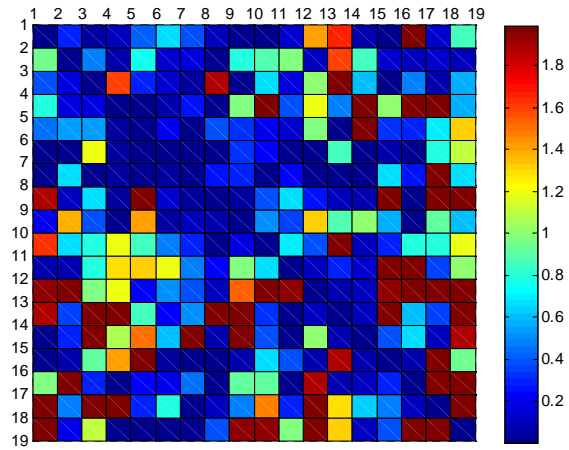


Figure 9.32: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:40-8:49

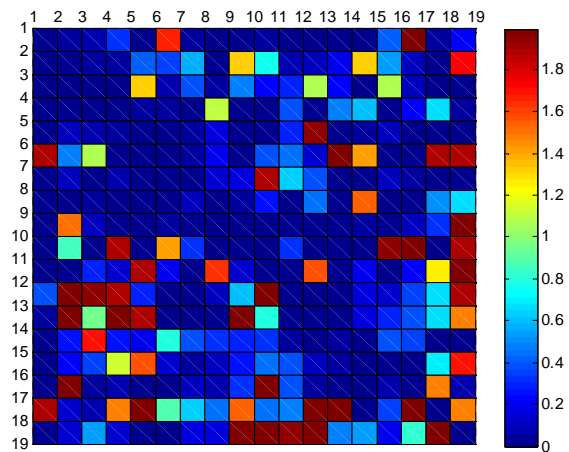


Figure 9.33: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. The whole time period 7:30-8:49

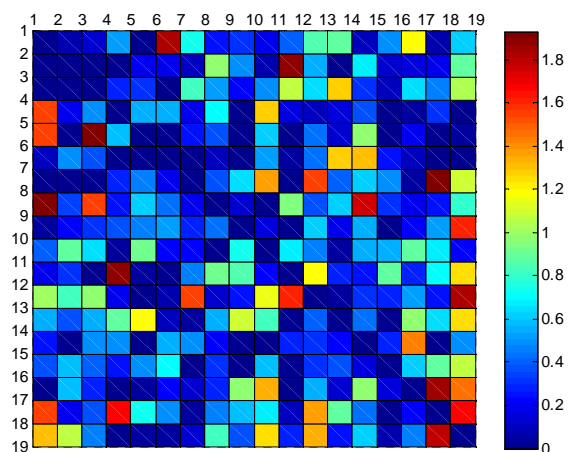


Figure 9.34: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:30-7:39

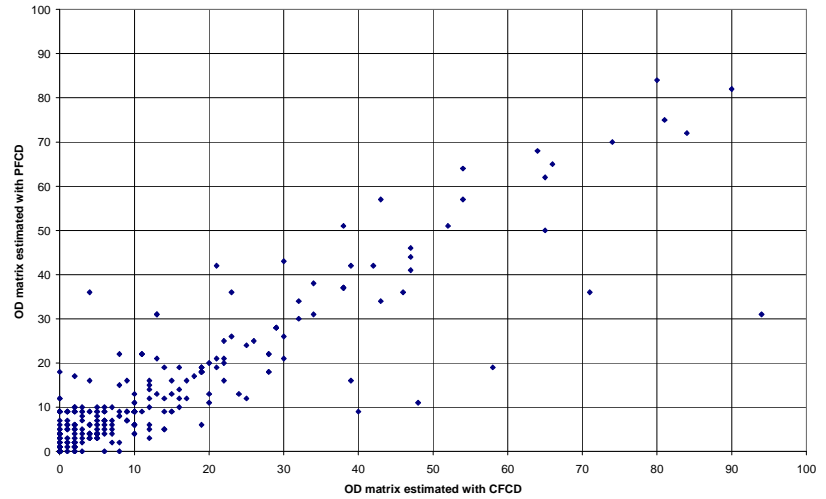


Figure 9.35: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:40-7:49

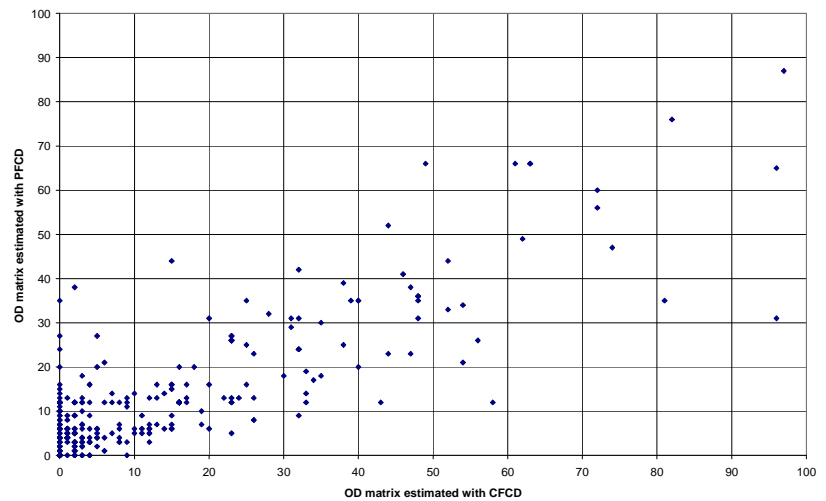


Figure 9.36: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 7:50-7:59

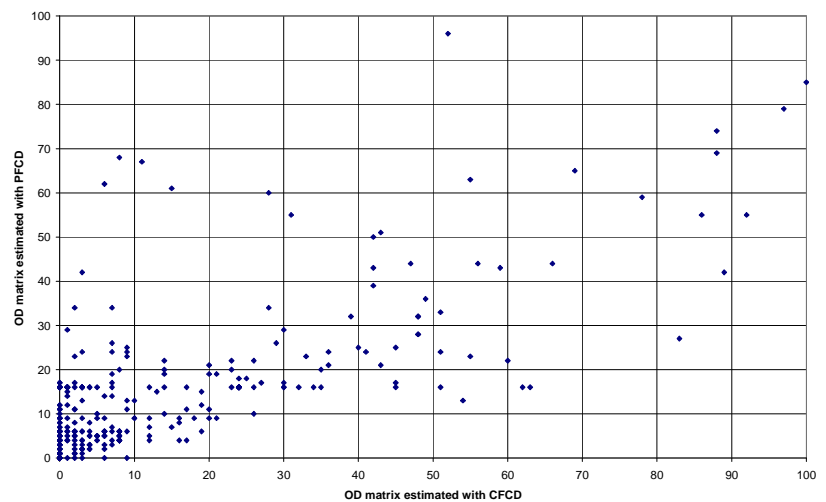


Figure 9.37: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:00-8:09

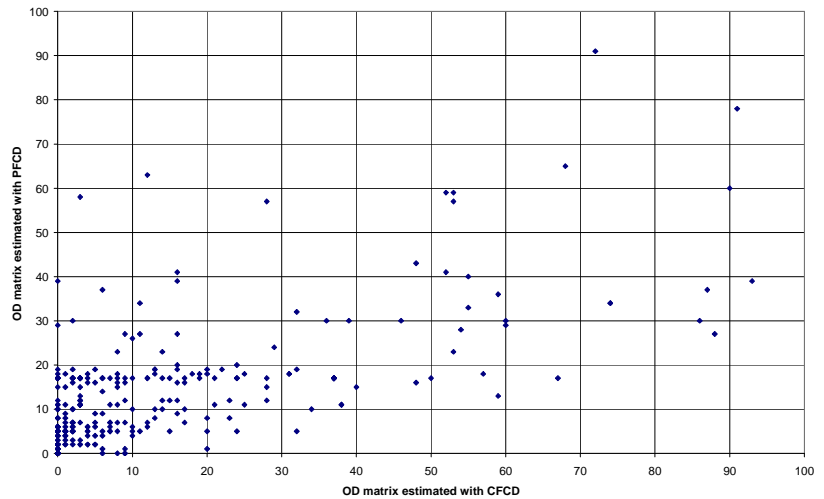


Figure 9.38: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:10-8:19

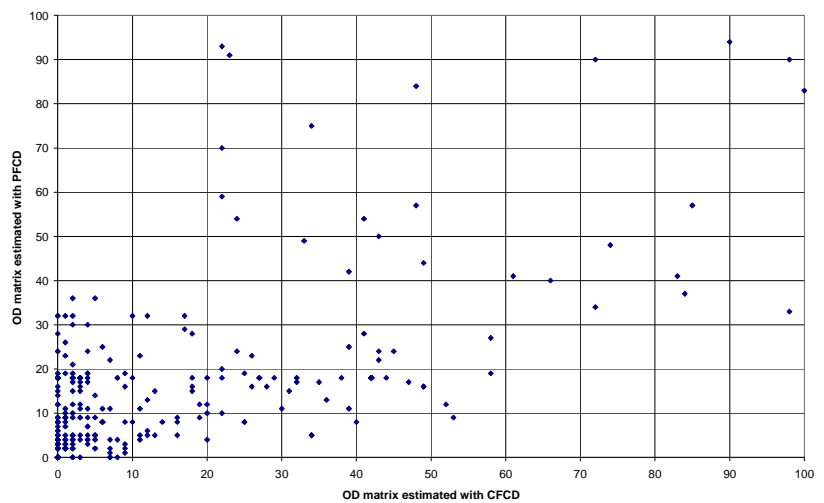


Figure 9.39: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:20-8:29

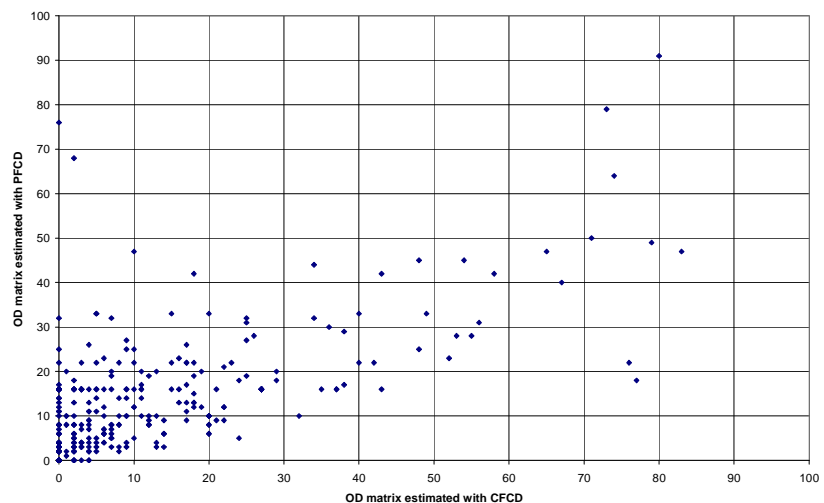


Figure 9.40: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:30-8:39

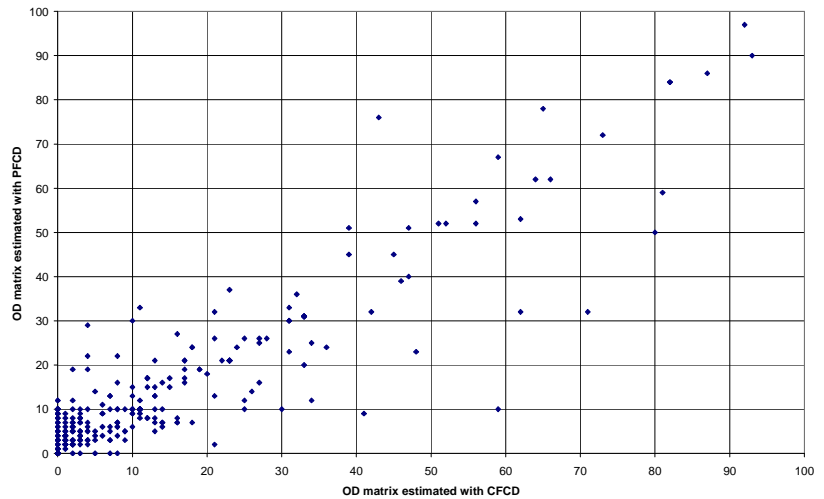


Figure 9.41: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Time period 8:40-8:49

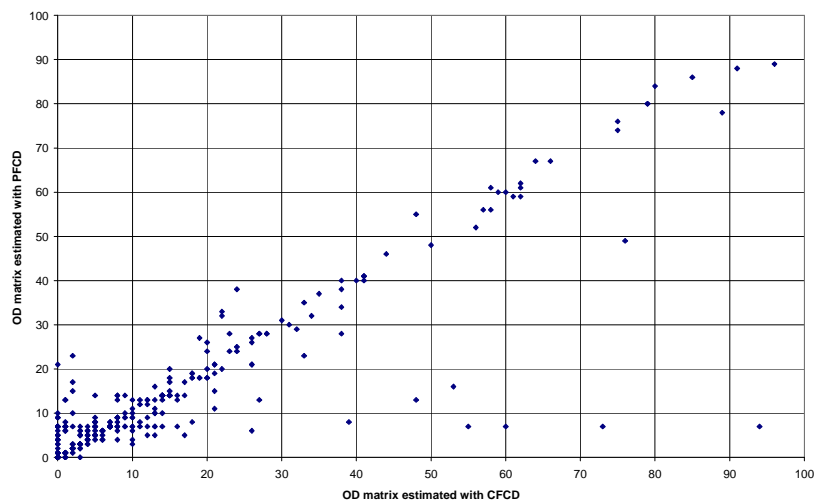
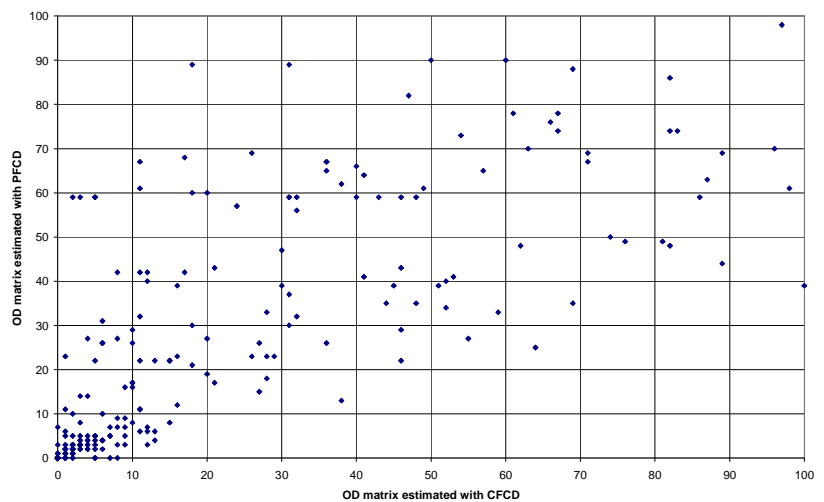


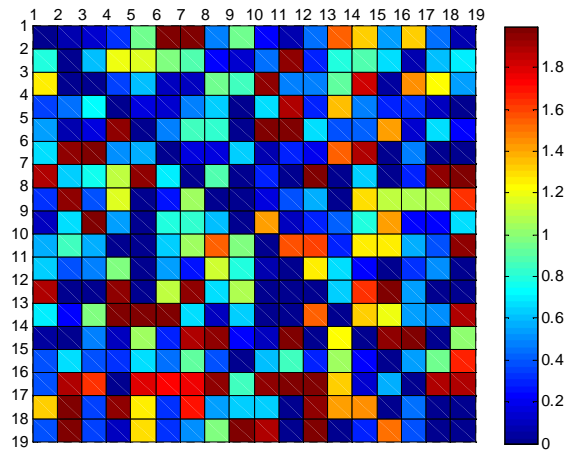
Figure 9.42: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and PFCD. Total time period



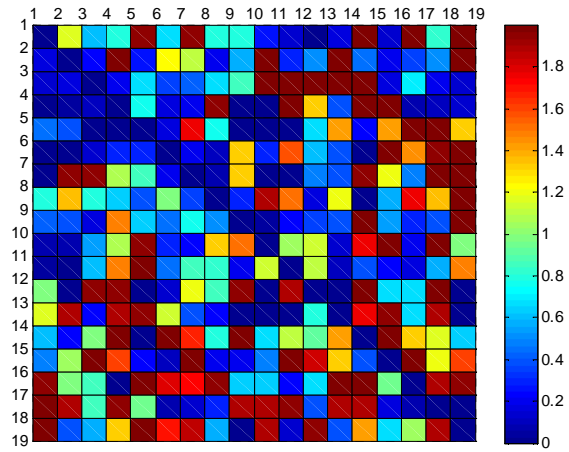
---

## CFCD vs. TC

.....  
Figure 9.43: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:30-7:39



.....  
Figure 9.44: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:40-7:49



.....  
Figure 9.45: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:50-7:59

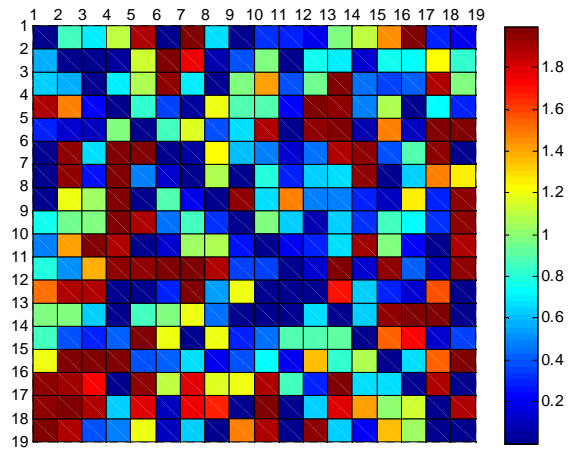




Figure 9.46: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:00-8:09

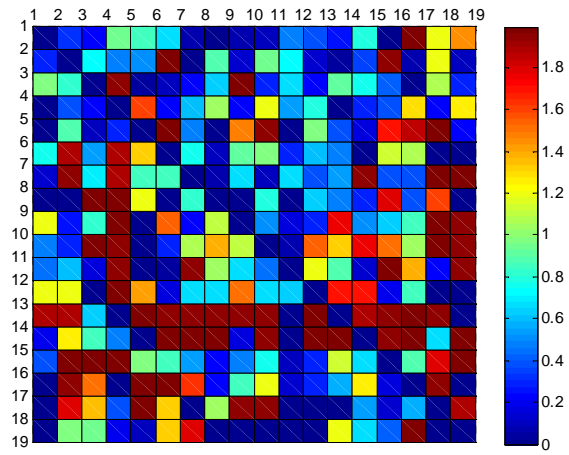


Figure 9.47: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:10-8:19

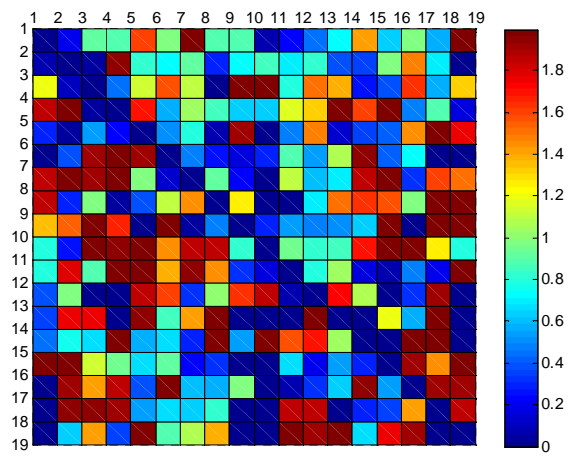


Figure 9.48: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:20-8:29

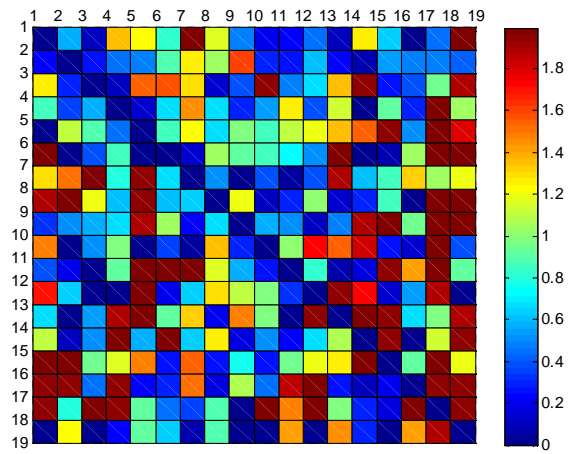


Figure 9.49: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39

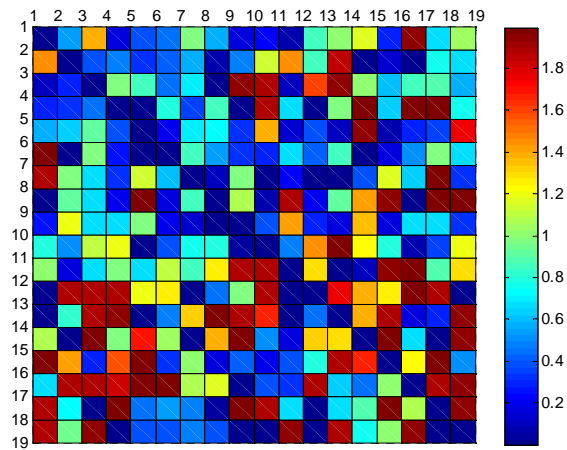


Figure 9.50: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:40-8:49

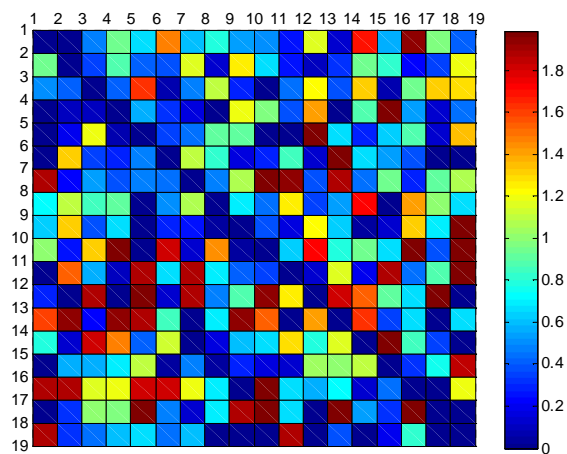


Figure 9.51: Comparison of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Total time period 7:30-8:49

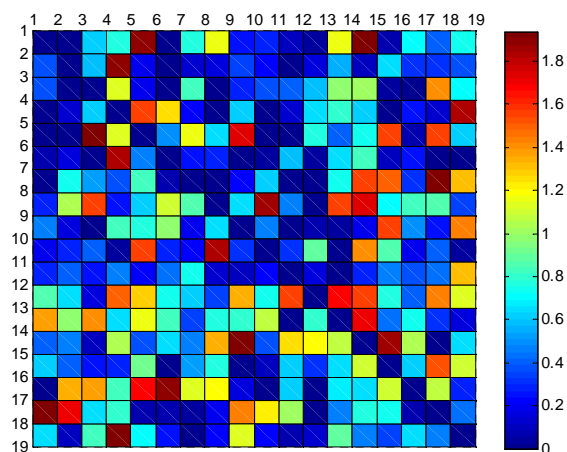


Figure 9.52: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:30-7:39

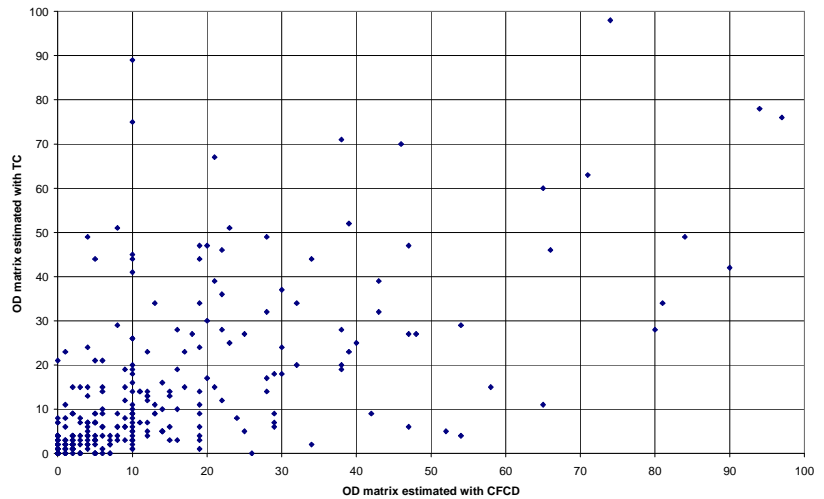


Figure 9.53: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:40-7:49

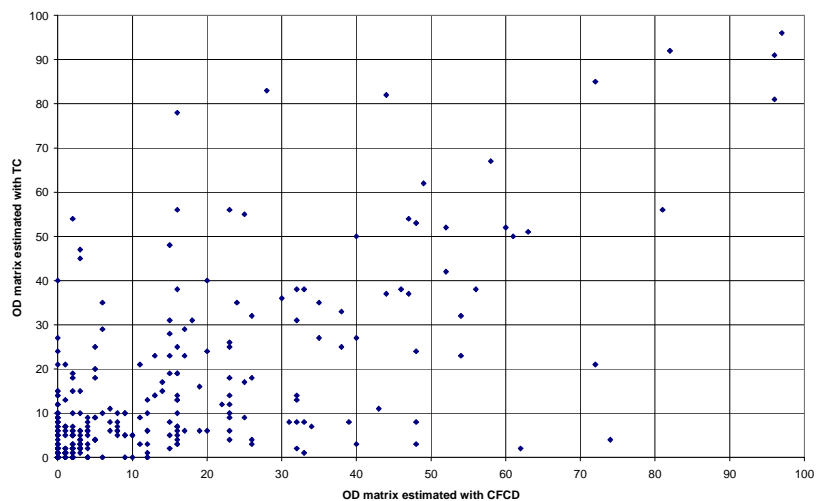


Figure 9.54: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 7:50-7:59

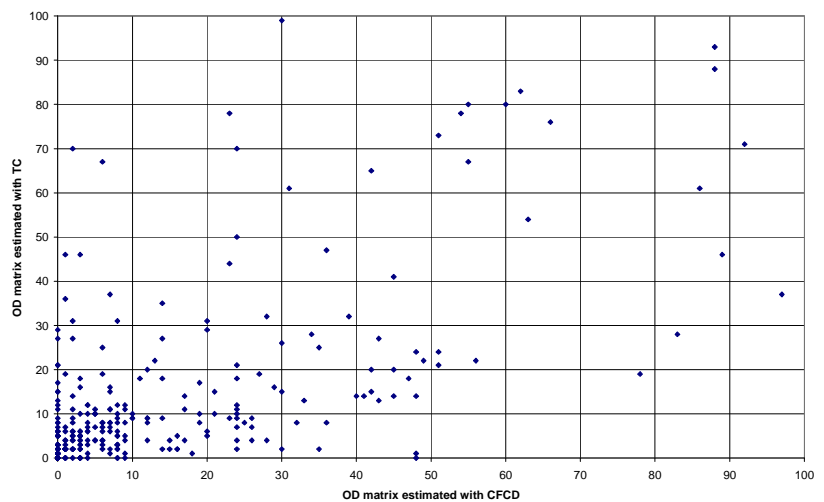


Figure 9.55: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:00-8:09

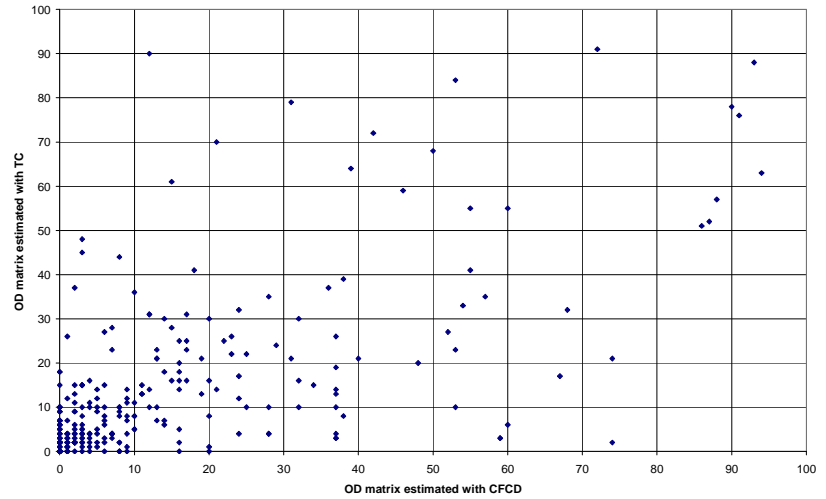


Figure 9.56: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:10-8:09

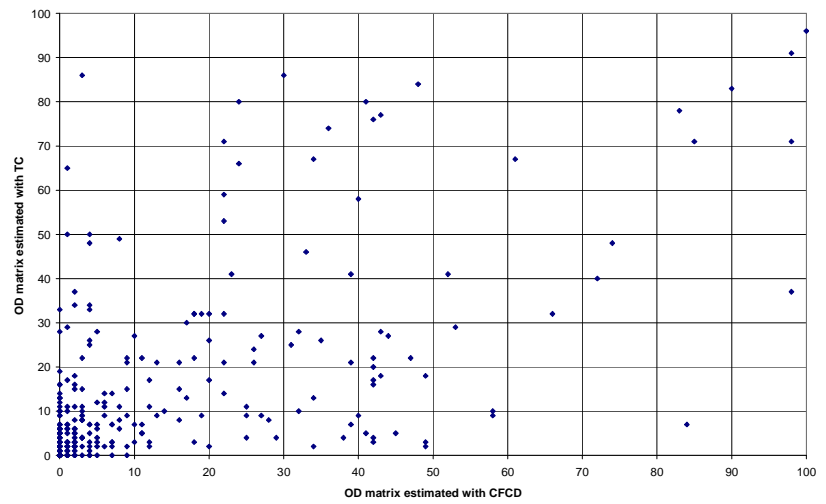


Figure 9.57: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:20-8:29

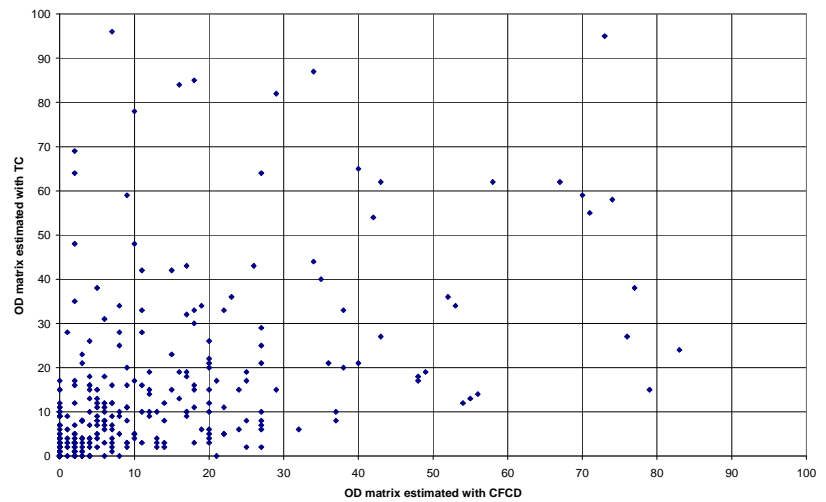


Figure 9.58: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:30-8:39

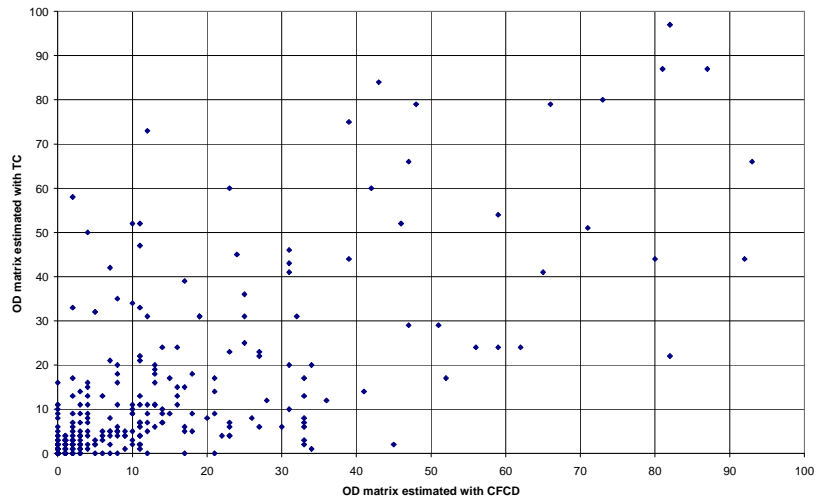


Figure 9.59: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Time period 8:40-8:49

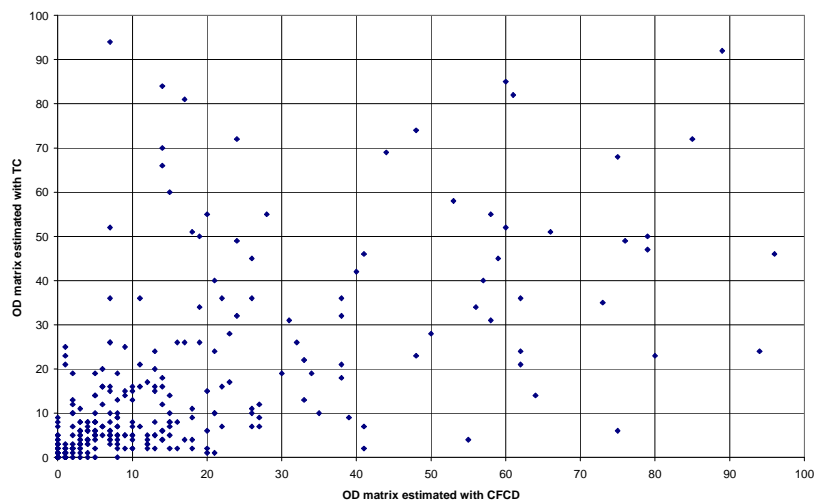
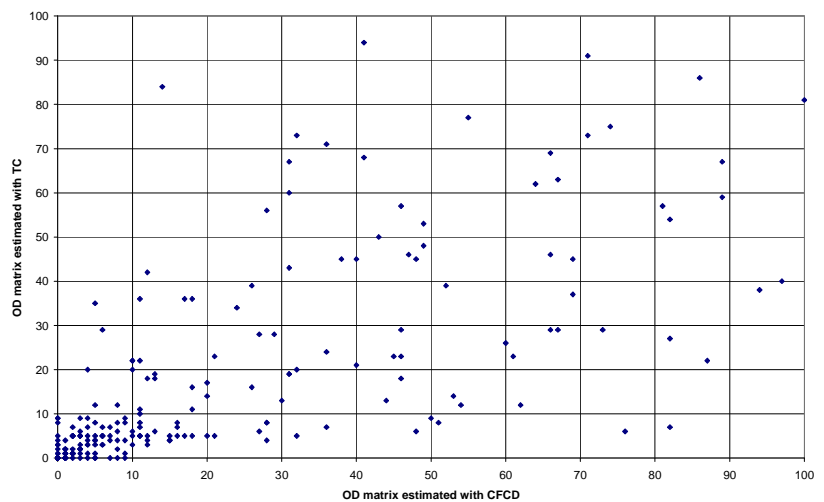


Figure 9.60: Comparison plot of the estimated OD matrices with two different a-priori matrices; CFCD and TC. Total time period



## 9.7 Appendix 7: Comparison of the structure of OD matrices estimated with the developed method and with REMDOE

Figure 9.61: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:30-7:39

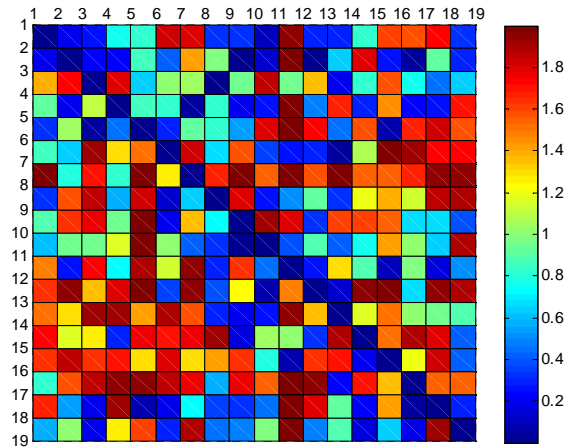


Figure 9.62: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:40-7:49

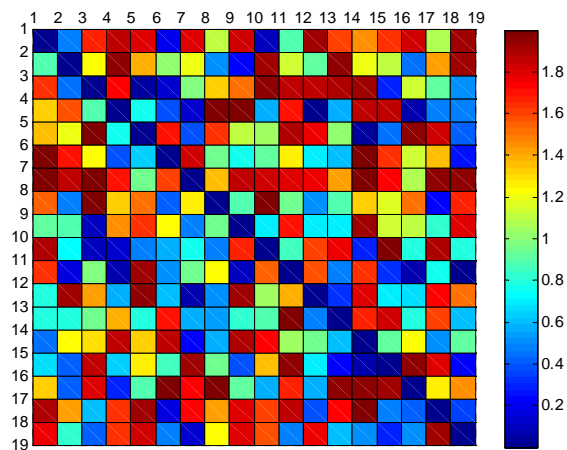
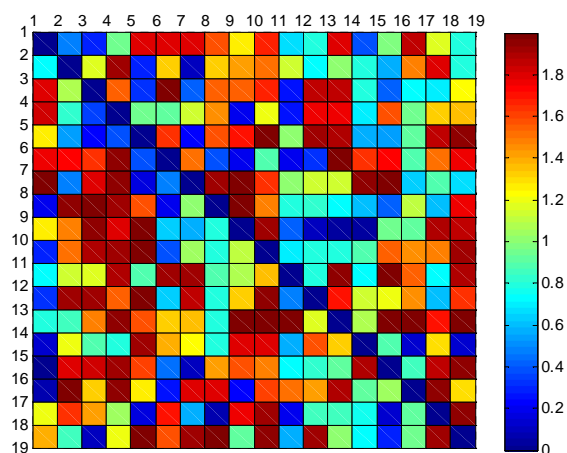
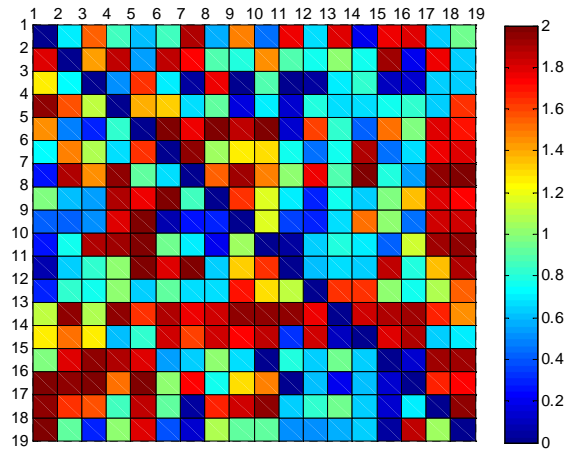


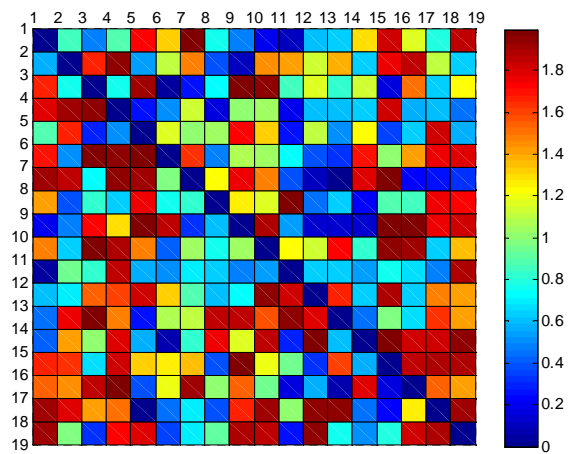
Figure 9.63: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 7:50-7:59



.....  
Figure 9.64: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:00-8:09



.....  
Figure 9.65: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:10-8:19



.....  
Figure 9.66: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:20-8:29

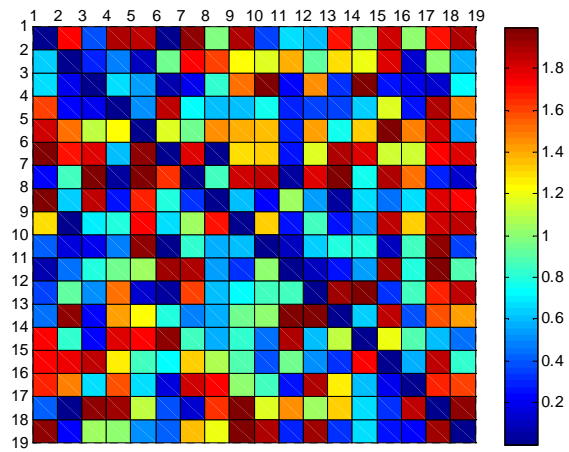


Figure 9.67: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:30-8:39

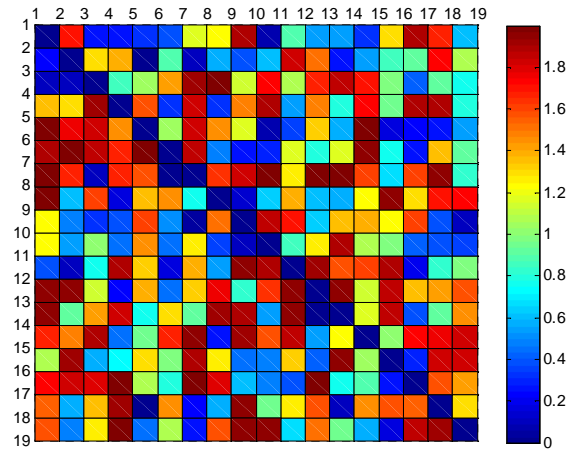


Figure 9.68: Comparison between OD matrix estimated here and OD matrix estimated with REMDOE. Time 8:40-8:49

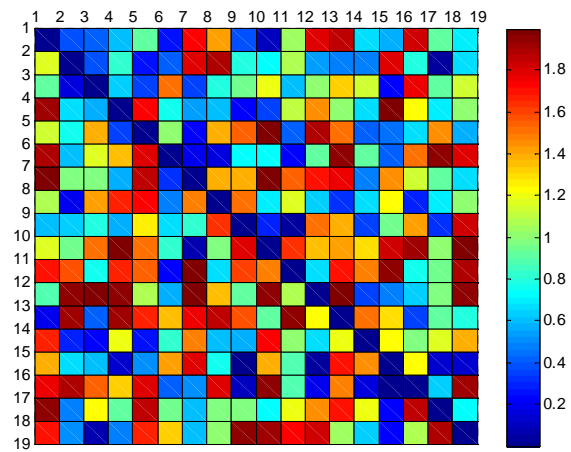


Figure 9.69: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:30-7:39

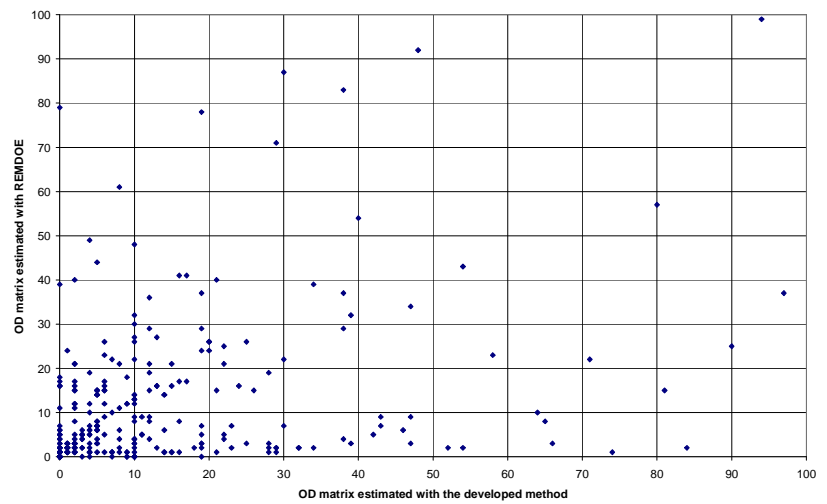




Figure 9.70: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:40-7:49

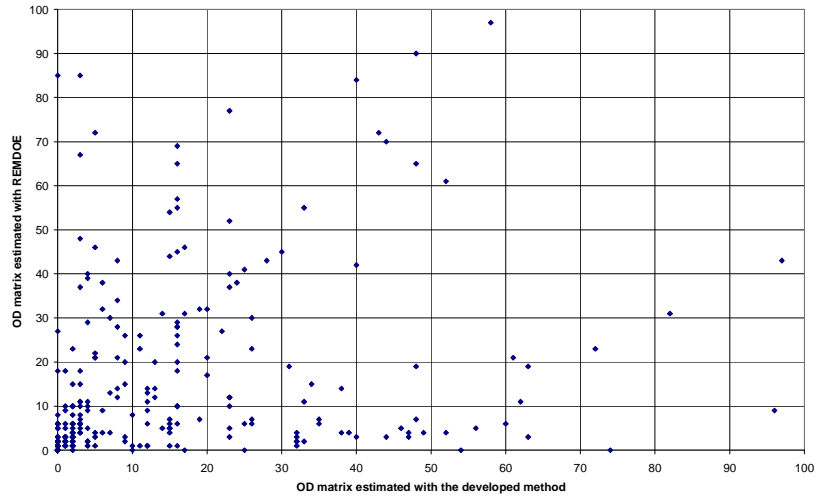


Figure 9.71: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 7:50-7:59

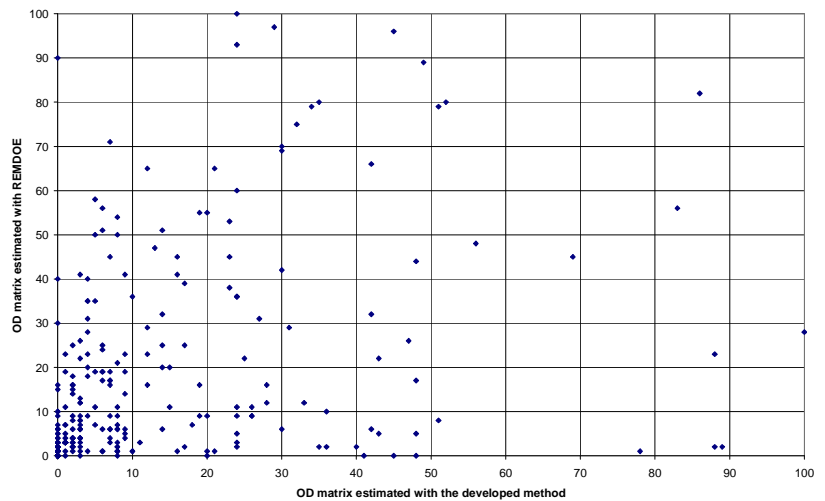


Figure 9.72: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:00-8:09

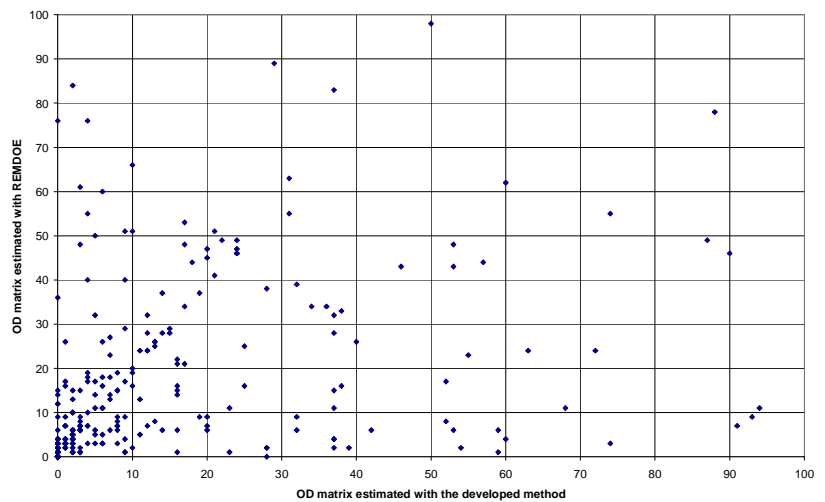


Figure 9.73: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:10-8:19

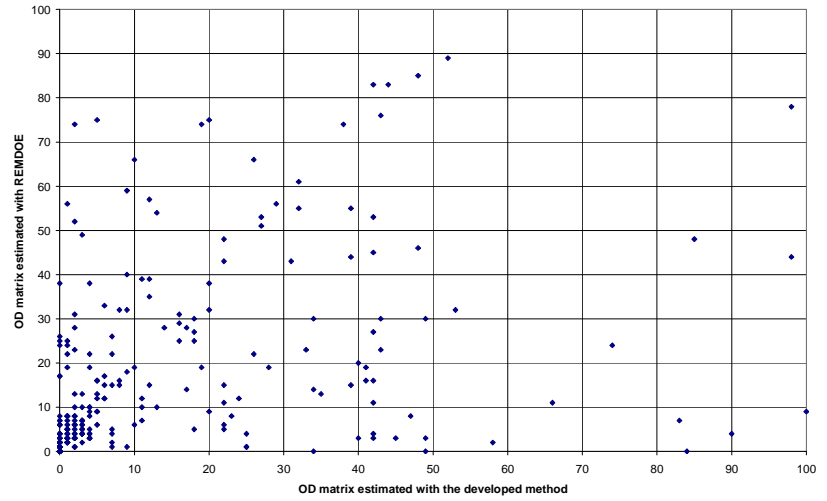


Figure 9.74: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:20-8:29

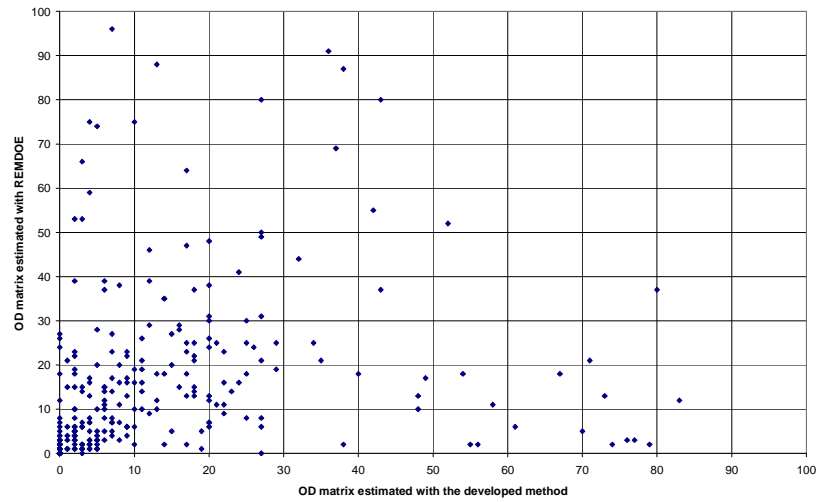


Figure 9.75: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:30-8:39

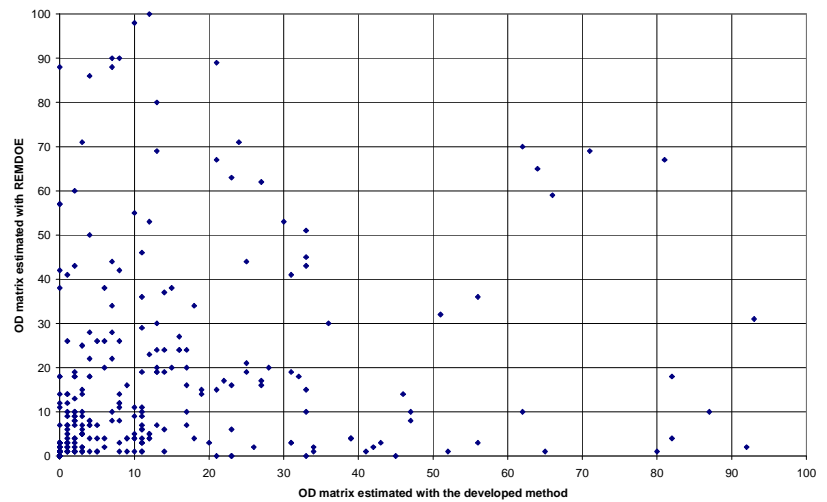


Figure 9.76: Comparison plot for the OD matrix estimated here and the OD matrix estimated with REMDOE. Time 8:40-8:49

