

Intelligent Roads

Project no. 012344

Specific Targeted Research Project
Sustainable Surface Transport, priority 1.6.2

D4.1 Performance Indicators needs and derivation: Single and multisource solutions

Due date of deliverable: 31st October 2006
Actual submission date: 20th September 2007

Start date of project: 1st March 2005 Duration: 36 months

Edward Chung, EPFL, Switzerland

Revision: Final Version 3.0

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Editor

Edward CHUNG (EPFL-LAVOC, Switzerland)

Work Package Leader

Edward CHUNG (EPFL-LAVOC, Switzerland)

Task Leaders

(In alphabetical order by institute)

Bernhard NOWOTNY (ARSENAL Research, Austria)
Nour-Eddin EL FAOUZI (INRETS-LICIT, France)
Rune KARLSSON (VTI, Sweden)

Contributors

(In alphabetical order by institute)

Elisabeth MRAKOTSKY (ARSENAL Research, Austria)
Bernhard NOWOTNY (ARSENAL Research, Austria)

Ashish BHASKAR (EPFL-LAVOC, Switzerland)
Edward CHUNG (EPFL-LAVOC, Switzerland)

Michal KARKOWSKI (FEHRL/IBDiM, Poland)

Maurice ARON (INRETS-GRETIA, France)
Olivier DE MOUZON (INRETS-LICIT, France)
Nour-Eddin EL FAOUZI (INRETS-LICIT, France)
Stéphane ESPIE (INRETS-MSIS, France)
Regine SEIDOWSKY (INRETS-GRETIA, France)

Jean-Marc MORIN (ISIS, France)

Josep Maria AYMAMÍ (TSS, Spain)
Josep PERARNAU (TSS, Spain)
Alexandre TORDAY (TSS, Spain)

Rune KARLSSON (VTI, Sweden)

ARSENAL RESEARCH
Geschäftsfeld Verkehrstechnologien /
Business Unit of Transport Technologies
Austria, 1030 Vienna, Faradaygasse 3, Object 219C
Tel: +43 (0) 50550-6338
Fax: +43 (0) 50550-6439
E-mail bernhard.nowotny@arsenal.ac.at
www.arsenal.ac.at

SWISS FEDERAL INSTITUTE OF TECHNOLOGY, LAUSANNE
EPFL-ENAC-ICARE-LAVOC
Station 18
CH- 1015 Lausanne
Switzerland
Tel +41 (0) 21 693 23 43
Fax +41 (0) 21 693 63 49
E-MAIL: edward.chung@epfl.ch
http://lavoc.epfl.ch/

ROAD AND BRIDGE RESEARCH INSTITUTE
Jagiellonska Street 80
Warsaw
3-301 Poland
Tel +48 (0) 22 811 29 36
Fax +48 (0) 22 811 17 92
E-MAIL: michal@ibdim.edu.pl
http://www.ibdim.edu.pl/

THE FRENCH NATIONAL INSTITUTE FOR TRANSPORT AND SAFETY RESEARCH
2 avenue du Général Malleret-Joinville
F-94114 ARCUEIL
Tel +33 (0)1 47 40 70 00
Fax +33 (0)1 45 47 56 06
E-MAIL: nour-eddin.elfaouzi@inrets.fr
http://www.inrets.fr

ISIS
78 rue de la Villette
F-69425 Lyon
Tel: +33 (0)4 78 71 89 65
Fax: +33 (0)4 78 71 89 64

E-MAIL: jm.morin@isis.tm.fr
www.isis-its.com

TSS-TRANSPORT SIMULATION SYSTEMS
Pg. de Gràcia 12
08007 Barcelona
Spain
Tel +34 (0) 933 171 693
Fax +34 (0) 933 171 716
E-MAIL: torday@aimsun.com
www.aimsun.com

SWEDISH NATIONAL ROAD AND TRANSPORT RESEARCH INSTITUTE
Olaus Magnus väg 35,
581 95 Linköping
Tel +46 (0) 13 20 41 35
Fax +46 (0) 14 14 36
E-MAIL: rune.karlsson@vti.se
http://www.vti.se/

Reviewed by:

Rigobert Optiz (ROC, Germany)
Jaime Ferrer (TSS, Spain)
Sigvard Tim (VTI, Sweden)

Final control:

Released by: Bengt Wälivaara (VTI, Sweden)

Signature:

Date: September 25:th 2007

Version	Date	Prepared by	Checked by	Modifications
1.0	28th October	E. Chung	Rigobert Optiz	1st draft sent to reviewers
1.4	21st November	E. Chung	Jaime Ferrer	2nd draft sent to reviewers
1.6c	27th November	E. Chung	Sigvard Tim	Final draft sent to QA manager
3.0	20th September	E. Chung		Chapter 3 by TSS, Arsenal & VTI

Table of Contents

Editor 2
Work Package Leader 2
Task Leaders 2
Contributors 2
Reviewed by: 4
Table of Contents 5
List of Abbreviations 7
List of Figures 8
Executive Summary 12
Key traffic and safety indicators 12
Travel time prediction using floating car data 12
Identified floating car 13
Unidentified floating car 13
Data fusion 14
1 Introduction 16
1.1 INTRO Project 16
1.2 Purpose of WP4 Traffic and Safety Monitoring 16
1.3 Structure of this report 17
2 Traffic and Safety Indicators 18
2.1 Inventory of indicators 19
2.2 Indicators related to road network global conditions (weather, road surface) 22
2.3 Indicators related to traffic demand 22
2.4 Indicators related to supply 22
2.5 Indicators related to the demand/supply interaction, i.e. the quality of flow 24
2.6 Indicators related to safety 32
2.7 Conclusion 36
2.8 References 36
3 Floating car based indicators 38
3.1 Estimation of travel times with FCD 38
3.1.1 Definition and concept of travel time 38
3.1.2 Travel time estimation 42
3.2 Identified Floating Car 45
3.2.1 Introduction 45
3.2.2 Methods 45
3.2.3 Comparison of historical or model timeseries 47
3.2.4 Investigation of variances of route speed and effect of linear Kalman filter 47
3.2.5 Comparison of floating car route speeds and AVI route speeds 50
3.2.6 Conclusion 50
3.3 Unidentified Floating Car 53
3.3.1 Introduction 53
3.3.2 Test bed 54
3.3.3 Application of the unidentified floating car 56
3.3.4 Development of new algorithms for travel time estimation 56
3.3.5 Results for simulated data from Gothenburg network 67
3.3.6 Results for Vienna floating car data 76
3.3.7 Conclusions 79
3.4 Floating Car Travel Time Estimation for Urban Network 82
3.4.1 Introduction 82
3.4.2 Equipment ratio 82
3.4.3 Link definition 83
3.4.4 Aggregation period 87
3.4.5 Combined effect of the aggregation period and link definition choice 88
3.4.6 Substitution method for missing 89
3.4.7 Handling missing data 90

3.5 Simulation 95
3.5.1 Introduction 95
3.5.2 Use of microsimulation 95
3.5.3 Use of AIMSUN 96
3.5.4 Network description: The Lausanne network 96
3.5.5 Experiment description 100
3.5.6 Description of studied scenarios 101
3.6 Comparison of travel time results from identified and unidentified floating car data 104
3.6.1 Methodology and task organisation 104
3.6.2 Performance Indicators 104
3.6.3 Comparison of results for Lausanne Centre 106
3.6.4 Results comparison for Lausanne agglomeration 121
3.6.5 Conclusions 125
3.7 High Resolution Floating Car Data 127
3.7.1 Introduction 127
3.7.2 Test bed 127
3.7.3 Investigation of the relation between high resolution floating data and the traffic state 129
3.7.4 Conclusion 133
3.7.5 References 133
4 Data Fusion for Traffic Indicators/Travel Time 135
4.1 Introduction 135
4.2 Data Fusion: Concepts and methods 135
4.2.1 Basic concepts 135
4.2.2 Main techniques 144
4.2.3 Conclusion 151
4.3 Data Fusion in Transportation (Traffic and Safety) 151
4.3.1 Introduction 151
4.3.2 Opportunities and challenges of Road Traffic Data fusion 152
4.3.3 Data fusion in Road Traffic engineering 153
4.3.4 Conclusions and prospective analysis 158
4.4 Data Fusion for Traffic Monitoring Indicators 158
4.4.1 Data fusion for travel time 159
4.4.2 Estimation of dynamic OD matrices based on floating car data 194
4.5 Final Conclusion on Data Fusion for Traffic Indicators 207
4.6 References 207
5 Conclusions 214
6 Annexe I - Overview of traffic performance indicators and safety indicators 217
References 226
Literature related to traffic performance indicators and safety indicators 226
7 Annex II 233

List of Abbreviations

3-D	3 dimension.
ACC	Adaptive Cruise Control.
ADAS	Advanced Driver Assistance Systems.
AID	Automatic Incident Detection.
ALT	Average Link Travel Time.
AP	Aggregation Period.
APRR	Autoroute Paris-Rhin-Rhône. A French toll motorway company.
APT	Average Path Travel Time.
ATIS	Advanced Traveller Information Systems.
ATMS	Advanced Traffic Management Systems.
AVI	Automatic Vehicle Identification.
BPT	Built Path Travel Time.
CAS	Collision Avoidance System.
CCTV	Closed-Circuit Television.
CTC	Cumulative Traffic Counts. This method, used for travel time estimation, is also named stock method.
Dn.m	Deliverable number m of WPn.
DF	Data Fusion.
DTA	Dynamic Traffic Assignment.
EKF	Extended Kalman Filter.
ETC	Electronic Toll Collection.
FCD	Floating Car Data.
FDSA	Finite Difference Stochastic Approximation.
GPS	Global Positioning System.
HGV	Heavy Goods Vehicle.
INRETS	Institut National de REcherche sur les Transports et leur Sécurité. The French National Research Institute on Transport and Security is a member of INTRO.
INTRO	INTelligent ROads. Name of this European project.
ITS	Intelligent Transportation Systems.
MCTC	Modified CTC.
OD	Origin-Destination.
QoS	Quality of Service.
RA	Rest Areas.
RMTS	Remote Traffic Microwave Sensor.
RRSE	Root Relative Squared Error.
SA	Service Areas.
SA	Stochastic Approximation.
SE	Speed Extrapolation. This method, used for travel time estimation, is also named speed method.
SPSA	Stochastic Perturbation Stochastic Approximation.
TPI	Temps de Parcours Instantané. Instantaneous TT. Tool developed by ISIS, member of INTRO.
TMC	Traffic Management Centre.
Tn.m	Task number m of WPn.
TT	Travel Time.
UKF	Unscented Kalman Filter.
UT	Unscented Transformation.
WPn	Work Package number n (of INTRO).

List of Figures

Fig. 2-1: Example for analysis of indicators 19
Fig. 2-2: Interaction of parameters / indicators 21
Fig. 2-3: Forecasting travel times based on stock method 28
Fig. 2-4: Correction of link travel times 30
Fig. 2-5: Calculation of queue density 31
Fig. 3-1: Difference between BPT (red) and APT (blue) calculation 40
Fig. 3-2: Aggregation of LT data for averages or medians for 15 minutes [1] 41
Fig. 3-3: Calibration of time series model for historical and current speed values (route 29) 48
Fig. 3-4: Calibration of time series model for historical and current speed values (route 69) 49
Fig. 3-5: Timeseries of single vehicles speeds in (route 29, 2005-09-28, 14:00 to 18:45) 51
Fig. 3-6: Timeseries of single vehicle speeds in (route 69, 2005-09-28, 14:00 to 18:45) 51
Fig. 3-7: Comparison of route speeds (route 29, 2005-09-28, 14:00 to 18:45) 52
Fig. 3-8: Comparison of route speeds (route 69, 2005-10-06, 14:00 to 18:00) 52
Fig. 3-9: The Gothenburg network 55
Fig. 3-10: Method suggested and investigated by Westman [22]. O1, O2, O6 denote observations from unidentified probe vehicles. 56
Fig. 3-11: Illustration of the method proposed by Brorsson [21] 57
Fig. 3-12: Illustration of θ-functions 60
Fig. 3-13: Two different examples of velocity profiles along links 61
Fig. 3-14: Illustration of the asymmetry problem 62
Fig. 3-15: θ-function method for low velocities. The blue line denotes the θ-function and the red line denotes the true velocity profile 63
Fig. 3-16: The definition of true travel time during the time slice $[t_2, t_3]$ is consistent with the probability that an observation (in a position on link AB) occurs during this time slice 66
Fig. 3-17: Cumulative diagrams for the relative error in travel time estimates based on single observations 68
Fig. 3-18: Cumulative diagrams for the relative error in travel time for the observations in the network 70
Fig. 3-19: Cumulative diagram for the relative error for various spatial combination methods. 71
Fig. 3-20: Estimated travel time versus exact travel time for different estimation methods. 72
Fig. 3-21: Comparisons between constant velocity method (Harmonic flat) and θ-function method (Harmonic theta) 77
Fig. 3-22: Travel time estimates for various lengths of aggregation periods (Equipment ratio is 5% and Harmonic theta method has been used) 78
Fig. 3-23: Travel time estimates for various equipment ratios (In all cases Harmonic theta has been used. In a) results for Route 29 and in b) for Route 69 are shown) 79
Fig. 3-24: Classical representation of link and nodes 83
Fig. 3-25: Individual link travel time records for a link with signalized intersections upstream and downstream [1] 84
Fig. 3-26: Individual link travel time records gathered per entrance and exit links pairs [1] 85
Fig. 3-27: Subdivided representation of links 86
Fig. 3-28: Short and medium term variability cumulated 88
Fig. 3-29: Average standard deviation of the individual travel time records populations for the Lausanne centre network 89
Fig. 3-30: Percentage of missing data in function of the aggregation period (AP) and the equipment ratio (classical link definition on left and in-out one on right) [1] 90
Fig. 3-31: Evolution of the relative approximation error (with ER=1%) and standard deviation of the record population in function of the number of recording days [1] 92
Fig. 3-32: Lausanne Centre Network 97
Fig. 3-33: Lausanne Agglomeration Network 98
Fig. 3-34: Example of stability with 80% flow in Lausanne centre network 101
Fig. 3-35: Flow-chart for comparison process of simulated FCD 102
Fig. 3-36: Sensitivity of MSLE to equipment ratio and reporting interval 107

Fig. 3-37: Sensitivity of MSLE to equipment ratio and reporting interval (source [1]) 108
Fig. 3-38: Sensitivity of MALE to equipment ratio and reporting interval 109
Fig. 3-39: Average network travel time for ER=10% and RI=15 sec 111
Fig. 3-40: Average network travel time for ER=5% and RI=30 sec 111
Fig. 3-41: Sensitivity of AILE to equipment ratio and reporting interval (source [1]) 112
Fig. 3-42: Sensitivity of AILE to equipment ratio and reporting interval 113
Fig. 3-43: Sensitivity of AIPE to equipment ratio and reporting interval (source [1]) 114
Fig. 3-44: Sensitivity of AIPE to equipment ratio and reporting interval 115
Fig. 3-46: Error ranges for absolute travel time difference [s] 116
Fig. 3-47: Link-to-link differences between simulated and estimated (UFCS) 120
Fig. 3-48: Sensitivity of MSLE to equipment ratio and reporting interval 122
Fig. 3-49: Sensitivity of AILE to equipment ratio and reporting interval 123
Fig. 3-50: Sensitivity of AIPE to equipment ratio and reporting interval 124
Fig. 3-52: Map describing the southern 8 km part of the test bed 128
Fig. 3-53: Examples of speed profiles and associated traffic state parameters 130
Fig. 3-54: Examples of wavelet decompositions of a speed profile 132
Fig. 4.1: Fusion vs. fission (Adapted from [7]) 137
Fig. 4.2: Integration (Adapted from [7]) 139
Fig. 4.3: Fusion 140
Fig. 4.4: Aggregation 140
Fig. 4.5: Generic process of fusion in 4 consecutive steps 141
Fig. 4.6: Parallel architecture (Adapted from [7]) 142
Fig. 4.7: Serial architecture (Adapted from [7]) 142
Fig. 4.8: Main theories in data fusion and their relationships (Adapted from [7]) 144
Fig. 4.9: Scenarios 1-4 161
Fig. 4.10: Origin-Destination counts from 0:00 am to 12:00 pm, scenario 2, 24/11/2005 162
Fig. 4.11: From toll stamps to a reference travel time 163
Fig. 4.12: Reference travel time by vehicle category 164
Fig. 4.13: Loop detectors flows, scenario 1 165
Fig. 4.14: Loop detectors speeds, scenario 1 165
Fig. 4.15: Enriched reference travel time compared to direct-link data, scenario 1 166
Fig. 4.16: Speed Extrapolation scenario 1 167
Fig. 4.17: CTC (off-line) compared to the reference, scenario 1 168
Fig. 4.18: MCTC (on-line) using direct link data, scenario 1 169
Fig. 4.19: MCTC (on-line) compared to the reference, scenario 1 170
Fig. 4.20: Travel time estimators: Comparisons in congestion / free flow, scenario 1 170
Fig. 4.21: Effect of fusion on on-exit travel times, scenario 1 172
Fig. 4.22: Effect of fusion on on-start travel times (real time travel time), scenario 1 173
Fig. 4.23: Loop detectors flows, scenario 2 174
Fig. 4.24: Loop detectors speeds, scenario 2 174
Fig. 4.25: Enriched reference travel time compared to direct-link data, scenario 2 175
Fig. 4.26: Speed Extrapolation scenario 2 176
Fig. 4.27: CTC (off-line) compared to the reference, scenario 2 176
Fig. 4.28: MCTC (on-line) using direct link data, scenario 2 177
Fig. 4.29: MCTC (on-line) compared to the reference, scenario 2 178
Fig. 4.30: Travel time estimators: Comparisons in congestion / free flow, scenario 2 178
Fig. 4.31: Loop detectors flows, scenario 3 179
Fig. 4.32: Loop detectors speeds, scenario 3 180
Fig. 4.33: Enriched reference travel time compared to direct-link data, scenario 3 181
Fig. 4.34: Speed Extrapolation scenario 3 182
Fig. 4.35: CTC (off-line) compared to the reference, scenario 3 183
Fig. 4.36: MCTC (on-line) using direct link data, scenario 3 184
Fig. 4.37: MCTC (on-line) compared to the reference, scenario 3 184
Fig. 4.38: Relaxed MCTC (on-line) using direct link data, scenario 3 185
Fig. 4.39: Relaxed MCTC (on-line) compared to the reference, scenario 3 186
Fig. 4.40: Travel time estimators: Comparisons in congestion / free flow, scenario 3 187
Fig. 4.41: Loop detectors flows, scenario 4 188

Fig. 4.42: Loop detectors speeds, scenario 4 189
Fig. 4.43: Enriched reference travel time compared to direct-link data, scenario 4 189
Fig. 4.44: Speed Extrapolation scenario 4 190
Fig. 4.45: CTC (off-line) compared to the reference, scenario 4 191
Fig. 4.46: MCTC (on-line) using direct link data, scenario 4 192
Fig. 4.47: MCTC (on-line) compared to the reference, scenario 4 192
Fig. 4.48: Travel time estimators: Comparisons in congestion / free flow, scenario 4 193
Fig. 4.49: Comparison of travel time estimators, all scenarios (congestion and free flow) 194
Fig. 4.50: DTA overview 195
Fig. 4.51: Ring subnetwork representation 202
Fig. 4.52: Calibrated speed-density relationship 203
Fig. 4.53: Calibration results (6am-7am) 205
Fig. 4.54: Calibration results (7am-8am) 205
Fig. 4.55: Calibration results (8am-9am) 206
Fig. 4.56: Calibrated link flows 206

List of Tables

Table 3-1 Differences between urban and interurban network characteristics 42
Table 3-2: Comparisons between travel time estimates using the θ- function method and the constant function method for different traffic loads. The percentages denote percentage of peak hour traffic load 69
Table 3-3: Absolute error per observation for travel time estimates for different classes of T and $\tau(x)$. Travel time estimates have been computed using θ-functions 69
Table 3-4 Relative error per observation for various equipment ratios and for different estimation methods 73
Table 3-5: Relative error per observation for various reporting intervals and for different estimation methods 73
Table 3-6: Relative error per observation for various period lengths and for different estimation methods 73
Table 3-7 The average relative error for travel time estimates using harmonic average with a fixed number of observations 74
Table 3-8 The average relative error for travel time estimates using a combination of periods and various values for α. Time period lengths of 5 minutes have been used 75
Table 3-9 Average relative errors in travel time for various levels of disturbances in positions for the observations 75
Table 3-10: Average relative error of travel time estimates for different perturbations in dense a priori function 76
Table 3-11: Average relative error of travel time estimates during peak hour traffic for various levels of dense a priori functions. 76
Table 3-12: Cost function (MSE) value for the combination method neural network 94
Table 3-13: Basic characteristics for Lausanne networks 97
Table 3-14: Basic parameters for Microsimulation 98
Table 3-15: Route choice for Lausanne networks 99
Table 3-16: Vehicles characteristics for Lausanne networks 100
Table 3-17: Experiments definition 103
Table 3-18: Sensitivity of MSLE to equipment ratio and reporting interval 107
Table 3-19: Sensitivity of MALE to equipment ratio and reporting interval 109
Table 3-20: AILE for UFCS and FCE 113
Table 3-21: Sensitivity of AIPE to equipment ratio and reporting interval 115
Table 3-22: Sensitivity of MSLE to equipment ratio and reporting interval 122
Table 3-23: Sensitivity of AILE to equipment ratio and reporting interval 123
Table 3-24: Sensitivity of AIPE to equipment ratio and reporting interval 124
Table 3-25: MSLE for 5, 10 & 15 min AP (for UFCS) 125
Table 3-26: Correlations between various parameters. All observations included 131
Table 3-27: Correlations between various parameters. Aggregations to minutes and radar stations. 131
Table 3-28: Correlations between various parameters for speeds exceeding 50 km/h. Aggregations to minutes and radar stations 131
Table 3-29: Correlations between the number of sign changes per unit time and various traffic state parameters 132
Table 3-30: Correlations between wavelet components of velocity profiles and some traffic state parameters for various wavelet types, speed conditions and aggregation types. 133
Table 4-1: Travel time estimators: Comparisons in congestion / free flow, scenario 1 171
Table 4-2: Travel time estimators: Comparisons in congestion / free flow, scenario 2 179
Table 4-3: Travel time estimators: Comparisons in congestion / free flow, scenario 3 187
Table 4-4: Travel time estimators: Comparisons in congestion / free flow, scenario 4 193
Table 4-5: Speed-density calibration estimated parameters 203

Executive Summary

Efficient traffic management relies on up-to-date information about the state of the road network. As more traffic and other data sources become available, more accurate estimation and prediction of road traffic conditions utilising all data sources is essential to reduce congestion and increase traffic safety. This is one strand of the INTRO (Intelligent Roads) project.

The INTRO project is aimed at demonstrating how safety, capacity, road operation and maintenance problems can be alleviated by the use of existing and new sensor technologies in a holistic way, to add intelligence to road networks.

It is a three year project which started in March 2005, with a budget of 3.5 million Euro supported by the European Commission under the 6th Framework Programme. The INTRO project consists of three technical work packages (WP2: Novel concepts of surface safety monitoring; WP3: Pavement conditions from intelligent pavements and intelligent vehicles; and WP4: Traffic and safety monitoring) and with Work Package 1 (WP1: Current Knowledge and Implementation Strategies) interacting with the technical packages. Further information on the overall aims and activities of the project can be found at the INTRO website http://intro.fehrl.org/

Work Package 4 of the INTRO project focuses on developing methodologies for predicting traffic conditions and for monitoring traffic safety using data measured from traffic sensors and floating car data. Data fusion methodologies on combining data from multi sources to predict travel time and traffic safety are investigated. The effect of weather on traffic conditions is also an important factor; and the development of safety indicators to monitor traffic safety using traffic sensor data are also addressed in this work package.

Deliverable D4.1 is the first of three deliverables of WP4. The scope of this deliverable are:

- Identification of key traffic and safety indicators to characterise traffic conditions,
- Development of methodologies to predict travel time using floating car data
- Development of data fusion methodologies using data from multiple sources to predict travel time.

Key traffic and safety indicators

As a first preparatory step of the project, an inventory of traffic and safety indicators, methods for their computation or estimation and existing data sources for sensor-based measurement of traffic parameters have been compiled. At this state of analysis one indicator related to road network global conditions (stopping distance), one indicator related to supply (road capacity), two indicators related to demand-supply interaction (traffic density, travel time) and three indicators directly related to safety accident risk or accident rate, macroscopic accident probability, microscopic speed in homogeneity have been identified and discussed.

Travel time prediction using floating car data

Floating car also referred to as probe car is vehicle equipped with positioning equipment such as GPS to log the position of the vehicle and other data collected on-board. Unlike fixed point traffic sensors that collected traffic data of all the vehicles passing through the sensors, floating car data collects continuous trajectory data of the equipped vehicle. However, from the floating car data, it is not possible to know the number of vehicles (demand) at any given time. A fleet of floating car is needed to generate enough data for traffic management such as travel time prediction. Taxis are the most common fleet used for floating cars. Given that privacy is a sensitive issue, as the data collected by the on-board unit transmits the vehicle ID and positions, some trial experiments with unidentified floating car data. This system transmits only the position and speed of the vehicle but not the vehicle

ID. This means that when all the data from an unidentified floating car fleet are combined, we will have a network with speed values at different points on the network. Clearly, methodologies to predict travel time for identified and unidentified float cars are different.

This deliverable describes two travel time prediction models, one for identified FCD and the other for unidentified FCD.

Identified floating car

Data analysis was based on data from the FLEET system in Vienna, which has been operational since 2004. A taxi fleet of 800 taxis equipped with GPS receivers delivers positions and time stamps to a taxi management centre at an interval of 15 to 30 sec. From the taxi management centre, data are transmitted and evaluated at 15 min intervals.

Message data from floating cars are cleaned in a three-step process in order to reduce spatial and temporal errors of taxi messages.

- A map matching algorithm projects trip data (message points) to the road graph.
- If only start and stop of trips are known or if GPS messages are sparse, an optimised shortest path algorithm is applied for route search. The algorithm uses several labels in order to reduce spatial errors of routing. After routing, average velocity of each trip can be calculated.
- A travel time calculation algorithm calculates the estimated speed value for every road on a route. The average speed for the route is distributed to each road on the route. In order to take into account different speed levels of each road, average route velocity is distributed to each road proportional to its free speed or historical speed. Then space mean speed is calculated from all speed values available for a road in a time interval (of currently 15 min).

The investigation of route speeds has shown that variance plays an important role in the interpretation of results from route speed measurements. Average route speed on intra-urban roads can be estimated using time series modelling and linear Kalman filtering. Therefore, estimated route speed can be regarded as the relevant measurand which is related to other traffic parameters such as traffic density and traffic flow (concept of equilibrium speed).

As route speeds from a taxi fleet only constitute a small sample of experienced drivers, accuracy of route speed measurements has to be compared to an external reference. AVI measurements based on license plate recognition which represent a random sample of the normal population of drivers were used to evaluate whether route speeds represent the whole or average population of drivers. Route speed data which have been filtered with a Kalman filter are closely related to the average route speed of drivers measured by AVI. Floating car route speeds are only slightly overestimated average route speed of drivers measured by AVI. Mean error between average floating car route speed and average AVI route speed on one lane was +5.1 km/h on route 29 and +5.8 km/h on route 69.

Unidentified floating car

In Chapter 3.3 a new method is proposed to estimate travel times on links from unidentified probe observations of position and velocity. The new method is based on modelling the velocity profile as a continuous function along each link in a network (rather than piecewise constant functions in existing travel time estimation methods). The spatial modelling is done by first establishing a priori velocity profiles for sparse and dense traffic conditions, respectively, using historical data. Each new observation from an unidentified floating car is then combined with these a priori profiles (using linear interpolation) to obtain a current velocity profile. Various approaches to combine travel time estimates from such individual velocity profiles are investigated and compared, as well as comparisons with existing methods.

The θ function associated with the observation of speed, v at position, x is used to obtain the current speed profile using the equation:

$$f_\theta(x) \equiv g(x) + \theta \cdot (h(x) - g(x)) = v$$

where:

$g(x)$ Corresponds to historical speed profile for light traffic conditions;

$h(x)$ Corresponds to historical speed profile for heavy traffic conditions; and

θ is a parameter for linear interpolation.

The figure below illustrates the θ function with $M(x)$, the maximum velocity curve, and $m(x)$, the minimum velocity curve. The star denotes an observation from an unidentified vehicle and the blue line denotes the θ function associated with the observation, computed by linear interpolation between the surrounding a priori functions for light and heavy traffic conditions.

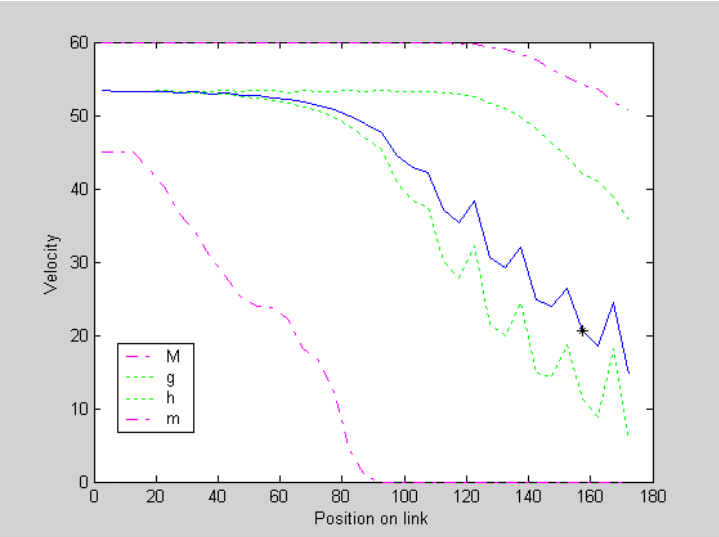


Illustration of θ functions

The results based on simulated data show that the new method generally reduce the relative error in travel time estimate as compared to existing methods based on piecewise constant velocity profiles. However, implementation details, especially the handling of low velocities, may strongly affect results, particularly in very congested traffic situations.

Data fusion

Increasing data sources such as sensors, databases and experts, often inconsistent and sometimes contradictory, make the synthesis tremendously difficult. In order for the data sets to be fully useful, data sources need more and more complex processing techniques, to take into account data qualities and imperfections of each source. One solution is based on data fusion. Data fusion approach is quite appealing mainly in heterogeneous data sources settings.

In this deliverable, the focus is on data fusion for two traffic performance indicators: Travel time and traffic density. The data fusion processes developed in the INTRO project aim to provide a consistent and comprehensive picture of network conditions.

Firstly, it is shown that the combination of toll travel times of different links can result in a more reliable reference travel time. Secondly, very encouraging results are then obtained for travel time estimation by the fusion of loop detector data and toll collection data. The integration of toll collection data into

the cumulative traffic counts method (classically based on loop detector data only) improves both the accuracy and the robustness of travel time estimation:

- Accuracy:
  - During congestion, the estimated travel time is better than with the classical cumulative traffic counts method or with the speed extrapolation method (which sometimes misses the congestion).
  - During free flow traffic, the estimated travel time is better than with the classical cumulative traffic counts method and close to the speed extrapolation method.
  - Overall, the estimated travel time is closer to the reference.
- Robustness:
  - Travel time estimation is correct or quickly returned to correct, even when there are counting mistakes in the loop detector data.
  - The same applies for wrong initialization too.

Overall, this research shows the potential of toll collection data for travel time purposes, both for reference travel time computation and for travel time estimation. The results must still be confirmed and evaluated on a larger sample of real situations including various distributions of available toll travel times.

Finally, travel time and link flows fusion, under heterogeneous data source configurations, are also examined. It is found that, the outlined combining strategies improve the quality of the estimation and provide a comprehensive traffic estimate using floating car data as the primary source of information for a Dynamic Traffic Assignment (DTA) model.

This work shows how estimates of link flows can be obtained using floating car data as the primary source of information for a DTA model. While no measurements of the link flows are available to compare with the estimated flows, the results are encouraging and demonstrate that the calibrated model accurately matches the historical travel times. This is an interesting finding in terms of traffic management applications, as travel times are one of the most important outputs of DTA and traffic estimation and prediction models.

As more network operators and cities are moving away from conventional loop detector surveillance systems (that are expensive and tedious to install, maintain and operate) and towards emerging data collection systems, such as floating cars, the issue of using such data for traffic management purposes becomes increasingly relevant. The ability to use such data for the calibration and operation of traffic estimation and prediction systems may be an important driving force for these technologies and change the way traffic data is collected and used.

In summary, data fusion has proven very useful for estimating and predicting traffic indicators. The main innovation here is the use of alternative traffic data sources such as toll collection data and more generally, floating car data, which is becoming more and more available.

Building on this success, data fusion will be applied within INTRO to safety perspectives as well. The aim is to fuse already operational safety indicators in order to achieve a robust and comprehensive characterisation of traffic situation from a safety point of view. The results will be reported in deliverable D4.3.

1 Introduction

Everyday traffic control centres are faced with operational issues such as traffic congestion and traffic incidents. Traffic sensors are installed along road networks to collect traffic data and often for one specific application. If these data are converted into useful traffic indicators describing the traffic conditions, appropriate traffic management actions can be activated. However, it is quite common that collected data are not optimally used even though it could be very useful when these data are integrated. An example is the queue lengths caused by incidents, which are collected in France on rural motorways from the observation of patrols. These data are transmitted to the operator at the regional traffic centre and is generally only used for producing congestion statistics. On Autoroute Paris-Rhin-Rhône (APRR) network, it is now used for the enhancement of real time traffic management tools such as updating of its estimation travel time method. Another example is toll ticket data which is only used for tolling. In France, only one toll motorway company is currently using this data for real-time traffic management, namely travel time prediction, although other motorway companies are now considering this use.

Efficient traffic management relies on up-to-date information about the state of the road network. As more traffic and other data sources become available, more accurate estimation and prediction of road traffic conditions utilising all data sources is essential to reduce congestion and increase traffic safety. This is one strand of the INTRO (Intelligent Roads) project.

1.1 INTRO Project

The INTRO (Intelligent Roads) project is aimed at demonstrating how safety, capacity, road operation and maintenance problems can be alleviated by the use of existing and new sensor technologies in a holistic way, to add intelligence to road networks.

It is a three year project which started in March 2005, with a budget of 3.5 million Euro supported by the European Commission under the 6th Framework Programme. The INTRO project consists of three technical work packages (WP2: Novel concepts of surface safety monitoring; WP3: Pavement conditions from intelligent pavements and intelligent vehicles; and WP4: Traffic and safety monitoring) and with Work Package 1 (WP1: Current Knowledge and Implementation Strategies) interacting with the technical packages. Further information on the overall aims and activities of the project can be found at the INTRO website http://intro.fehrl.org/

1.2 Purpose of WP4 "Traffic and Safety Monitoring"

Work Package 4 of the INTRO project focuses on developing methodologies for predicting traffic conditions and for monitoring traffic safety using data measured from traffic sensors and floating car data. Data fusion methodologies on combining data from multi sources to predict travel time and traffic safety are investigated. The effect of weather on traffic conditions is also an important factor; and the development of safety indicators to monitor traffic safety using traffic sensor data are also addressed in this work package.

WP4 is divided into 5 tasks and a brief description of each task is given below. The results from Task 4.1 to Task 4.3 are reported in this deliverable D4.1.

- **Task 4.1: Traffic parameter needs and single sources estimation**
  The first task is to identify indicators that are most pertinent to road authorities and operators, in describing the traffic conditions and traffic safety.

- **Task 4.2: Innovative use of sensors for improving traffic parameters estimation and forecasting**
  The focus of this task is to develop new and improved methodologies to estimate indicators using floating car data. To compliment the field data collected from floating car, microscopic simulation is used for evaluating the newly developed methodologies.

- **Task 4.3: Traffic parameters estimation by using data fusion**
  In task 4.2, only data from single source is used for estimating indicators. This task explores the benefits that data fusion offers in combining data from multi data sources.

- **Task 4.4: Traffic parameters forecasting taking into account weather effects**
  Effects of weather on drivers behaviour and therefore road performance are often not taken into account. This task analyses the effect of inclement weather on drivers behaviour such as speed and headway. The task also includes driving simulator study on the effect of fog on drivers behaviour. The results of this task will be reported in deliverable D4.2.

- **Task 4.5: Safety indicators deducted from traffic parameters**
  This task develops safety indicators for monitoring traffic safety. Microscopic simulation is used to test the safety indicators for a range of scenarios for evaluating the robustness and effectiveness of the safety indicators. The task also includes driving simulator study on the effect of fog on drivers behaviour. The results of this task will be reported in deliverable D4.3.

Deliverable D4.1 is the first of three deliverables of WP4. The scope of this deliverable are:

- Identification of key traffic and safety indicators to characterise traffic conditions,
- Development of methodologies to predict travel time using floating car data
- Development of data fusion methodologies using data from multiple sources to predict travel time.

1.3 Structure of this report

Chapter 2 of this deliverable provides a selected inventory of traffic and safety indicators, methods for their computation and existing data sources for measuring the traffic parameters. Full list of the traffic and safety indicators considered are provided in Annexes I and II.

Chapter 3 covers the use of identified and unidentified floating car data (FCD) to predict travel time. Floating car is also commonly referred to as probe car. Traffic simulation is employed to augment field data for further testing of the methodologies developed and comparative studies of simulated versus estimated results are presented. The use of high resolution FCD for predicting traffic state is also discussed.

Chapter 4 provides an overview of the concepts and methods of data fusion. Applications of data fusion for travel time prediction and for dynamic OD estimation are described.

Finally, Chapter 5 draws conclusions about the research conducted for this deliverable.

2 Traffic and Safety Indicators

The overall objective of Task 4.1 is to establish an inventory of traffic indicators, methods for their computation or estimation and existing data sources for sensor-based measurement of traffic parameters. The basic needs of traffic indicators regarding the different possible applications have been defined in WP 1. Stakeholder interviews have been conducted and evaluated in order to summarize the most important targets of road operators (see [4]). The results of the interviews have underlined the importance of information on traffic performance, weather and road surface state.

Task 4.1 has evaluated and selected the most relevant indicators characterising traffic performance and traffic safety. The selection of indicators was based on the experience of the participating organisations and a literature review. The most important publications have been collected in a reference list (see Annexe I - Overview of traffic performance indicators and safety indicators). Each indicator was characterised by its definition and by its spatial and temporal attributes. Additional attributes, like reliability or accuracy were taken into account as well. These attributes were analysed for each indicator, and methods for their computation or estimation were identified. The requirements of these indicators were the basis for selection of parameters from each data source.

Task 4.1 was dedicated to conventional single data sources measuring traffic flow conditions (e.g. traffic flows, traffic density, current speed, travel time, etc.). The analysis included a description of the data quality (e.g. spatial and temporal resolution of data) and data availability which influences the quality of the derived traffic indicators. Basic data correction methods were described. The result is a selection of the most relevant indicators for traffic performance and traffic safety, a description of existing and new methods for computing or estimating them, and an inventory of available data sources. These outputs also offer a common language for the whole WP. This common basis is necessary to ensure an optimal compatibility of the other tasks outputs, especially for practical applications.

Fig. 2-1 shows a typical approach for indicator analysis (in this case, the level of service, LOS).

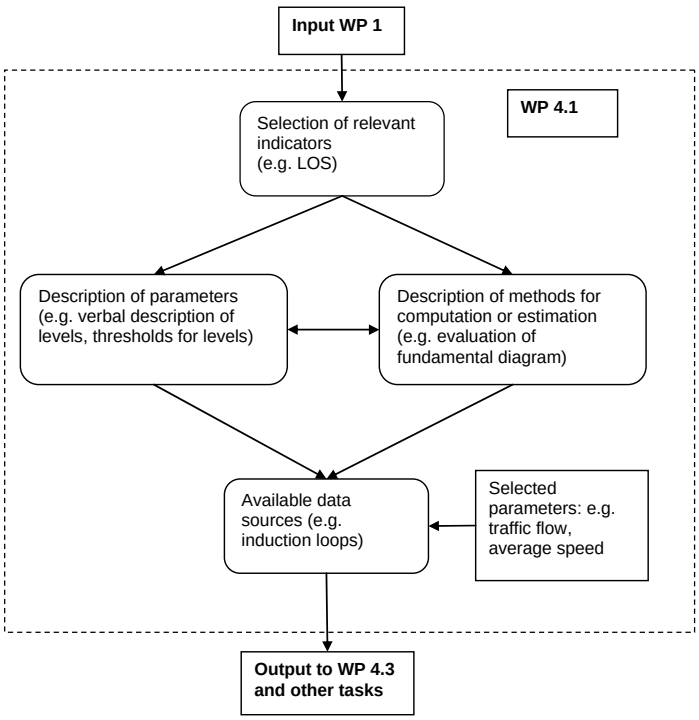


Fig. 2-1: Example for analysis of indicators

2.1 Inventory of indicators

In this report, the most relevant traffic performance indicators and safety indicators were categorized, and their parameters were listed. They have been identified as particularly important for analysing traffic performance and traffic safety, because they provided aggregated information concerning traffic performance and safety.

For each indicator, this analysis relied on a comprehensive grid including:

- its definition, its relevance for traffic performance and safety and its relationship with other indicators
- its spatial and temporal attributes (geographical and temporal scope).
- measured raw parameters
- indirect estimation and calculation techniques from available parameters

The analysis was based upon more detailed information on available raw traffic parameters collected by measurement stations and traffic management centres see Annexe I - Overview of traffic performance indicators and safety indicators).This list formed the information pool for further analysis and selection of indicators.

At this state of analysis one indicator related to road network global conditions, one indicator related to supply, two indicators related to demand-supply interaction and three indicators directly related to safety have been identified and discussed. In Fig. 2-2 a graphical overview of the investigated indicators, their corresponding parameters and relations among them is presented.

To be in line with real world applications we asked several users for their feed-back on selected indicators. The replies received from Austria (ASFINAG), Germany (Polizeipräsidium Oberbayern), and Poland (General Directorate of National Roads and Motorways and City Road Administration Warsaw) have been analysed and integrated into the description of each indicator.

The recent development of ITS applications have improved the existing means of traffic data collection and produced new sources of information. Frequently more than one parameter for a given traffic situation is available. The calculation of safety indicators usually requires more than a single parameter because various aspects of road safety have to be taken into account.

In order to combine all available information sources, methods of data fusion have to be applied. Methods of data fusion and existing applications are being investigated in the framework of Task 4.3. The applications of data fusion that are tested within the framework of the INTRO project mainly depend on the availability of data sources at specific test sites. Therefore, a subgroup of relevant traffic performance and safety indicators which are eligible for data fusion (performance indicators eligible for data fusion, PIEF) will receive a special focus.

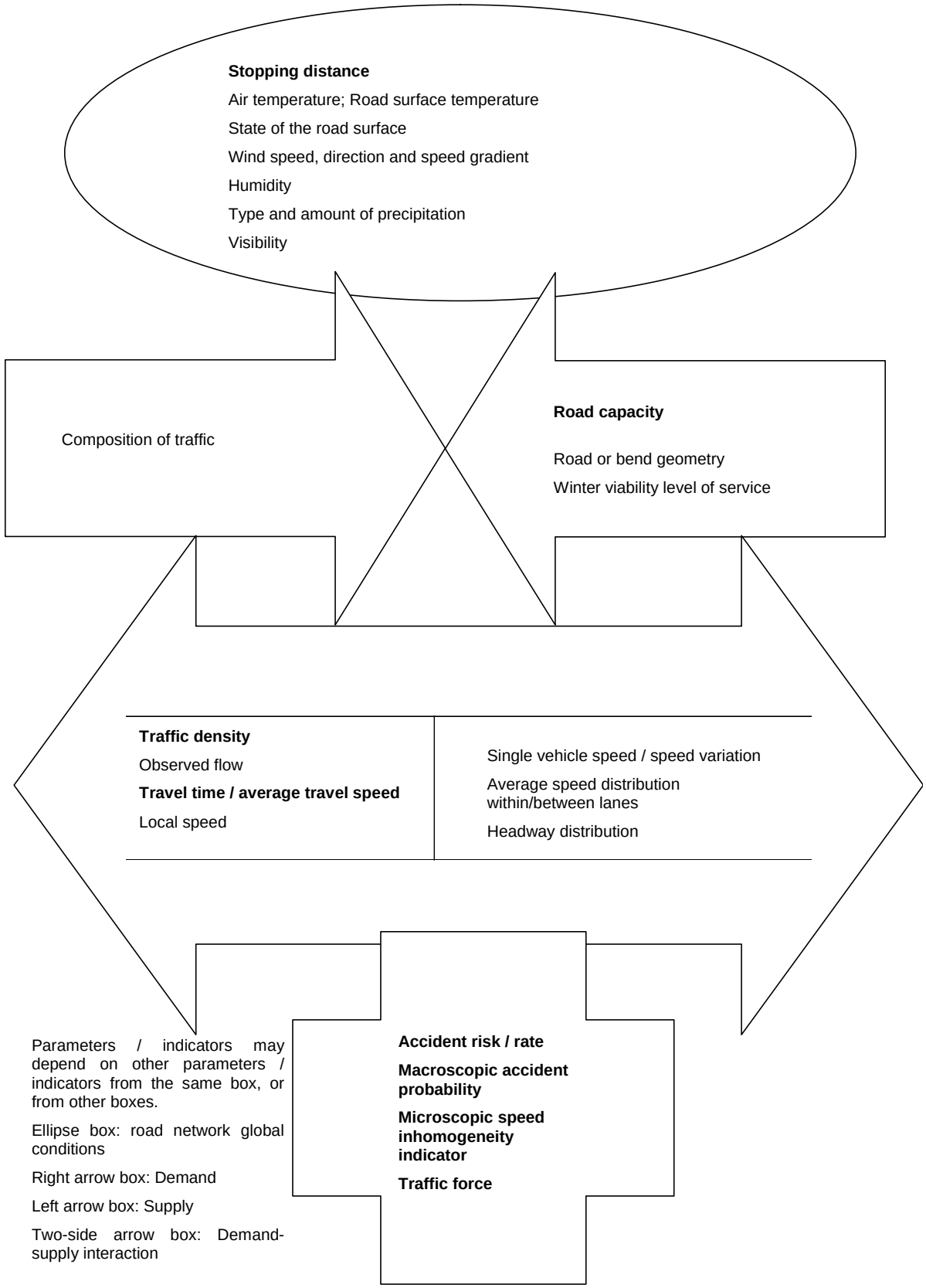


Fig. 2-2: Interaction of parameters / indicators

2.2 Indicators related to road network global conditions (weather, road surface)

As it will be demonstrated by task 4.4, an accurate use of weather data improves traffic performance indicator estimation and forecasting. Therefore, the most important parameters describing weather and road surface were identified and described. This subtask relies on initial outputs given by WP2, but also from practical experience of the different partners. Most of them have dealt with data on road network global conditions and faced important limitations in actual indicator description. As information from individual speed, time or space headway and road network global conditions determine the individual stopping distance of a car, methods of data fusion can be used in order to calculate stopping distance.

Indicator R1: Stopping distance

The stopping time in relation to time headway determines the ability of the driver to stop without colliding with the front vehicle. Stopping distance has to be compared with space headway. Basically, stopping distance must be less than space headway, and stopping time must be less than twice the time headway, considering that average speed during stopping is half of the initial speed.

R1	Stopping distance / space and time headway
Geographical scope	segment
Temporal scope	minute
Parameters	• individual speed • road network global conditions
Elaboration method	1) calculation from individual speed, additional car equipment data and road network global conditions

Elaboration method 1

Stopping distance is calculated from individual speed, additional car equipment data such as breaking system (if available) and road network global conditions. A model for stopping distance is being worked out in WP 2 ([5]).

2.3 Indicators related to traffic demand

Different parameters describing traffic demand and traffic composition play an important role as input parameters for traffic models and traffic safety (mainly traffic composition and percentage of trucks).

Whereas parameters related to traffic demand play an important role for traffic performance indicators, no indicator of high importance was identified that was based exclusively on traffic demand.

2.4 Indicators related to supply

Indicators related to supply of road infrastructure constitute another important input for analysis of traffic performance and safety. These include mainly static road parameters (e.g. road geometry) and dynamic properties of roads (e.g. capacity or reduced capacity due to roadwork and incidents).

Indicator Su1: road capacity

An important indicator for traffic performance is road capacity. It is defined as the maximum possible flow of vehicles and depends mainly on road geometry (i.e. number of lanes, lane width, link length, number of signalised or unsignalised intersections) and road network global conditions. Some methods of capacity estimation are based upon traffic state and therefore, have to fuse traffic flow

information with traffic state information. If capacity is estimated by model calibration, several traffic parameters are used in combination.

Su1	road capacity
Geographical scope	segment, point
Temporal scope	hour, day
Parameters	• observed traffic flow • traffic state
Elaboration method	1) capacity as largest observed traffic flow in steady-state traffic 2) capacity as traffic flow before traffic breakdown 3) capacity as calibrated model parameter

Elaboration method 1: capacity as mean flow of steady-state speed classes

Source: [6]

1) Define the exogenous conditions and the composition of traffic. Typically, for road traffic it is possible to measure traffic composition on the basis of the proportion of heavy vehicles, discretised in classes of 2 to 5%
2) Define the speed classes for the exogenous conditions identified in stage 1. Typically, classes of 5 km/h are suitable for road traffic.
3) Group observation periods together in batches, on the basis of exogenous conditions and speed classes. Unstable periods, in which the speed varies considerably, should be eliminated from each batch. This can be done by eliminating the periods h in which the mean speed differs from that of the adjacent periods, e.g. the periods that violate the following stability conditions: $|v_h-v_{h-1}|+|v_h-v_{h+1}| \leq 8$ km/h
4) Compute the mean flows within each batch.
5) Compare the speed classes for each exogenous condition in order to identify discrete and continuous traffic regimes.
6) For each steady state batch, the mean flow measures the capacity under the particular exogenous conditions and speed.
7) The maximum mean flow of the steady state batches, under given exogenous conditions, provides a measurement of the road maximum capacity, i.e. its critical flow. The speed at this flow level is the critical speed. For a motorway this is typically between 85-95 km/h.
8) To measure the passenger car equivalent of a lorry, it may be possible to use the mean flows per class in the case of steady state batches

Elaboration method 2: capacity as traffic flow before traffic breakdown

Traffic flow is measured and usually aggregated in 5 min intervals. Traffic flow from intervals which are followed by traffic breakdown is selected, and probability of traffic breakdown is calculated for every traffic flow value. The expectation value of traffic flow for this distribution is chosen as capacity (see [3]).

A method based on analysis of measured time headways in the traffic stream was developed at Delft University of Technology [8] to estimate total and per lane capacity. Carlsson et al. [9] applied this method to estimate traffic capacity on an arterial road with several lanes in each direction (Essingelink) in Sweden.

Recent research by Lorenz and Elefteriadou [10] and Brilon et al. [11] incorporates the probabilistic nature of freeway breakdown process and treated capacity as a random variable. A methodology for the estimation of capacity distribution functions based on empirical data can be found in [11].

Elaboration method 3: capacity as calibrated model parameter

In urban traffic networks the capacity of a street is mainly determined by the capacity of its signalised intersections. As traffic flow measurements are only available for arterial roads, capacity is frequently estimated from a few historic measurements or from signal timing.

In traffic assignment models, capacity of a segment is a supply parameter for the traffic performance function (e.g. queuing model) together with other supply parameters (e.g. parameters of the speed-density relation). If sufficient measurement data are available (e.g. traffic flow, local speed, travel times from floating cars, etc), capacity can be calibrated together with other supply parameters. The supply parameters are adapted until the model output corresponds to actual measurement data (see [1]).

User perspective – feedback from highway providers

In Austria, the use of the indicator road capacity is planned to provide service and information for drivers and operators, as well as for statistical data collection.

In Traffic Control Centres (TCC) this indicator is used, if traffic flow is close to or exceeds capacity. In this case maximal speed on VMS is reduced or overtaking of HGVs is prohibited for specified motorway sections (e.g. A9 near Nürnberg).

In Poland, the indicator is used for maintenance (winter and summer), road network development, and planning of future investments.

2.5 Indicators related to the demand/supply interaction, i.e. the quality of flow

The most important indicators related to demand/supply interaction describe traffic density or are connected to traffic density. If high-quality traffic flow measurements are available, traffic density can be calculated from local measurements or from a combination of two measurements for a road section (e.g. method of cumulative number of vehicles). As the estimation of traffic density is affected by measurement and estimation errors, additional information e.g. travel times from tolling systems has to be used for improving density estimation. Traffic models based on Dynamic Traffic Assignment can be applied for estimating traffic density from measurement data as well.

As average travel speed is limited by traffic density average, travel speed relative to free flow speed or normal speed can be used as indicator for traffic density. The main challenge is how to derive travel time data from traffic detectors and from mobile sensors (floating cars is specifically addressed in Chapter 3). Even if various existing methods offer satisfactory results, substantial improvements are still possible. For example, this is the case for the method based on the cumulative number of vehicles (stock method) and other methods of data fusion. Microsimulation is used to develop and validate these methods for specific test sites.

Indicator D+S1: Traffic density (number of vehicles per road length unit)

Traffic density defines space headway between vehicles. Its relevance for safety is under investigation (some types of accident, e.g. getting off the road due to speed, may decrease during congestion). In urban areas traffic flow is only measured on arterial streets. Therefore methods to estimate density for the whole street network are of considerable importance.

D+S1	Traffic density
Geographical scope	segment
Temporal scope	minute, hour
Parameters	• occupancy • flow and local speed, route speed (from floating cars)
Elaboration method	1) Estimation of traffic density by cumulative number of vehicles 2) Estimation of traffic density by Dynamic Traffic Assignment

Elaboration method 1: Estimation of traffic density by cumulative number of vehicles

If roads are equipped with a measurement station on every section, traffic density can be calculated by the cumulative number of vehicles method (stock method). As traffic flow measurements are subjected to systematic errors, calculated density values tend to differ from real traffic density and have to be corrected. This can be accomplished by data fusion of travel times estimates with measured travel times from toll data (see chapter 4.4.1).

Elaboration method 2: Estimation of traffic density by Dynamic Traffic Assignment

Source: [2]
A mesoscopic dynamic traffic assign model (e.g. DynaMIT) is calibrated using travel times and parameter estimates of the density-speed function and capacity of each street segment. The OD flows from node to node (equivalent to traffic flows) are adapted until simulated travel times approach measures travel times (see chapter 4.4.2 Estimation of dynamic OD matrices based on floating car data).

User perspective – feedback from highway providers
The use of the indicator traffic density is planned for service and information for drivers and operators, and for statistical data collection.

In Traffic Control Centres (TCC) this indicator is used, if traffic density increases beyond a certain threshold, maximal speed on VMS is reduced or overtaking of HGVs is prohibited for specified motorway sections (e.g. A9 near Nürnberg).

In Poland the indicator is used for maintenance (winter and summer), road network development, and planning of future investments. Furthermore, it will be a basis for the planned service for drivers (in 2008 after the Traffic Management System has been launched).

Indicator D+S2: Route travel speed / Travel time

As travel speed is limited by traffic density, route travel speed or travel time is used as indicators for traffic performance. These indicators are also a suitable form for traffic information to drivers.

D+S2	Average travel speed / Travel time
Geographical scope	segment
Temporal scope	minute, hour
Parameters	• route travel speed (current / historical) • traffic flow
Elaboration method	1) enhanced estimation of travel time based on the cumulative number of vehicles

Elaboration method 1: enhanced estimation of cumulative number of vehicles

This description is a summary of experiences concerning travel time estimation on the Autoroute Paris Rhin Rhône (APRR) motorway network in France. Detailed results can be found in Annex II

A - The different travel time calculation methods based on stationary traffic data

The speed method

The most popular method for travel time estimation is known as the speed method: the speed measured at the location of each traffic station is assumed to be the speed experienced by drivers along the stretch allocated to this station. Then, it is easy to calculate the crossing time of the stretch by dividing its length by the speed on the station. For the estimation of the travel time on a given itinerary, one then states that, at a given time t, this travel time is equal to the sum of the elementary travel times at time t on all the sections forming this itinerary. This travel time is known as an instantaneous travel time .

$$TT_{Speed}(t)=\frac{L}{2}\times\left[\frac{1}{V_{upstream}(t)}+\frac{1}{V_{downstream}(t)}\right] \quad (2.1)$$

$V_{upstream}$, $V_{downstream}$ traffic speed of the upstream and downstream traffic data station
L length of the section

This method requires a high density of data collection, so that the approximation of the section speed by the data station speed is not too inaccurate (in order to limit the spreading effect). Typically, data stations on motorways are spaced at 500 m intervals.

In addition, due to the instantaneous feature of this travel time, the estimation is always late during changing traffic situations compared to the real travel time experienced by drivers (underestimation during increasing congestion, and overestimation during decreasing congestion). This bias prevents the use of this method on long itineraries along which the traffic situation may change strongly while the driver is proceeding to his destination.

On such highly equipped motorway sections, very satisfactory performance has been obtained.

The stock method

On interurban motorways, intervals between data stations are generally equal to 20-25 km except on some locations where stations can be spaced at 5 km intervals (e.g. in the Rhône valley on A7 motorway in France). This 5 km interval is still suitable for the use of the speed method in the situations of general congestion, but generally speaking such low traffic data measurement densities are inappropriate for the use of the speed method in some other situations such as an accident happening during medium or low traffic level.

ISIS has designed and developed the so-called stock method which relies on traffic volume measurements instead of traffic speed measurements: as soon as the exit traffic volume differs from the entry traffic volume, congestion can be inferred somewhere on the motorway stretch. An instantaneous travel time can then be calculated by dividing the number of vehicles in the stretch (the stock) by the exit volume: if 500 vehicles are present at time t and the exit traffic volume is 1000 veh/h, the last one will experience a travel time equal to (500/1000)*60 = 30 min. The method requests the measurements of all entering and exiting volumes (so-called closed stretches).

Equations are the following:

$$Stock(t)=Stock(t_0)+\Delta\times\sum_{k=t_0}^{k=t}\left[Q_{upstream}(k)-Q_{downstream}(k)\right] \quad (2.2)$$

$$TT_{Stock}(t)=Stock(t)/Q_{downstream}(t) \quad (2.3)$$

where

$Q_{upstream}$ and $Q_{downstream}$ denote the traffic volume of the upstream and downstream traffic data station respectively,
L the length of the section,
Δ the elementary time slice (e. g. 6 minutes),
Stock(t) and Stock(t_0) the current and initial numbers of vehicles in the section.

Note that this formula is only applicable during congestion and $Q_{downstream}$ is equivalent to bottleneck capacity during congestion. If the formula is to be applied for free flow conditions, modification to the formula is necessary.

This estimation is still instantaneous in the sense that the exit volume is taken at time t, whereas it can change in the future (especially in case of an incident, when the carriageway is cleared). In this respect, as for the speed method, it is best fitted to the calculation of travel times on short itineraries or the calculation of queue crossing times. However, the method is based on a true traffic model (even if it is very simple) and it contains a predictive dimension: the stored vehicles will have to exit at some time in the future.

Compared to the speed method, the strength of the stock method is that, a good estimation of the travel time can be given very quickly (the delay necessary for the wave of decreasing volume to propagate until the downstream station is short) without having to wait for the congestion to step back on to the upstream station (even if this happens). Furthermore, no undesirable effect as the spreading of congestion along the whole stretch occurs.

Its weakness lies in the potential inaccuracy of the stock estimation: biased traffic volume measurements can make this estimation diverge after a period of several hours. The other limitation of the method lies in the fact that it can only work on closed sections , i.e. sections where all entering and exiting traffic is measured.

The combined method

This method designed by ISIS consists in using the speed method during free-flow traffic, switching to the stock method as soon as congestion is detected and coming back to the speed method when the congestion disappears. Experience shows that, on most occasions, the possible measurement biases have generally not enough time to significantly hamper the travel time estimation.

The detection makes use of several modules: detection of measurement drift, detection of congestion appearance, detection of decreasing congestion, detection of return back to free-flowing.

The combined method has been implemented in the TPI (Temps de Parcours Instantané) tool in the Autoroute Paris Rhin Rhône (APRR) Control Centre in Dijon.

The improvements of the combined method developed within the frame of INTRO are responses to the 2 limitations underlined above: lack of forecast and closed sections.

B - The introduction of forecast into the stock method

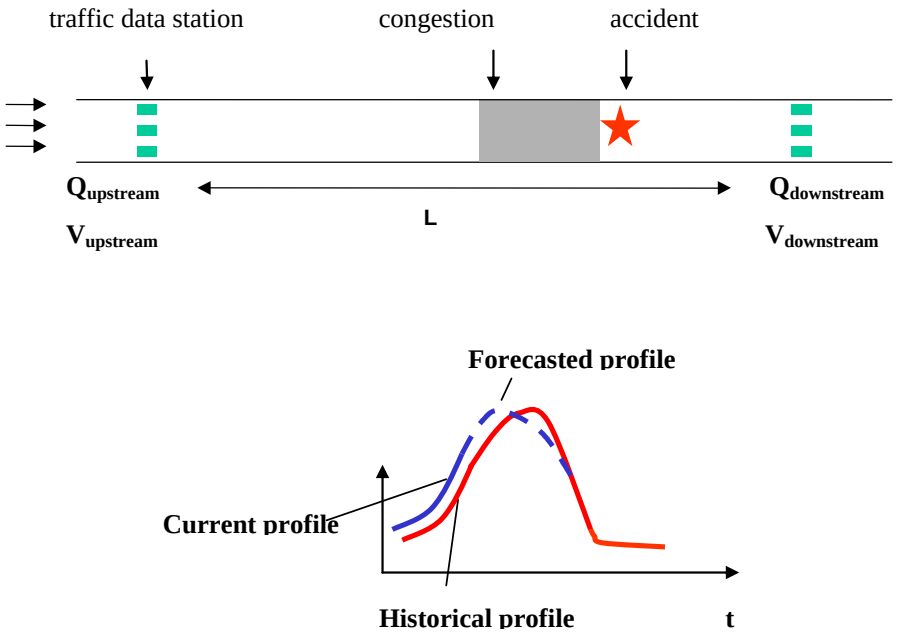


Fig. 2-3: Forecasting travel times based on stock method

Forecasting the stock means that we must forecast the upstream volume and the downstream volume of the section. The forecast of the downstream volume is specific because it is related to the incident characteristics. The problem is then to forecast the traffic volume on any counting station on the same time slice basis as the one used for the travel time calculation, i.e. 6 min (see Fig. 2-3).

In the case of APRR application, we have developed a simple method of traffic volume forecast built on the hourly traffic volumes (hourly traffic volume = number of vehicles passing during 1 hour) forecast that APRR elaborates 1 year in advance using a calendar-based method. This historical profile is available for each counting station and for each day of the coming year.

We first update on real-time at the end of any hour the next 2 forecasts (1 hour and 2 hour in advance) according to the last real-time hourly measurements already made. We then produce each 6 min, the 6 min-based traffic volume forecasts on a 2-hour moving horizon, on the basis of the latter forecast of 1 hour-based traffic volumes. Running the stock method on these forecasted stocks, we then produce each 6 min the profile of the 20 travel times forecasted from now until 2 hours ahead.

1 hour-based forecast of 6 min traffic volumes on a 2 hour horizon

For the hourly traffic volume forecast, several heuristic approaches have been tested from which we have extracted the two presented below:

a) Transposition method

This simple method consists in adding to the last measured traffic volume the predicted difference between the historical forecasts at time t and time t + 1 hour:

$$Prev(t+1)=Mes(t)+[Prev_histo(t+1)-Prev_histo(t)] \quad (2.4)$$

where:

Prev(t+1) = hourly traffic volume forecasted for the next hour
Mes(t) = hourly traffic volume measured during the last hour

Prev_histo(t) and Prev_histo(t+1) = historical forecasted hourly traffic volumes for the current and next hours.

b) Pattern method

The idea here is to assume that the general shape of the historical traffic volume is correct and to make use of the first hours of the day to fit this historical pattern in the best way with the first measured traffic volumes. Then, the transposition method is applied on the updated historical traffic volumes instead of the historical volumes themselves:

$$Prev(t+1) = Mes(t) + [Dernière_Prev_histo_recalée(t+1) - Dernière_Prev_histo_recalée(t)] \quad (2.5)$$

where:

Dernière_Prev_histo_recalée(t) and Dernière_Prev_histo_recalée(t+1) = updated historical forecasted traffic volumes for the current and next hours.

Two different ways have been investigated: the moving pattern method, where the updating of the historical pattern is made each hour, and the frozen pattern method, where the pattern is kept unchanged from the sixth hour after the beginning of the moving method.

6 min-based forecasts on a 2 hour horizon

The first step is to transform the 1-hour based forecast traffic volume profile into a 6 min-based traffic volume profile. Several methods can be used: the first one, called constant breaking-up, simply consists to transform the 1-hour traffic volume into constant 6 min traffic volume slices, each one being equal to this 1-hour traffic volume (when expressed in v/h). The second one, called linear breaking-up, consists in assuming that, at half time of 1-hour time slice (say 10:30 am, 11:30 am,), the 6 min-based traffic volume (expressed in v/h) is equal to the traffic volume of the 1-hour based, linking all these points by straight lines and dividing each 1-hour slice into 10 6-min slices accordingly.

Then, 2 different main ways for producing the 6 min-based traffic volume forecast are possible:

- method A : make the updating of the 1-hour historical demand only at changing hours (10:00 am, 11:00 am,) on the next 3 hours, transform it into 6 min profile and keep this 6 min profile until the next changing hour, simply reading the 6 min volumes on this profile along a moving 2 hours window
- method B: use at any time the current 6 min volume measurement (averaged with the one before in order to smooth random variations) for updating the 6 min profile on the next 2 hours.

Several variants have been also tested.

Forecast of the exit volume

When the cause of the congestion is geometric, for example a decrease of the number of lanes, then the exit volume is known, and is constant from the beginning of the congestion until its end. In that case, the instantaneous travel time is also a forecasted travel time and there is no need to do any forecast of the exit volume.

When the cause of the congestion is an accident, there are methods to predict the duration of the capacity restriction in relation with the characteristics of the accident (i.e. number and type of vehicles involved, presence of injured people, existence of spilled load, etc). ISIS has developed a duration prediction tool for APRR (PREDIRE) based on statistical relationships from 2 years of historical data (more than 1500 accidents in the log).

The principle is the following: knowing that the anticipated time when the carriageway is cleared is T_1 (T_1 > current time t), then as soon as the travel time $TT_{Stock}(t)$ exceeds the time lasting until T_1, the exit

volume must, during the period *TT(t) - T_1* , be taken different from the current exit volume and equal to the nominal capacity *Capa_nom*. The formula is then the following:

$$TT_{Stock}(t)=(T_1-t)+[Stock(t)-(T_1-t)\times Q_{downstream}(t)]/Capa_nom \quad (2.6)$$

Fig. 2-4 shows the correction obtained, which leads to a more realistic estimation of the real travel time:

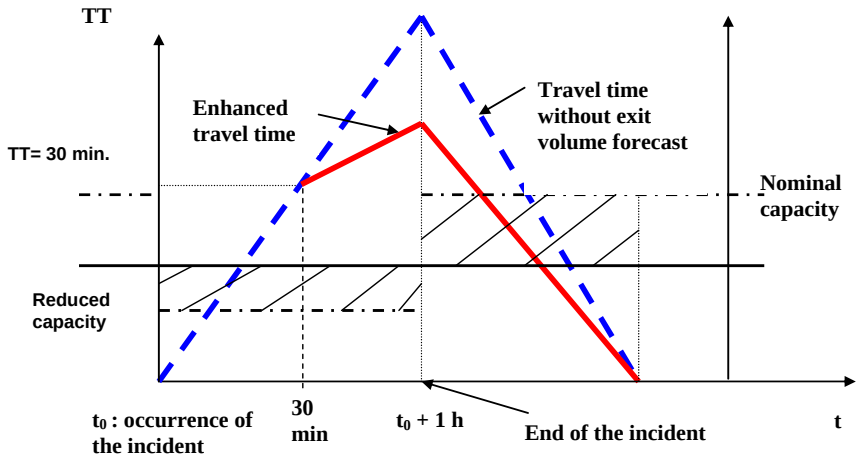


Fig. 2-4: Correction of link travel times

C - The extension to open sections

The extension of the stock method to open sections (i.e. sections where some exit or entry volumes are not measured because of a lack of counting stations or their breakdown) is another important enhancement, presented below. The idea is that the data of queue length observed on the field can be transformed into an estimation of the number of vehicles in the section, then, dividing by the exit volume gives the travel time. The principle is illustrated in Fig. 2-5. If Q denotes the exit volume, the density in the queue Queue_density is given by the fundamental diagram of the section.

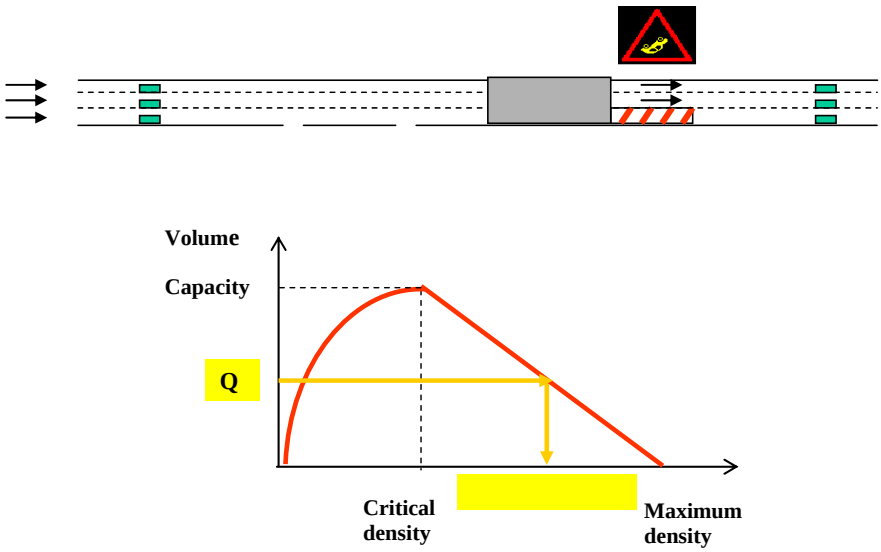


Fig. 2-5: Calculation of queue density

If L denotes the length of the queue, then the number of vehicles in the queue is given by Eq.(2.7). The crossing time of the queue is given Eq.(2.8)

$$N = \frac{L}{Queue_density} \quad (2.7)$$

$$TT = \frac{N}{Q} \quad (2.8)$$

The travel time of the section is then calculated by adding to TT the travel time on the upstream and downstream free parts of the section from their lengths (known from the kilometre points of the head and the tail of the queue) divided by the upstream and downstream speeds respectively.

The method is effective and can give far better results than the speed method when the congestion is not seen by the traffic data stations, provided that the queue lengths are measured accurately by the patrol man on the field. This is not always the case, because he may falsely deduce that the congestion is over, because the traffic flow is free at his level, whereas this is an acceleration wave and an upstream congestion still remains. This ends the considerations about the stock method and how to make the best use of traffic data and event data.

Detailed results for each method are included in the Annex II In chapter 4.4.1 a method will be presented using toll ticket data as a valuable additional source to contribute to robust real-time travel time estimations on motorways.

User perspective – feedback from highway providers

There is growing number of travel time information systems, operated in several EU countries. Selected examples are:

- Travel time estimation system on an access route to Valencia, Spain
- Travel time calculation and dissemination on the Paris ring road, France
- Travel time and delay estimation on the motorway network in the Copenhagen region, Denmark
- Travel time estimation and dissemination on French inter-urban motorways
- On-trip information and trial of journey time collection systems in Glasgow, Scotland

- Travel time estimation in real time using toll stamps on Italian motorways

Travel time services have been investigated and tested in Euro-regional projects and the TEMPO programme of DG-TREN of the European Commission.

In Poland the indicator average travel speed/ travel time will be used for the planned service for drivers, which will be implemented in some areas (in 2008 after the Traffic Management System has been launched).

2.6 Indicators related to safety

In this subtask, relevant safety indicators have been defined and evaluated. Most effort was focused on the parameters which are necessary to calculate them and their availability. Both macroscopic parameters (e.g. average speed, flow,) and microscopic parameters (vehicle interaction parameters like time and space headways, difference of speed,) were analysed. These results serve as basis for task 4.5 investigations and will be reported in deliverable D4.3.

Indicator Sa1: accident risk / accident rate

One cardinal safety indicator is the risk of accident which can be regarded as a function of several relevant parameters. The indicator used for comparing the relevant parameters is the observed accident rate.

Sa1	accident risk / accident rate
Geographical scope	point, segment
Temporal scope	minute, hour
Parameters	• Road geometry (road section before bend, width, radius and inclination of bend, intersection, ramp, weaving section) • Road network global conditions (Road surface conditions and weather conditions) The existence of (even light) rain limits the resistance and thus the stopping performance of the vehicles (although frequently drivers do not modify their behaviour). Road surface conditions (wet carriageway) may be observed by specific sensors which are still rarely implemented on motorways • Driver reaction time (alcohol, tiredness) • Manoeuvres (lane changing, overtaking) • Platoon formation (local traffic flow by lane, percentage of trucks) • Time headway, speed and relative speed of vehicles (stopping time) • Type of traffic participants (truck, two-wheeled, pedestrian) • Day category and daytime (Night/day) • Speed or capacity reduction (jam creation) • Type of collision (single vehicle, rear-end collision, perpendicular collision, frontal collision)
Elaboration method	1) investigation of scenarios for combinations of relevant parameters and estimation of accident risk

Elaboration method 1: investigation of scenarios for combinations of relevant parameters and estimation of accident risk

By varying the level of parameters (mainly daytime, road configuration, road global conditions, traffic parameters and traffic state), different scenarios are formed. Then frequency and type of accidents are

summarized per scenario or group of scenarios. The risk of accident for each scenario or group of scenario is compared to a base scenario in order to calculate the added risk of accident.

Modelling of vehicle distribution can be accomplished by microsimulation in order to better understand the phenomena for each scenario. Expected frequency of accident (accident risk) will be calculated on basis of a multifactorial regression function. This function will be investigated in Task 4.5 and will be reported in deliverable D4.3.

User perspective – feedback from highway providers

In Austria the use of the indicator accident risk/ accident rate is planned to provide information on accident cost and accident cost rates for the Austrian statistical data collection.

In Southern Germany the accident rate per road section and time period (1 year, 3 years) is used for evaluation of road sections and necessary improvements of road construction and maintenance.

Furthermore, there exists an ongoing research project in order to improve and automate data retrieval and management of accident data (EU project FAIR, German project VULKAN).

In Poland the indicator is used for safety evaluation of existing roads, road construction and winter maintenance. This is in the framework of an ongoing program of reducing road sections with increased accident risk on public roads.

Indicator Sa2: macroscopic accident probability (based on current speed in relation to equilibrium speed)

A possible further safety indicator is based on the concept of equilibrium speed . The assumption is that when the current speed is above the equilibrium speed longer than a given duration, the risk of collision rises. This is caused by the fact that drivers still drive at the same speed for a while, although the local density has risen (and the headways decreased). So they provisionally accept shorter headways than usual for this level of speed, until a new equilibrium is reached through a general decrease of the flow speed. This overfusion phase is risky, and the slightest disturbance (e.g. if a car breaks or changes its lane) can lead to an accident.

Sa2	macroscopic accident probability
Geographical scope	point, segment
Temporal scope	minute
Parameters	• traffic flow, local speed, occupancy rate • percentage of HGV • weather conditions
Elaboration method	Time during which current local speed exceeds equilibrium speed

User perspective – feedback from highway providers

In Germany the necessary data is available, but indicator is not calculated.

Indicator Sa3: microscopic speed inhomogeneity indicator

In task 4.5 a microscopic model will be calibrated with available traffic performance and road global condition parameters. Microscopic indicators for accident risk will be calculated and compared to actual accident rate at the planned test sites.

The number of lane changing constitutes the risk exposure to lane changing accidents. On motorways, lane assignment varies with the number of lanes and traffic flow. In order to arrive at the desired lane,

some cars change lanes after each entrance, and, on the other hand, some cars change lanes in order to join the exits. Moreover, the attractiveness of a speedier lane implies some lane changing.

Sa3	microscopic speed inhomogeneity indicator
Geographical scope	segment
Temporal scope	minute
Parameters	• time headway • speed, relative speed • position of vehicle in platoon
Elaboration method	1) investigation of scenarios combining different levels of speed and speed inhomogeneity

Elaboration method 1: investigation of scenarios combining different levels of speed and speed inhomogeneity

Unfortunately lane changing cannot be easily measured. A minimum number of lane changes can be computed from the traffic flow on each entrance and on each exit in order to reach the observed lane assignment structure. The speed difference between lanes implies additional lane changing, a few literature citations report some numerical elements on this phenomenon.

A safety indicator may result from the combination of floating car data (speed, speed variation, headway) with road sensors data (average speed or occupancy on a road section) with data fusion methods.

Two uses of data fusion for safety indicators are proposed below addressing

1) Speed

The fusion of the floating car speed, with speed recorded by the road sensor, will show the traffic homogeneity, which is linked to safety (at least for high traffic flow), for :

- the frequency of rear-end collisions increases when there are important differences between the speeds of two consecutive bunched vehicles on the same lane
- the lane changing frequency also increases when there are important differences between the speeds on two adjacent lanes

Differences of speeds between vehicles using the same lane or between lanes are usual. First these differences of speeds have to be studied on an historical baseline case, based on a set of data where the homogeneity of the traffic speed is already assessed as normal . Then, in the real time, we have to compare the actual differences of speeds with those of the baseline case, in order to state if the traffic is more or less homogeneous than in the baseline case.

High differences between the floating car speed and the speed deduced from the road sensor will show inhomogeneity.

If several floating cars are travelling on the same road section on the same period, differences between their speeds will also show inhomogeneity.

As floating cars are commercial vehicles (taxis, hauliers, buses), their speeds may be different from common cars : therefore an empirical study is needed in order to set, for the baseline case, the speed of the floating cars related to the average speed measured on the road sensor.

In the case of urban networks, this empirical study is particularly needed, because the speed recorded by the road sensor is related to the position on the lane of the sensor and is not the average speed of the vehicles on the whole road section.

The temporal resolution of the floating car data is an important topic: It is not the same information if the floating car speed is available each half of a second or for a longer period (e.g. 5 minutes).

1) In the first case, several consecutive speeds are available from the same floating car on a same road section; the fusion of these different floating car speeds with the road sensor speed will enhance the knowledge of the evolution of the speed of that section.

- the speed of a floating car vehicle at the very place where the road sensor (inductive loop or radar) may be correlated to the average speed recorded by the road sensor.
- the average speed of a floating car vehicle on the section may also be correlated to the average speed recorded by the road sensor

This fusion process addresses two indicators: the traffic homogeneity indicator (linked to safety) and also the travel time indicator.

2) In the second case (when the floating car data period is higher, e.g. 5 minutes)
The average speed of a floating car vehicle must be correlated to the set of road sensor speeds for the sections travelled by the vehicle during the period.

In this case, the speed variation will be particularly needed (see the next point).

The crossing between the enhanced value of speed and density (provided by the occupancy observed with the loops) is a safety indicator (situations characterized by high speed and high density are unsafe).

2) Speed variation

In the fusion process, the introduction of the speed variation of the same floating car on a road section will enhance the traffic speed inhomogeneity indicator, because an average speed of the floating car may hide the case where this vehicle experiences sequences of high and low speeds. The baseline study will provide which speed variations are normal . From this, low and high speed variation levels may be identified.

With three levels of speed and three levels of traffic speed homogeneity, nine scenarios appear:

Speed and Traffic speed inhomogeneity

		Speed		
		Low	Baseline	High
Traffic speed inhomogeneity	Low	L-L (1)		
	Baseline			
	High	L-H (2)		H-H (3)

(1) corresponds to a congestion
(2) corresponds to the beginning of a congestion
(3) unsteady case, potentially unsafe
Note that for low traffic flow situations, the traffic speed inhomogeneity does not result in an unsafe situation.

User perspective – feedback from highway providers

In Germany, for the microscopic lane changing indicator there are no single vehicle data available.

User perspective – additional indicators from highway providers

In addition, the indicator *Time to Collision (TTC)* was quoted which is planned to be used as a microscopic indicator for research purposes. (method of calculation: headway/relative speed; main

parameters: headway, relative speed between following vehicles). Another additional indicator quoted by highway providers was *Headway* Control Warning which is already in use for collective and individual traffic control with VMS (method of calculation: time headway compared to a given threshold; main parameters: average and individual time headway, average and maximum speed, traffic flow).

Indicator Sa4: transport force

In WP1, transport force has been identified as a possible indicator for severity of accidents (cf. D1.1 § 5.3.16, Transportkraft).

Sa4	transport force
Geographical scope	point, segment
Temporal scope	second, minute
Parameters	• weight • speed • time headway
Elaboration method	1) calculation of transport force

Elaboration method 1: calculation of transport force

Transport force is calculated from weight times speed divided by time headway of each vehicle ($G_i * \frac{v_i}{t_i}$) and summed up for all vehicles during one hour. This value can be compared to a given threshold [7]. If only single vehicle data and no weight measurements are available, the related transport acceleration can be computed from speed by time headway($\frac{v_i}{t_i}$).

User perspective – feedback from highway providers

In Germany, WIM measurement systems are in operation on selected road sections for statistical data collection and 2 semi automatic weight control stations are in operation (Project CITRA and TOP TRIAL), however information is not integrated with traffic control data available.
In France a network of over 100 WIM sites is operated.

In Poland, the indicator transport force may be used in the future (by 2010) after the weight enforcement system will be installed. In some areas, it is used locally in the place of road weight installations.

2.7 Conclusion

As a first preparatory step of the project, an inventory of traffic indicators, methods for their computation or estimation and existing data sources for sensor-based measurement of traffic parameters have been compiled. At this state of analysis one indicator related to road network global conditions (stopping distance), one indicator related to supply (road capacity), two indicators related to demand-supply interaction (traffic density, travel time) and three indicators directly related to safety accident risk or accident rate, macroscopic accident probability, microscopic speed inhomogeneity have been identified and discussed.

2.8 References

[1] Antoniou C. (2004). On line Calibration for Dynamic Traffic Assignment (PhD, MIT, Boston, September 2004)

[2] Antoniou C., Nowotny B., Rechbauer A., Linauer M. (2006). Calibration of DTA models using floating car data: An application of DynaMIT in Vienna, International Symposium of Transport Simulation 2006 (ISTS), September 2006, Lausanne, Switzerland

[3] Brilon W., Regler M., Geistefeldt J. (2005): Zufallscharakter der Kapazität von Autobahnen und praktische Konsequenzen (Randomness of capacity on motorways and its practical consequences). Straßenverkehrstechnik 3 (2005)

[4] Excerpt from State-of-the-art and User Needs Report. INTRO WP1 report, Chapter 2.3 (2005)

[5] LCPC, arsenal research, VTI (2006). Draft report of INTRO D 2.1 Model for Estimating Expectable Braking Distance

[6] Leurent F. (2004): A model and a rigorous measurement method model for flow capacity , Aron M., Boillot F. and Lebacque JP. (eds), Modélisation du trafic, actes du groupe de travail 2002. Actes INRETS #97, pp. 9-30. INRETS, France

[7] Franke, L., Brannolte, U,. (2000), Transportkraft als dreidimensionale Verkehrsgröße (transport force as three-dimensional traffic measurand). Sonderdruck Straßenverkehrstechnik Heft 11/2000 (www.lotharfranke.de)

[8] Hoogendoorn, S. P. och Botma, H. (1997). Modeling and estimation of headway distributions, Transportation Research Record, 1591, pp. 14-22.

[9] Carlsson, A., Wiklund, M. and Yahya, M-R. (2006). Estimation of traffic capacity on the Essingelink. VTI Manuscript 2006-03-31, VTI, Sweden.

[10] Lorenz, M. and Elefteriadou, L. (2000). A probabilistic approach to defining freeway capacity and breakdown. Proceedings of the 4th International Symposium on Highway Capacity, pp. 84-95. TRB-Circular E-C018, Transportation Research Board, Washington, D.C.

[11] Brilon, W., Geistefeldt, J. and Regler, M. (2006). Randomness of capacity idea and application. Proceedings of the 5th International Symposium on Highway Capacity and Quality of Service, pp. 147-157, Yokohama.

3 Floating car based indicators

3.1 Estimation of travel times with FCD

Real time estimation and forecasting of traffic parameters is one of key issues in traffic management. In particular journey times are of significant interest for traffic engineers and also for normal road users. One of non conventional sensor based methods is Floating Car Data (FCD) systems.

FCD-based parameter estimation have successfully been demonstrated in previous research projects and during the last years also been implemented in operational traffic systems. By improved data processing there is a great potential for more reliable parameter estimates and more efficient use of available probes (equipped and reporting vehicles). The number of probes and total communication cost can then be reduced without serious effects on accuracy.

This chapter goes deeper into two different floating car data systems, evaluating them and comparing their performance via the usage of microsimulation, which have become an invaluable tool to describe the dynamic traffic behaviour of traffic, and also to analyse this methods. Developed methods have been verified by micro simulation studies.

3.1.1 Definition and concept of travel time

The concept of travel time for a single vehicle is the time spent for going from a point A to a point B. In those cases where these two points are the boundaries of a link in a network, travel time is called Link travel time (LT). However, this concept is not so trivial when it is referred to flow of vehicles.

3.1.1.1 The deterministic approach

The deterministic approach defines the idea of travel time for a flow as a unique value. It is a simplification or model of the phenomena, by means of an aggregation of vehicle travel times during that period of time. So the travel time for that period is considered to be constant.

The deterministic approach is then an aggregation of individual travel time. Instinctively the average travel time seems to be the most representative of the individual data collected. Accordingly the average travel time between a point A and a point B can be defined as:

$$\overline{T}_{AB,\tau} = \frac{1}{N_\tau} \sum_{n=1}^{N_\tau} T_{AB}^{n} \qquad (3.1)$$

Where N_τ is the number of vehicles that has passed through point B during the time interval $T = [\, t\, ,\, t+\Delta t]$, Δt as the aggregation period (abbreviated AP).

This approach to the travel time concept is the most used, with A and B as the boundaries of a link in a network, or those of a complete path (a group of consecutive links). For the first case it is defined as the Average Link Travel Time (abbreviated ALT). For each path there is more than one itinerary to join points A and B. Both approaches (taking in consideration or not the itinerary followed by the vehicles) have their utility, so they have to be different. Then, the Average Path Travel Time (abbreviated APT) for the itinerary *p,* is defined as:

$$APT_{p,\tau} = \frac{1}{N_\tau} \sum_{n=1}^{N_\tau} PT_p^n \qquad (3.2)$$

Where PT_p^n is the travel time for individual vehicle n (part of the N_τ vehicles that have finished going through all the links composing the path).

Then it is essential to show the difference between the APT and the Built Path Travel Time (abbreviated BPT) for the itinerary p, which is the sum of ALT calculated for every link in the path *p:*

$$BPT_{p,\tau} = \sum_{a=a_0}^{a_k} ALT_{a,t} \quad \text{with } p=\{a_0, a_1, ..., a_k\} \qquad (3.3)$$

Among the main differences between both concepts, the most important ones are:

- It may be the case that only a part of the vehicles whose LT is used for the ALT calculation would finish the complete path p.
- The vehicle PT_P whose link a_j LT is used for the ALT_T calculation of that same link, cannot be used for $APT_{p,\tau+i}$, and being larger than the $APT_{p,\tau}$ is larger than AP.

This second point is shown in the Fig. 3-1, where 3 vehicles trajectories are marked in a space-time diagram. If the itinerary p is made of links A1, A2 and A3, then only V3 vehicle is taken into consideration for the APT_T calculation, which is equal to T in this case. During this same period T, only a vehicle has exit all 3 sections so $ALT_{A1,\tau}$ = t1, $ALT_{A2,\tau}$ = t2 and $ALT_{A3,\tau}$ = t3. Finally, BPT_T = t1+t2+t3 is different to APT_T.

The BPT is yet frequently used as an approximation to APT, as in practice, the calculation of ALT is clearly easier than those of APT. The higher ALT time variation gets and the lower AP gets, the higher the approximation error gets.

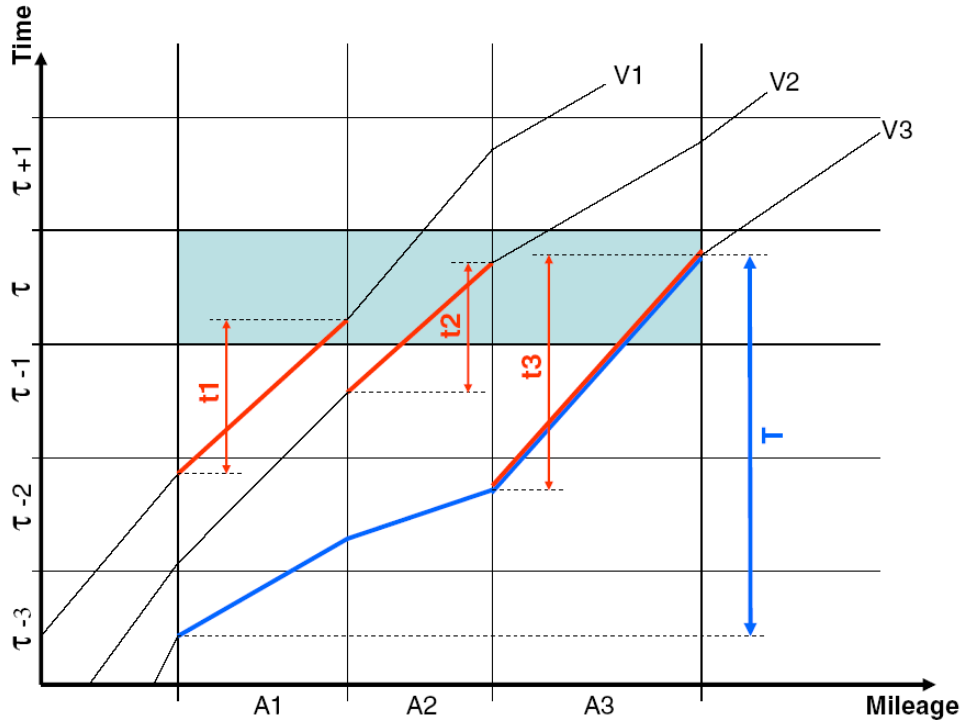


Fig. 3-1: Difference between BPT (red) and APT (blue) calculation

Fig. 3-2 shows the individual LT data of a link in a model representing the Lausanne city centre. This experiment was carried out by Torday in [1]. Model description is detailed further in chapter 3.5. It remarks the discretisation and aggregation explained above and representing the travel times per averages and medians.

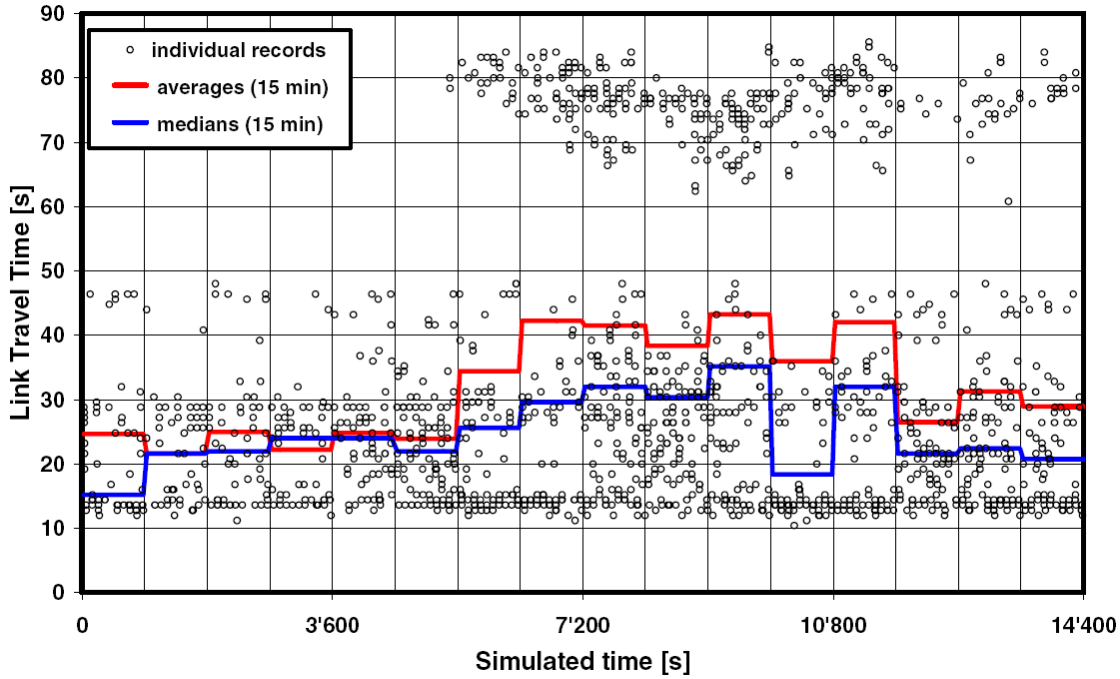


Fig. 3-2: Aggregation of LT data for averages or medians for 15 minutes [1]

The median is the justified representation of the travel time phenomena. It has a main advantage in front of the average as it is less sensible to extreme values,

The median is an accurate representation of the travel time. It is much less sensitive than the mean that has sometimes aberrant values. On the other hand for a LT bimodal distribution, the median can be quite variable from one aggregated period to another one if the sharing out of the population between two modes is balanced. The period going from 9 ' 900 to 10 ' 800 seconds of the graph of the Fig. 3-2 is an example. If the problem of the bimodality is raised here, it is that it is about a frequent scenario for an arc ending by a signalised junction and that they find as a result for an urban network. This subject will be treated in more detail later.

3.1.1.2 Urban networks particularities

A great part of studies carried out about travel time estimation are based on highways or rapid urban carriageways. These routes offer to drivers uninterrupted traffic conditions but the conditions may change with time. This continuity is interrupted only by incidents, and it is always treated as exceptional situations.

For two vehicles with similar driving characteristics (aggressiveness, desired speed, etc) going one near the other, then this time continuity in traffic conditions will mean similar travel times for the different links. Variance in measured LT population will be then explained by differences in driver s behaviour. On the contrary, urban type networks are under discontinuous traffic conditions due to the existence of signalized intersections. In fact, two consecutive vehicles may have very different link travel times if one goes through the link-ending intersection during the green phase and the other stops due to the red phase.

Then, the difference between extreme values in measured LT population during an aggregation period is similar to the link-ending intersection s red phase duration. As link lengths in this type of networks is

significantly shorter compared to interurban network type it is frequently noted that the variance coefficient standard deviation divided by the average exceeds 100%.

In Table 3-1 these main differences between the two network types are summarized.

Besides signalized intersections, urban networks also involve non signalized intersections usually with one primary road and other secondary entries/exits or roundabouts. Travel times distribution changes according to these junctions types.

These particularities have a great influence on FCD travel time estimation and prediction.

Table 3-1 Differences between urban and interurban network characteristics

	Interurban	Urban (signalized)
Traffic conditions	Continuous	Discontinuous
Link length	Long (from 0.5 to many km)	Short (10 to 200 meters)
Variance coefficient	<< 100 %	≥ 100%

3.1.2 Travel time estimation

3.1.2.1 State of the art

A key area of research within floating car data is determination of the minimum number of equipped vehicles present in the network, along with the associated filtering and analysis techniques employed.

There are two main data acquisition methods. In the first one, the data obtained by FCD systems are the regular transmission of vehicle position (i.e. every 3 minutes in [1]). Travel times for each link are then deducted in a traffic control centre using a digital GIS and a non-trivial time travel calculation algorithm for travel time calculation. The second method equips each vehicle with geographical information, so that they are able to register travel times between key points in the network (i.e. roundabouts, highway entrances or exits, points where number of lanes change in a link), which are already prescribed within the database.

Measurements by the second method are generally more precise but require a more complex on-board system, which means a higher investment.

The key travel time estimation is question is

> How representative is the travel time information obtained from equipped vehicles of travel times experienced by all vehicles in the network?

Many statistical models have been proposed to answer these questions. Van Aerde et al. [3] have developed an analytical expression to determine the FCD travel time s reliability. This expression, based on the hypothesis of statistical independence of measures, shows that if the percentage of floating vehicles grows, the average of the sample tends towards the average of the population. Other research work is based on this hypothesis to calculate the necessary equipment rate [3], [4] and [5].

Contrary results are obtained from analysis of FCD obtained in projects [4], [5] and [6], in which the authors show that samples are not independent. This results in covariance between their speed values. Average Link travel time is stabilised above a certain equipment ratio, although it may not match the whole population average.

It seems, based on the studies carried out on this subject [7], [8], [9] and [10] that it is impossible to define an ideal equipment ratio for all situations and network types. In fact, parameters such as the reporting interval the frequency of data transmission - and the length and structure of the links in the network have an important influence on determination of a representative global value.

Depending on the specific situation, an ideal equipment ratio between 3 and 7% was recommended.

3.1.2.2 The FCD estimation question

The approach normally used consists of approximating travel times calculated from all the vehicles that have gone through a link or a path using only registered data from a limited and usually low number of - floating vehicles - thereby obtaining an approximation to a representative parameter, an average or median TT of the population by means of a sample.

According to the determinist approach and using ALT as a representative parameter, FCD technique tries to approximate to the average of the a link travel time values Nτ with those of the sample floating vehicles Estimated Link Travel Time, abbreviated ELT - :

$$ALT_{a,\tau} = \frac{1}{N_{a,\tau}} \sum_{n=1}^{N_{a,\tau}} LT_a^n \qquad (3.4)$$

$$ELT_{a,\tau} = \frac{1}{F_{a,\tau}} \sum_{f=1}^{F_{a,\tau}} LT_a^f \quad \text{with } F_{a,\tau} £ N_{a,\tau} \text{ and } E_{a,\tau} = \frac{F_{a,\tau}}{N_{a,\tau}} \qquad (3.5)$$

where Fa,T is the number of floating vehicles in link a and period T

This is the most common expression. The percentage of floating vehicles is varies from one link to another, as well as varying over time from one aggregation period to the next, even though the equipment ratio is fixed.

The estimation error εa,τ is then defined as the difference between the estimated value and the target one:

$$ALT_{a,\tau} = ELT_{a,\tau} + \varepsilon_{a,\tau} \qquad (3.6)$$

The statistical distribution form influences evidently in the error. In fact, for a fixed population size and a fixed percentage of floating vehicles, estimation error will be more important than population s variance. It might be considered an average error, as the Fτ floating vehicles are chosen randomly from the population Nτ vehicles, so each experiment means a different sampling average, so a different error.

As previously mentioned, urban networks have the particularity of having a high LT population variance. Then, a less performant approximation compared with those in highway networks will be expected, or a higher percentage of floating vehicles would be needed in order to have a similar estimation quality.

Decreasing LT population variance is one of the ways to improve the quality of FCD travel time estimation. It is evident that observed traffic can not be influenced by the measurement technique and only a modification of spatial and time limits allows a variance reduction. So it is especially interesting to study:

- Aggregation period duration

- Links definition

These two elements play a major role in the estimation process.

In conclusion, it is necessary to use simulation evidence for methods and studies comparison. In fact, simulation allows floating car data to be obtained, and well as those for the remaining vehicles the complete population while real world experiments only provide floating vehicles measures.

Although simulation can provide floating car data as well as data for the rest of the population, the observed data will only cover that of the floating cars.

Moreover, this advantage allows parametric analysis to be made based on certain performance indicators.

3.2 Identified Floating Car

3.2.1 Introduction

Identified floating cars deliver direct information on current travel speeds for routes. If the vehicle fleet is sufficiently large in order to cover all important streets of a city and if cars can be tracked in short intervals, reliable speed measurements with low deviations can be expected. Several theoretical and practical investigations have been conducted in order to define the necessary fleet size or data frequency for a target data quality (see e.g. [15] and [16]). Under practical conditions, the actual tracking interval is limited by technical constraints (e.g. performance of fleet management system) or financial constraints (e.g. transmission costs for traffic data). Therefore, data from lower fleet sizes and with higher tracking intervals are delivered which nevertheless provide an important insight in traffic situations. For the description of data quality, simple and descriptive measurands have to be used in order to support practitioners in traffic information centres.

During the last decade a number of pilot trials have been carried out in different parts of the world (e.g. Europe, Japan, USA) in order to produce and evaluate floating car data. Important scientific projects in EU countries include

- project WAY-FLOW in the Rhein/Main region, Germany
- project OPTIS in Gothenburg, Sweden
- project RTMS (Road Traffic Monitoring by Satellite) in Rotterdam region, Netherlands
- Project STRIP on French motorways (within SERTI programme)

Only few projects entered into a commercial phase (e.g. iTIS Holdings in England with a fleet of 50.000 vehicles, FLEET system in Vienna with 800 taxis, PRO-NAVI system in Nagoya, Japan with 7.000 taxis and other commercially oriented systems based on taxi fleets).

In this study, speed data from identified floating cars have been investigated. As measurements of route speed in street networks are influenced by many factors, methods for calculation of average route speed have to take into account its variability. Vehicles on rural roads and intercity motorways can be regarded as individual vehicles which choose their speed independently from each other. On the contrary, vehicles on well-travelled intraurban motorways and intraurban roads are frequently influenced by other vehicles (in vehicle platoons), or they arrive at traffic signals at the same time. These influences may lead to similarity of their travel times which results in correlation of the individual route speeds. This similarity, which can be measured by the auto-covariance of route speeds, has been observed in a field test in the ADVANCE program [7]. The real variance of average route speed is a very important measurand for the reliability of the average route speed and therefore a central input parameter for further applications of route speeds. These may range from using travel time information in a traffic information centre or data fusion with other traffic parameters.

As route speeds from a taxi fleet only constitute a small sample of experienced drivers, accuracy of route speed measurements has to be compared to an external reference. AVI (Automatic Vehicle Identification) measurements, e.g. based on license plate recognition, can be used to evaluate whether route speeds represent the whole or average population of drivers.

3.2.2 Methods

Data analysis was based on data from the FLEET system in Vienna, which has been operational since 2004. A taxi fleet of 800 taxis equipped with GPS receivers delivers positions and time stamps to a taxi management centre at an interval of 15 to 30 sec. From the taxi management centre, data are transmitted every 15 to 60 sec and evaluated in 15 min intervals.
Message data from floating cars are cleaned in a three-step process in order to reduce spatial and temporal errors of taxi messages.

1) A map matching algorithm projects trip data (message points) to the road graph. Distance to the nearest road and heading of trip in relation to road heading are taken into account for map matching. Taxi message points, whose distance exceed a defined distance (of currently 200 m) and whose heading deviate more than a defined angle (of currently +/- 90 degrees), are discarded. Therefore the identification of the floating cars is very important, in order to be able to take into account heading information.
2) If only start and stop of trips are known or if GPS messages are sparse, an optimised shortest path algorithm is applied for route search. The algorithm uses several labels in order to reduce spatial deviation of routing. Trips are discarded, if the distance between two message points exceeds a defined length (currently 4 km) and if the trip length after routing is longer than airline distance times a defined factor (currently 2.5).
3) A travel time calculation algorithm calculates the estimated speed value for every road on a route. In order to take into account different speed levels of roads, average route velocity is distributed to each road proportionally to its free speed or historical speed. Then space mean speed (corresponding to ELT, estimated link travel time) is calculated from all speed values available for a road in a time interval (of currently 15 min).

As a basis for calculation of differences and variance of route speeds two models for average route speed were compared.

With a linear model (Eq.(3.7)) the variance of average route speed in each time interval $x_{d,t}$ is decomposed into an effect of the average speed, of the day class and the effect of the time interval.

$$x_{d,t} = \eta + \alpha_d + \beta_t + \varepsilon_{d,t} \qquad (3.7)$$

where:

η ... effect of grand average

α_d ...effect of day class average

β_t ..effect of time interval

$\varepsilon_{d,t}$. residual of model

In addition, historical time series of each route were calculated using the harmonic average of route speeds for each time interval in each day class. A classification scheme with 4 day classes was used for calculation (see also [18]).

Variance of average route speed was estimated by variance analysis. The squared residuals of the model ($\varepsilon_{d,t}$ in Eq.(3.7)) were used as an estimate of variance of route speeds within each time interval.

Time series analysis was applied to current and historical time series of route speeds [13]. As current and historical speed time series showed non-stationary behaviour, the speed differences between time steps were investigated. In previous investigations of speeds on motorways, an ARIMA (1,1,1) model, i.e. a model with an autoregressive operator of order 1 and a moving average operator of order 1 has shown a realistic description of the temporal speed evolution [14]. This ARIMA model was fitted to speed time series of 5 successive working days. The residuals of the model showed no remaining trends, and no autocorrelation of residuals was detected. Two related ARIMA models (ARIMA(2,1,1), ARIMA(1,1,2)) also were fitted to the data, but additional estimated weights of the autoregressive operator and the moving average operator were small in comparison to their standard error and the AIC (Akaike Information Criterion) even slightly increased. So the ARIMA (1,1,1) model was used and transformed to a state-space model for linear Kalman Filtering.

A Kalman filter was used to calculate the speed estimate with minimal variance (see e.g. [13]).The Kalman filter provides an appropriate solution for predicting the value at the next time step and calculating a corrected value from the predicted value and the current measurement value. Therefore,

this method does not only provide a solution for optimal estimation, but also for data fusion of consecutive values. Variance of process noise (Q) was estimated from a first run of the Kalman filter which outputs an estimate of the variance of the state vector, i.e. route speed deviation. This estimate is equivalent to the residual error of a least square regression of the transition equation. Variance of the measurement noise (R) was estimated by variance analysis. The squared residuals of the model ($\varepsilon_{d,t}$ in Eq.(3.7)) were used as estimate of variance of route speeds within each time interval.

Average route speed data of two major routes in Vienna, the Margaretengürtel with a length of 3.0 km (Route 69) and the Weißgerber Lände (Route 29) with a length of 2.4 km were analysed. Data were available from July to September 2005 and aggregated to intervals of 15 min. On these two routes, travel speeds were also measured by an AVI system on the right lane for approximately 4 hours on each route.

3.2.3 Comparison of historical or model timeseries

Both fitted values of the linear model (see Eq.(3.7)) and average values of historical timeseries represented a good model for current route speeds. In both cases residuals of the models were approximately normally distributed, and residuals showed no relation to the level of speed or the day class. Variance analysis on basis of the linear model showed that between 40 to 50 % of speed variance could be explained by the factors day category, time and their interaction.

3.2.4 Investigation of variances of route speed and effect of linear Kalman filter

Variance of average route speed was calculated from the squared deviations between the harmonic average of the timeseries models and current route speed from taxi data. On route 69 root of mean squared error (RMSE) for both models ranged between 5.2 and 5.6 km/h, on route 29 RMSE for both models was slightly higher (between 5.9 and 6.5 km/h). RMSE was used as an estimate for measurement standard deviation on a route within a time interval on a single day. On both routes a typical historical trend for speed evolution of working days could be identified (see CorrHistVal in Fig. 3-3 and Fig. 3-4). Current speeds of the observed ordinary working (see CorrMeasVal in Fig. 3-3 and Fig. 3-4) days were higher than historical speeds (between 3 to 4 km/h on route 29, between 2 and 9 km/h on route 69) and followed a trend similar to those of historical timeseries.

Standard deviation (sd) of estimation error converged to approx. 1.4 to 1.6 km/h on route 29 and to approx. 1.1 to 1.2 km/h on route 69. This is due to the fact that the Kalman filter outputs an estimation of speed trend. This trend is expected to have stable variance in contrast to variance of single speeds which depends on traffic density.

Missing values, which occasionally occurred for up to 3 time intervals, were replaced by predictions of the Kalman filter.

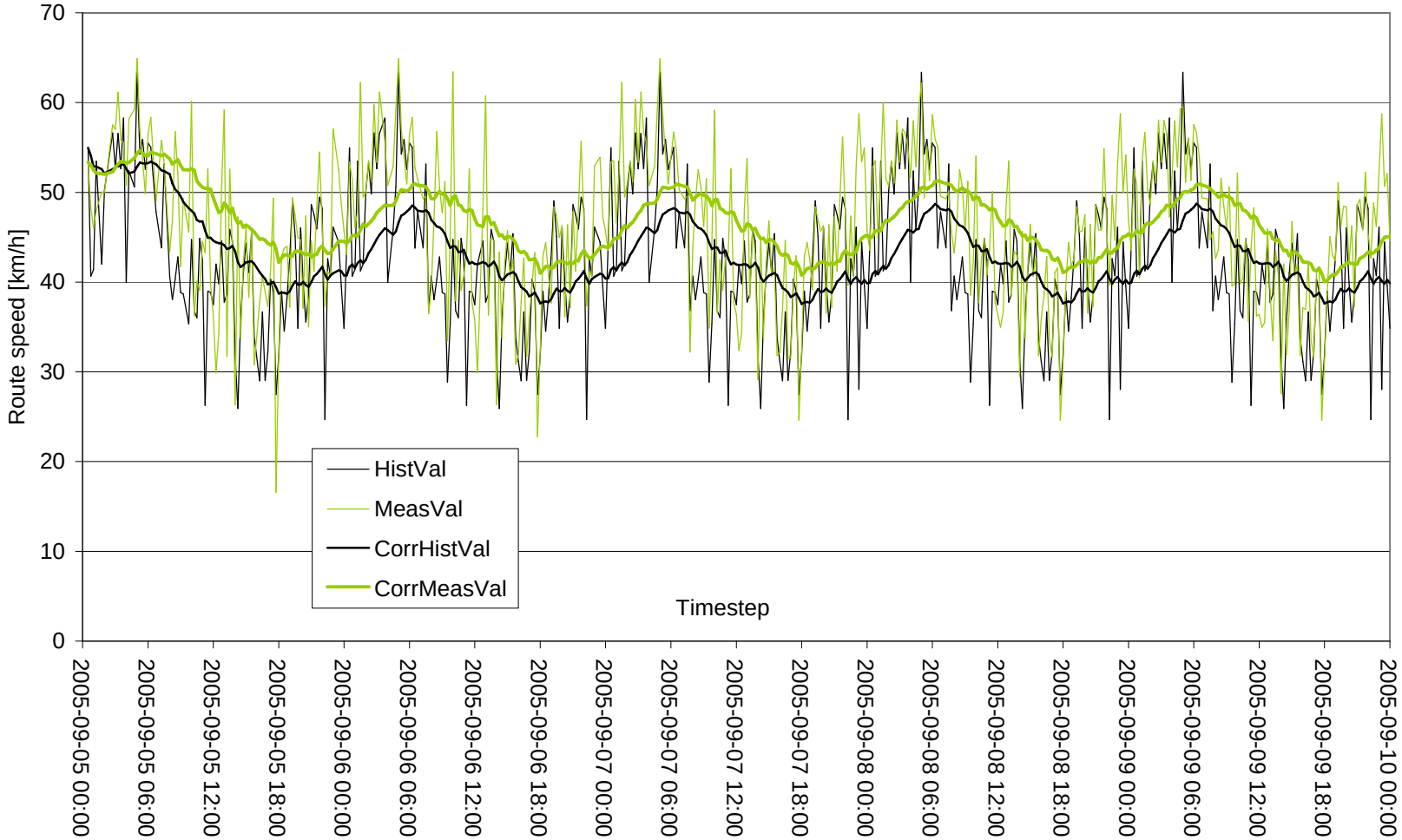


Fig. 3-3: Calibration of time series model for historical and current speed values (route 29)

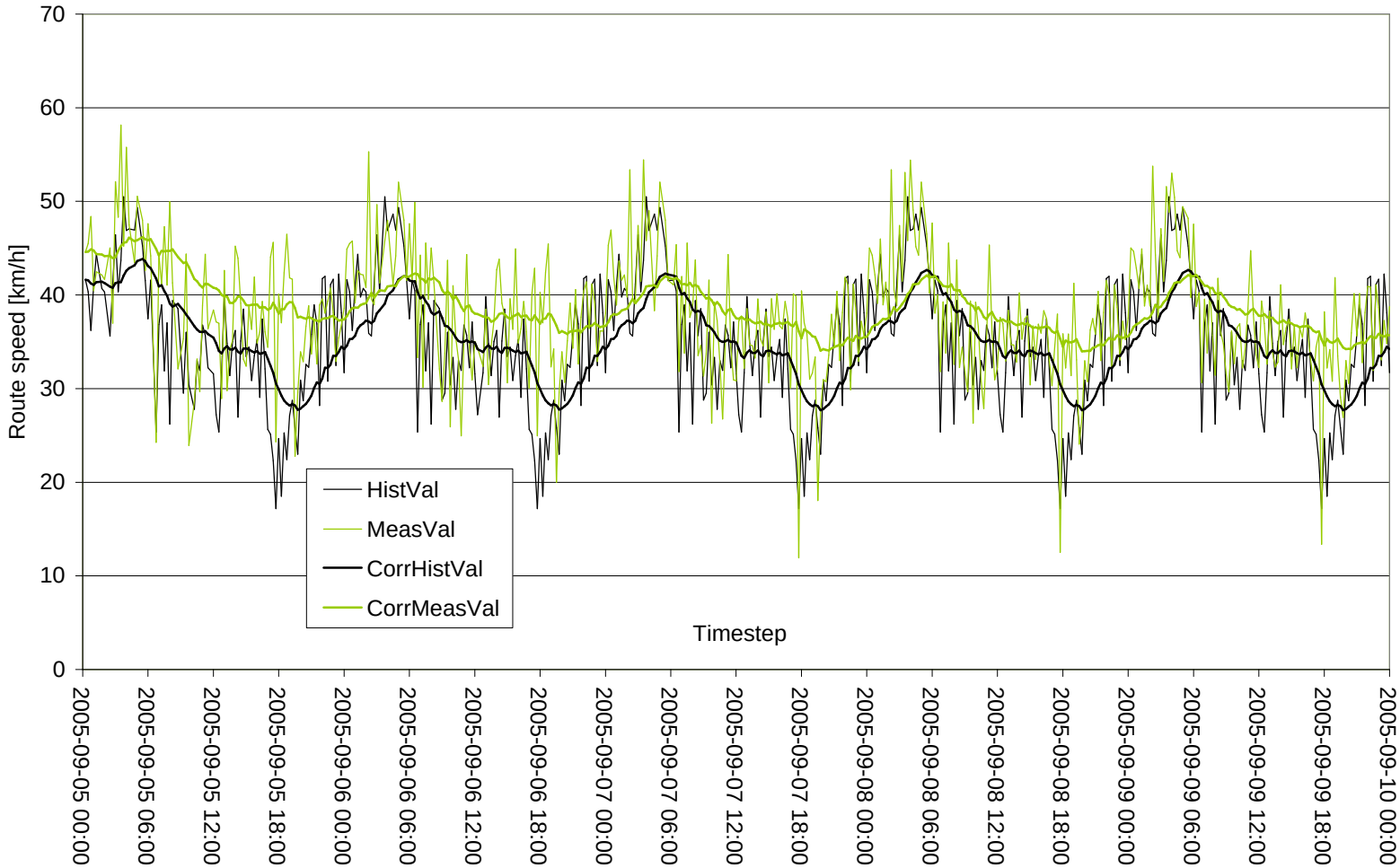


Fig. 3-4: Calibration of time series model for historical and current speed values (route 69)

3.2.5 Comparison of floating car route speeds and AVI route speeds

As route speeds from a taxi fleet only constitute a small sample of experienced drivers, accuracy of route speed measurements has to be compared to an external reference. AVI measurements based on license plate recognition which represent a random sample of the normal population of drivers were used to evaluate the difference between average route speed (corresponding to Average Link Travel Time ALT) and estimated route speed (corresponding to Estimated Link Travel Time ELT).

In the timeseries of both investigated routes variation of route speeds was high during normal hours and decreased substantially during peak hours (approx. 16:30 to 18:00 h on route 69 and 16:00 to 17:30 h on route 29). So standard deviation of route speed ranged between 6.7 during normal hours to 1.2 during peak hours on route 69, and between 13.1 during normal hours and 2.6 during peak hours on route 29 (see Fig. 3-5 and Fig. 3-6). Variance data show the wide range of possible single route speeds during normal hours, which is only reduced during peak hours.

Route speed data which have been filtered with a Kalman filter are closely related to the average route speed of drivers measured by AVI. Floating car route speeds are only slightly overestimated average route speed of drivers measured by AVI. Mean error between average floating car route speed (CorrMeasVal) and average AVI route speed on one lane (AvgspeedAvi) was +5.1 km/h on route 29 and +5.8 km/h on route 69 (see Fig. 3-7 and Fig. 3-8).

As no additional traffic parameters were measured, the high RMSE of approx. 9 km/h on route 29 and 6 km/h on route 69 between floating car route speed and average AVI route speed on one lane could not be explained. This may be due to differences in lane speeds in front of signalised intersections.

3.2.6 Conclusion

The investigation of route speed showed that variance plays an important role in the interpretation of results from route speed measurements. Average route speed on intra-urban roads can be estimated using timeseries modelling and linear Kalman filtering. Therefore, estimated route speed can be regarded as the relevant measurand which is related to other traffic parameters as traffic density and traffic flow (concept of equilibrium speed).

Historical speeds show typical trends of speeds. If their relation to current speed is described, this relation can be used for prediction and for imputation of missing values. One possible approach is based on the observed relation of deviations between current and historical route speed values. This model uses the temporal development of deviations from the typical link speeds on each route (see Eq.(3.8)).

$$\Delta x_{d,t} = \Delta x_{d,t-1} + \lambda_{t-1}$$
$$\Delta xm_{d,t} = \Delta x_{d,t} + \mu_t \qquad (3.8)$$

$\Delta x_{d,t} = x_{d,t} - x^{H}_{d,t}$ Deviation between estimated and historical value (of day d and time t)

$\Delta xm_{d,t} = xm_{d,t} - x^{H}_{d,t}$ Deviation between measured and historical value (of day d and time t)

λ_t Process noise

μ_t Measurement noise

State space model for deviations of route speeds

Variance of route speed deviations has a contribution from variance of route speeds and variance of the typical value of the historical timeseries. As the typical value of the historical timeseries is an

average of single route speed values, variance of single route speeds and (possible) covariance of historical timeseries are the dominant components for calculating variance of route speed deviations.

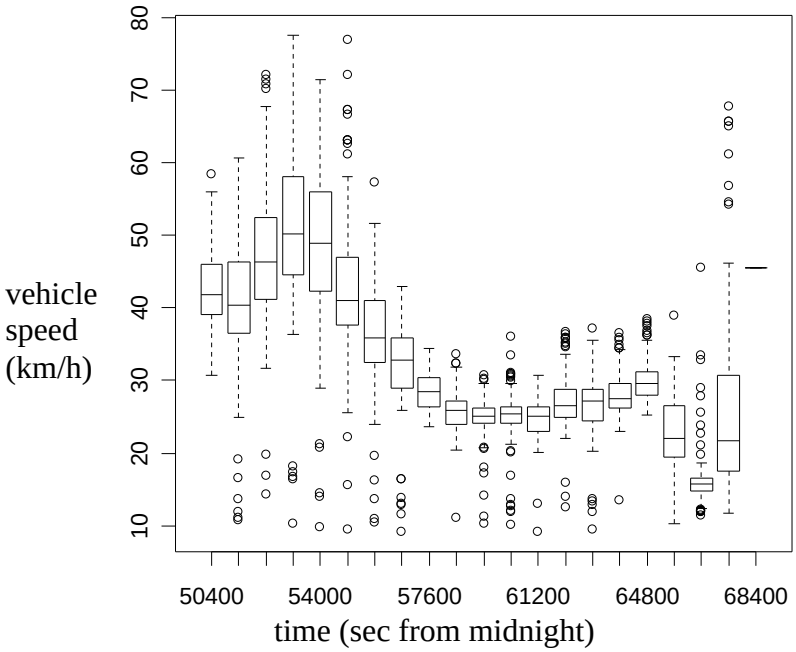


Fig. 3-5: Timeseries of single vehicles speeds in (route 29, 2005-09-28, 14:00 to 18:45)

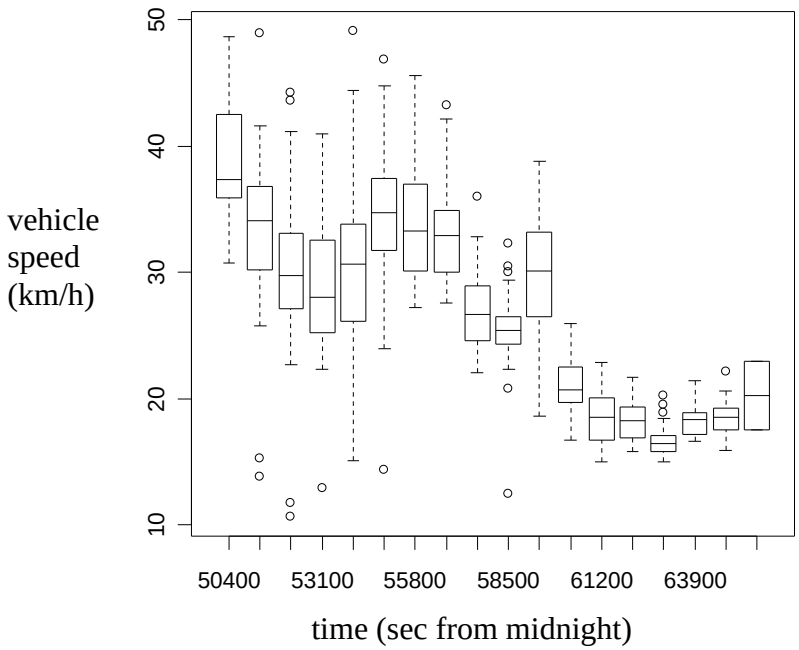
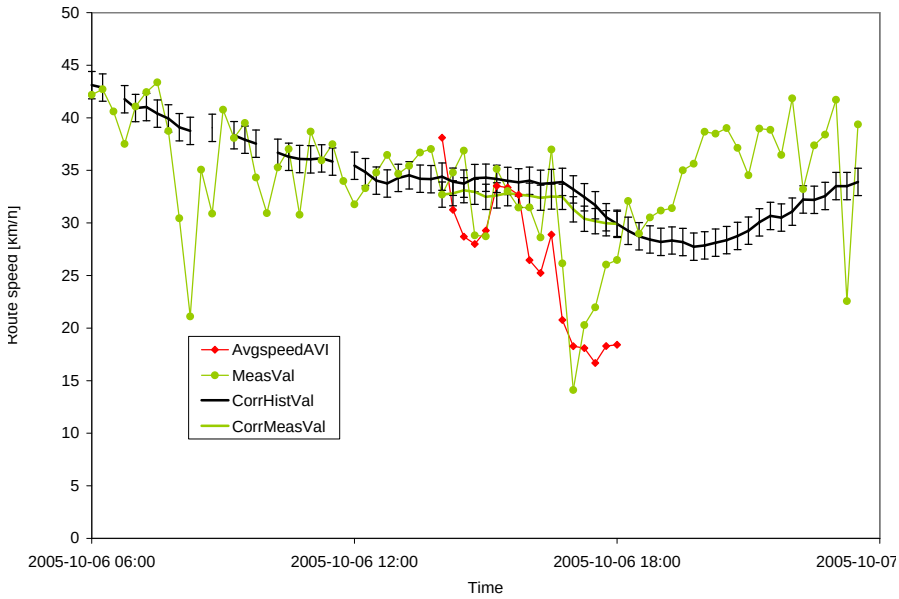
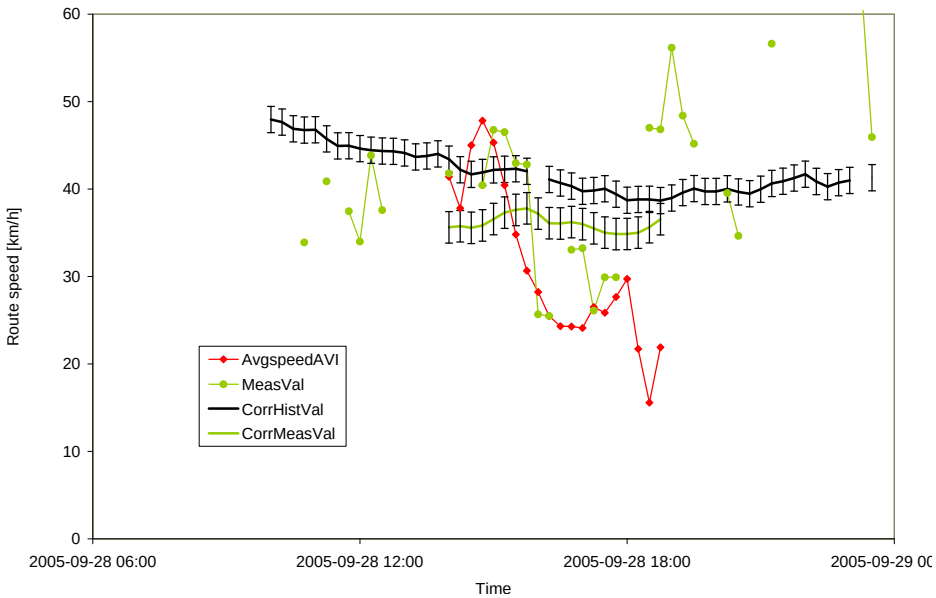


Fig. 3-6: Timeseries of single vehicle speeds in (route 69, 2005-09-28, 14:00 to 18:45)



AvgspeedAvi ... current route speed measured by AVI system
MeasVal ... current route speed measured by floating cars
CorrHistVal ... filtered value of historical timeseries
CorrMeasVal ... filtered route speed measured by floating cars

Fig. 3-7: Comparison of route speeds (route 29, 2005-09-28, 14:00 to 18:45)



AvgspeedAvi ... current route speed measured by AVI system
MeasVal ... current route speed measured by floating cars
CorrHistVal ... filtered value of historical timeseries
CorrMeasVal ... filtered route speed measured by floating cars

Fig. 3-8: Comparison of route speeds (route 69, 2005-10-06, 14:00 to 18:00)

3.3 Unidentified Floating Car

3.3.1 Introduction

3.3.1.1 Defination of unidentified floating car

The rapid development of information technology and communication systems provides us with a whole new set of tools for monitoring the traffic state everywhere in a network. It is easy to imagine a future scenario in which data from individual vehicles are exploited to estimate congestion state, travel times, incident detection, etc. In principle, each vehicle in the network could deliver frequent position and velocity data which would yield a rather complete picture of the traffic. It can however be anticipated that, for integrity reasons, many drivers will be unwilling to reveal their whereabouts in detail. Also, transportation companies may not want to expose their routes. The acceptance of a monitoring system of individual vehicles will probably be much higher if it can be guaranteed that no tracking of the position of the vehicles will be possible. This could be achieved by deidentifying data the moment they leave the vehicle. An *unidentified floating car* is a vehicle that delivers data that are completely anonymous, i.e., there is no way to deduce from which vehicle they are generated. As a consequence, there are no means to connect two separate observations generated by the same vehicle. All observations are of equal footing independent of which vehicle they originate from.

3.3.1.2 Type of data collected

In principle, data generated by an unidentified floating car can be of various complexities. One may imagine an advanced system where equipment in the unidentified vehicle collects large amounts of data (e.g., position data every second) and using knowledge of the road network, performs the data processing to compute the travel time for the previous link passed by the vehicle. Finally, the floating car sends the computed link travel time without identification to a traffic control unit. This system presupposes that the vehicle carries not only a position measurement device but also a data processing unit as well as road network data. Besides the extra cost of such processing devices, the drivers might regard them with suspicion. In a less sophisticated system needing no extra data processing, an unidentified observation might consist of the current position as well as position data for one (or even several) previous occasion. In both these examples travel time could be estimated using the same techniques as for identified floating cars discussed in section 3.2.

The simplest possible data from unidentified floating cars useful for travel time estimation might consist of position data only. The density of observations may then possibly reveal something about the density of traffic and, hence, also the travel time. The travel time estimated this way will, however, be very uncertain. By measuring also the current velocity, the travel time estimates can no doubt be improved without much increasing the complexity of the system.

Important other factors, besides the integrity issues, when choosing between possible data collecting systems for unidentified data will be the capacity and cost for data transfers as well as technological advances.

3.3.1.3 Experiment sites

Systems for travel time estimations have been (or are being) developed in many cities. Most such systems rely, however, on stationary detectors. Only a few seem to take unidentified floating cars into account.

In Sweden (Stockholm, Gothenburg and Malmö) a system (STRESS, see [19] and [20]) for estimation and prediction of travel time and congestion rates has been developed during a number of years. Four types of data sources are used in the system:

- Stationary detectors such as inductive loops and radar stations

- Dynamical detectors (probe vehicles) that produce unidentified observations (this subsystem was previously called PROBES)
- Camera systems yielding travel times along specific routes
- Probe vehicles yielding travel times along arbitrary routes (identified observations)

Travel time estimates from the various sources are fused together resulting in an overall estimate for each link.

Although the probe vehicle system yielding unidentified observations has existed since 2002, data are still not being used in the fusion process. This is partly due to the very limited number of vehicles connected to it, partly because it has not been sufficiently validated for operational use. Until summer 2006 only 50 vehicles were active, mainly buses and delivery cars. In September 2006, 800 taxi cars have been equipped with GPS devices and connected to the system. By January 2007, the system will include a total of 1500 taxi cars. The system has originally been designed to exploit velocity data. Unfortunately, due to low bandwidth, it is currently not possible to transfer velocity data from these vehicles, except for time and position of these vehicles. Therefore, it seems that these data will only be useful if the vehicles are considered as identified floating cars. The STRESS system still relies mainly on observations from stationary detectors.

In the present study the probes vehicle subsystem of the STRESS system in Stockholm is taken as a starting point and as a reference project.

3.3.1.4 Purpose of this study

In this study an observation from an unidentified floating car is assumed to consist of four instantaneous quantities:

- Time
- Position
- Direction
- Velocity

This calls for entirely different methods for travel time estimation than for identified floating car data. The purpose of this INTRO subproject is to improve existing methods for estimating travel times from such data. The focus has been on the algorithmic development. In section 3.3.4.3 we propose a new method for travel time estimation using spatial models of the velocity profiles of each link obtained from a priori data. From these velocity profiles the travel time along the links can easily be estimated. In the subsequent sections various properties of this method as well as implementation details is discussed. In section 3.3.5 a comparison between the new method and existing methods is carried out based on simulated data.
No effort has been made to develop methods for travel time predictions; only the travel time at the current time point is estimated.

3.3.2 Test bed

3.3.2.1 Test bed used for the study

It was the original intention in this project to use the Stockholm STRESS system as test bed. The idea was to use data from the unidentified probe vehicles to improve on existing methods for computing travel times on links. The evaluation of these methods would then be done by comparisons with true travel times obtained from the other data sources within the system (e.g. from the camera systems).

However, during the course of work it became clear that the amount of available Stockholm data as well as their quality were insufficient. The idea to instead use simulated data as a basis for the methodological development started to grow. There are a number of advantages of this approach; a much better control over what is happening is achieved and much more data is available for validation and evaluation purposes. On the other hand, various sources of disturbances and other deficiencies of real data are lacking in simulated data. An additional study of real data is therefore desirable.

Thus, two test beds have been used in the development work: a small network modelling a region in the southern suburbs of Gothenburg and real data from the taxi fleet in Vienna. (In addition, in chapter 3.6, the methods developed here will be applied to simulated data on a larger network from Lausanne.) Each of these is described in more detail below.

3.3.2.2 Simulated data

The network models two traffic junctions along Highway E6 and their surroundings, located 6 km south of Gothenburg centre (see Fig. 3-9). The distance between the northern and southern ends of the network is approximately 3 km. The network consists of 286 straight sections, which make up 92 polysections. These are regarded as the links in the network. Two OD matrices have been used for simulations, one corresponding to peak hour traffic in the morning and another corresponding to more sparse traffic conditions. Other OD matrices have been constructed by combining these original matrices and by scaling them. Simulations have usually been done for a period of 90 minutes. During the last 30 minutes the OD matrix is set to zero in order to empty the network. In the dense traffic scenario the maximum number of vehicles simultaneously in the network is 640 and the total number of vehicles appearing in the network during a 60 minute simulation is 10314.

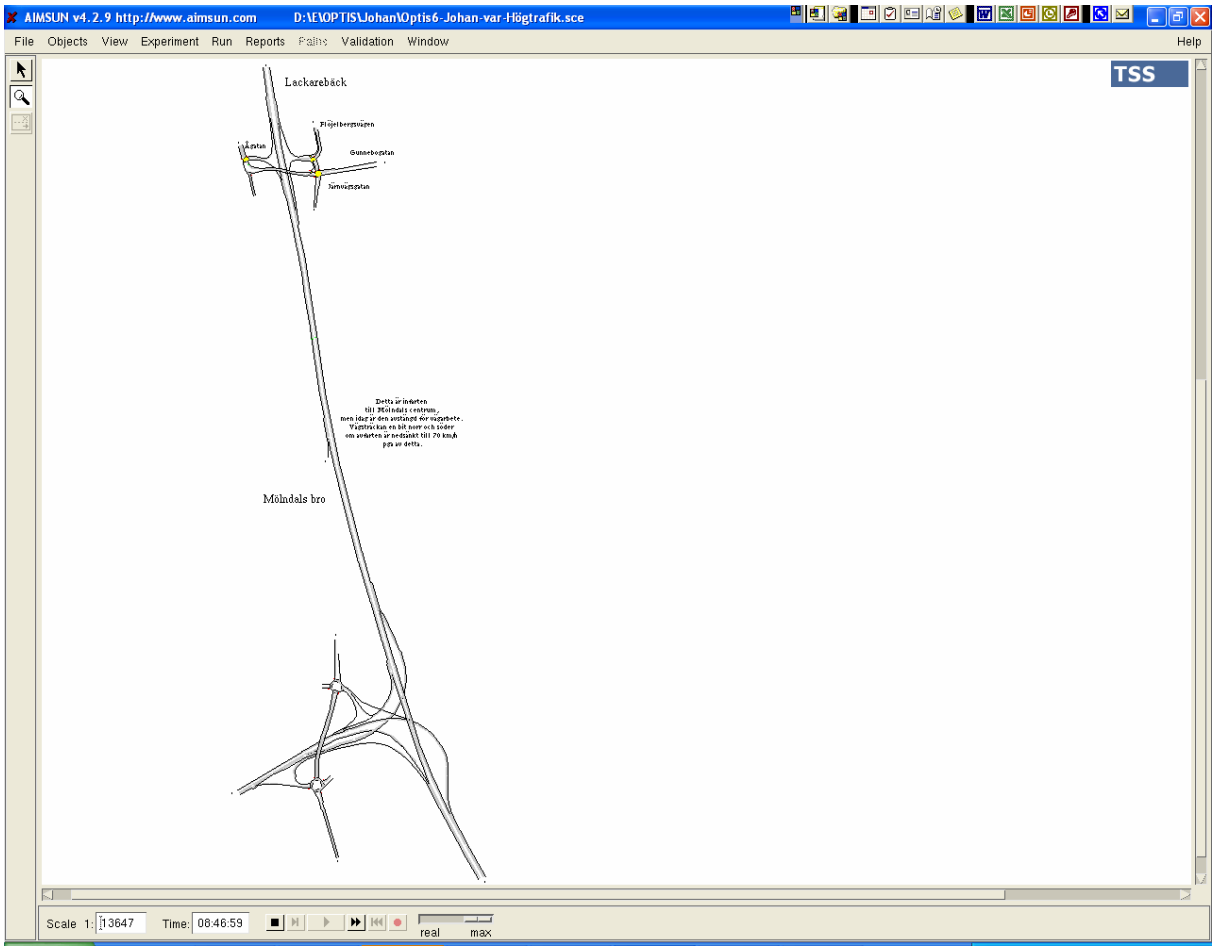


Fig. 3-9: The Gothenburg network

In order to be able to evaluate the methods thoroughly, not only the four basic quantities mentioned above were collected, but also vehicle identity, section identity, position within the section, lane and section entry time were collected. Instead of using vehicle coordinates for the observation the

positions within the sections are used. Map-matching to a stylistic network has not been performed, giving rise to some other problems discussed in section 3.3.4.9.

The sampling period has been chosen to 0.2 seconds, rather than the usual 0.6 seconds, thereby achieving higher accuracy in the velocity profiles and in the positions of the vehicles. The total number of samples during a 90 min simulation for the dense traffic scenario is 11 million and the file containing them is 900 MB. Although the network is small, a rather large amount of data is produced.

3.3.2.3 Real floating car data from Vienna

In Vienna, a large number of taxi cars have been equipped with GPS devices. Data from these cars can be considered as identified FCD and has been used in chapter 3.2. By deidentifying these data (i.e., removing the identity of the vehicle that generated them) they can be used also as unidentified FCD. Floating car data, in form of position and velocity, from taxi cars have been collected during two days. Since data were collected for the identified case, the speed data is not instantaneous, but describes the average velocity between two events usually separated by approximately 30 seconds, but in some cases even more. Thus, data are not perfectly adapted to the unidentified case.

Simultaneously, for validation purposes, AVI data for two different routes are available. These have been obtained from stationary cameras at the end points of the routes and have been applied to all vehicles passing travelling along the routes.

3.3.3 Application of the unidentified floating car

The purpose of travel time estimates from unidentified floating car observations is essentially the same as for identified floating cars (see section 3.2).

3.3.4 Development of new algorithms for travel time estimation

In this section a new approach to compute link travel time estimates from unidentified probe vehicle data is presented. The basic idea is to model the (current) velocity profile along each link. A number of relevant issues and problems are also discussed.

3.3.4.1 Existing methods

The methods for estimating travel times from unidentified probe observations used in the Stockholm STRESS system are based on piecewise constant velocity functions. One approach suggested and investigated by Westman [22], consists in partitioning each link in a fixed number of subsegments of equal lengths (see Fig. 3-10). For a specific time period the travel time is computed for each subsegment. The travel time within each subsegment is computed by $tt = s / \tilde{v}$, where s denotes the length of the subsegment and $\tilde{v}$ is the harmonic average of the observed velocities.

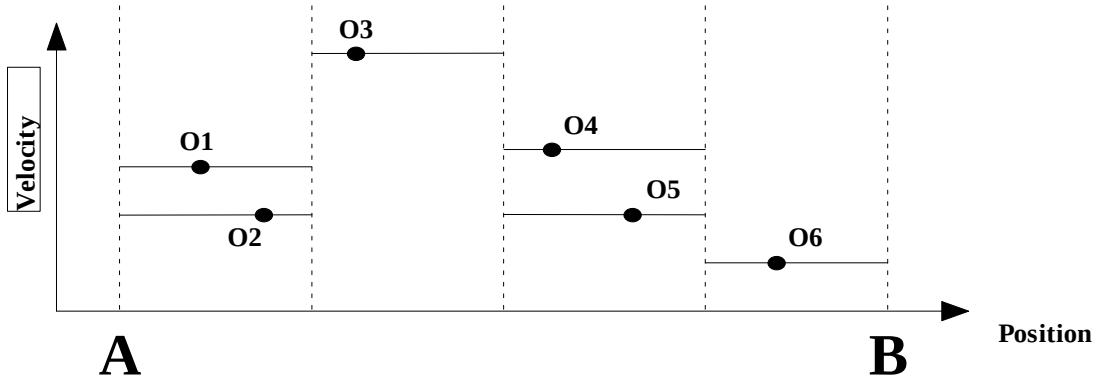


Fig. 3-10: Method suggested and investigated by Westman [22]. O1, O2, O6 denote observations from unidentified probe vehicles.

Another suggestion, proposed by Brorsson [21] and implemented in the current version of the Stockholm STRESS system, is based on a segmentation of the links, defined by the actual observations (see Fig. 3-11). The travel time is computed by $tt=\frac{s_0}{v_1}+\sum_{i=1}^{n}\frac{s_i}{v_i}$

where

s_0 denotes the distance between point A and the first observation

s_i denotes the distance between observations Oi and $O(i+1)$

v_i denotes the velocity for observation Oi .

The velocity in the first observation also determines the estimated travel time during the initial subsegment.

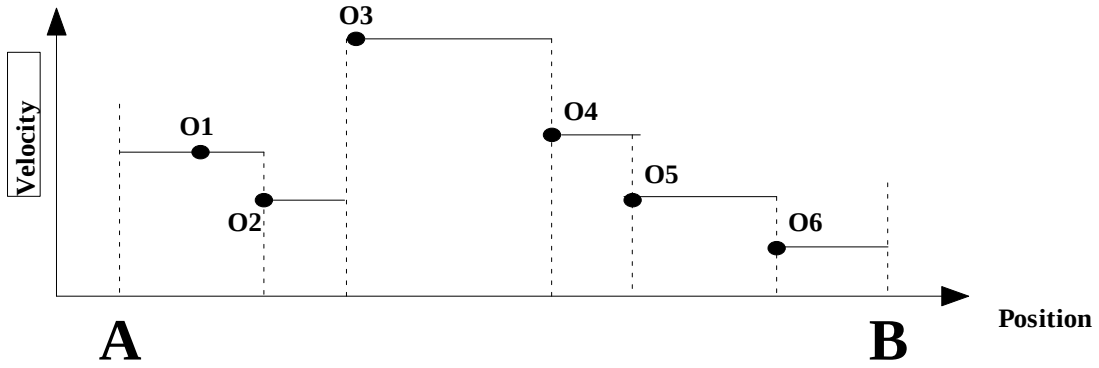


Fig. 3-11: Illustration of the method proposed by Brorsson [21]

In [22], it is reported that the difference in travel time computed with the methods of Brorsson and Westman is not very significant. This conclusion is, however, based on studies of a single link only in the Gothenburg network.

A weakness in the described computation schemes for unidentified probe vehicles is the assumption that the speed is a piecewise constant function of the position. (An immediate alternative would be to use a continuous piecewise linear function instead, this would possibly be an improvement.) A more fundamental argument against the schemes is that no a priori information is exploited in the computation of current link travel time, i.e. no information of the normal traffic conditions is exploited. As a consequence, especially if the number of observations is low, the statistical fluctuations in travel time estimates can be expected to be large and the travel time estimates will be vulnerable for erroneous data. To some extent a priori information is actually introduced into the STRESS system in connection with the concept of normal travel times . However, these normal travel times are essentially based on current travel time data from the immediately preceding time periods, which in practice is limited to the last hour(s) since older data fades away rather quickly. They do not include the very large amount of data that may describe how things usually are.

3.3.4.2 A strategy for developing new methods

The basic idea in this study has been to try to improve on existing link travel time computation schemes by more efficiently exploiting available information about the normal velocity distribution along the link. One might, for instance, establish some standard velocity profile (not necessarily piecewise constant) along the links and somehow compute current velocity profiles by relating or connecting these standard profiles with the observed probe velocities. Travel time for the link could then be estimated from these current profiles.

There is indeed a very large variety of possible implementations of this rather general idea.

An important part of the work in this study has been to try to sort out appropriate candidates for new methods. Our main concern has been to minimize the deviation between estimated travel times and true travel times. But there are also other considerations to make.

One important guide in this process is that the new methods should not be too complicated. If for example, one considers a parameterization of the standard velocity profile it is desirable to keep the number of parameters low. Simplicity often facilitates a theoretical analysis of the method. Stability and robustness are also important requirements. The method should not produce unreasonable travel times.

Another important matter is availability of data. The following example illustrates this point. From *identified* vehicles data it is possible to compute conditional velocity distributions along the link, i.e., given the velocity in one specific point, the velocity distributions in all other points along the link may be determined. Having established such conditional velocity distributions and given an observed velocity in one particular point on the link, it is possible to compute expected (in the sense of expectation values) velocities in all other points along the link, i.e. to compute an expected velocity profile. The conditional velocity distributions could, however, not be determined from unidentified vehicles data since these observations are independent. Possibly, hybrid methods combining data from identified and unidentified vehicles may be considered. No such attempts have been done in this project.

Although we have focused on the spatial distribution of velocities, the temporal aspect is also important. One issue in this context is how to suitably weight together recent observations with older observations.

3.3.4.3 Spatial modeling using a one-parameter family of functions

It is a very natural idea to collect data from unidentified floating cars during a longer time period to establish some sort of standard velocity profile for each link. An immediate difficulty that arises here is that the standard velocity profile may vary a lot under different traffic conditions. Another difficulty is to find a procedure to combine each new observation from an unidentified floating car with the given standard profiles to obtain an estimate of the current travel time for the link.

One way to deal with the first problem would be to partition the different traffic conditions into a number of classes and to create a standard velocity profile for each link and each such class. This approach gives rise to some new problems. Firstly, the classes have to be defined. One possibility would be to use specific time periods during which the traffic intensity is usually within some predefined level. Another possibility may be to let the above time periods dynamically depend on the traffic intensity. This would require that the traffic intensity is somehow graded and measured, either globally (over some region) or for each specific link.

A more elegant way to eliminate (or at least reduce) the problem of time variation of the standard profiles is the following approach. Let us think of the standard velocity profiles (for a specific link) for different traffic conditions as a family of functions. This family of functions may be assumed to depend continuously on one single parameter, θ, which is somehow related to the traffic load (or congestion) on the link. Each particular function in this family can be thought of as being the average of all velocity profiles generated by infinitude of vehicles passing the link under one and the same traffic load. It is therefore reasonable to assume that the family of functions is decreasing with respect to the load parameter, θ. Let us distinguish two particular profiles from this family: one, $g(x)$, corresponding to low traffic conditions (where each vehicle can move unhindered from the others) and another one, $h(x)$, corresponding to peak-hour traffic. (We ignore for the moment that the specification of these two classes is not unique.) Let us further imagine a maximum velocity curve, $M(x)$, and a minimum velocity curve, $m(x)$. The purpose of introducing the two latter ones is that the entire space of possible observations should be spanned by these functions. Let us (as a convention) associate to each of these functions the following values of θ:

Function	θ
M	*0*
g	*1*
h	*2*
m	*3*

By collecting unidentified floating car data during a longer time period at, respectively, low traffic and peak hour conditions, the functions *M*, g, h and *m* can be determined for each link. This is done by segmenting the link into short sections (typically 5 m) and computing the average velocity for the observations within each such section. If from measurements we obtain $h(x) > g(x)$ for some section, then we set $h(x) = g(x)$, to assure that the family of curves is decreasing.

Thus, an arbitrary function within the family is uniquely specified by the parameter θ and the four special functions correspond to integer values of θ. The basic idea now is to associate to each new velocity observation one specific θ and to model the corresponding function by using simple linear interpolation. More precisely, if (x, v) is an observation of position and velocity, the associated parameter θ is determined by the equation:

$$f_\theta(x) \equiv g(x) + \theta \cdot (h(x) - g(x)) = v$$

We will call f_θ the θ-function associated with the observation (x, v). The idea is that it seems reasonable to assume that the velocity profile for the vehicle is similar to the function f_θ for a fixed value of θ, rather than having constant velocity along the link, see Fig. 3.12. The star denotes an observation from an unidentified vehicle. The blue line denotes the θ-function associated with the observation. It has been computed by linear interpolation between the surrounding a priori functions for sparse and dense traffic conditions (dotted green lines, g and h).

From the above it follows that $f_0 = M$, $f_1 = g$, $f_2 = h$ and $f_3 = m$.

The ideas described here may be generalized by introducing additional intermediary functions. This has, however, not been tried in this study.

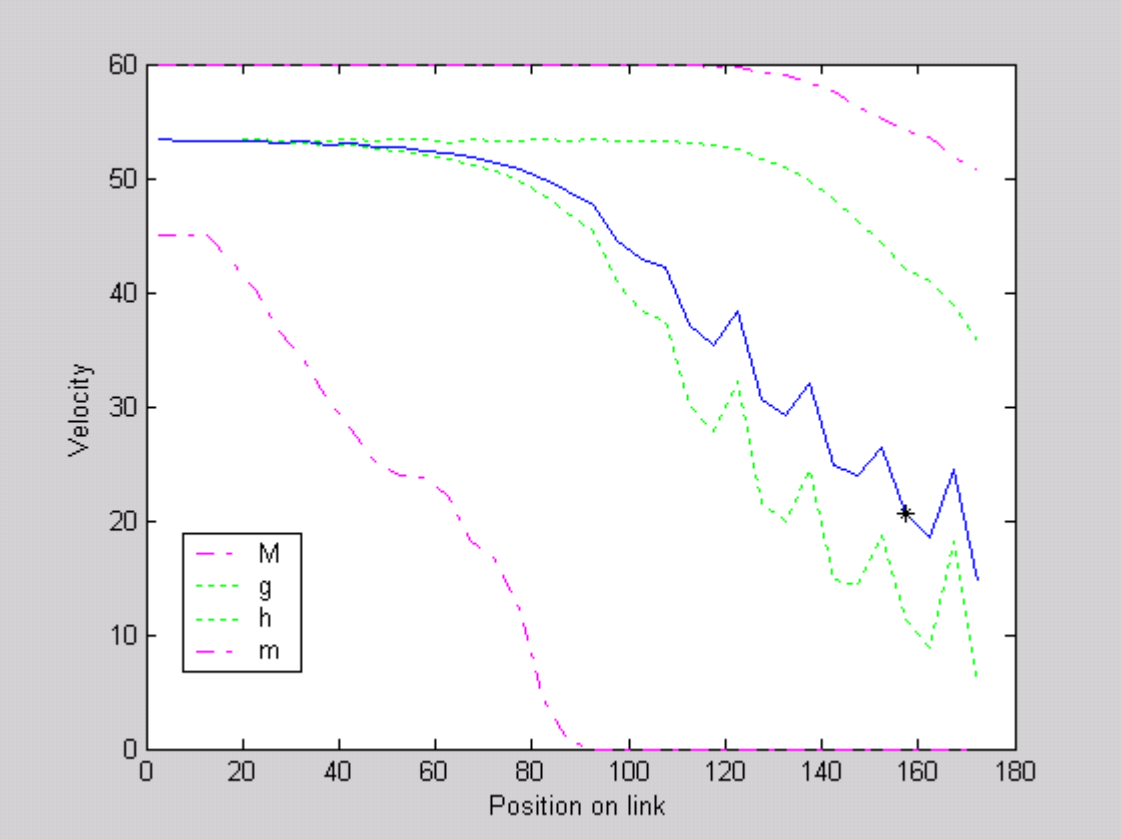
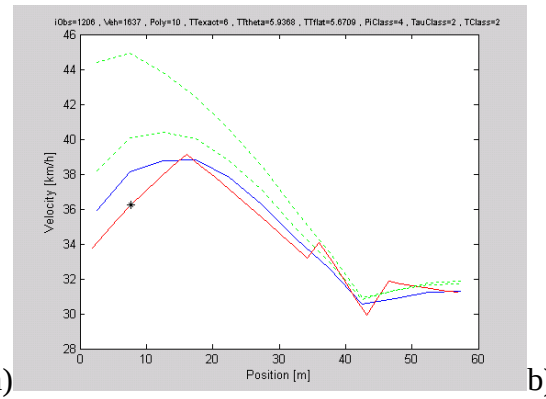


Fig. 3-12: Illustration of θ-functions

How well will θ-functions model true velocity profiles for individual vehicles? This is one of the issues that are investigated in this paper. Other important questions are: how sensitive are the computed travel times for disturbances of various kinds? (E.g., disturbances in a priori functions, *g* and *h*, or disturbances in the positions of the observations.) Our working hypothesis is that θ-functions in general yield better estimates of travel time than constant velocities. In Fig. 3-13 two examples of real (i.e. simulated) velocity profiles and modeled velocity profiles are shown, where θ-functions associated to an unidentified probe observation have been used. The star denotes a single observation from an unidentified probe observation. The dotted (green) lines describe a priori velocity profiles for sparse (upper line) and dense (lower line) traffic conditions. (The maximum and minimum functions are not shown in the figure.) The red line describes the true velocity profile for the probe car. The blue line describes the modeled velocity profile using θ-functions. TTexact, TTtheta and TTflat respectively denote exact travel time, travel time computed from θ-function and travel time computed from constant velocity function. The reason why the observed velocity does not cut the θ-function is explained in section 3.3.4.5.

a) 

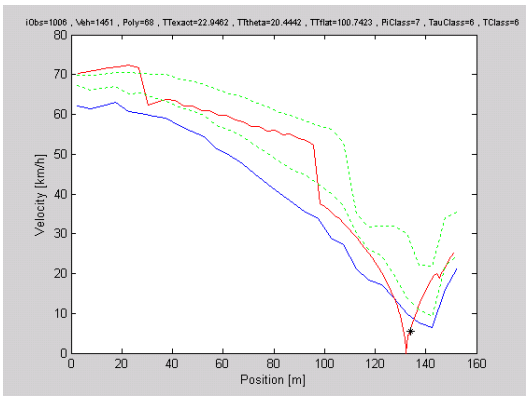
b) 

Fig. 3-13: Two different examples of velocity profiles along links

In both Fig. 3-13 a) and b) the θ-function approximates quite well the exact velocity profile. In a) the observed velocity is such that the constant velocity method also yields a good approximation for the travel time along the link. In figure b) the θ-function method yields a much better estimation for the travel time than the constant velocity method.

3.3.4.4 The asymmetry problem

There is an interesting asymmetry involved when using a one-parameter family of functions. This is illustrated in Fig. 3-14. In a), the observed velocity occurs in a position where there is a big gap between the two a priori curves. There is a good chance then that the θ-function method yields a good result, since the exact velocity profile is determined by the observation in the gap part of the profile and, with high probability, is bounded by the a priori functions in the narrow part of the profile. In b), on the other hand, it is difficult from the observed value to predict with confidence the drop in velocity during the last part of the link. Both the θ-function method and constant velocity method underestimate the true travel time. In c), the θ-function method overestimates the true travel time.

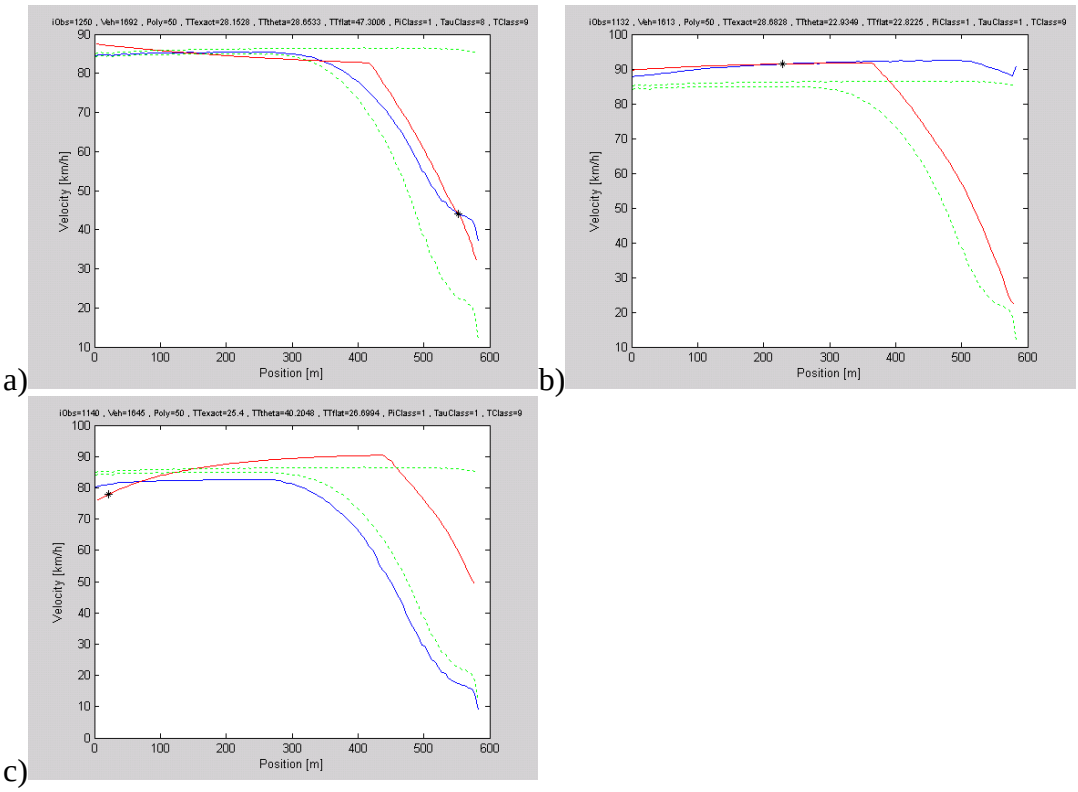


Fig. 3-14: Illustration of the asymmetry problem

In order to describe the asymmetry situation it is useful to introduce the function

$$\tau(x) = \frac{g(x) - h(x)}{g(x)} \qquad (3.9)$$

where *x* denotes the position of an observation along a link.

Observations for which $\tau \approx 0$ contain less information than those for which $\tau \approx 1$. This means that more accurate travel time estimates for larger values of τ could be expected than for lower values.

3.3.4.5 The low velocity problem

It is not unusual, especially in congested traffic situations, that very low velocities are observed - even zero velocity. If not handled in some special way, these low values will cause absurdly high over-estimates for the travel times, in particular when using constant velocity profiles. For the θ-function method a radical improvement can be achieved by modifying the computations for observations having low velocities, i.e., within the region $v < h(x)$. The parameter θ is then replaced by

$$\theta(\omega) = 2 + \frac{v - h(x)}{m(x) - h(x)} \times \omega\,, \qquad (3.10)$$

where *x* and *v* are the observed position and velocity. Thus, $\omega = 1$ corresponds to the unmodified case, while $\omega = 0$ means that f_θ will coincide with *h*.

A suitable value for the relaxation parameter has empirically been found to be $\omega = 0.5$ for peak hour traffic conditions.

For the constant velocity method a fixed lower limit for the velocity has been set. The value 3 km/h has been found an appropriate compromise (and is used throughout this report unless otherwise stated) although different aggregation methods have somewhat different optimal value.

In Fig. 3-15 a few examples are shown where velocity profiles have low velocities.

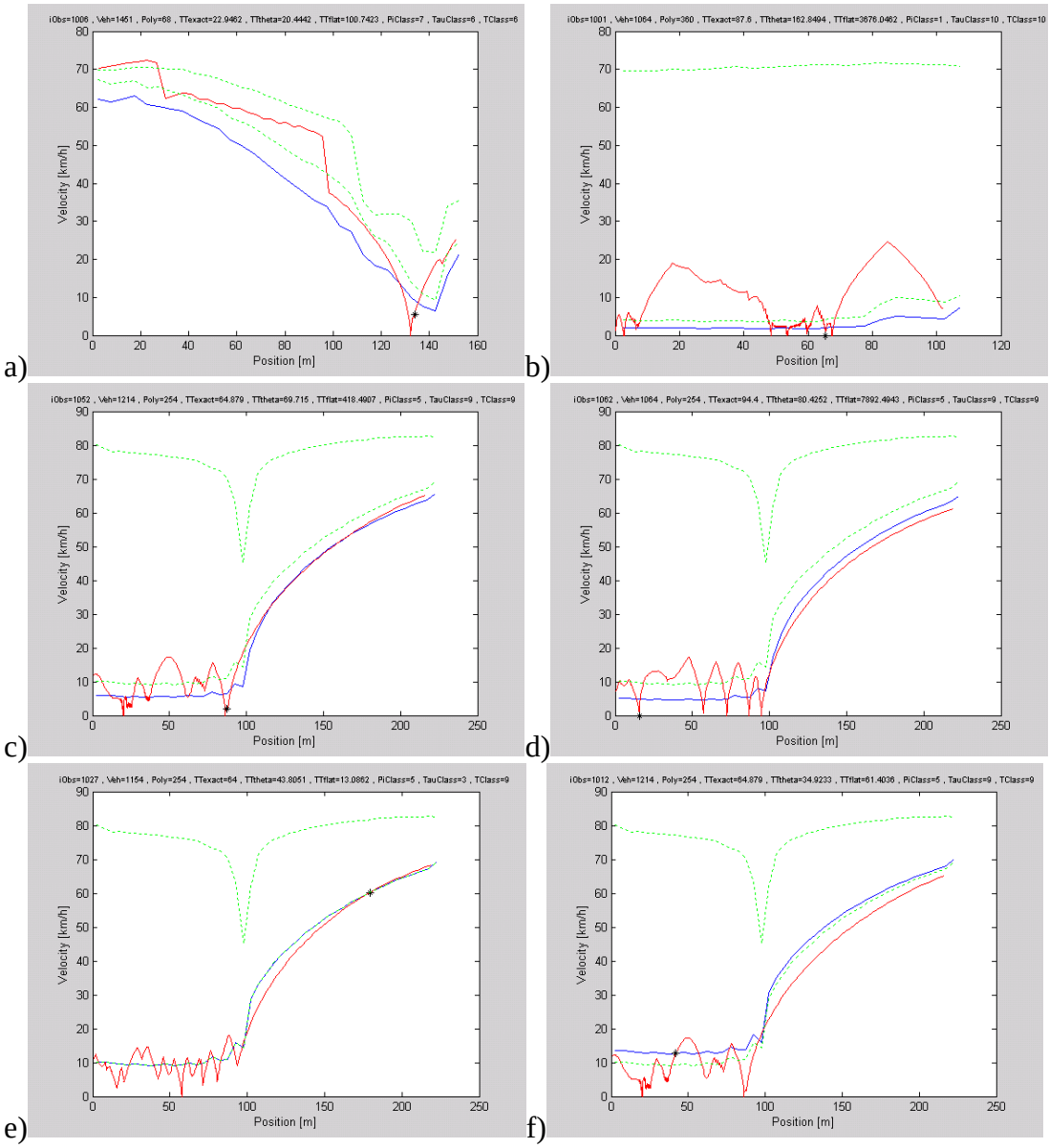


Fig. 3-15: θ-function method for low velocities. The blue line denotes the θ-function and the red line denotes the true velocity profile

In a) the θ-function method yields a good result although the observed velocity is small. Notice that the θ-function does not cross the observed value since θ has been adjusted. In b), the observed velocity is (almost) zero and both the constant velocity method and the θ-function method would have crudely overestimated the true travel time if no special adjustment for low velocities had been done. Similarly for c) and d). In e) and f), the θ-function method underestimates the true travel time although the velocity profiles seem to agree reasonably well.

3.3.4.6 Spatial combination of observations during a fixed time period

There is a large variety of possibilities to combine several probe observations or their corresponding velocity profiles. Let us assume that we have one specific time slice (e.g. a five minute period) in

which all probe observations for a specific link are considered having equal time. Also, we assume that some sort of velocity function over the link has been assigned to each measurement, e.g. constant velocity function, a θ-function or some other type of function. Only the spatial combination is thus considered here. Below is a list of examples describing how this can be done:

1. Predefined segmentation of the link in a fixed number of sections (for instance equidistant intervals).
2. Segmentation of the link using knowledge of normal velocity distributions along the link. For instance, the function τ defined by Eq.(3.1) can be used for finding a suitable partitioning of the link.
3. Segmentation using the observations as node points (the approach used by Brorsson [21]).
4. Compute travel time for each velocity profile, i.e., each observation. Then compute the average (arithmetic, harmonic or maybe a weighted average) of these travel times.
5. Compute a mean velocity profile by averaging over the velocity profiles for the respective observations. Then compute travel time for the mean velocity profile.

The first example has the disadvantage that the observations do not in general cover all sections of the link. Thus, results will have to be combined with older observations, i.e. from other time slices. A somewhat similar situation arises in the second example. The third example is the one used in the Stockholm STRESS system (for constant velocity components). The travel time computed in example 5 is never larger than that of example 4.

It turns out that averaging, especially the harmonic average, can, in some cases, improve travel time estimates remarkably.

3.3.4.7 Temporal combination of observations

In the previous section we discussed a fixed time period during which at least one observation was registered for a specific link. Sometimes, or even frequently if the equipment ratio is low, no observation at all occur along a link during a given time period. This has to be dealt with by using observations from previous time periods. More generally, there is a conflict between the desire to make the algorithm react quickly on changes in traffic state caused by changes in OD matrices or sudden changes in capacity due to incidents, and, on the other hand, to collect enough data to make reliable travel time estimations. In this section two different approaches to deal with this problem are discussed.

The first method consists in weighting together travel time estimates for the current time slice and the previous time slice. It is appropriate to adopt the terminology used in the STRESS project. We use current travel time for the most recent time slice where only spatial combinations of observations have been made, and the term normal travel time for a combination of current and older data. Let tt_i and TT_i respectively denote the current travel time and the normal travel time for time slice *i*. We then have:

$$TT_i = \alpha \times TT_{i-1} + \beta \times tt_i$$

where suitable values for the coefficients α and β are determined empirically. The length of a time slice will be fixed, typically five to fifteen minutes. The computation of current travel time for a time slice can be done by any method described in section 3.3.4.6.

The second method makes use of a fixed number of recent observations (rather than a fixed time period) which are combined spatially by any of the methods described in section 3.3.4.6. A suitable value for the number of observations used can be determined empirically. A possible advantage with this approach is that estimations can be done for arbitrary time points and not only for predefined time slices.

More advanced methods for temporal combinations will not be investigated in this project since this topic is more suitably studied in connection with travel time prediction.

3.3.4.8 Estimating the variance

It is important to obtain some measure of uncertainty in the travel time estimates, preferably the variance. This is needed when fusing data (see chapter 4) with data from other sources, e.g, travel time estimates from identified floating cars (see section 3.1). The spread of the observations during a time period provides us with a possibility to estimate the variance when the arithmetic average of travel times is used. Misleading results might occur when the distribution of the travel times is very skew, such as is the case when the low velocity problem is not handled properly.
When instead using the harmonic average the variance could be estimated as follows.

Let t_H denote the harmonic average of $t_1, t_2, \ldots, t_n$ and let $v = \sum_{i=1}^{n} \frac{1}{t_i}$. Let Δt_H and Δv denote the corresponding variances. The variance in t_H may be estimated by the simple heuristic:

$$\Delta t_H = -\frac{1}{v^2} \Delta v .$$

This estimate may be somewhat misleading since the variance in the harmonic average does not reveal anything about the (maybe systematic) difference between the harmonic and arithmetic averages. Thus, there is a real danger that the harmonic average yields a significant underestimation of the true travel time, although the above variance estimate might be very low.

For the methods based on joining segments together no known estimation method for the variance is available.

3.3.4.9 Some other practical problems

Non-unique link lengths

The very notion of travel time for a link requires that links (or at least their lengths) are specified uniquely. For real data, this requirement boils down to the map-matching problem, where observed coordinates have to be projected onto link positions in a stylistic network. For simulated data that have been used in this project, the position of a vehicle along a section within a polysection (=link) is obtained from AIMSUN (version 4.2). However, since each polysection is made up of one or several lanes, the length of a specific link will become dependent on the lane. This is particularly the case when a link is connected to another at a skew angle (such as for on- and off-ramps to highways) where the differences in link length may vary significantly between the lanes. As a consequence, uncertainties in the travel time estimates arise.

In order to make estimated travel times consistent with true travel time, we have chosen to compute the distance travelled by each individual probe vehicle on each link. The travel time estimates are based on these distances, rather than some fixed lengths of the links. In this way, the relative error in travel time estimates is correctly computed. On the other hand, some inconsistencies with a priori data may occur.
The above problems have their origins in the fact that no map-matching was done from the AIMSUN network onto a stylistic network.

No distinction of lanes

From simulated data we have the possibility to separate between different lanes. This possibility has not been exploited since real position data presently do not have enough precision to discern the separate lanes with certainty. Future improvement of precision may actualize this issue.

How to define travel time for a link during a time period?

Using simulated data has the advantage to provide us with the possibility to directly compare the estimated travel times with exact travel times for each link. As a consequence there is a need for a strict definition of what is meant by exact travel time on a link. For a single observation this is less of a

problem, but how to define the exact travel time (for a link) during a specific time slice of, say, five minutes?

The arithmetic mean of the travel times of all vehicles passing the link during the time slice would may be seem to be the obvious choice. However, the distribution of travel times for single vehicle is sometimes (especially for dense traffic) very asymmetric. The arithmetic mean of travel times is therefore not always very typical. A good alternative candidate is the harmonic average[1] of travel times which corresponds to the arithmetic average of mean velocities over the link. The distinction between arithmetic and harmonic means becomes more accentuated for the *estimated travel time* values from probe observations but it is also desirable that the exact definitions are consistent with the estimating procedures. One could therefore possibly distinguish between the concept *arithmetic travel time* and *harmonic travel time*.

The partitioning of time into time slices may lead to inconsistencies between the estimated travel times and exact travel times since the major part of the duration of a probe vehicle on a specific link may fall outside the current time slice although the observed event takes places within the current time slice (see Fig. 3-16). This is often the case for very short time slices (or long links). To avoid this kind of inconsistencies each true travel time result is weighted with the share of the path included in, or belonging to, the current time slice. More specifically, the weight coefficient for the left path in the figure is $\frac{t_3 - t_2}{t_4 - t_1}$.

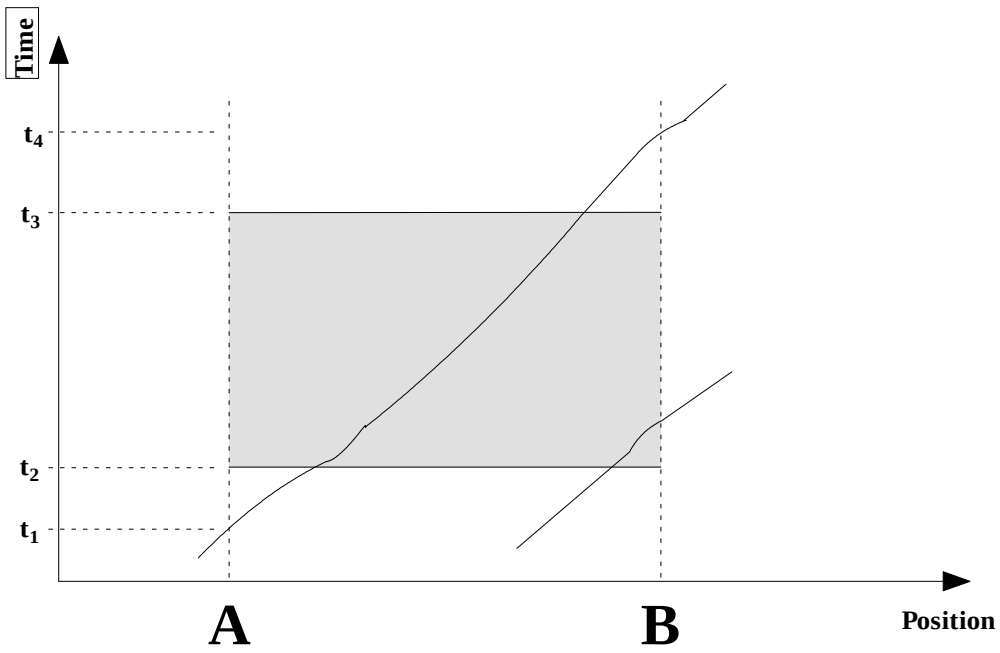


Fig. 3-16: The definition of true travel time during the time slice $[t_2,t_3]$ is consistent with the probability that an observation (in a position on link AB) occurs during this time slice

A further possibility would be to let the definition of true travel time be based on the probe vehicles only, rather than all the simulated vehicles. The vehicles, that are not probes, do not provide any information at all to the system and could therefore be considered as unobservable. The probe vehicles on the other hand do provide information and could therefore be a suitable reference. However, since the selection of probe vehicles is small, and hence subjected to larger uncertainties, we have chosen to use all vehicles when estimating true travel times.

[1] Harmonic mean is often used in connection with speeds representing space mean speed, while here, the harmonic mean is applied to the travel times.

3.3.5 Results for simulated data from Gothenburg network

In this section travel time estimates from (different formulations of) algorithms are compared with true travel times for simulated vehicles. To gain a better understanding of the various phenomena that occur, results for different levels of complexity are presented. We first focus on single observations and investigate how large is the potential to improve travel time estimates by using spatial modeling instead of constant velocities. Next, different methods for combining single observations during a specific time slice are compared. Then we present results using different methods for time recursions. Finally, a sensitivity analysis is done.

All results presented in this section are based on simulated data from the Gothenburg network. In later chapters results for simulated data from the Lausanne network (chapter 3.6) and real data from the Vienna network (chapter 3.3.6) are presented.

3.3.5.1 Comparisons between flat velocity profiles and θ-functions

What effect does it have to model the velocity profile for each individual observation using a function of constant θ instead of constant velocity? There are two reasons for studying this issue in more detail.

Firstly, it can give us some insight into what gain one can possibly hope while using spatial modeling. Secondly, single observations for a link during a specific time period constitute an important special case that often arises during sparse traffic conditions.

Cumulative diagrams over the relative error in computed travel times can give a good overview over the precision of different methods. In Fig. 3-17 comparisons between estimates using θ-functions and constant velocities are shown for different traffic loads.

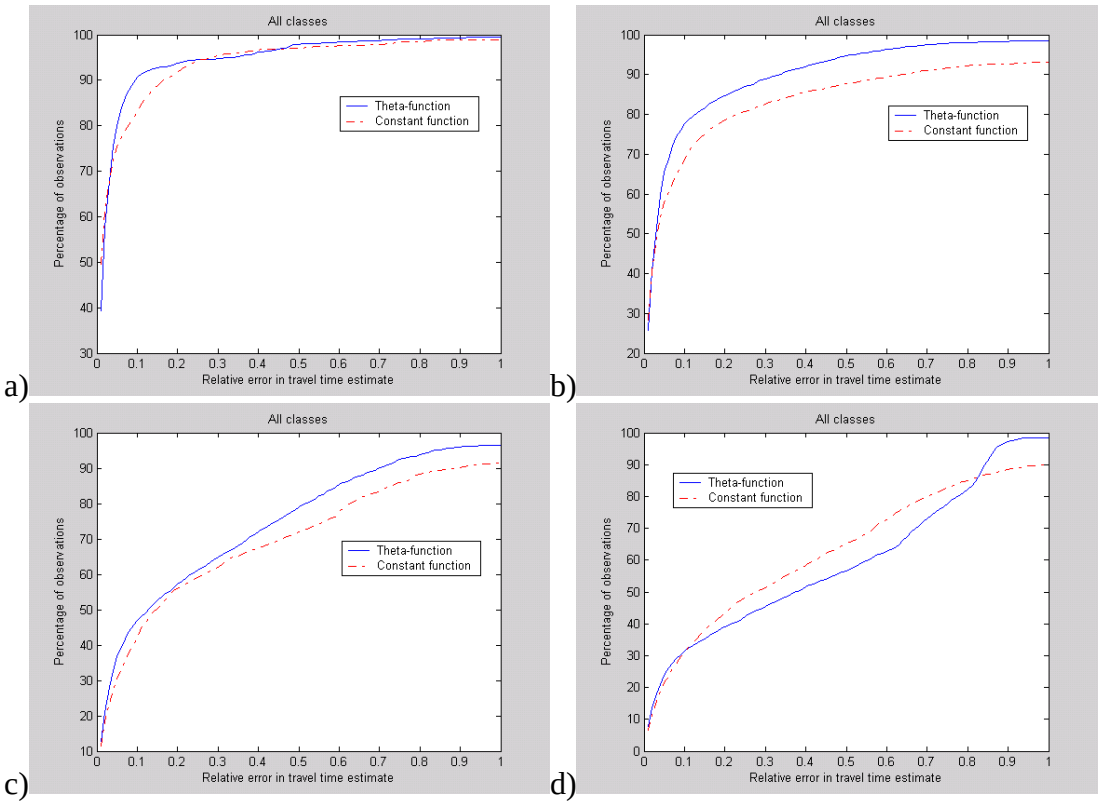


Fig. 3-17: Cumulative diagrams for the relative error in travel time estimates based on single observations

The dashed-dotted line denotes estimates using constant velocities while the solid line denotes results based on θ-functions. In a), sparse (uncongested) traffic conditions is used. In b) the OD-matrix has been set to 80% of peak-hour conditions, in c) peak-hour conditions and in d) 120% of peak-hour conditions. For sparse traffic conditions, (a), the θ-function method seems to give slightly better results than the constant function method. Since practically no congestion occurs in this case the difference is probably caused by acceleration and retardation segments near junctions and cross-roads. When the congestion is increased to 80% of peak-hour conditions, (b), the θ-function method is clearly better than the constant function method. Notice in particular the difficult cases : for the constant function method approximately 10% of the observations yield a relative error larger than 70%, while the corresponding number for the θ-function method is 2-3%. The θ-function method still seems better than the constant function method when the load is increased to peak hour conditions, (c), but the superiority decreases with the congestion. In very congested conditions, (d), the constant function method yields in general lower relative error. The results for this latter case are heavily dependent on how the particular implementation handles the low velocity problem.

In Table 3-2 some other measures to compare the θ-function method and constant function methods are displayed. Computing the sum of all estimated link travel times from all available unidentified observations and dividing these sums with the number of observations provides us with one such measure (Average travel time in Table 3-2). Average absolute errors and relative errors are other measures.

Table 3-2: Comparisons between travel time estimates using the θ- function method and the constant function method for different traffic loads. The percentages denote percentage of peak hour traffic load

	Sparse	80%	100%	120%
Number of observations	1832	3942	7868	11757
Average true travel time	24.97	28.39	132.80	158.10
Average travel time with theta method	24.53	28.85	130.66	91.84
Average travel time with constant functions	25.43	36.21	129.60	154.53
Average absolute error with theta method	1.69	3.69	53.38	85.92
Average absolute error with const functions	1.86	10.78	51.40	77.79
Average relative error with theta method	0.06	0.11	0.24	0.38
Average relative error with const functions	0.08	0.29	0.41	0.48

One may first notice that the (exact) travel time increases dramatically when the load is increased from 80% to 100% of peak hour traffic. This increase does *not* represent the increase in total travel time within the network (i.e., the sum of link travel times in the network)! The reason for this is that travel time estimates in Table 3-2 are computed from probe observations with constant reporting intervals. Since travel times are longer on congested links, these will be over represented among the probe observations and, thus, the total travel time within the network will be overestimated. Nonetheless, it is interesting to compare total (or average) travel times for different methods. For the very congested case (120%) the θ-function method yields a very low value compared to the true travel time. One credible explanation for this is the phenomenon illustrated in Fig. 3-17 e)-f). Although the velocity profile may be modeled well by the θ-function, the stops will not be modeled and the travel time is underestimated. These features are, however, very sensitive to the actual implementation of the methods.

The result can be expected to depend on how much the velocity usually varies along the specific link. To investigate the effect of the asymmetry problem we define $\mathrm{T} = \max_x \tau(x)$, i.e., T is the largest relative deviation between the standard low and high traffic velocity profiles. In Table 3-3 absolute errors per observation for different classes of T and $\tau(x)$ are shown. For a specific (class of) $\tau(x)$ the error tend to increase with increasing T. This is consistent with the asymmetry properties. For a specific (class of) T the error sometimes tend to increase with increasing $\tau(x)$, which is more difficult to explain. A possible explanation might be that low velocities influence results when $\tau(x)$ is of the same order as T.

Table 3-3: Absolute error per observation for travel time estimates for different classes of T and $\tau(x)$. Travel time estimates have been computed using θ-functions

	0 <= T <= 0.1	< T <= 0.2	< T <= 0.3	< T <= 0.4	< T <= 0.5	< T <= 0.6	< T <= 0.7	< T <= 0.8	< T <= 0.9	< T <= 1
0 <= tau <= 0.1	0.38	1.21	1.38	1.03	0.36	14.52	1.59	19.60	8.08	19.39
0.1 < tau <= 0.2	0	1.58	1.52	0.92	0.00	3.13	3.08	18.38	9.40	1.21
0.2 < tau <= 0.3	0	0	2.86	2.45	0.00	2.56	3.77	0.52	8.82	4.08
0.3 < tau <= 0.4	0	0	0	1.01	0.00	1.71	4.39	0.00	9.27	0.00
0.4 < tau <= 0.5	0	0	0	0	0.00	2.23	6.90	18.67	6.41	56.32
0.5 < tau <= 0.6	0	0	0	0	0	7.11	4.90	21.63	10.06	0.00
0.6 < tau <= 0.7	0	0	0	0	0	0	8.69	21.11	8.86	35.99
0.7 < tau <= 0.8	0	0	0	0	0	0	0	18.13	11.14	35.50
0.8 < tau <= 0.9	0	0	0	0	0	0	0	0	17.16	98.75
0.9 < tau <= 1	0	0	0	0	0	0	0	0	0	164.34

In Fig. 3-18 comparisons between estimates using θ-functions and constant velocities are shown. In a) all observations are shown. The dashed-dotted line denotes the constant velocity method and the solid line the θ-function method. In a) all links are considered. In b), c) and d) specific classes for T

and $\tau(x)$ are displayed. The θ-function method seems to be only slightly better than the constant function method. However, the really difficult cases, where the relative error exceeds 0.5, a significantly smaller part is represented by θ-function method than by constant velocity method.

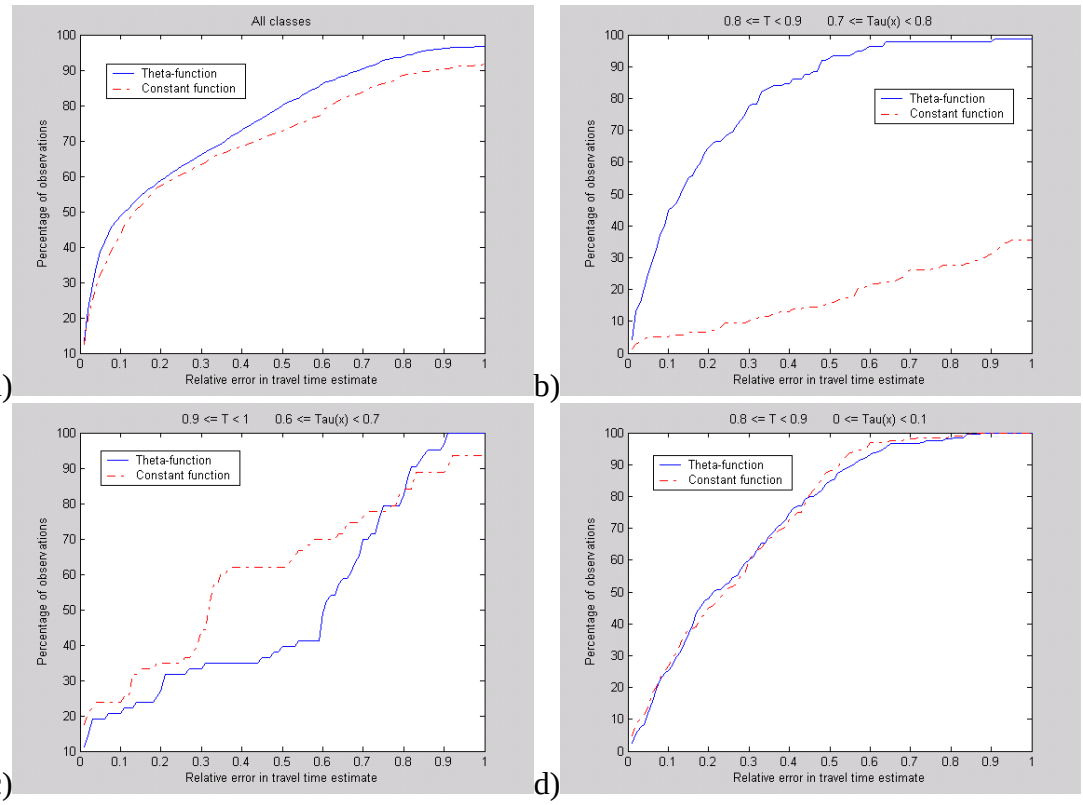


Fig. 3-18: Cumulative diagrams for the relative error in travel time for the observations in the network

In Fig. 3-18 b) an example is shown where the constant velocity is not very successful. This is in accordance with the asymmetry properties. In c) an example is shown where the θ-function method is less successful.

3.3.5.2 Comparisons between different strategies for spatial combinations

In this section results concerning aggregations of observations for a fixed time period (usually five minutes) is presented. Each such period (or aggregate) and link is referred to as one *link-period*, and each link-period corresponds to one observation. Since in the Gothenburg network there are 92 links (=polysections) and the simulation has been done for 90 minutes a maximum of 1656 observations are possible when the time period length is 5 minutes. Excluding warming up and cooling down periods as well as time periods without reports from probe vehicles leaves us with 897 observations on which the results in this section are based.

Unless otherwise mentioned the traffic load studied in this section is that of peak hour traffic, the equipment ratio is 5%, the report interval is 10 seconds and the aggregation period is 5 minutes.

The following spatial combination methods have been studied:

1. Arithmetic average of travel times using constant velocities (Average flat)
2. Arithmetic average of travel times using θ-functions (Average theta)
3. Harmonic average of travel times using constant velocities (Harmonic flat)

4. Harmonic average of travel times using θ-functions (Harmonic theta)
5. Segmentation of velocity profile by joining constant velocities, i.e., Brorssons method (Join flat)
6. Segmentation of velocity profile by joining θ-functions (Join theta)

In Fig. 3-19 cumulative diagrams of the relative error for each method are presented.

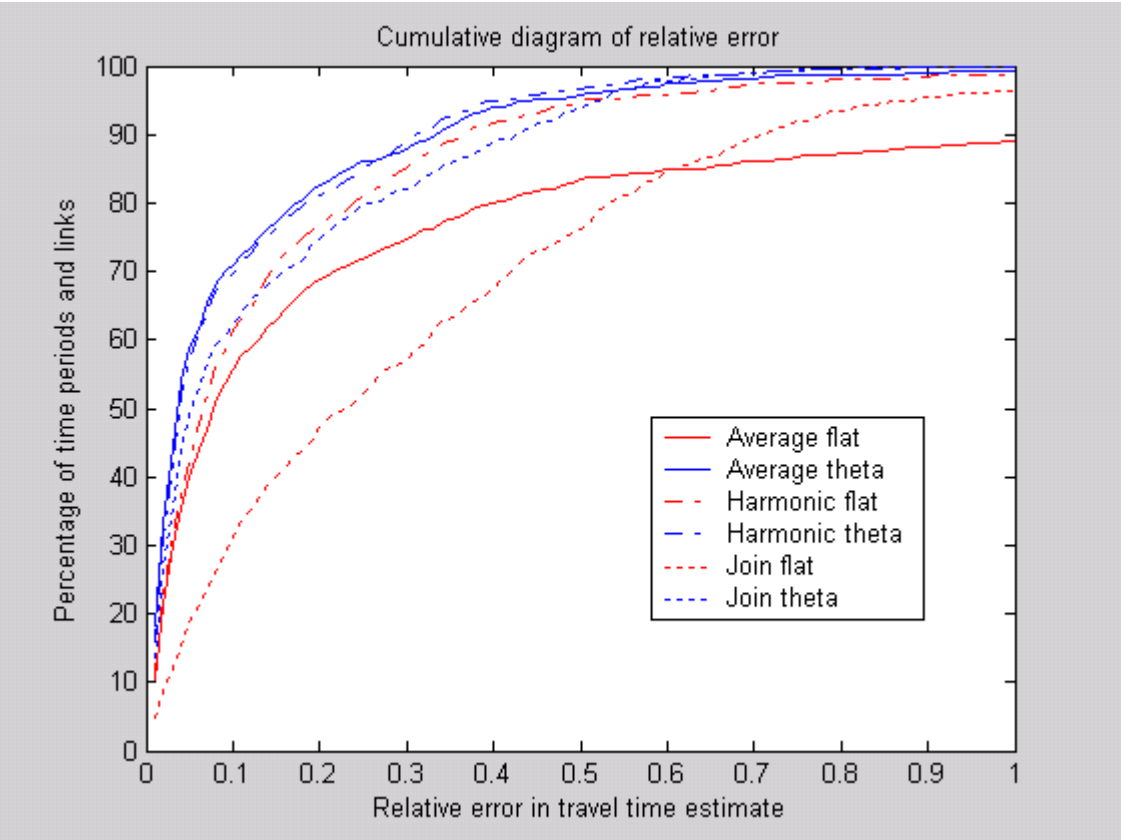


Fig. 3-19: Cumulative diagram for the relative error for various spatial combination methods.

The percentage of cases (link-periods) for which the relative error does not exceed a specific value (The equipment ratio is 5% and the report interval is fixed to 10 seconds)

It is seen that, in general, methods based on θ-functions have lower relative errors than those based on constant velocities. This is consistent with the results in the previous section.

Also, it seems preferable to use harmonic average of travel times rather than the arithmetic average at least when computations are based on constant velocity profiles. The reason for this gain in accuracy is the very asymmetrical distribution of travel times that arises during low speed traffic conditions. Using Average flat in these situations often leads to huge overestimations of travel times. In the case of Harmonic theta , the asymmetry in the distribution of travel times seems to have been avoided by preventing too low velocities by using ω =0.5 in Fig. 3-10.

From Fig. 3-19 it is also seen that using segmented (joined) velocity profiles seems to be less accurate than taking averages of travel times (at least when the relative error is not too high). For instance, Brorssons method is the least accurate method for a large number of cases. It seems that reducing errors in travel times by averaging operations is more efficient than by joining operations.

Is there any skewness or obliqueness in the distribution of the estimated travel times? In Fig. 3-20 the estimated travel times for each link-period are plotted against true travel times.

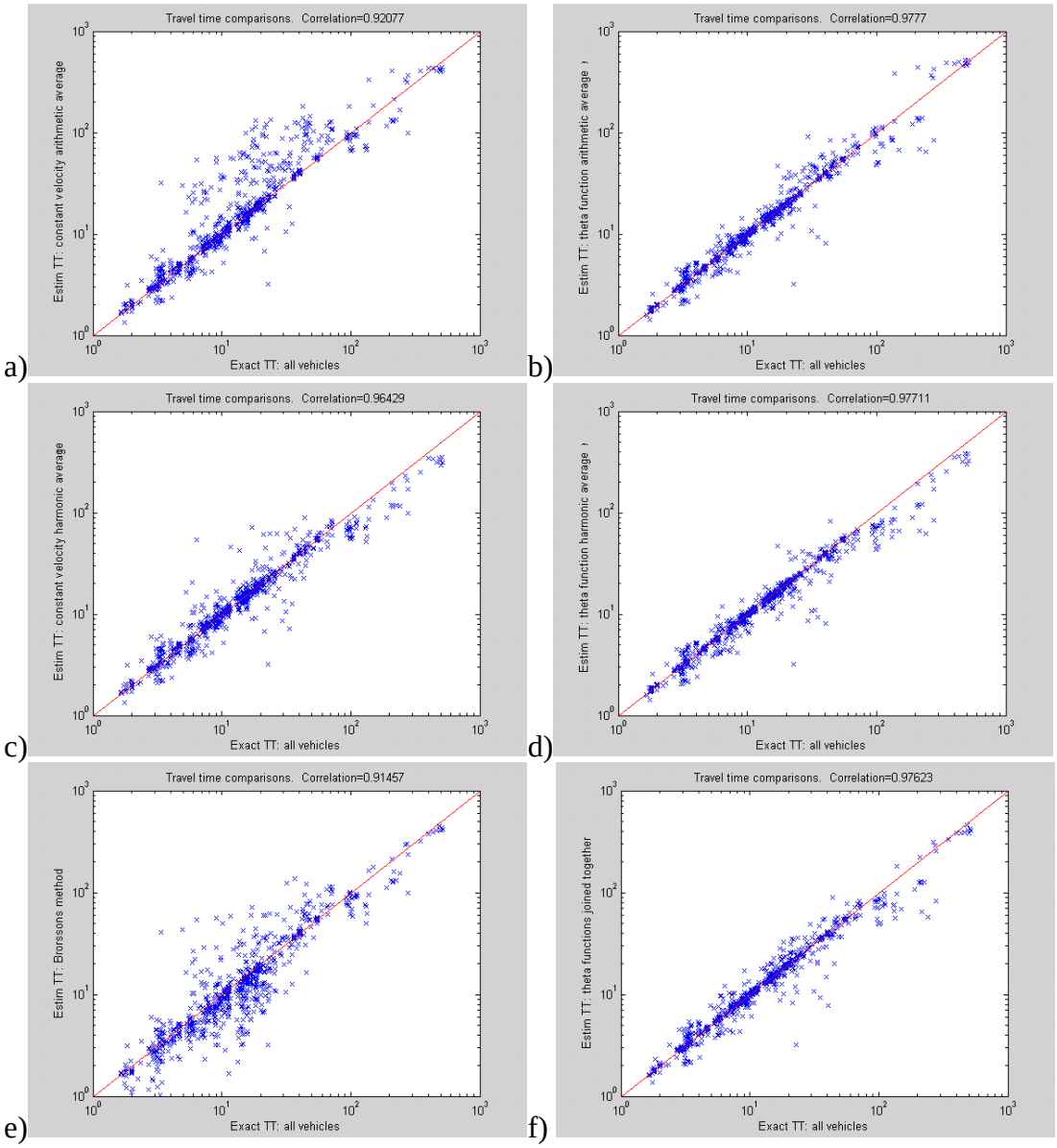


Fig. 3-20: Estimated travel time versus exact travel time for different estimation methods.

Notice that the correlation corefficient is computed from the lograthmic quantities.

From a) it is seen that Average flat tend to overestimate the travel times while Average theta seems more symmetric. The harmonic averages tend to decrease the estimates. This tendency is particularly strong for large travel times (i.e., for low velocities).

Are these conclusions still true for other parameter values? In Table 3-4 the relative error per observation for different equipment ratios is presented. For most of the methods the (average) relative error increases with decreasing equipment ratios. An exception is Average flat for which the relative error seems virtually independent of the rate. One might have expected a more radical increase in relative error. It should, however, be remembered that only such links and time periods for which there is any observations are considered. Since reporting intervals of 10 seconds have been used, there is, at least for longer links, a high probability that a probe vehicle reports more than one (and even several) observations. The forces of averaging are therefore still active.

Table 3-4 Relative error per observation for various equipment ratios and for different estimation methods

	Equipment ratios				
Method	0.1	0.05	0.03	0.01	0.005
Average flat	0.395	0.425	0.425	0.396	0.435
Average theta	0.103	0.112	0.124	0.126	0.180
Harmonic flat	0.142	0.169	0.204	0.217	0.233
Harmonic theta	0.099	0.110	0.123	0.138	0.152
Join flat	0.305	0.342	0.360	0.334	0.395
Join theta	0.095	0.110	0.123	0.120	0.163

In Table 3-5 the effect of varying reporting intervals is studied, i.e., the time interval between subsequent reports (observations) produced by the probe vehicles. Notice the increase in relative error for the Harmonic flat case. It seems that the radical decrease in error that the harmonic average operation causes in the constant velocity case is being lost when the reporting intervals increases. This is natural since two observations are seldom produced from the same vehicle on the same link if the reporting interval is long. Thus, for longer reporting intervals there is a higher probability for single observations and differences for different spatial combination methods should be small. The differences between methods based on θ-functions and on constant velocities is considerable and consistent with previous results.

Table 3-5: Relative error per observation for various reporting intervals and for different estimation methods

	Reporting intervals (seconds)			
Method	10	30	60	120
Average flat	0.406	0.532	0.591	0.557
Average theta	0.115	0.149	0.184	0.172
Harmonic flat	0.158	0.309	0.415	0.489
Harmonic theta	0.107	0.139	0.186	0.181
Join flat	0.350	0.481	0.492	0.546
Join theta	0.139	0.155	0.177	0.164

In Table 3-6 the relative error for various lengths of aggregation periods is shown. The period length is the time interval during which the observations are collected and the averages are computed. For short period lengths the number of observations is small and the statistical uncertainty increases. Notice in particular the Harmonic flat case where the increase is more rapid than for the other methods. Notice also that for shorter periods the number of cases grows.

Table 3-6: Relative error per observation for various period lengths and for different estimation methods

	Period lengths (minutes)							
Method	30	15	10	7	5	3	2	1
Average flat	0.466	0.414	0.423	0.395	0.406	0.390	0.403	0.408
Average theta	0.096	0.097	0.110	0.110	0.115	0.123	0.134	0.148
Harmonic flat	0.116	0.117	0.134	0.139	0.158	0.169	0.196	0.219
Harmonic theta	0.089	0.090	0.099	0.104	0.107	0.114	0.121	0.137
Join flat	0.287	0.291	0.343	0.333	0.350	0.341	0.368	0.400
Join theta	0.116	0.120	0.131	0.135	0.139	0.145	0.148	0.161

3.3.5.3 Comparisons between different strategies for temporal combinations

In order to achieve non-trivial dynamical effects, the simulation has been performed with a varying OD matrix. The weight factor 1 corresponds to peak hour load.

Time period	Weight on OD-matrix
0 10 min	*1*
10 20 min	*0.5*
20 30 min	*1.2*
30 40 min	*0.5*
40 50 min	*1.5*
50 60 min	*0.5*

Notice that, in this section all results are based on every time period within 0-60 minutes, except possibly the first ones where data might be lacking.

The two approaches for travel time estimates described in section 3.3.4.7 have been investigated. We first present results for the Harmonic theta method using a fixed number of (most recent) observations. In Table 3-7 the relative error is shown for different equipment ratios and number of observations. In general, the relative error increases with decreasing equipment ratio. The reason for this is probably that with lower rates the observations become scarcer and the n^{th} last observation older and less accurate due to dynamical effects. Likewise, the relative error decreases with increasing number of observations up to an optimal value and then increases again. The more the observations the higher the accuracy until the observations becomes obsolete. The optimal number of observations seems to decrease with decreasing equipment ratio, which is very natural. Certainly, the rate of change of the traffic state must also affect this optimum.

Judging from the table n=4 or 5 seems not an unreasonable choice in the present case.

Table 3-7 The average relative error for travel time estimates using harmonic average with a fixed number of observations

	Equipment ratio				
Number of observations	0.1	0.05	0.03	0.01	0.005
1	0.211	0.224	0.225	0.180	0.253
2	0.178	0.191	0.195	0.157	0.246
3	0.167	0.177	0.188	0.172	0.247
4	0.153	0.165	0.185	0.167	0.248
5	0.142	0.159	0.188	0.171	0.261
6	0.139	0.159	0.182	0.180	0.297
7	0.137	0.162	0.177	0.176	0.310
8	0.137	0.158	0.169	0.172	0.314
9	0.138	0.158	0.171	0.178	0.302
10	0.141	0.160	0.177	0.177	0.325

Let us consider the other method based on combining the current period with previous ones.

In Table 3-8 the relative error is shown for different equipment ratio and different values for the coefficient of conservatism , α . It seems that low values for α are to be preferred in this case (in particular when the equipment ratio is low) which means that the main emphasis is on the current time period. The optimum value for α depends probably on the period length, the rate of change of traffic state, the travel time estimation method for the time period and the variance for the observations within it as well as on the ordinary parameters equipment ratio and reporting interval.

Notice also the sudden drop in relative error for $\alpha = 1$. The explanation for this is that the method then assigns the previous time period to the present one.

Table 3-8 The average relative error for travel time estimates using a combination of periods and various values for α. Time period lengths of 5 minutes have been used

	Equipment ratio				
α	0.1	0.05	0.03	0.01	0.005
0.1	0.140	0.161	0.165	0.144	0.215
0.2	0.138	0.162	0.166	0.154	0.236
0.3	0.139	0.166	0.168	0.170	0.265
0.4	0.141	0.170	0.174	0.195	0.302
0.5	0.146	0.177	0.182	0.230	0.349
0.6	0.153	0.187	0.193	0.279	0.408
0.7	0.165	0.202	0.209	0.346	0.487
0.8	0.185	0.229	0.235	0.436	0.590
0.9	0.221	0.278	0.276	0.559	0.731
1	0.168	0.169	0.175	0.176	0.344

It is not easy to draw from the above results any final general conclusions for which approach is the best. The level of the average relative error is slightly less for the combined method. The greater flexibility of the fixed number of observation -method is appealing.

Another issue concerning dynamic effects is that the harmonic average of θ-functions tends to underestimate peaks in travel times when the congestion is large. The harmonic average of constant velocities seems less prone to these underestimations, although they yield larger errors.

3.3.5.4 Sensitivity analysis

How sensitive are the computed travel times for measurement errors in vehicle positions?

Real observed position data from unidentified vehicles are impaired by measurement errors. For example, for the Vienna data the uncertainty in measured positions are estimated to be in the order of 3-15 meters. It is interesting to study how sensitive the methods are for such disturbances. We can immediately conclude that methods based on flat velocities along the whole link are not affected as long as the disturbed positions belong to the correct link.

Let us impose a uniformly distributed error (bounded by δ) in each observed position, with the restriction that the disturbed and undisturbed position belongs to the same link. In Table 3-9 the sensitivity in estimated travel times for disturbances in positions is shown.

Table 3-9 Average relative errors in travel time for various levels of disturbances in positions for the observations

Estimation method	$\delta = 5$	$\delta = 10$	$\delta = 20$	$\delta = 40$
Aver_flat	0.4250	0.4250	0.4250	0.4250
Aver_theta	0.1117	0.1123	0.1157	0.1221
Harm_flat	0.1687	0.1687	0.1687	0.1687
Harm_theta	0.1113	0.1126	0.1159	0.1202
Join_flat	0.3457	0.3338	0.3423	0.3617
Join theta	0.1113	0.1108	0.1157	0.1173

As expected the methods based on constant velocity functions (except Join_flat) remain unaffected by the disturbances. Also the effect on the methods based on θ-a function seems surprisingly small.

How do results depend on a priori data?

A difficulty with the methods based on θ-functions are that the result depends on the actual choice of a priori data. There are mainly two issues that should be clarified in this context:

- How much a priori data are needed to yield acceptable results? Or one should rather ask: How are results affected when the amount of a priori data is diminished?
- Getting data from sparse traffic should be no difficulty. But for dense traffic there are different degrees of congestion. How do results depend on the traffic density when a priori data are collected?

In Table 3-10 we display the error for methods based on θ-functions for different values of random perturbations in the dense a priori function. For each 5-meter interval the function $h(x)$ has randomly been multiplied or divided by the factor $1+r\cdot\eta$ where *r* is a random number between 0 and 1.

Table 3-10: Average relative error of travel time estimates for different perturbations in dense a priori function

Estimation method	$\eta=0.2$	$\eta=1$	$\eta=5$
Aver_theta	0.114	0.141	0.268
Harm_theta	0.112	0.135	0.220
Join theta	0.114	0.153	0.366

It seems that the harmonic average is the most robust method with respect to disturbances in the dense a priori function. The result in Table 3-10 may give some clues on how many observations are needed to form the dense a priori function.

To study the other issue we compare results for different levels of OD demands used when the dense a priori function is computed (see Table 3-11). The OD matrix corresponding to peak hour traffic has been multiplied by the factor γ to generate various dense a priori functions. The sparse a priori function has not been changed. The computations have been done for peak hour traffic conditions

Table 3-11: Average relative error of travel time estimates during peak hour traffic for various levels of dense a priori functions.

Estimation method	$\gamma=0.8$	$\gamma=1$	$\gamma=1.2$
Aver_theta	0.178	0.112	0.165
Harm_theta	0.176	0.110	0.128
Join theta	0.283	0.110	0.236

The average relative error is smallest when the dense a priori and current results are generated for the same traffic load. It seems that one should avoid using uncongested data when computing the dense a priori function. This is consistent with the results in Table 3-2 and Table 3-10 where the θ-function method proved less successful when current traffic condition is more congested than for the dense a priori function.

3.3.6 Results for Vienna floating car data

In this section some results are presented where travelling time estimation methods have been applied to real floating car data from taxi cars in Vienna (see section 3.3.2.3). In particular, the focus in

this section will be on the dynamical properties of the estimation methods, i.e. how do the travel time estimates vary with time.

Two routes, for which the true travel times for each vehicle have been obtained from AVI data, have been studied. The first route (nr 29) is 2.4 km long and consists of 20 links, the other route (nr 69) is 3.0 km long and consists of 21 links.

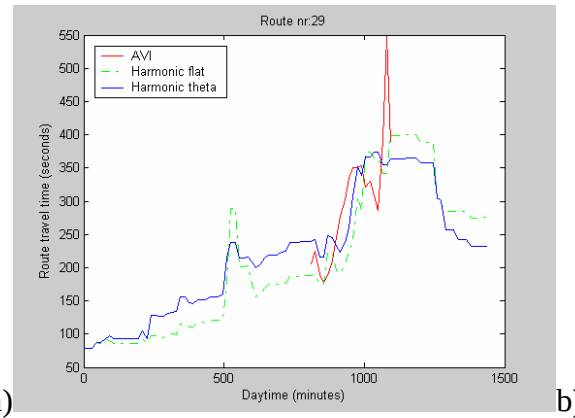
FCD from taxi cars have been collected during two days (2005-09-28 and 2005-10-06) from the links that are part of these routes. Simultaneously, although restricted to four hours in the afternoon, 2 p.m. 18.30 p.m., AVI data have also been collected.

The total number of observations along the two routes during the two days is 7111. In order to compute dense and sparse a priori velocity profiles for each link (see section 3.3.4.3) the observations have been divided into two groups: sparse and dense traffic. This division is based on the mean velocity during different hours. Dense traffic conditions is defined as occurring between 4 p.m. and 8 p.m., while sparse traffic occurs else but excluding the morning rush hour and the afternoons after 2 p.m.. The actual experiment is computed for only one day: 2005-09-28 for route nr 29 and 2005-10-06 for route nr 69.

For each link along the routes travel time estimates for some aggregation period (usually 15 minutes) are computed from taxi FCD using methods described in sections 3.3.4.3 - 3.3.4.7. From the corresponding AVI data true travel time estimates are obtained by (arithmetic) averaging of travel times for each individual vehicle travelling along the route during the same time period. (The moment when the vehicle enters the route defines the time period.)

Note that, in this section, $\omega = 0.9$ has been used for the θ-function method. A lower value on ω (which means that low velocities are cut off at a higher level) seems to yield results that underestimate the true travel times.

In Fig. 3-21 a comparison between flat velocity method and θ-function method is shown applied to the two different routes. The equipment ratio is 5% in all cases and the aggregation period is 15 minutes. The lag between AVI data and the estimates that occurs for Route 29 is probably due to unsynchronized clocks during measurements. Unfortunately, AVI data used for validation was only available for a few hours in the afternoon. It seems that for Route 29, the constant velocity method manages somewhat better to model the increase in travel time that occurs around 5 p.m. (minute = 1020). On the other hand the θ-function method responds better to the increase for Route 69.

a) 

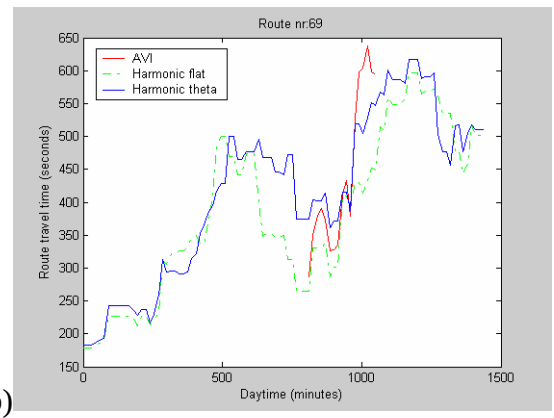
b) 

Fig. 3-21: Comparisons between constant velocity method (Harmonic flat) and θ-function method (Harmonic theta)

How do results depend on the lengths of aggregation periods? From Fig. 3-22 it is seen that the estimated travel times is very stable and is not much affected by the decrease in period lengths while the fluctuations in AVI data increase with shorter period lengths. One reason for the stability in estimated travel times is that the temporal combination method (with $\alpha = 0.3$ used here) filters the fluctuations that would otherwise occur.

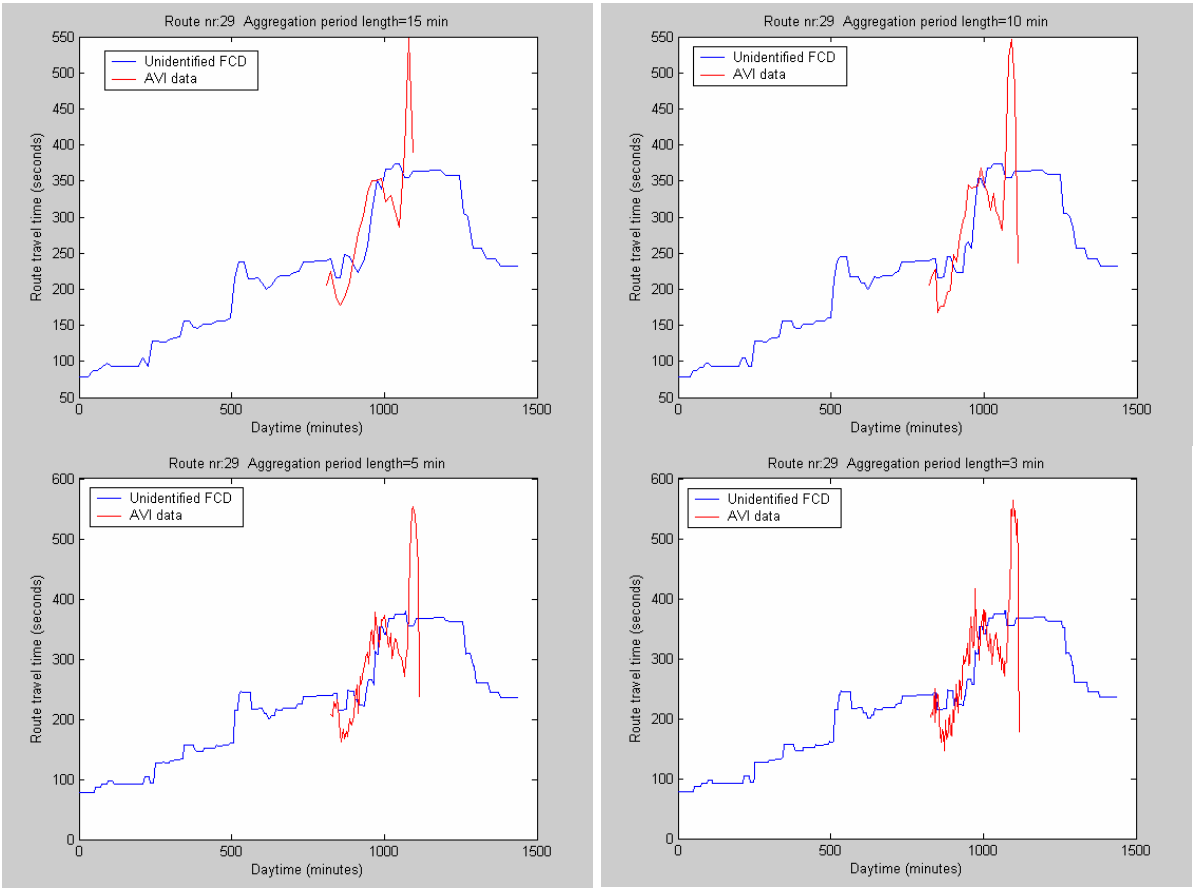
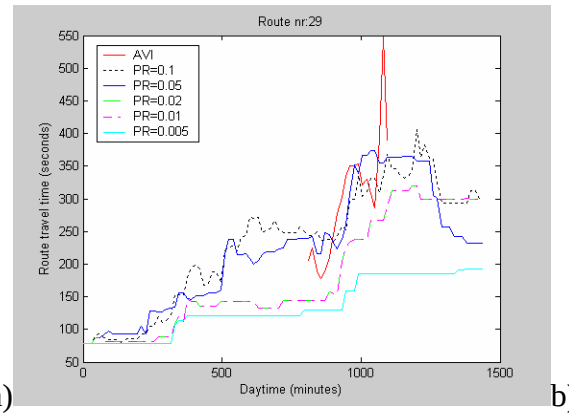


Fig. 3-22: Travel time estimates for various lengths of aggregation periods (Equipment ratio is 5% and Harmonic theta method has been used)

How do results depend on the equipment ratio? In Fig. 3-23 it is seen that results may vary much depending on the equipment ratio.

a) 

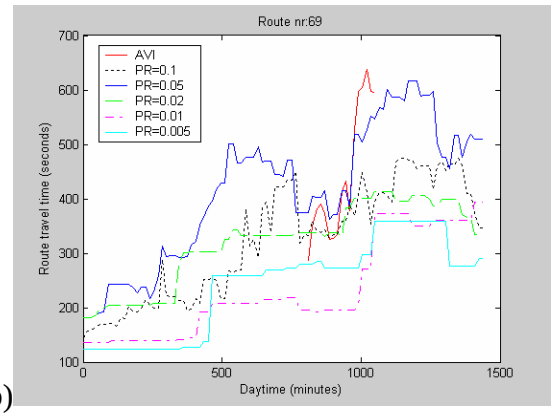
b) 

Fig. 3-23: Travel time estimates for various equipment ratios (In all cases Harmonic theta has been used. In a) results for Route 29 and in b) for Route 69 are shown)

It is difficult to draw any final conclusions from the above examples. A fact that may have affected the results is that the observed velocities reported by the taxi cars are not instantaneous, but instead average velocities over periods of 30 seconds or more. This implies that very low velocities are (often) avoided and there is a risk that the true velocity becomes underestimated. Notice also that a constant velocity minimizes the travel time (between two points) when the mean velocity (over the distance) is fixed.

Concerning the conflict between stability and fast response time there is a complex interplay between a numbers of parameters:

- α *:* increased α generally leads to better stability but the response time is increased
- ω *(or lower bound of velocities):* if the lower bound of permitted velocities is reduced the travel time estimates can be grossly overestimated. Probably, although not investigated here, the adaptation to very congested travel times can then also be quicker.
- Harmonic versus average mean: harmonic average of travel times have a tendency to be more stable than arithmetic average and the huge (unstable) overestimates that can occur in with the latter method are usually avoided. However, harmonic averages has a tendency to underestimate the travel time (at least when ω is set too small) and the large travel times during very congested periods may be missed.

3.3.7 Conclusions

In this chapter various approaches to estimate travel times on links from unidentified probe observations of position and velocity have been investigated. The work has essentially been based on data generated by the (micro-) simulation program AIMSUN on a small network. We feel that simulated data is totally adequate and suitable for this kind of analysis. The possibility to compare the results from various formulations of algorithms with the true travel times is invaluable. Characteristic features can be discovered and studied which would be impossible using only real data. The low information content in data from unidentified vehicles gives rise to large uncertainties in estimated travel times, probably overshadowing the modeling errors from the simulation program.

However, using simulated data is not entirely without problems. No match-making of observations to the network has been done in this study and this gives rise to (some of) the difficulties discussed in section 3.3.4.9.

The results in section 3.3.5.1 indicate that spatial modeling of velocity profiles along links using θ-functions has a good potential to improve travel time estimates as compared to methods based on constant functions. The θ-functions often seem to model the true velocity profile better than a constant velocity profile. In Table 3-2 the average relative error in travel time estimation during peak

hour conditions is 0.24 when using the θ-function method and 0.41 when using the constant velocity method.

From section 3.3.5.2 it may be concluded that the travel time estimates depend on how the individual observations are combined during some fixed or flexible time period. Three different spatial combination methods have been investigated: arithmetic average of travel times, (A), harmonic average of travel times, (H), and piecewise joining of segments of velocity profiles, (J). Applied to constant velocity functions, (H) has in general produced more accurate results than the other two methods. Applied to θ-functions the results are less unambiguous but (H) and (J) often seem more accurate than (A).

An important issue (especially in very congested traffic conditions) is how to handle low velocities. Even small changes in implementation details may affect results significantly. There exists an intricate interrelationship between these effects and the spatial combination methods. For instance, if the low velocity problem is neglected (or a very low lower bound for the velocity is set) the distribution of estimated (individual) travel times will be skewed and (sometimes very) large over-estimates will occur. Using (A) or (J) will then in general often yield over-estimated travel times, while this is often avoided when using (H). If instead, the lower bound is increased (in the form of a constant bound or the method with ω, see section 3.3.4.5), the skewness of the distribution of travel times tend to decrease. As a consequence, the differences in results from (A), (J) and (H) diminish. On the other hand, when the lower bound increases, a situation similar to the one in Fig. 3-23 e-f often seems to occur: the velocity profile is modeled quite well, but the stops (which can be of significant lengths) will not be compensated for and the estimated (individual) travel time underestimated. In these cases (H) often tend to yield larger underestimations than the other methods.

Thus, increasing the lower bound on velocity causes the travel time to decrease. Applying (H) has a similar effect while (A) does not. To attain the best result a delicate balancing between these factors are needed. These issues are further dependent on the number of observations available. If we only have one observation, methods (A), (H) and (J) will yield identical results (providing the implementation details are identical). In this case the choice of lower bound constitutes the single control mechanism. If on the other hand many observations are available (for the current time period) the differences between (A), (H) and (J) may be significant.

Two different approaches for temporal combination have been studied, the first using a fixed number of latest observations and the second combining the current travel time with the normal travel time . No decisive results are obtained for which of these approaches is the most accurate. Both methods depend on parameters that must be specified. Optimal values for these have been estimated empirically, but it would be desirable to construct a model where these parameters are estimated dynamically. Among the approaches for travel time estimation from unidentified observations that have been studied here we suggest using harmonic averaging with the θ-function method. For temporal combinations we suggest the method of combination with $\alpha = 0.2$ or 0.3. This algorithm is the one suggested for use in section 3.6

Results in section 3.3.5.4 indicate that the θ-function method is very robust for disturbances in vehicle positions (providing the observations are assigned to the correct link). As to the computation of the dense a priori function, results indicate that the congestion rate should be large enough when the dense function is computed (see Fig. 3-17 d), Table 3-2 and Table 3-11). If the dense function is computed using data with too low congestion rate, the θ-function method seems not to handle congested situations very well. On the other hand, the explicit dealing of the low velocity problem (see section 3.3.4.5) also strongly affects the situation and makes decisive conclusions difficult.

Overall, travel time estimations based on unidentified vehicle data is not very suitable for handling very congested traffic with low velocities. The resulting estimates will have large uncertainties.

There exist a number of possibilities to further develop the ideas discussed here. For instance, instead of using only two a priori traffic states, the sparse and dense, a more refined division could be done.

(E.g., one might separate between dense traffic and very congested traffic.) Also, data from other sources could be used to improve travel time estimates for particularly difficult links. The temporal combination methods could certainly be improved by more advanced methods. The very important issue of travel time prediction is closely related to this problem.

3.4 Floating Car Travel Time Estimation for Urban Network

3.4.1 Introduction

For the study of probe vehicle based travel time estimation, it is essential to study the different variables involved. Specifically in urban networks, there are some peculiarities in this study whose implications will be developed in this chapter.

Modelling was used to study behaviours of the different parameters involved in floating car travel time estimation, such as the equipment ratio, link definition, aggregation period and some substitution methods for filling missing values. As a case study, Lausanne city centre model was selected and tested, so all figures and results were obtained for that network.

Information about the network model itself and experimental characteristics are detailed in chapter 3.5

3.4.2 Equipment ratio

The equipment ratio (thereafter ER) is obviously the main parameter of interest in travel time estimation with probe vehicles or floating car data (thereafter FCD) techniques. As on-board units or equipment are the economically sensitive aspect, obtaining a travel time accuracy level acceptable for the application using these data while minimizing the number of equipped vehicles, is the goal behind this type of techniques.

This research does not aim to determine an acceptable accuracy level, as this criterion is more subjective than really quantitative. Therefore, efforts will be concentrated on determining the relationship between the ER and the expected benefits, mostly in terms of travel time estimation.

The ER notion needs however to be clarified as it is not univocal as it can be seen by comparing its various definitions present in the literature. If only limited to a materialist point of view, the ER should be considered as the ratio between the number of equipped vehicle and the total number of vehicles belonging to people living in a limited area. Such definition is only considering the technical and economical aspects (like the ABS or automatic gear box penetration rates could be considered). However, this approach does not allow us to determine the ratio between the number of equipped vehicle and the total number of vehicles driving in a specific area and at a fixed time. The latter definition is in fact the one needed to correctly link the ER with the FCD technique parameters.

For this reason and in the context of this simulation-based research, the ER is defined in the following way: the ER is the proportion of vehicles generated at the origin points (centroids) that will be used to get individual travel time information records. Thus, for each generated vehicle, a randomly based process determines if it is considered as equipped or not. The ER, chosen at the beginning of the simulation, is therefore the probability of being considered as equipped. As this random process will rely on an external random number generator, all simulation runs will be different, even for running the same replication.

3.4.3 Link definition

The usual way to model a road network is to represent intersections as nodes and the roads as links linking the nodes (one per traffic sense). Fig. 3-24 is a typical example of this type of road network representation with two nodes linked by two links, on per direction.

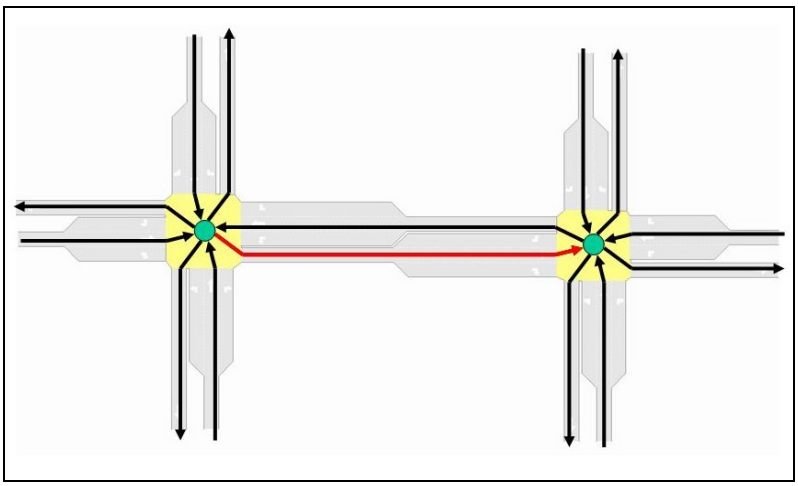


Fig. 3-24: Classical representation of link and nodes

Fig. 3-25 represents the individual link travel time measurements recorded on a specific link of the Lausanne simulation model. This link is situated in between two signalized intersections and the record period is 14 400 seconds, or 4 hours. The important variability of the measurements population (22 seconds standard deviation) is characteristic for a urban link, in particular one with a signalized intersection downstream. The average variance coefficient is 0.71 (calculated on the basis of 16 periods of 15 minutes).

This variance can be partly explained by the fact that all vehicles that used this link have not been exposed to the same or, better said, to similar traffic conditions. Indeed, vehicles can access the observed link using different upstream links (called entrance links). Likewise, there are different downstream links that can be used to leave the observed link (called exit links). For each pair of entrance and exit links, traffic conditions are different for two main reasons:

- Traffic signal synchronization of upstream and downstream intersections implies different waiting times for each pair.
- The queue lengths for the different turning movement used to exit the link maybe different, especially if lanes are dedicated to specific turning movements.

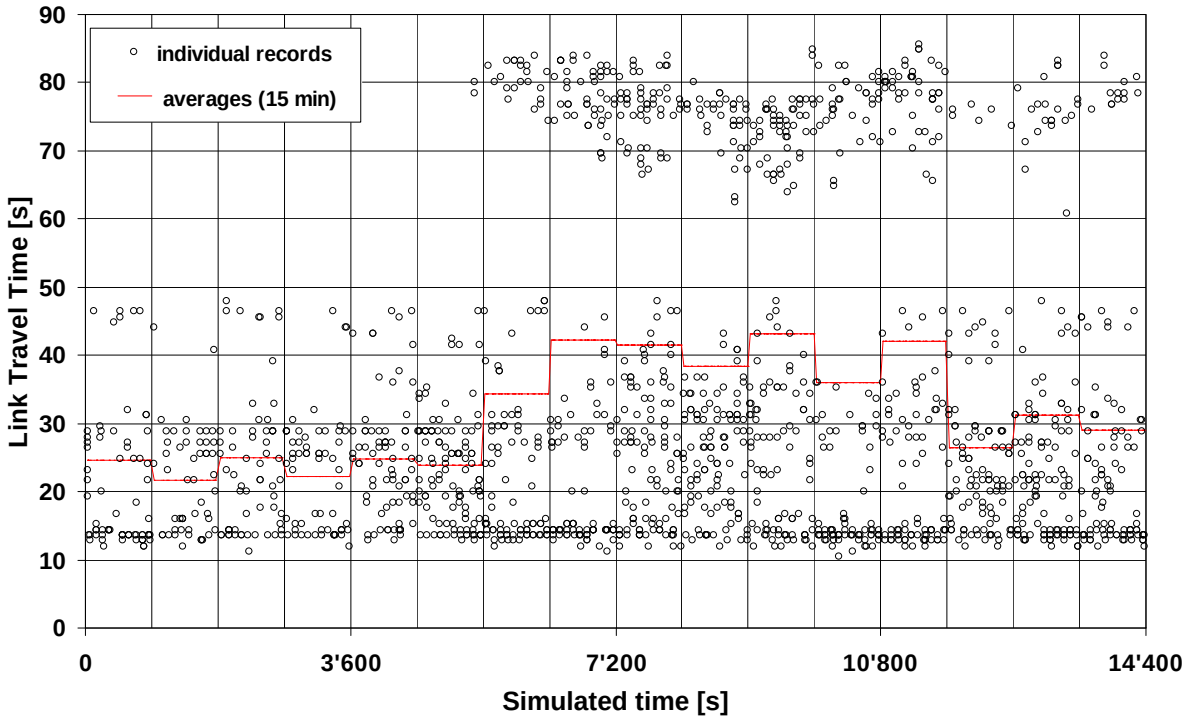


Fig. 3-25: Individual link travel time records for a link with signalized intersections upstream and downstream [1]

For that reason, it would be interesting to split the link travel time records in sub categories, each category gathering the vehicles using the same entrance-exit link pair. In the example illustrated by Fig. 3-26, the upstream intersection is accessible by two entrance links and, likewise, the downstream intersection can be left using two exit links. Fig. 3-26 shows individual link travel time records for each entrance-exit pair with their respective average standard deviations.

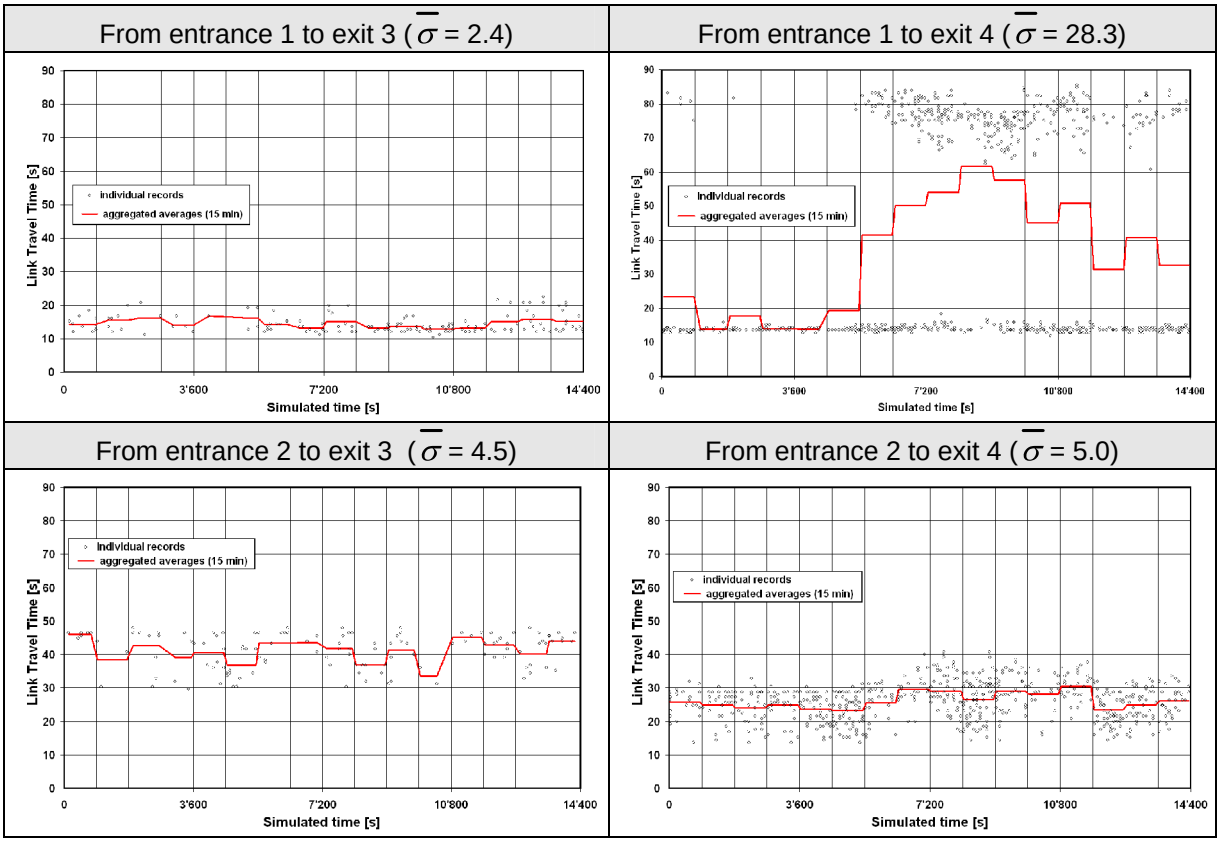


Fig. 3-26: Individual link travel time records gathered per entrance and exit links pairs [1]

This methods allows confirmation that, in three of the four cases, sub-populations variance is relatively low (the average variance coefficients being lower than 0.18). This means that for this three entrance-exit pairs, vehicles experience similar traffic conditions.

Only the pair composed by entrance 1 and exit 4 is subject of a large average variance coefficient (0.79), higher than that of the total population. This situation can be explained by the fact that the first vehicles of the platoon can reach the end of the observed link when the downstream traffic light is still green while the vehicles at the end of the platoon have to stop because it turns to red. This phenomenon is only observed after one and half hours of simulation. Until this time, the vehicles platoon size was small enough to allow all the vehicles to leave the section during the green phase.

This discontinuity in the traffic conditions experienced by the vehicles implies bimodal sub-population distribution, a typical characteristic of signalized links. The average travel times (represented by the red line in the graphs of Fig. 3-26) have, no real meaning as no vehicle experiences a link travel time close to these values. These values only give information about the ratio of vehicles belonging to one or the other mode of the bimodal distribution. Such example clearly highlights the limitation of a deterministic approach based exclusive on average travel times. Moreover, in a bimodal distribution, even median does not provide useful values to estimate travel times, and therefore for bimodal distribution it could be better to calculate the average of each mode and give a probability to belong to one or the other.

However, as explained before, FCD travel time estimation accuracy is higher when the measurements population variance is smaller. One can expect a good accuracy level for travel time estimation of the three of the entrance-exit pairs while the fourth pair level would be similarly bad as the total population s one. Therefore, from a global point of view, the sub-division into sub-links (one per

entrance-exit pair) should improve the performances of FCD travel time estimation techniques. The network representation of Fig. 3-24 would therefore be replaced by that in Fig. 3-27.

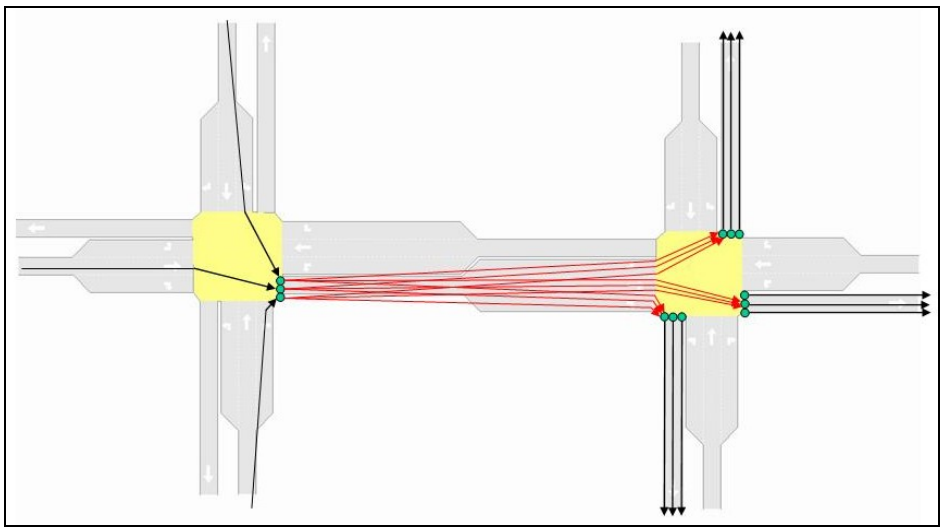


Fig. 3-27: Subdivided representation of links

However such links sub-division presents some disadvantages. Firstly, network complexity is dramatically amplified due to the increase in the number of links. As an example, such an operation for the Lausanne centre network implies a 3.6 multiplication in the number of links. Nevertheless, this aspect is, above all, limited to computational time augments, mainly for shortest path calculations. Thus, it can be considered as a technical limitation and not a fundamental one.

Secondly, this subdivision implies a decrease in the absolute size of the measurement population per link which is much more problematic. Indeed, if it has been said before that decreasing the population variance improves the travel time estimation, this statement is only valid when the ER and the population size is identical! When subdividing links, sub-population s ER may vary from the main population but, for certain, sub-population sizes are smaller than the main population. Impacts of such phenomena should consequently be examined.

A direct consequence of the population sizes decrease implied by the subdivision method is an significant increase in the number of links without probe vehicle measurement (no equipped vehicle using the link during the considered aggregation period), hereafter referred to as missing. Indeed, for a given ER, the smaller the population size, greater the probability that no probe vehicles record. Presence of missing implies that a substitution method has to be applied in order to obtain a link travel time for the considered aggregation period. Generally, but not always, this substituted values are supposed to be less accurate than the FCD ones. The specific topic of the substitution methods is treated later on.

In summary, subdivision of links depending on the entrance and exit link used offers advantages and disadvantages. The ratio between them is a function of various parameters. The upcoming parametric analysis should provide answers on when the subdivision method offers better performances than the classic approach.

At this point, only a complete subdivision approach has been considered. However, partial subdivisions should also be considered as they may offer better advantages/disadvantages ratio. Partial subdivision, means to differentiate links in function of the exit links (without considering the entrance ones) or in function of the entrance links (without considering the exit ones). This would be particularly beneficial if the population variance decrease is for the most part due to one of these two differentiation types.

In order to get information on these considerations, four link definitions will be taken into account for this research:

- Classical (usual representation)
- In (partial subdivision in function of entrance links)
- Out (partial subdivision in function of exit links)
- In-Out (complete subdivision)

Concrete results based on the Lausanne centre model simulation will be provided later on, once the aggregation period topic has been treated. They will permit quantification of the advantages and disadvantages of each link definition under various conditions.

3.4.4 Aggregation period

As explained earlier, the link travel time deterministic approach implies a time discretisation in periods during which link travel times (or, more precisely, their model) are supposed to be constant. The duration of these aggregation periods (AP hereafter) plays an important role in FCD techniques. Indeed, like the subdivision method, AP has an influence on travel time measurements population variance and size.

In order to better understand how AP may influence the population variance, one should consider two types of variability in the time dependent context:

- short term variability
- medium term variability

Short term variability is only due to the different traffic conditions experienced by the vehicles as described earlier. This variability always exists, even if the traffic demand is perfectly constant.

Inversely, the medium term variability is directly implied by the time dependent traffic demand changes. Global traffic conditions are changing as a function of traffic flows variations, and general tendencies of link travel time increase or decrease may be observed.

Fig. 3-28 illustrates clearly the difference between these two types of variability. During the first 6000 seconds of simulation, only the short term variability is observed, no global increase or decrease tendency being observed. During the rest of the simulation, the short term variability is always present, but the medium term one also appears. Indeed, global tendency for travel time increase or decrease are clearly observable.

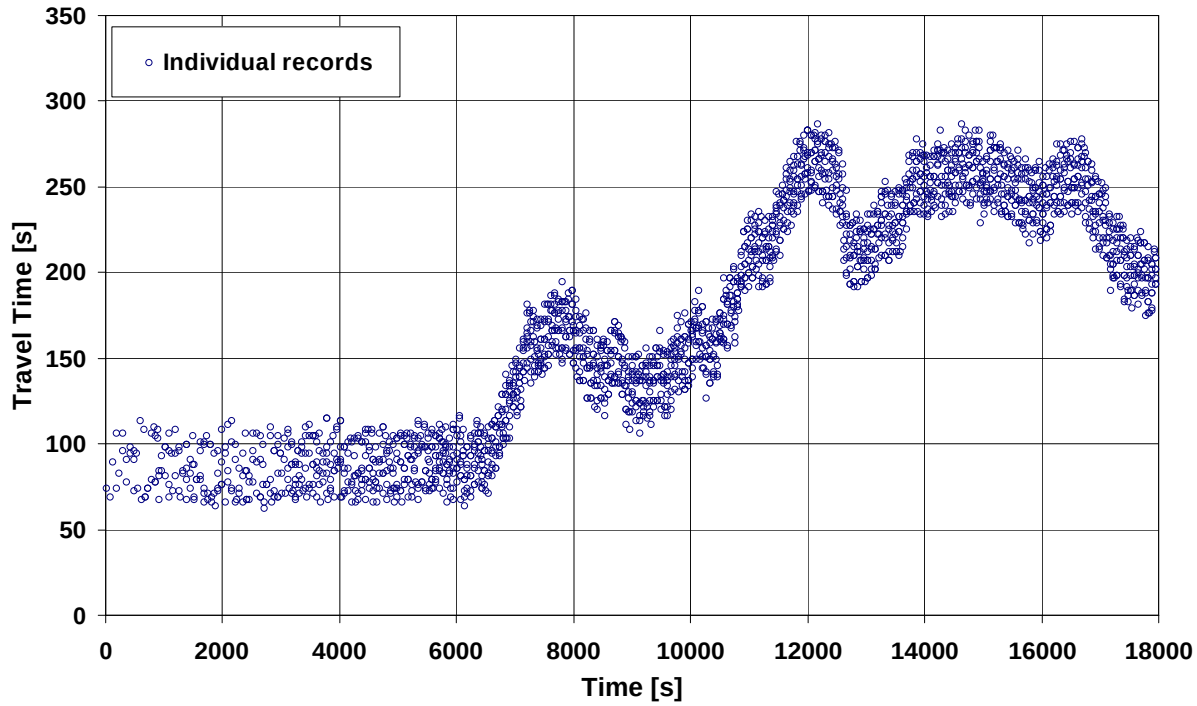


Fig. 3-28: Short and medium term variability cumulated

Consequently, the longer the aggregation period, the more the link individual travel time population variance presents an important medium term component. Logically, the shorter the AP, the more limited to the short term the population variability is, implying a decrease in population variance. As an example, if, as shown in Fig. 3-28 the five hour records would be aggregated in 30 minutes periods, the average standard deviation per period would be 18.8 seconds. This value would decrease to 13.5 seconds if aggregated in 5 minutes periods. Once again, this decrease in population variance should imply an improvement in the travel time estimation accuracy. Additionally, the short aggregation period allows better capture of the temporal variation of link travel time with a frequent information refresh. That is fundamental for applications like dynamic route guidance.

However, as in the case of the subdivision method, reduced aggregation period duration implies a decrease in measurements population size which, as seen before, leads to an increase in missing values. Once more, this approach has then advantages and disadvantages. The AP offering the best performances would have to be selected as a function of other parameters playing a role in the FCD technique.

As for the ER, a series of values to be tested has to be determined for the parametric analysis. For this research, the following AP values have been chosen: 60, 120, 300, 600, 900, 1200 1800 and 3600 seconds. They cover in a representative way the range of values that can be accepted as reasonable.

3.4.5 Combined effect of the aggregation period and link definition choice

In order to confirm the two previous chapters findings, a first analysis based on the Lausanne centre network simulation has been undertaken. Fig. 3-29 represents the average population s standard deviation calculated over all the time periods and links of the network. Each curve represents a different link definition.

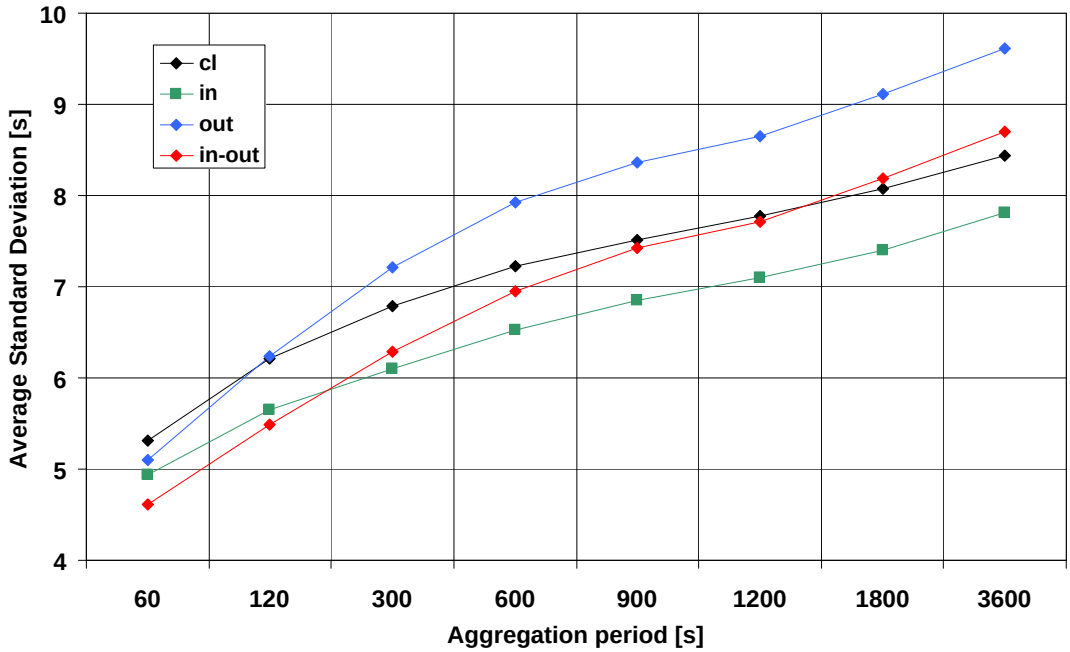


Fig. 3-29: Average standard deviation of the individual travel time records populations for the Lausanne centre network

Firstly, these results confirm the fact that the population variability decreases with smaller AP. Indeed, the four curves present a clear decrease tendency going from the highest values of standard deviation for 3600 seconds AP to the lowest value for 60 seconds AP. Then, an observation of the four curves for short AP (when only short term variability is considered) confirms that a complete subdivision offers the best performance (4.6 seconds against 5.3 for the classical definition) while partial subdivisions present in between results.

However, when long term variability plays an important role (for higher AP), conclusion is different. The advantage of a complete subdivision over the classical approach tends to decrease and finally disappear for AP higher than 1200 seconds. For partial subdivisions, conclusions are even more contrasted. For AP values higher than or equal to 300 seconds, the out subdivision description is the worse performing option while the in subdivision is the best one.

At this point, one should not forget that AP and link definition choice has an important impact on the lack of measurements i.e. periods or links without FCD, that this graph doesn t show.

3.4.6 Substitution method for missing

Missing (specific period and link for which no probe travel time has been recorded) is a fundamental feature of FCD techniques. Their proportion, as it will be shown later, can be very important and, obviously, will mainly depend on the ER.

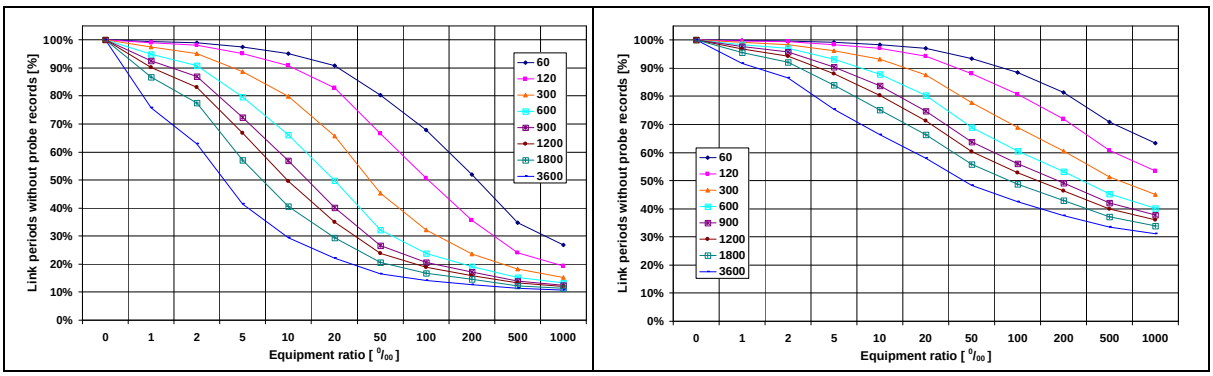


Fig. 3-30: Percentage of missing data in function of the aggregation period (AP) and the equipment ratio (classical link definition on left and in-out one on right) [1]

An analysis taking into consideration the ER, the AP and the link definition has to be undertaken in order to know what the missing proportion for each parameters combination is. Fig. 3-30 presents these missing proportions for a classical link definition on the left and an in-out one on the right.

3.4.7 Handling missing data

3.4.7.1 The static method

The static method, also called free flow thereafter, relies on replacing missing values by the theoretical travel time of the observed link under free traffic conditions (this explains the free flow denomination). These free flow values are the ones used by static navigation systems and are calculated based on the link length and a link category dependent average speed. In the case of urban links, most of the commercial navigation systems (ViaMichelin, MapSonic, Garmin) use a 30 km/h (8.3 m/s) average speed. As the Lausanne centre network is mainly composed by urban links, the static link travel time (Free flow travel time, thereafter FLT) is calculated in the following way:

$$ELT_{a,\tau} = FLT_{a,\tau} = \frac{L_a}{FS} \qquad (3.11)$$

with L_a being the link length and FS the free flow speed (8.3 m/s in the Lausanne centre case). It must be highlighted that FLT are independent of the considered period, thus explaining why it is called static .

If replacement of missing per FLT values offer the possibility of assessing the intrinsic performance increase resulting from the probe records only, this approach cannot be considered as a satisfactory method in the objective of maximising the FCD technique performances from a global point of view.

3.4.7.2 The stationary method

The stationary method is based on the following concept: if no probe measurement is available for the current time period, the travel time estimated on the previous period is considered as valid. Thus, if no probe data is recorded during period τ of link a, then:

$$ELT_{a,\tau} = ELT_{a,\tau-1} \qquad (3.12)$$

By applying this method, a stationary hypothesis is consequently formulated. Thus, traffic conditions are considered as stable in time. This approach may be accepted if the medium term variability is limited (which is usually not the case, especially around peak hours) or for very short AP.

3.4.7.3 Historical profile method

The idea upon which this approach relies on is that an important similarity exists in term of traffic conditions from one day to another. These similarities are especially present when comparing two same days of the week (the Tuesdays for example). This can be explained by similar traffic demand for each week day.

By cumulating probe data records from the past, an average daily travel time profile can be obtained. One will then refer to the typical Tuesday profile. Differences between the typical day profile and the current day one are varying. They will be important if a special event makes the current day incomparable to normal ones, as is the case with incidents, road closures, sport events, etc

The probe vehicles records are consequently stored in a database in order to calculate the estimated historical link travel time (thereafter EHLT) which can be defined as:

$$EHLT_{a,\tau} = \frac{\sum_{d=1}^{D} \frac{\sum_{f=1}^{F_{d,a,\tau}} LT_a^f}{F_{d,a,\tau}}}{D} \qquad (3.13)$$

For later comparison, the real historical link travel time (hereafter HLT) calculated with the individual link travel time records of all vehicles (not only the equipped ones) are defined as:

$$HLT_{a,\tau} = \frac{\sum_{d=1}^{D} \frac{\sum_{n=1}^{N_{d,a,\tau}} LT_a^n}{N_{d,a,\tau}}}{D} \qquad (3.14)$$

In both definitions, D is the number of storage days used for calculating the historical profile.

In the practice, historical profiles are calculated on a more complex way, for example by taking into account seasonal tendencies or by putting different weights for the most recent recorded days. However, as this research is based only on one OD matrix, there is no sense in applying these types of refinements and, therefore, the primary definitions are applied.

The use of historical profile is without doubt a very important feature for FCD techniques. Indeed, it can be shown that, for an individual link travel time records population with a fixed variance and ER, the estimation error decreases if the population size increases.

Consequently, the link travel time of a link during the period τ of a normal day of the week can be considered as a unique phenomena with a fixed variance and with a population size increasing with the number of storage weeks. So, even with a very limited ER, one can get an accurate estimation of the HLT with the EHLT if the number of days stored in the historical data base is sufficient. To illustrate this property, Fig. 3-31 shows the evolution of the variance (represented by the standard deviation) of the cumulated population (pink curve) and the relative difference | EHLT-HLT | / HLT (blue curve) in function of the number of accumulated days in the historical data base. The link used to

get this graph is the Lausanne model s Cesar-Roux street link during the 17h00 to 17h15 time period. The considered ER is 10 $^{0}/_{00}$.

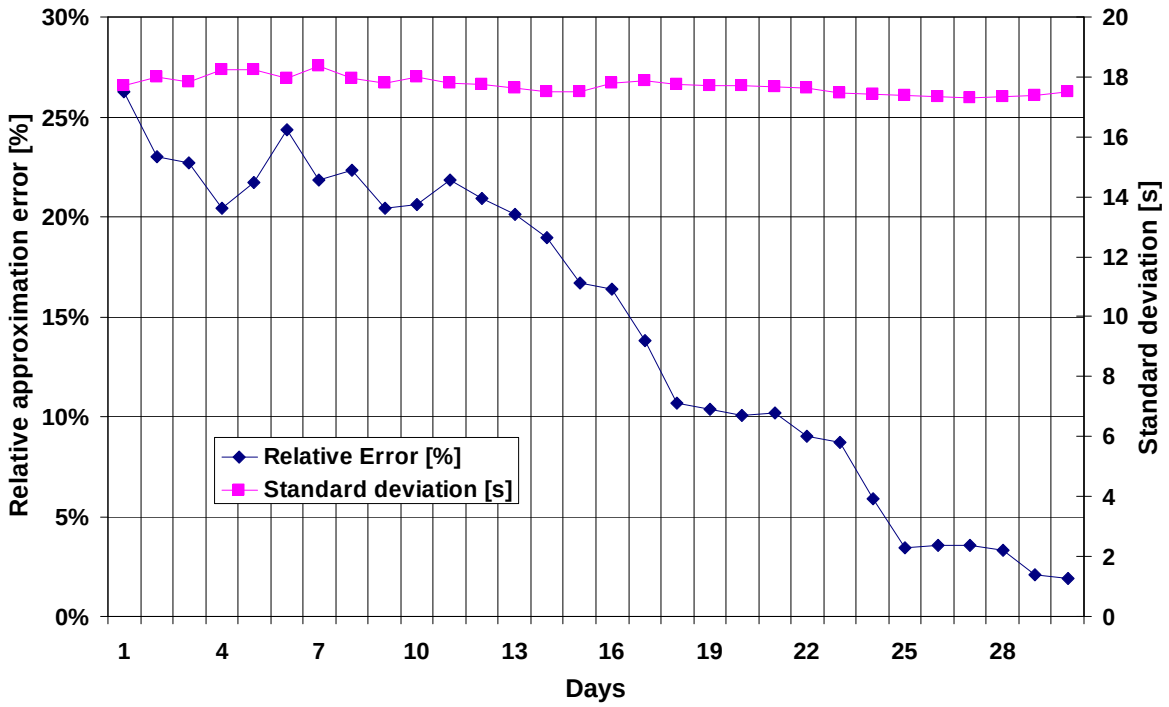


Fig. 3-31: Evolution of the relative approximation error (with ER=1%) and standard deviation of the record population in function of the number of recording days [1]

Fig. 3-31 confirms that the variance of the population is stable even if its size increases with the number of days stored. On the other hand, the difference between the estimated and the real historical link travel time decreases as the number of recorded days increases, reaching a value close to zero after 30 days (30 weeks if each week day is considered separately which is not the case for this simulation based study).

It is important to highlight that these simulated results have been obtained relying on the following hypothesis: one OD matrix is considered as representing the demand of a typical day (Tuesday for example), the difference between the specific days (3rd of April, 10th of April, 17 of April, etc.) being reproduced by changing the random seed (starting point for the random number generator) used for a specific replication. Consequently, the 30 simulated days have been obtained using the same matrix but 30 different random seeds. At this point, it would be useful to check if the changes in traffic conditions from one random seed to another are representing adequately the ones observed in the real life. This verification has not been done in this research framework.

The substitution method based on the use of the historical profile consists of replacing the missing value by the EHLT:

$$ELT_{a,\tau} = EHLT_{a,\tau} \qquad (3.15)$$

3.4.7.4 Predictive method

Many separate research projects have been providing travel time forecasting methods based on fixed location measurements (loops) or mobile ones (probes). That means that during a period τ, these

types of method offers as output predicted link travel times (thereafter PLT) for the next time periods: $PLT_{\tau,\tau+1}$,$PLT_{\tau,\tau+2}$, , $PLT_{\tau,\tau+i}$. The idea of this substitution method is to replace the missing link travel time by the one forecast during the previous time period:

$$ELT_{a,\tau} = PLT_{a,\tau-1,\tau} \quad (3.16)$$

For this research, the predictive methods that have been used are based on a neural network approach. For more information on these methods, please refer to [1]). Two types of neural network are used: one considering only the ELT as input, the other one combining ELT and EHLT. Both forecasting methods producing different PLT as output and two different substitution methods are considered for the upcoming parametric analysis.

3.4.7.5 Combined method

Until now, the ELT has been considered from individual travel time record coming from equipped vehicles or from one of the substitution method if no records have been obtained for a specific link and time period. However, there is no guarantee that the estimation based on the equipped vehicles records is the best ELT to be used. Indeed, the smaller the ER, the less the sample average matches that for the total population.

For example, it is possible that in certain cases, the historical profile (EHLT) offers a better estimation than the one coming from the records of the current period only.

Instead of having to choose between these two values case by case, a linear and weighted combination of both values seems more adequate. Thus, if one calls ELT_f the estimation based on equipped vehicles record only, then the proposed estimated value can be defined as:

$$ELT = \alpha \cdot ELT_f + (1-\alpha)EHLT \quad (3.17)$$

The main obstacle to this approach is to correctly choose the α factor. Explicitly, the higher is the ELT_f performance, higher the α should be. Consequently, there is a clear relation between this factor and the equipment ratio. However, this is not the only parameter to be taken into account. Indeed, a highly weighted EHLT (a small α) will only be preferable if the current day traffic conditions are reasonably close to the normal (represented by the historical profile), a condition difficult to check a priori. One can however assume that the closer ELT_f and EHLT are, the greater is the probability that this condition is true. Inversely, if these two values are clearly different, the probability that the current day traffic conditions are different from the normal day is high. In this case, using the combination approach would decrease the performance of the estimation method.

The influence on α on this difference between the ELT_f and the EHLT, as well as determining an upper limit over which the combination method is no longer justified is not an easy task. They are both dependent of the individual travel time population variance. To take it into account, a function α=f(ER, ELTf, EHLT) should be calibrated for each link and time period, each AP, each link definition, etc. This would represent a considerable amount of work.

Consequently, a more generic approach must be considered. It should then consider as inputs: the ER, the ELT_f, the EHLT, the AP, the link definition, the link length, the link type, the time period and the downstream junction type (signalized or not). Taking into account the number of parameters and their difference in type, an analytical expression seems impossible to develop. For this type of problem, neural networks (NN) have been proven to perform well. For this reason, a one hidden layer MLP (Multi layer perceptron) has been used to tackle this combination problem (for more details refer to [19]]). The output of this NN being the $ELT_{a,\tau}$, the $ALT_{a,\tau}$ is used as target value for the training process.

Four data sets of 15 000 input/outputs packages (one package being all the inputs and outputs recorded by simulation for a specific link and time period) has been used: the first for the training process, the three others for the validation. The cost function to be minimized is the MSE (mean square error). Stabilization during the training process has been obtained after 100 to 150 iterations. The Table 3-12shows the results of this training process and for the three validation data sets. Considering the stability of the validation results and their proximity to the training ones, the NN can be considered as well calibrated.

Table 3-12: Cost function (MSE) value for the combination method neural network

	Training	Validation 1	Validation 2	Validation 3
MSE (ELT/ALT) [s^2]	13.9	16.2	15.3	15.1

In order to assess the combined method performance, the MSE indicator has also been calculated (based on the first data set) by comparing the target value ALT to the ELT_f and EHLT. The results, respectively 35.2 and 69.3 s^2, show a clear advantage in using the combination method instead of only the equipped vehicles records or the historical profile. The excellent determination coefficient (R^2 = 0.97) obtained with this method has also to be highlighted. Finally, a NN that only uses the ELT_f and EHLT as input has also been calibrated. A 36.6 s^2 MSE obtained shows the importance of including the other inputs listed above.

3.5 Simulation

3.5.1 Introduction

Microsimulation has been found to be an invaluable tool to analyse and evaluate FCD systems. It can complement costly field trials and can be even more efficient in comparing different methods for data processing and studying the effect of varying the different parameters involved, such as the equipment ratio (i.e. the proportion of probes in the network) or the reporting interval(i.e. how often reports are sent from the probes).

Simulation studies are generally less expensive than field trials and give unique possibilities to compare estimated journey times with the complete picture of all individual journey times. In real trials there is always an uncertainty about what the actual reference journey time is.

The purpose of this chapter is to show how microsimulation can be used to model a network and collect data for different FCD systems, allowing the creation of different experiments which can easily be associated with typical traffic situations on a network. It also allows the trial, in a well studied and controlled environment, the effect of certain essential parameters for FCD, so their relative effect can be specifically determined, such as the equipment ratio or the reporting interval.

Thus it will be easier to determine the amount of data necessary to get a proper estimation, with a well established confidence margin. It is also possible to compare different methods under the same controlled conditions, such as comparison between identified and unidentified floating car data methods

3.5.2 Use of microsimulation

There are three kinds of simulators: macrosimulators, mesosimulators and microsimulators. Each of them has advantages and disadvantages, and in this section the decision to use a microsimulator as one of the best ways to compare different FCD methods and techniques is explained.

Macrosimulators are static i.e. supply and demand not changing over time, and modelling traffic as a homogeneous flow characterised by parameters such as capacity, speed and density. Flow is determined by macroscopic physical laws, specifically the conservation of flow at junctions (Kirchoffs Law), and link speed flow density graphs based on empirical studies. Traffic assignment is determined by an iterative process with respect to Wardrop s static equilibrium

Mesosimulators are dynamic, so traffic variation with time is considered. Traffic representation is based on an individual description of vehicles but the rules are applied to a vehicle group. The outflow of these groups is determined by similar rules to those used for macroscopic models.

Microsimulators are also dynamic with vehicles treated as individuals, and vehicle flow is determined by behavioural laws between individual vehicles and their infrastructure. Principle models concern car following and lane changing.

In this study the acquisition of vehicle data, as well as equipped vehicles guidance, has to be realistic, so the simulator has to possess the following characteristics:

- Represent each vehicle individually
- Model the flow discontinuities related to the traffic lights at intersections
- Show differences in travel times experienced by each vehicle
- Allow each vehicle to be given a different route
- Allow itineraries to be modified during the trip

So microsimulators are the best option in order to perform experiments with a simulated floating car fleet, obtain data from them and establish a comparison between the different existing methods.

3.5.3 Use of AIMSUN

Different microsimulator softwares are available, and AIMSUN, developed by the Universitat Politècnica de Catalunya and later commercialised by the company Transport Simulation Systems was chosen for this study.This software has been chosen for the following reasons:

- As it is a commercial software package, it has a well-developed user interface which permits an easy and precise network editing.
- It allows user modification of the traffic parameters to evaluate solutions other than the default one.
- The API (Application Programming Interface) allows a complete interaction with external applications and thus it is possible to get all data concerning the floating car fleet and convert this into the right format for each external algorithm that evaluates the FCD to determine travel times.
- It has an easy to use interface which permits interaction with the different parameters that may affect the experiments in Floating Car Data experiments, such as the penetration rate and the reporting interval.
- It is also very successful in obtaining statistics for speed and travel time with different aggregation intervals.
- The software was developed by TSS, partner in this project, and their support is been essential to modify the simulator and obtain the correct data to be used as input for the different FCD methods to generate travel times.

3.5.4 Network description: The Lausanne network

To perform the experiments two types of network were chosen. This was essential to test Floating Car methods in different situations, in effect evaluating the influence of the network type itself.

Both networks are parts of the Lausanne network, although they have very different characteristics, and each one is focused on different aspects, and characteristics, as shown in Table 3-13. The most noticeable difference is a clear difference in size, but other differences are described below

It should be noted that the smaller (100% urban) network corresponds to the centre of the bigger (mixed urban and interurban) network.

Table 3-13: Basic characteristics for Lausanne networks

	Lausanne Centre	Lausanne Agglomeration
Polysections	653	4178
Total Length of Sections (km)	54	677
Matrix O/D Size	80 x 80	292 x 292
Nodes	230	1.476
Signalized Intersections	49	148
Peak Hour Traffic (veh/h)	16.437	61.594
% Of PH Traffic in experiments	80	50
Simulated Hours	4	1

The following figures show the schemes for both models. Fig. 3-32corresponds to the centre of Lausanne city, a dense and non-aligned succession of streets in a typical configuration of an old town. The area also encompasses the surroundings of the main railway station.

Trips in this network are not long, along quite congested streets for the simulation made and there are a high percentage of signalized intersections, so traffic conditions are discontinuous.



Fig. 3-32: Lausanne Centre Network

Fig. 3-33 corresponds to the complete area of Lausanne city, including its ring highway and its access and main arterials. This model is useful for showing the interface between interurban and urban

networks, with some larger and longer arterials, and correspondingly a sparser dispersal of intersections.

From the floating car data aspect, this network allows longer trips, in more continuous traffic conditions, as a result of a lower proportion of signalized intersections.

Public transport lines were also implemented and simulated in the larger model, due to the significant impact they have on real traffic conditions.

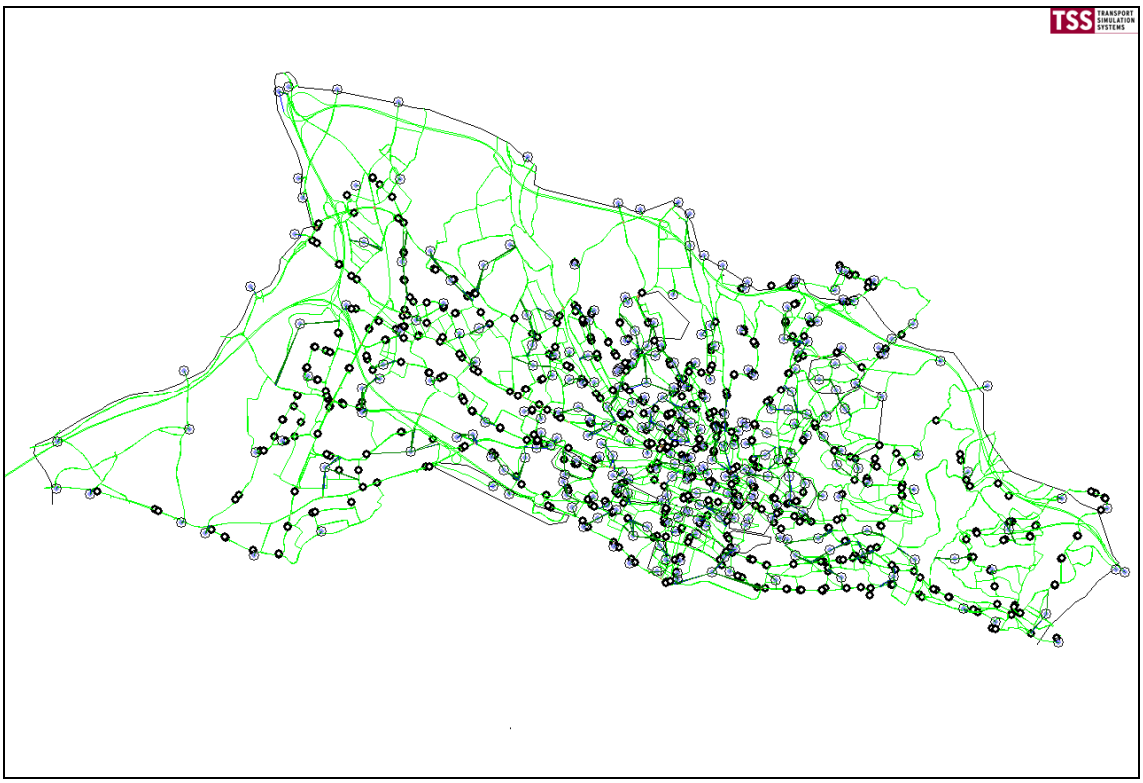


Fig. 3-33: Lausanne Agglomeration Network

3.5.4.1 Calibration

The calibration process was already finished for both networks, as they were the test bed in [1]. However, several simulations were carried out to ensure the consistency of experiments since experiments were carried out in a upgraded version of the simulator, specifically from Getram v4.2 to AIMSUN NG 5.1.

3.5.4.2 Parameters considered:

The basic parameters adopted for the microsimulation after calibration are shown in Table 3-14

Table 3-14: Basic parameters for Microsimulation

	Lausanne Centre	Lausanne Agglomeration
Reaction Time (sec)	0.75	0.75
Reaction Time at Stop (sec)	1.30	1.30
Simulation Step (sec)	0.75	0.75

Values in Table 3-14 are considered as average. In Table 3-15 route choice is described extensively. Parameters are individually explained below, with information taken from [2].

Table 3-15: Route choice for Lausanne networks

	Lausanne Centre	Lausanne Agglomeration
Cycle (minutes)	3	3
Number of Intervals	3	3
Default Cost function	AIMSUN 5.0	AIMSUN 5.0
Capacity Weight	2	2
User Defined Cost	0	0
Provide Travel Times	Yes	No
Route Choice Model	C-Logit	C-Logit
Dynamic (En-route Assignment)	No	Yes
Initial K Shortest Paths	1	1
Max number to keep	10	10
Maximum number of routes	10	5
Scale factor	15	6
Beta factor	0.15	0.15
Gamma factor	1	1
Use Past Costs from Replication	Not used	Not used

Cycle defines the interval between shortest path calculations and 3 minutes represents a fast path recalculation, necessary in large and complex networks.

Number of intervals defines the number of previous cycles over which travel times are collected for shortest path calculation.

Default cost function was set for 5.0 (networks had been previously calibrated with this cost function).

Capacity weight defined the weight of the link capacity on the total cost of routes in comparison to travel times. This value was set based on calibration.

User defined costs is set to zero since no user defined costs are used.

Travel times determine that travel times are stored in addition to cost and path lengths for display within the environment, and have no impact on the dynamic traffic assignment. Their storage has been omitted in the larger network to reduce memory usage.

The C-Logit route choice model was applied in both networks, a model which is normally recommended for urban networks to reduce route choice oscillations caused by the assignment to overlapping paths.

The scale factor reflects the tendency of drivers to select lower cost paths. It is recommended by TSS that the scale factor is set to equal to 60 divided by the average travel time in minutes. Within the Lausanne centre, a relatively high scale factor of 15 indicates that drivers are more likely to select the lower cost route. In larger networks, drivers are likely to take wider variety of paths.

Beta was set to a low value of 0.15 based on calibration. Gamma is left at the default value of 1.

While a low number of initial shortest paths is present, a low cycle time and presence of a warm up period lead to early generation of many path alternative based on applied traffic, with both networks maintaining a maximum of 10 shortest path trees.

Within Lausanne centre, path selection takes place between the best 10 routes available, while on the larger network, path selection is between the best 5 routes.

Table 3-16: Vehicles characteristics for Lausanne networks

	Lausanne Centre				Lausanne Agglomeration			
	Mean	Dev.	Min	Max	Mean	Dev.	Min	Max
Length (metres)	4	0.2	3.5	4.5	4	0	4	4
Width (metres)	2	0	2	2	2	0	2	2
Max. Desired Speed (km/h)	100	20	80	150	100	20	80	150
Max. Acceleration (ms^{-2})	2.8	0.3	2	4	2.8	0	2.8	2.8
Normal Deceleration(ms^{-2})	4	0.3	2.5	5.5	4	0	4	4
Max Deceleration(ms^{-2})	8	1	6	10	8	0	8	8
Speed Acceptance	1.05	0.05	0.95	1.2	1	0	1	1
Min Distance between vehicles (metres)	1	0	1	1	1	0	1	1
Give Way Time (seconds)	15	15	10	30	15	15	10	30

Equipped Vehicles (%)	To be fixed for each experiment	
PCU s	1	0

3.5.5 Experiment description

The experiments were tested and run in the two networks shown. The length of simulation for the larger network was reduced to 1 hour due to computational requirements and to limit the amount of Floating Car Data generated and transferred to both FCD systems.

3.5.5.1 Definition of Experiments

One of the main points of interest of FCD systems is their adaptation to get right values under a changing flow - and therefore travel times situations. To have an idea of the evolution of travel times over the duration of the experiment and in front of variations in flow, in one of the networks a time dependent set of matrices was chosen. This adds variability in traffic conditions.

The detailed definition of the experiments is shown below:

- **Lausanne centre**
  - Length of Simulation: 4 hours
  - Warm up period: 10 minutes
  - Percentage and evolution of Demand matrices used:
    1st hour: 50% of Peak Hour Demand

2nd hour: 80% of Peak Hour Demand
3rd hour: 70% of Peak Hour Demand
4th hour: 50% of Peak Hour Demand

- **Lausanne agglomeration**
  - Length of Simulation: 1 hour
  - Warm up period: 15 minutes
  - Percentage and evolution of Demand matrices used:
    1st hour: 50% of Peak Hour Demand

3.5.5.2 Tests and consistency of experiments

All two networks were tested in order to check that traffic flow was as expected. To do this selected macro indicators were considered, specifically the density stability. This was also used to estimate the warm-up period. In the small network, a 10 minutes warm-up was sufficient, while for the larger network it was extended to 15 minutes. During the warm-up no data was collected from probe vehicles. Density stability was assured for the length of the experiments, as proof of consistency in traffic performance.

An example of density stability during an experiment is shown in Fig. 3-34

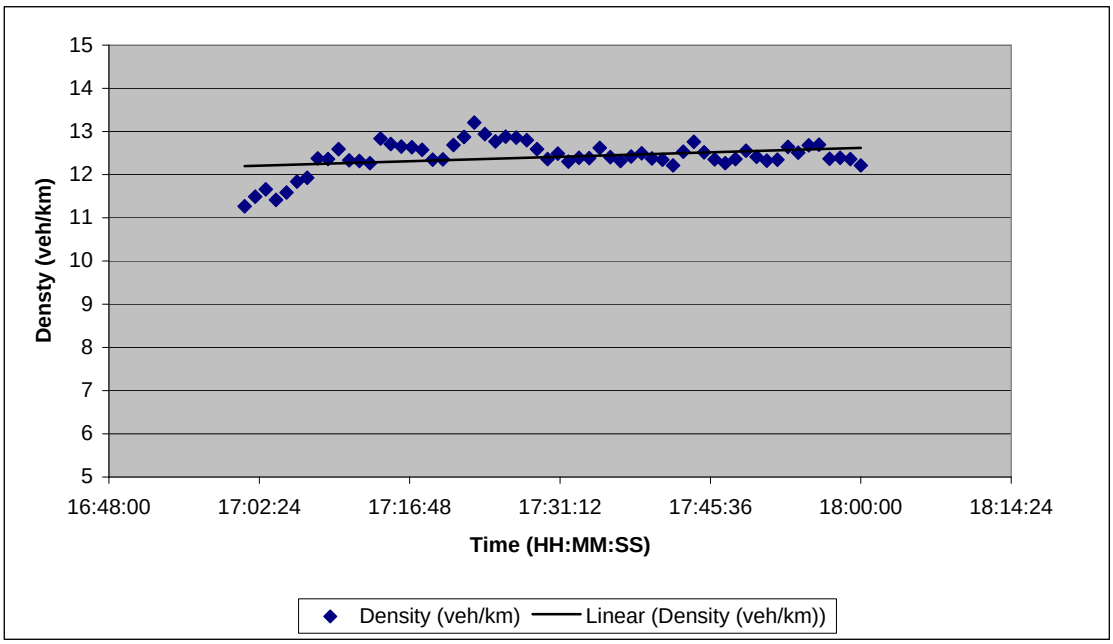


Fig. 3-34: Example of stability with 80% flow in Lausanne centre network

3.5.6 Description of studied scenarios

The microsimulation tool AIMSUN was used to simulate the movements of all vehicles in a network for various scenarios. The simulation results were stored in an AIMSUN database. Probe vehicles were randomly selected from the new vehicles entering the network with a probability equal to the equipment ratio for the current scenario. For each probe vehicle (probe) data were sampled every RI second, where RI is the reporting interval of the current scenario. These (probe) data correspond to data generated by a real floating car. Each data sample contains the following information: time, vehicle position, vehicle velocity. These samples constitute the input to the FCE and UFCS systems

respectively. There is one important difference, however; for the FCE system the identity of the probe vehicle that generated the data is also supplied.

Another difference between the systems is the way the vehicle position is represented. In the FCE system a map matching procedure is used and the position must be given in form of X and Y coordinates. For the UFCS system the position is represented as the distance from the beginning of the link.

The FCE and UFCS systems use the data to estimate the average travel time on each link. The averaging is done over an aggregation period. Using AIMSUN data for all vehicles exact average link travel times were computed. Hence, various error measures can then be computed for the FCE and UFCS travel time estimates respectively.
A flow-chart of the comparison procedure is shown in Fig. 3-35.

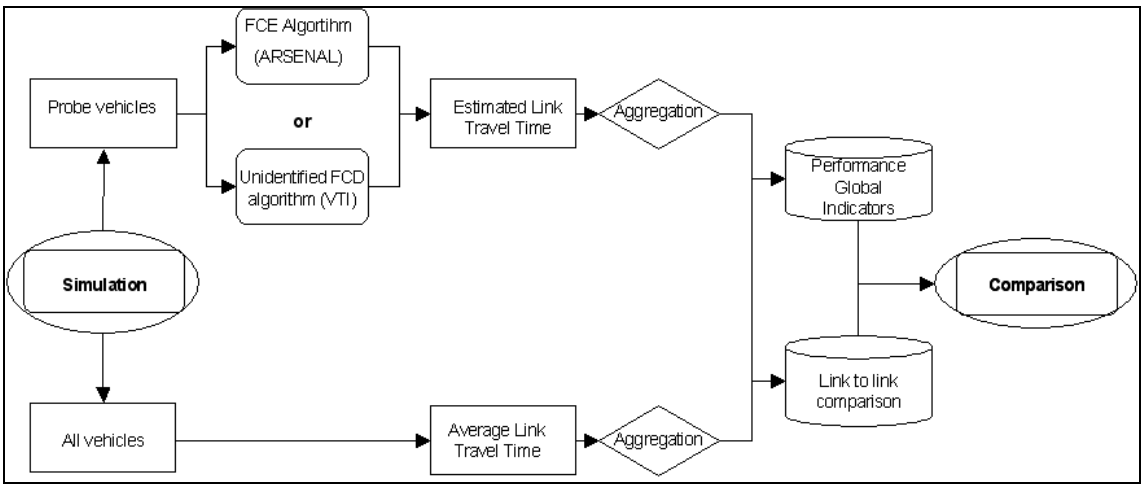


Fig. 3-35: Flow-chart for comparison process of simulated FCD

An important detail in the comparison procedure is the handling of missing values. Travel time estimates are only available for those link-periods where data is available. This is strictly true for the UFCS method. For the FCE method all intermediate links between two data reports will not be missing. This means that, in general, the UFCS method generates more missing values than the FCE method. For the FCE method map-matching and routing errors may occur, resulting in some missing values on otherwise complete paths.

There are at least two ways to deal with missing values:

- keep only those link-periods for which travel time estimates exist for both FCE and UFCS method
- extend results to all link-periods by using some substitute values.

An advantage using the first approach is that the comparison will be independent of any substitute values and describe the pure travel time estimates. A disadvantage is that when having few observations the number of link-periods to be compared will be small yielding large uncertainties in the comparison. Also, the number of links that are being compared will vary with the scenarios.

Two different kind of substitution values have been used: the temporal combination method described in Section 3.3.4.6, and travel time computation using free flow speed. Note that the temporal combination method is actually more than just a substitution method. It not only fills in the missing values but affects also the existing travel time estimates.

Due to some practical problems implementing the interface between AIMSUN and the FCE and limited time, only the scenarios described in Table 3-17 have been computed. For the Lausanne Centre network, comparison was done without substitute values in most cases. Free flow speed was used to

compute missing travel times, when travel times were computed on paths and the spatial distribution of the errors were studied. In Lausanne Agglomeration network, only results of the UFCS method were computed, and hence no comparison could be made. The temporal combination was done using $\alpha = 0.3$. (This method presupposes that there exists a value for the first aggregation period of the experiment. If this is also missingthe a priori function corresponding to sparse traffic condition was used.) In all cases UFCS method was implemented using the θ-function method with harmonic averaging of individual observations over an aggregation period. The low velocity parameter ω was set to 0.5.

The output was intended to study the influence of some of the basic parameters in FCD systems:

- Aggregation period
- Equipment ratio (sometimes referred to as penetration rate)
- Reporting interval

Table 3-17 shows the different parameters used in each experiment.

For comparison of both methods an aggregation interval of 15 minutes was used, as this was the standard setting for the Floating Car Evaluator (FCE). Different aggregation periods were compared in the Unidentified Floating Car System (UFCS), in order to evaluate the known influence of the aggregation period.

Wheras calculation of travel times with the UFCS was carried out on basis of the microsimulation tool AIMSUN, microsimulation output had to be converted to a different coordinate sysystem (WGS 84) used by the FCE. As several tests were necessary to complete the interface, only experiments for Lausanne centre were run with the FCE.

Table 3-17: Experiments definition

	Lausanne Centre		**Lausanne Agglomeration**
	FCE	UFCS	UFCS
Equipment ratio	0.5%,1%, 2%, 5%,10%		0.5%,1%, 2%, 5%,10 %
Reporting period (sec)	15,30,60		15,30,60
Aggregation period (min)	15	5,10,15	5,10,15
Experiment length (hours)	**4**		**1**
Substitute values	free flow speed		temporal combination

3.6 Comparison of travel time results from identified and unidentified floating car data

3.6.1 Methodology and task organisation

As outlined in chapter 3.4 and 3.5, microsimulation was used to create a series of floating car data. The output of microsimulation was transmitted to the Floating Car Evaluator (FCE) developed by arsenal research and the Unidentified Floating Car System (UFCS) developed by VTI, in order to provide them with simulated but realistic floating car data for an urban network.

After calcluating travel times using the two different approaches, average travel times or average speeds for each network link were returned to TSS for comparing both results with true average link travel time calculated from microsimulation output. Microsimulation experiments carried out in a previous study on floating car data for navigation systems were taken into consideration for drawing conclusions from the results of the current experiment [1].

3.6.2 Performance Indicators

To evaluate the influence of the parameters which play a role in the ELT estimation process, it was necessary to define performance indicators.

For this purpose, the term link-period was defined. It refers to a link *a* during a specific aggregation period τ. In the same way, path-periods *p*,τ were defined for paths, referring to a path *p* during a specific aggregation period τ.

Three main indicators were used to compare both FCD systems.

3.6.2.1 Mean Square Link Travel Time Error

This first indicator (Mean Square Link Travel Time Error (MSLE) can be formulated as the estimation error of travel times over all the A links of the network and over the T time periods considered:

$$MSLE = \frac{\sum_{a=1}^{A}\sum_{\tau=1}^{T}\varepsilon_{a,\tau}^{2}}{A \times T} \qquad (3.18)$$

where $\varepsilon_{a,\tau}$ is the difference between the average and the estimated link travel times for each link and aggregation period, as it is shown in equation (3.1)

$$ALT_{a,\tau} = ELT_{a,\tau} + \varepsilon_{a,\tau} \qquad (3.19)$$

Some of the links are not used by simulated vehicles during a certain period of time. Only the arc-periods having a population larger than zero are considered in this calculation.

This indicator is the most frequently used to evaluate the performance of travel time estimation by FCD. It is based on the principle that the average of the population of LT is the true value and that the average of the sample must be close to it.

3.6.2.2 Mean Link Travel Time Error

The Mean Absolute Link Travel Time Error (MALE) also measures the absolute deviations between ELT and ALT, with less emphasis on large deviations. So differences between MSLE and MALE are due to large deviations, which get more weight in the calculation of the MSLE.

$$MALE = \frac{\sum_{a=1}^{A}\sum_{\tau=1}^{T}\left|\varepsilon_{a,\tau}\right|}{A*T} \tag{3.20}$$

The Mean Link Travel Time Error measures deviations between ELT and ALT, taking their sign into account. The MLE indicates whether ALT was overestimated or underestimated on average, pointing to systematic errors of travel time estimation.

$$MLE = \frac{\sum_{a=1}^{A}\sum_{\tau=1}^{T}\varepsilon_{a,\tau}}{A*T} \tag{3.21}$$

3.6.2.3 Average Individual Link Travel Time Error

The average of the population of LT should not be considered as the real value to estimate. It might be considered as a simplification or, more precisely, a modelling of the phenomena travel time for a section. Therefore the MSLE is comparing only two models (ELT vs. ALT), and it is required to quantify its aptitude for modelling the phenomena itself.

For this reason, the comparison between each individual vehicle LT for every section and its model (ELT) seems more appropriate. The Average Individual Travel Time Error (AILE) is defined as:

$$AILE = \frac{\sum_{a=1}^{A}\sum_{\tau=1}^{T}\sum_{n=1}^{N_{a,\tau}}\frac{\left|LT_a^n - ELT_{a,\tau}\right|}{ELT_{a,\tau}}}{\sum_{a=1}^{A}\sum_{\tau=1}^{T}N_{a,\tau}} \tag{3.22}$$

So it measures the average relative error in the estimation of LT of a vehicle.

3.6.2.4 Average Individual Path Travel Time Error

The reflection that leads to the AILE definition can be applied to paths as well. The average individual path travel time error (AIPE) is defined as:

$$AIPE = \frac{\sum_{p=1}^{P}\sum_{\tau=1}^{T}\sum_{n=1}^{N}\left|\frac{PT_p^n - EBPT_{p,\tau}}{EBPT_{p,\tau}}\right|}{\sum_{p=1}^{P}\sum_{\tau=1}^{T}N_{p,\tau}} \tag{3.23}$$

3.6.3 Comparison of results for Lausanne Centre

3.6.3.1 Preliminary considerations

As mentioned in chapter 3.4 and 3.5, methods for filling missing data in some estimated travel times is essential for floating car systems, as this is one of the most important challenges when a typical phenomenon appears: the lack of data from any floating vehicles for a specific link and aggregation period.

In order to have a complete and realistic comparison, all link quality indicators (MSLE, MALE, AILE) were only calculated for link-periods, for which floating cars were available in each scenario. So the important task of replacing missing values has not been considered for this comparison.
For the path indicator AIPE, free-flow travel time was used to substitute all missing values on the actual path of the simulated vehicle. The amount of missing values was expected to have a main effect on travel time estimation for the whole network, especially in situations with low equipment ratio and long reporting intervals.

3.6.3.2 Parametric analysis

This analysis allows estimating the influence that each important parameter in FCD systems plays in the accuracy of results, thus:

- To compare the different indicators previously defined and determine their success in evaluating travel time estimation methods.
- To show the influence of the different parameters under the different estimation methods, evaluating their combined effects.
- To obtain the optimal combination of parameters, according to the chosen indicator and for each equipment ratio

3.6.3.2.1 MSLE

As it was previously remarked, this indicator is the most frequently used in FCD travel time estimation research. Its results for different scenarios are shown in Table 3-18 and Fig. 3-36.

MSLE values for the FCE method showed an elevated level. Although only link-periods with available FCD were considered, MSLE ranged between 200 and 400 s^2. In a previous study, estimated values ranged between 50 and 300 s^2 after substituting free-flow travel time for missing values (see Fig. 3-37). As MALE values based on the same absolute deviations ranged between 7 and 11 s (see Table 3-19 and Fig. 3-38), this elevated level of MSLE was caused by a few extremely high deviations. The cause for large deviations could be the influence of spatial imprecision and routing errors. As the simulation environment AIMSUN uses a simplified street network, coordinates were different from the real street network in the GIS data (with deviations between 10 to 30 m). These spatial differences can cause map-matching errors. Map-matching errors and additional routing errors have a larger effect, when reporting intervals are low. This effect has been frequently observed, when GPS tracking data were used. A more detailled analysis would be necessary in order to detect the specific causes.

For most scenarios, the FCE method had a lower MSLE for the same reporting interval than the UFCS method. The UFCS method exceeded the performance of the FCE method only for ER = 10% and RI = 15 sec. This could be explained by the fact that the UFCS used historical information or by harmonic averaging which effectively reduced erroneous estimates.

Table 3-18: Sensitivity of MSLE to equipment ratio and reporting interval

ER	RI	FCE	UFCS
0.5%	15	288.2	400.2
0.5%	30	307.2	1 427.3
0.5%	60	311.9	2 971.3
1.0%	15	278.6	409.0
1.0%	30	345.6	1 276.1
1.0%	60	365.5	2 711.4
2.0%	15	275.5	350.6
2.0%	30	334.6	1 120.0
2.0%	60	371.0	1 882.0
5.0%	15	258.9	338.2
5.0%	30	262.3	738.9
5.0%	60	320.6	1 439.2
10.0%	15	262.0	164.2
10.0%	30	238.9	442.7
10.0%	60	299.6	1 107.9

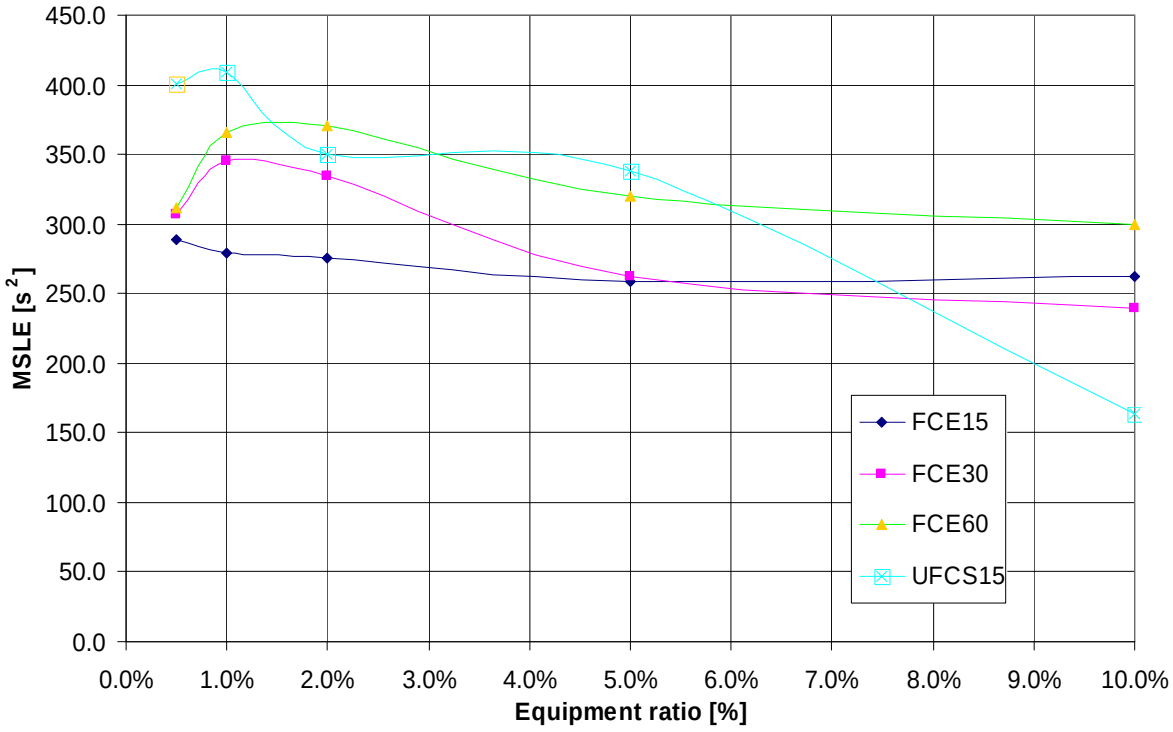


Fig. 3-36: Sensitivity of MSLE to equipment ratio and reporting interval

MLSE values for the UFCS method exhibited a high sensitivity to the reporting interval (see Table 3-18). This is a natural result since the error should depend much on the number of observations per link, which should grow in inverse proportion to the length of the reporting interval. In contrast, the

MSLE values show a rather small (but noticeable) sensitivity to the equipment ratio. An explanation for this is that additional probe vehicles often use other routes than the original vehicles and, therefore, the number of observations per link grows slowly with the equipment ratio until most of the links in the network are covered during an aggregation period. This is in accordance with the theoretical results presented in Fig. 3-37.

Only a very small sensitivity to reporting interval and nearly no sensitivity to equipment ratio was observed for the MALE values (see Table 3-19 and Fig. 3-38). For all scenarios, the UFCS method had a lower MALE than the FCE method for the same reporting interval.

Since the MSLE error measure is more sensitive than MALE to accidental large errors for some links, these results indicate that the UFCS method often (i.e. on many links) produces more exact result than the FCE method but for some links produce (much) worse results. The bad performance of the UFCS method on some links can easily be understood in light of its big vulnerability to low velocities. The FCE method should be more stable when low velocities are encountered. It may come as a surprise that the UFCS method on many links yields better results than the FCE method. An explanation for this is that the UFCS method does, in contrast to the present FCE method, effectively exploit much a priori information concerning the normal velocity distribution along the links.

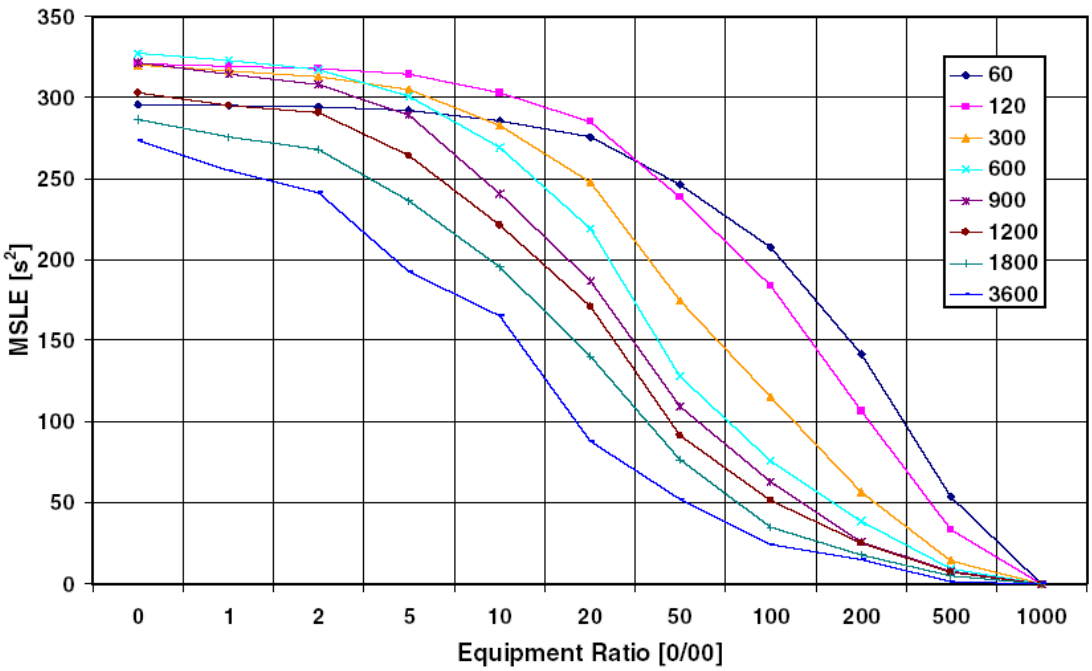


Fig. 3-37: Sensitivity of MSLE to equipment ratio and reporting interval (source [1])

Fig. 3-37 shows results from microsimulation for different aggregation periods and different equipment rates. It can clearly be seen that MSLE value usually decreases with increasing aggregation period, and that all curves converge to zero for 100% equipment ratio, as the entire population is considered in the sample. Trends for all aggregations display similar shapes but with increasing degree of steepness with decreasing aggregation period length, except for the 60 seconds case which has a slightly different behaviour.

Table 3-19: Sensitivity of MALE to equipment ratio and reporting interval

ER	RI	FCE	UFCS
0.5%	15	7.9	4.9
0.5%	30	8.6	6.5
0.5%	60	9.8	6.7
1.0%	15	7.9	5.2
1.0%	30	8.7	6.8
1.0%	60	10.4	9.2
2.0%	15	7.8	4.0
2.0%	30	8.6	5.8
2.0%	60	10.4	8.1
5.0%	15	7.6	3.7
5.0%	30	8.0	5.5
5.0%	60	10.1	6.4
10.0%	15	7.4	3.3
10.0%	30	7.6	4.4
10.0%	60	9.7	6.2

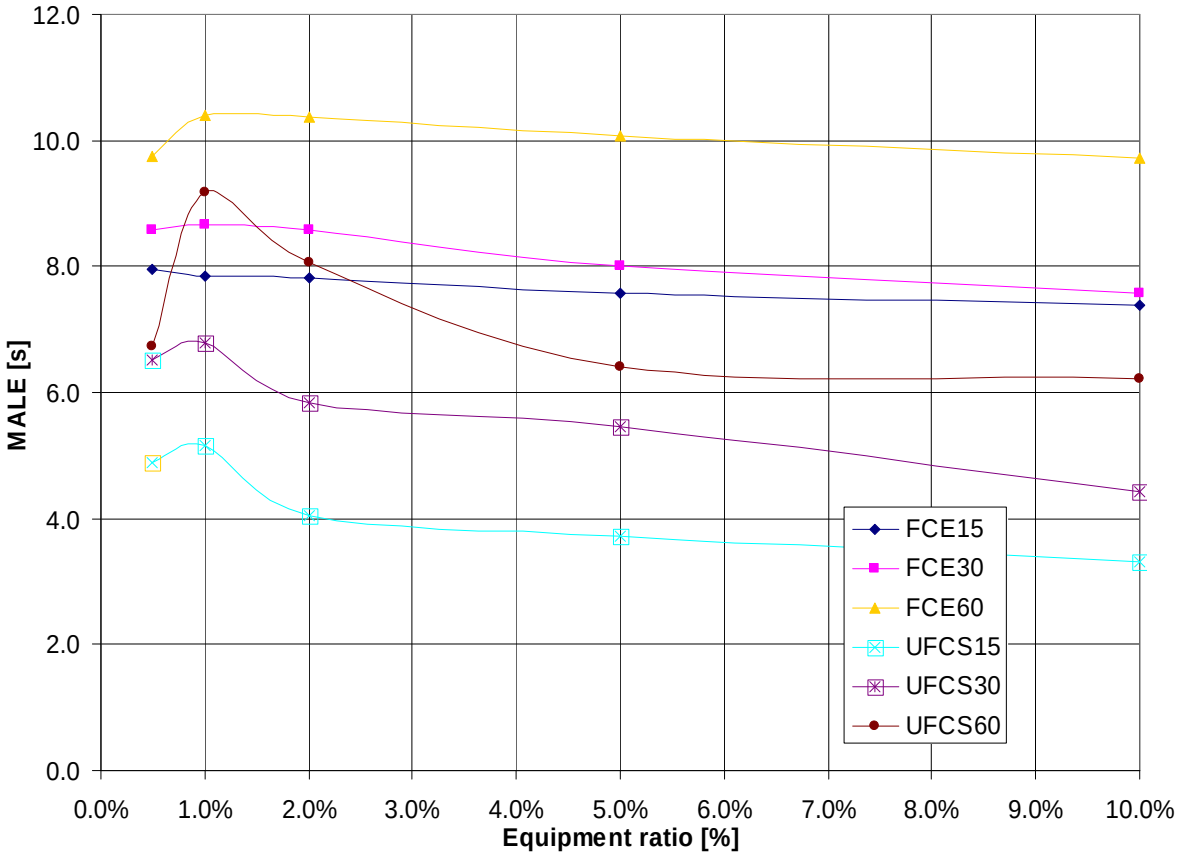


Fig. 3-38: Sensitivity of MALE to equipment ratio and reporting interval

3.6.3.2.2 Evolution of network average travel time

This indicator allows having a general idea of an average travel time for all sections. (see Fig. 3-39 and Fig. 3-40). It clearly showed temporal evolution of traffic, whose behaviour was defined by microsimulation scenarios (see chapter 3.5). Differences between average network travel times and estimated network travel times reflected systematic errors of travel time estimation.

ELT from the UFCS method overestimated average network travel times from microsimulation with AIMSUN. This overestimation is probably caused by the method s inability to deal with the low velocity problem , see Section 3.3.4.5. UFCS may for some links yield huge overestimates of the travel time, which may overestimate the overall average network travel time. Particularly when there is only one observation during a link-period, risk for overestimation is large since the damping effect of harmonic averaging is then not present. The overestimation is affected by the value of the parameter ω as well as the choice of substitute values. For a specific traffic condition, suitable choices of ω and substitute values (by calibration) may eliminate this overestimation. However, this balancing depends on the traffic conditions and it is not possible to apply it for varying conditions.

ELT from the FCE method underestimated average network travel times. However, their trend exhibited a clear correlation with the increasing and decreasing level of travel time simulated in the microsimulation scenarios.

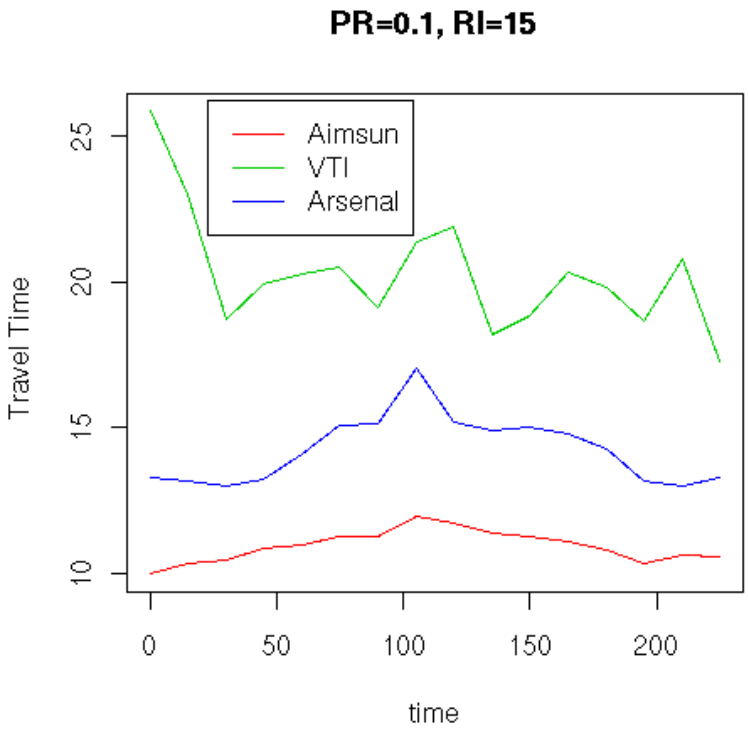


Fig. 3-39: Average network travel time for ER=10% and RI=15 sec

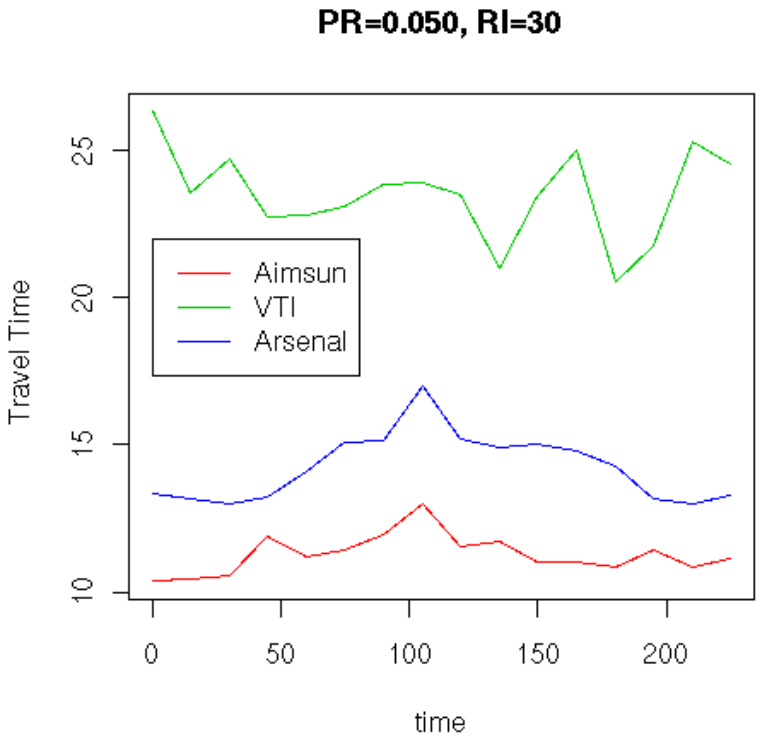


Fig. 3-40: Average network travel time for ER=5% and RI=30 sec

3.6.3.2.3 AILE (Average Individual Link Travel Time Error)

The usage of AILE provided more qualified results (see Table 3-20 and Fig. 3-42). AILE values from FCE and UFCS experiments ranged between 1.78 and 0.49 and showed a high sensitivity to different equipment ratios. .

As expected, and it is shown in previous experiments shown in Fig. 3-41, only experiments with ER over 5% give a good approximation to individual travel times. This is due to the elevated number of missing values that low ER generate. Trends for the different RI are similar for each method, showing better performances as RI increases.

Major differences can be observed between both methods, with a slight better performance for UFCS method in front of FCE. Best performance can be obtained with a 10% of ER and 15 seconds RI for UFCS. The major differences could be observed between reporting intervals and between methods. AILE values always decreased with increasing RI.

AILE confirmed that ELT values estimated by both FCD and UFCS were a good approximation to individual link travel times with high ER. The range of AILE values is a bit higher or similar to the range of AILE values observed in a previous approach, in which missing values were replaced by free-flow travel time.

Compared to the high variation of travel times on urban links, which are due to the influence of traffic lights, this can be considered as a satisfactory result, which should be a good starting point for estimating average link or path travel times.

In Fig. 3-41, the effect of aggregation period length observed in a previous study is shown. The error decreased with the length of the aggregation period, until the curves for all AP converged at an AP of 50 %. For higher equipment ratios an inverse trend could be observed. This is explained by the fact that replacement of missing values does no longer play an important role and a more refined discretisation offers better results.

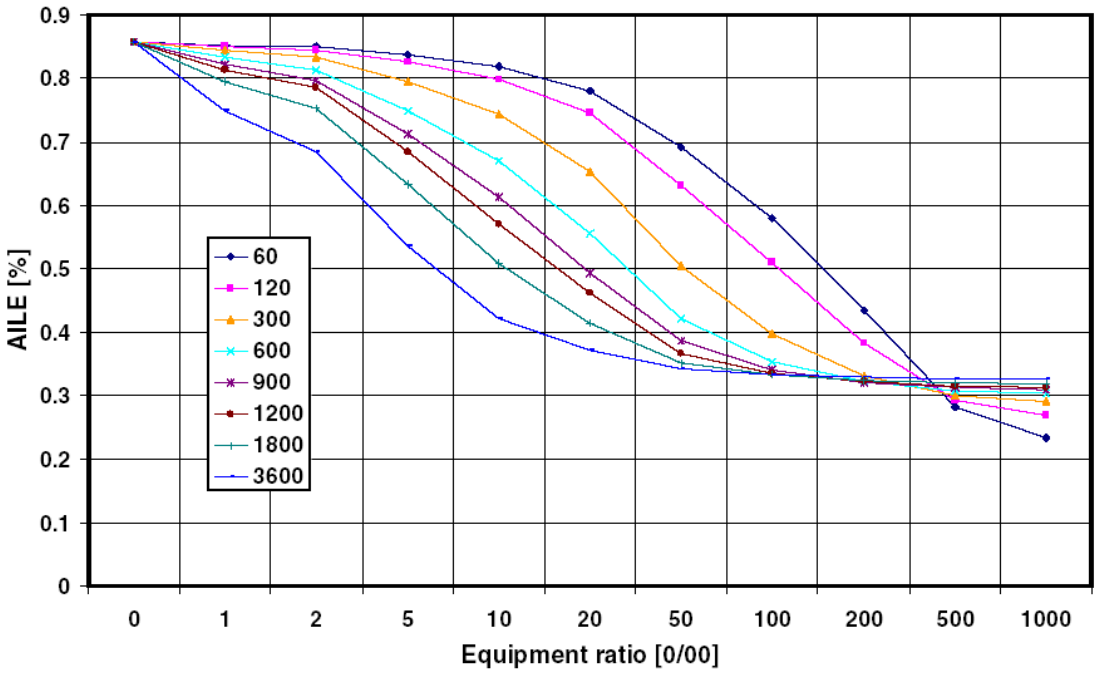


Fig. 3-41: Sensitivity of AILE to equipment ratio and reporting interval (source [1])

Table 3-20: AILE for UFCS and FCE

ER	RI	FCE	UFCS
0.5%	15	1.78	1.73
0.5%	30	1.77	1.83
0.5%	60	1.78	1.92
1.0%	15	1.60	1.46
1.0%	30	1.58	1.59
1.0%	60	1.61	1.74
2.0%	15	1.36	1.10
2.0%	30	1.34	1.26
2.0%	60	1.35	1.46
5.0%	15	1.14	0.68
5.0%	30	1.07	0.84
5.0%	60	1.05	1.04
10.0%	15	1.04	0.49
10.0%	30	0.97	0.62
10.0%	60	0.92	0.79

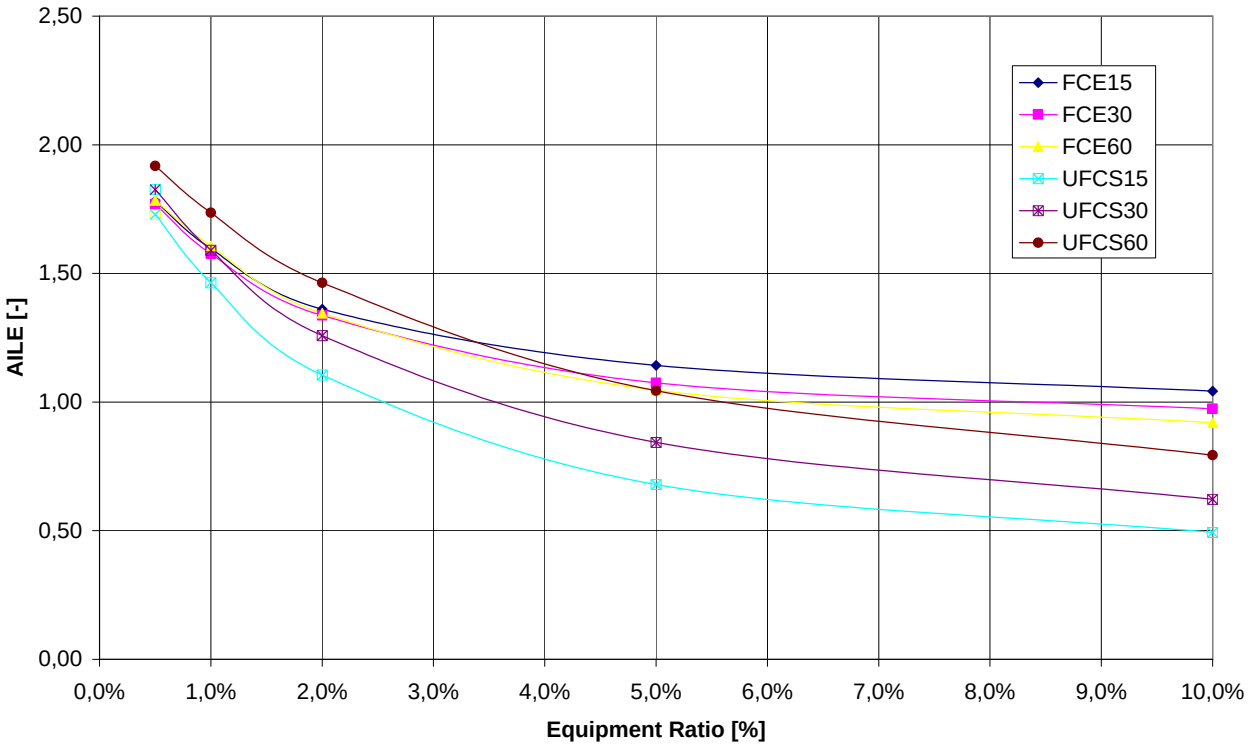


Fig. 3-42: Sensitivity of AILE to equipment ratio and reporting interval

3.6.3.2.4 AIPE (Average Individual Path Travel Time Error)

All results of the AIPE indicators are presented in Table 3-21 and Fig. 3-44. Both methods showed a high sensitivity to the equipment ratio up to an equipment ratio of 5 %. At this point, both methods decrease the sensitivity but keep improving their performances in estimating travel times. In all cases, results show lower levels of error, compared with AILE values. ,

This result was in the line of the observations in a previous study (see Fig. 3-43), where the relative error was significantly lower for paths than for links. This phenomenon can be explained by the fact that estimation errors for links do not accumulate along a path and that they may compensate each other (whereas the AILE takes into consideration the absolute value of errors). The dependence of travel times between links also plays a role in this context.

For lower ER, FCE and UFCS results are very similar, due to the high percentage of missing values which may be replaced with the substitution method, based in free flow travel times. Also due to this substitution method, an unexpected result may be explained, the less error shown for lower RI in FCE method.

The error of the AIPE in Fig. 3-43 decreased with the length of the aggregation period until the curves for all AP converged at an AP of 10 to 20% (lower than observed for AILE in Fig. 3-41). For higher equipment ratios, an inverse trend could be observed.

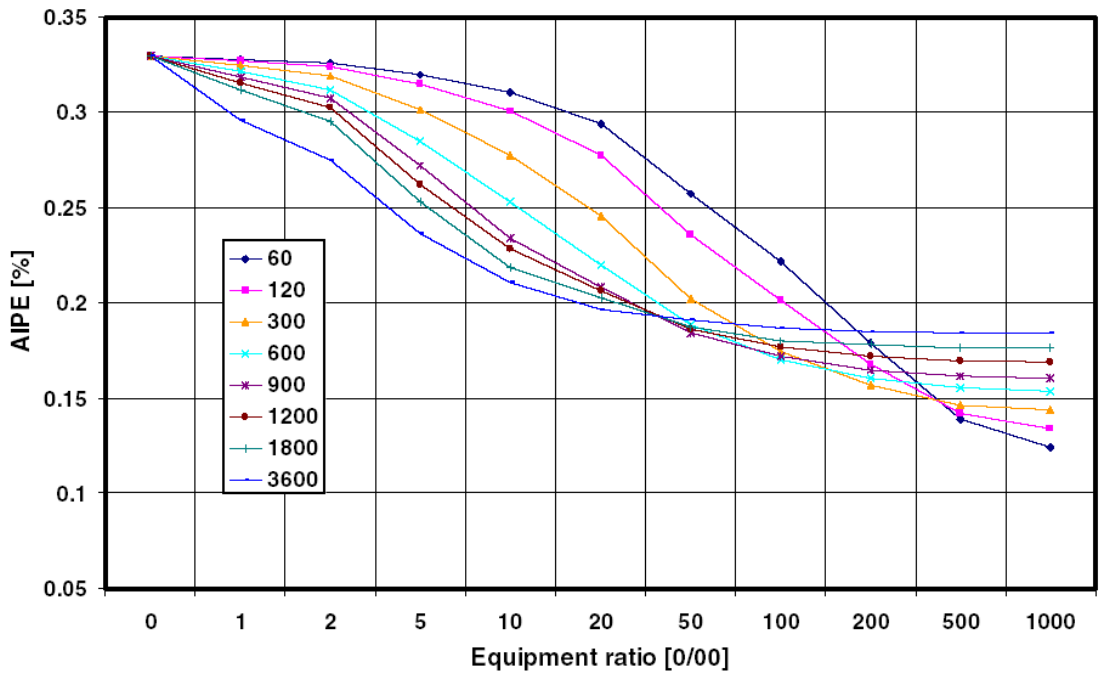


Fig. 3-43: Sensitivity of AIPE to equipment ratio and reporting interval (source [1])

Table 3-21: Sensitivity of AIPE to equipment ratio and reporting interval

ER	RI	FCE	UFCS
0.5%	15	0.96	1.00
0.5%	30	0.88	1.04
0.5%	60	0.88	1.12
1.0%	15	0.82	0.75
1.0%	30	0.73	0.80
1.0%	60	0.68	0.87
2.0%	15	0.70	0.53
2.0%	30	0.56	0.57
2.0%	60	0.50	0.66
5.0%	15	0.60	0.37
5.0%	30	0.43	0.39
5.0%	60	0.35	0.45
10.0%	15	0.57	0.30
10.0%	30	0.39	0.33
10.0%	60	0.29	0.38

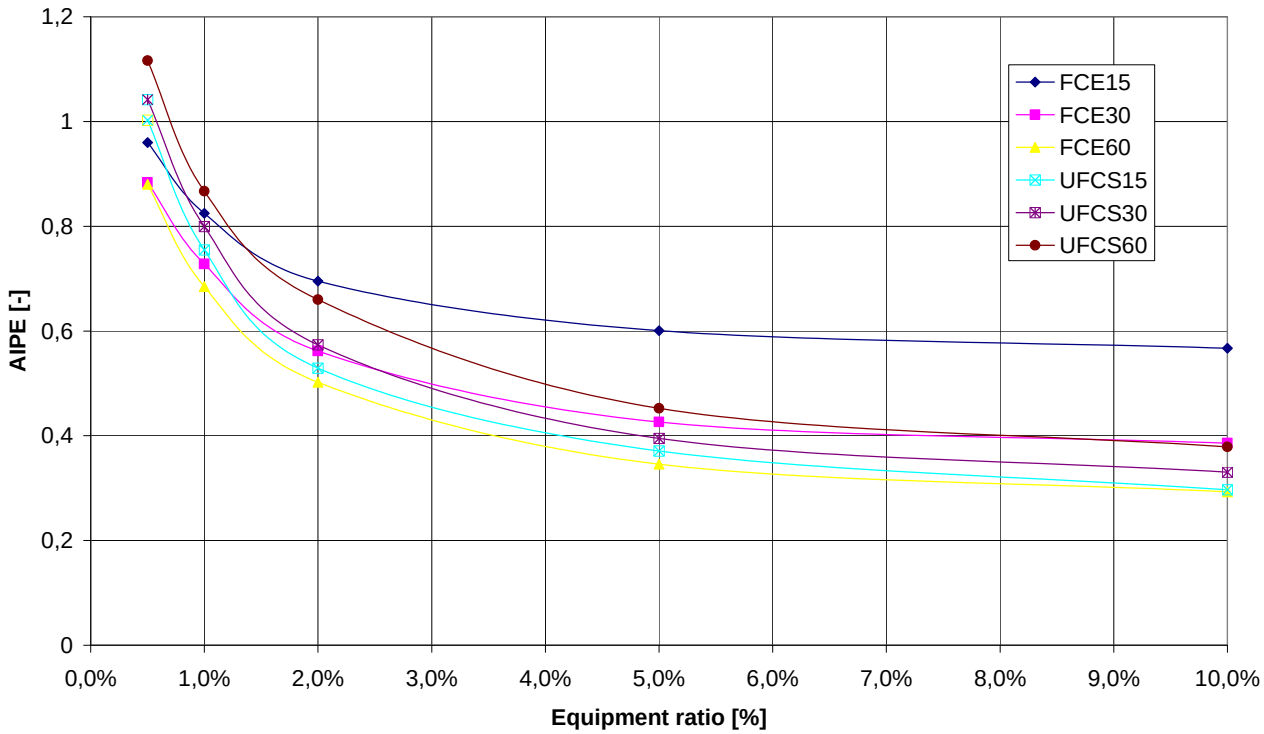


Fig. 3-44: Sensitivity of AIPE to equipment ratio and reporting interval

3.6.3.2.5 Spatial distribution of link travel time errors

In order to investigate the spatial distribution of travel time estimation errors, absolute differences between simulated average link travel time ALT and estimated ELT for each link were plotted on a map (by means of an AIMSUN interface). The comparison is based on the error ranges shown in Fig. 3-45.

The main considerations about this graphical representation are:

- links coloured in green, corresponding to differences from 0 to 1 second, can be appreciated as excellent travel time estimations especially for long sections
- yellow and orange sections show larger differences - from 1.5 to 10 seconds - typical for urban networks, showing the influence of the intersections, especially signalized ones
- red sections show larger deviations between ALT and ELT
- sections marked in black are those with high differences, which indicate anomalies in travel time estimation. These values must be discarded, as they do not show realistic travel times and should be treated separately.

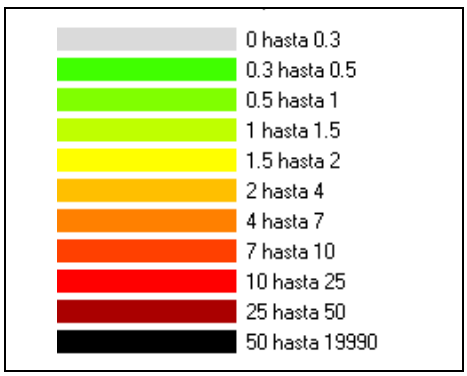
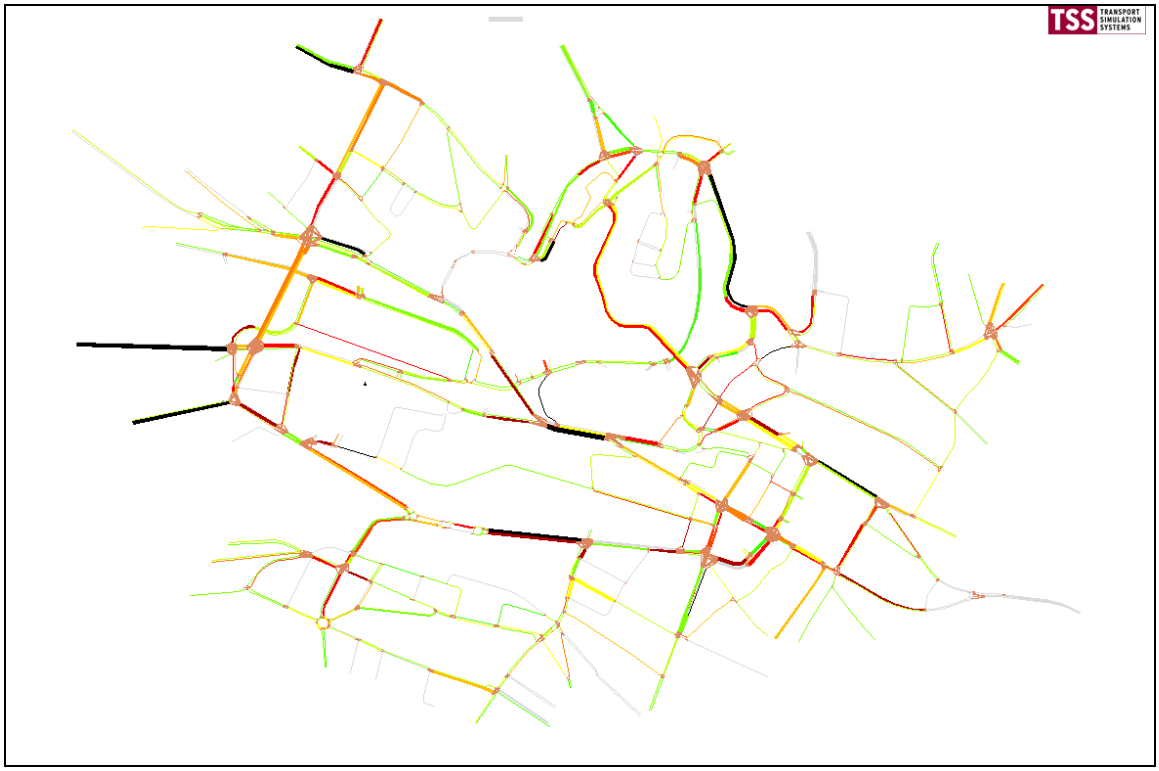


Fig. 3-45: Error ranges for absolute travel time difference [s]

Graphical presentations could only be produced for the UFCS method (see Fig. 3-46) and showed variation in error depending on the equipment ratio, traffic conditions (dense or sparse) and reporting interval. Eight situations were chosen.

For the UFCS method a higher equipment ratio combined with a lower reporting interval (e.g. ER=5%; RI=30 s) seemed to perform better than a lower equipment ratio combined with a higher reporting interval (e.g. ER=2%; RI=15 s). This may appear contradictory to the conclusions drawn in section 3.6.3.2.1, where the length of the reporting interval played a more important role than the equipment ratio. In this comparison, however, the situation is quite different, since substitute values have been used in order to obtain values for all links. As more values were available for each link-period, erroneous or erroneous substitute values had a smaller effect on the resulting ELT.

Although it is difficult to draw final conclusions from these selected results, they indicate that certain links cause problems in many scenarios and should be investigated in greater detail.



UFCS ER=0,5%;RI=60 sec. Sparse situation

Fig. 3.46 (a): Link-to-link differences between simulated and estimated (UFCS)



UFCS ER=10% ; RI=15 sec. Sparse situation

Fig. 3.46 (b): Link-to-link differences between simulated and estimated (UFCS)



UFCS. PR=5% ; RI=30 sec. Sparse situation

Fig. 3.46 (c): Link-to-link differences between simulated and estimated (UFCS)



UFCS. ER=2% ; RI=15 sec. Sparse situation

Fig. 3.46 (d): Link-to-link differences between simulated and estimated (UFCS)



UFCS ER=0,5%;RI=60 sec. Dense situation

Fig. 3.46 (e): Link-to-link differences between simulated and estimated (UFCS)



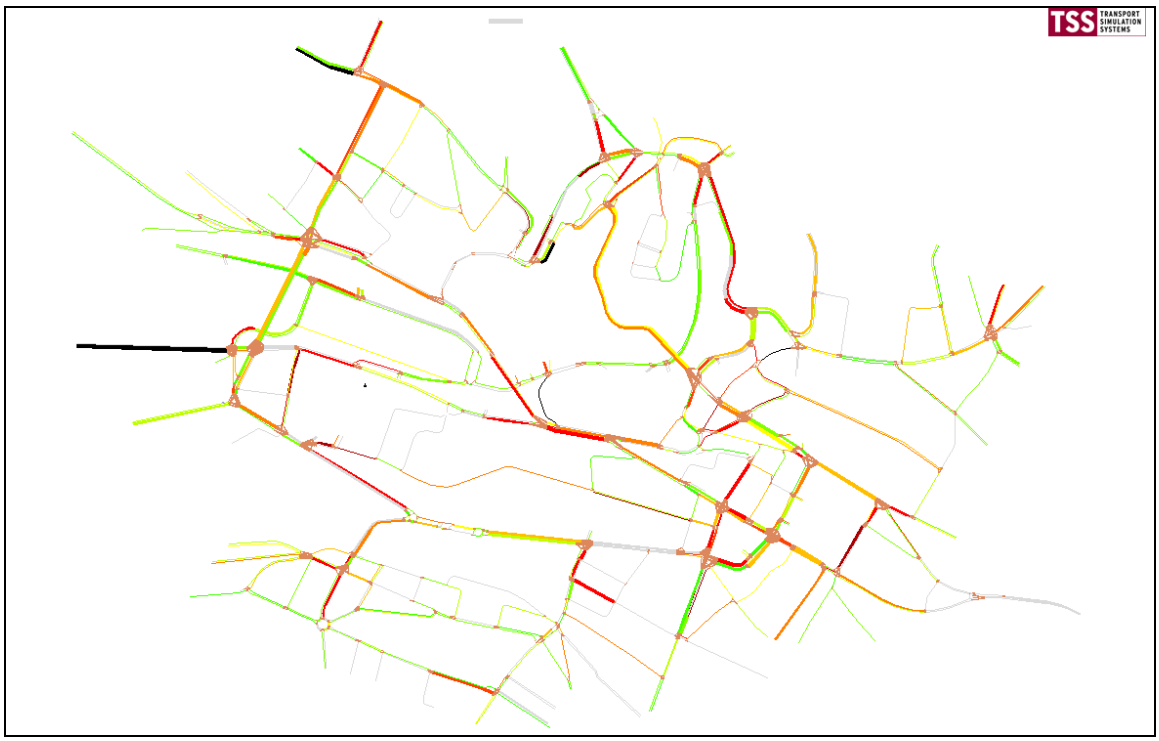
UFCS. ER=10% ; RI=15 sec. Dense situation

Fig. 3.46 (f): Link-to-link differences between simulated and estimated (UFCS)



UFCS. ER=5% ; RI=30 sec. Dense situation

Fig. 3.46 (g): Link-to-link differences between simulated and estimated (UFCS)



UFCS. ER=2% ; RI=15 sec. Dense situation

Fig. 3.46 (h): Link-to-link differences between simulated and estimated (UFCS)

Fig. 3-46: Link-to-link differences between simulated and estimated (UFCS)

3.6.4 Results comparison for Lausanne agglomeration

3.6.4.1 Preliminary considerations

The original considerations taken for simulation, collection of results and data processing were exactly the same as for the smaller Lausanne Centre model.

Travel time calculations for Lausanne agglomeration was intended to be carried out with the FCE method and the UFCS method in parallel. For the FCE method, GIS data and vehicle data had to be converted to a different coordinate system. Due to initial problems with the conversion of coordinates, the first results for Lausanne agglomeration mainly reflected problems with coordinate transformation. Because of time constraints, it has not been possible to repeat the large run of the FCE with corrected coordinates. Therefore no results with the FCE method for Lausanne agglomeration can be presented in this report.

3.6.4.2 Parametric analysis

As in Lausanne Centre simulation, the same indicators were used to get tendencies and to establish relationships between the different reporting intervals and equipment ratios. As the UFCS method used substitute values for missing values that were derived from previous observations results reflected the importance of correct substitute values for the UFCS method.

3.6.4.2.1 MSLE

Table 3-22 shows the values for MSLE for UFCS methods and all experiments.

MSLE value for UFCS, shown in Fig. 3-47, were generally lower than the ones for Lausanne Centre (except for a RI of 15 s). The reason for these results were related to traffic conditions for this experiment (low and constant traffic density during the 1-hour-simulation) together with a substitution method based partially on a sparse situation. As explained in Lausanne centre network, single abnormal values can affect results in several aggregation period in cases with long RI and low ER situations, if no new data is collected for a link. This may be the explanation for the large changes in trend for a RI of 60 s.

3.6.4.2.2 AILE (Average Individual Link Travel Time Error)

AILE values only showed a slight sensitivity to equipment ratio, although the tendency is clearly decreasing for all RI (Table 3-23 and Fig. 3-48). Results were very similar to the lowest values obtained in a previous study (see Fig. 3-41**Fel! Hittar inte referenskälla.**). Compared to Lausanne centre, values were lower for the UFCS method.

3.6.4.2.3 AIPE (Average Individual Path Travel Time Error)

A comparison of the AIPE indicators for each experiment is shown in table Table 3-24 and Fig. 3-49. AIPE values are always lower than AILE values for the same RI. As discussed in a previous study [1], this phenomenon can be explained by the fact that estimation errors for links do not accumulate along a path and by dependence of travel times between links.

Table 3-22: Sensitivity of MSLE to equipment ratio and reporting interval

ER	RI	UFCS
0.50%	15	342.4
0.50%	30	389.6
0.50%	60	532.8
1.00%	15	353.9
1.00%	30	563.7
1.00%	60	1066.7
2.00%	15	423.0
2.00%	30	418.5
2.00%	60	614.4
5.00%	15	311.5
5.00%	30	516.0
5.00%	60	991.5
10.00%	15	287.6
10.00%	30	426.2
10.00%	60	616.5

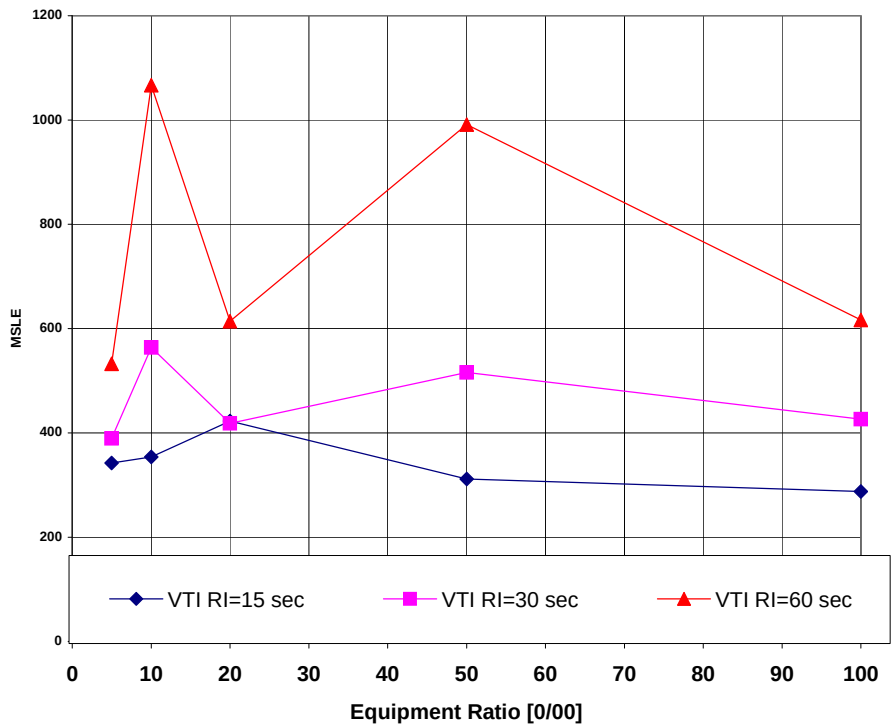


Fig. 3-47: Sensitivity of MSLE to equipment ratio and reporting interval

Table 3-23: Sensitivity of AILE to equipment ratio and reporting interval

ER	RI	UFCS
10.0%	15	03470
10.0%	30	0.3491
10.0%	60	0.3479
5.0%	15	0.3510
5.0%	30	0.3532
5.0%	60	0.3535
2.0%	15	0.3551
2.0%	30	0.3543
2.0%	60	0.3566
1.0%	15	0.3684
1.0%	30	0.3635
1.0%	60	0.3649
0.5%	15	0.3752
0.5%	30	0.3754
0.5%	60	0.3739

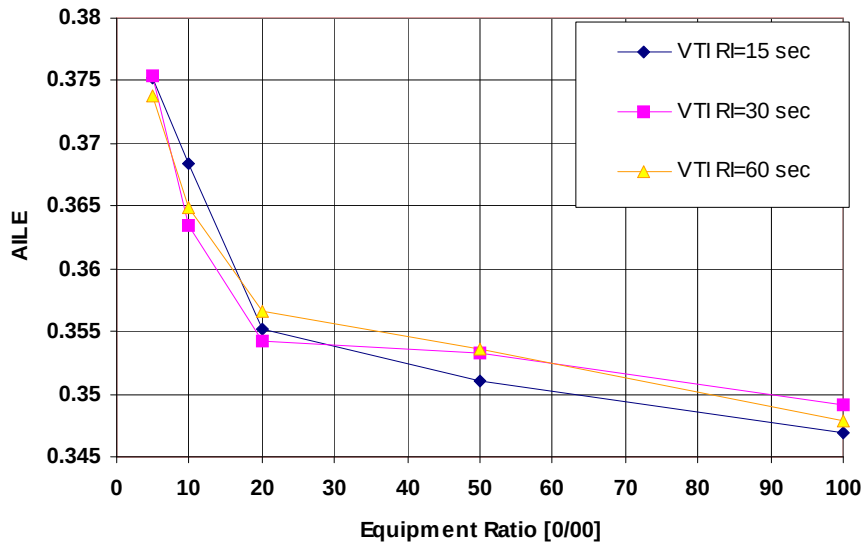


Fig. 3-48: Sensitivity of AILE to equipment ratio and reporting interval

Table 3-24: Sensitivity of AIPE to equipment ratio and reporting interval

ER	RI	UFCS
10.0%	15	0,1641
10.0%	30	0,1591
10.0%	60	0,1573
5.0%	15	0,1618
5.0%	30	0,1569
5.0%	60	0,1591
2.0%	15	0,1604
2.0%	30	0,1556
2.0%	60	0,1631
1.0%	15	0,1643
1.0%	30	0,1659
1.0%	60	0,1719
0.5%	15	0,1720
0.5%	30	0,1712
0.5%	60	0,1705

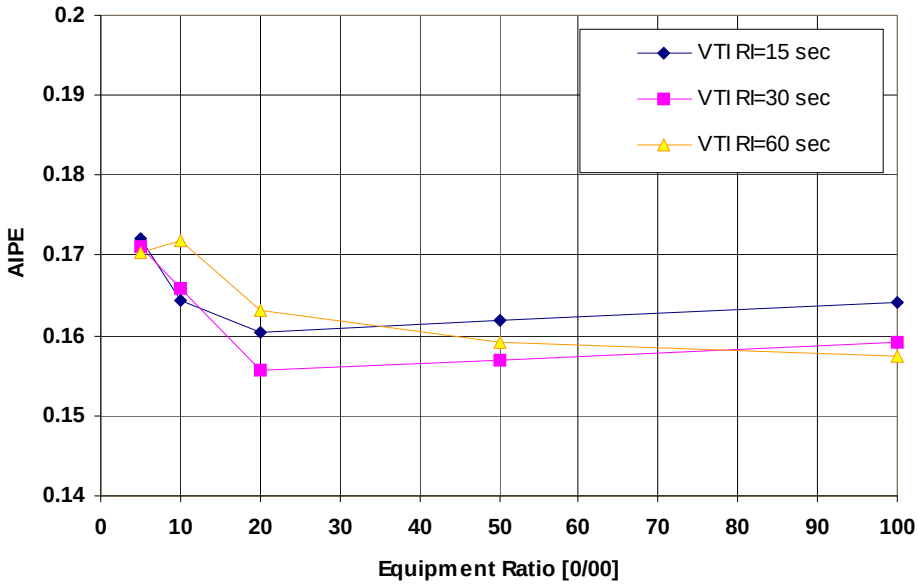


Fig. 3-49: Sensitivity of AIPE to equipment ratio and reporting interval

3.6.4.2.4 Aggregation period

Experiments with different aggregation periods were tested with the UFCS method as well. Error was described by the MSLE indicator, shown in Table 3-25. Error rates increased, as aggregation period decreased. This error was especially high for a 5 min aggregation period.

Table 3-25: MSLE for 5, 10 & 15 min AP (for UFCS)

PR	RI (min)	15 min	10 min		5 min	
		MSLE UFCS	MSLE UFCS	diff	MSLE UFCS	MSLE UFCS
0.50%	15	342,4	382,0	11,56%	342,4	382,0
0.50%	30	389,6	515,5	32,29%	389,6	515,5
0.50%	60	532,8	600,8	12,77%	532,8	600,8
1.00%	15	353,9	420,6	18,84%	353,9	420,6
1.00%	30	563,7	631,0	11,93%	563,7	631,0
1.00%	60	1066,7	1165,0	9,21%	1066,7	1165,0
2.00%	15	423,0	367,1	-13,21%	423,0	367,1
2.00%	30	418,5	455,0	8,74%	418,5	455,0
2.00%	60	614,4	801,4	30,43%	614,4	801,4
5.00%	15	311,5	420,0	34,82%	311,5	420,0
5.00%	30	516,0	420,0	-18,61%	516,0	420,0
5.00%	60	991,5	1054,7	6,37%	991,5	1054,7
10.00%	15	287,6	349,9	21,65%	287,6	349,9
10.00%	30	426,2	605,8	42,16%	426,2	605,8
10.00%	60	616,5	765,6	24,18%	616,5	765,6

3.6.5 Conclusions

A comparative study of the performance of two different link travel time estimation techniques, FCE and UFCS has been presented. The comparison was essentially based on microsimulated data for a network covering Lausanne Centre, modelling a four hour time period with varying traffic congestion. Both algorithms were fed with random samples of floating car data, each sample consisting of time, position and velocity. The FCE method also required the identity of the vehicle, in sharp contrast to the UFCS method. Various scenarios have been studied, mainly describing different equipment ratios (the share of probe vehicles) and different reporting interval (time period between two data samples from one floating car).

From the point of view of the simulation, it has been tested as a useful tool to evaluate new floating car methods. One of the most important aspects to be considered is the geometry of the modelled network. The results show some links which might not be suitable to estimate travel times with enough accuracy using floating car methods. These are for example, too short links in signalized intersections. So taking it into consideration, results show the consistency of the comparison between the two methods, showing in the simulated scenario some of the hypothesis about minimal equipment ratio or reporting interval.

Some very clear conclusions can be drawn from the comparisons. An evident difference between the methods is that UFCS results strongly depend on the reporting intervals while FCE results are less sensitive to them (see example Table 3-18). This is consistent with intuition since for UFCS all data samples contribute independently of the vehicle that generated them and therefore the error should depend much on the number of observations per link. For FCE, however, an additional data sample from a probe vehicle does not necessarily improve on the precision of a particular link travel time estimate.

Another difference between the methods is that the UFCS method has comparatively larger MSLE errors and smaller MALE errors than FCE (see Tables Table 3-18 and Table 3-19). This result indicates that the UFCS method often (i.e. for many link-periods) produces more exact results than the FCE method, but it produces (much) worse results for some links. The bad performance of the UFCS method on some links can easily be understood in light of its big sensitivity to the low velocity

problem (see Section 3.3.4.5). The FCE method is less vulnerable to low velocities and hence produce more stable results.

A further difference between the methods is that FCE method seems to produce better estimates of the overall network travel time than UFCS (see Fig. 3-39 and Fig. 3-40). A probable explanation for the rather large deviation for the UFCS method is again its inability to handle the low velocities often occurring in congested traffic. The UFCS method often overestimates the link travel time and this bias causes an overestimation of the total network travel time. Although this bias depends on the choice of the parameters of the UFCS method, this is an important weakness of the method. The network travel time error for the FCE method is much smaller. With the FCE method, it was possible to observe the simulated increase and decrease of average network travel times during a simulation period of 4 hours. The results indicate, however, that FCE tends to systematically overestimate the network travel time.

For both methods the errors seem to vary rather little with the equipment ratio at least for equipment ratios less than 5%. This is in accordance with theory (see e.g. Fig. 3-37). An explanation for this is that, for low equipment ratios, an additional probe vehicle often uses other routes than the original vehicles and, therefore, the number of observations per link grows slowly with the equipment ratio until most of the links in the network are covered during an aggregation period.

Intuitively, considering the fact that the FCE method takes advantage of the identity of the probe vehicles, it could be expected that the FCE method would outperform the UFCS method. However, Table 3-19 reveals that the MALE error is smaller for UFCS method than for FCE method for all Lausanne Centre scenarios. This means that the error for the UFCS method is smaller than the error for the FCE method for most link-periods. An explanation for this somewhat surprising result could be that the UFCS method, in contrast to the present FCE method, effectively exploits huge amounts of a priori information concerning the normal velocity distribution along the links. The use of the a priori information is evidently a powerful technique, which in many cases compensates for the lack of knowledge of vehicle identity. It could be a fruitful approach to try to combine a priori information with the FCE method. And this fact shows the importance of substitution methods to complete that information not provided by floating vehicles for some links or time intervals.

Concerning the UFCS method, the low velocity problem is a fundamental impediment for major improvements. We believe that to overcome this problem we need additional data sources or combine the method with other methods.

In addition to the parameters that have been investigated in this study, appropriate methods for replacing missing values have been shown to have a great impact on results. A large amount of missing values is inevitable under practical conditions (i.e. equipment ratios below 5 %). Although microsimulation provides a good basis for investigation, it has not been possible to carry out comparisons concerning the influence of several replacement methods. Possible approaches based on historical data have been described in chapter 3.2 on identified floating cars, in chapter 3.3 on unidentified floating cars and in a previous study [1].

Spatial imprecision is an important topic for travel time applications using GPS data. Although spatial imprecision of position measurement has been quantified, impact on travel time estimation has rarely been analysed and quantified. For the UFCS method some analysis were made in Section 3.3.5.4. The results indicate that the method was insensitive for spatial imprecision (at least as long as the disturbed position belong to the correct link. The effect of the map matching algorithm used in the FCE system has not been analysed.

3.7 High Resolution Floating Car Data

3.7.1 Introduction

3.7.1.1 High resolution floating car

A floating car is a vehicle that regularly collects vehicle data, such as position and velocity. It is understood that these data are transmitted to a central unit, either in raw or in refined form. A high resolution floating car is a vehicle that collects data with short enough sampling periods to contain high frequency components of data. The meaning of the expression high frequency in this context is not strictly defined but it should include the range of frequencies that embraces the normal velocity fluctuations of a traffic stream. Extreme events such as sudden breaking are not of central interest here. Likewise, very low frequency components of the velocity and the average velocity will also be disregarded.

3.7.1.2 Potential for safety applications

The special interest in high frequency components is based upon the assumption that the fluctuations ought, somehow, to contain information about the current traffic state and, hence, in some sense say something about the level of safety. Intuitively, it is tempting to believe that in dense flows the drivers ought to be more inclined to make faster adjustments towards the ideal or preferred distance between the vehicles than would be the case in more sparse flows. Consequently, the fluctuations in velocity ought to have different characteristics in dense and sparse flows. If this is the case, high resolution data might contain additional information (besides what can be deduced from the average velocity) about the current traffic state.

It would be of particular interest to relate high frequency data to the flow when the traffic stream is unstable but has not yet collapsed. A relation of this kind would be very useful for predicting collapses in the traffic stream. Once the traffic stream has already collapsed the average velocity is probably a more reliable indicator of the traffic state.

In order to investigate such hypothetical relations data are needed. One can hardly expect to find data that directly link the high resolution fluctuations with traffic safety parameters. Instead, the aim must be to relate the high resolution data to traffic state parameters such as flow or density. This presupposes a simultaneous collection of high resolution data and traffic state data.

3.7.1.3 Purpose of the present study

The purpose of the present study (subtask 4.2.3) has been to explore the possibility to exploit high frequency fluctuations in velocity to determine the traffic state and safety conditions. In the first place, it has to be established whether such fluctuations exist. If so, the correlation between these fluctuations and parameters describing the traffic state should be investigated.

3.7.2 Test bed

3.7.2.1 Test bed used for this study

For this study it has been necessary to collect data from a test bed where stationary detectors are distributed along an entire road section. One such test bed exists at Essingeleden, the main arterial road through Stockholm. Since 2004 a system of radar stations has been in operative use here. VTI has for a long time regularly been carrying out floating car measurements along this section. Thus, both floating car data (FCD) and traffic state data are available.

3.7.2.2 Description of the study area

In Fig. 3-50 a picture of (part of) the 12 km road segment used as testbed for this study is shown. There are 43 radar stations along this road segment, i.e., approximately 300 meters distance between them. The number of lanes is for most parts three or four, but along some segments only two. The northern end of the area is situated a few km west from the city centre of Stockholm.

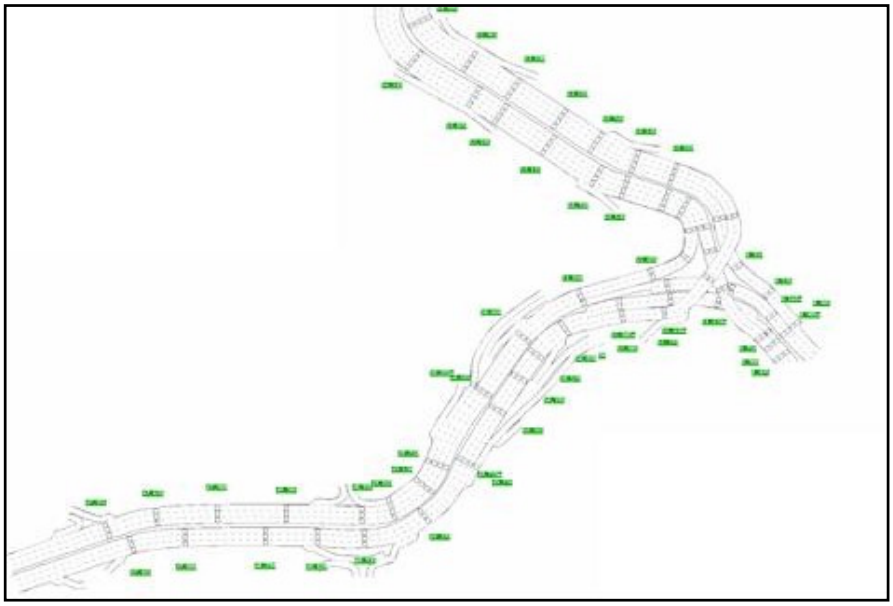


Fig. 3-50: Map describing the southern 8 km part of the test bed

3.7.2.3 Data collected

Two types of data have been collected simultaneously:

- Traffic state data from the radar stations
- Floating car measurements with a specially equipped vehicle from VTI.

From the radar stations traffic flow and velocity data are obtained aggregated to one-minute period for each lane. The floating car collects driving distances with a sampling frequency of 2 Hz. It has been estimated that this frequency is sufficient to capture the phenomena that are of interest here. From the driving distances the velocity is computed by using a backward difference scheme. Also, the starting time (and position) for every run is registered, making it possible to synchronize with the radar data.

Data for one day, 2004-12-07, has been used in the analysis. During this day 15 runs in each direction where performed with the floating car.

3.7.2.4 Preparation of data

A radar station provides information about the traffic state for an arbitrary time point in the place where it is situated. It is necessary to match the FCD with radar data for exactly those time points when the floating car passes the radar station. From the floating car the velocity of the vehicle and from the radar stations the flow of the traffic streams are obtained. There are some issues to consider when the matching is done:

1. Radar data is aggregated to one minute periods while FCD are in principle instantaneous.
2. Radar data refer to certain positions while FCD are spread out continuously along the whole road segment.
3. Radar data yields the flow in each lane but the lane in which the floating car is running might be unknown (or at least uncertain).

4. The computation of densities might be difficult in those segments where there are on- or off-ramps.

The first problem is handled by matching all FCD to the current minute flow. One negative consequence of this is that the FCD can be connected to flows occurring after the floating car has already passed. Also there is the question whether the flows should be computed for longer time periods, such as 2, 3 or even 5 minutes, thereby yielding more stable flows.

The next problem is handled by dividing the road segment into a number of sub segments by choosing the middle points between the stations as delimiters. In this way FCD will always be matched against the radar station that is closest to the floating car for each moment. This requires that the starting position of the vehicle is known. Alternatively, flow data from the two nearest radar stations could be weighted together by using linear interpolation, thereby avoiding the discontinuities in flow data that otherwise appear.

The third problem has been handled by using the average flow per lane instead of the total flow.

Computation of densities (for a specific subsegment) is often done by using flow data from the end points of the subsegment. This approach becomes difficult to apply when the segment contains on- or off-ramps. If FCD are available it is convenient to use the average velocity (between two radar stations) of the floating car to translate from flow to density. Notice that the density is then computed from a mixture of radar data and FCD.

3.7.3 Investigation of the relation between high resolution floating data and the traffic state

3.7.3.1 Method

The primary goal in this study is to investigate if there is any correlation between the high frequency components and the density or flow of the traffic stream. Of particular interest is the case where the flow is still unstable and large enough that the formation of a queue can soon be anticipated.

The quantification or characterization of high frequency components could be done in a number of different ways. For instance:

1. Using a simple measure of the speed variation, such as the absolute variation, or the mean square deviation from a local mean.
2. Analysing the frequency content of velocity data, e.g., wavelet analysis or Fourier analysis.
3. Counting the number of sign changes (in the acceleration curve) per time unit.

In the following subsections some of these methods will be applied in the quest for a relation between high resolution data and traffic flow or density.

3.7.3.2 Using the mean square deviation from a local velocity trend

A simple measure of the velocity fluctuations is the mean square deviation from an average of a fixed number of recent observations. This measure is, however, somewhat misleading when there is a steady drop in velocity during some time. Therefore, instead of the average of recent values we use least square method to estimate the linear trend of recent observations and compute mean square deviation from this trend estimate. In this report this measure is called Deviation.

In Fig. 3-51 four examples of speed profiles (Speed) and deviation from moving average trend (Deviation 10) are shown. A moving average of 10 samples has been used, i.e., over the five previous seconds. Also plotted are the total flow (FlowTot) over all lanes in the same direction as the floating car and the average flow per lane (FlowPerLane). All quantities are plotted against the driving distance (meters) run by the floating car. The first plot (Morning_1_643_Northwards) represents a run in the morning started at 6:43 am, the next plot a run at lunch hour and the last two in the afternoon.

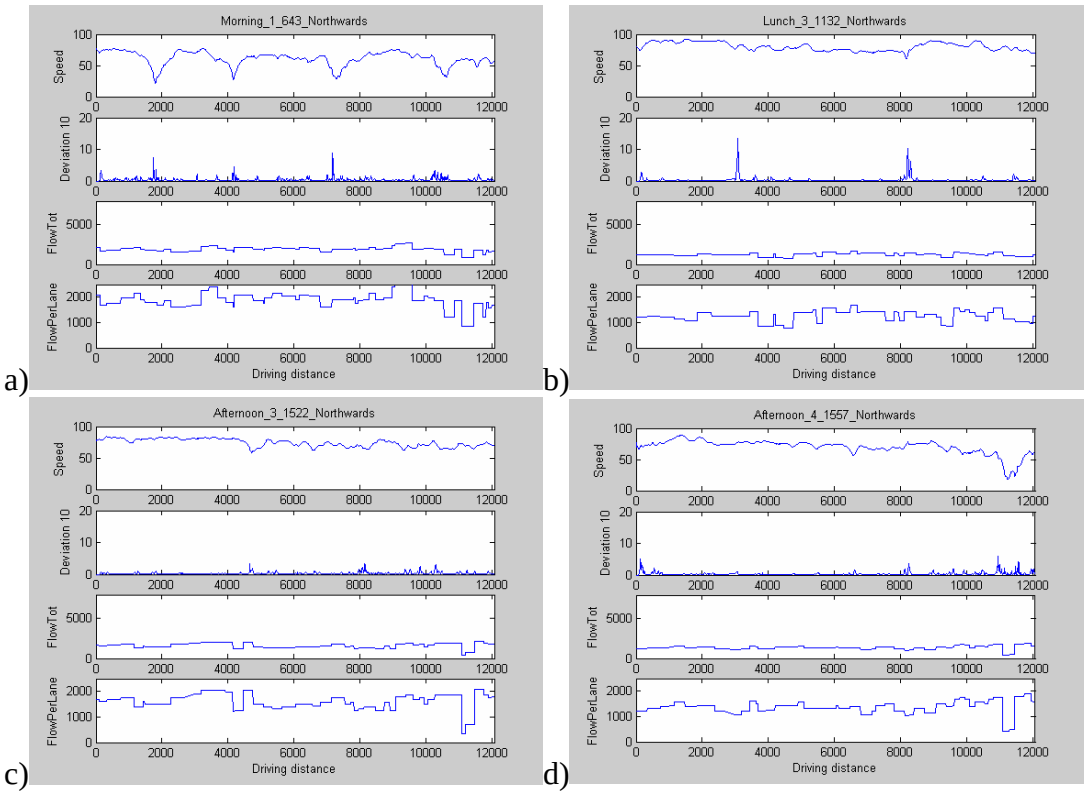


Fig. 3-51: Examples of speed profiles and associated traffic state parameters

The flow curves are piecewise constant since flow data are obtained from the closest radar station in each moment.

Longitudinal studies of this kind give rise to some problems not encountered in the ordinary case. Usually, flow is considered and measured at only one location. But here, matters are more complicated since the flow and other circumstances vary from one location to another. Which one of the parameters total flow and flow per lane best represent the ordinary quantity flow ? The total flow varies along the road section because of on- and off-ramp traffic along the section. The flow per lane varies because the total flow varies but also because the number of lanes varies along the road section. Sudden drops and increases in flow per lane are often caused by changes in the number of lanes. A corresponding difficulty arises for the density.

From a) and d) one can suspect that there is a connection between the Deviation and low velocities. Whenever there is a (strong) dip in the velocity there seem to be some response in the Deviation. However, from b) it is seen that the Deviation may be large even when the velocity remains on a fairly even level.

Although the relation between the velocity level and the Deviation may be of some interest, the primary goal here is to find connections between the Deviation and the flow or density of the traffic stream. In Table 3-26 correlations between these quantities are presented, based on all 15 northward runs with the floating car. Each individual observation from the vehicle is included.

Table 3-26: Correlations between various parameters. All observations included

	Deviation	*Speed*	*FlowTot*	*FlowPerLane*	*Density*
Deviation	1.00	-0.25	-0.08	-0.08	0.22
Speed	-0.25	1.00	0.27	0.27	-0.79
FlowTot	-0.08	0.27	1.00	1.00	-0.10
FlowPerLane	-0.08	0.27	1.00	1.00	-0.10
Density	0.22	-0.79	-0.10	-0.10	1.00

There is very little correlation between the Deviation and the flow. As could be guessed from Fig. 3-51 some correlation exists between Deviation and speed. Consequently, since the correlation between speed and density is rather large, there is also some correlation between Deviation and density.

The variation between individual observations could be expected to be large. Averaging data over specific radar stations and time periods (one minute) could be expected to increase the correlations between Deviation and the traffic state parameters. This is confirmed by the results in Table 3-27.

Table 3-27: Correlations between various parameters. Aggregations to minutes and radar stations.

	Deviation	*Speed*	*FlowPerLane*	*Density*
Deviation	1.00	-0.42	-0.07	0.35
Speed	-0.42	1.00	0.11	-0.80
FlowPerLane	-0.07	0.11	1.00	0.05
Density	0.35	-0.80	0.05	1.00

Is there any correlation for the case when the (average) speed exceeds 50 km/h? From Table 3-28 it can be seen that the correlations between Deviation and speed (or density) is significantly lower than in the unrestricted case.

Table 3-28: Correlations between various parameters for speeds exceeding 50 km/h. Aggregations to minutes and radar stations

	Deviation	*Speed*	*FlowPerLane*	*Density*
Deviation	1.00	-0.24	0.03	0.15
Speed	-0.24	1.00	-0.31	-0.64
FlowPerLane	0.03	-0.31	1.00	0.81
Density	0.15	-0.64	0.81	1.00

Some attempts have been made to study situations immediately preceding an actual decrease in velocity. For example, the velocity 30 seconds ahead (or alternatively, a fixed number of meters ahead) is computed and all cases where the current velocity is larger than 60 km/h and the ahead velocity is less than 40 km/h are extracted and correlations are computed. From the few such experiments done no significant differences to the results in the tables above can be reported. This indicates that high resolution data do not seem to contain information useful for detecting unstable flows and predicting collapses in the traffic stream.

3.7.3.3 Counting the number of sign changes per time unit

The number of sign changes in acceleration per time unit is another measure or characteristic of the fluctuations. This measure will be called SignChanges here.

In Table 3-29 the correlation between SignChanges and traffic state parameters is shown. The correlation is in general very low.

Table 3-29: Correlations between the number of sign changes per unit time and various traffic state parameters

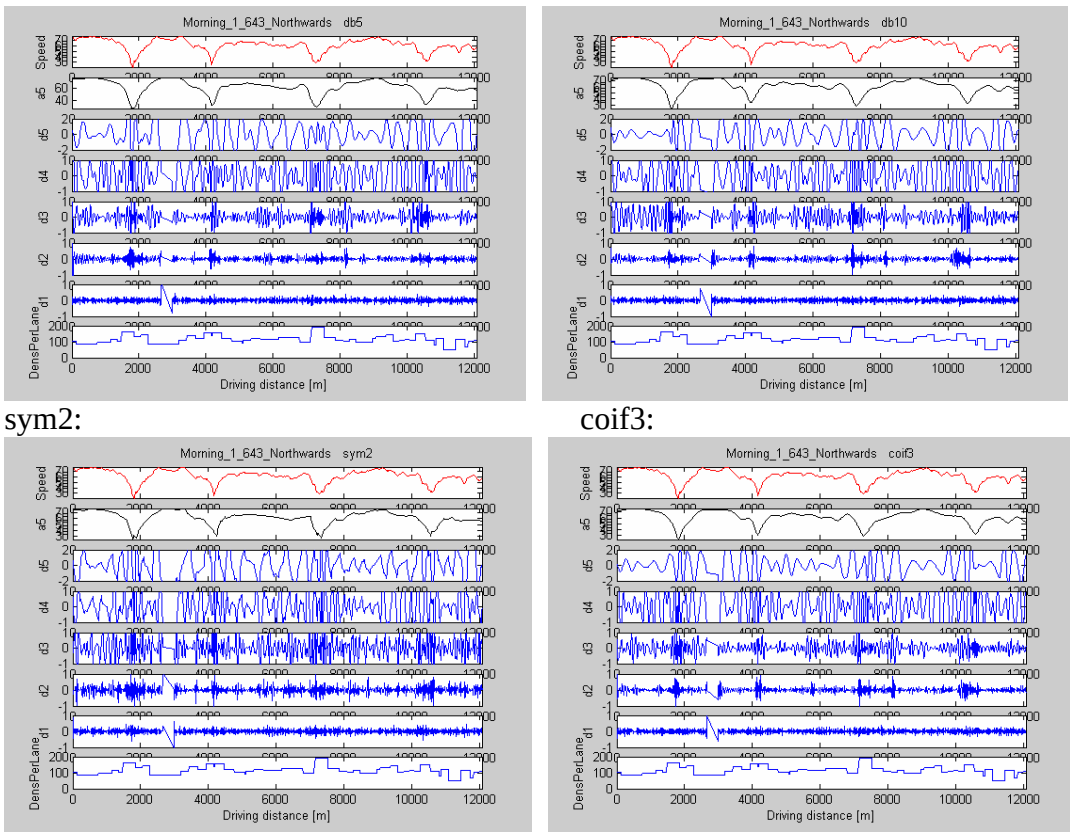
		Correlations			
Restrictions	*Aggregation*	*SignChanges*	*Speed*	*FlowPerLane*	*Density*
No	no	1	0.03	-0.06	-0.02
No	yes	1	0.09	-0.06	-0.06
speed > 50	no	1	0.10	-0.03	-0.08
speed > 50	yes	1	0.12	-0.05	-0.10

3.7.3.4 Using wavelet decompositions

A practical tool for analysing the frequency components of a function such as the velocity profile is wavelet decomposition. An advantage of wavelets as compared to ordinary Fourier analysis is that information about locality is maintained. The frequency components can vary from one position to another while, in contrast, Fourier components are the same on the whole interval considered.

There exist a great number of different wavelet decompositions, each characterized by the choice of a basis function and the depth of the decomposition (i.e. the level or frequency at which to stop the decomposition). In Fig. 3-52 an example is shown with four different wavelet decompositions of the same velocity profile (morning at 6:43 am).

Wavelet = db5: db10:



sym2: coif3:

Fig. 3-52: Examples of wavelet decompositions of a speed profile

In the upper graph (Speed) the original velocity profile is plotted and in the lower graph (DensPerLane) density per lane. In d5-d1 wavelet components of different frequencies are plotted, d5 is the

component with longest wavelength and d1 with shortest wavelength. The components of most interest here are probably d2 and d3, with wavelengths in the range of approximately 2-6 seconds.
One can see that the dips in the velocity curve are clearly reflected in a corresponding increase in the amplitude of some of the frequency components. The absolute value of a wavelet component could be used as a measure or characteristic of the velocity fluctuations (of a certain wavelength). This measure will be called ComponentSize.

In Table 3-30 correlations between ComponentSize and some traffic state parameters are shown. Evidently the correlations are smaller for speeds exceeding 50 km/h. Also aggregations yield higher correlations than individual values. Use of the component d2 instead of d3 yields approximately the same result.

Table 3-30: Correlations between wavelet components of velocity profiles and some traffic state parameters for various wavelet types, speed conditions and aggregation types.

			Correlations			
Wavelet method	*Restrictions*	*Aggregation*	*ComponentSize*	*Speed*	*FlowPerLane*	*Density*
db5, d3	no	no	1.00	-0.26	-0.08	0.20
db5, d2	no	no	1.00	-0.20	-0.11	0.21
db5, d3	speed > 50	no	1.00	-0.11	-0.01	0.05
db5, d3	no	yes	1.00	-0.35	-0.05	0.25
db5, d3	speed > 50	yes	1.00	-0.17	0.03	0.08

Similar attempts were done in section 3.7.3.2 to study the situations occurring before the velocity collapses No increase in correlation has been detected.

3.7.4 Conclusion

All results in the previous sections point in the same direction. The relation between high resolution data and the traffic state is rather weak if at all existing for the Stockholm data. It does not seem to be a feasible approach to try to derive non-collapsed traffic state data from high resolution data. A possible explanation for this may be that the drivers have a good view far ahead and adapt there driving behaviour to the vehicles far down-stream. This may result in a coherence in the traffic stream even when the density is large.

Judging from the results in section 3.7.3.4, a weak correlation seems to exist between the fluctuations in the range of 2-6 seconds and the density (as well as the level of speed), while the correlation with the flow is significantly lower. Thus, there exists undoubtedly some relation between the size of the fluctuations and the density or the level of velocity. Although this connection seems too weak to be exploited for determining the traffic state, it might be useful for other purposes, such as testing and improving car-following models.

It should be added that one and the same vehicle and driver was used in the experiments. Our test driver has driven along this road section at many occasions and may therefore be expected to be representive for the daily user of the road. Possibly, a less experienced (or more nervous) driver might show other characteristics in the driving behaviour and in the high resolution pattern.

3.7.5 References

[1] Torday, A., Elaboration d un système de navigation auto-alimenté (SNA) et évaluation de ses performances , Thesis nº 3.147, École Polytechnique Fédérale de Lausanne, 2005.
[2] AIMSUN 5.1 Users Manual 2006. Transport Simulation Systems
[3] Van Aerde, M., et al., INTEGRATION: An overview of traffic simulation features, in Department of Civil Engineering. 1996, Queen's University: Kingston, Canada.

[4] Boyce, D., J. Hicks, and A.K. Sen. In-vehicle navigation system requirements for monitoring link travel times in dynamic route guidance system. Proceedings of 70th Annual Meeting of the TRB. 1991. Washington D.C.
[5] Boyce, D.E., A.M. Kirson, and J.L. Schofer, ADVANCE - The Illinois dynamic navigation and route guidance demonstration program, in Advanced technology for road transport: IVHS and ATT. 1994.
[6] Sen, A., et al. Arterial Link Travel Time Estimation: Probes, Detectors and Assignment-type Models. Proceedings of 76th Annual meeting of the TRB. 1997. Washington D.C.
[7] Sen, A., et al., Frequency of probe reports and variance of travel time estimates. Journal of Transportation Engineering, 1997. 123(4): p. 290-297
[8] Sanwal, K. and J. Walrand, Vehicles as Probe. 1995, California PATH Working Paper UCB-ITS-PWP-95-11.
[9] Sen, A., ADVANCE Evaluation: Frequency of probe reports. 1996, Urban Transportation Research, University of Illinois: Chicago.
[10] Soot, S. and H. Condie, ADVANCE Quality of Probe Reports. 1996, Urban Transportation Research, University of Illinois: Chicago.
[11] Westerman, M., R. Litjens, and J.-P. Linnartz, Integration of Probe Vehicle and Induction Loop Data - Estimation of Travel Times and Automatic Incident Detection. 1996, California PATH Research Report.
[12] Hellinga, B. and L. Fu, Assessing the expected accuracy of probe vehicle travel time report. Journal of Transportation Engineering, 1999. 125(6): p. 524-530.
[13] Box G.E.P., Jenkins G.M., Reinsel G.C., 1994. Time Series Analysis, Prentice Hall, New Jersey
[14] Bogenberger K., Belzner H., Kates R.: Ein hybrides Modell basierend auf einem Neuronalen Netz und einem ARIMA-Zeitreihenmodell zur Prognose lokaler Verkehrsgrößen (A hybrid model based on a neural network and an ARIMA timeseries model for prediction of local traffic measurands), Straßenverkehrstechnik 1 (2003)
[15] Davidsson F., Matstoms P. et al. OPTIS Statistical analysis and PROBE simulation study, Centrum för trafiksimulering, Royal Institute of Technology, Sweden, 2002
[16] Office of Highway Information Management, Texas Transportation Institute. Travel Time Data Collection Handbook, Report No. FHWA-PL-98-035, U.S. Department of Transportation, March 1998
[17] Laborczi P., Nowotny B. et al. Optimal Route Guidance Based on Floating Car Data, World Congress of Transport Research 2004, Istanbul, July 2004
[18] B. Nowotny, J. Asamer, K. Din, K. Rouzbeh, 2003. Classification of traffic data time series by cluster analysis, artificial neural networks and ANOVA, 10th World Congress on Intelligent Transport Systems, Madrid, Spain, November 2003
[19] Vägverket, *STRESS Tjänstekoncept, Version 1.0*, Internal PM, Swedish National Road Administration. (In Swedish)
[20] Vägverket, *Specifikation Tjänstekoncept för restidsberäkning, Version 1.0*, Internal PM, Swedish National Road Administration. (In Swedish.)
[21] Brorsson, J., Probes i Stockholm, Ett praktiskt försök att mäta restider i realtid . AerotechTelub, 2001. (In Swedish)
[22] Westman, F., Modeller för restidsuppskattning baserat på Floating Car Data. , Master s thesis, LITH-ITN-KTS-EX-02/13-SE, Linköping University, 2002. (In Swedish.)

4 Data Fusion for Traffic Indicators/Travel Time

4.1 Introduction

Providing real-time traffic information is one of the critical issues for the success of Intelligent Transportation Systems (ITS) programs and becomes a major challenge for public institutions and private companies [1].

Many ITS applications, such as Advanced Traffic Management Systems (ATMS), have been and are still developed in order to improve traffic performance and safety. They mainly rely on real-time data or, more precisely, indicators that are used to manage traffic flows, react on emergency cases, and inform road users. Nevertheless, most conclusions in the studies and researches of this field point out that the weak spot of such applications is the poor quantity and quality of real-time data available for input.

As economical aspects are critical for ITS developments, new sensor development and implementation are more and more limited, particularly for the infrastructure part. Consequently, a better use of actual sensor equipment becomes a strategic topic in ITS applications. This can be achieved through data fusion.

In this section we concentrate on data fusion for two traffic performance indicators: Travel time and traffic density. The data fusion processes developed in the INTRO project aim to provide a consistent and comprehensive picture of network conditions.

This chapter is organized into five sections. Section 2 describes general introduction on concept of data fusion, different approaches and some of its related theory. Section 3 describes the comprehensive framework for data fusion from traffic monitoring indicators perspective adopted here. Section 4 describes the data fusion applications developed within INTRO: The first part concerns travel time and the innovative use of ETC data to regularly correct the Cumulative Traffic Counts method (CTC). The fusion of data from different tolls is also developed for reference travel time. A case-study is conducted on 4 scenarios on some French highway sections, using real-world data. The second part concerns the fusion of Floating Car Data (FCD) for the dynamic estimation of Origin-Destination (OD) matrices. A case-study is conducted on a test bed ring in the region of Vienna, Austria. Conclusions are drawn in section 5.

4.2 Data Fusion: Concepts and methods

Decision problems are faced with increasing data sources (sensors, databases, experts), often inconsistent, sometimes contradictory, making a synthesis tremendously difficult to build.

In order to be fully useful, data sources need more and more complex processing techniques, to take into account data qualities and imperfections of each source. One solution is based on data fusion. Data fusion approach is quite appealing mainly in heterogeneous data sources settings.

This part is an attempt to define data fusion, both on a conceptual standpoint (definitions, principles) and from a practical point of view (list of different techniques and their characteristics).

4.2.1 Basic concepts

Data fusion has been considered very early in different fields (Military, Weather). Each field has started developing its own nomenclature in data fusion and different techniques, although the problem was basically the same. Some attempts have been made in order to provide generic terms and framework of data fusion, but to date, no widely accepted nomenclature exists.

This chapter defines a common definition and generic framework for data fusion. The most widely used data fusion techniques are presented in the next chapter.

4.2.1.1 Definition

Data and Information fusion are routinely exploited in daily applications. An illustration of data fusion is given by the human system. For instance, we are able to identify a given scene on the road from information coming essentially from our two eyes and/or our two ears (i.e. 1 to 4 sensors). Even with only one sensor , changing head (hence sensor) position can lead to more information (than just a picture or a sound record). Indeed, all available data or information is processed within the brain, taking into account other information as memory, experience and a priori knowledge. This process is referred to as a *fusion*: Data fusion exploits the synergy offered by the information originating from various sources.

Prior to defining data fusion, it is necessary to clarify the concepts of *data*, *information* and *knowledge* (used in the former example):

DEFINITION 4.1: Data denotes raw observations, measurements, or/and assertions. Information denotes *structured* data in the context of a study. Knowledge denotes information that is *understood* and *explained*.

Note that all three concepts may present some imperfections (uncertainty and/or imprecision, as discussed in 4.2.1.3.2) that will have to be modelled within the fusion process and more specifically at knowledge modelling step (see Fig. 4.1).

With these concepts in mind, it is possible to define the notions of data fusion and other related concepts.

4.2.1.1.1 Generic definition

Data fusion is a recent term. It means an approach to information extraction, spontaneously adopted in several domains before this was expressed as "data fusion". Two related concepts are worth distinguishing:

- **Fission** is the process that splits observations of a single source into several fragmented views. Hence, collecting data from a given scene can be seen as a particular case of fission, [2]: Each sensor gives a fragmented view of this scene, for instance, an estimate of the state of the underlying system. Yet, our sensors are the only means of getting an idea of the scene (or more generally the real world).

- **Fusion** is thus the opposite process that tries to rebuild a real world view, using as much information as possible (or at least the information that is sought) on the scene, from the available sensors and other information such as a priori knowledge.

Fig. 4.1 draws the relation between fission and fusion. It helps conceptualize the operation of fusion, although it does not take into account the necessary sorting of data observation that concern the system under study (from those that concern other systems). This point will be further developed.

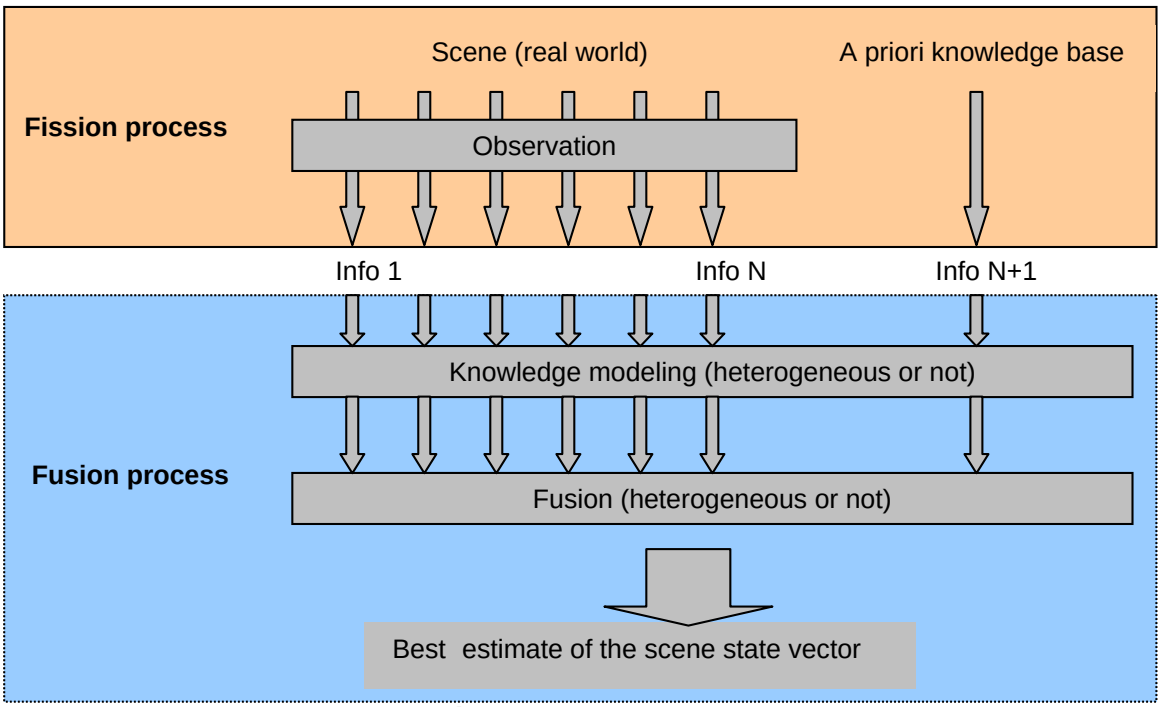


Fig. 4.1: Fusion vs. fission
(Adapted from [7])

Hence the concept of data fusion is quite easy to understand by the examples given above. And yet, there is no consensus to date on a definition of data fusion.

Let aside the battle of words and unclear or very restrictive definitions, there are two major definitions to date:

The JDL[2] definition, [4], was emerged from the first field to consider data fusion (Military). It is both rather general and very popular in the military community. Yet it is still military-oriented: mobile tracking and association.

DEFINITION 4.2 – *Data fusion is a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats and their significance.*

More recently, a fully general definition has been searched, leading to several proposals, which show to be in accordance, [5], [6], [7] and [8] In the context of INTRO WP4, T4.3 (fusion of traffic and safety indicators), we shall use the following generic definition:

DEFINITION 4.3 – *Data fusion denotes the process that enables to best combine multi-source data, possibly heterogeneous, for resulting information of higher quality. The exact meanings of best and higher quality are application-dependent*.

This definition, adapted from [7], is chosen because it is a generic definition that puts strength on three essential points in data fusion:

1. Data and information can be heterogeneous.
2. Pieces of information are used jointly, leading to more reliable results/decisions.

[2] JDL stands for Joint Directors of Laboratories, U.S. Department of Defense. The definition is issued from a Data Fusion Subpanel.

3. The data fusion method is goal-oriented. Hence, the fusion can be applied for reactive state estimate (using current information) and for proactive state estimate.

4.2.1.1.2 Related concepts

Other concepts connected to data fusion can be encountered when dealing with data fusion. This section clarifies the definition of the main concepts in data fusion field and their relationships: association, integration, fusion and aggregation.

These concepts are presented in the following subsections, but basically:

1. Association is a pre-process for fusion but can be one goal of the fusion itself (military problems).
2. Integration is a particular case of fusion (incremental concatenation of knowledge).
3. Fusion is the general process.
4. Aggregation is a fusion (that can follow the general process) but applied to decisions derived from individual source (distributed fusion) instead of raw data from available data sources (as this is not always do-able).

4.2.1.1.2.1 Association

Association consists of grouping information referring to the same object (phenomenon, entity) of the scene. Basically, associating two representations requires that they refer to the same object. The union of the representations is then called *association* or concatenation. In other words, association is made of an increase of the size of the state vector of the object.

This process contains a first step towards fusion, which is called alignment (see Section 4.2.1.2.1). For instance, when trying to identify road traffic congestion, data should be associated by road section and direction. There is no mean and no benefit to fuse data from opposite directions or disconnected sections. Hence, association would associate video data, inductive loops data and probe data that concern the same section, direction and time period. These data are then fused to compute a level of congestion.

Note that association can take place at a first level of sensors, but that it really takes place at the level of information. For instance, association should link the time stamps of two video sensors (placed at two different points of the section) when they have recognized the same car plate.

Association is independent of the semantic level of the information. Association is performed by an analysis of the degree of correlation / relation between the information to be fused and the entity under concern. Some examples can be easily found, where sources are not exactly referring to the same object. In that case, though the sources are aligned, the representations cannot be associated. It can be a matter of period of observation for example. If one is observing a moving target within a limited period, any information well before and well after is poorly correlated to the dynamics of the target. On the contrary, quasi-simultaneous observations of different or same parts of the same human spine by X-rays scanner and nuclear-magnetic resonance imaging system refer to the same object. Once aligned for units (or grey level dynamics) and geometric superimposability, these observations can be fused for, e.g., a 3-D reconstruction of the spine, possibly given some additional knowledge.

Data concatenation is accomplished by juxtaposing all the data into the state vector, hence augmenting it. A straightforward example is given by a time-series of images from the meteorological geostationary satellites, which are taking a picture of the Earth every half-hour or less. The raw data are processed by the meteorological offices, and are spatially superimposable once delivered to the customer. In that case, at each pixel, one can define a state vector by the concatenation of all the observations made at this pixel in the period under concern. Because the data provider has performed the alignment of data, the customer deals in this case with concatenation and subsequent analysis.

In some cases, the problem can be the selection of sub-sets of sensors, which are the most relevant for a given problem. A metric should then be defined for the comparison between sensors, and the choice of the most appropriate ones.

4.2.1.1.2.2 Integration

In the case where sources are in sequence, the fusion of associated data can be seen as a concatenation of available information from the different sources. Hence, every new piece of information is integrated to improve the knowledge, which leads to a more complete estimation of the state of the system, as shown in Fig. 4.2.

This particular case of fusion is generally named integration.

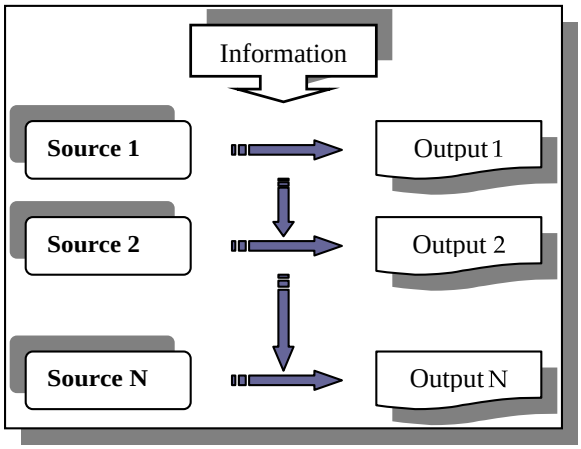


Fig. 4.2: Integration
(Adapted from [7])

A typical example of such process is Kalman filtering in which more recent pieces of information are integrated for the state estimation revisions, [9]. Such filter can be viewed as temporal fusion.

Note that if sources are not independent, integration ignores their interactions. Besides, if one source is missing, part of the estimation of the state of the system will be missing too. Hence, a more general fusion process than integration can be sought.

4.2.1.1.2.3 Fusion

When sources are not independent, it is necessary to take into account their inter-correlation. Such process can be achieved by data fusion. This solution is very efficient, as data fusion takes into account the complementariness and redundancy of sources.

As in integration, fusion takes into account pieces of information as they are available, but the determination of the whole state of the system is not necessarily sequential: Sources can be in sequence or simultaneous. All sources are taken into account for the fusion. See Fig. 4.3.

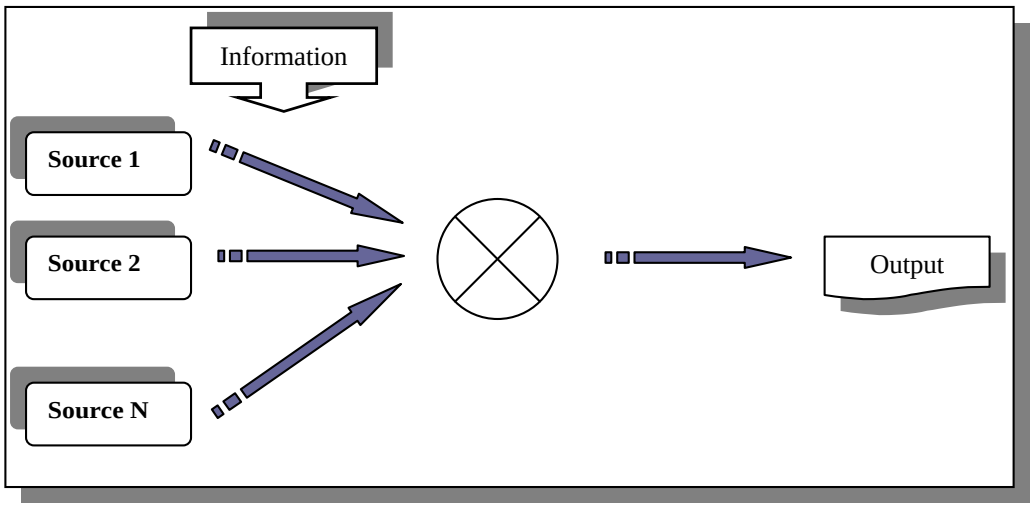


Fig. 4.3: Fusion

4.2.1.1.2.4 Aggregation

In the special case where fusion operates not on raw data/information but on individual state estimations or decisions (derived from individual sources), the fusion is named aggregation, as in Fig. 4.4. This is also called distributed fusion as the state/decision is taken on the basis of individual outputs of each source (processed locally).

Aggregation is especially useful when raw data or the methods used to compute (some of) the available state estimators are not available. For instance: Aggregation of opinions of experts.

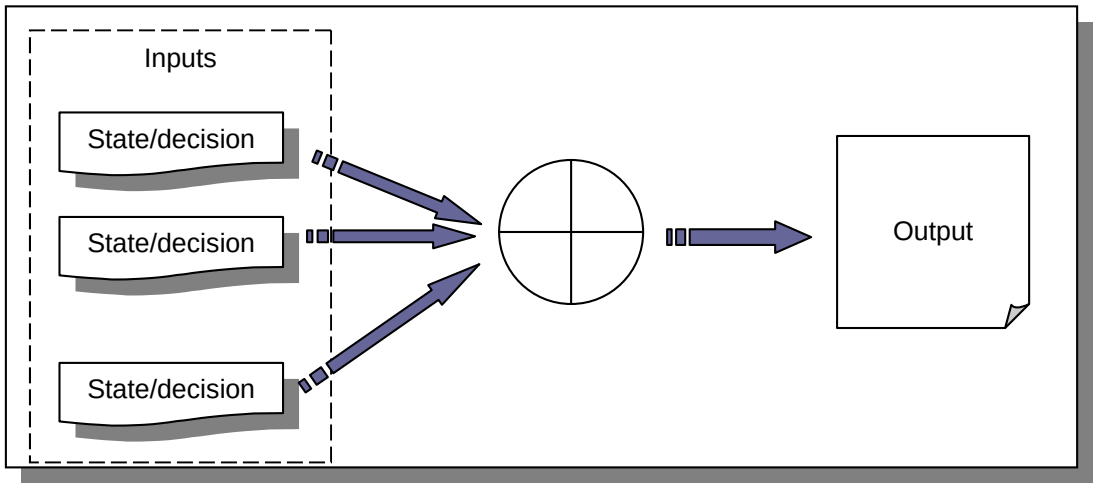


Fig. 4.4: Aggregation

Hence, aggregation also makes a combination of individual sources (State 1 to N in Fig. 4.4), which can be a general fusion or integration. For instance, a first source will lead to one state of the system and a second source will lead to another state of the system (N = 2). Aggregation consists of deciding the state of the system from this information. Information on the sources can also be taken into account, as their reliability, for instance. Note that the reliability of a state estimator can be a function of several parameters (it is not necessarily constant).

4.2.1.2 Generic process

Deciding the state of a system often relies on numerous sources, heterogeneous, complementary and redundant, sometimes contradictory, more or less precise, more or less certain. The aim is then to extract as much information as possible from these sources, as well as gaining in precision and certainty.

The generic process of data fusion can be split into 4 consecutive steps (the output of each step serves as input of the next step), as shown in Fig. 4.5: Alignment, knowledge modelling, combination and decision.

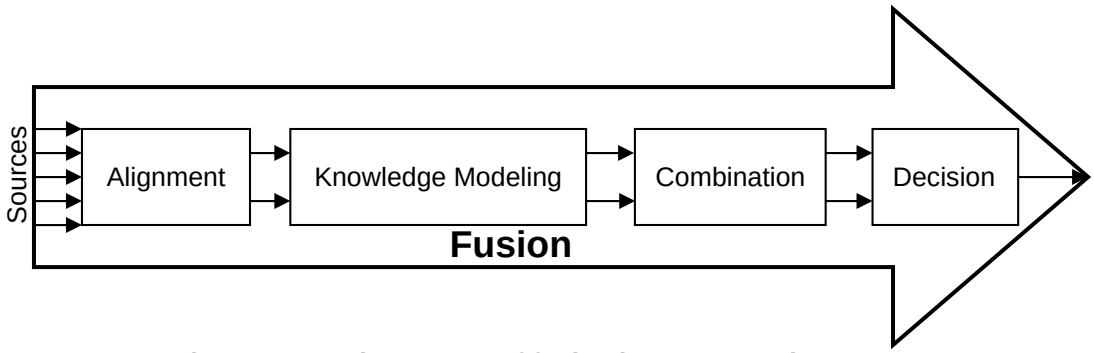


Fig. 4.5: Generic process of fusion in 4 consecutive steps

Thus, fusion is reached step by step. Each step is explained in the following subsections, but basically:

1. Alignment puts data in a common representation space (in time and in space).
2. Knowledge modelling takes into account imperfections of sources and derives knowledge representation metrics, conveyed by each source, via source likelihood measures.
3. Combination merges likelihood measures according to a specific combination rule.
4. Decision selects the best state (according to a given strategy/criterion).

4.2.1.2.1 Alignment (spatial and temporal)

Alignment can be viewed as a sub-process which matches and maps a given data source to a common representation space in which the fusion will take place. More precisely, this step transforms each source into equivalent information in a common representation space. This space is chosen so that the different sources give homogeneous information (same object, time and space).

Basically, alignment defines and applies a homomorphism source by source, both in time and in space. This step is necessary when the different data sources to fuse are heterogeneous. In the case of Fig. 4.1, no alignment step appeared, as data was supposed to be already aligned.

4.2.1.2.2 Knowledge modelling

The data from each source not only needs alignment, but also in-depth modelling, especially when information is imperfect[3]. Indeed, any source even as imperfect it may be conveys information that should be used to get the best picture possible of reality. So modelling sources and the knowledge they contain are fundamental.

In fact, data fusion consists mainly of combining likelihood or distance measures computed from each individual source. Knowledge modelling is precisely the step where these measures are defined: probability, possibility, belief function, distance

So this step will lead to the choice of a specific fusion framework, based on the following considerations:

- How to model imperfections of available data sources?
- How to quantify and model the reliability of sources?
- How to best combine available information, given the sources complementariness and redundancy?

[3] Imperfections are discussed in 4.2.1.3.2.

4.2.1.2.3 Combination

This step is the one that really fuses information: Aligned, modelled information is combined, following a specific combination rule, which depends on the chosen fusion framework (cf. 4.2.1.2.2). Indeed, combining probabilities, possibilities or belief functions leads to distinct combination rules. So the combination rule depends on 4.2.1.2.2 choice of a likelihood measure. Note that a combination rule must be able to deal with contradictory information and to model potential conflicts among sources.

4.2.1.2.4 Decision

This final step computes the estimation of the state of the system, making the best choice (for a specific criterion) from fused information. This selection relies on fused likelihood measures, which are derived from the previous process and are assigned to each state of the system. The final decision is subject to a decision strategy (e.g. maximum likelihood criterion).

4.2.1.3 Classification: Architecture and Environment

For real world applications, the method to use for fusion mainly relies on the characteristics of the sources. These characteristics can be analysed from several points of view, leading to several classifications of fusion methods. For instance, such a classification can be based on a system point of view (as the architecture of the system of fusion) or on a functional point of view (as the nature of the environment in which a decision is searched). As the key step in the fusion process depends mainly on knowledge modelling (cf. 4.2.1.2.2), Section 4.2.2 will focus mainly on the environment classification, grouping the techniques by environment. Furthermore, for each technique, the corresponding architecture will also be described.

4.2.1.3.1 Architecture / Schemes

From a system point of view, a data fusion application can be build according to 2 main architectures: parallel and serial.

In a parallel configuration, each source gives information simultaneously on the system state. Moreover, sources are exchangeable (no order) and similar (in the data they deliver). See Fig. 4.6. Hence, this type of fusion is denoted as symmetric; see for instance [10].

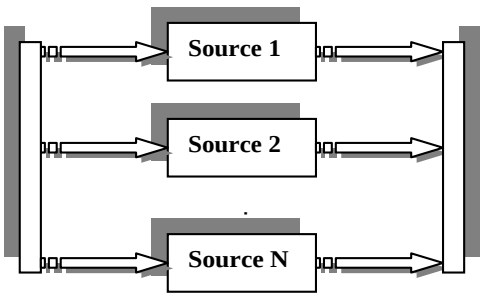


Fig. 4.6: Parallel architecture
(Adapted from [7])

In a serial configuration, the data from each source is combined with the decision taken at the former source level, see Fig. 4.7. Here, the order is important, so this type of fusion is referred to as asymmetric fusion. For instance, Kalman filter and (more generally) integration are asymmetric fusion.

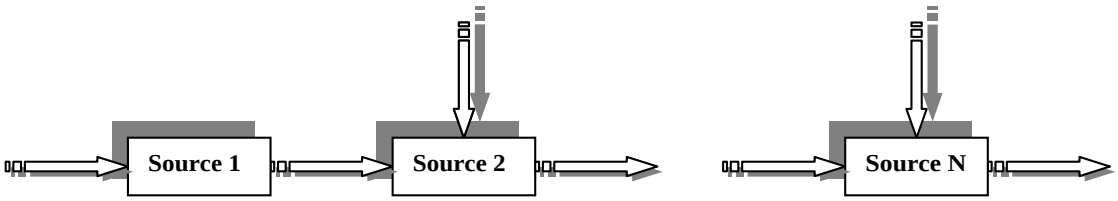


Fig. 4.7: Serial architecture
(Adapted from [7])

The type of fusion architecture mainly depends on the level of availability of multiple sources and their similarity. Note that hybrid configurations are possible (containing symmetric and asymmetric parts or subparts).

As such, the architecture classification does not list the different fusion techniques, the choice of which mainly depends on the environment. So architecture classification can be interesting on a system point of view, but at a secondary level.

4.2.1.3.2 Environment

For a primary classification level, we propose the environment classification that has a more functional point of view, as it is directly linked to the fusion techniques.

To do so, we need to define the following terms:

- Certainty is linked to veracity of information. The major source of uncertainty is ignorance due to both:
  - Bias (systematic error).
  - Randomness (random error).

- Precision is linked to an approximation degree (for the result). Imprecision can be of two types:
  - Vague information, when information is ambiguous.
  - Fuzzy information, when information frontiers are not given.

Note that there is a duality between imprecision and uncertainty. For instance, in Probability, the higher is the level of confidence, the larger is the confidence interval.

So, one can distinguish 4 main environments for the decision process:
- Perfect environment: The data to fuse are both certain and precise.
- Uncertain environment: The data to fuse are uncertain but precise.
- Imprecise environment: The data to fuse are imprecise but certain.
- Imperfect environment: The data to fuse are both uncertain and imprecise.

In each environment there are different fusion techniques that model the type of data considered, see **Fig. 4.8**. These techniques are presented in the next section.

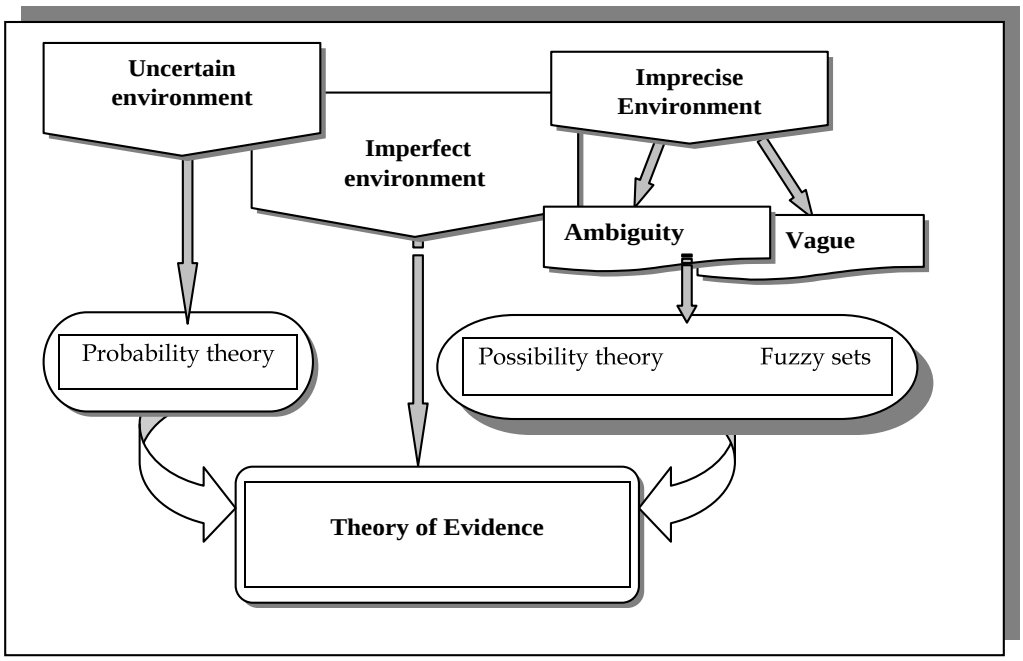


Fig. 4.8: Main theories in data fusion and their relationships
(Adapted from [7])

4.2.2 Main techniques

4.2.2.1 Uncertainty-based framework

4.2.2.1.1 Bayesian fusion

Probability theory has proven its suitability in modelling uncertainty within the fusion context. Most of the fusion techniques are based on the Bayesian approach.

As all probabilistic methods, the Bayesian approach enables to build a model that takes into account uncertainty. In the case of fusion, this approach models the uncertainties of the different sources and combines these inputs into a fused uncertainty . Hence, the final output is based on a maximum likelihood decision (i.e. the value of the system that has the lowest fused uncertainty).

Namely, let k be the number of sources, $\mathrm{O} = (\varsigma^1, \ldots, \varsigma^k)$ be the observations gathered from all the sources[4], χ be the characteristic of the system that is studied, and U be the set of all possible values for χ. $\forall u \in U,\ p(\mathrm{O} \mid u)$ is the likelihood function for u : Probability of observing O when $\chi = u$. $\forall u \in U,\ p(u \mid \mathrm{O})$ is the probability that $\chi = u$ when O is observed.

Bayes formula writes:

$$p(u \mid \mathrm{O}) = \frac{p(\mathrm{O} \mid u)\, p(u)}{p(\mathrm{O})}, \qquad (4.1)$$

where $p(u)$ is the probability that $\chi = u$ and $p(\mathrm{O})$ is the probability of observing O.

When Θ is finite, Eq.(4.1) writes:

[4] Note that each source may provide several observations: $\forall s \in [1, k], \varsigma^s = (o_1^s, \ldots, o_{n_s}^s)$.

$$\forall i \in [1, r],\ p(u_i \mid \mathrm{O}) = \frac{p(\mathrm{O} \mid u_i)\, p(u_i)}{\sum_{j=1}^{r} p(\mathrm{O} \mid u_j)\, p(u_j)}, \tag{4.2}$$

where $r = card(U)$ and $U = \{u_1, \ldots, u_r\}$.

As stressed in [11], the fusion process is implicit in those Bayes formulas. Indeed, it takes place in $p(\mathrm{O} \mid u)$ computation: Combination of the likelihood functions associated to each single source, $p(\varsigma^s \mid u),\ s = 1, \ldots, k$, possibly taking into account other information as sources dependencies and reliabilities.
In most of the cases, sources are assumed to be conditionally independent, leading to:

$$p(\mathrm{O} \mid u) = \prod_{s=1}^{k} p(\varsigma^s \mid u) \tag{4.3}$$

4.2.2.1.2 Evidential fusion

Theory of Evidence was first outlined by Dempster in the sixties and formalized by Shafer in the seventies: [12],[13] and [14]. It is also known as Dempster-Shafer Theory, Belief Functions Theory

This theory can be seen as an extension of Bayes approach, accounting not only for uncertainty but also for imprecision.

The fundamental of this theory is that it can deal with non exclusive hypothesis, contrarily to probability theory. Indeed, it is possible to associate a degree of belief not only to atomic hypothesis but also to disjunctive sets of hypotheses without any additive axiom.

Namely, let U be the universe of decision (e.g. list of the different states of the system). U is also named *frame of discernment*. In this theory, U is supposed to be *finite* and the hypotheses it contains are both exhaustive and mutually exclusive: $U = \{u_1, \ldots, u_r\}$. The theory also supposes that one and only one hypothesis is true at a time (closed world assumption).
Now, the theory is based on the use of a basic belief assignment function that associates a degree of belief to each subset of U:

$$\begin{aligned} m_U :\ P(U) &\to [0,\ 1] \\ A &\mapsto m_U(A) \end{aligned}, \tag{4.4}$$

where $P(U)$ is the power set of U.
This function can be freely defined provided it is consistent with the two following properties:

- $m_U(\varnothing) = 0$ (Closed world assumption).
- $\sum_{A \in P(U)} m_U(A) = 1$.

In probability theory, a measure of probability is assigned to atomic propositions $u_i,\ i \in [1, r]$ and this measure is additive.
Here, m_U is usually neither additive nor monotonic. Besides, the following properties usually hold: $m_U(U) < 1$ and $m_U(A) + m_U(\bar{A}) < 1$, where A is a subset of U and $\bar{A}$ is its complement in U. Indeed, $\forall A \in P(U),\ m_U(A)$ is the part of belief that supports A, but does not support anything more specific (i.e. strict subsets of A).

Namely, $\forall i \in [\![1,r]\!]$, $m_U(u_i)$ represents the degree of belief that the true value is u_i for sure. In case of imprecision, a degree of belief can be associated to a set containing several propositions, not taking into account the degree of belief already associated to more precise subsets (i.e. strict subsets of this set). Hence, $\forall A \in P(U)$, $m_U(A)$ represents the degree of belief associated to A that cannot be transferred to (strict) subsets of A. A subset A for which $m_U(A) > 0$ is called a *focal element*.

If all focal elements are singletons of U, then the basic belief assignment function is infect a probability function and Theory of Evidence meets the Bayesian approach (Section 4.2.2.1.1), [14]. This happens when there is no imprecision to model. This fact underlines that Bayesian approach does not model imprecision but only uncertainty. Hence, Theory of Evidence can be seen as an extension of the Bayesian approach for modelling imprecision.

In the end, use of this methodology only requires to have a finite universe (the frame of discernment) and to define the degrees of belief that pertain to each subset.
This is done for each source. The fusion of sources is then obtained by Dempster orthogonal sum combination rule, [14].

In the case of two sources, ς^1 (modelled by m_U^1) and ς^2 (modelled by m_U^2), the Dempster orthogonal sum combination rule writes $m_U^1 \oplus m_U^2$:

$$\begin{cases} m_U^1 \oplus m_U^2(\varnothing) = 0 \\ \forall B \in P(U) \setminus \{\varnothing\},\ m_U^1 \oplus m_U^2(B) = \dfrac{1}{1-\Lambda_c} \displaystyle\sum_{\{(A_1,A_2) \in P(U)^2 / A_1 \cap A_2 = B\}} m_U^1(A_1) \cdot m_U^2(A_2) \end{cases}$$

where $\Lambda_c = \displaystyle\sum_{\{(A_1,A_2) \in P(U)^2 / A_1 \cap A_2 = \varnothing\}} m_U^1(A_1) \cdot m_U^2(A_2)$ is a normalization term, with $0 \le \Lambda_c \le 1$, which can be viewed as a measure of the conflict between the two sources.

The Dempster s combination rule has been justified theoretically by several authors [15], [16] However, the normalization step was also criticized [17], [16] In fact, it is very important to take into account the value of Λ_c: When Λ_c is high (≈ 1), combining the sources is a non-sense leading to incoherence and involving counter-intuitive behaviours.

The Dempster orthogonal combination rule can straightforwardly be extended to more than 2 sources: Let k be the number of sources, $\{\varsigma^1, \ldots, \varsigma^k\}$. For each source ς^s, $s \in [\![1,k]\!]$, let m_U^s be the basic belief assignment function that models ς^s. Then, the general Dempster orthogonal sum combination rule writes $\overset{k}{\underset{s=1}{\oplus}} m_U^s$:

$$\begin{cases} \overset{k}{\underset{s=1}{\oplus}} m_U^s(\varnothing) = 0 \\ \forall B \in P(U) \setminus \{\varnothing\},\ \overset{k}{\underset{s=1}{\oplus}} m_U^s(B) = \dfrac{1}{1-\Lambda_c} \displaystyle\sum_{\{(A_1,\ldots,A_k) \in P(U)^k / \bigcap_{s=1}^{k} A_s = B\}} \prod_{s=1}^{k} m_U^s(A_s) \end{cases}$$

where $\Lambda_c = \displaystyle\sum_{\{(A_1,\ldots,A_k) \in P(U)^k / \bigcap_{s=1}^{k} A_s = \varnothing\}} \prod_{s=1}^{k} m_U^s(A_s)$.

Going back to Fig. **4.5**, the last step of decision can follow several strategies in the case of using Theory of Evidence:

1. Maximum credibility:
   Credibility of a set A is defined as:
   $$Cr_U(A) = \sum_{B \subseteq A} m_U(B),\ \forall A \in P(U).$$
   Credibility represents the quantity of evidence that supports for sure a solution in A. Note that $Cr_U(U) = 1$ and, more precisely, $Cr_U\left(\bigcup_{A \in \{B \subset U / m_U(B) > 0\}} A\right) = 1$.

   Maximum credibility strategy consists of selecting the hypothesis $u_{Cr} \in U$ such that $Cr_U(\{u_{Cr}\})$ is maximum (among the $Cr_U(\{u_i\}),\ \forall i \in [1, r]$):
   $$u_{Cr} = \arg\max_{u \in U} Cr_U(\{u\}).$$
   This strategy is somehow pessimistic, as it does not take into account the evidence that could support a solution in A, namely: The $m_U(B), \forall B$ with $A \not\emptyset\ B \subseteq U$. For these B, the mass is supposed to belong to $B \setminus A$ (and not at all to A).
   Note that $\forall i \in [1, r], Cr_U(\{u_i\}) = m_U(\{u_i\})$. Hence, u_{Cr} is the hypothesis that has the highest guaranteed mass (i.e. not taking into account the mass affected to sets containing u_{Cr}).

2. Maximum plausibility:
   Plausibility of a set A is defined as:
   $$Pl_U(A) = \sum_{B \in \{C \subset U / A \cap C \neq \emptyset\}} m_U(B),\ \forall A \in P(U).$$
   Plausibility represents the quantity of evidence that could possibly support a solution in A.
   Note that plausibility and credibility are dual: Evidence either possibly supports A or discards A for sure (i.e. supports for sure $\overline{A} = U \setminus A$):
   $$\forall A \subset U, Pl_U(A) + Cr_U(\overline{A}) = 1.$$
   Note also that the following property holds:
   $$\forall A \subset U, 0 \leq Cr_U(A) \leq Pl_U(A) \leq 1.$$

   Maximum plausibility strategy consists of selecting the hypothesis $u_{Pl} \in U$ such that $Pl_U(\{u_{Pl}\})$ is maximum (among the $Pl_U(\{u_i\}),\ \forall i \in [1, r]$):
   $$u_{Pl} = \arg\max_{u \in U} Pl_U(\{u\}).$$
   This strategy is optimistic, as $\forall B \subseteq U$ with $A \cap B \neq \emptyset$, $m_U(B)$ is taken as belonging to A (in fact, $A \cap B$) only (and not at all to $B \setminus A$).
   Note that $\forall i \in [1, r], Pl_U(\{u_i\}) = \sum_{A \in \{B \subseteq U / u_i \in B\}} m_U(A)$. Hence, u_{Pl} is the hypothesis that has the highest possible mass (i.e. taking into account the whole mass affected to all sets containing u_{Cr}).

3. Best belief interval $[Cr_U, Pl_U]$:

$\forall i \in [1, r], \left[Cr_U\left(\{u_i\}\right), Pl_U\left(\{u_i\}\right) \right]$ represents the domain containing the real belief that should be associated to u_i. The larger the interval, the higher the uncertainty on the real belief that should be associated to u_i. Hence, the more this interval tends towards singleton $\{1\}$, the more surely hypothesis u_i can be selected.

Namely, this strategy selects $u_{BI} \in U$ such that $u_{BI} = \arg\max_{u_M \in U_{Pl}} Cr_U(\{u_M\})$, where $U_{Pl} = \{u_M \in U / Pl_U(\{u_M\}) = Pl_U(\{u_{Pl}\})\}$. Note that this strategy extends the maximum credibility strategy to the case where U_{Pl} is not a singleton (where it is difficult to select one specific u_{Pl}). Indeed, $Pl_U(\{u_{BI}\})$ is maximum (i.e. equal to $Pl_U(\{u_{Pl}\})$. In the case where $card(U_{Pl}) = 1$, $u_{BI} = u_{Pl}$. When $card(U_{Pl}) > 1$, u_{BI} is chosen among U_{Pl}, so that $Cr_U\left(\{u_{BI}\}\right)$ is maximum in U_{Pl}.

Note that $Pl_U(\{u_{BI}\}) = Pl_U(\{u_{Pl}\}) = 1$ if and only if the source is coherent (i.e. $\forall (A, B) \subset U^2, A \cap B = \varnothing \Rightarrow m_U(A) \cdot m_U(B) = 0$).

When $card(U_{Pl}) = 1$, the condition $Pl_U(\{u_{BI}\})$ is maximum can be relaxed in order to select the hypothesis that gives the shortest interval $[Cr_U, Pl_U]$, among the hypotheses with high plausibility. Combination of interval size and plausibility can also give another strategy among the hypothesis with high plausibility.

4. Maximum pignistic probability:
Pignistic[5] probability was first introduced in [18]. It consists of equally distributing the mass of a set on each of its elements:

$$\forall A \subseteq U, Pg_U(A) = Cr_U(A) + \sum_{B \in \{C \subseteq U / A \not\varnothing C\}} \frac{m_U(B)}{card(B)}.$$

Pignistic probability represents the quantity of evidence that probably supports a solution in A. Note that the following property holds:

$$\forall A \subset U, 0 \le Cr_U(A) \le Pg_U(A) \le Pl_U(A) \le 1.$$

Maximum pignistic probability strategy consists of selecting the hypothesis $u_{Pg} \in U$ such that $Pl_U(\{u_{Pg}\})$ is maximum (among the $Pg_U(\{u_i\})$, $\forall i \in [1, r]$):

$$u_{Pg} = \arg\max_{u \in U} Pg_U(\{u\}).$$

This strategy is neither optimistic nor pessimistic. It deals with uncertainty as in probability models, with equal distribution of the mass, in order to compute the hypothesis that probably best fits.

Note that $\forall i \in [1, r], Pg_U(\{u_i\}) = \sum_{A \in \{B \subseteq U / u_i \in B\}} \frac{m_U(A)}{card(A)}$.

[5] *Pignus* is the Latin for *bet*.

When all focal elements are singletons of U, then the basic belief assignment function is in fact a probability function and all former criteria are equal to the mass:

$$\forall A \subseteq U, m_U(A) > 0 \Rightarrow card(A) = 0$$
$$\Downarrow$$
$$\forall B \subseteq U, Cr_U(B) = Pl_U(B) = Pg_U(B) = \sum_{v \in B} m_U(v)$$

Besides, as written earlier, the basic belief assignment function is, in this case, a probability function. Hence, Evidence Theory is truly an extension of Bayesian approach.
Finally, it can be shown that this is the only case where credibility is a probability, [19]:

Cr_U is a probability if and only if $\forall A \subseteq U, m_U(A) = \begin{cases} Cr_U(A), \text{ if } card(A) = 1 \\ 0, \text{ otherwise} \end{cases}$.

4.2.2.2 Possibilistic fusion

In 1965, Zadeh introduced the Fuzzy Set Theory:[20]. This theory has been used in many fields where relations between continuous variables (such as speed, distance) are fuzzy relations. Quantification of these fuzzy relations is modelled through a (fuzzy) membership function, also called possibility distribution. Indeed, a possibility distribution is the distribution of the membership degrees (in $[0,1]$) for a given class, depending on the value of the variable. A possibility degree of 0 states that the considered value is excluded from the class. A possibility degree of 1 states that the considered value is fully in the class. An in-between possibility degree denotes the partial membership degree of the considered value to the class.

Possibility Theory ([21]) is based on possibility distributions ([22]). Let U be the set of possible values for a given variable. In the context of fusion, U contains the list of the different states of the system. As in Theory of Evidence, U is supposed to be *finite* and the propositions it contains are both exhaustive and mutually exclusive: $U = \{u_1, \ldots, u_r\}$. The theory also supposes that one and only one proposition is true at a time (closed world assumption).

Now, the theory is based on the use of a possibility distribution $\pi_U : U \to [0,1]$ that expresses for each u_i ($i = 1 \ldots r$) the possibility that u_i is the current state of the system. Note that $\exists i \in [1, r], \pi_U(u_i) = 1$: At least one state is fully possible (closed world assumption).

Then, possibility measure Π_U is defined:

$$\begin{array}{rccc} \Pi_U : & P(U) & \to & [0,1] \\ & A & \mapsto & \sup_{a \in A} \pi_U(a) \end{array}. \quad (4.5)$$

Note that here $\sup_{a \in A} \pi_U(a)$ is equivalent to $\max_{a \in A} \pi_U(a)$, as U is finite.

Π_U follows these properties:

- $\Pi_U(\varnothing) = 0$.
- $\Pi_U(U) = 1$.
- $\forall A, B \in P(U),\ \Pi_U(A \cup B) = \max(\Pi_U(A), \Pi_U(B))$.

Necessity measure N_U, a dual measure of Π_U, can also be defined:

$$\begin{array}{rccc} \mathrm{N}_U: & P(U) & \to & [0,1] \\ & A & \mapsto & \inf_{a\notin A}\left(1-\pi_U(a)\right)=1-\sup_{a\notin A}\pi_U(a) \end{array} \quad (4.6)$$

Again, the following properties hold:

- $\mathrm{N}_U(\varnothing)=0$.
- $\mathrm{N}_U(U)=1$.
- $\forall A,B\in P(U),\ \mathrm{N}_U(A\cap B)=\min\left(\mathrm{N}_U(A),\mathrm{N}_U(B)\right)$ (Notice the duality).

The duality between Π_U and N_U also reads:

- $\forall A\in P(U),\ \mathrm{N}_U(A)=1-\Pi_U\left(\overline{A}\right)$, where $\overline{A}=U\setminus A$. Equivalently:

  $\forall A\in P(U),\ \Pi_U(A)+\mathrm{N}_U\left(\overline{A}\right)=1$.
- $\forall A\in P(U),\ \mathrm{N}_U(A)\le\Pi_U(A)$.
- $\forall A\in P(U),\ \mathrm{N}_U(A)>0\Rightarrow\Pi_U(A)=1$ or $\forall A\in P(U),\ \Pi_U(A)<1\Rightarrow\mathrm{N}_U(A)=0$ (equivalently).

Going back to the fusion of different sources of information represented by a possibility distribution, three classes of combination operators (named Δ: $\overleftarrow{\Delta}$, $\overrightarrow{\Delta}$ and $\overleftrightarrow{\Delta}$) can be considered,[23], [24].
Namely, let k be the number of sources to fuse and $\forall\varsigma\in[\![1,k]\!],\ \pi_U^{\varsigma}:U\to[0,1]$ be the possibility distribution that models the information of source ς.
The three combination operator classes are:

1. $\overleftarrow{\Delta}$: *Conjunctive operators*: These operators are pessimistic, as their values are always lower than the lowest value among the sources:
   $\forall k\in\mathbb{N}^*,\forall\left(\Pi_U^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]},\forall u\in U,\overleftarrow{\Delta}\left(\left(\Pi_U^{\varsigma}(u)\right)_{\varsigma\in[\![1,k]\!]}\right)\le\inf_{\varsigma\in[\![1,k]\!]}\left(\Pi_U^{\varsigma}(u)\right)$. E.g., $\min$ is a conjunctive operator. T-norms, used in fuzzy logic, are also conjunctive operators. In fact, $\min$ is the smallest T-norm.
2. $\overrightarrow{\Delta}$: *Disjunctive operators*: These operators are optimistic, as their values are always greater than the greatest value among the sources:
   $\forall k\in\mathbb{N}^*,\forall\left(\Pi_U^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]},\forall u\in U,\overrightarrow{\Delta}\left(\left(\Pi_U^{\varsigma}(u)\right)_{\varsigma\in[\![1,k]\!]}\right)\ge\sup_{\varsigma\in[\![1,k]\!]}\left(\Pi_U^{\varsigma}(u)\right)$. E.g., $\max$ is a disjunctive operator. T-co-norms, used in fuzzy logic, are also disjunctive operators. In fact, $\max$ is the largest T-conorm.
3. $\overleftrightarrow{\Delta}$: *Compromise operators*, which have a cautious behaviour:
   $\forall k\in\mathbb{N}^*,\forall\left(\Pi_U^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]},\forall u\in U,\ \inf_{\varsigma\in[\![1,k]\!]}\left(\Pi_U^{\varsigma}(u)\right)\le\overleftrightarrow{\Delta}\left(\left(\Pi_U^{\varsigma}(u)\right)_{\varsigma\in[\![1,k]\!]}\right)\le\sup_{\varsigma\in[\![1,k]\!]}\left(\Pi_U^{\varsigma}(u)\right)$ E.g., means and weighted means are compromise operators:
   $\forall k\in\mathbb{N}^*,\quad\forall\left(\Pi_U^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]},\quad\forall\alpha\in\mathbb{R},\quad\forall\left(w^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]}\in\mathbb{R}^k,$ with $\sum_{\varsigma\in[\![1,k]\!]}w^{\varsigma}=1,\quad\forall u\in U,$

   $\overleftrightarrow{\Delta^{\alpha}_{\left(w^{\varsigma}\right)_{\varsigma\in[\![1,k]\!]}}}\left(\left(\Pi_U^{\varsigma}(u)\right)_{\varsigma\in[\![1,k]\!]}\right)=\left(\sum_{\varsigma\in[\![1,k]\!]}w^{\varsigma}\cdot\left(\Pi_U^{\varsigma}(u)\right)^{\alpha}\right)^{\frac{1}{\alpha}}$ is a compromise operator.

For instance, for classical means, $\forall \varsigma \in [1,k], w^{\varsigma} = \frac{1}{k}$, and the most used compromise operators are:

- Arithmetic mean ($\alpha = 1$).
- Harmonic mean ($\alpha = -1$).
- Geometric mean ($\alpha = 0$, by extension, using Taylor expansion).
- Note that using limits of α in $-\infty$ and $+\infty$ leads to border operators:
  - $\min$ ($\alpha \to -\infty$): Zadeh s T-norm, a conjunctive operator.
  - $\max$ ($\alpha \to +\infty$): Zadeh s T-conorm, a disjunctive operator.

Other alternative operators have been proposed in the literature (see [25]).

4.2.3 Conclusion

The decision making aid requires the development of increasingly precise, synthetic and reliable information. The decision maker faces more and more a surge of data and information of various origins, which they come from experts, data bases or means of measurements, often incoherent, sometimes contradictory, of which it is quite difficult to make the synthesis.

This problem of decision under a multi-source environment is a common problem to a vast field of application. To be fully exploitable, this abundance of sources of information requires increasingly complex techniques of treatment, able to take into account qualities and the imperfections of each source. The techniques of fusion of data constitute an interesting answer to address this need. They make it possible to make profitable a significant number of data by holding account of the diversity of their imperfections and while trying to mitigate the weaknesses of some with the strong points of the others. The knowledge of qualities and the weaknesses of each data source must be regarded as a precondition to any operation of fusion.

It is by carrying out this assessment that one can determine the best structure of the system and the data to be amalgamated. Therefore, one will be able to choose the most efficient methodology.

In front of the profusion at the same time of concepts which proclaim data fusion of data and methodologies of fusion, this section draws up a panorama of the principal concepts relating to fusion of data by clarifying at the same time their fields of actions and their network of interdependence. In addition, it offers a detailed description of the characteristics of the informational elements to take into account in a system of fusion and specifies the guiding principles of the fusion of data as of the mechanisms which control this one. It offers finally introductory presentations to the principal mathematical theories used in fusion of data while placing them from the point of view of practical implementation.

4.3 Data Fusion in Transportation (Traffic and Safety)

An objective of this section is to present an analysis of recent applications of data fusion (DF) in road traffic engineering. First, we report the most significant applications of data fusion techniques in road traffic engineering area: Traffic monitoring, signal control, Automatic Incident Detection, traffic forecasting, Intelligent Transportation Systems ..., as well as the extent and direction of DF interest inthe field. Second, a classification including applications, fusion goals and mathematical tools are proposed. This section is an update of [26].

4.3.1 Introduction

Transportation systems are motivated to obtain reliable information for monitoring and control operations. Providing traffic information is becoming a major challenge for the public institutions and private companies within important Intelligent Transportation Systems (ITS) programs [27]

At the same time, the emergence of these new information technologies and the transformation that has occurred in road traffic management have both increased drivers' already considerable need for a

very accurate road traffic information. So, in order to provide a better picture (more comprehensive and complete) of the state of traffic on a road network, the traffic sensors that are usually used to measure traffic conditions are ineffective and other sources of data (cameras, GPS, cell phone tracking, probe vehicles etc.) are increasingly needed in order to supplement the information provided by that conventional measurement systems.

For many applications, the information provided by individual sensors are incomplete, inconsistent or/and imprecise. Additional sources may provide complementary data, and fusion of different information can produce a better understanding of the observed situation, by decreasing the uncertainty related to the single source. The fusion of multiple sources is perceived, rightly, as a well-adapted answer to the operational needs of traffic management centres and traffic information operators, allowing them to achieve their goal more efficiently.

The primary goal of this chapter is to acquaint the reader with the most significant applications of DF techniques in road traffic engineering area (traffic monitoring, signal control, Automatic Incident Detection, traffic forecasting, Intelligent Transportation Systems), and to indicate the directions of future research in this area.

Section 4.3 is organized into four subsections. Subsection 4.3.2 describes basic of traffic engineering operations with emphasis on data sources available. The DF applications to the traffic engineering area are presented in Subsection 4.3.3. Conclusions are drawn in Subsection 4.3.4.

4.3.2 Opportunities and challenges of Road Traffic Data fusion

Technological advances in the area of road telematics (electronic systems aboard vehicle, vehicle localization, Telecommunications, data processing) have favoured an improvement of existing means of traffic data collection, and sometimes the appearance of new sensors, or new architectures: equipment on board vehicles, off-set on the side of roads, multiform data collection

Basic traffic data is largely based on road sensors embedded in the pavement. Such sensors mostly use inductive loop detectors and are able to measure temporal traffic characteristics. They detect basic parameters needed by operational traffic engineer, such as traffic volume, occupancy and speed at a given point. Other fixed sensors such as optical detectors, ultrasonic based detectors were developed and used for network surveillance.

These fixed sensors are very useful, but they fail in measuring spatial behaviour of traffic. Besides that, they suffer from their limited reliability, estimated to be between 10% and 20 %, with their prohibitive cost in attaining significant coverage of the roadway network. Other fixed sensors with spatial capabilities have been developed and used to supplement loop detectors data: cameras, RMTS (Remote Traffic Microwave Sensor)

The recent deployment of ITS applications and the induced needs for real-time data to fulfil a wide range of purposes (real-time traffic operations monitoring, incident detection and route guidance applications) have given rise to another complementary source of data allowing traffic parameters estimation: One of the actual trends is probe vehicle data collection technique. According to this technique, cars on the road shift from a passive attitude to an active one and act as moving sensor, continuously feeding information about traffic conditions to a traffic management centre.

Different Automatic Vehicle Identification (AVI) systems, based on different technologies can be used as detection devices: automatic vehicle tags identification, automatic license plate matching techniques, global positioning system (GPS) More recently, with the important growth of wireless communications and the spread of cellular phones, technical developments in cellular positioning provides the opportunity to track cell phone equipped drivers as traffic probes. Many research studies have demonstrated the feasibility of using cell phones as traffic probes (see e.g. [28], [29]).

Therefore, a wide spectrum of data and heterogeneous sources of information are of potential use for a given traffic situation. As a result, many problems of traffic engineering become a typical **data fusion problem**. Indeed, some of the basic road traffic engineering used to meet operators needs

have been gained through the conventional fixed detectors. However, in the context of traffic operations where highly accurate information are needed, ITS framework for instance, the information provided by fixed detectors data alone may be not sufficient except in some special situations (for some network configuration and/or with a high detector coverage). Among reasons, the lack of full traffic information at the network wide scale is of importance. So, for traffic operations improvements, other sources of data are increasingly used in order to supplement the information provided by conventional measurement techniques.

The purpose of DF is to produce an improved model or estimate of a system from a set of independent data sources. For traffic applications, the desired model is the state vector of the traffic phenomenon. These estimates may include statements about current or future vehicular speeds, average speeds, travel time, vehicle classification and similar topics of interest to travellers and traffic operators.

4.3.3 Data fusion in Road Traffic engineering

Advanced telematics systems provide a data-rich, real-time environment which gives rise to new challenging problems for the traffic modelling community. In this framework, new models and systems are currently being developed for traffic engineering problems, based on multisource data in which DF process is recognized as a central module.

If the application of DF techniques to complex systems modelling is not new (see e.g. [30]), there is a growing interest in their use in transportation systems. This interest gives rise to an emerging and evolving field and some scientific research has promoted its application in the study of vehicle navigation and related areas. Nevertheless, road traffic could be considered as a field where benefits expected from the application of DF techniques are not only potentially the most relevant, but also as the most challenging field for assessing feasibility, effectiveness and usefulness of such approaches [2], [7].

In traffic engineering literature, the interest for DF is quite new, and it coincides with ITS advent. The first paper which mentioned DF was the one of Sumner in the early 1990s [31] He acknowledged the importance of DF for effectiveness of ITS systems.

Regarding the application of DF in engineering, the book [32] is of utmost importance, as is the recently published book by Klein [35] dedicated to sensor technologies within ITS framework.

4.3.3.1 Data fusion approaches

Data fusion has found widespread application in diverse areas ranging from local civilian application (agricultural disease tracking, environmental diagnostics, cartography, ocean temperature profiling, and urban planning) to global military theatre defence (see [34], [35], [36], [37]). Several methodologies have been proposed in the literature for the purpose of multisensor fusion and aggregation under heterogeneous data configuration. These techniques were drawn from a wide range of areas including artificial intelligence, pattern recognition, statistical estimation, and other areas. Traffic engineering field is naturally benefited from this abundant literature. For instance, independent from applications, a variety of techniques were used ranging from a simple arithmetic mean to a more complex DF approach. More precisely, a three-way split of the proposals could be suggested:

- Techniques based on statistical nature: Weighted combination, multivariate statistical analysis and its most up-to-date form data mining engine. Among statistical techniques, the arithmetic mean approach is the simplest which is used for information combination. This approach is not suitable when the information at hand is not exchangeable or when estimators/classifiers have dissimilar performances (see [38], [39], [40] and [41]).
- Probabilistic-based or probabilistic-like approaches. For instance, Bayesian approach with Bayesian network and state-space models (see [42]). Possibility theory [43], evidential reasoning and more specifically evidence theory ([12], [13] and [14]) are also an important

modelling framework under multisource information. This later technique could be viewed as a generalization of Bayesian approach (see [13]).

- And finally neuromimetic networks and artificial cognition including artificial intelligence, genetic algorithms and neural networks. In many applications, this later approach serves both as a tool to derive classifiers or estimators and as a fusion framework of classifiers/estimators.

All these techniques were applied at different levels of complexity in traffic problems with a special mention of Kalman Filtering. Indeed, application of Kalman filter to road traffic problems has been a long tradition. The early application has been reported about three decades ago (see e.g [44] and [42]).

The DF process investigated in traffic literature involves basic functions such as temporal or/and spatial alignment of input data, data association and data mining for a knowledge extraction purpose. This later purpose is also one of potential objectives of a multisource information fusion (see [45]) other class of problems, where DF was used as a modelling tool, refers to multisensor fusion and classification/estimation fusion.

4.3.3.2 Main data fusion domains

[38] and [46] are the first published papers in which a practical traffic and transportation problem was solved using DF approach. In the late 1990s and the beginning of the 2000s, different authors made significant contributions to the field of DF in transportation systems.

Naturally most research in this area is put into the context of ITS problems. Indeed, ITS offers indisputably the most relevant framework for DF and also the most challenging (see e.g. [2] and [47]). Others conventional problems in transportation modelling have also been concerned with multisource processing. Namely, planning problems, demand estimation, traffic estimation (See [48] and [49]).

4.3.3.3 Intelligent Transportation Systems

Many countries focus on Intelligent Transportation Systems (ITS) to solve their traffic congestion and safety problems. To be effective, a variety of functions were assigned to the ITS. These functions were designed to achieve a specific task in order to cope with an operational problem. One can mention advanced traveller information systems (ATIS), automatic incident detection (AID), advanced driver assistance. Each of these sub-systems can potentially work with a broad set of source data information; combine them to yield to a better decision.

4.3.3.3.1 Advanced Traveller Information Systems

Advanced Traveller Information Systems (ATIS) is one of several ITS sub-systems that offer users integrated traveller information. In ATIS, different automatic data collection techniques are of potential use to apprehend traffic conditions and derive relevant indicators for user s information and guidance (see. e.g. [2] for objectives statement for the research on data Fusion for delivering ATIS).

One form in which the user s information is presented is the travel time, and a number of systems are based on its dissemination. In this context, this indicator is used as a measure of impedance (or cost) for route choice strategies. Travel time is also used by network operators as an indicator of quality of service (QoS).

This raises the problem of estimating travel times with an acceptable degree of accuracy, which is a particularly difficult task in urban areas as a result of difficulties of a theoretical, technical and methodological nature. Thus, in order to find out the traffic conditions that prevail on an urban road, the traffic sensors that are usually used to measure traffic conditions are almost ineffective. New measurement devices proliferation (cameras, GPS or cell phone tracking, etc.) mean that other sources of data are increasingly used in order to supplement the information provided by conventional measurement techniques and improve the accuracy of travel time estimates. As a result, travel time estimation becomes a typical DF problem.

[50], [51] and [52] discussed statement of requirements of Data fusion for ADVANCE program. ADVANCE was an in-vehicle advanced traveller information system (ATIS) providing route guidance

in real time that operated in the north-western portion and northwest suburbs of Chicago. It used probe vehicles to generate dynamically travel time information about expressways arterials, and local streets. A general framework was proposed for combining data from loop detectors and travel time reports of probe vehicles using inference rules. Evaluations of the proposed algorithms found that probe data greatly improve static (archival average) link travel time estimates by time of day.

[53] reports a detailed description of a current data amalgamation (fusion) within ITS project and the presentation of a new quantitative data fusion algorithm to estimate speed from volume and occupancy measurements.

Another work is that of [54] which presents a framework for the evaluation of ITS effectiveness, based on data fusion. Data fusion approaches are presented for various types of measures of effectiveness, and techniques for handling biases of various kinds are developed.

[38] and [40] have proposed an estimation framework of travel time based on multisource data. Two sources were considered: Data from conventional loop detectors, data from probe vehicles. A third source was added to these two travel time measurements collected by license plate matching technique. These later travel time measurements were considered as a reference to reconstitute and were used for validation purposes only. The first technique used is of statistical nature and can be viewed as a distributed estimation problem: Each source derives an estimator of travel time and the individual estimates are then combined according to weighted mean strategy. The weights were derived from variance-covariance estimation errors. Results display propensity of proposed schemes for estimation accuracy improvement. More recently, the evidence theory was used to solve the same problem ([47], [41] and [55]) In these papers, travel time was broken into classes in order to formulate the estimation problem as a classification one. Various strategies for classifiers fusion were proposed and their evaluation showed some improvements capabilities in terms of classification rate.

[56] reported work on travel time forecasting where data from automatic vehicle identification devices were used to correction purposes. Dynamic route Guidance systems are also an area where DF is of potential use. [57] has proposed a framework for fusing information from various sources within the Dynamic Route Guidance systems. Once again, the objective is Travel Time estimation and prediction. The data consists of loop detectors, probe vehicle and QoS indicator with some exogenous information: Information on road works and incidents. The proposed solution was based on a weighted mean scheme. The weights were derived according to the source reliability.

[58] and [59] have tackled the problem of generating travel time from loop detectors, probe vehicles and video-camera sources. They proposed a fuzzy logic based approach with its evaluation on a theoretical example.

4.3.3.3.2 Automatic Incident Detection

Incident detection methods for automatic recognition of incidents, accidents and other road events requiring emergency responses have existed for more than three decades. Most of the developed and implemented algorithms rely on loop detectors data. However, these algorithms work with mixed success. Recently, there has been renewed interest in incident detection algorithms partly because of the availability of new sensors and data sources. One of these sources is probe vehicles. Hence, Automatic Incident Detection (AID) belongs to the class of problems that can be solved by DF techniques.

Applications of several data fusion techniques to traffic management to support incident detection have occurred, and the data fusion algorithms used include Dempster-Shafer inference, Bayesian inference, and voting logic. Most of these applications have explored the use of probe vehicles data with the conventional traffic data for incident detection purposes. As an example of such work, Ivan et al. ([60], [61]) developed an AID system using surveillance data from two different sources: fixed detectors (e.g. inductive loop detectors) and probe vehicle specially equipped to report link travel time. The neural network approach was considered and two strategies were tested. The first one combined observed traffic directly to determine whether or not an incident occurred. In the second, separate incident detection algorithms individually pre-process data from each source, reporting scores which

are combined by neural network. Different neural network representations were studied in [61] and [62] and results found that probe and detector based incident detection on arterial networks show considerable promise for improved performance and reliability.

[63] has investigated this problem from a multiple attribute decision making point of view, with Bayesian scores. She proposed an approach which utilizes as input the combinations of probe travel times, number of probe reports, detector occupancies and volume. She pointed out that models based solely on probe data lack in performance due to excessive overlaps in class distributions. Models based on detector occupancies and vehicle counts by lane perform outstandingly. The probe data is shown to enhance the performance of detector data based models.

More recently, [64] studied the application of Dempster-Shafer inference at a traffic management centre to support incident detection and the identification of other events of concern to traffic managers. The application of the Dempster-Shafer inference algorithm to incident detection and verification is illustrated with an example consisting of three possible events, where data are supplied from three different types of sources. The available information is combined using Dempster s rule and the most probable event is identified.
Incident detection algorithms fusion is another direction for classification accuracy improvement. This direction was investigated by [65] and [66]. In this later paper three aggregation schemes were investigated: A logical aggregation, a neural network fusion and a veto procedure. From the validation step, which was carried on real-world data, results demonstrated that both logical aggregation and veto procedure outperform the single best algorithm.

4.3.3.3.3 Advanced Driver Assistance

Tremendous progress has been made with regards to vehicle safety. Early safety approaches emphasised precaution and focused on such passive devices as seat belts, air bags and lighting. In spite of crash-related injury severity rate reduction, drivers demand greater improvements in vehicular transportation safety. Research trend is to migrate from passive safety systems to active ones. Advanced Driver Assistance Systems (ADAS) and Collision Avoidance Systems (CAS) are an illustration of such a trend. Automated highways are another research topic where DF is an important area.

Both ADAS and CAS are two domains where DF is applied with some success. The main objective assigned to these systems is to provide a more reliable description of the traffic scene surrounding the vehicle in order to vulnerable road users, pre-Crash situations, Adaptive Cruise Control (ACC), Collision avoidance Systems

Operational systems are based on several sensor systems which are complementary and redundant achieving via a DF process, a fused description of the traffic scene. This fusion incorporates the data of the available sensors into a single description. The problems to solve here are:

- Data association: Sensor data have to be associated with environment description, which requires synchronization of the sensor data and associated object state.
- Multisensor fusion for tracking and estimation: Examples of such applications can be found in automated highways; namely autonomous navigation and line tracking, three-dimensional motion estimation

Several works report some results within this topic. For example, [67] discussed sensor fusion's role in vehicle guidance and navigation, and proposed general methods for fusing data, and organization of sensor-fusion activities within a robot architecture. In their work, Pei and Liou [68] proposed three-dimensional vehicle motion estimation by the fusion of multisource information. Image point and line features were considered for fusion.

[69] investigated DF systems for Automatic Cruise control with the stop and Go phenomenon, and [70] reported DF framework for obstacle detection and tracking.

4.3.3.4 Network control

Data fusion techniques were also recently applied (to the issues) in the area of road network control. Mueck [71] investigated the problem of constructing an adaptive online traffic control in urban road networks. He derived a model that determines queue length on the basis of vehicle counts from detectors located close to the stop line and on the basis of signal timings.

Along this direction, [72] introduced a new approach based on in queuing theory models for real time queue length determination. In this later method, Mueck s model serves as a quasi measurement with Kalman filtering technique.

4.3.3.5 Accident Analysis and prevention

Although there has been a steady reduction in the number of accidents, they continue to sustain heavy losses in both human and economic terms. Reduction of the number of accidents could be due to multiple efforts: Road infrastructure improvement, regulations on alcohol and speed, and an improvement in vehicle safety.

Many studies have been carried out in order to explain the circumstances and the characteristics of traffic accidents. One way to conduct such explorative studies is to utilize retrospective data available such as traffic accident records. Along this direction, Sohn and Shin ([73]) employed both neural network and decision tree algorithms to find the classification model for road traffic accident severity (bodily injury or property damage) as a function of potentially related categorical factors. They noted that the classification accuracy of the individual algorithm was relatively low. Recognizing that, Sohn and Lee ([74]) use data fusion and ensemble algorithms to increase the accuracy. DF techniques are used to perform classifiers fusion using the evidence theory.

Data ensemble combines various results obtained from a single classifier fitted repeatedly based on several bootstrap samples (see e.g. [75], [76] and [77]). More precisely, three different approaches were tried: classifier fusion based on the Dempster Shafer algorithm, the Bayesian procedure and logistic model; data ensemble fusion based on arcing and bagging; and clustering based on the k-means algorithm. Empirical results indicate that a clustering based classification algorithm works best for road traffic accident classification.

4.3.3.6 Traffic demand estimation

One of the most important problems in the field of transportation planning and control is the problem of origin-destination estimation from link counts. In order to decrease the cost of passenger surveys, traffic count are undertaken on certain links of the transportation network. An estimation of a most likely origin-destination matrix is then derived from counts. In the last two decades a greater number of models have been developed for Origin-destination (OD) estimation from link counts.
Some of the proposed schemes derived the OD matrices from data combination of different sources. An illustration of this class of problem is the dynamic OD estimation initiated by [78] and [79]. Further developments along the same line were pursued later by many authors (see e.g. [42] and [80]). The Kalman filtering is of common practice in this class of problems.

Ben Akiva and Morikawa [81] have explored the OD estimation methods that combine different data sources (stated preference data and traffic measurements).

This problem of traffic demand estimation has been considered within the INTRO project, see section 4.4.2 (page 194).

4.3.3.7 Traffic forecasting and traffic monitoring

Traffic flow forecasting has received increasing attention in the past few years and different techniques have been developed mainly for traffic surveillance and control. Many prediction schemes of traffic flow were obtained by means of classic autoregressive models, especially time series techniques. Some authors have tackled this problem in the context of Bayesian framework [82]. Some others used Kalman filtering technique [83], or neural networks and system identification [84] and more recently a nonparametric paradigm was adopted via kernel technique [85]

None of these proposals allow one to achieve highly accurate predictions except in some special situations (for some network configuration and/or with high detector coverage). This is induced to some extent by traffic dynamic which cannot be formalized by a single procedure. Therefore, in the context of traffic operations where highly accurate forecasts are needed, one can obtain different forecasts of the same quantity (the underlined assumption here is that different predictors are measures of the same quantity and/or various aspects of the same thing) by two or more different methods. The set of available methods may consist of alternative models, different forecasters, or a mixture of models and forecasters.

Often, the approach used is to find the single best predictor in some sense (most accurate values, most appropriate models of the underlying process, most cost-effective, etc) among the available forecasting methods. Another approach consists of combining these individual forecasts.

The idea of combining estimators instead of selecting the single best model has a long history and has generated intensive theoretical works since the seminal article of Bates and Granger ([86]). In their work, they showed that the linear combination of several predictors from a single data set can outperform the individual predictors, methodological and practical issues related to combining forecasts produced by different methods has been investigated extensively in various contexts with notable successes.

In traffic forecasting under heterogeneous data sources configuration, [87] has provided a methodological framework to combine various forecasts of the same quantity. He derived two predictors using a nonparametric traffic flow via kernel estimator and predicting scheme based on a propagation of a lagged upstream traffic flow. Encouraging results were achieved.

Data integration and data fusion were applied for other purposes. Cremer and Schrieber [88] studied the integration problem of in-vehicle information and data provided by loop detectors. The core of the integration step was the use of extended Kalman filtering. [89] to examine the problem of missing data estimation and proposed a framework for missing data inference based on evidential reasoning.

4.3.4 Conclusions and prospective analysis

Needs for DF has emerged from transportation applications for at least one decade and give rise to an emergent field which is somewhat in its infancy. This chapter has described the state of the art and practice of amalgamation problem of road traffic data from various sources.

For most of the applications reported in this article, DF techniques seem promising. However, these encouraging results should not conceal the difficulties that still remain to be solved before any operational widespread deployment of DF in the transportation field. These challenges include the accuracy necessary for the effective application, dynamic aspect of traffic and data quality, real time dimension.

The assessment of the benefits of DF will be more readily performed with the increasing number of successful practical applications of DF in transportation field. It is most definite, however, that there are real opportunities for greater DF applications in road transportation systems. Prospects include the increased collection of usable data from different sources other than that to installed sensors for traffic surveillance. Wireless technologies, which offer the potential for easier reporting and access to customized information, may accelerate needs for DF operational systems. The ability of tracking individual vehicles and information collected by probe vehicles will enrich the available information on traffic situations.

4.4 Data Fusion for Traffic Monitoring Indicators

Conventional loop detector surveillance systems are expensive, and tedious to install, to maintain and to operate. Hence more and more network operators and cities are moving towards emerging data collection systems, such as CCTV cameras for traffic detection, as well as newer methods using

Automatic Vehicle Identification (AVI), floating car data (FCD) and automatic vehicle classification. Using such data for traffic management purposes becomes increasingly relevant.

Additional data sources may also be available, such as detectors for monitoring weather conditions and manual data that describes major events. Hence, for providing a consistent and comprehensive picture of network conditions the complete fusion and integration of complementary multi-sources data has to be performed. The ability to use conventional as well as emerging data collection systems for the calibration and operation of traffic estimation and prediction systems may be an important driving force for these technologies and change the way traffic data is collected and used.

The approach developed within INTRO aims to develop operational methods for road operators. These methods depend on the available sources and are tested for several source configurations for different types of networks. They allow an optimal use of data from all available sources. Namely:

- Traffic data coming from loop detectors, collected in Austria, France, Poland and Sweden.
- Floating car data (FCD) and toll stamps data, collected in Austria, France and Sweden.
- Other types of sources, such as weather or road surface conditions. Available in France, Poland and Switzerland.

The added value of synergistically using and combining different data sources lies within shared costs for shared infrastructure components as well as in intelligent services taking into account available data from more than one sensor type. Indeed, combination of different sensor data (infrastructure or in-vehicle based, weather, etc.) enables the estimations of entirely new real-time safety parameters and performance indicators to be used in traffic and safety monitoring systems. These performance indicators will allow warning users in real-time of accident risks, bad weather conditions, etc. Such systems will be implemented within INTRO in future work.

A major innovative feature of the current work is that it considers fusion of sensor data from vastly different sources and time lines. This is a challenge and opens new perspectives, considering the amount of data collected each year.

In the following, data fusion is first applied for travel time and then for traffic density estimation. Most of this work has been already disseminated through recent publications. An overview is available in [90]. [91] gives a more complete version of INRETS work within INTRO on data fusion for travel time. Fusion for road safety monitoring is still under development and will be described in the D4.3 report.

4.4.1 Data fusion for travel time

This section reports all the work that has been done within INTRO on data fusion for travel time, including ISIS contribution on the first scenario. [91] contains INRETS work on the use of toll collection data for travel time: That is the work that has been done within INTRO, as well as some extensions done outside INTRO but cited in this section.

Travel time estimation and prediction problems have generated extensive research in traffic engineering literature, and with the recent advent of ITS there is a huge need for a real-time accurate travel time estimation. The proposals were drawn from traffic flow theories ([92], [93], [94], [95], [96] and [97]) or directly from flow measurements ([46], [92], [98], [99], [100], [101], [102], [103], [104], [105], [106] and [107]). This latter class is more desirable for ITS implementation ([92]). The proposed approach lies within this class as it exploits data from conventional detectors to derive travel time by combining a cumulative number of vehicle approach, and data from electronic toll collection (ETC) system in order to improve the whole performance of estimation process. More specifically, the main objective of this work is to develop an efficient algorithm for motorway travel time estimation using information provided by traffic surveillance system. The data considered in this section is mostly from inductive loop detectors deployed along the highways and from ETC data, [108].

Two different sources of data are considered here: Traffic data collected from loop detectors, and ETC data for motorway networks. This data (traffic and toll data) has been provided by the toll motorway company Autoroute Paris Rhin Rhône. Toll data is a very rich source data as far as each vehicle is a

probe: The toll ticket is stamped with the current time when it is taken at the entrance barrier and when it is given back at the exit barrier. From this source one can get time spent in motorway sections by all vehicles and via specific designed filtering algorithms, one can derive travel time. Although travel times from toll data are only known at the time when vehicles exit the motorway, it can be used in real time to enhance the travel time estimation. It can also be used off-line as a reference for the evaluation of travel time estimation and for motorway operation assessment.

This section is divided in 4 subsections. Section 4.4.1.1 describes the correction process integrating ETC data into the Cumulative Traffic Counts method. Section 4.4.1.2 describes the enrichment process that enables to get a reference travel time, taking into account the data from all tolls. The results of the two previous processes are given in Section 4.4.1.3, on 4 scenarios, using real-world data. Section 4.4.1.4 gives the final conclusions of this work.

4.4.1.1 Travel time estimation and correction strategies

4.4.1.1.1 Travel time estimation

The main objective of this work is to develop an efficient algorithm for motorway travel time estimation using information provided by traffic surveillance system. The classical cumulative traffic counts method (CTC) is used here as a basic framework for travel time estimation and the effort focuses in developing innovative schemes for improving accuracy of CTC method using ETC data.
This method has been introduced in Chapter 2, as the Stock Method (Section 2.5).

4.4.1.1.2 Correction Strategy Using ETC Data

One way to perform the correction process within CTC method is to combine, in a projective manner, speed extrapolation method (travel time method from dual loop speeds) and CTC outputs. More precisely, acknowledging the fact that the velocity extrapolation technique gives a quite good estimates of travel time only under homogenous traffic conditions assumption (see an illustrative example of such a situation in[99]) and that CTC method performs fairly well under congestion conditions ([93], [94], [95], [96] and [97]), Ben Aïssa et al. [110] proposed an automatic switching algorithm according to the prevailing traffic conditions.

The reported work here puts forward another step by extending the previous work and designing, in a unified framework, a corrective strategy of CTC method to estimate more accurately travel time for toll-operated networks. Hence, this method is called the Modified Cumulative Traffic Counts (MCTC). The correction process uses ETC data measured for the section under consideration during the same time period and operates differently according to the type of the section. More specifically, the correction process is designed to cope with the following situations:

1. *S1-type: section without on-ramp or off-ramp, and without rest or service areas*. This is in fact the perfect situation where the simple CTC performs well, provided that loop detector data is valid. MCTC will be of help when the data is not valid, by reconstructing the correct data from ETC observations.

2. *S2-type: section with only rest or service areas.* This situation is more challenging, as rest or service areas act as small entries and exits: A vehicle making such a stop is still counted as present in the section by CTC. If several vehicles stop when the flow is low (for instance at night), CTC will detect a false congestion. Conversely, if CTC is initialized at night (free flow), CTC might predict very low (or even negative) travel times in the morning, as the vehicles leaving the rest areas were not counted in the initialization. MCTC copes with rest and service areas, maintaining a minimum travel time (equivalent to speed extrapolation) when CTC would predict a value too low, and returning to the minimum travel time (still equivalent to speed extrapolation), when CTC is not consistent with an observed low ETC. Note that S2 can also be thought as S1 where the loop detector data is not always valid.

3. *S3-type: section with on-ramps and off-ramps.* This situation is more complex, as the entries and/or exits can be of much greater importance than in S2. Hence, it is necessary to use ETC

data to return to a situation similar to S1 or S2. Yet, destination is not known in advance, not even the direction: The vehicle after the toll may enter the network in one direction or the other. In [111], the direction is forecasted from historical data on similar type of days. Another solution would be to enable MCTC to modify the loop detector data from ETC data also during congestion. This will be studied in future researches. This is a more complex procedure but should be much more practical for S3, as it does not need historical data and constantly adjusts CTC to ETC data.

The current correction process operates in real time (with the same time step as the loop detectors) during free flow traffic conditions. Basically, the correction process acts dynamically on cumulative curves (raising or lowering one curve, the other or both), so that the estimated travel time from these curves is as close to actual travel time as possible. The correction is based on the three following considerations, leading to the MCTC method:

1. CTC should not lead to unreasonably low (or even negative!) travel times. So, when the CTC is likely to derive such travel times, it is corrected accordingly.

2. CTC should not lead to unreasonably high travel times, when much lower travel times are found in the ETC data (during the same time period). ETC data are then used to adjust any drift in the standard CTC method.

3. During stable free flow traffic, the number of vehicles in the section can be estimated from loop detector data. The CTC estimation of this number has to be consistent with the loop estimation.

Hence, the correction process combines those considerations in order to adjust any drift in outputs of the standard CTC method. Such process is done dynamically and has proved to be efficient.

The selected scenarios in INTRO are sections of S2-type of the APRR network. Part of this work has been published [90].
Results for scenarios of the other types have been obtained on another motorway [111].

Here, the scenarios concern three lane motorway sections and are presented in Fig. 4.9.

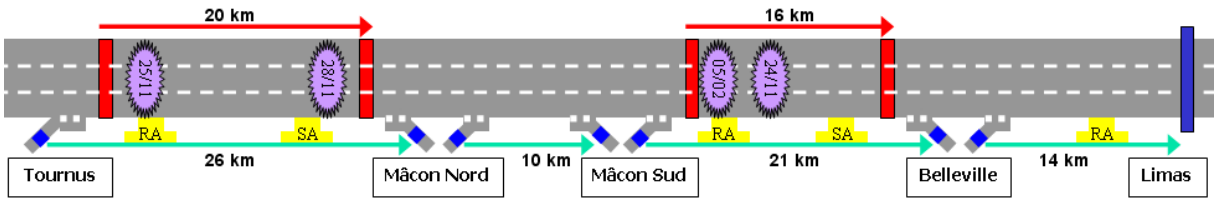


Fig. 4.9: Scenarios 1-4

Loop detectors are the rectangles on the road. Rest Areas (RA) and Service Areas (SA) are also represented, as well as main toll gates (for instance Limas) and on/off-ramp tolls. The purple zone represents the location of the congestion in each scenario.

The selected S2-type section scenarios are numbered according to the day of the scenario where congestion was observed on the test site shown on Fig. 4.9. Scenario 1 (05/02/2005) and Scenario 2 (24/11/2005) from Mâcon Sud to Belleville. This route is approximately 21 km long and contains a rest area and a service area in-between the two loop detectors (see right part of Fig. 4.9). Scenario 3 (25/11/2005) and Scenario 4 (28/11/2005) from Tournus to Mâcon Nord. This route is approximately 26 km long and contains a rest area and a service area in-between the two loop detectors (see left part of Fig. 4.9). In both cases, the congestion is only in-between the loop detectors, hence the total travel time can be computed straightforwardly from the MCTC travel time in-between the two loops and a free flow travel time outside the loops.

The previous correction process is tested and evaluated using real-life data. Reference travel time is derived from historical ETC data and enrichment process for the considered route section, see below (Section 4.4.1.2).

Travel time estimators are compared, using the Root Relative Squared Error (RRSE) of the different TT estimators, both for congestion and free flow traffic. This performance measure is defined as follow:

$$RRSE_m = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i^{(m)} - z_i\right)^2}{\sum_{i=1}^{n}\left(x_i\right)^2}} \qquad (4.7)$$

where $y_i^{(m)}$ is the i^{th} value of predicted TT by method m, x_i is the actual travel time and z_i is either x_i (during congestion) or given by following expression (during free flow traffic):

$$z_i = \begin{cases} x_i & \text{if} \quad y_i^{(m)} \in CI\left(x_i\right) = \left[x_{\min}^{(i)}; x_{\max}^{(i)}\right] \\ x_{\min}^{(i)} & \text{if} \quad y_i^{(m)} \le x_{\min}^{(i)} \\ x_{\max}^{(i)} & \text{if} \quad y_i^{(m)} \ge x_{\max}^{(i)} \end{cases} \qquad (4.8)$$

For a perfect fit, the numerator of Eq.(4.7) is equal to 0 and RRSE = 0. So, the RRSE index ranges from 0 to infinity, with 0 corresponding to the perfect match between method s output and actual travel time.

4.4.1.2 Reference travel time derivation from toll stamps – Enrichment process

Before any fusion strategy for travel time estimation, a reference travel time is necessary in order to compute and to compare performances of travel time estimations. This reference can be computed off-line from toll data collection by suitable filtering and smoothing strategies. Indeed, there can be misleading data in the toll data collections as the total time is rather time spent on the network (including stops at rest areas) rather than travel time. Hence, vehicles stopping for too long have to be filtered as they do not represent traffic dynamics such as congestion.

The computation of the reference travel time presented in this section will make use of an innovative enrichment process. This process can be applied to any network with tolls. Here, it is exemplified on scenario 2. In order to make the explanation more general, tolls have been renamed A, B and C on Fig. 4.10.

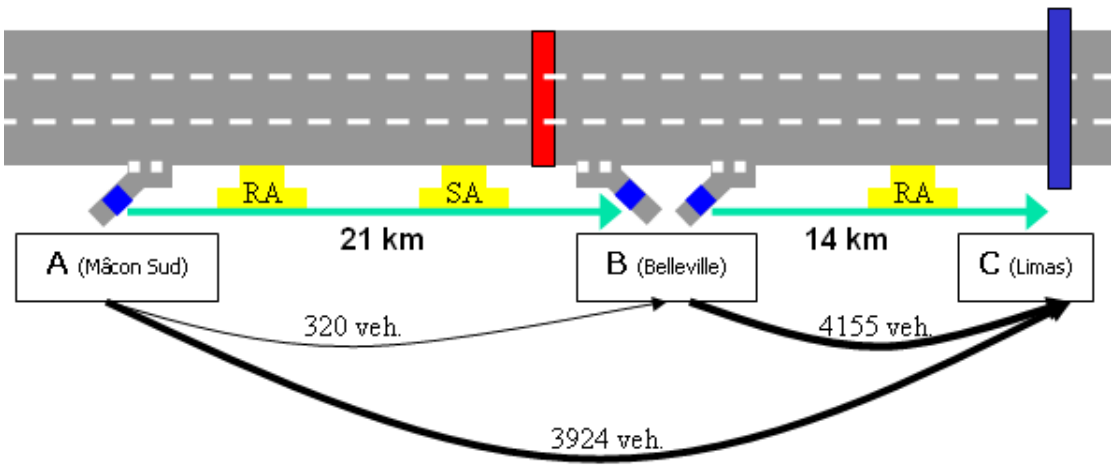


Fig. 4.10: Origin-Destination counts from 0:00 am to 12:00 pm, scenario 2, 24/11/2005

As shown on Fig. 4.10, A and B are on/off-ramp tolls and C is a main toll gate. Yet, the same process would apply even if C was an on/off-ramp toll. Rest Areas (RA) and Service Areas (SA) are also

represented. The red rectangle is a loop detector located slightly before toll B. Finally, the number of vehicles is given for each link on November, 24th 2005 (scenario 2, whole day: from 0:00 am to 12:00 pm). The aim here is to compute a reliable reference travel time from A to B, using toll stamps.

Fig. 4.11a shows the result of the filtering and smoothing procedure on the direct link travel time. As is also shown, the 70 % confidence interval is rather large, especially during congestion. This is due to the fact that few vehicles made this precise link (entry in A and exit in B) during the day: only 320 vehicles.

Yet, many more vehicles went from A to C (3924) and from B to C (4155), see Fig. 4.10. As the reference travel time can be computed off-line, it is possible to apply the smoothing and filtering procedure to the two links above and deduce the travel time from A to B. As shown on Fig. 4.11b, the 70 % confidence interval is much thinner during congestion. Thus, this travel time shape is closer to reality and is used as the reference.

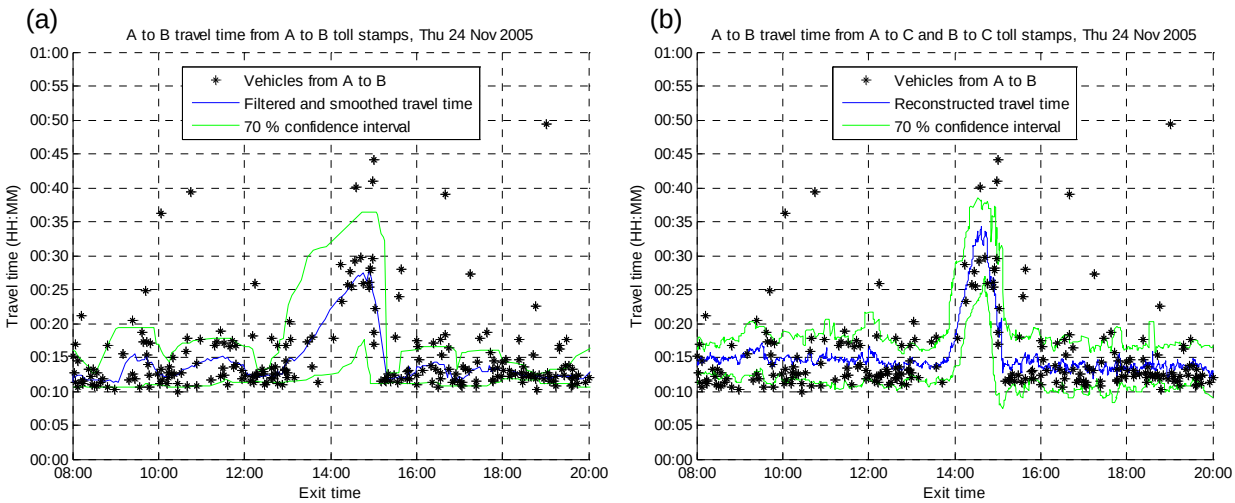


Fig. 4.11: From toll stamps to a reference travel time

The same process applies to any other toll C before A, in-between A and B, or after B. Yet it is not useful to try other links in the case of scenario 2, as A to C and B to C are the only links with several thousands of vehicles. However, for other cases, travel time using C and travel time using C can be fused using their confidence intervals, leading to a more precise reference. More generally, the fusion can use all the C of the network.

Note that the enrichment process has to use link travel time by vehicle category in order to avoid bias due to differences in traffic composition (vehicles that enter in A and exit in C, vehicles that enter in B and exit in C, vehicles driving on the network between A and B, whatever their entry and exit tolls). Hence, the reference travel time has first been computed by enrichment for each vehicle category (light vehicles on the one hand and heavy vehicles on the other hand), as presented in Fig. 4.12. Then the references by category have been merged in a single reference (Fig. 4.11b), using the loop detector information on the traffic composition driving on the network before the exit (red rectangle in Fig. 4.10). This fusion process guaranties that the computed travel time can be taken as the reference experienced travel time.

It is this enrichment process that has been applied in the 4 scenarios in order to compute the reference travel time.

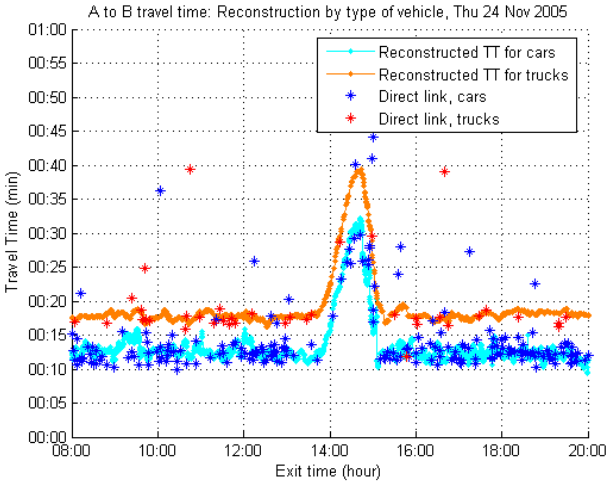


Fig. 4.12: Reference travel time by vehicle category

In Fig. 4.12, it is interesting to note that free flow travel time is more stable for trucks (between 16 and 19 minutes) than cars (between 9 and 16 minutes). The average difference is about 6 minutes (on this 21 km section). It is interesting to see that this difference is kept during congestion. This is because most of the section experienced free flow: Only a small part of it (less than 10%) suffered from congestion, as will be shown later in scenario 2 (Section 4.4.1.3.2.1). Actually, the travel time difference between trucks and cars seems even more than 6 minutes during most of the congestion, which could be explained by the differences in the dynamics of the two categories of vehicles, as it takes trucks more time to reach their free flow speed.

4.4.1.3 Results and Comparative Analysis

4.4.1.3.1 Scenario 1: Mâcon Sud -> Belleville, 05/02/2005

4.4.1.3.1.1 Scenario description

In this scenario, the loop detectors have experienced losses, especially the upstream detector. When no value is given, the last known value is kept; except at the beginning, when no previous measurement is available. Fig. 4.13 and Fig. 4.14 give the flows and speeds (respectively) which have been used for CTC and MCTC.

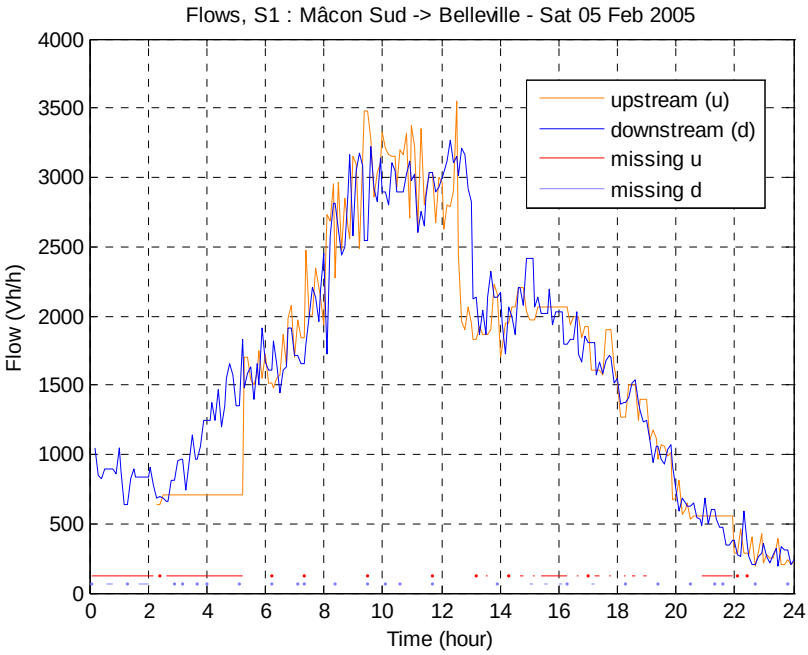


Fig. 4.13: Loop detectors flows, scenario 1

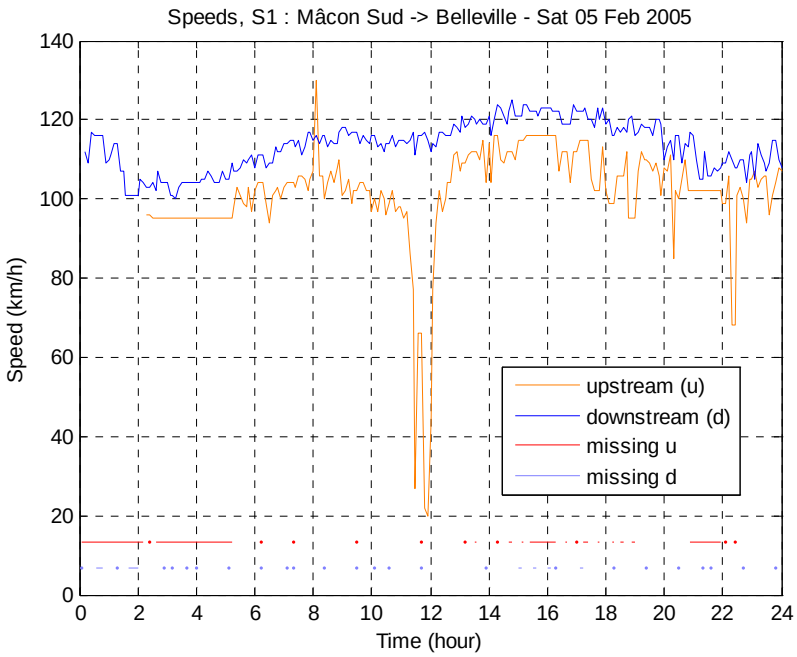


Fig. 4.14: Loop detectors speeds, scenario 1

As stated in the patrol based data, congestion appeared around 10:00 am and lasted 3 hours, see the red bars in Fig. 4.15. This congestion is located 3.7 km after the upstream detector (see also Fig. 4.10). The height of the red bars corresponds to the number of km of congestion (also written in blue in Fig. 4.15).

For instance, the length of the congestion was measured 3.9 km just before noon, which means that the congestion backed up to the upstream detector. This information is perfectly coherent with the upstream speed observed at that time in Fig. 4.14.

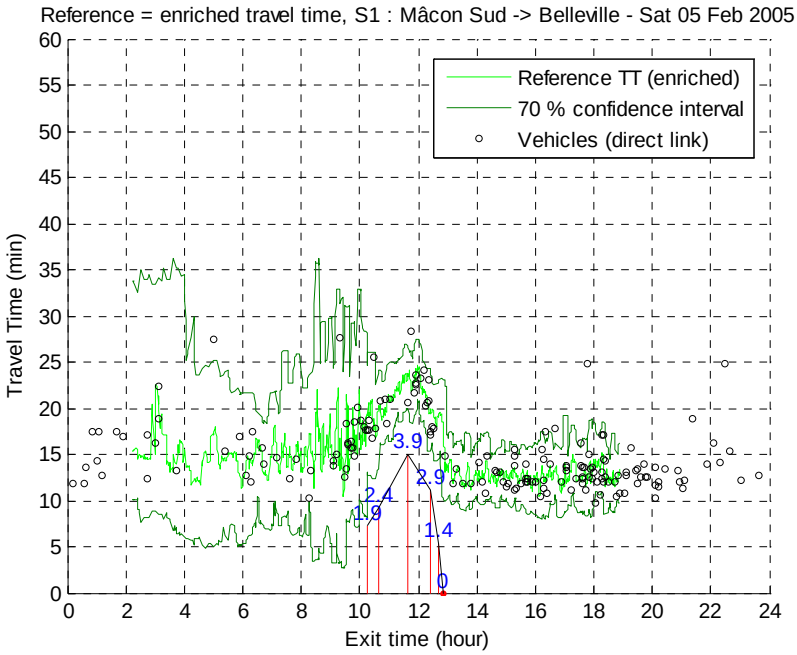


Fig. 4.15: Enriched reference travel time compared to direct-link data, scenario 1

Fig. 4.15 also gives the reference travel time, computed through the enrichment process (see Section 4.4.1.2). As shown on this figure, the reference is coherent both with the direct-link data and with the patrol based data. Yet, the confidence interval is rather wide at the beginning of the day, so the reference will be used for assessment purposes of travel time estimators only later in the day, when its reliability is better (10:30 am 7:00 pm, see also Fig. 4.20).

4.4.1.3.1.2 Speed Extrapolation (SE)

Fig. 4.16 gives the estimation of the travel time, based on the classical Speed Extrapolation (SE) method: The average speed on the section is estimated from the detectors, resulting in the computation of a travel time. This estimation is very good during free flow. Yet, as the density of speed detectors is low (16 km between the 2 detectors), this estimation is quite far from the reference during congestion. In fact, the congestion starts and ends in-between the two detectors. So its beginning and end have no effect on the detectors, resulting in wrong free flow travel time estimation. The upstream detector is affected only when the congestion goes back up to it (between 11:00 am and noon, see Fig. 4.14). At that time, the upstream detector might give the average speed of the congestion, but the congestion is only 4 km long, while the section is 16 km long. So the average section speed is biased, resulting in travel time estimation much higher than reality.

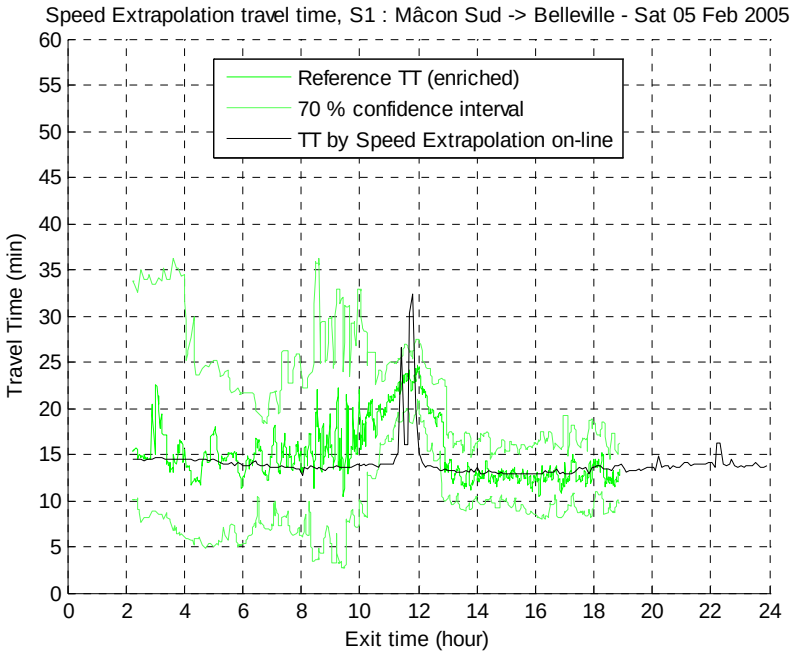


Fig. 4.16: Speed Extrapolation scenario 1

4.4.1.3.1.3 Cumulative Traffic Counts (CTC)

Since speed extrapolation cannot be applied when detector density is low, Cumulative Traffic Counts (CTC, also known as Stock Method) can be tested as an estimator alternative, which is also classical. Yet, it is made for sections with no entry or exit. Here, there is a rest and a service area, which can lead to perturbations. Indeed, the initial stock is more difficult to guess, as it has to take into account the vehicles parked in the rest and service areas. Otherwise, when they leave, the stock might be negative! In order to simplify initialization phase, the initial stock is computed off-line, using the day data, so that the stock remains positive.

As a result, Fig. 4.17 definitely shows a CTC travel time that is always greater or equal to the free flow travel time. It has also identified the congestion period, although it is overestimated by 5 minutes for a free flow travel time of 13 minutes (i.e. 40%). But the night estimations are even more discarding this method. The explanation is simple: Many vehicles are parked in the rest and service areas for the night. Besides, not so many vehicles are driving at that time. So the stock is high and the flow is low, resulting in unlikely high travel time estimation (Eqs. (2.2) and (2.3)). CTC not only fails at night and has low performances during the day, but this result could be obtained only off-line, because stock initialization is not possible on-line.

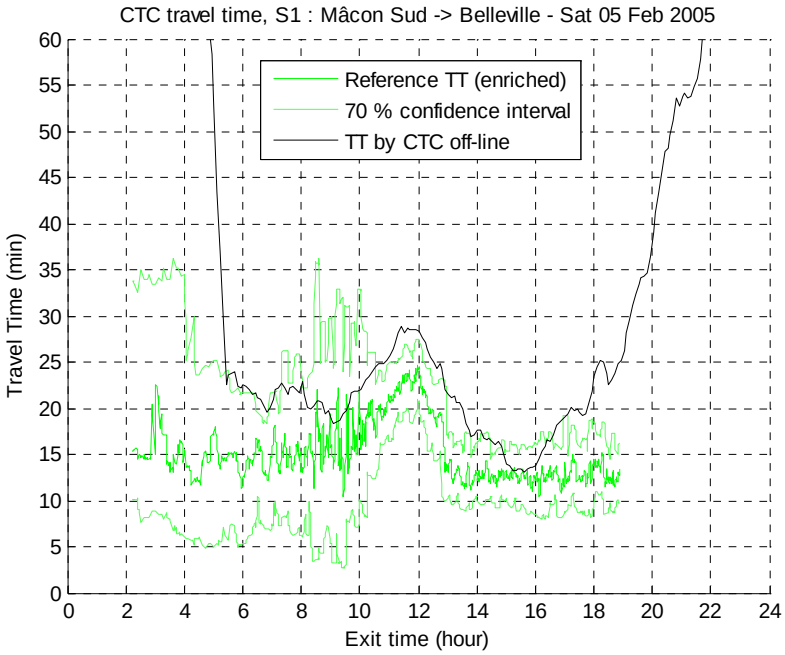


Fig. 4.17: CTC (off-line) compared to the reference, scenario 1

4.4.1.3.1.4 Modified Cumulative Traffic Counts (MCTC)

As explained above, both classical travel time estimators (speed extrapolation and cumulative traffic counts) fail. Yet, cumulative traffic counts only failed because it does not take into account the rest and service areas in order to compute the real number of vehicles that are driving in the section. Speed extrapolation would need supplementary sensors expenses. Yet, the modified cumulative traffic counts (MCTC) leads to fair results (see Fig. 4.19) without adding any sensor. Apart from the cumulative traffic counts, it only uses ETC data to regularly re-estimate the number of driving vehicles (and those who stopped at rest or service areas), as shown on Fig. 4.18.

Here, MCTC started only at around 8:00 am, due to the missing loop detector data at the beginning of the day (see Fig. 4.13). Initialization also needs some stability in the measured data.

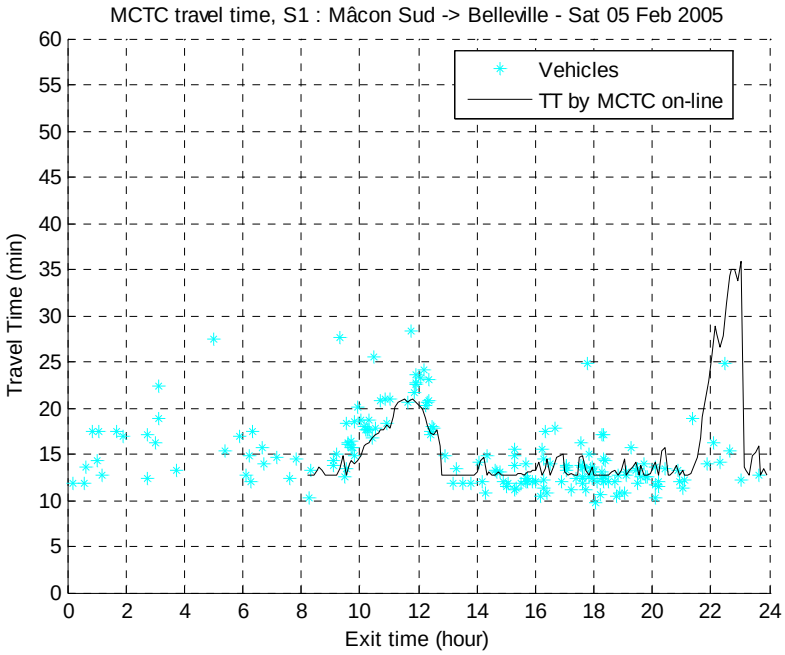


Fig. 4.18: MCTC (on-line) using direct link data, scenario 1

Fig. 4.19 shows that MCTC is quite close to the reference travel time. The only real problem is at night between 9:00 pm and 11:00 pm. Although no reference is available at this time of the day, such a high travel time is not expected. The speed extrapolation, Fig. 4.16, seems to have spotted something (small) around 10:00 pm, but it is probably really something small. This shows a limit of MCTC: The stock cannot be recomputed because there are only very few ETC data and the only data available indicate that free flow conditions do not hold (around 10:00 pm), Fig. 4.18.

In order to solve this problem, enriched ETC data could be used, in order to have more data. Moreover, MCTC could be extended in order to modify CTC not only in free flow conditions buts also in congestion. This cannot be done in the timeline of INTRO but should be investigated in the future.

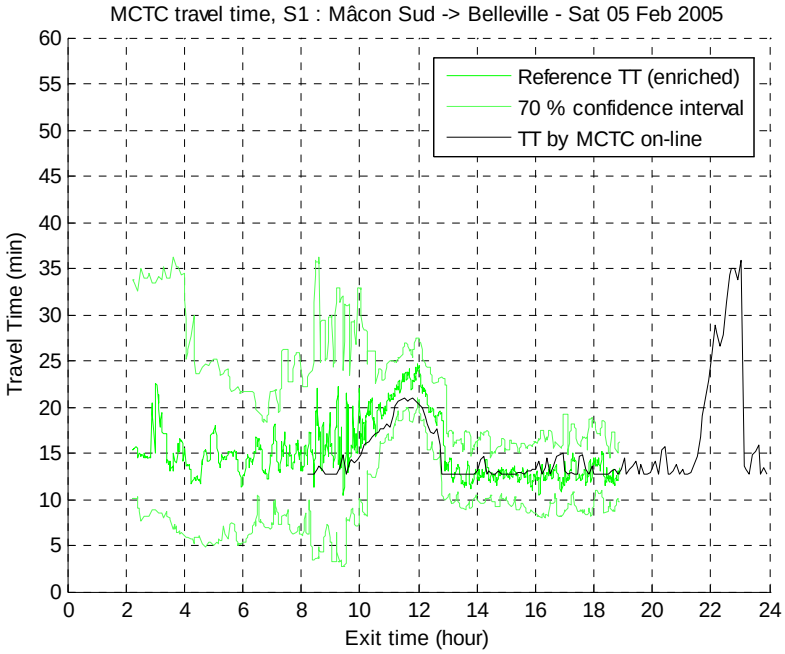


Fig. 4.19: MCTC (on-line) compared to the reference, scenario 1

4.4.1.3.1.5 Results

Fig. 4.20 shows the different estimators for travel time, the reference and the time periods used to compare estimators to the reference in congestion and in free flow conditions.

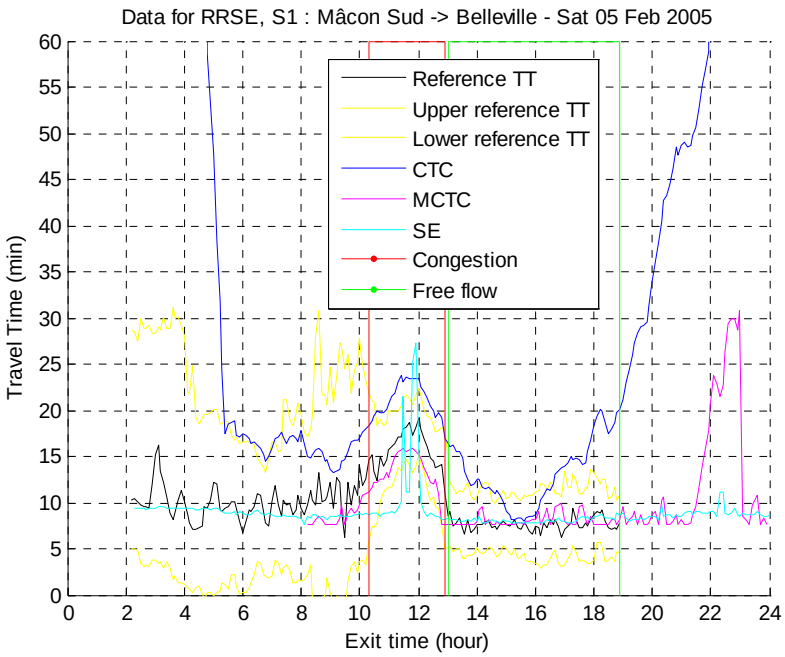


Fig. 4.20: Travel time estimators: Comparisons in congestion / free flow, scenario 1

As summarized in Table 4-1, RRSE-results for the time periods given in Fig. 4.20 show that CTC is as bad as SE in congestion and equally bad in free flow. If the reference had been available at night, CTC would have been totally discarded in free flow. SE and MCTC are very good in free flow. MCTC would not have performed so perfectly, if the night reference had been available (as explained above),

but it would not have been discarded. This sole false alarm can be treated manually. As for congestion, MCTC performs much better than the others (factor 2 at least).

Table 4-1: Travel time estimators: Comparisons in congestion / free flow, scenario 1

RRSE	Congestion	Free flow
CTC	0.342	0.467
MCTC	0.175	0.000
SE	0.405	0.000

Hence, MCTC has shown most reliable to detect beginnings, ends and levels of congestion, with only one alarm that was too high. As explained above, MCTC can be improved to be able to use more ETC data (enrichment) and to use it also during congestion.

4.4.1.3.1.6 Results of the integrative fusion applied to ISIS combined method

Here, we can present the results achieved by fusion applied to ISIS combined method implemented in the TPI tool on the same scenario (for the description of the combined method, refer to Chapter 2).

As already mentioned above, two different kinds of fusion applied to travel time estimation can be developed:

- First, an *aggregation approach* aiming to derive an enhanced estimation from the competing estimation schemes.
- Second, an *integration approach* trying to correct bias of one method by integrating output of another method.

Along the first direction, two different approaches can be elaborated:

- A projective approach which is able to choose at any moment the best of two (or more) estimators and then produce a more robust travel time less dependent on data quality (such as breakdown of data collection stations, insufficient number of toll travel time data) than a single source estimator. Such an approach was applied to travel time estimation problems through a switching procedure using two travel time estimation algorithms and both ISIS and INRETS elaborated switching processes based on traffic patterns and/or quality indices [110]. ISIS recently implemented its own method on a motorway network in France [112].
- An aggregative approach in which constant weight (at least for linear strategy) is assigned to each estimator over the whole range of input space to minimize the overall error. The main advantage of such a strategy is that errors from different estimators may cancel out one another [113].

Within INTRO, the stress was therefore put by ISIS on the integration approach. More specifically, the approach followed here is to correct the stock method (or its hybrid version) through the use of toll travel times.

As a matter of fact, although special automatic updating procedures of the stock are already used in the combined method to prevent long term drifting (the drawback of the stock method because of data collection measurement inaccuracies), there is room for further enhancement through the use of the additional information procured by the travel times of the vehicles currently exiting the motorway.

The idea is to calculate what should have been the value of the stock in the past (had the stock travel time been equal to the toll travel time) and then update the stock from this time in the past onto the current time. Once validated, this correction procedure will be incorporated into the real-time estimation TPI tool already deployed on APRR motorway network, [114]. Of course, before obtaining toll travel times, a filtering and smoothing process has to be used, and the number of individual travel times has to be sufficient for statistical robustness of the travel time estimation. ISIS has developed a

simple filtering method together with a general algorithm generating toll travel times on any itinerary by combination of both an adding process (the itinerary is made of several sub-itineraries added) and a subtracting process (the itinerary is made from encompassing itineraries from which exceeding parts are subtracted, as explained above in the enrichment process). These developments are currently in operation on the AREA toll company network since the beginning of Year 2006.

To assess the implemented correction approach, several scenarios elaborated from several real situations have been tested. It follows that the correction approach has proved to be very efficient.

The figures below show the example of the results obtained on scenario 1:

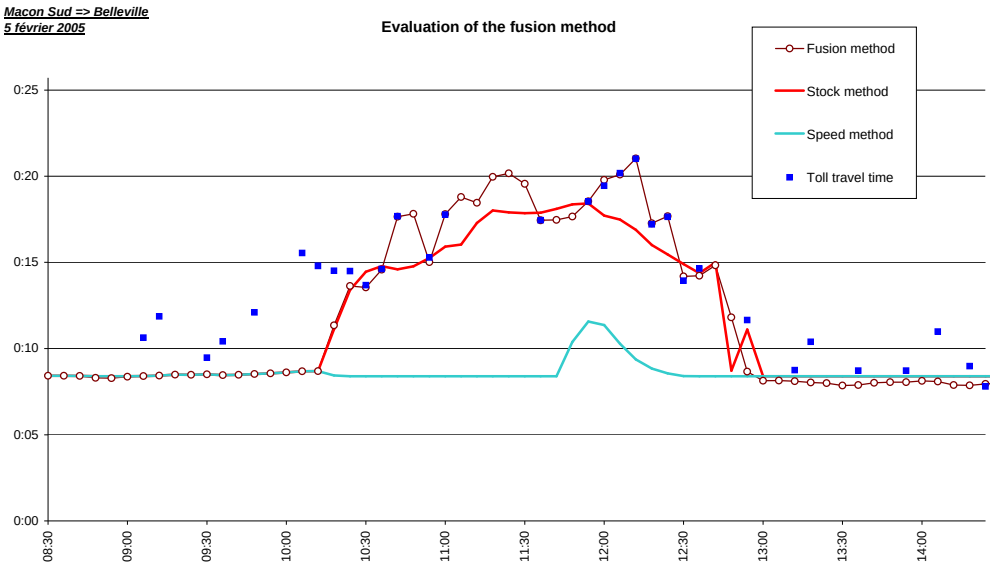


Fig. 4.21: Effect of fusion on on-exit travel times, scenario 1

- Reference travel times are the toll ticket travel times (blue dots represented at their exit times).
- The speed method (bottom curve) fails to calculate the travel time: it only reacts when the congestion steps back onto the upstream loops.
- The stock method starts a bit late and presents a rather satisfactory shape, although it can be seen somewhat below the actual travel times.
- The fused stock method (curve linking circles) show the exact travel times at times when the latter are used for fusion (blue circles) and keeps this updating between these times (see the five successive dots between 11:00 and 11:30).

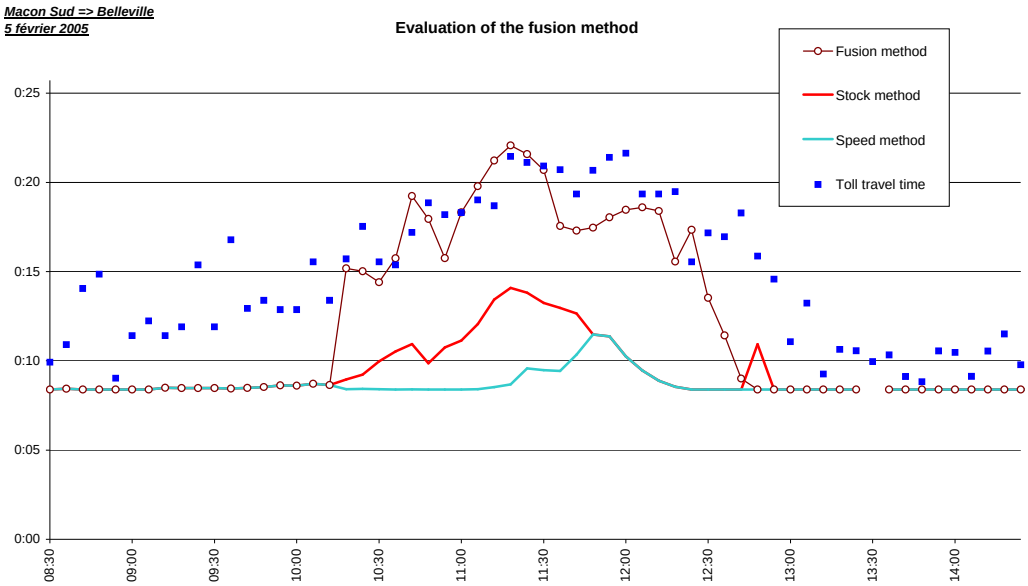


Fig. 4.22: Effect of fusion on on-start travel times (real time travel time), scenario 1

- The reference travel times are plotted at their departure times; so they represent the actual times for drivers entering the section at that time.
- From the automatic (late) start of the stock method (10:12 am), each reference time is used for updating the stock; the enhancement is evident when compared with the stock method which underestimates the actual travel times (probably because the initial stock at 10:12 is underestimated).
- The stock method switches to the speed method at 11:48 because at that time the congestion steps back onto the upstream station and because the volume measurements in these conditions are considered too inaccurate to keep the stock method. This is not a problem when the fusion with toll tickets is used: the stock is entirely estimated from the toll ticket travel times and the stock travel time is then deduced using the downstream traffic volume (which corresponds to free flow conditions).
- Furthermore, the knowledge of toll ticket travel times allows launching the stock method as soon as these travel times exceed a given threshold. In this way, the stock method would have started much sooner.

In fact, the conditions of flow at the beginning of this morning were stop-and-go style, along hundreds of kilometres, and the threshold system of detecting congestion (with the waiting of confirmations) failed to detect these conditions. In addition, this probably explains why the initial stock was underestimated (because the speeds on the stations, which are used for the estimation of the average density along the section, were probably high at the very time when the stock was initialised).

The general conclusion about the fusion of the stock method and the toll tickets measurements is that there is a very significant potential of improvement of the stock method in order to get a more precise and robust estimation of travel times: at any time (even in case of traffic data measurement failure) and in any section (open or closed) provided there is a sufficient number of toll tickets available.

4.4.1.3.2 Scenario 2: Mâcon Sud -> Belleville, 24/11/2005

4.4.1.3.2.1 Scenario description

In this scenario, the loop detectors have experienced only very few losses, as shown on Fig. 4.23 and Fig. 4.24. So the flows and speeds which have been used for CTC and MCTC are mostly real measurements. In the same scope, the reference travel time (Fig. 4.25) could be computed earlier than in scenario 1 (Fig. 4.15).

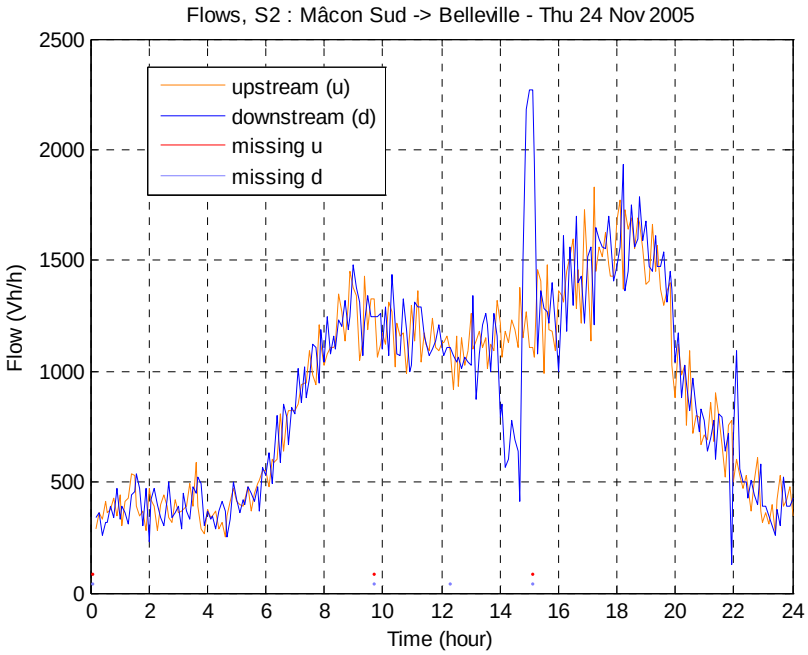


Fig. 4.23: Loop detectors flows, scenario 2

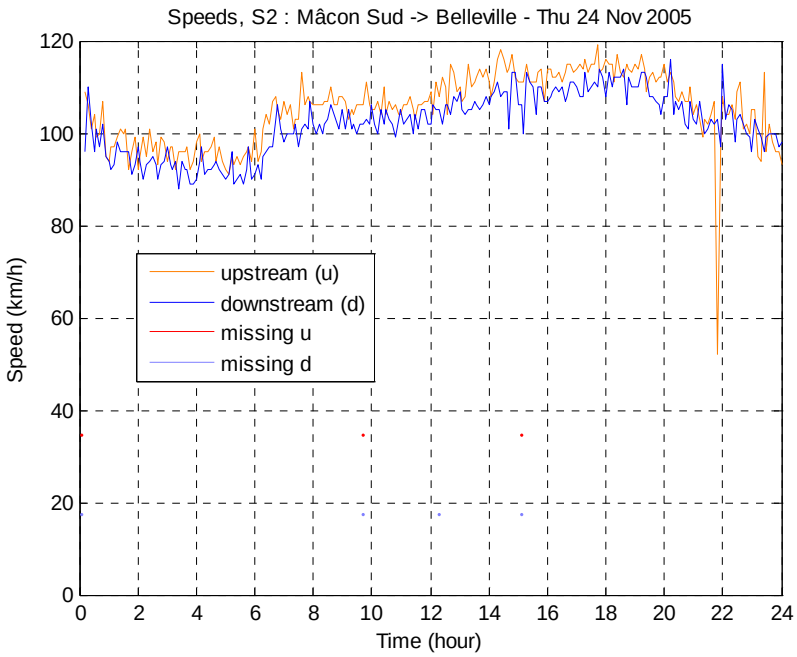


Fig. 4.24: Loop detectors speeds, scenario 2

As stated in the patrol based data, congestion appeared around 2:00 pm and lasted 1 hour, see the red bars in Fig. 4.25. This congestion is located 6.8 km after the upstream detector (see also Fig. 4.10). The height of the red bars corresponds to the number of km of congestion (also written in blue in Fig. 4.25).

For instance, the highest length of the congestion was measured 2.1 km around 2:20 pm, which means that the congestion was never located on a detector (upstream or downstream). This information is perfectly coherent with the speeds observed on Fig. 4.24: Free flow speed was measured on the detectors during the congestion period. Yet, the congestion can clearly be identified

on the flows (Fig. 4.23). This is why Speed Extrapolation should fail and why Cumulative Traffic Counts might be successful.

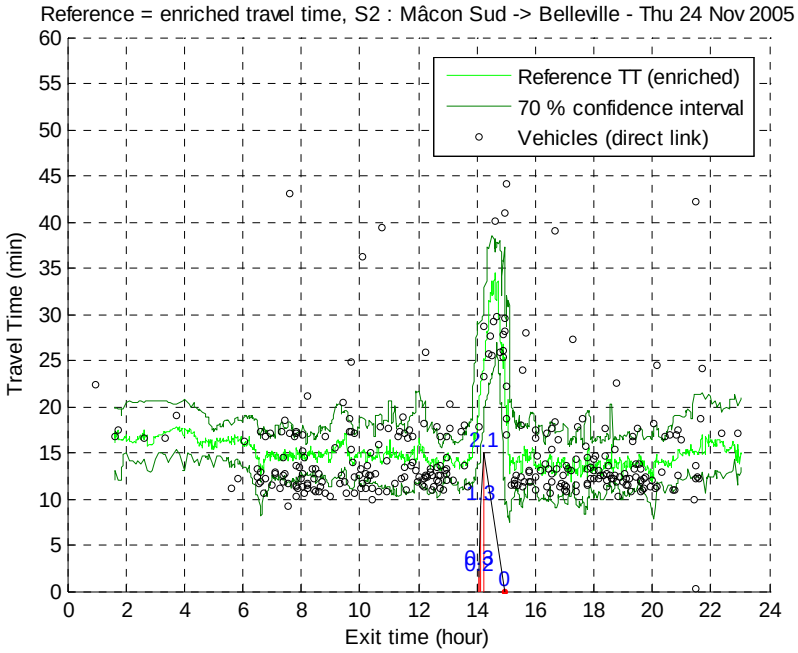


Fig. 4.25: Enriched reference travel time compared to direct-link data, scenario 2

Fig. 4.25 also gives the reference travel time, computed through the enrichment process (see Section 4.4.1.2). As shown on this figure, the reference is coherent both with the direct-link data and with the patrol based data. Besides, the confidence interval is excellent throughout the day. So this scenario is an excellent case for comparing the different travel time estimators (see also Fig. 4.30).

4.4.1.3.2.2 Speed Extrapolation (SE)

Fig. 4.26 gives the estimation of the travel time, based on the classical speed extrapolation method. Yet, due to the low density of speed detectors, this method sees free flow (almost) everywhere and misses the congestion. It only detects a slight problem just before 10:00 pm, which is coherent with the reference, but can still be considered as free flow (only a 4 minutes delay).

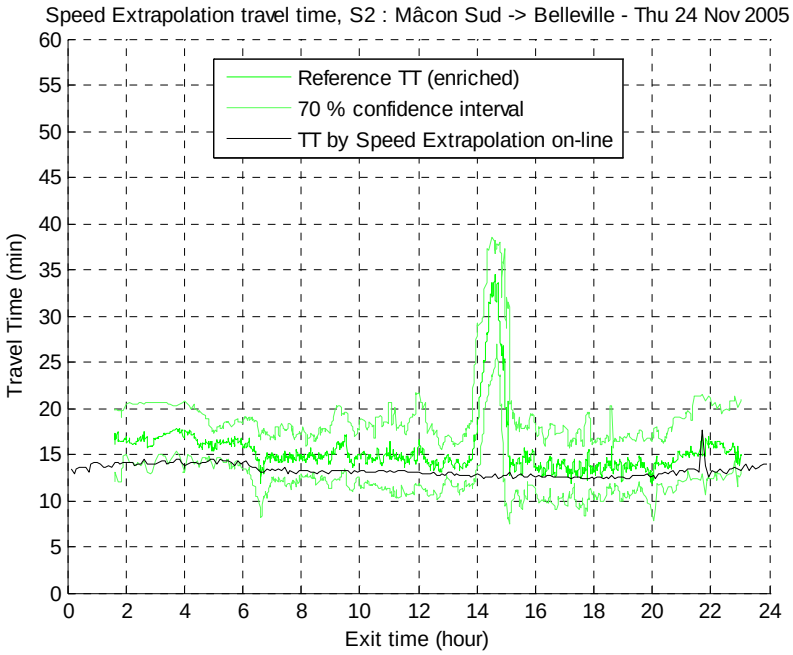


Fig. 4.26: Speed Extrapolation scenario 2

4.4.1.3.2.3 Cumulative Traffic Counts (CTC)

As in scenario 1 (see Fig. 4.17), Fig. 4.27 shows a CTC travel time that has identified both the congestion period and its level for scenario 2. In fact, CTC performs even better during the day in scenario 2 than in scenario 1. Yet the night estimations are still discrediting: CTC can definitely not deal with rest and service areas. Besides, CTC is still only available off-line.

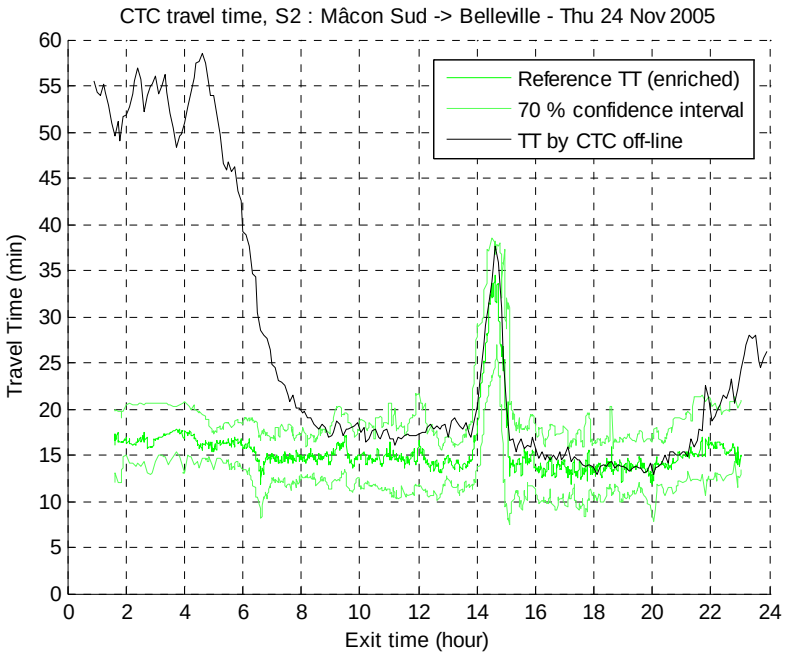


Fig. 4.27: CTC (off-line) compared to the reference, scenario 2

4.4.1.3.2.4 Modified Cumulative Traffic Counts (MCTC)

Here also, as in scenario 1, MCTC leads to fair results (see Fig. 4.29). It is even better than in scenario 1, as no false alarm was raised at night. This is because there are more vehicles on this day than on the day of scenario 1 (compare number of blue stars in Fig. 4.18 and Fig. 4.28).

Here, MCTC also started at the beginning of the day, as the loop detector data is better in scenario 2.

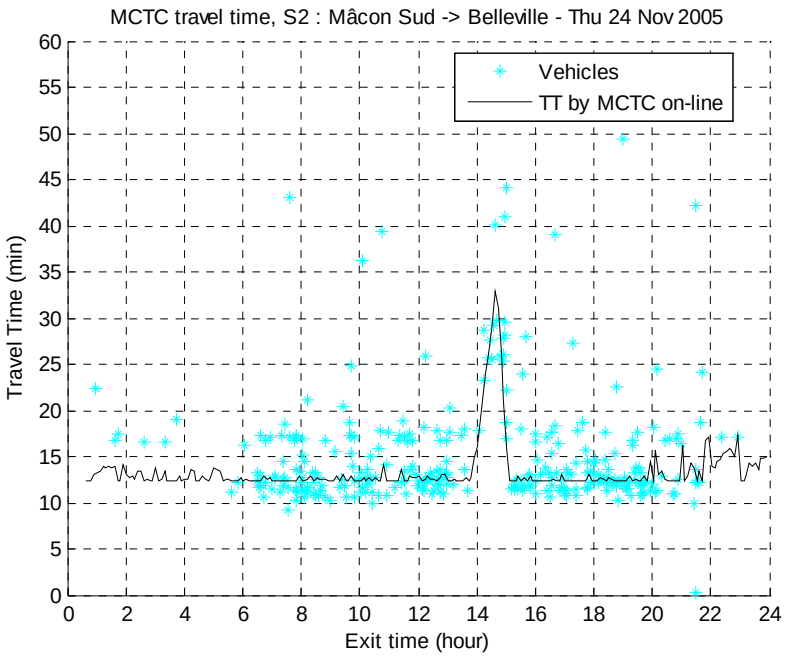


Fig. 4.28: MCTC (on-line) using direct link data, scenario 2

Fig. 4.29 shows that MCTC is really close to the reference travel time. The only slight problem is that it tends to underestimate free flow travel time at the beginning of the day. This is because there are mainly heavy vehicles at this time on the network and MCTC is initialized with the free flow travel time of light vehicles. This small bias will be corrected in the future as in the case of enrichment, using traffic composition to compute the current free flow travel time.

The speed extrapolation, Fig. 4.26, seems to have spotted something (small) around 10:00 pm, which is also identified by MCTC and coherent with the reference. In fact, the reference shows a small increase in free flow travel time after 8:00 pm, which is better identified by MCTC than speed extrapolation.

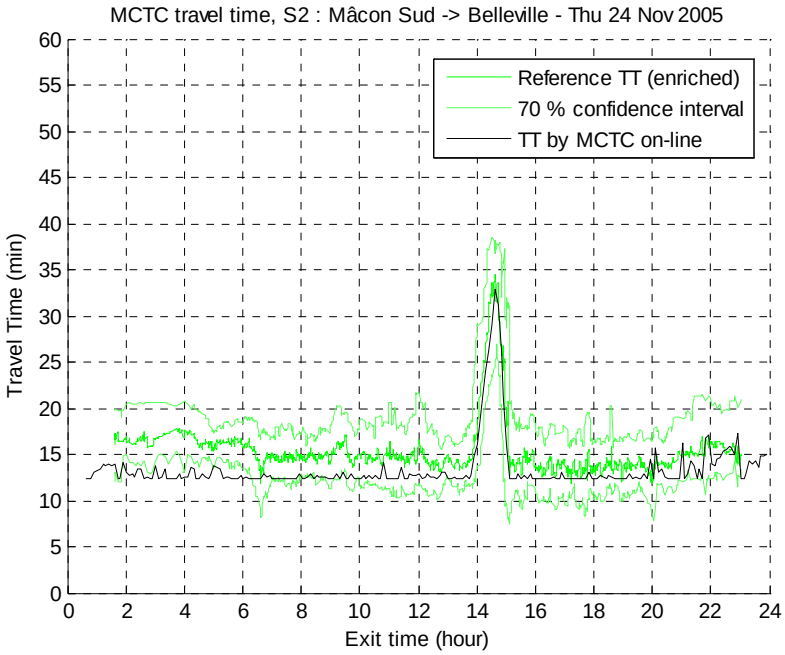


Fig. 4.29: MCTC (on-line) compared to the reference, scenario 2

4.4.1.3.2.5 Results

Fig. 4.30 shows the different estimators for travel time, the reference and the time periods used to compare estimators to the reference in congestion and in free flow conditions.

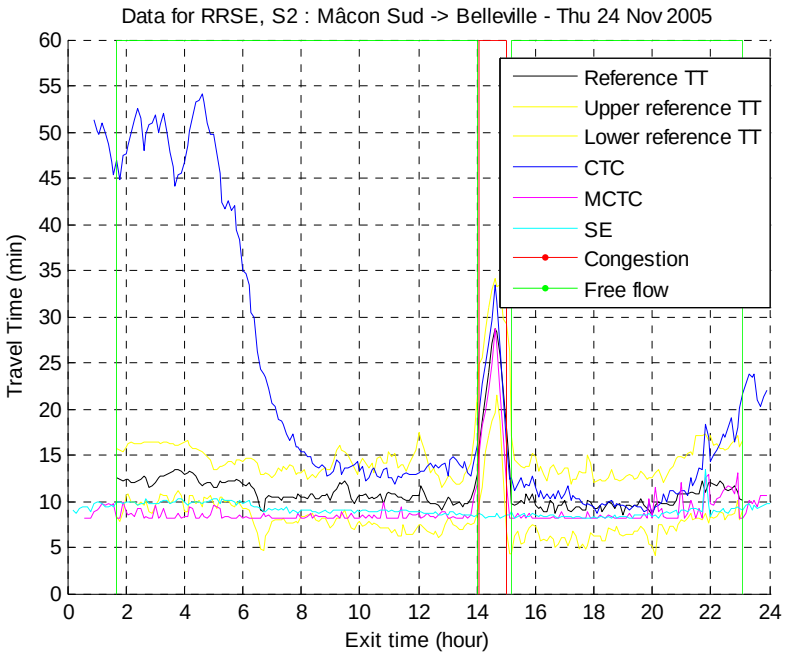


Fig. 4.30: Travel time estimators: Comparisons in congestion / free flow, scenario 2

As summarized in Table 4-2, RRSE-results for the time periods given in Fig. 4.30 show that CTC is totally discarded in free flow. SE and MCTC are very good in free flow. SE is slightly better than MCTC in free flow, but MCTC performs much better than SE in congestion. As the aim is both not to detect congestion in free flow and to detect most accurately the level of congestion during congestion, the only applicable method for scenario 2 is MCTC.

Table 4-2: Travel time estimators: Comparisons in congestion / free flow, scenario 2

RRSE	Congestion	Free flow
CTC	0.132	1.434
MCTC	0.131	0.069
SE	0.651	0.018

4.4.1.3.3 Scenario 3: Tournus -> Mâcon Nord, 25/11/2005

4.4.1.3.3.1 Scenario description

In this scenario, the loop detectors the loop detectors have measured continuously all day long (no data loss). Nevertheless, they show very disturbed flow and speed curves at night, before 8:00 am, Fig. 4.31 and Fig. 4.32. Information was not available, but it could be due to road works and/or detector or data collection malfunctions.

Yet, the data after 8:00 am seems valid. It is also coherent with patrol based data, which states that congestion appeared around 9:30 am and lasted 1 hour, see the red bars in Fig. 4.33. The congestion is located 5.5 km after the upstream detector (see also Fig. 4.10). The height of the red bars corresponds to the number of km of congestion (also written in blue in Fig. 4.33). For instance, the highest length of the congestion was measured 2.5 km around 10:00 am, which means that the congestion was never located on a detector (upstream or downstream). This information is perfectly coherent with the speeds observed on Fig. 4.32: Free flow speed was measured on the detectors during the congestion period. Yet, the congestion can clearly be identified on the flows (Fig. 4.31). This is why Speed Extrapolation should fail and why Cumulative Traffic Counts might be successful.

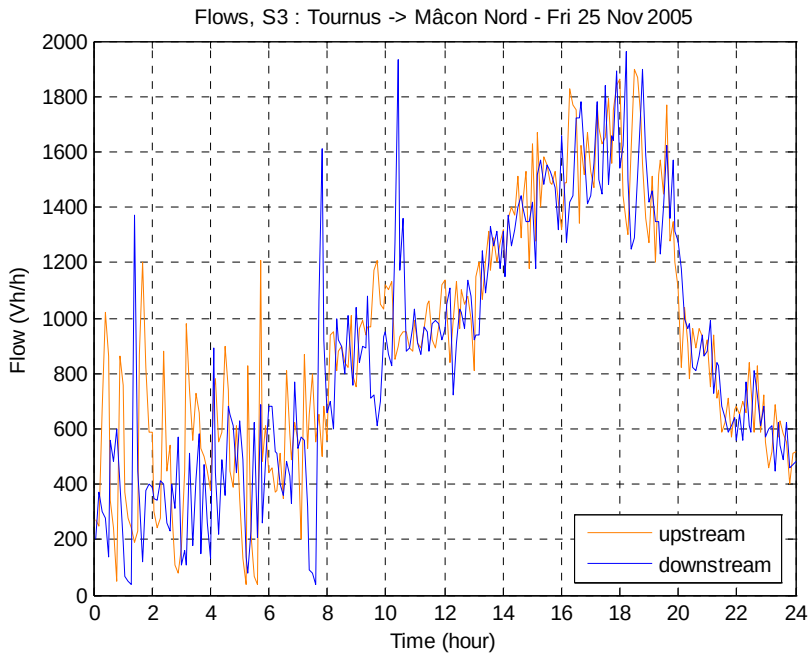


Fig. 4.31: Loop detectors flows, scenario 3

Note also in Fig. 4.32 that during most of the valid period, the upstream speed is always greater than the downstream speed. The mean difference is around 20 km/h. This is very different from the previous scenarios (about 8 km/h in scenario 1, Fig. 4.14, and less than 5 km/h in scenario 2, Fig. 4.24). The explanation of such differences is because in scenarios 1 and 2 the road profile is flat, contrarily to scenarios 3 and 4, where the upstream detector is at the bottom of a steep section and the down stream detector is at the top of another steep section. So the speed is easily maintained

throughout the section in scenarios 1 and 2, whereas it changes in scenarios 3 and 4, due to the road profile. Another consequence is that the speed extrapolation method will be even more difficult to use in scenarios 3 and 4. This is because the measured speeds are not representative of the space mean speed on the section, as they are biased by the road profile.

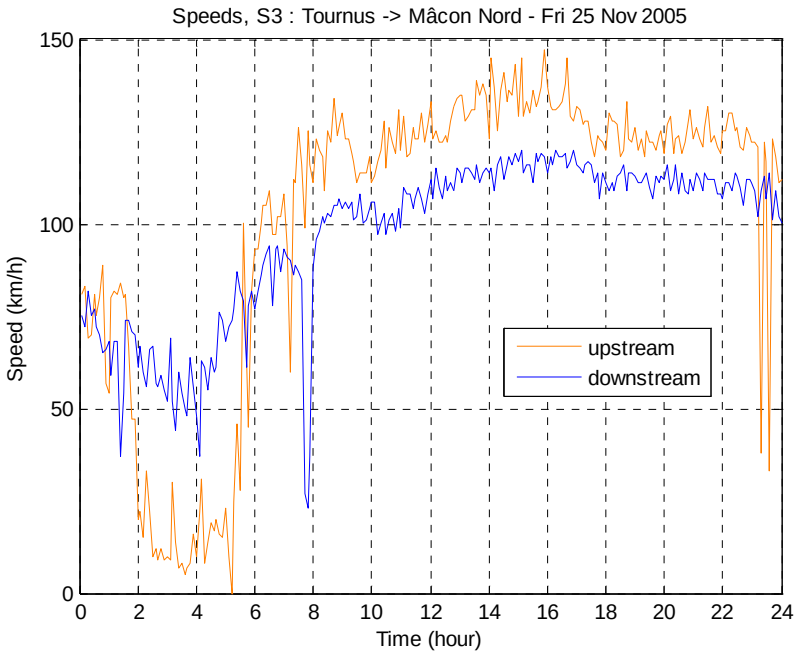


Fig. 4.32: Loop detectors speeds, scenario 3

Fig. 4.33 also gives the reference travel time, computed through the enrichment process (see Section 4.4.1.2). As shown on the figure, the reference is rather coherent both with the direct-link data and with the patrol based data, although the congestion in the reference travel time clearly lasts a little too long, compared to the direct-link data and the patrol based data.

Note also that the reference could be computed only between 6:00 am and 9:00 pm, because there are too few trucks at night on this section. Should a reference really be needed at night, the enriched reference for cars could be used instead, hence neglecting trucks.

Finally, the confidence interval for the reference is rather large (around +/-10 minutes), which also explains the already mentioned imprecision in the ending of the congestion. Such a low reliability in reference travel time is obtained because there are fewer vehicles that can be used in the enrichment process for this section. Hence, scenarios 3 and 4 will be less discriminating when comparing the travel time estimators. As scenarios 1 and 2 have already shown the performance of MCTC, it is not so important to have two other scenarios with highly reliable references. Scenarios 3 and 4 will rather point to the extensions that could improve even more MCTC. Yet, for a more reliable reference, enrichment process could use not only 3-toll information but all toll information, as mentioned in Section 4.4.1.2.

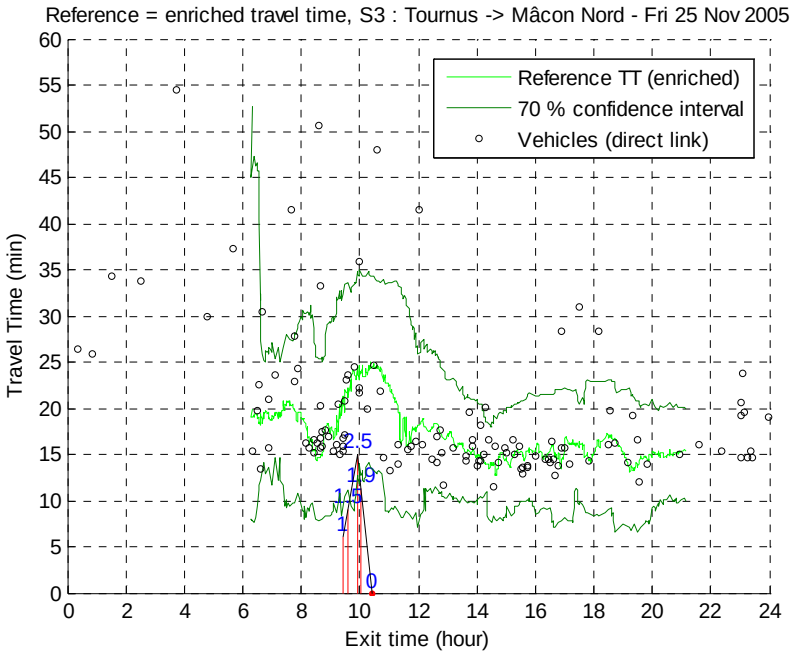


Fig. 4.33: Enriched reference travel time compared to direct-link data, scenario 3

4.4.1.3.3.2 Speed Extrapolation (SE)

Fig. 4.34 gives the estimation of the travel time, based on the classical speed extrapolation method. Yet, due to the low density of speed detectors, this method sees free flow (almost) everywhere during the valid time period (i.e. after 8:00 am) and misses the congestion. Following loop data, the estimated travel time is quite fluctuating at the beginning of the day, with picks above four times the free flow traffic travel time. As explained already in scenario 1 (see Section 4.4.1.3.1.2), the travel time computed from SE can be underestimated or overestimated, simply because the density of detectors is low: This is even worse for scenarios 3 & 4 where there are 20 km in-between the 2 speed measurements (see Fig. 4.10).

Hence, speed extrapolation does not see the congestion and the high travel times at the beginning of the day only draw our attention to a possible problem (road works, bad loop measurement, local epiphenomenon) but are not reliable. The same applies for the small picks at the end of the day. Note also the missing travel time around 5:00 am. This happens because upstream data measured a zero speed (see Fig. 4.32).

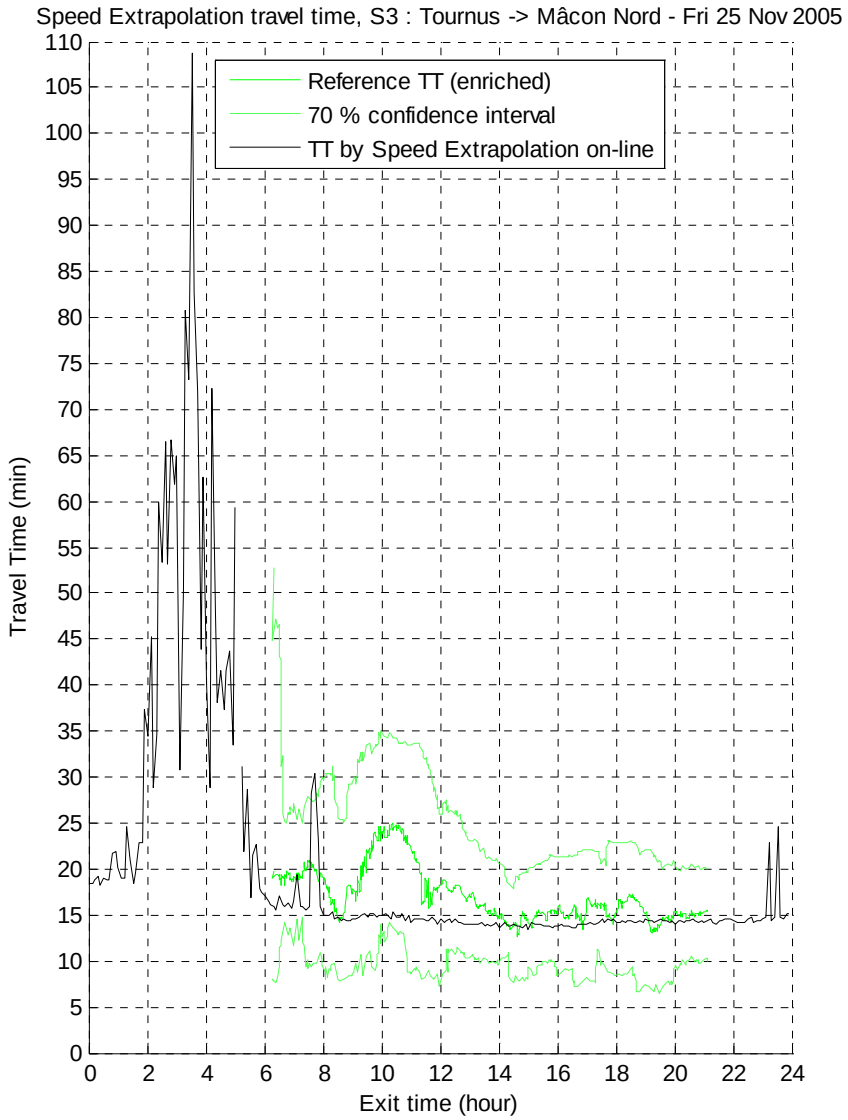


Fig. 4.34: Speed Extrapolation scenario 3

4.4.1.3.3.3 Cumulative Traffic Counts (CTC)

Here (Fig. 4.35), the travel time also rises very high at night. As the travel time does not go back to normal at the beginning of the valid period (8:00 am), this demonstrates that the night loop data is faulty (the stock computed from this data is too high at 8:00 am). It would be interesting to know weather this is due to night road work or to some temporary detector miss measurements.

Nonetheless, the period of the congestion can be identified (see red zone in Fig. 4.35): Beginning and ending are coherent with patrol data. The increase of 10 minutes in the red zone is also coherent with the reference increase.

Yet, it is also very clear that CTC alone cannot distinguish between the false picks and the real ones. Neither can it give the real travel time. And remember that CTC is only an off-line procedure.

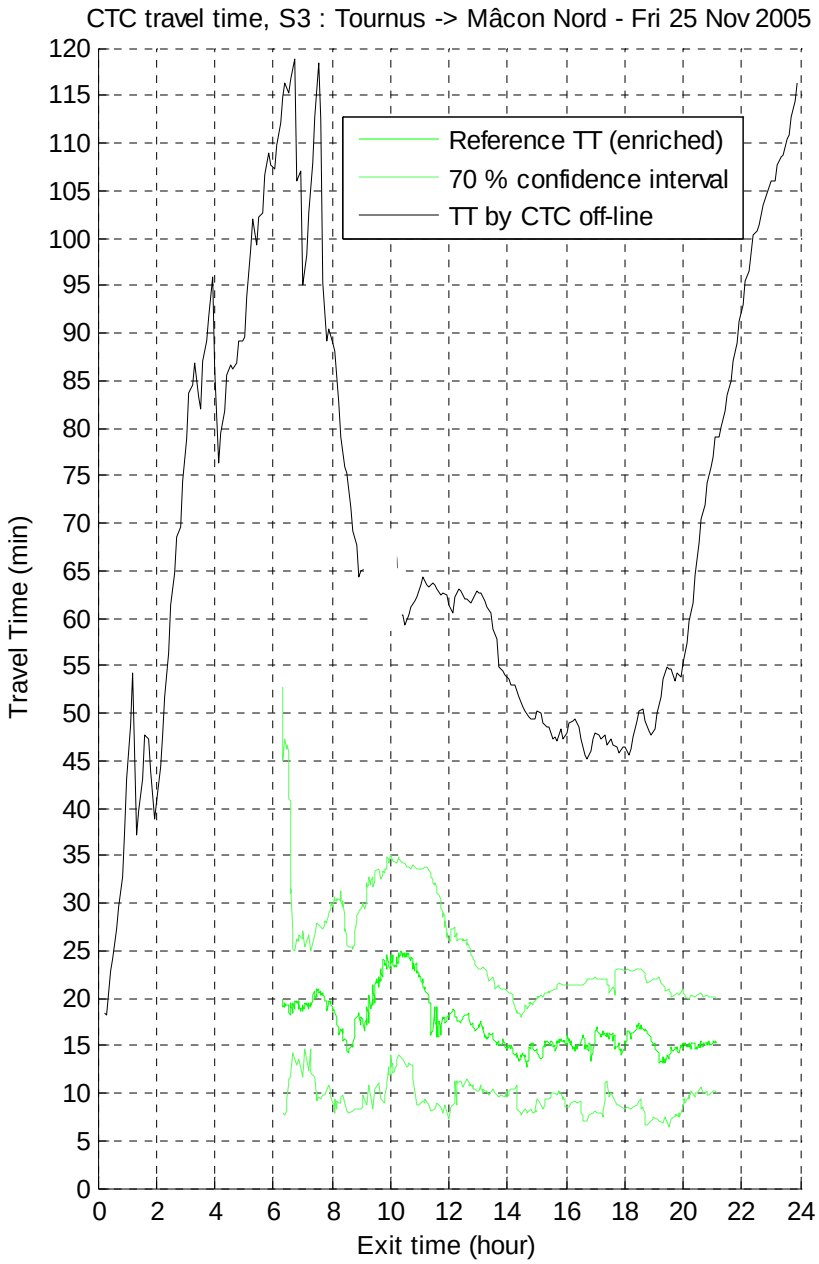


Fig. 4.35: CTC (off-line) compared to the reference, scenario 3

4.4.1.3.3.4 Modified Cumulative Traffic Counts (MCTC)

So our best hopes are in MCTC, the on-line procedure that has shown better performances than both SE and CTC in the previous scenarios. And yes, here also (Fig. 4.37), MCTC travel times are perfectly coherent with the reference. As shown on Fig. 4.36, there is also enough direct link data to show that the high travel times estimated by SE at the end of the day are only very local epiphenomenon and that travel time, on the whole, remains free flow for this section and time of the day.

But MCTC has delivered travel times only after 3:00 pm, which is after the congestion period! This is due to the initialization step of MCTC, which was possible only late in the day because of the very

unstable loop detector data. In fact, initialization step is more constrained than correction process, because the real evaluation of the number of vehicles driving in the section was searched.

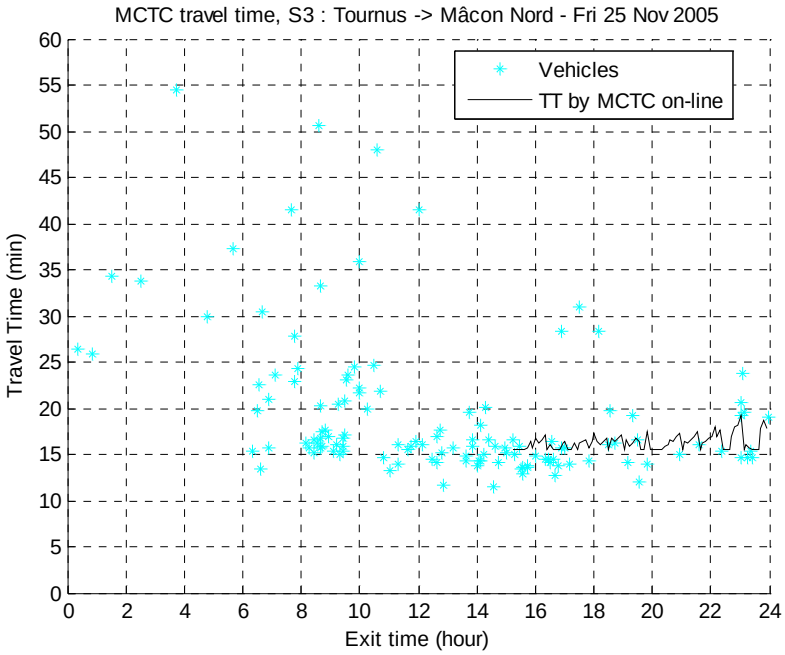


Fig. 4.36: MCTC (on-line) using direct link data, scenario 3

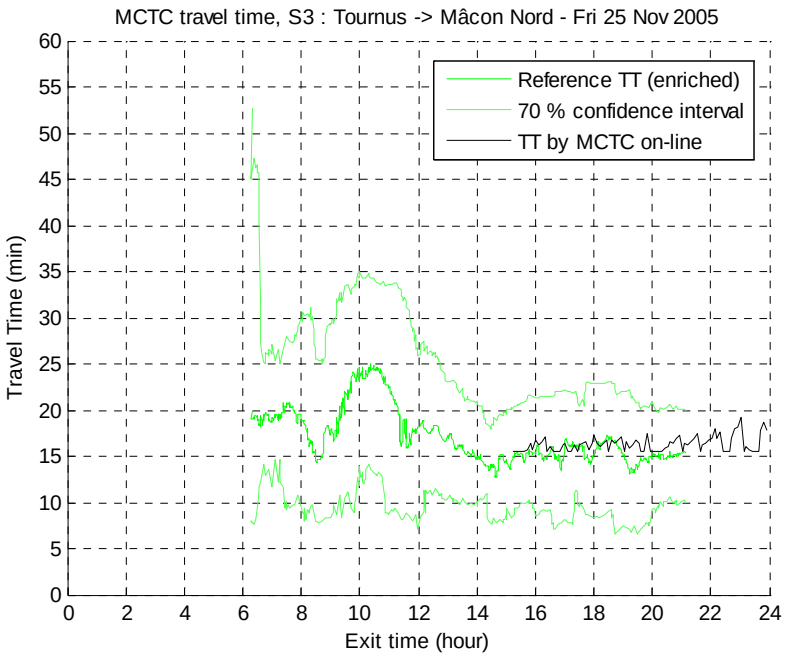


Fig. 4.37: MCTC (on-line) compared to the reference, scenario 3

Yet, had the initialization been done the day before (at a stable traffic period), MCTC would have been able to show results for the whole day, and, most interesting, for the congestion period.

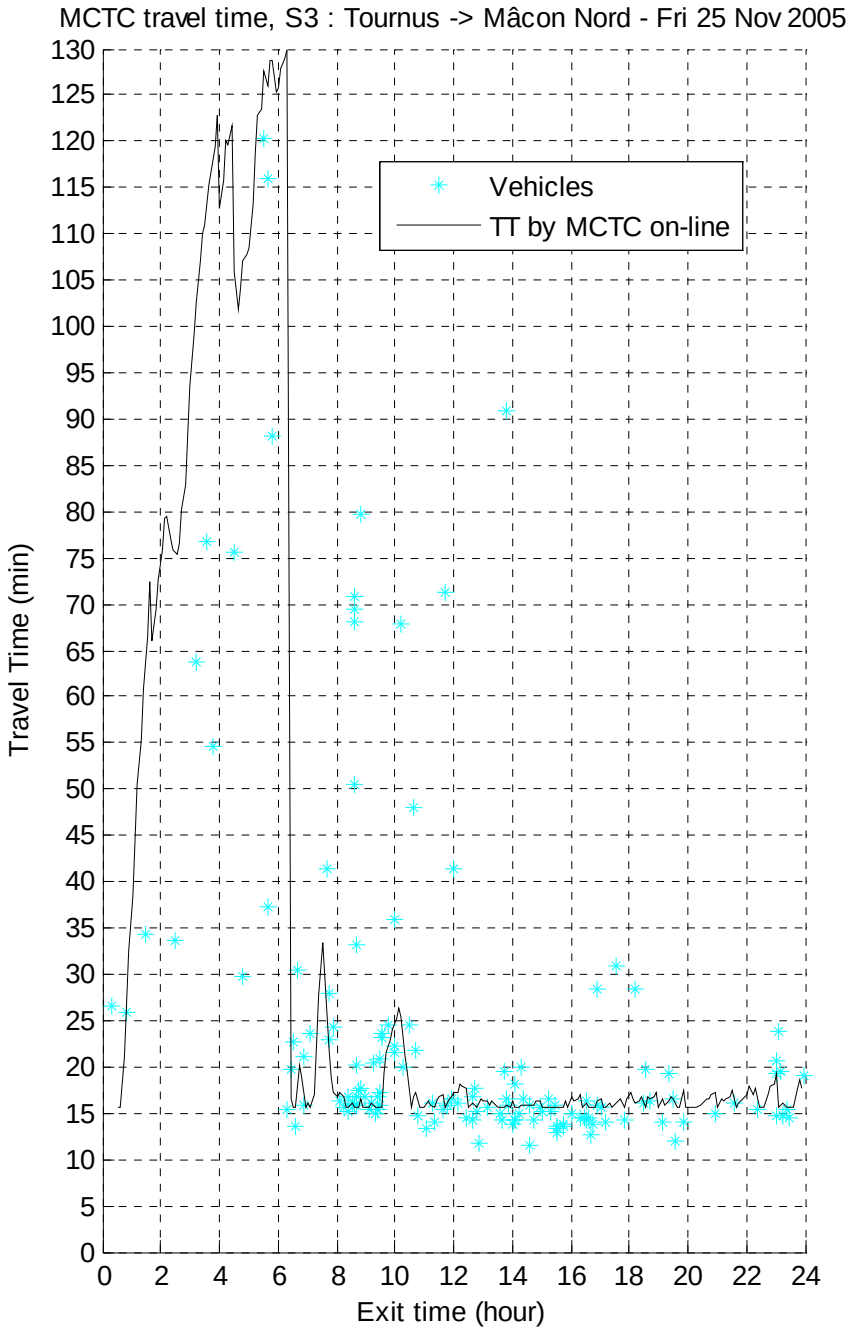


Fig. 4.38: Relaxed MCTC (on-line) using direct link data, scenario 3

Another approach would be to relax the constraints of the initialization step. In fact, MCTC is quite robust to wrong initializations, as the stock is regularly corrected from the feedback of the toll data. The strong initialization step is useful only to guaranty MCTC results from the beginning.

So in order to both demonstrates this robustness and test MCTC on the whole day, MCTC has been manually initialized as if the section were absolutely empty at the beginning of the day. The result is shown in Fig. 4.38 and Fig. 4.39.
The congestion is properly identified: Its beginning and end are very coherent with the direct link data and the patrol data, even more than in the case of the reference (see Section 4.4.1.3.3.1). Its level is coherent with the reference (up to 10 minutes lost during congestion).

Free flow is detected after the congestion and right before it. The only outliers that remain are always the same ones: Those at the beginning of the day, because of the strange loop detector data. Yet, they would have been erased if there had been enough direct link data supporting free flow conditions. But, as shown in Fig. 4.38, no vehicle supports this before 6:00 am, and only very few of them before 8:00 am. So it really seems that there was something special on this section before 6:00 am. Yet, to get the right travel time level, it should be necessary to extend MCTC once more in order to make corrections not only during free flow but also during congestion. This should be done in future work in order to cope with bad detector data in all traffic conditions.

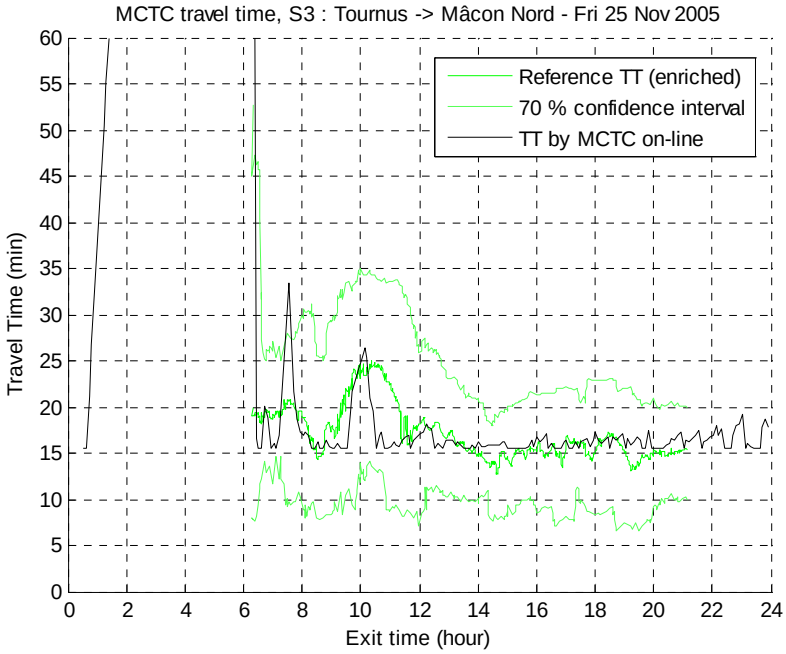


Fig. 4.39: Relaxed MCTC (on-line) compared to the reference, scenario 3

4.4.1.3.3.5 Results

Fig. 4.40 shows the different estimators for travel time, the reference and the time periods used to compare estimators to the reference in congestion and in free flow conditions. As already pointed out in Section 4.4.1.3.3.1, the reference reliability is not very good for this scenario. Hence, only a few time zones are used for estimator comparison: Those where the reference confidence interval is thin enough, resulting in 1 time zone for congestion and 4 for free flow, one of which being between 6:00 am and 8:00 am. The use of this specific time zone is questionable for this scenario, as detector data validity is assumed after 8:00 am only. Yet, it is the only free flow period available for comparison purposes before congestion and it has been ranked as enough reliable to be used. Besides, this will not have a big impact on the results: In fact it will only show that SE and MCTC are very good but not always 100% perfect during free flow, which is absolutely true.

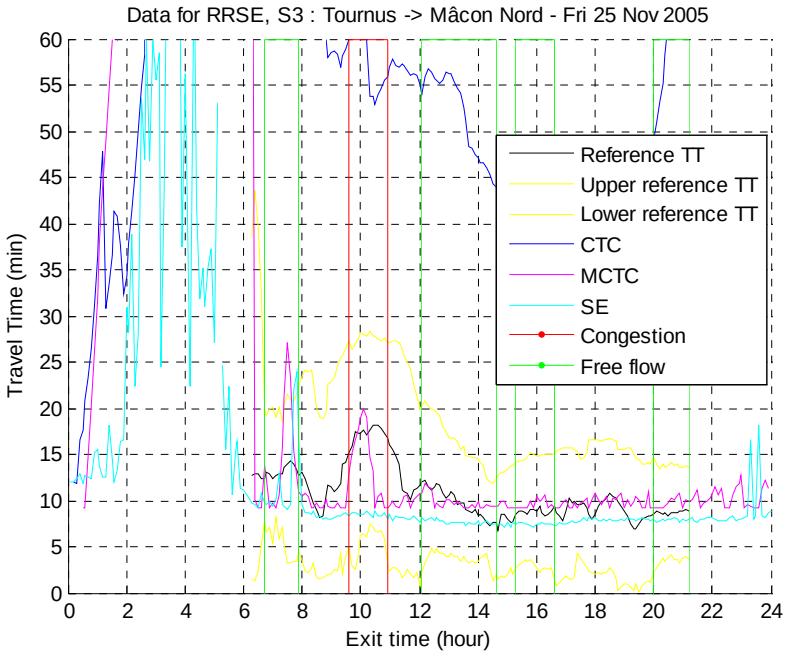


Fig. 4.40: Travel time estimators: Comparisons in congestion / free flow, scenario 3

As summarized in Table 4-3, RRSE-results for the time periods given in Fig. 4.40 show that CTC is totally discarded both in free flow and in congestion. SE and MCTC are very good in free flow. SE is still slightly better than MCTC in free flow (although it could have turned out the other way around, had the reference been available around 11:00 pm too), but MCTC performs much better than SE in congestion. As the aim is both not to detect congestion in free flow and to detect most accurately the level of congestion during congestion, the only applicable method for scenario 3 is MCTC (in its relaxed version for the initialization step).

Table 4-3: Travel time estimators: Comparisons in congestion / free flow, scenario 3

RRSE	Congestion	Free flow
CTC	2.344	4.721
MCTC (relaxed)	0.294	0.073
SE	0.510	0.031

It would be interesting to know what happened at the beginning of the day, even though the reference is not available. It could also be interesting to have MCTC extended to corrections during congestion. This way, the estimated travel time would certainly remain below 40 minutes at the beginning of the day, leading to better results than SE or CTC for this time period too.

4.4.1.3.4 Scenario 4: Tournus -> Mâcon Nord, 28/11/2005

4.4.1.3.4.1 Scenario description

This scenario is very similar to scenario 2 except that the congestion appears in 2 steps: It first starts just after 10:00 am (congestion about 1km long), and then worsens at lunch time (12:00 am 1:00 pm, congestion up to 2.5 km long). Finally, it ends just before 2:00 pm, as stated by the patrol data (red bars and corresponding blue kilometre values in Fig. 4.43).

As in scenario 2, the loop detectors have experienced only very few losses, as shown on Fig. 4.41 and Fig. 4.42. So the flows and speeds which have been used for SE, CTC and MCTC are mostly real measurements.

The congestion is located 1.2 km before the downstream stream detector (see also Fig. 4.10). The height of the red bars in Fig. 4.33 corresponds to the number of km of congestion (also written in blue). For instance, the highest length of the congestion was measured 2.5 km around 1:00 pm, which means that the congestion was never located on a detector (upstream or downstream). This information is perfectly coherent with the speeds observed on Fig. 4.42: Free flow speed was measured on the detectors during the congestion period, although, when comparing to the 2 hour time periods just before and after the congestion, the downstream speed seems to be 10 km/h less during congestion (probably because not all vehicles have reached their cruise speed 1.2 km after congestion). In fact, the congestion can much more clearly be identified on the flows (Fig. 4.41): First step from 10:00 am to 12:00 am, second and greater step from 12:00 am to 1:00 pm, and ending from 1:00 pm to 2:00 pm. This is why Speed Extrapolation should fail and why Cumulative Traffic Counts might be successful.

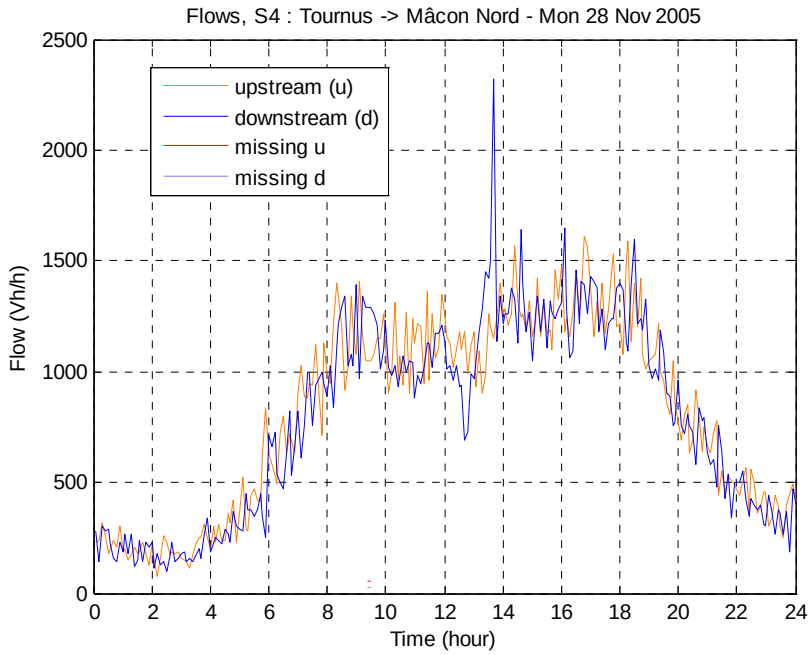


Fig. 4.41: Loop detectors flows, scenario 4

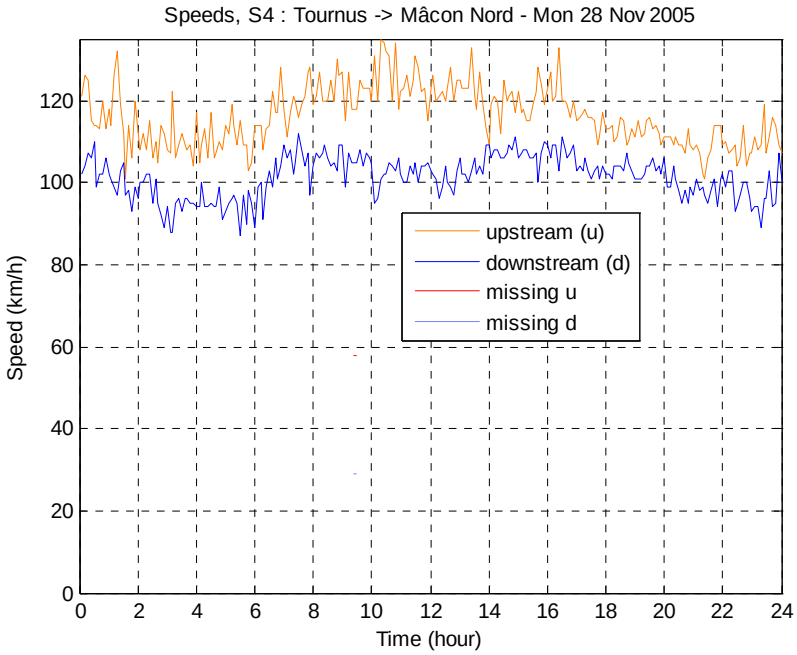


Fig. 4.42: Loop detectors speeds, scenario 4

Note also in Fig. 4.42 that the upstream speed is always greater than the downstream speed, as in scenario 3. The mean difference is around 15 km/h. As explained in scenario 3, this is due to the road profile and the position of the sensors downhill for the upstream detector and uphill for the downstream one. Again, this makes even less reliable the speed extrapolation method, as the only 2 measured speeds are biased by the road profile.

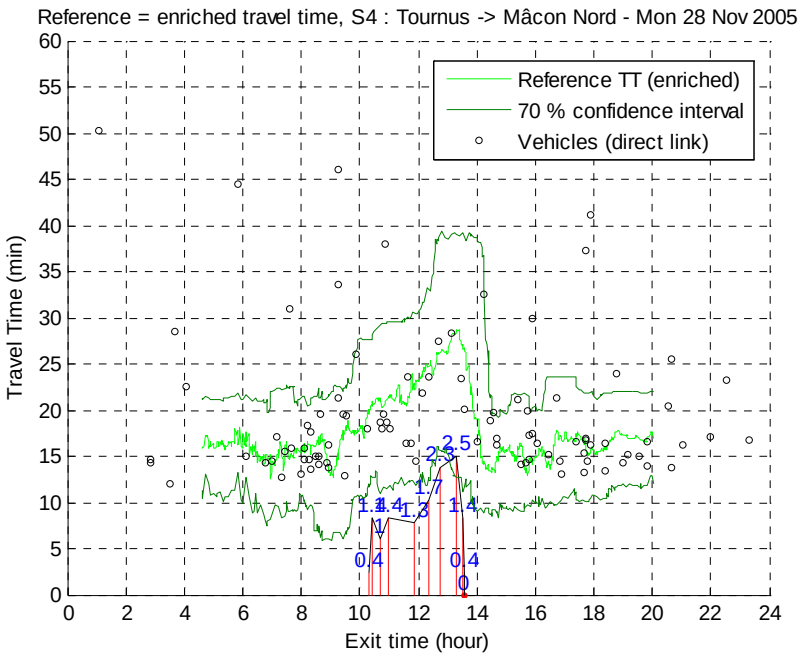


Fig. 4.43: Enriched reference travel time compared to direct-link data, scenario 4

Fig. 4.43 also gives the reference travel time, computed through the enrichment process (see Section 4.4.1.2).

As shown on this figure, the reference is rather coherent both with the direct-link data and with the patrol based data, even though some vehicles from the direct link seem to have managed a free flow travel time around noon. Note also that the reference could be computed only between 4:00 am and 8:00 pm, because there are still too few trucks at night on this section. Should a reference really be needed at night, the enriched reference for cars could be used instead, hence neglecting trucks.

Finally, the confidence interval for the reference is rather large, especially during congestion (up to +/- 10 minutes). Such a low reliability in reference travel time is obtained because, during day time also, there are fewer vehicles that can be used in the enrichment process for this section. Hence, as for scenario 3, scenario 4 will be less discriminating than scenarios 1 and 2 when comparing the travel time estimators. Hence, scenario 4 will rather point out some extensions that could improve even more MCTC. Yet, for a more reliable reference, enrichment process could use not only 3-toll information but all toll information, as mentioned in Section 4.4.1.2.

4.4.1.3.4.2 Speed Extrapolation (SE)

Fig. 4.44 gives the estimation of the travel time, based on the classical speed extrapolation method. Yet, due to the low density of speed detectors, this method sees free flow everywhere and misses the congestion.

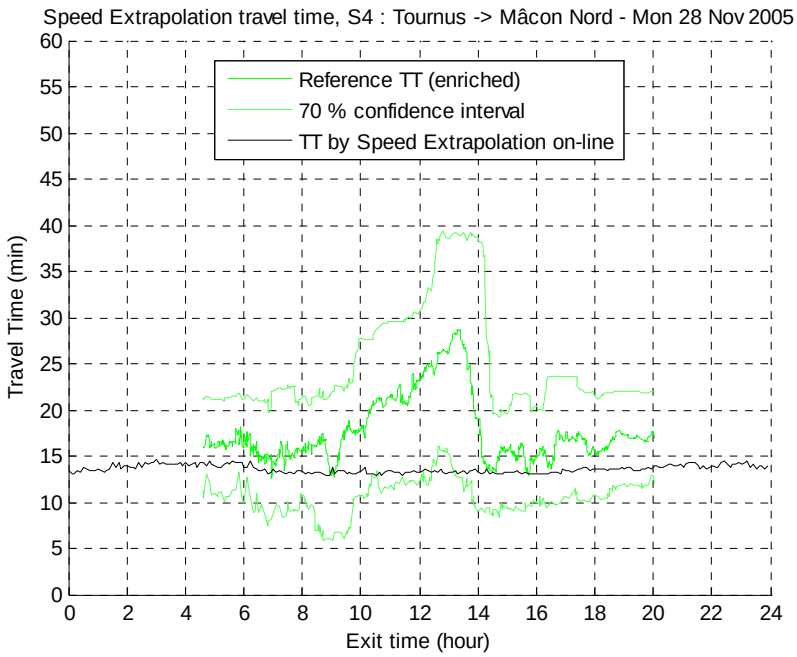


Fig. 4.44: Speed Extrapolation scenario 4

4.4.1.3.4.3 Cumulative Traffic Counts (CTC)

As in scenarios 1 and 2 (see Fig. 4.17 and Fig. 4.27), where loop data is valid all day long, Fig. 4.45 shows a CTC travel time that has identified the congestion period and is also close enough to the reference confidence interval during day time, although the travel time estimation during congestion was really satisfying only for scenario 2. In fact, CTC is really far from the reference travel time only in presence of not valid data (scenario 3, Fig. 4.35). Nevertheless, the night estimations are discrediting in all scenarios: CTC can definitely not deal with rest and service areas. Besides, CTC is still available off-line only.

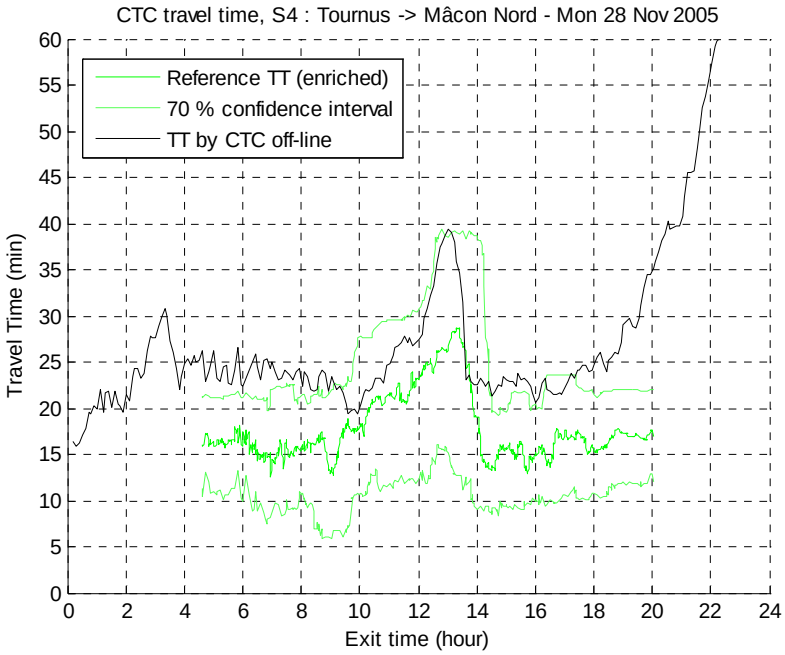


Fig. 4.45: CTC (off-line) compared to the reference, scenario 4

4.4.1.3.4.4 Modified Cumulative Traffic Counts (MCTC)

Here, relaxed MCTC was not necessary, as strong initialization was doable at the beginning of the day, thanks to stable traffic and valid loop detector data (see Fig. 4.41). As in scenario 1, MCTC performs much better than SE or CTC, despite some slight problems persisting at night. Actually, the reference is not available for these short periods of the night (at the beginning and at the end of the day), but free flow travel time is most probable even though MCTC leads to higher travel times. These slight problems could not be corrected by MCTC because there are not enough vehicles travelling on this section at night in order to make corrections regularly enough (Fig. 4.46). The use of enriched data (as for the reference) could enable MCTC to make corrections not only based on direct link data but on the data from all tolls, thus enabling more frequent corrections.

As for day time, MCTC proves very efficient already, with direct link data only. The two step congestion is seen, the second step having more impact on travel time, which is coherent with both patrol data and the reference travel time. The only questionable point during congestion is the return back to free flow travel time in-between the two steps of the congestion (around noon). As shown on Fig. 4.46, this is because 2 vehicles of the direct link managed to drive at free flow travel time around noon. So, either there have been two successive congestions with only a small free flow period in-between, or the two vehicles escaped someway the congestion, using for instance the emergency lane. In order to decide whether a vehicle is representative of the flow or not, it is necessary to have the exit travel time of as many vehicles as possible. As there are only two vehicles on the direct link, the use of reliable enriched data would improve MCTC results. In the present case, the enriched reference is not reliable enough to decide between the two options, as the confidence interval is coherent with both options, Fig. 4.47.

Nonetheless, the overall result is fair, as MCTC and the reference are coherent during most of the two step congestion. The only main difference is the return to free flow or not at noon. Both options are coherent with the reference confidence interval. Yet, if MCTC were mistaking, its quick return to good travel time estimation during the second congestion step would demonstrate even more robustness from MCTC.

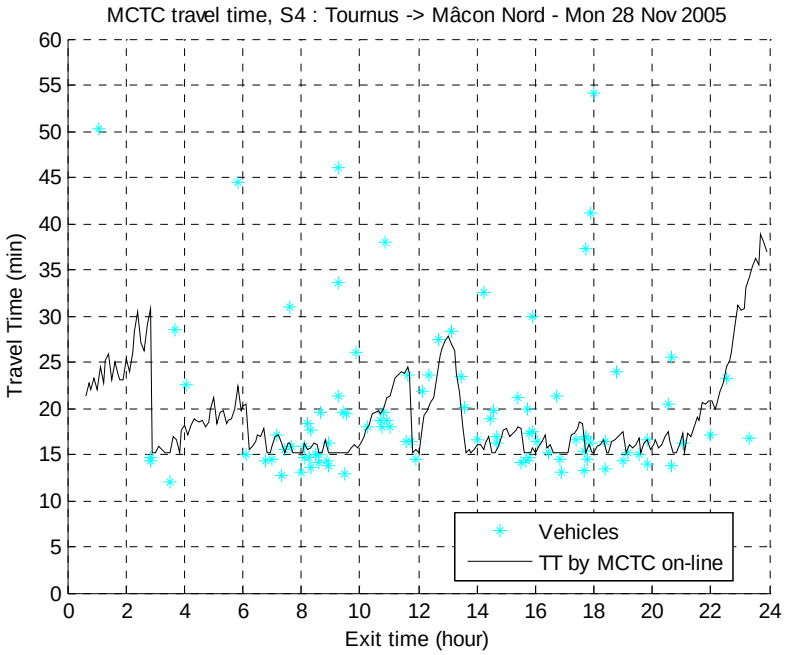


Fig. 4.46: MCTC (on-line) using direct link data, scenario 4

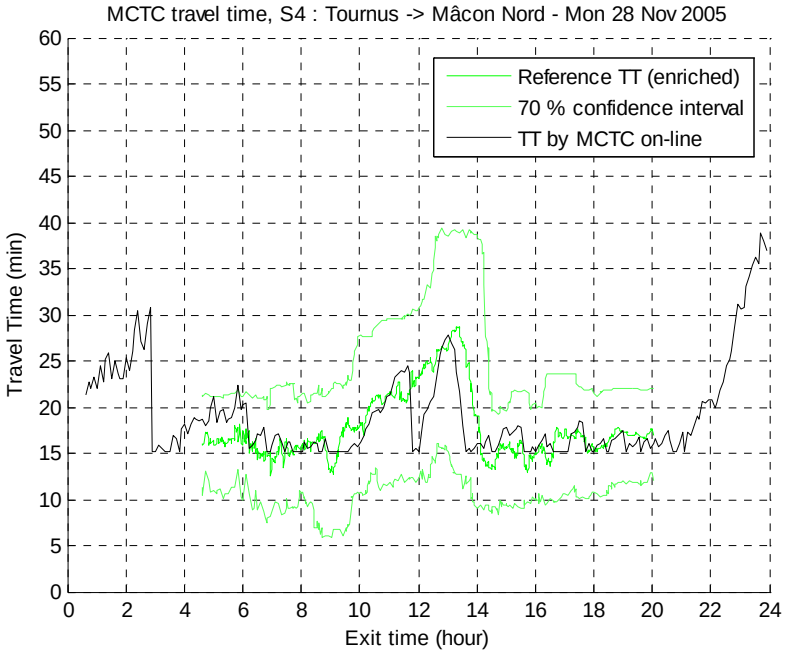


Fig. 4.47: MCTC (on-line) compared to the reference, scenario 4

Using enriched data (as also explained in scenario 1) could help both avoid night false alarms and possibly false congestion end detection. This will be investigated in future work.

4.4.1.3.4.5 Results

Fig. 4.48 shows the different estimators for travel time, the reference and the time periods used to compare estimators to the reference in congestion and in free flow conditions. As in scenario 3, the reference reliability is not always good enough for this scenario (because there are fewer vehicles on this section for enrichment purposes). So, only time zones where the reference confidence interval is thin enough are used for estimator comparison. Yet, scenario 4 has a better reference than in scenario

3. So, most of the reference can be used to compare the different estimators, before, during and after congestion.

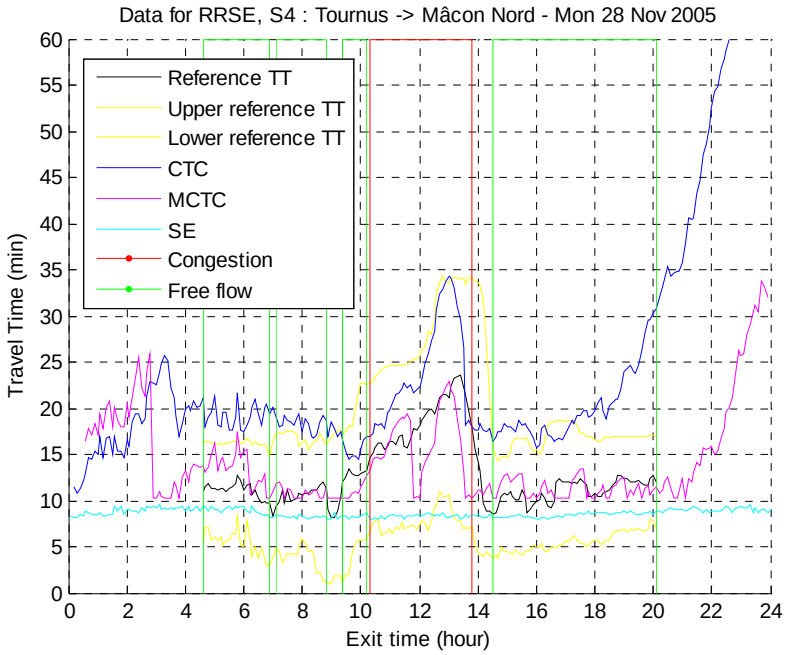


Fig. 4.48: Travel time estimators: Comparisons in congestion / free flow, scenario 4

As summarized in Table 4-4, RRSE-results for the time periods given in Fig. 4.48 show that CTC was rather successful for both congestion and free flow of scenario 4, although still way below SE and MCTC in free flow. MCTC also performs better than CTC in congestion. So MCTC is always better than CTC, as in all previous scenarios, both for congestion and free flow. SE is still slightly better than MCTC in free flow, but MCTC performs much better than SE in congestion. As the aim is both not to detect congestion in free flow and to detect most accurately the level of congestion during congestion, the best method for scenario 4 is MCTC, as in all scenarios.

Table 4-4: Travel time estimators: Comparisons in congestion / free flow, scenario 4

RRSE	Congestion	Free flow
CTC	0.355	0.363
MCTC	0.244	0.005
SE	0.565	0.000

4.4.1.4 Conclusion

This section has developed a new methodology for highway travel time estimation and short-term prediction using a modified cumulative traffic counts method. The idea is to make CTC corrections so that the experienced travel time estimated by MCTC follows the real one obtained from ETC data. The correction process is thus on the experienced travel time only. Then, as CTC has been corrected/validated into MCTC, the classical prediction (presented in Annex II-A) applies.

The results are very promising and MCTC can be easily implemented in real-time. MCTC is also a travel time estimation that integrates CTC, speed extrapolation, and ETC data into a unified framework.

A schematic representation of all RRSEs given in the previous tables for each scenario (Table 4-1, Table 4-2, Table 4-3, and Table 4-4) is presented in Fig. 4.49. It supports the discussions and

conclusions of this section on travel time estimation on S2-type sections (no entry or exits, but rest and service areas):

- CTC is outperformed by MCTC as RRSEs are always better for MCTC than CTC both in congestion and free flow, in all scenarios and as MCTC estimation is on-line whereas CTC estimation is only off-line.
- Neither CTC nor SE is satisfying: CTC performs better than SE in congestion for the 3 scenarios where loop data is valid all day long. In fact SE misses the congestion in 3 of the scenarios and only finds some hint about it in scenario 1. During free flow traffic, though, SE is the best estimator in the 4 scenarios and CTC the worst. Besides CTC is off-line and SE is on-line.
- MCTC is the best estimator: It can particularly well distinguish congestion from free flow conditions. During congestion, MCTC is the best estimator. It most accurately estimates the level of congestions. For free flow, it performs closely to SE. Finally, MCTC estimates travel times on-line.

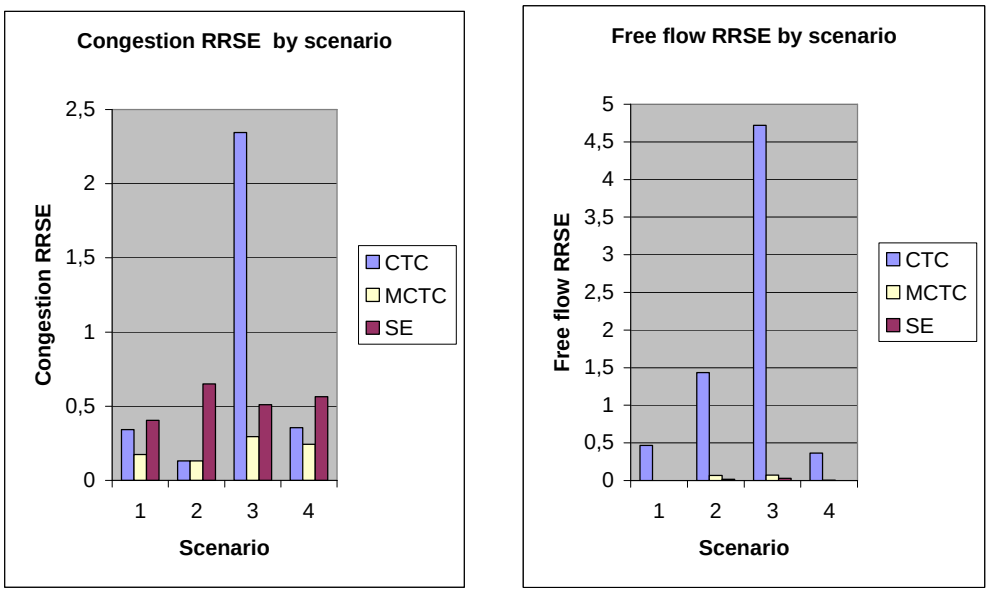


Fig. 4.49: Comparison of travel time estimators, all scenarios (congestion and free flow)

So MCTC has proven its reliability and that integrating toll data to the loop detector data is very efficient for travel time estimation. MCTC has also demonstrated its robustness to initialization (scenario 3) and to suspicious toll data (scenario 4).

Yet the success of MCTC depends on the availability of toll data and future work should concentrate on integrating not only the direct link toll data but all the toll data (as enrichment for the reference travel time). Moreover, the actual correction process of MCTC applies for the moment during free flow conditions only. But investigations have already started to make corrections during congestion too, as presented in Section 4.4.1.3.1.6. This will enable to generalize MCTC to any section for toll-operated networks (even of S3-type, i.e. with entry and exits, and without the need of historical data, which otherwise is necessary, as presented in [111]).

4.4.2 Estimation of dynamic OD matrices based on floating car data

4.4.2.1 Introduction

Traffic data collection has been dominated by loop detectors for several decades. Over the past years, however, there have been considerable advances in data collection technologies, including Automatic Vehicle Identification systems, floating car (or probe vehicle) deployments, and alternative emerging sensor technologies, such as microwave/acoustic/laser sensors.

Dynamic traffic assignment (DTA) models can be used for the fusion of multiple, diverse data sources for the estimation and prediction of traffic conditions in a network. Antoniou et al. ([115], [116]) have demonstrated the combination of loop detector data with emerging data technologies for the OD estimation and prediction problem.

A flexible methodology that can be used for fusion of any available data sources, including Automated Vehicle Identification data (e.g. from floating cars, transponder based systems, or CCTV-based license plate recognition systems), as well as point sensor data (e.g. data from loop detectors or microwave/radar/acoustic sensors) is presented. An application of the methodology for the calibration of the DynaMIT DTA model in a network in Vienna using floating car data is also presented.

4.4.2.1.1 Dynamic traffic assignment framework

Dynamic Traffic Assignment (DTA) systems typically reside at Traffic Management Centres (TMC) and can support both planning and real-time applications. Planning applications may include the off-line evaluation of incident management strategies, the evaluation of alternative traffic signal and ramp meter operation strategies and the generation of evacuation and rescue plans for emergencies (e.g. natural disasters) that could affect the traffic network.

Real-time applications make use of the traffic prediction capabilities of DTA systems and may include on-line evaluation of guidance and control strategies, real-time incident management and control, support of real-time emergency response efforts and optimization of the operation of TMCs through the provision of real-time predictions. Real-time DTA systems are also often referred to as *traffic estimation and prediction systems* (the two terms will be used interchangeably in this document). An overview of the state-of-the-art Dynamic Traffic Assignment framework is shown in Fig. 4.50.

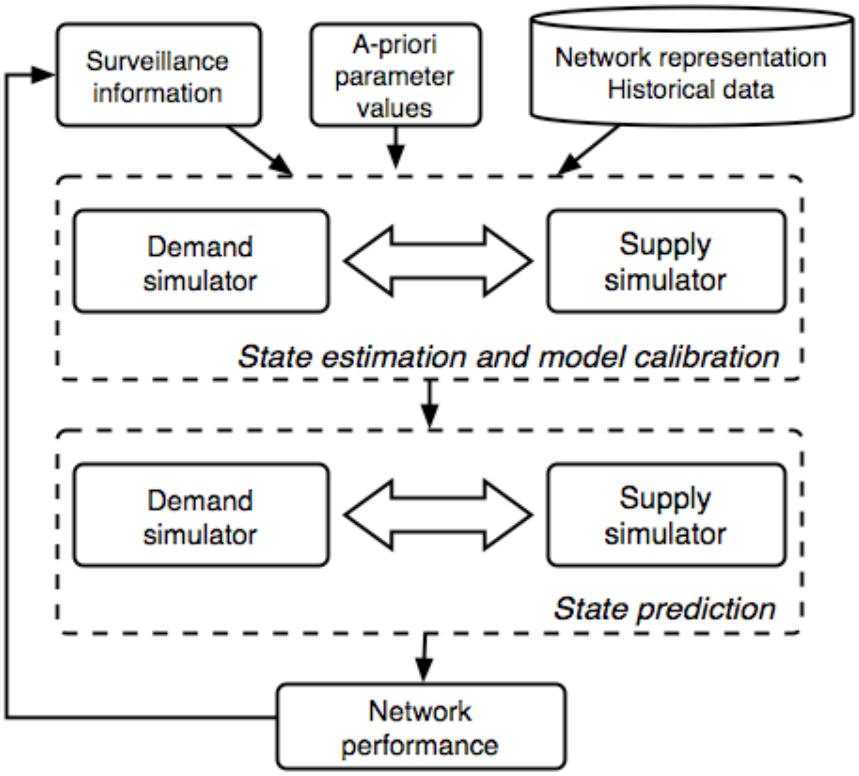


Fig. 4.50: DTA overview

Real-time DTA systems typically comprise two main functions ([117], [118]):

- State estimation; and
- Prediction-based information generation

During the *state estimation* phase, real-time information is combined with historical data to capture the current traffic conditions in the network. Detailed traffic information that is obtained through the instrumented portions of the network is used to infer the conditions in the parts of the network for which no real-time information is available. This is achieved through an iterative simulation of demand-supply interaction designed to reproduce real-time observations from the surveillance system.

The role of the *prediction-based information generation process* is to generate unbiased and consistent traffic information for dissemination to travellers. Information based on predicted network conditions (i.e. anticipatory information) is likely to be more effective than information based on current traffic conditions because it accounts for the evolution of traffic conditions over time, which is what travellers will experience. A detailed treatment of the demand-supply interactions within a state-of-the-art DTA system can be found in [117] and [118].

One of the key components of dynamic traffic assignment is the Origin-Destination (OD) estimation and prediction process [119]. OD estimation combines historical and real-time information to obtain dynamic i.e. time-dependent demand matrices. Furthermore, OD prediction exploits estimated behavioural patterns to anticipate the short-term evolution of demand [120] Based on the predicted demand, it is possible to generate and evaluate response strategies and generate anticipatory guidance [121].

The supply simulator is usually based on high-level (mesoscopic or macroscopic) models that represent traffic dynamics using speed-density relationships, kinematics representation of traffic elements of queuing theory, etc. Such models are used in real-time applications, e.g. DynaMIT ([117] and [118]) and DYNASMART [122]. Proper use of these models requires calibration of a number of parameters.

4.4.2.1.2 Research objective

The objective of this research is to motivate and demonstrate the use of floating car data for DTA model calibration. A flexible methodology that can be used for fusion of any available data sources, including Automated Vehicle Identification data (e.g. from floating cars, transponder based systems, or CCTV-based license plate recognition systems), as well as point sensor data (e.g. data from loop detectors or microwave/radar/acoustic sensors).

One particular aspect of using floating car data is that due to the inherent sampling of vehicles associated with the data collection technology it is not possible to obtain direct estimates of link flows. Link flow information is very important in order to be able to generalize the sample information from the equipped vehicles to the entire traffic population. The application that is presented in this deliverable focuses on the estimation of link flows from floating car data (in particular travel time information).

4.4.2.1.3 Structure for the dynamic OD matrix estimation

The dynamic OD matrix part of this deliverable is structured as follows. Section 4.4.2.2 provides an overview of relevant literature, while Section 4.4.2.3 outlines the methodology. The overall DTA calibration framework is presented first, followed by an extension focussing on the use of floating car data. The formulation of the problem, along with appropriate solution approaches is discussed next. The DynaMIT-P DTA system is presented in Section 4.4.2.4. The setup of the pilot demonstration is presented in Section 4.4.2.5 and results of the calibration using travel times are presented in Section 4.4.2.6. Section 4.4.2.7 concludes this part of the deliverable with a discussion on the application and directions for next steps.

4.4.2.2 Literature review

The problem of OD estimation and prediction has been well studied over the last two decades (e.g. [123], [42], [124], [125], [126], [80], [127], [119] and [128]). In the majority of the relevant literature, OD estimation is based on historical demand information and link count data. However, today, a number of traffic surveillance technologies have been developed, with different technical characteristics,

including data collected, accuracy of the measurements, levels of maturity, feasibility and cost, and network coverage.
Based on their functionality traffic sensors can be categorized as point, point-to-point, and area wide. Point sensors are the most widely used type of detector today. (Inductive) loop detectors have been in use for several decades, while other technologies, including radar, infrared, acoustic, and video sensors have also been developed. Point-to-point sensors are an emerging group of technologies that allow for the detection of vehicles in multiple locations as they traverse the network. This feature makes the re-identification and tracking of these vehicles possible, which in turn may (under certain conditions) provide information such as point-to-point travel times, route choice fractions, paths, and OD flows. Area-wide technologies are promising technologies that are still under research and include airborne sensors, i.e. unmanned aircraft continuously flying over and surveying a traffic network.

Ashok [127] introduced the notion of direct measurements for the incorporation of probe vehicle information for the solution of the OD estimation and prediction problem. By definition, a direct measurement provides a preliminary estimate of an OD flow. In the case of information obtained from probe vehicles (with known origin and destination), direct measurements may be represented by the product of the flow of probe vehicles with an appropriate expansion matrix to account for the fact that probe vehicles constitute only a fraction of the total number of vehicles in the network.

[129] combined volume counts with trajectory information obtained from automated license-plate surveys for the estimation of OD flows. A measurement equation for the trajectory counts is specified and split probabilities are estimated from combined link volume counts and trajectory counts. The method of estimating split probabilities from combined data was tested by the author in experiments using synthesized data. These experiments showed that using combined data consistently led to lower errors relative to the case with only traffic counts.

Mishalani et al. ([130]) evaluate the role of various types of surveillance data in the real-time estimation of dynamic OD flows at the intersection level. Turning fraction and travel time data was collected from video cameras at three adjacent intersections. OD estimation using link counts was used as the base case. Additional scenarios involved the incorporation of turning fractions at the intersections, travel times, and combination of turning fractions and travel time. The main finding of the study was that intersection turning fraction data significantly improves the quality of the OD flow estimates, while the use of link travel time data in addition to turning fractions can also be useful.

Oh and Ritchie ([131]) develop a methodology for the design of advanced traffic surveillance systems based on microscopic traffic simulation. The authors demonstrate the proposed methodology through the development of an inductive-signature-based anonymous vehicle tracking system for the origin-destination flow estimation problem.

Dixon and Rilett ([132]) propose a method for using sample link choice proportions and sample OD matrix information derived from AVI data sampled from a portion of vehicles to estimate population OD matrices with the AVI data collection points acting as the origins and the destinations. This assumption, however, is limiting in terms of the applicability of the model. The authors also propose a method that extends their approach to the estimation of ramp-to-ramp freeway OD volumes.

Kwon and Varaiya ([133]) develop a statistical OD estimation model using partially observed vehicle trajectories obtained with vehicle re-identification or automatic vehicle identification (AVI) techniques such as electronic tags. The authors derive an unbiased estimator of the O-D matrix based on the method of moments and also develop the bootstrap standard error estimate of the O-D matrix estimator.

Antoniou et al. ([115]) presented a methodology for the incorporation of AVI information into the OD estimation and prediction framework, which however was limited with respect to the type of data that could be considered. Antoniou et al. ([116]) extend the early formulation, by developing a general framework that can accommodate any type of sensor data, and by relaxing approximations (e.g. linear relationship between OD flows and flows).

4.4.2.3 Methodology

4.4.2.3.1 DTA calibration framework

Calibration of the parameters of the demand and supply models within DTA is an important and crucial step in any DTA application, as it ensures that the model accurately captures the traffic conditions prevailing in the network and is defined as follows. Calibration addresses the following problem: *Given a set of initial values for various parameters and typically available (aggregate) measures of flows, speeds and densities at sensor locations, determine the OD flows, route choice parameters, capacities and speed-density relationships, so that the discrepancy error between DTA output and observed values is minimized*.

Following [134], [135] and [136] the calibration problem is formulated as an optimization problem:

$$\min_{\mathbf{x},\beta,\gamma} \left[z_1\left(\mathbf{M}^{sim},\mathbf{M}^{obs}\right)^2 + z_2\left(\mathbf{x},\mathbf{x}^a\right)^2 + z_3\left(\beta,\beta^a\right)^2 + z_4\left(\gamma,\gamma^a\right)^2 \right] \quad (4.9)$$

where β represents the route choice parameters, γ represents the parameters in the supply simulator, $\mathbf{x}$ are the OD flows, M^{obs} are the observed measurements and M^{sim} are their simulated counterparts. Superscript α indicates a priori values and z_1, z_2, z_3 and z_4 are error functions. The objective function minimizes the discrepancy between various simulated and observed measurements M (for example, flows and speeds can be used as measurements) as well as deviations from a priori values, appropriately weighted. The above problem is solved iteratively using appropriate algorithms (see [136]).

A common special case of the problem involves the estimation of the OD flows $\mathbf{x}_h$ for interval *h*. In this case the general formulation reduces to:

$$\mathbf{x}_h = \arg\min \left[z_1\left(\mathbf{x}_h,\mathbf{x}_h^{\alpha}\right) + z_2\left(\sum_{p=h-p'} \mathbf{a}_h^p \mathbf{x}_p, \mathbf{y}_h^a \right) \right] \quad (4.10)$$

Function z_1 measures the Euclidian distance of the estimated flows $\mathbf{x}_h$ from their *a priori* values $\mathbf{x}_h^a$, while z_2 measures the distance of the measured counts $\mathbf{y}_h$ from their simulated counterparts. The assignment matrices $\mathbf{a}_h^p$ required by the OD estimation module are outputs of the supply simulator in DynaMIT-P, and are functions of the as yet unknown OD flows, the equilibrium travel times on each link (tt_l^{eq}), the route choice parameters β and the supply-side parameters γ:

$$\mathbf{a}_h^p = g\left(\mathbf{x}_p,\beta,\gamma,tt_l^{eq}\right) \quad (4.11)$$

The determination of consistent equilibrium travel times is part of the calibration process. Equilibrium is achieved when motorists follow the perceived shortest path to their destination. Travel times on those paths are time-dependent. These equilibrium travel times are themselves a function of the OD flows, route choice parameters and supply-side parameters:

$$tt_l^{eq} = h\left(\beta,\gamma,\mathbf{x}_p\right) \quad (4.12)$$

Equations (4.10) and (4.12) clearly demonstrate the fixed-point nature of the calibration problem. An iterative approach is therefore used to converge to a consistent calibration of the parameters and equilibrium conditions (see [137] and [136]).

4.4.2.3.2 Formulation and solution approaches

The problem can be formulated as a constrained optimization problem (as introduced above) and also as a state-space model. A large literature on optimization algorithms is available, and therefore we will not present them in detail. It is simply stated that traditional gradient based approaches may not be easily applicable for this problem, as the simulation-based assignment model means that there are no analytical gradients. Instead, numerical derivatives need to be computed, which require a large number of function evaluations, each of which implies a separate run of the simulator.

Another approach is to use stochastic approximation (SA) techniques. The widely used Finite Difference Stochastic approximation (FDSA) is an option with proven performance, but considerable computational complexity. Recently, an alternative algorithm called Stochastic Perturbation SA (SPSA) with excellent computational properties has been proposed by Spall ([138]).

In SA (stochastic approximation) algorithms, the gradient vector estimate is computed by a stochastic perturbation to the decision vector iterate. One SA algorithm that contrasts the SPSA in the way the gradient vector estimate is computed is the finite-difference stochastic approximation method (FDSA). In the FDSA, the gradient is computed by individually perturbing (by a stochastic perturbation) every parameter in the decision vector iterate of n dimensions, while keeping the other parameters in the decision vector fixed, and approximating the partial derivative. This random perturbation can follow any standard random variable distribution. Therefore, for an estimate of the entire gradient vector at any iteration, 2*n computations of the objective function are required. For multivariate problems with large dimensions for the decision vector, the computation time per iteration can itself be pretty large, thereby increasing the total computation time of the algorithm.

In comparison, the SPSA algorithm requires only two computations of the objective function at a given iterate for obtaining the estimate of the gradient vector, irrespective of the dimensional size of the decision vector. This approximate estimate is obtained by simultaneous perturbation of all the parameters in the decision vector. Although the FDSA is expected to give a more accurate and reliable estimate of the gradient vector at the end of every iteration, it does so at a high computational cost. Also, theoretical and empirical studies have shown that the convergence properties of the SPSA in terms of the number of iterations needed to come within an acceptable margin of the global optimum is comparable to that of the FDSA. Considering that the computational time per iteration for FDSA grows significantly with an increase in the number of dimensions of the decision vector, the SPSA is understandably a more efficient algorithm for solving large multivariate stochastic optimization problems. It has also been empirically shown, through application of these candidate algorithms on several standard stochastic problems, that the SPSA outperforms the FDSA in terms of the overall convergence speed, and was also a more robust option with increasing data size.

Having said that, the FDSA algorithm has been used for this application. Given the reasonable problem dimensions the use of the FDSA algorithm was practical, allowing the run of an entire iteration in a few minutes. (Note that each iteration involves 2*n runs of the simulator). As the problem size increases, however, the computational requirements of the FDSA algorithm are expected to become impractically large and then the SPSA will likely be a more reasonable approach.

The optimization problem can also be recast as a state-space problem. Most of the methodologies for the solution of state-space models are based on the Kalman Filter algorithm [9]. The original Kalman filter was developed for the solution of linear, dynamic state-space models and is based on a prediction-correction cycle. An initial estimate of the state vector and its covariance is obtained at each time interval by applying the transition equation to *predict* the state (and covariance) for the new interval (time update). The filter then uses the available measurements to *correct* this prior estimate of the state vector and its covariance (measurement update). The outcome of this process is a *filtered* state vector and updated covariance.

The problem of calibration of DTA parameters, however, is highly non-linear with respect to its measurement equations (which do not even have analytical representation). Hence, extensions of the original Kalman filter need to be employed for its solution. The most straightforward extension is the Extended Kalman Filter (EKF), which employs a *linearization* of the non-linear relationship to

approximate the measurement equation with a first-order Taylor expansion [139]. This algorithm can be applied to simulation-based systems (which do not have an analytical representation) by performing the linearization step using numerical derivatives. Naturally, the use of numerical techniques requires a large number of function evaluations (each implying a run of the simulator) thus making this technique potentially computationally expensive.

The Limiting Extended Kalman Filter is a special case of the EKF that offers considerable computational advantages. The most *computationally intensive* step in the EKF algorithm is the linearization of the measurement equation, as it requires the use of numerical derivatives. In real-time applications, it may be possible to replace the Kalman Gain matrix by a constant gain matrix considerably decreasing the computation time. The *limiting* Kalman Filter (LimEKF) is defined by replacing the Kalman gain matrix with its limit , called the *limiting* (or *stable*) *Kalman gain matrix* [139]. The limiting Kalman gain matrix can be computed off-line. Several strategies can be developed to improve the quality of the limiting Kalman gain. For example, the EKF could be run off-line, with each run producing a new Kalman gain matrix. These Kalman gain matrices could then be used to update the *limiting* Kalman gain matrix. Another strategy would be to consider only the last few Kalman gain matrices, i.e. use a type of moving average. Weighted averages (e.g. using lower weights for older gain matrices) can also be considered.

The Unscented Kalman Filter (UKF) [140] is an alternative filter for dynamic state-space models. The UKF uses a deterministic sampling approach (Unscented Transformation, UT) to represent a random variable using a number of deterministically selected sample points (often called *sigma points*). These points capture the mean and covariance of the random variable and, when propagated through the *true* nonlinear system, capture the posterior mean and covariance accurately to the second order (Taylor series expansion).

A presentation of the extensions of the Kalman filter for the on-line calibration of DTA models is presented in [141] while an application on the on-line calibration of speed-density relationships is available in [142].

4.4.2.4 The DynaMIT-P system

The general DTA framework has been presented above. In this section we introduce in more detail the DynaMIT-R state-of-the-art DTA system ([117] and [118]), which was used in this application. The key to the functionality of DynaMIT-P is its detailed network representation, coupled with models of traveller behaviour. Through an effective integration of historical databases with real-time inputs, DynaMIT-P is designed to efficiently achieve:

- Real time estimation of network conditions.
- Rolling horizon predictions of network conditions in response to alternative traffic control measures and information dissemination strategies.
- Generation of traffic information and route guidance to steer drivers towards optimal decisions.

To sustain users' acceptance and achieve reliable predictions and credible guidance, DynaMIT-P incorporates unbiasedness and consistency into its core operations. Unbiasedness guarantees that the information provided to travellers is based on the best available knowledge of current and anticipated network conditions. Consistency ensures that DynaMIT-P's predictions of expected network conditions match what drivers would experience on the network. DynaMIT-P has the ability to trade-off level of detail (or resolution) and computational practicability, without compromising the integrity of its output. Its important features include:

- A microscopic demand simulator that generates individual travellers and simulates their pre-trip and en-route decisions (choice of departure time and route) in response to information provided by available Advanced Traveller Information Systems (ATIS).
- Estimation and prediction of origin-destination flows.
- Simulation of different vehicle types and driver behaviours.
- A mesoscopic supply simulator that explicitly captures traffic dynamics related to the development and dissipation of queues, spillbacks, and congestion.

- Time-based supply simulation that simulates traffic operations at a user-defined level of detail that facilitates real-time performance. The level of detail could be determined by the choice of time steps and the level of aggregation of vehicles into homogeneous packets.
- Traveller information and guidance generation based on predicted traffic conditions to account for driver over-reaction to incident congestion. The system iterates between predicted network state, driver response to information and the resulting network state, towards the generation of consistent and unbiased information strategies.
- Adaptable to diverse ATIS requirements.
- Distinguishes between informed and uninformed drivers.
- Fuses historical, surveillance and O-D data to generate reliable OD estimates in real-time. The system records the results from previous O-D estimations to update O-D databases.
- Uses a rolling horizon to achieve efficient and accurate real-time estimations and predictions.
- Handles real-time scenarios including incidents, special events, weather conditions, highway construction activities and fluctuations in demand.
- Integrates with the MITSIMLab microscopic traffic simulator for offline evaluation and calibration.
- Ready for integration with external global systems (such as a TMC).

The state estimation module provides estimates of the current state of the network in terms of O-D flows, link flows, queues, speeds and densities. This step represents an important function of DTA systems, since information obtained from the traffic sensors can vary depending on the type of surveillance system employed. In the presence of *floating* or *probe* vehicles in the network, or the existence of Automated Vehicle Identification (AVI) information systems, detailed information about vehicle location and possibly origin and destination could be obtained. The DynaMIT-P system models can incorporate information obtained from such advanced surveillance systems ([115] and [116]).

4.4.2.5 Case study setup

4.4.2.5.1 Network and data

The state-of-the-art DTA model DynaMIT ([117] and [118]) has been calibrated using floating car data for a network in the centre of the city of Vienna. The 5.3km-long inner ring of Vienna has been selected, based on data availability. Most of the ring has three lanes, with few links having four lanes, and the condition of the road is fairly homogeneous, e.g. in terms of lane widths and surface conditions. There are also several traffic signals. The network has been coded in the DynaMIT format using 17 links (Fig. 4.51).

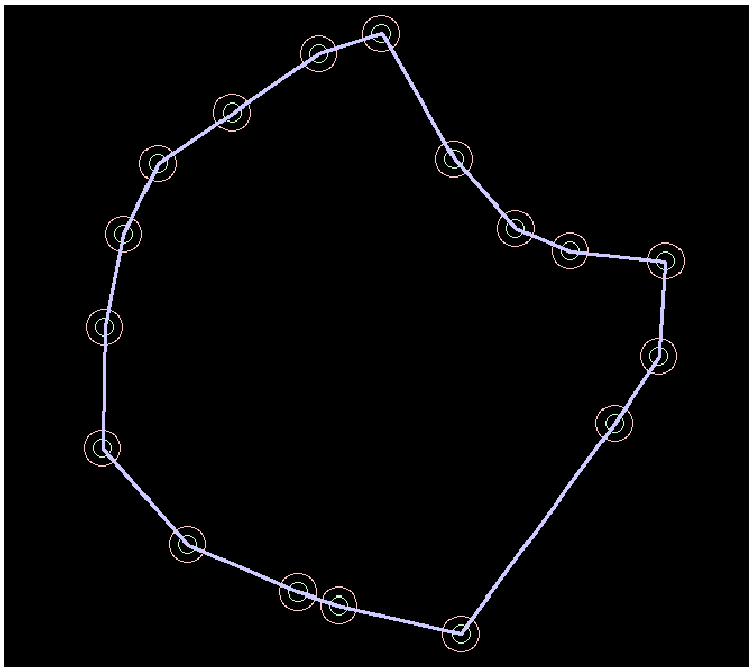


Fig. 4.51: Ring subnetwork representation

Sequential calibration has been performed, i.e. the supply was calibrated first and then the demand was calibrated using the output of the supply calibration as inputs. Only floating car data were available for the ring, as there were no reliable loop detectors operating in the network area.

Travel time data were available from the FLEET system, which uses a fleet of approximately 500 taxis operating in the entire Vienna area, equipped with GPS units 4. Data are collected and processed in a central system. The system can generate a large number of outputs based on this information, such as point to point travel time information, turning fractions, and partial OD information. However, at the time of this application only travel time information was available and therefore only this information has been used. Data from 2.5 months of regular week-days, i.e. no holidays, were used. Due to the small sample size in the considered network (average number of equipped vehicles in each 60-min interval was between two and four per link) the link travel times for 15-min intervals showed a high variability. Hourly travel time averages showed more stable patterns and were therefore chosen for this application.

4.4.2.5.2 Supply calibration

The calibration of the supply side involves the estimation of segment capacities and the estimation of the parameters of the speed-density relationships for the various segments. As it was mentioned above, no reliable loop detector data were available for the study network. As a result, calibration of the speed-density relationships was based on data from a nearby link with similar characteristics as the ring. The argument that supports this decision is that in general the same population of drivers in the two locations and since the two facilities have similar characteristics the driving dynamics experienced at the two locations are expected to be comparable. Capacities were determined using Highway Capacity Manual guidelines, and average green times for the signals in the ring.

The speed-density relationship used by DynaMIT has the following form:

$$u = \begin{cases} u_f & if \quad k < k_{\min} \\ u_f \left[1 - \left(\dfrac{\max(0, k - k_{\min})}{k_{jam}} \right)^{\beta} \right]^{\alpha} & otherwise \end{cases} \tag{4.13}$$

where u denotes the space mean speed, u_f is the free flow speed, k is the density, $k_{\min}$ is the minimum density, k_{jam} is the jam density, and α and β are model parameters. There are, therefore, five parameters to calibrate for each segment: u_f, $k_{\min}$, k_{jam}, α and β Given the lack of detailed information about the driving patterns throughout the network, it was assumed that the traffic dynamics (as far as the speed-density relationship is concerned) is uniform throughout the network. Therefore, a single speed density relationship has been calibrated and applied for the entire network. The non-linear function has been fit to the data and the results are visually presented in Fig. 4.52. The calibrated parameter values are shown in Table 4-5.

Table 4-5: Speed-density calibration estimated parameters

Parameter	Estimated value	Unit
uf	45	kmh
kmin	26	veh/km/lane
kjam	180	veh/km/lane
a	3	(unitless)
b	0.8	(unitless)

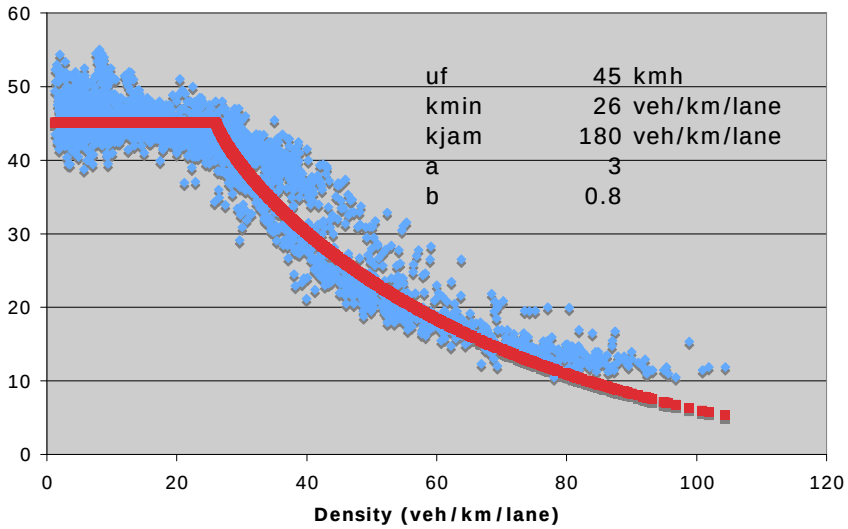


Fig. 4.52: Calibrated speed-density relationship

4.4.2.5.3 Calibration of OD matrices

One important input to the OD estimation problem is a set of starting values for the optimization. In some applications, a planning level OD matrix is available and is used as a starting point. Such a historical OD matrix with an appropriate resolution for the test network in the city of Vienna was not available. Therefore, arbitrary (yet reasonable, given the level of traffic and capacity constraints) starting demand values were selected. In the next paragraph, an approach to obtain a potentially better estimate of a starting demand matrix in the presence of link counts is outlined.

The partial OD matrix of the floating cars can be combined with point link counts to obtain an estimate of the OD matrix. This process involves two steps and an assumption. The assumption is that the

overall demand pattern for the floating car fleet is similar to that of the overall population. As the sample increases, this assumption becomes more and more reasonable. The first step in the process is to establish an OD matrix for the floating cars, in this case the equipped taxi fleet. For this purpose all recorded trips of a long period (e.g. 6 months) would be filtered, and the dynamic OD flow of taxis via the investigated OD paths would be recorded. As OD flows of taxis only constitute a small proportion of total OD flow, data have to be expanded. An expansion factor can be estimated from the relation of total link flows and link flows of taxis on the investigated links. The measured taxi OD flows times the expansion factor yields an a priori estimate of OD flows.

$$x_h^a = E_h * x_h^{pr} = x_h + u_h$$

x_h^a ... (n_{OD}*1) vector of a priori estimates of OD flows

E_h ... (n_{OD}*n_{OD}) matrix of expansion factors for partial OD flows of probe vehicles

x_h^{pr} ... (n_{OD}*1) vector of partial OD flow of probe vehicles

x_h ... (n_{OD}*1) vector of OD flows

u_h ... (n_{OD}*1) vector of measurement errors

The problem of OD estimation in this context is essentially an optimization problem where the unknown quantity is the matrix of OD flows that when loaded into the simulation model would result in measurements equal to the observed surveillance. Since the only available measurements in this context are travel times, this is what was used for the calibration.

Historical speed time-series of each link were calculated using the harmonic average of link speeds for each time interval in each day class. A classification scheme with 4 day-classes was used for grouping. The historical time-series of a working day (outside holiday periods) was used as indirect measurement for the historical OD matrix. Estimated OD flows were used as input for DynaMIT in order to simulate travel times for each time interval. Loading an OD matrix x_h to the simulation model S would then result in a set of simulated travel times tt_h^s for that interval h:

$$tt_h^s = S(x_h)$$

The measured travel times tt_h^m are equal to the simulated travel times $tt_h^s = S(x_h)$ corrupted by a vector of errors v_h :

$$tt_h^m = tt_h^s + v_h = S(x_h) + v_h$$

where

tt_h^m ... (n_l * 1) vector of measured travel times

$tt_h^s = S(x_h)$ ... (n_l * 1) vector of travel times simulated by DynaMIT

v_h ... (n_l * 1) vector of random error terms

The estimation of (true) OD flows can be solved as a constrained optimization problem. If all measurements of several time periods are used simultaneously in order to make optimal use of the information contained in the measurements, a simultaneous optimization formulation is obtained:

$$(\hat{x}_1,...,\hat{x}_n) = \arg\min\left[f1(x_1,...,x_n; x_1^a,...,x_n^a) + f2(tt_1^m,...,tt_n^m; tt_1^s,...,tt_n^s) \right]$$

for $x_h \geq 0 \quad \forall h = 1,...,n$.

In many practical applications, a sequential optimization approach is preferred. In that case, the OD flows for each interval are estimated once and kept fixed to their estimated values during the optimization of the following intervals. This can be justified as most information about each OD flow is obtained the first time it is estimated.

4.4.2.6 Results

The results of the calibration are shown in Fig. 4.53, Fig. 4.54 and Fig. 4.55. Each of these graphs shows the simulated travel times (obtained from the calibrated DTA model) against the observed travel times. Each point in these figures indicates the travel time for a particular link (the thin diagonal lines indicate where the points would lie in case of perfect fit). The results suggest that the calibration has indeed captured the prevailing travel times. A mild downward bias can be detected, which suggests that the calibrated travel times tend to be lower than the respective measurements.

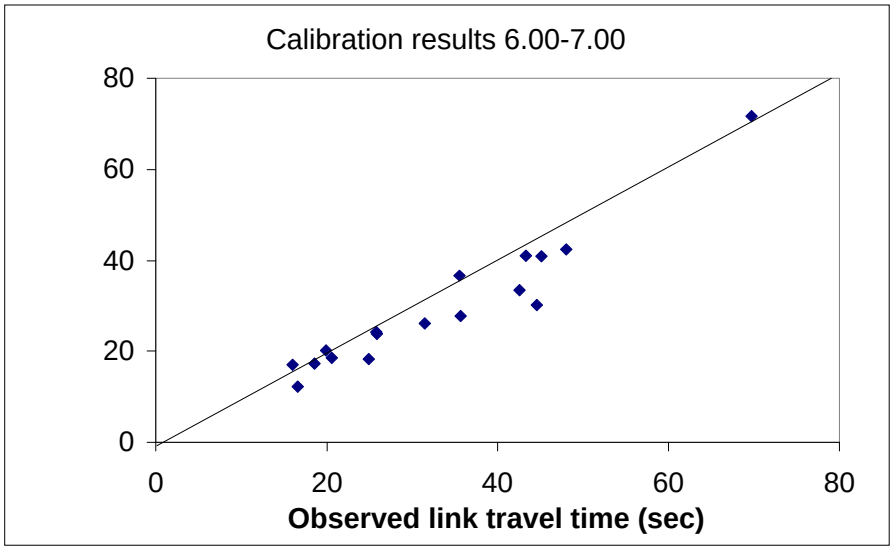


Fig. 4.53: Calibration results (6am-7am)

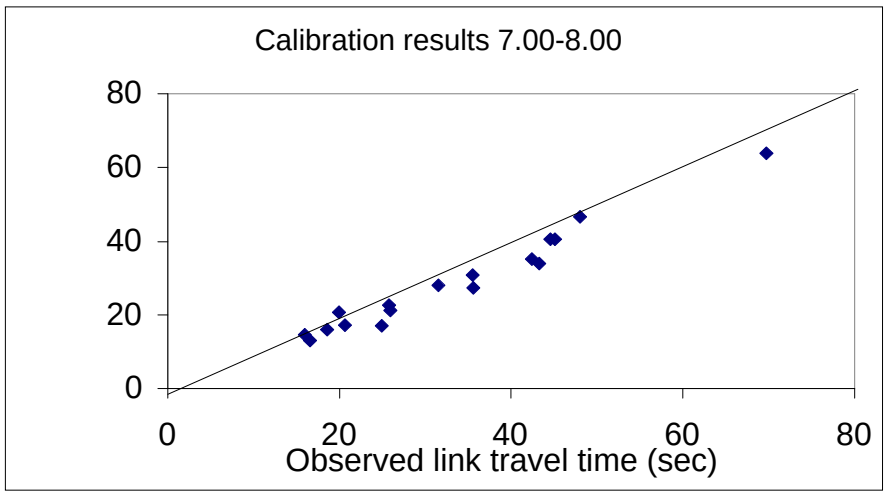


Fig. 4.54: Calibration results (7am-8am)

The fit was quantified using the Normalized Root Mean Square Error statistic:

$$RMSN = \frac{\sqrt{N \sum_N (y - \hat{y})^2}}{\sum_N y}$$

where N is the number of observations, y is an observation and y is a simulated value. The obtained $RMSN$ values for the time periods 6am-7am, 7am-8am, and 8am-9am were 0.166, 0.155, and 0.129 respectively.

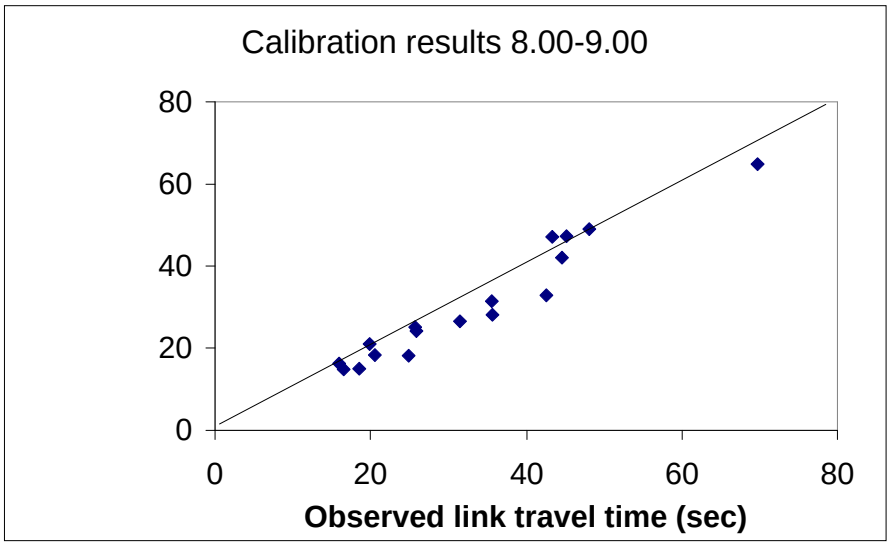


Fig. 4.55: Calibration results (8am-9am)

The calibrated flows for each link (all lanes) and time period are shown in Fig. 4.56. The estimated flows are reasonable and within the capacity of the network.

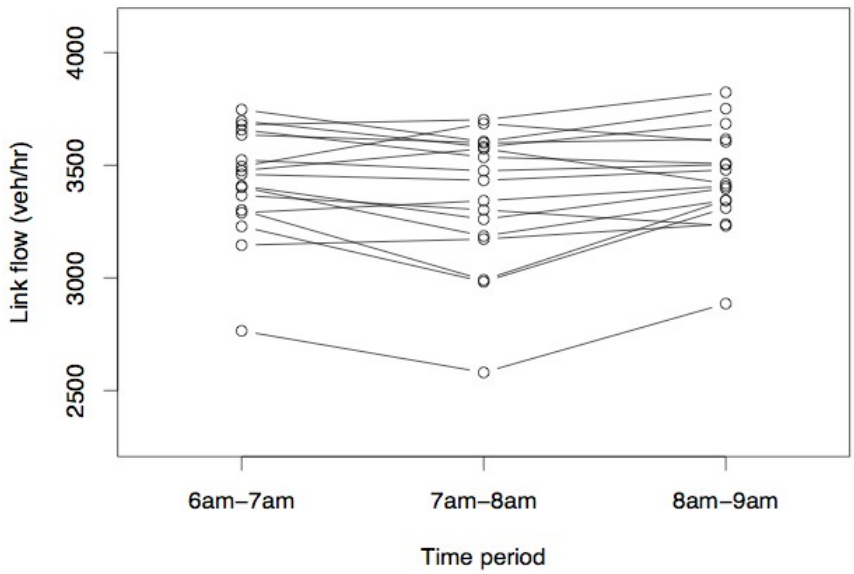


Fig. 4.56: Calibrated link flows

4.4.2.7 Conclusion

The use of floating car data for calibration of DTA models has been demonstrated. In particular, the methodological framework has been presented, and an application in a network from Vienna, Austria, has been described. The objective of the application was to illustrate how estimates of link flows can be obtained using floating car data as the primary source of information for a DTA model. While no measurements of the link flows are available to compare with the estimated flows, the results are encouraging and demonstrate that the calibrated model accurately matches the historical travel times. This is an interesting finding in terms of traffic management applications, as travel times are one of the most important outputs of DTA and traffic estimation and prediction models.

As more network operators and cities are moving away from conventional loop detector surveillance systems (that are expensive and tedious to install, maintain and operate) and towards emerging data collection systems, such as floating cars, the issue of using such data for traffic management purposes becomes increasingly relevant. The ability to use such data for the calibration and operation of traffic estimation and prediction systems may be an important driving force for these technologies and change the way traffic data is collected and used.

Next steps include the incorporation of further surveillance information as it becomes available. In particular, additional data can be inferred from floating car data, such as turning fractions, partial OD flows and sub-path flows. Furthermore, data from point sensors can also be incorporated to the framework, thus providing direct measurements of link flows, time-mean speeds and densities. This information is particularly interesting in the case of floating car data, as it can be used to generalize the information collected from the equipped vehicles (usually a small sample of the traffic) to the entire population.

Applications of this methodology to extended networks (both in Vienna as well as other cities with similar data) will be valuable in strengthening the impact of these findings.

4.5 Final Conclusion on Data Fusion for Traffic Indicators

In this chapter, data fusion proves to increase the accuracy of performance indicators, such as travel time, and to enable the estimation of performance indicators, for instance through DTA calibration.
Very encouraging results are obtained for travel time estimation by the fusion of loop and toll travel times. These results must still be confirmed and evaluated on a larger sample of real situations including various distributions of available toll travel times. It is also shown that the combination of toll travel times of different links can result in a reliable reference travel time. At this stage, the algorithms prove to be efficient although further improvements, already identified, can still be and will be made in the future.

Travel time and link flows fusion, under heterogeneous data source configurations, are also examined. It is found that, the outlined combining strategies improve the quality of the estimation and provide a comprehensive traffic estimate using floating car data as the primary source of information for a DTA model.

Hence, data fusion has proven very useful for traffic indicators. Building on this success, data fusion will be applied within INTRO to safety perspectives as well. The aim is to fuse already operational safety indicators in order to achieve a robust and comprehensive characterisation of traffic situation from a safety point of view. The results will be reported in D4.3.

4.6 References

[1] Federal Highway Administration, Data fusion for Delivering Advanced Traveller Information Services, US DoT and ITS Joint Program, pp. 247-254, 2003.
[2] Dasarathy B. V., *Decision Fusion*. IEEE Computer Society Press, 1994.
[3] Nifle. *Modélisation Comportementale en Fusion de Données: Application à la Détection et l Identification d Objets ou de Situations*. PhD thesis, University Paris-Sud, September, 1998.
[4] White, F. E., Data Fusion Lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories, Oct. 1991.

[5] L. Wald. A European proposal for terms of reference in data fusion. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 7, pp.651-654, 1998.
[6] A.N. Steinberg, C.L. Bowman, F.E. White. *Revisions to the JDL Data Fusion Model*. Proc. of SPIE AreoSense, Orlando, Florida, USA, pp.430-441, 1999.
[7] N.-E. El Faouzi. *Fusion de données: Concepts et méthodes*. INRETS-LICIT report, No 2002-10-15, December, 2000. EL Faouzi N.-E., Data *Fusion: Concepts and Methods*, Monograph: LICIT, INRETS-ENTPE, 2000 (In French).
[8] Durant-Whyte H. Sensor Models and Multisensor Data Fusion, *J. Robotics Res*., 6, pp. 97-113., 1988.
[9] Kalman, R. E., *A New Approach to Linear Filtering and Prediction problems*, Transactions of the ASME, Serie D: Journal of Basic Engineering, 82(1), pp. 35-45, 1960.
[10] D. Dubois, H. Prade, R.R. Yager. *Merging fuzzy information*. In: *Fuzzy Sets in Approximate Reasoning and Information Systems*. J. Bezdek, D. Dubois, H. Prade (Eds.), Kluwer, Boston, Mass., The Handbooks of Fuzzy Sets Series, 335-401, 1999.
[11] D. Dubois, H. Prade, H. Hguyen. *Possibility Theory, Probability and Fuzzy Sets: Misunderstandings, Bridges and Gaps*. In: Fundamentals of Fuzzy Sets. D. Dubois and H. Prade (Eds.), The Handbook of Fuzzy Sets Series, Kluwer Academic Pulications, Dordrecht, 1999.
[12] A.P. Dempster. *Upper and Lower Probabilities Induced by Multivalued Mapping*. Annals of Mathematical Statistics, Vol. 38, pp. 325-339, 1967.
[13] A.P. Dempster. *A Generalization of Bayesian Inference*. Journal of the Royal Statistical Society, 30B, pp. 205-247, 1968.
[14] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, Princeton, vol. 2702, N. J., 1976.
[15] D. Dubois and H. Prade. On the unicity of Dempster rule of combination. Int. J. Intelligent System, pages 133 142, 1996.
[16] F. Voorbraak. On the justification of Dempster s rule of combination. Artificial Intelligence, pages 171 197, 1991.
[17] L.A. Zadeh. On the Validity of Dempster s Rule of Combination of Evidence. Univ. California, Berkeley, 1979. ERL Memo M79/24.
[18] P. Smets. Constructing the pignistic probability function in a context of uncertainty. Uncertainty in Artificial Intelligence 5, M. Henrion, R.D. Shachter, L.N. Kanal, and J.F. Lemmer (Editors), Elsevier Science Publishers, pp. 29-39, 1990.
[19] G. Klir, T. Folger. Fuzzy Sets, Uncertainty and Information. Printice-Hall Inc., 1988.
[20] L. Zadeh. *Fuzzy Sets*. Information and Contol, N° 8, pp. 338-353, 1965.
[21] D. Dubois, H. Prade. The Treatment of Uncertainty in Knowledge-based Systems Using Fuzzy Sets and Possibility Theory. International Journal of Intelligent Systems, 3(2), pp. 141-165, 1988.
[22] L. Zadeh. *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets and Systems, Vol. 1, pp. 3-28, 1978.
[23] D. Dubois, H. Prade. A Review of Fuzzy Set Aggregation Connectives. Information Science, Vol. 36, n° 1-2, pp. 85-121, 1985.
[24] R.R. Yager. *Connectives and Quantifiers in Fuzzy Sets*. Fuzzy Sets and Systems, Vol. 40, n°1, pp. 39-75, 1991.
[25] Bloch, I. and Hunter (Eds) (2000). Fusion: General Concepts and Characteristics. International Journal of Intelligent Systems, 2000.
[26] El Faouzi, N.-E. *Data Fusion in Road Traffic Engineering: An overview*. Multisource, Multisensor Information Fusion Conference, Defense and Security Symposium, Orlando, USA, 12-16 April 2004.
[27] Federal Highway Administration. ITS System Architecture and Standards. *FHWA Ruling on Architecture Standards and conformance of projects to ITS Architecture Standards*, August, 2001.
[28] Asakura Y., Tanabe J., Lee Y. H. Characteristics of positioning data for monitoring travel bahvior. *7th World Congress on I.T.S*., Torino, Italy, 2000.
[29] Ygnace J.-L. et al.. Travel Time/Speed Estimates on the French Rhone Corridor Network Using Cellular Phones as Probes: STRIP Project. Final Report, European Community Research Program SERTI, December 2001.

[30] Waltz E., Llinas J., *Multisensor Data fusion*, Artech House, Boston, 1990.
[31] Sumner R. Data Fusion in PathFinder and TravTek. *Proceedings of the Vehicle Navigation and Information Systems Conference: VNIS 91*, pp 71-75, 1991.
[32] M. A. Abidi and R. C. Gonzalez. Data Fusion, in Robotics and Machine Intelligence. Academic. Press, 1992.
[33] Klein L. A., *Sensor technologies and data requirements for ITS*, Artech House, Boston, 2001.
[34] Durant-Whyte H. Sensor Models and Multisensor Data Fusion, *J. Robotics Res.*, 6, pp. 97-113., 1988.
[35] Manyika J., H. Durrant-Whyte., *Data Fusion and sensor Management: A decentralized Information- Theoretic Approach*, Ellis Horwood, London, 1994.
[36] Harris C., A. Bailey, T. Dodd. Multisensor data Fusion in Defense and Aerospace. *Journal of Royal Aerospace Society*, 162(1015), pp. 229-244, 1998.
[37] Dasarathy B. V. Sensor fusion, potential exploitation: Innovative architectures and illustrative applications. Proceedings of the IEEE, 85, pp. 24-38, 1997.
[38] EL Faouzi N.-E., J.-B. Lesort. Travel Time Estimation on Urban Networks from Traffic Data and On-Bord Trip Characteristics., .Proceedings of the second World Congress on ITS, Yokohama, Japan, pp. 88-93, 1995.
[39] Hashem S. Optimal linear combinations of neural networks. Neural Networks, 10(4), pp. 599-614, 1997.
[40] EL Faouzi N.-E., Heterogeneous data source fusion for impedance indicators, *IFAC Symposium on Transportation Systems*, Chania, Greece, Volume 3, pp 1375-1380, 1997.
[41] EL Faouzi N.-E.. Evidential Data Fusion for travel time estimation: a discrete case. *Preprint,* submitted for publication INRETS-LICIT, 2003.
[42] Okutani I. The Kalman Filtering approaches in some Transportation and road traffic problems. *Proceedings of the 10th ISTTT (International Symposium on Transportation and Traffic Theory)*, Gartner N. and Wilson N. Eds., Elsevier Science Publishing, pp. 397-416, Oxford, 1987.
[43] Dubois D., Prade H. *Possibility theory*. Plenum Press, New-York, 1988.
[44] Szeto M. W. and Gazis D. C. (1972) Application of Kalman Filtering to the Surveillance and Control of Traffic Systems, Transportation Science, Vol. 6, No. 4, pp. 419-439.
[45] B. V. Dasarathy, Information fusion, data mining, and knowledge discovery, Information Fusion, 4(1), 2003.
[46] Sethi V., N. Bhandari, F. Koopelman, J. Schofer. Arterial Incident Detection Using Fixed Detector and Probe Vehicle Data, *Transp. Research*,2(C)., pp. 99-112., 1995.
[47] EL Faouzi N.-E.. Travel Time Estimation via Evidential Data Fusion. *Recherche Transports Sécurité*, N° 68, pp. 15-30, 2000.
[48] Sussman J. *Introduction to Transportation Problems*, Artech House, Norwood, Massachusetts, 2000.
[49] EL Faouzi N.-E., Eds.. *Multiform Traffic Data Collection and Data Fusion in Road Traffic Engineering*. Proceedings of the INRETS workshop, N°87 INRETS publishing, 2003.
[50] Tarko A., N. Rouphail. Travel Time Fusion in Advance, Pacific *Rim TransTech Conference: A Ride into the Future*, July 25-28, 1993, Seattle, Washington USA.
[51] Thakuriah P., A. Sen. Data Fusion for Travel Time Prediction: A Statement of Requirements., *Advance Working Paper Series*, No. 22, Task TRF-TT-07, Urban Transportation Center, University of Illinois, Chicago, 1993.
[52] Berka S., Tian X., A. Tarko. Data fusion algorithm for Advance. Advance Working Paper, N° 48, may 1995.
[53] Dailey, D. et al. ITS Data Fusion. Research ProjectT9903, Task 9: ATIS/ATMS Regional IVHS Demonstration. Report N° WA-RD 410.1. Washington State, WADOT, 1996.
[54] Evanco, W.M. (1996). A Data Fusion Framework for Meta-Evaluation of Intelligent Transportation System Effectiveness. MITRE Center for Information Systems.
[55] Ben Aissa A., N.-E. EL Faouzi, E. Lefevre. Fusion of Classifiers via Evidence theory: a traffic engineering application. *Proceedings of the EGC 04 Conference (KDD and DM Conference),* 20-24 January 04, Clermont-Ferrand, France., 2004. (In French).
[56] Abe A. The correction of the forecasting travel time by using AVI data. *Proceedings of the I.V.H.S.*, 1997.

[57] Kühne R. D. Data fusion for DRG systems. *IFAC Symposium on Transportation Systems*, Chania, Greece, Volume 3, 1997.
[58] Choi K. Data fusion for generating the link travel time with insufficient data sources. Proceedings of the I.V.H.S., 1997.
[59] Choi K., Y. Chung. Travel Time Estimation Algorithm Using GPS probe and Loop Detector Data Fusion. *Proceedings of the 80th TRB Annual Meeting* (CD-Rom), Paper no. 01-0374. Washington, DC, Transportation Research Board, 2001.
[60] Ivan J. N., Schofer J. L., Coppelman F. S. , Massone L. L. E. Real-time Data fusion for arterial street Incident detection using neural Networks, *Transportation Research Record*, 1497, pp. 27-35., 1995.
[61] Ivan J. N. Neural Network Representations for arterial street Incident detection Data Fusion., *Transportation Research* C., pp. 247-254., 1997.
[62] Ivan J. N., Schofer J. L., Coppelman F. S. , Massone L. L. E. Real-time Data fusion for arterial street Incident detection using neural Networks, *Transportation Research Record*, 1497, pp. 27-35., 1995.
[63] Thomas, N. E., Multi sensor, multivariate, and multiclass incident detection system for arterial streets, *Proceedings of the 13th International Symposium on Transportation and Traffic theory (ISTTT)*, Lyons, J.-B. Lesort (Ed.), Pergamon, Elsevier, pp. 315-339, 1996.
[64] Klein, L. Dempster-Shafer data fusion at the traffic management center. *Transportation Research Board, 79th Annual Meeting*. Paper no. 00-1211. Washington, DC, Transportation Research Board, 2000.
[65] Cohen S., S. Bettiol. Algorithm fusion for better incident detection. Recherche-Transport-Sécurité (RTS), N° 55, pp. 39-46, 1997. (In French).
[66] Cohen S. Fusion of incident detection Algorithms. Proceedings of the INRETS workshop, N.-E. El Faouzi (Eds), N°87, pp. 109-117, 2003. (In French).
[67] Murphy, R. R. Sensor and Information Fusion for Improved Vision-Based Vehicle Guidance. IEEE Intelligent Systems. FL**:** 49 - 56, 1998.
[68] Pei, S.C. and Liou, L.G., Vehicle-Type Motion Estimation By the Fusion of Image Point and Line Features, Pattern Recognition, 31(3), pp. 333-344, Mar. 1998.
[69] Langheim, J., Buchanan, A. J., Lages, U. and Wahl, M. , CARSENSE New environment sensing for advanced driver assistance systems, in Proceedings of the IEEE Intelligent Vehicle Symposium 2001, Tokyo, Japan, 2001, pp. 89 94.
[70] C. Stiller, J. Hipp, C. Rössig, and A. Ewald, Multisensor obstacle detection and tracking, Image and Vision Computing, vol. 18, pp. 389 396, 2000.
[71] Mueck, J. Estimation methods for the state of traffic at traffic signals using detectors near the stop-line. *Traffic Engineering and Control*, December 2002.
[72] Friedrich B., R. Minciardi. Data Fusion Techniques for Adaptive Traffic Signal Control, *IFAC Symposium on Transportation Systems*, Tokyo, Japan, pp 53-55, 2003.
[73] Sohn S.-Y., Shin H. W. (2001) Data mining for road traffic accident classification. *Ergonomics*, 44, pp. 107-117, 2001.
[74] Sohn S. Y., S. H. Lee. Data fusion, ensemble clustering to improve the classification accuracy for the severity of road traffic accident in Korea. *Safety Science*, 41, pp. 1-14., 2003.
[75] Breiman L. Bagging Predictors. *Machine Learning, 26, pp. 123-140*, 1996.
[76] Breiman L. Arcing classifiers. *The Annals of Statistics* 26 (3), pp. 801 849, 1998.
[77] Wolpert D. H., Stacked Generalization. *Neural Networks*, n° 5, 1992, p. 241-259.
[78] Cremer M.and H. Keller. Dynamic identification of OD flows from counts at complex intersections. *Proceedings of the 8th International Symposium on Transportation and Traffic theory (ISTTT)*, Toronto 1981, University of Toronto Press.
[79] Cremer M. Determining Time dependent trip distribution in a complex intersection for traffic responsive control. *Proceedings on Control in Transportation Systems*, IFAC-IFIP-IFORS, Beden-Beden, Germany, 1983.
[80] Ashok K.,M. Ben-Akiva. Dynamic O-D matrix estimation and prediction for real-time traffic management systems.*Proceedings of the 12th ISTTT*., Daganzo, C. Ed., pp. 465-484, Elsevier Science Publishing, Oxford, 1993.
[81] Ben-Akiva, M. and Morikawa, T., Data Fusion Methods and Their Applications to Origin-Destination Trip Tables, Selected Proc. of the 5th World Conference on Transport Research, Yokohama, pp.279-293, 1989.

[82] Harrison J., F. Stevens. A Bayesian Approach to Short-Term Forecasting. *Operational Research Quarterly*, 22(4), pp. 341-362, 1971.
[83] Okutani I., J. Stephanedes. Dynamic Prediction of Traffic Volume through Kalman Filtering theory. *Transportation Research B*., 18(1), pp. 1-11, 1984.
[84] Vythoulkas P. C. Alternative Approaches to Short Term Traffic Forecasting for Use in Driver Information Systems. *Proceedings of the 12th ISTTT*, Berkeley, pp. 485-506, 1993.
[85] EL Faouzi N.-E., Nonparametric Traffic Flow Prediction using Kernel estimator, *Proceedings of the 13th International Symposium on Transportation and Traffic theory (ISTTT)*, Lyons, J.-B. Lesort (Ed.), Pergamon, Elsevier, p. 41-54, 1996.
[86] Bates, M. and Granger, W. J., The combination of forecasts, Operational Research Quarterly, 20, pp. 451-468, 1969.
[87] EL Faouzi N.-E., Combining Predictive Schemes in Short-Term Traffic Forecasting, *Proceedings of the 14th International Symposium on Transportation and Traffic theory (ISTTT)*, Jerusalem, 26-28 July 1999, A. Ceder (Ed.), Pergamon, Elsevier, p. 471-487.
[88] Cremer M., Schrieber S. Monitoring Traffic Load Profiles With Heterogeneous Data Source Configurations, Proceedings of the thirteenth isttt, Lyon, France, 1996.
[89] Choi D. B. On multisensor data fusion using attributes association for Intelligent traffic Congestion Information Inference. Proceedings of the I.V.H.S., 1989.
[90] De Mouzon O., El Faouzi N.-E., Nowotny B., Morin J.-M., Chung E. Data fusion for traffic and safety indicators: The Intelligent Roads perspectives. ITS Conference, London, 2006.
[91] De Mouzon O., El Faouzi N.-E., Innovative Processing of Toll Collection Data: Filtering, enrichment and real-time fusion with loop detectors for travel time estimation, INRETS-LICIT report n°0604, Oct. 2006.
[92] Nam D. H., Drew D. Traffic Dynamics : Method for Estimating Freeway Travel Times in Real Time From Flow Measurements., *Journal of Transportation Engineering*, pp. 185-191, 1996.
[93] Lawson, T.W., D.J. Lovell and C.F. Daganzo. Using the Input-Output Diagram to Determine the Spatial and Temporal Extents of a Queue Upstream of a Bottleneck. In *Transportation Research Record 1572*, TRB, Washington, D.C., 1997.
[94] Lin, W. H. and C. F. Daganzo, 1997, A Simple Detection Scheme for Delay-Inducing Incidents, *Transportation Research* A, 31 No. 2, pp. 141-155.
[95] Erera A., Lawson T., Daganzo C. A Simple, Generalized Method For Analysis Of A Traffic Queue Upstream Of A Bottleneck. In *Transportation Research Board, paper N*° 981292, Washington, D.C., 1998.
[96] Cassidy, M. & J. Windover. Methodology for assessing dynamics of freeway traffic flow, Trans. Res. Rec. 1484, 73-79., 1995.
[97] Daganzo C. F**.** A Simple Traffic Analysis Procedure*, Networks and Spatial Economics*, 1: 2001, pp. 77-101.
[98] Wang Y., and Nihan N. L. *Freeway Traffic Speed Estimation Using Single Loop Outputs.,* Department of Civil Engineering. University of Washington.
[99] Oh J.-S., Jayakrishnan R., Recker W., Section Travel Time Estimation from Point Detection Data, Technical report, UCI-ITS-WP-02-11, ITS, University of California, Irvine.
[100] Coifman, B., Improved Velocity Estimation Using Single Loop Detectors, *Transportation Research* A, 35 No 10, pp. 863-880, 2001.
[101] Coifman, B. Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors, Transportation Research A, 36 No 4, pp. 351-364, 2002.
[102] Dailey D. J. Travel Time Estimates using a series of Single Loop volume and occupancy measurements. 76th Annual Meeting of T.R.B., paper n° 970378, Washington D. C.
[103] Dailey D.J. A statistical algorithm for estimating speed from single loop volume and occupancy measurements. *Transportation Research* Part B 33, pp. 313-322, 1999.
[104] Hall F. L. and Bhagwant N. Persaud **-** *Evaluation* of Speed Estimates Made with Single Detector Data from Freeway Traffic Management Systems, *Transportation Research Record* 1232, pp. 9-16.
[105] Hellinga Bruce R., P.E. Improving Freeway Speed Estimates from Single-Loop Detectors*, Journal of transportation Engineering*, Vol. 128, No.1, January 1, 2002, pp56-67.
[106] Rice J., Erik Van Zwet**.** *A Simple and Effective Method for Predicting Travel Times on Freeways*, Department of Mathematics at the University of California. Berkley, USA.

[107] Zhang X., J. A. Rice**.** Short-term travel time prediction, *Transportation Research Part* C, pp.187-210, 2003.
[108] Hensher, D. A. Electronic toll collection. *Transportation Research*, Vol. 25A, no. 1, pp. 9-16. 1991.
[109] Ben Aissa A., El Faouzi N.-E., De Mouzon O., Sau J. Highway Travel Time Estimation and Prediction from Electronic Toll Collection System. INRETS-LICIT report, 2005.
[110] Ben Aissa A., El Faouzi N.-E., de Mouzon O., Sau J. Motorway travel time estimation by switching algorithms and its assessment using toll data. Proceedings of the 1st International Symposium on Freeway & Tollway Operations, Athens June 4-7 2006.
[111] Olivier DE MOUZON, Nour-Eddin EL FAOUZI, Anis BEN AISSA, Jacques SAU , Innovative travel time estimation for toll motorways by combining detectors and ETC data, LICIT, INRETS-ENTPE report, 2006.
[112] Morin J.-M., Morello S. (2006). Tools to Facilitate Operator Decision in Freeway and Tollway Traffic Control Centres. In Proceedings *1st International Symposium on Freeway and Tollway Operations*, Athens.
[113] El Faouzi N.-E. (2004). Aggregative Schemes for Estimation fusion: a road traffic application. In Proceeding *SPIE conference on Multisensor, Multisource Information Fusion, Architectures, Algorithms and Applications*, Orlando, Florida, USA.
[114] Chevreuil M., Buvat J. (2001). Travel Time Calculation on the SAPRR Network: the Stock Method and Future Prospects. In Proceedings *SERTI Meeting*. Avignon.
[115] Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos: Incorporating Automated Vehicle Identification data into Origin-Destination Estimation , Transportation Research Record, 2004, 1882, pp. 37-44.
[116] Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos (2006). Dynamic Traffic Demand Prediction Using Conventional and Emerging Data Sources. IEE Proceedings Intelligent Transport Systems, March.
[117] Ben-Akiva, M., Bierlaire, M., Burton, D., Koutsopoulos, H. and Mishalani, R. (2001). Network state estimation and prediction for real-time transportation management applications. Networks and Spatial Economics 1(3/4).
[118] Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. N., and Mishalani, R. (2002). Real-time simulation of traffic demand-supply interactions within DynaMIT. In Gendreau, M. and Marcotte, P., editors, Transportation and network analysis: current trends, pages 19-36. Kluwer Academic Publishers. Miscellenea in honor of Michael Florian.
[119] Ashok, K. and Ben-Akiva, M.: Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows , Transportation Science, 2000, 34(1), pp. 21-36.
[120] Antoniou, C., M. Ben-Akiva, M. Bierlaire and R. Mishalani. Demand simulation for dynamic traffic assignment. Proceedings of the 8th IFAC Symposium on Transportation Systems, Chania, Greece, June 1997.
[121] Bottom, J., Consistent Anticipatory Route Guidance; PhD Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, September, 2000.
[122] Mahmassani, H. S. (2001). Dynamic network traffic assignment and simulation methodology for advanced system management applications. Networks and Spacial Economics, 1(3):267 292.
[123] Cascetta, E.: Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator , Transportation Research, 1984, 18B(4/5), pp. 288-299.
[124] Ben Akiva M. Methods to combine different data sources and estimate origin-destination matrices. *Proceedings of the 10th ISTTT (International Symposium on* Transportation and Traffic Theory), Gartner, N. and Wilson, N. (Eds.) Transportation and Traffic Theory (Elsevier Science Publishing, Oxford, 1987), pp. 459-481, Cambridge, MA.
[125] Ben-Akiva, M. and Morikawa, T.: Data fusion methods and their applications to origin-destination trip tables . Proceedings WCTR, 1989.
[126] Cascetta, E., Inaudi, D., and Marquis, G.: Dynamic estimators of origin-destination matrices using traffic counts , Transportation Science, 1993, 27(4), pp. 363-373.
[127] Ashok, K.: Estimation and Prediction of time-dependent Origin-Destination Flows , PhD thesis, Massachusetts Institute of Technology, 1996.

[128] Ashok, K. and Ben-Akiva, M.: Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows , Transportation Science, 2002, 36(2), pp. 184-198.
[129] van der Zijpp, N.: Dynamic Origin-Destination Matrix Estimation on Motorway Networks , PhD thesis, Transportation Planning and Traffic Engineering Subsection of the Faculty of Civil Engineering of Delft University of Technology, 1996.
[130] Mishalani, R. G., Coifman, B., and Gopalakrishna, D.: Evaluating real-time origin-destination flow estimation using remote sensing-based surveillance data . Proceedings of the Transportation Research Board 82th Annual Meeting, Washington, D.C., 2003.
[131] Oh, C., and Ritchie, S. C.: Development of a methodology for designing advanced traffic surveillance systems toward origin-destination based traffic information . Transportation Research Record 1935, pages 37-46, 2005.
[132] Dixon, M. P. and Rilett, L. R.: Real-time origin-destination estimation using automatic vehicle identification data . Proceedings of the 79th Transportation Research Board Meeting, Washington, D.C., 2000.
[133] Kwon, J., and Varaiya, P.: Real-Time Estimation of O-D Matrices with Partial Trajectories from Electronic Toll Collection Tag Data . Transportation Research Record 1923, pages 119-126, 2005.
[134] Balakrishna, R. (2002) "Calibration of the Demand Simulator within a Dynamic Traffic Assignment System." Master's thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
[135] Kunde, K., Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assigment; MS Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, September, 2002.
[136] Balakrishna, R., H. N. Koutsopoulos and M. Ben-Akiva (2005) "Calibration and Validation of Dynamic Traffic Assignment Systems." H. S. Mahmassani (ed.) 16th International Symposium on Transportation and Traffic Theory, Maryland, pp. 407-426.
[137] Sundaram, S. Development of a Dynamic Traffic Assignment System for Short-Term Planning Applications; MS Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, June, 2002.
[138] Spall J.C. Implementation of the Simultaneous Perturbation Algoritm for Stochastic Approximation, IEEE Transactions on Aerospace and Electronic Systems, Vol.34 (3), 1998.
[139] Chui, C. K. and Chen, G.: Kalman Filtering with Real-Time Applications (Springer-Verlag, Berlin, 1999).
[140] Julier, S. and Uhlmann, J. (1997). A new extension of the kalman filter to nonlinear systems. In Proceedings of Aerosense: The 11th International Symposium of Aerospace/Defense Sensing, Simulation and Controls, Orlando, FL., volume Multi Sensor Fusion, Tracking and Resource Management II.
[141] Antoniou, C. "Online calibration for dynamic traffic assignment", PhD Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, September 2004.
[142] Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos. On-Line Calibration of Traffic Prediction Models. Transportation Research Record 1934, pp. 235-245, Washington D.C., 2005.
[143] Laborczi P., Nowotny B. (2004). Optimal Route Guidance on the Basis of Floating Car Data, 10th World Conference on Transport Research, Istanbul, Turkey, July 2004.

5 Conclusions

This report is the first of three deliverables of WP4. The scope of this deliverable are:

- Identification of key traffic and safety indicators to characterise traffic conditions.
- Development of methodologies to predict travel time using floating car data.
- Development of data fusion methodologies using data from multiple sources to predict travel time.

Identification of key traffic and safety indicators to characterise traffic conditions

As a first preparatory step of the project, an inventory of traffic and safety indicators, methods for their computation or estimation and existing data sources for sensor-based measurement of traffic parameters have been compiled. At this state of analysis seven indicators related to traffic state and safety have been identified and discussed. The indicators are stopping distance (related to road network global conditions), road capacity (related to supply), traffic density, travel time (related to demand-supply interaction), accident risk or accident rate, macroscopic accident probability, and microscopic speed inhomogeneity (directly related to safety).

Development of methodologies to predict travel time using floating car data

A fleet of floating car is needed to generate enough data for traffic management such as travel time prediction. Taxis are the most common fleet used for floating cars. Given that privacy is a sensitive issue, as the data collected by the on-board unit transmits the vehicle ID and positions, some trial experiments with unidentified floating car data. The unidentified floating car data system transmits only the position and speed of the vehicle but not the vehicle ID. This means that when all the data from an unidentified floating car fleet are combined, we will have a network with speed values at different points on the network. Clearly, methodologies to predict travel time for identified and unidentified float cars are different. This deliverable described two travel time prediction models, one for identified FCD and the other for unidentified FCD.

Identified floating car

The identified float car methodology was developed based on data from the FLEET system in Vienna, which has been operational since 2004. A taxi fleet of 800 taxis equipped with GPS receivers delivers positions and time stamps to a taxi management centre at an interval of 15 to 30 sec. From the taxi management centre, data are transmitted and evaluated at 15 min intervals.

Message data from floating cars are cleaned in a three-step process in order to reduce spatial and temporal errors of taxi messages.

- A map matching algorithm projects trip data (message points) to the road graph.
- If only start and stop of trips are known or if GPS messages are sparse, an optimised shortest path algorithm is applied for route search. The algorithm uses several labels in order to reduce spatial errors of routing. After routing, average velocity of each trip can be calculated.
- A travel time calculation algorithm calculates the estimated speed value for every road on a route. The average speed for the route is distributed to each road on the route. In order to take into account different speed levels of each road, average route velocity is distributed to each road proportional to its free speed or historical speed. Then space mean speed is calculated from all speed values available for a road in a time interval (of currently 15 min).

The investigation of route speeds has shown that variance plays an important role in the interpretation of results from route speed measurements. Average route speed on intra-urban roads can be estimated using timeseries modelling and linear Kalman filtering. Therefore, estimated route speed can be regarded as the relevant measurand which is related to other traffic parameters such as traffic density and traffic flow (concept of equilibrium speed).

As route speeds from a taxi fleet only constitute a small sample of experienced drivers, accuracy of route speed measurements has to be compared to an external reference. AVI measurements based on license plate recognition which represent a random sample of the normal population of drivers were used to evaluate whether route speeds represent the whole or average population of drivers. Route speed data which have been filtered with a Kalman filter are closely related to the average route speed of drivers measured by AVI. Floating car route speeds are only slightly overestimated average route speed of drivers measured by AVI. Mean error between average floating car route speed and average AVI route speed on one lane was +5.1 km/h on route 29 and +5.8 km/h on route 69.

Unidentified floating car

In Chapter 3.3 a new method is proposed to estimate travel times on links from unidentified probe observations of position and velocity. The new method is based on modelling the velocity profile as a continuous function along each link in a network (rather than piecewise constant functions in existing travel time estimation methods). The spatial modelling is done by first establishing a priori velocity profiles for sparse and dense traffic conditions, respectively, using historical data. Each new observation from an unidentified floating car is then combined with these a priori profiles (using linear interpolation) to obtain a current velocity profile. Various approaches to combine travel time estimates from such individual velocity profiles are investigated and compared, as well as comparisons with existing methods.

The θ function associated with the observation of speed, v at position, x is used to obtain the current speed profile using the equation:

$$f_\theta(x) \equiv g(x) + \theta \cdot (h(x) - g(x)) = v$$

where

$g(x)$ corresponds to historical speed profile for light traffic conditions;

$h(x)$ corresponds to historical speed profile for heavy traffic conditions; and

θ is a parameter for linear interpolation.

The results based on simulated data show that the new method generally reduce the relative error in travel time estimate as compared to existing methods based on piecewise constant velocity profiles. However, implementation details, especially the handling of low velocities, may strongly affect results, particularly in very congested traffic situations.

Development of data fusion methodologies using data from multiple sources to predict travel time.

Increasing data sources such as sensors, databases and experts, often inconsistent and sometimes contradictory, make the synthesis tremendously difficult. In order for the data sets to be fully useful, data sources need more and more complex processing techniques, to take into account data qualities and imperfections of each source. One solution is based on data fusion and this deliverable focused on data fusion for two traffic performance indicators: Travel time and traffic density. The data fusion processes developed in the INTRO project aim to provide a consistent and comprehensive picture of network conditions.

Firstly, it is shown that the combination of toll travel times of different links can result in a more reliable reference travel time. Secondly, very encouraging results are then obtained for travel time estimation by the fusion of loop detector data and toll collection data. The integration of toll collection data into the cumulative traffic counts method (classically based on loop detector data only) improves both the accuracy and the robustness of travel time estimation:

- Accuracy:

  - During congestion, the estimated travel time is better than with the classical cumulative traffic counts method or with the speed extrapolation method (which sometimes misses the congestion).
  - During free flow traffic, the estimated travel time is better than with the classical cumulative traffic counts method and close to the speed extrapolation method.
  - Overall, the estimated travel time is closer to the reference.
- Robustness:
  - Travel time estimation is correct or quickly returned to correct, even when there are counting mistakes in the loop detector data.
  - The same applies for wrong initialization too.

In this project, toll collection data is used both to make corrections to the classical cumulative traffic counts method and to decide when to use this method or the speed extrapolation method (which still performs slightly better during free flow traffic). Overall, this research shows the potential of toll collection data for travel time purposes, both for reference travel time computation and for travel time estimation. The results must still be confirmed and evaluated on a larger sample of real situations including various distributions of available toll travel times.

Finally, travel time and link flows fusion, under heterogeneous data source configurations, are also examined. It is found that, the outlined combining strategies improve the quality of the estimation and provide a comprehensive traffic estimate using floating car data as the primary source of information for a Dynamic Traffic Assignment (DTA) model.

This work shows how estimates of link flows can be obtained using floating car data as the primary source of information for a DTA model. While no measurements of the link flows are available to compare with the estimated flows, the results are encouraging and demonstrate that the calibrated model accurately matches the historical travel times. This is an interesting finding in terms of traffic management applications, as travel times are one of the most important outputs of DTA and traffic estimation and prediction models.

Building on this success, data fusion will be applied within INTRO to safety perspectives as well. The aim is to fuse already operational safety indicators in order to achieve a robust and comprehensive characterisation of traffic situation from a safety point of view. The results will be reported in deliverable D4.3.

6 Annexe I - Overview of traffic performance indicators and safety indicators

(1) category
S static (network-related parameter)
I instantaneous
T traffic configurations (summer migration, vacation, and special events)
P overall performance and system monitoring (derived from the two other categories and from archived data using statistical analysis)

(2) type
Ro Road network global conditions
De Demand
Su Supply
D+S Demand/Supply interaction
Sa Safety

(3) relevance for road safety indicator

(4) relevance for traffic performance indicator

(5) Geographical scope
point, segment, path, sub-network and network

(6) Temporal scope (typical measurement period)
s, min, hour, day, month, year

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
S	Su	Winter viability level of service							

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
S	Su	Road or bend geometry	parameters influencing accident rate	-	Road section before bend, width, radius, inclination, road surface state	segment	Update upon change in geometry	record of road operator	-
I	Ro	Air temperature (Temperature of road surroundings)	Important parameter for glazed frost forecasting	-	[° C]	Sub-network	min, hour, day	measurement	-
I	Ro	Road surface temperature	Important parameter for glazed frost forecasting	-	[° C]	Segment	min, hour, day	measurement	-
I	Ro	State of the road surface	?	Parameter for road maintenance, winter viability	dry, damp, wet, icy, salted, etc.	Segment	min, hour, day	Observation, measurement (road surface temperature, air temperature, wind, humidity, dew point temperature)	classification or calculation from several parameters
I	Ro	Wind speed, direction and speed gradient	Useful parameter for glazed frost forecasting	-	[m/s], [direction]	Sub-network	hour, day	measurement	-
I	Ro	Humidity	Useful parameter for glazed frost and precipitation (rain, snow) forecasting	-	[%]	Sub-network	min, hour, day	measurement	-
I	Ro	Dew point temperature (temperature for a given humidity at which relative humidity will reach 100 %)	useful parameter for glazed frost and precipitation (rain, snow) forecasting	-	[° C]	Sub-network	min, hour, day		calculation from humidity

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
I	Ro	Type and amount of precipitation	increases rate of accidents	decreases capacity and speed, total trip (demand) also decreases	dew, rain or snow hard, normal, little hourly, daily amount [l/m2], continuous hours of rain [h], etc.	Sub-network	min, hour, day	measurement	-
I	Ro	Visibility (greatest distance in a given direction at which it is just possible to see and identify an object on the road)	?	decreases capacity and speed	[m]	Sub-network	min, hour	Measurement and time of Sunrise & Sunset	-
I	De	Composition of traffic (ratio of trucks, two-wheeled vehicles, pedestrians)	Severity of accident increases when heavy vehicles are involved	Parameter for calculation of capacity	percentage of cars / heavy vehicles [%]	Point, segment, path	hour	Measurement of vehicle length and number of axles (e.g. induction loops, infrared, laser,) and weight (e.g. WIM)	Classification according to vehicle length and/or weight and/or number of axles
T	De	Origin-Destination matrix (matrix of origins and destinations of traffic participants) / type of trips (transit, commuters,)	-	Input parameter for traffic models	,	network	hour	Traffic surveys (questionnaires), toll data, traffic data	Estimation method
T	De	Predicted volume [veh/h]				Network, sub-network		Simulation	Estimation method

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
I/T	De	Directional proportion [%]				Sub-network, intersection		Simulation or vehicle count (e.g. induction loops, infrared, laser)	Estimation method or calculation from vehicle count
I	Su	Type and duration of incidents and restrictions (e.g. work zones)	Important parameter for two-wheeled traffic	Relevant for reduction in capacity	number of lanes and length of road closed [m], time to clear incidents [h], residual static road capacity [veh/h]	segment	hour, day	record of the incident kept by traffic operator or traffic message	-
I	Su	Free flow speed / speed limit (maximum observed mean speed)	?	Parameter for density-speed relation	[km/h]	Segment, point	Hour, day	traffic flow and local speed, travel speed of floating cars	distribution analysis of flow and speed data
I	Su	Signal timing	-	Limits intersection capacity on urban streets	Length of green and red cycle [s], offset [s], control type fixed time, actuated, coordinated, etc.	point	min	controlled value	-
I	D+S	Number of vehicles in platoon	Relevance for safety under investigation	-	n	Point, segment	min	-	Modelled by microsimulation
I	D+S	Observed flow	Relevant for time headways, time to collision	-	[veh/h]	Point, segment	min, hour	measurement	-

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
I	D+S	Travel time / average travel speed (Travel time or average travel speed for a road section)	-	Average travel speed and its variance are parameters depending on traffic density	Value [s] or [km/h], standard deviation	section, path, given queue (crossing time)	min	measurands at data collection stations, travel speed of identified floating cars	1) modelling (speed method, stock method, real-time traffic simulation, statistical methods) 2) travel speed of identified floating cars 3) mix (fusion)
I	D+S	Local speed	-	depending on traffic density	Value [km/ h], space/time mean, (median), standard deviation, distribution	point	min, h	local measurement (based on dual loops and vehicle length or inferred from travel times)	-
I	D+S	Speed in relation to road geometry	Relation to accident rate (especially in bends)	-	Value [km/ h]	segment	Month, year	Speed, road geometry	Speed related to road geometry

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
I	D+S	Single vehicle speed / speed variation (single vehicle speed and speed variation for a single vehicle along its route)	Relevance for traffic safety under investigation	Depending on traffic density, travel time reliability	Value [km/h], standard deviation, distribution at single point	Point, segment	min	Single vehicle speed	Speed variation
I	D+S	Average speed distribution within lane / between lanes	Relative speed in combination with headway determines time to collision, speed difference between lanes implies some lane changing	?	Speed distribution within lane, distribution between lanes [km/h]	Point, segment	min, hour	measurement	Calculation of distribution
I	D+S	Headway distribution	determines the ability of the driver to stop without colliding with the front vehicle. High percentage of short headways implies higher accident rates	-	Distribution [s/veh]	segment	min	Single vehicle measurement at local detector stations	Calculation of distribution
T	Ro	Weather conditions, season	Useful if no instantaneous indicators are available for investigated road section	-	Type and characteristics of season	network	day, month	observation	-
I/T	De	Current or predicted demand profile		-	Volume time series	network	Hour, day, week or month	Measurement, or estimation and hour-day-week-month category	-

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
T	Su	Legal restrictions for users (e.g. concerning vehicle types and vehicle occupancy)	-	additional information on vehicle composition or modal split		network	year	Work legislation and regulation, day of the week	-
T	Sa	Rate of accidents	Important parameter for investigation of accident risk Rate of accident during weekend is higher than weekday. Supposed reason: driver familiarity with the road and different type of trips	-	[accidents per location/link/area], [accidents per daytime/week/year]	Segment, sub-network, network	month, year	Police reports, hour-day-week-month category, visibility, weather conditions	Calculation of frequency
P	D+S	Number of vehicles in the network (total number of vehicles simultaneously present)	-	Parameter for traffic performance	Value [veh]	network, path, segment	Min, hour, defined periods (e.g. peak-periods)	density measured at detector stations, flow and local speed at data collection stations	1) estimation from density measured at detector stations 2) stock estimation (input/output method) 3) simulation results

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
P	D+S	Total Mileage / Vehicle-kms travelled (VKT)	-	Parameter for traffic performance	Value [veh * km]	network	hour, defined periods (e.g. peak-periods)	flow and local speed at data collection stations	1) measurement of mean traffic volume in each section multiplied by the section length 2) simulation
P	D+S	Total time spent in the system	-	Parameter for traffic performance (integral of density)	Value [veh * h]	network	hour, defined periods (e.g. peak-periods)	flow and local speed at data collection stations	1) estimation of mean number of vehicles in each section multiplied by the time period 2) simulation

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
P	D+S	Mean travel speed	-	Parameter for traffic performance	[km/h]	network	hour, defined periods (e.g. peak periods)	flow and local speed at data collection stations	1) Total distance travelled divided by total time spent 2) simulation
P	D+S	Lost time or delay due to congestion	-	Parameter for traffic performance	Value [veh * h]	Segment, sub-network (e.g. at event), network	defined periods (e.g. peak periods, events)	flow and local speed at data collection stations	1) Average number of additional vehicles in the congestion (in comparison to normal traffic) times congestion duration 2) sum of delays multiplied by traffic volumes 3) simulation

Category (1)	Type (2)	Indicator / parameter	Relevance for road safety indicator (3)	Relevance for traffic performance indicator (4)	attributes (value, distribution, trend, reliability,)	Geographical scope (5)	Temporal scope (6)	Input parameters/ measured data	Calculation or estimation method
P	D+S	Delay to mileage ratio (total delay due to congestion in relation to total mileage / VKT)	-	Parameter for traffic performance	[h / (veh*km)]	network	hour, defined periods (e.g. peak periods)	flow and local speed at data collection stations	1) total delay due to congestion in relation to total mileage 2) simulation
P	D+S	Amount of congested road network (Number of congestions and total amount of queue length over a given period, normalised per lane)	-	Parameter for traffic performance	Number [n / month], intensity [hour * km], duration [h], extent [km], variability, accessibility	Segment, network	month, defined periods (e.g. peak periods)	Operators reports, flow and local speed at data collection stations	1) calculation from operators reports, 2) calculation from flow and local speed at data collection stations 3) simulation

References

Literature related to traffic performance indicators and safety indicators

The selected literature references to some of the following Performance and Safety Indicators:

Su1 Road Capacity

D+S1 Traffic State
D+S2 Traffic Density
D+S3 Level of Service
D+S4 Queue Length
D+S5 Ratio of AverageTraffic Speeds/Travel Time Delay
Sa1 Accident Risk/Accident Rate
Sa2 Macroscopic Accident Probability
Sa3 Stopping Time/Time Headway
Sa4 Microscopic Lane Changing Indicator
Sa5 Transport Force
M Method

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
R1 Stopping distance	Calibration of the bunched exponential distribution of arrivals headway	Akcelik, R.	Chung, E.		Road & Transport Research 3(1),pp 42-59	E	1994	Headway
R1 Stopping distance	Time Headway on French Rural Roads. 2003, session traffic theory , Tokyo, Août 2003	Aron, M.	Biecheler, M-B.	Peytavin, J-F.	Proc of 10th IFAC Symposium on Control in Transportation Systems,	E	2003	Headway
R1 Stopping distance	Close Following Drivers on Two-Lane Highways	Rajalin, S.	Hassel, S.O.	Summala H.	Accidents Analysis and Prevention Vol 29 N° 6, pp 723-729	E	1997	Short headway & accident
R1 Stopping distance	Risky driving related to driver and vehicle characteristics	Evans, L.	Wasielewski, P.		Accidents Analysis and Prevention Vol 15 pp 121-136	E	1983	headway violation & accident
R1 Stopping distance	Do accident-involved drivers exhibit riskier everyday driving behaviour?	Evans, L.	Wasielewski, P.		Accidents Analysis and Prevention Vol 14 N°1, pp 57-64	E	1982	headway violation & accident
R1 Stopping distance	Calibration of headway distributions	Hagring, O.			9th Meeting of the EURO Working Group on Transportation & ITS	E	2002	headway calibration

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
Su1 - Road Capacity	Empirical Study of traffic Features at a Freeway Lane Drop	Robert, L.	Bertini, M.	Leal, T	Journal of Transportation Engineering, June 2005, Vol 131,N°6	E	2005	Capacity and infrastructure
Su1 - Road Capacity	Moving Bottlenecks in the LWR Model+C1	Leclercq, L.	Chanut, S.	Lesort, J-B.	Transp. Research Record Vol 1883 (2003) PP 3-13	E & F	2003	Capacity and traffic unhomogeneity
Su1 - Road Capacity	Un modèle de la capacité d'écoulement et une méthode de mesure (a model and a rigorous measurement method model for flow and capacity)	Leurent F.			Modélisation du trafic, actes du groupe de travail 2002. Actes INRETS #97, pp. 9-30, INRETS, France	E & F	2002	
D+S2 Route speed / Travel time	Hastighetsförändringar och trafiksäkerhetseffekter (Speed changes and traffic safety effects,)	Nilsson, G.			VTI	S	2000	The relation between average speed and traffic safety
D+S2 Route speed / Travel time	Travel Time Estimation on Urban Networks from Traffic Data and On-Bord Trip Characteristics	El Faouzi, N.-E.	Lesort, J.-B.	-	Proceedings of the 2nd World Congress on ITSYokohama, Japan	E	1995	travel time estimation
Sa1 Accident Risk/Accident Rate	The Potential for Real-time Traffic Crash Prediction	Abdel-Aty, M.	Pande, A.	Hsia,L. Y. ; Abdalla, F.	The journal on the web, December 2005	E	2005	loop detector data for predicting crashes
Sa1 Accident Risk/Accident Rate	On establishing relationship(s) between freeway safety and peak period operations: performance measurement and methodological considerations	Hugues, R.	Counci,l F.		78th Annual meeting of the Transportation research Board(TRB), Washington,DC,USA,1999	E	1999	loop detector data for predicting crashes

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
Sa1 Accident Risk/Accident Rate	Real-time crash prediction model for the application to crash prevention in freeway traffic	Lee, C.	Saccomanno, F.F.	Hellinga, B.	82th Annual meeting of the Transportation research Board(TRB), Washington,DC,USA,2003	E	2003	loop detector data for predictin g crashes
Sa1 Accident Risk/Accident Rate	Relationships among urban freeway accidents, traffic flow, weather and lighting conditions	Golob, T.	Recker, W. W.		California PATH Working paper UCB-ITS-PWP-2001619, Institue of transportation Studies, university of California, Berkeley, 2001	E	2002	traffic & accidents
Sa1 Accident Risk/Accident Rate	Safety aspects of freeway weaving sections	Golob, T.	Recker, W. W.	V. M. Alvarez	82th Annual meeting of the TRB, Washington DC,USA,2003	E	2003	classifica tion of weaving sections
Sa1 Accident Risk/Accident Rate	A method for relating type of crash to traffic flow	Golob, T.	Recker, W. W.		82th Annual meeting of the TRB, Washington,DC,USA,2003	E	2003	traffic & accidents
Sa1 Accident Risk/Accident Rate	Analysis of Crash Precursors on Instrumented Freeways.	Lee, C.	Saccomanno, F.F.	Hellinga, B.	Transportation Research Record 1784 2002 pp 1-8	E	2002	traffic indicator s & accidents
Sa1 Accident Risk/Accident Rate	Factors Affecting Speed Variance and Its Influence on Accidents.	Garber, NJ.	Gadiraju, R.		Transportation Research Record 1213, 1989, pp. 64-71	E	1989	Speed variance and accidents
Sa1 Accident Risk/Accident Rate	Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments	Lord, D.	Manar, A.	Vizioli, A.	Accidents Analysis and Prevention Vol 37 n°1(January 2005), pp 185-189	E	2005	effect of traffic flow & density on accidents and their severity
Sa1 Accident Risk/Accident Rate	A method for identifying rear-end collisions risks using inductive loop detectors	Cheol, O.	Park, S.	Ritchie, S.G.	Accidents Analysis and Prevention Vol 38 n°2(March 2006), pp 295-301	E	2006	loop detector data for predictin

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
								g crashes
Sa1 Accident Risk/Accident Rate	Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model	Yan, X.	Radwan, E.	Abdel-Aty, M.	Accidents Analysis and Prevention Vol 37 n°6 (Nov 2005), pp 983 -995	E	2005	influence of different factors on accidents
Sa1 Accident Risk/Accident Rate	Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows	Hiselius, L.W.			Accidents Analysis and Prevention Vol 36 (2004), pp 985 -992	E	2004	impact of traffic on accidents
Sa1 Accident Risk/Accident Rate	Effect of roadway geometrics and environmental factors on rural freeway accident frequencies	Shankar, V.	Mannering, F.	Barfield, W.	Accidents Analysis and Prevention Vol 27 N°3 (1995), pp 371-	E	1995	impact of infrastructure on accidents
Sa1 Accident Risk/Accident Rate	Real Time Hazardous Traffic Conditions Warning System : Framework and Evaluation	Oh, C.	Oh, J-S.	Ritchie,S.G.	IEEE Transactions on ITS Vol 6 N°3 pp265-272	E	2005	traffic conditions and safety
Sa2 Macroscopic Accident Probability	Exploring the Relationship Between Freeway Speed Variance, L+C18ane Changing, and Vehicle Heterogeneity	Park, S.	Ritchie S.G.		83th Annual meeting of the Transportation research Board(TRB), Washington,DC,USA,2003	E	11-15 January 2004	speed variance as a function of vehicle heterogeneity and driver behaviour
Sa2 Macroscopic Accident Probability	Evaluating speed consistency between successive elements of a two-lane rural highway	Park, Y-J-.	Saccamann,o F.F.		Transp. Research A. Vol 40 n°5 (June 2006) pp 375-385	E	2006	Speed homogeneity
Sa2 Macroscopic Accident Probability	Impact of Differential Speed Limits on the Speed of Traffic and the Rate of Accidents	Garber, N. J.	Gadiraju, R.		Transportation Research Record 1375	E		Speed and accident
Sa2 Macroscopic Accident Probability	Synthesis of studies on Speed and Safety.	Feng, C.			80th Annual Meeting of the Transportation Research Board	E		

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
Sa3 Microscopic Speed Inhomogeneity Indicator	Extended time-to-collision measures for road traffic assessment	Minderhoud M. M.	Bovy, P. H. L.		Accident Analysis and Prevention 33 (89-07)	E	2000	Microscopic safety indicators
Sa3 Microscopic Speed Inhomogeneity Indicator	Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling	Archer, J.			Department of Infrastructure, Royal Institute of Technology, Stockholm	E	2005	Microscopic safety indicators
Sa3 Microscopic Speed Inhomogeneity Indicator	Modeling drivers' acceleration and lane changing behaviour	Ahmed, K.I			Thesis BUET (Bengladesh Univ. of Eng. and Technology 1991	E	February 1999 MIT	microscopic traffic and drivers performance models
Sa3 Microscopic Speed Inhomogeneity Indicator	Models of freeway lane changing and gap assistance behaviour	Ahmed, K.I	Ben-Akiva, M.E.	Koutsopoulos, H.N. ; Mishalani , R.G.	Transportation and Traffic Theory, PP. 501-515, Pergamon in J. lesort (Ed.)1996	D	1996	Lane changing behaviour
Sa3 Microscopic Speed Inhomogeneity Indicator	A model for the structure of lane changing decisions	Gipps,P.G			Transportation research 20B(5), 403-414, 1986	E	1986	lane changing behaviour
Sa3 Microscopic Speed Inhomogeneity Indicator	Lane Changing in traffic streams	Laval, A.J.	Daganzo, D.C		Transp. Research B Vol 40 n°3 (March 2006) pp251-264	E	2006	capacity restriction
Sa4 Transport Force	Time headway and speeds: what safety stake on Normandy roads?	Aron, M.	Biecheler, M-B.	Peytavin, J-F.	Internal INRETS Report	F	2004	accidents and safety indicators
Sa4 Transport Force	Evaluation of traffic safety using microsimulation	Huguenin, F.	Torday, A.	Dumont, A-G.	5th Swiss Transport Research Conference	E		microscopic simulation and accidents
Sa4 Transport Force	A microscopic analysis of traffic conflict	Uno, N.	Iida, Y.	Isubo, S.	9th Meeting of the EURO	E	2002	traffic

Referenced Performance and Safety Indicator	Title	Author 1	Author 2	Author 3	Distributor	Language (E English, F French, S Swedish)	Date	Topic 1
	caused by lane-changing vehicle at weaving section				Working Group on Transportation & ITS			and accidents
Sa4 Transport Force	Traffic Conflict Analysiis using Vehicle Tracking System with Digital VCR and New Conflict Indicator under High Speed and Congested Traffic Environment	Wakabayashi, H.	Renge, K.			E	2003	Conflict analysis
Sa4 Transport Force	Using Individual Measurements of traffic on motorways	Aron M.	Seddiki E.	Hau-Chow-Wah L.	1st World Congress ATT/IVHS,Paris	E	1994	Microscopic traffic & safety indicators
Overview of accident analysis	Observational before-after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety.	Hauer, E.			Book in Pergamon	E	1997	statistical methods for safety assessment
Overview of accident analysis	Confidence and prediction intervals for generalized linear accident models	Wood, G.R.			Accidents Analysis and Prevention Vol 37 n°2(March 2005), pp 267-273	E	2005	Confidence interval for accident prediction

7 Annex II

A - The different travel time calculation methods based on station traffic data

The speed method

The most popular method for travel time estimation is known as the speed method: the speed measured at the location of each traffic station is assumed to be the speed experienced by drivers along the stretch allocated to this station. Then, it is easy to calculate the crossing time of the stretch by dividing its length by the speed on the station. For the estimation of the travel time on a given itinerary, one then states that, at a given time t, this travel time is equal to the sum of the elementary travel times at time t on all the sections forming this itinerary. This travel time is known as an instantaneous travel time .

Equations are the following:

$$TT_{Speed}(t) = \frac{L}{2} \times \left[\frac{1}{V_{upstream}(t)} + \frac{1}{V_{downstream}(t)} \right]$$

where $V_{upstream}$ and $V_{downstream}$ denote the traffic speed of the upstream and downstream traffic data station respectively, and L the length of the section.

This method requires a high density of data collection so that the approximation of the section speed by the data station speed is not too inaccurate (in order to limit the spreading effect). Typically, data stations on Ile-de-France motorways and the Boulevard Périphérique around Paris are spaced at 500 m intervals. In addition, due to the instantaneous feature of this travel time, the estimation is always late during changing traffic situations compared to the real travel time experienced by drivers (underestimation during increasing congestion, and overestimation during decreasing congestion). This bias prevents the use of this method on long itineraries along which the traffic situation may change strongly while the driver is proceeding to his destination. On such highly equipped motorway sections, very satisfactory performance has been obtained (inaccuracy of the order of magnitude of a couple of minutes on the Boulevard Périphérique during steady congestion).

The stock method

On interurban motorways, intervals between data stations are generally equal to 20-25 km except on some locations where stations can be spaced at 5 km intervals (e.g. in the Rhône valley on A7 motorway in France). This 5 km interval is still suitable for the use of the speed method in the situations of general congestion as it is the case in the Rhône valley during Summer holiday rushes, but generally speaking such low traffic data measurement densities are inappropriate for the use of the speed method during other situations like an accident happening during medium or low traffic level.

ISIS has designed and developed the so-called stock method which relies on traffic volume measurements instead of traffic speed measurements: as soon as the exit traffic volume differs from the entry traffic volume, congestion can be inferred somewhere on the motorway stretch. An instantaneous travel time can then be calculated by dividing the number of vehicles in the stretch (the stock) by the exit volume: if 500 vehicles are present at time t and the exit traffic volume is 1000 v/h, the last one will experience a travel time equal to (500/1000)*60 = 30 min. The method requests the measurements of all entering and exiting volumes (so-called closed stretches).

Equations are the following:

$$Stock(t) = Stock(t_0) + \Delta \times \sum_{k=t_0}^{k=t} \left[Q_{upstream}(k) - Q_{downstream}(k) \right]$$

$$TT_{Stock}(t) = Stock(t) / Q_{downstream}(t)$$

where $Q_{upstream}$ and $Q_{downstream}$ denote the traffic volume of the upstream and downstream traffic data station respectively, L the length and Δ the elementary time slice (e. g. 6 minutes), Stock(t) and Stock(t_0) the current and initial numbers of vehicles in the section.

This estimation is still instantaneous in the sense that the exit volume is taken at time t whereas it can change in the future (especially in case of an incident, when the carriageway is cleared). To this respect, as for the speed method, it is best fitted to the calculation of travel times on short itineraries or the calculation of queue crossing times. However, the method is based on a true traffic model (even if it is very simple) and it contains a predictive dimension: the stored vehicles will have to exit at some time in the future.

Compared to the speed method, the strength of the stock method is that a good estimation of the travel time can be given very quickly (the delay necessary for the wave of decreasing volume to propagate until the downstream station is short) without having to wait for the congestion to step back on to the upstream station (if even this happens). Furthermore, no undesirable effect as the spreading of congestion along the whole stretch occurs.

Its weakness lies in the potential inaccuracy of the stock estimation: biased traffic volume measurements can make this estimation diverge after a period of several hours. This is the reason why ISIS has developed the combined method. The other limitation of the method lies in the fact that it can only work on closed sections , i. e. sections where all entering and exiting traffic is measured.

The combined method

This method designed by ISIS consists in using the speed method during free-flow traffic, switching to the stock method as soon as a congestion is detected and coming back to the speed method when the congestion disappears. Experience shows that, on most occasions, the possible measurement biases have generally not enough time to significantly hamper the travel time estimation. The detection makes use of several modules: detection of measurement drift, detection of congestion appearance, detection of decreasing congestion, detection of return back to free-flowing. The combined method has been implemented in the TPI tool in the Autoroutes Paris Rhin Rhône Control Centre in Dijon.

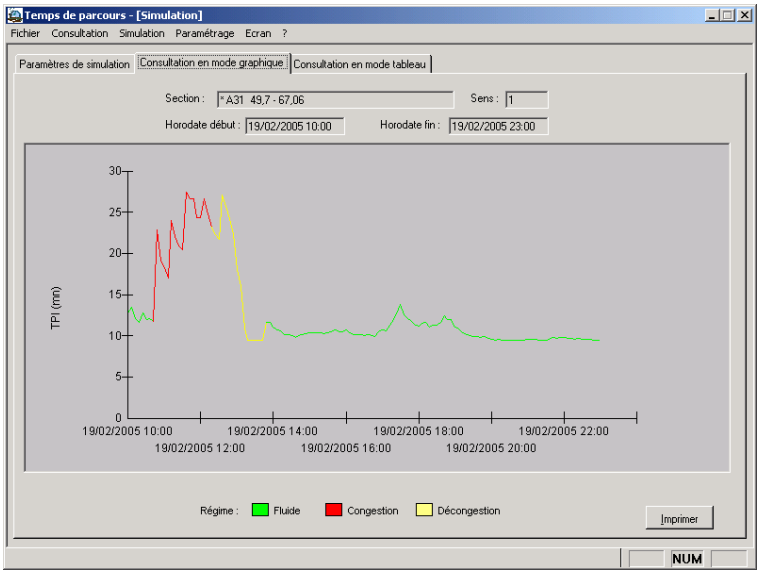


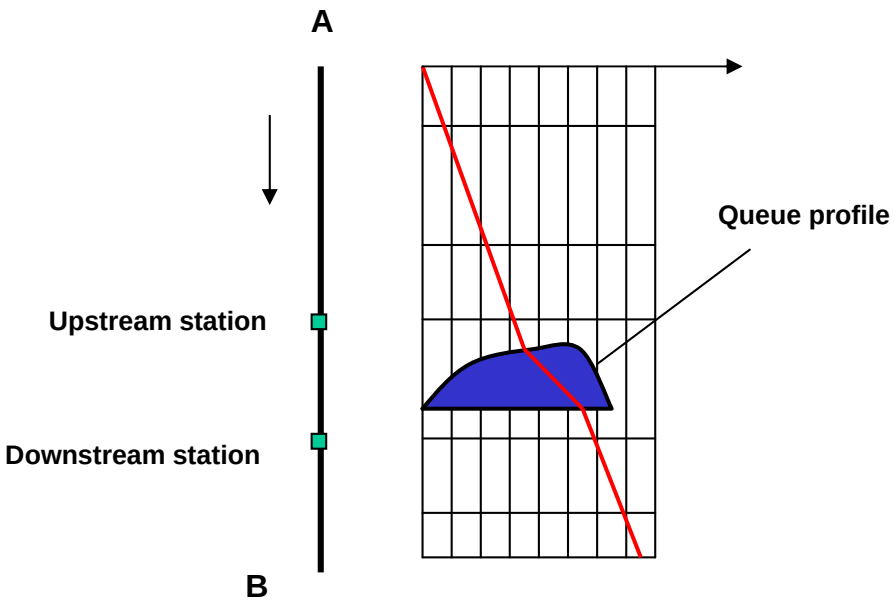
Fig. 7.1 An example of TPI tool display

The improvements made

The improvements of the combined method developed within the frame of INTRO are responses to the 2 limitations underlined above: lack of forecast and closed sections.

B - The introduction of forecast into the stock method

The introduction of forecast is aimed at enhancing the current stock which is penalised by its instantaneous feature. This is illustrated by the figure below:

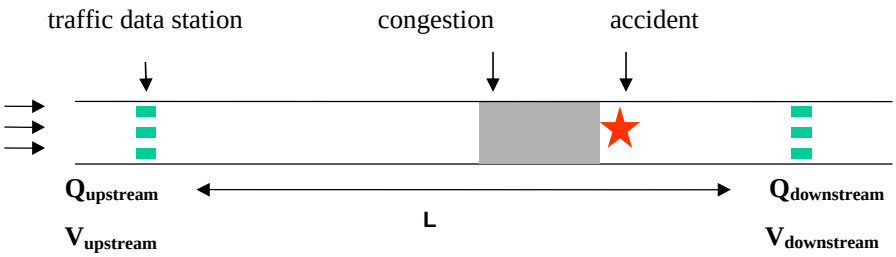


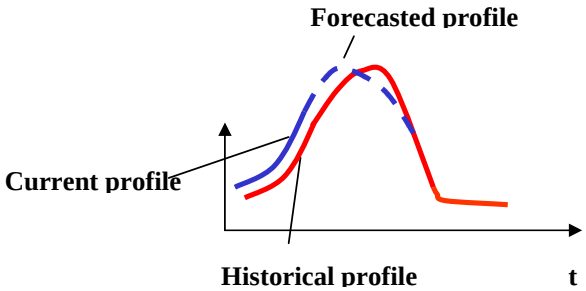
Clearly, the travel time of the vehicle whose trajectory is displayed on the figure (red line) will be better than the instantaneous travel time because it will take account of the real queue length met when it arrives to the congestion site.

Forecast can be introduced both into the stock (the upper part of the ratio) and into the exit volume (the lower part of the ratio).

- ***Forecast of the stock***

Forecasting the stock means that we must forecast the upstream volume and the downstream volume of the section. The forecast of the downstream volume is specific because it is related to the incident characteristics and will be presented below. The problem is then to forecast the traffic volume on any counting station on the same time slice basis as the one used for the travel time calculation, i. e. 6 min.





In the case of APRR application, we have developed a simple method of traffic volume forecast built on the 1 hour-based traffic volume forecast that APRR elaborate 1 year in advance using a calendar-based method.

We first update at any hour the 2 next forecasts (1 hour and 2 hour in advance) according to the last real-time measurements already made.

We then produce each 6 min the 6 min-based traffic volume forecasts on a 2-hour slipping horizon, on the basis of the latter forecast of 1 hour-.based traffic volumes. This allows producing each 6 min the profile of the 20 travel times forecasted from now until 2 hours ahead.

- **1 hour-based forecast on a 2 hour horizon**

For the 1 hour-based traffic volume forecast, several heuristic approaches have been tested from which we have extracted the two presented below:

The transposition method

This simple method consists in adding to the last measured traffic volume the predicted difference between the historical forecasts at time t and time t + 1 hour:

$$\Pr ev(t+1) = Mes(t) + \left[\Pr ev_histo(t+1) - \Pr ev_histo(t)\right]$$

where
Prev(t+1) = traffic volume forecasted for the next hour
Mes(t) = traffic volume measured during the last hour
Prev_histo(t) and *Prev_histo(t+1)* = historical forecasted traffic volumes for the current and next hours.

The pattern method

The idea here is to assume that the general shape of the historical traffic volume is correct and to make use of the first hours of the day to fit this historical pattern in the best way with the first measured traffic volumes. Then, the transposition method is applied on the updated historical traffic volumes instead of the historical volumes themselves:

$$\Pr ev(t+1) = Mes(t) + \left[Dernière_\Pr ev_histo_recalée(t+1) - Dernière_\Pr ev_histo_recalée(t)\right]$$

where
Dernière_Prev_histo_recalée(t) and *Dernière_Prev_histo_recalée(t+1)* = updated historical forecasted traffic volumes for the current and next hours.

Two different ways have been investigated: the slipping pattern method, where the updating of the historical pattern is made each hour, and the frozen pattern method, where the pattern is kept unchanged from the sixth hour after the beginning of the slipping method.

Results

The 3 methods have been tested on traffic data coming from 20 traffic data collection stations placed on the APRR network and several months, totalising 205 days from January to July 2005.

The following table summarises the results obtained during the period 7:00 am-7:00 pm (the period when traffic reaches interesting levels):

Period	Number of days	7:00 am - 7/00 pm Forecast 1h	Transposition 1h	Slipping pattern 1h	Frozen pattern 1h	Forecast 2h	Transposition 2h	Slipping pattern 2h	Frozen pattern 2h
January	31	158	91	118	89	158	116	125	122
Winter-Saturdays	5	369	217	221	213	369	283	283	283
Winter-Sundays	5	224	126	158	129	224	169	177	174
Winter-other days	20	158	90	120	86	158	108	114	111
February-other days	4	132	82	115	81	132	97	108	103
March-other days	25	146	99	134	97	146	121	131	126
Spring-Saturdays	5	431	160	163	158	431	225	228	226
Spring-Sundays	5	338	140	143	138	338	203	210	205
Spring-other days	20	224	105	125	101	224	135	141	139
April-other days	8	237	111	128	107	237	144	150	147
May-other days	23	163	93	117	90	163	114	122	118
June-other days	30	151	96	123	93	151	113	121	119
July-Saturdays	4	184	141	155	136	184	170	168	168
July-Sundays	4	169	125	163	125	169	149	158	156
July-other days	16	190	122	137	117	190	151	156	154
TOTAL	**205**	**186**	**106**	**130**	**103**	**186**	**132**	**139**	**137**

It appears that the simple transposition method proves to be the most effective for 1-hour forecast and equivalent to the others for 2 hour-forecast.

Calculated on the whole data set, the accuracies of the 3 methods can be compared (7:00 am-7:00 pm period):

Absolute Mean Square Error	1-hour forecast	2-hour forecast
Historical forecast	186 v/h	186 v/h
Transposition method	106 v/h	132 v/h
Slipping pattern method	130 v/h	139 v/h
Frozen pattern method	103 v/h	137 v/h

The transposition method has been chosen among the 3 methods tested because of it is generally better accuracy and its simplicity.

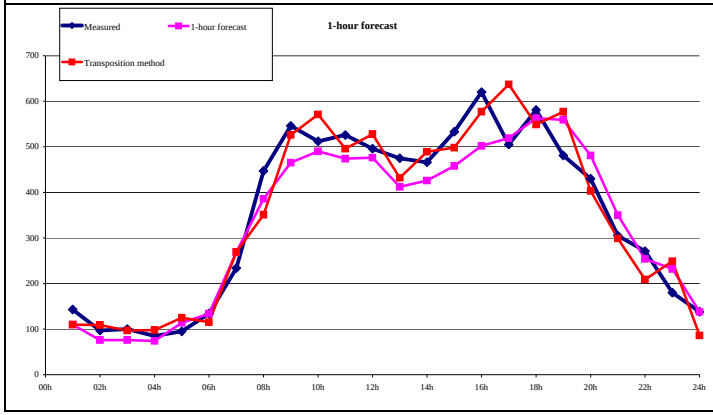
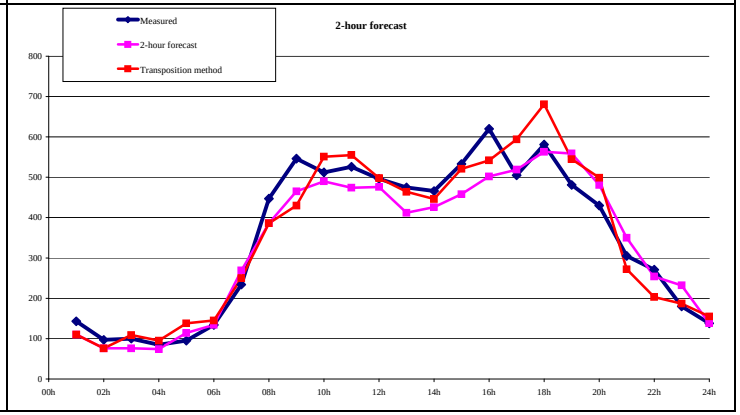
Results restricted to the transposition method are presented in the following table:

Period	1h forecast			2h forecast		
	Historical forecast	Transposition method	Benefit	Historical forecast	Transposition method	Benefit
January (31)	158	91	43%	158	116	26%
Winter-Saturdays (5)	369	217	41%	369	283	23%
Winter-Sundays (5)	224	126	44%	224	169	25%
Winter-other days (20)	158	90	43%	158	108	32%
February-other days (4)	132	82	38%	132	97	27%
March-other days (25)	146	99	32%	146	121	17%
Spring-Saturdays (5)	431	160	63%	431	225	48%
Spring-Sundays(5)	338	140	58%	338	203	40%
Spring-other days (20)	224	105	53%	224	135	39%
April-other days (8)	237	111	53%	237	144	39%
May-other days (23)	163	93	43%	163	114	30%
June-other days (30)	151	96	36%	151	113	25%
July-Saturdays (4)	184	141	24%	184	170	8%
July-Sundays(4)	169	125	26%	169	149	12%
July-other days (16)	190	122	36%	190	151	21%
TOTAL	**186**	**106**	**43%**	**186**	**132**	**29%**

It appears that:

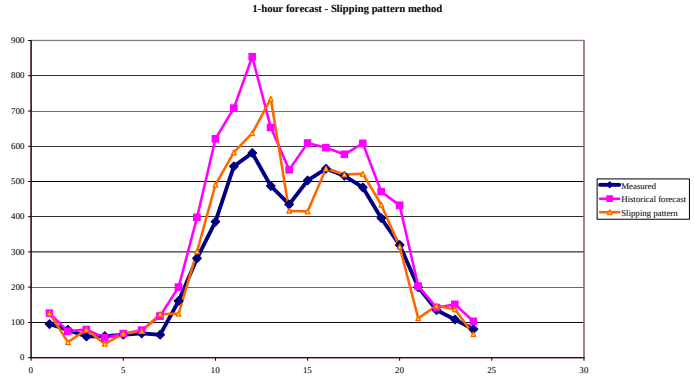
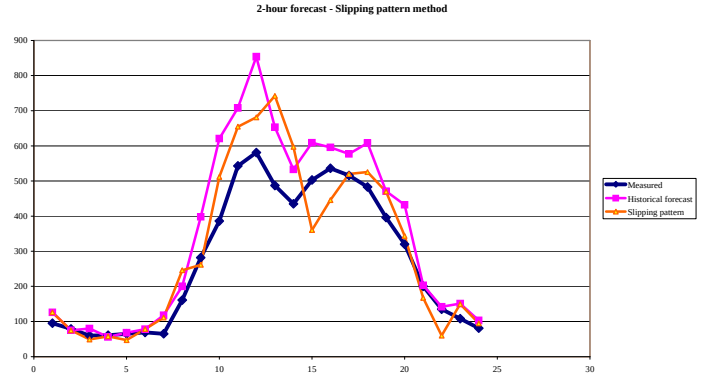
- main inaccuracies are concentrated on Saturdays, the days with higher traffic levels
- the real-time updating significantly enhances the accuracy of prediction compared with the historical method, although this one is already fairly good: 40 % increase for 1h forecast and 30 % for 2h forecast.

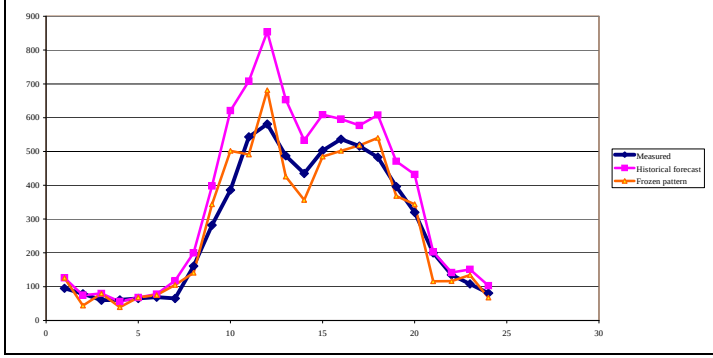
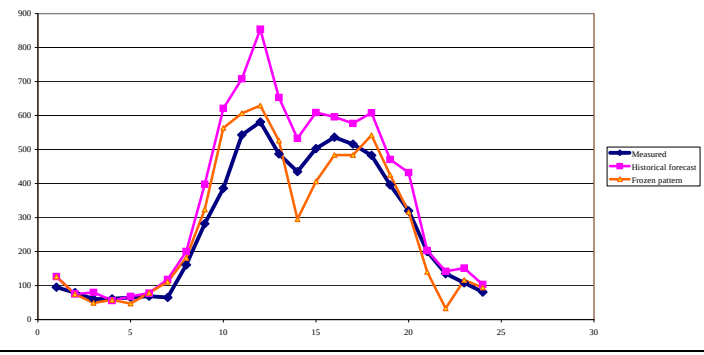
The following figures present some examples of the result of the transposition, slipping pattern and frozen, pattern methods on some days and data stations of the test sample used:

Transposition method A39 motorway 18 May 2005. K. Point 124.764. Southbound	
1-hour forecast	2-hour forecast
	

Absolute Mean Square error on 7:00 am-7:00 am period	**1-hour forecast**	**2-hour forecast**
Historical forecast	62 v/h	62 v/h
Transposition method	63 v/h	60 v/h

The transposition method gives here the same accuracy as the historical forecast.

Slipping pattern method on A39 motorway 23 April 2005. K. Point 23.65. Southbound	
1-hour forecast	2-hour forecast
	

Frozen pattern method on the same data as above	
1-hour forecast	2-hour forecast
	

Absolute Mean Square error on 7:00 am-7:00 am period	1-hour forecast	2-hour forecast
Historical forecast	106 v/h	106 v/h
Slipping pattern method	66 v/h	88 v/h
Frozen pattern method	50 v/h	63 v/h

As can be seen on the figures, the frozen pattern method is slightly better than the slipping pattern method in this case.

- **6 min-based forecast on a 2 hour horizon**

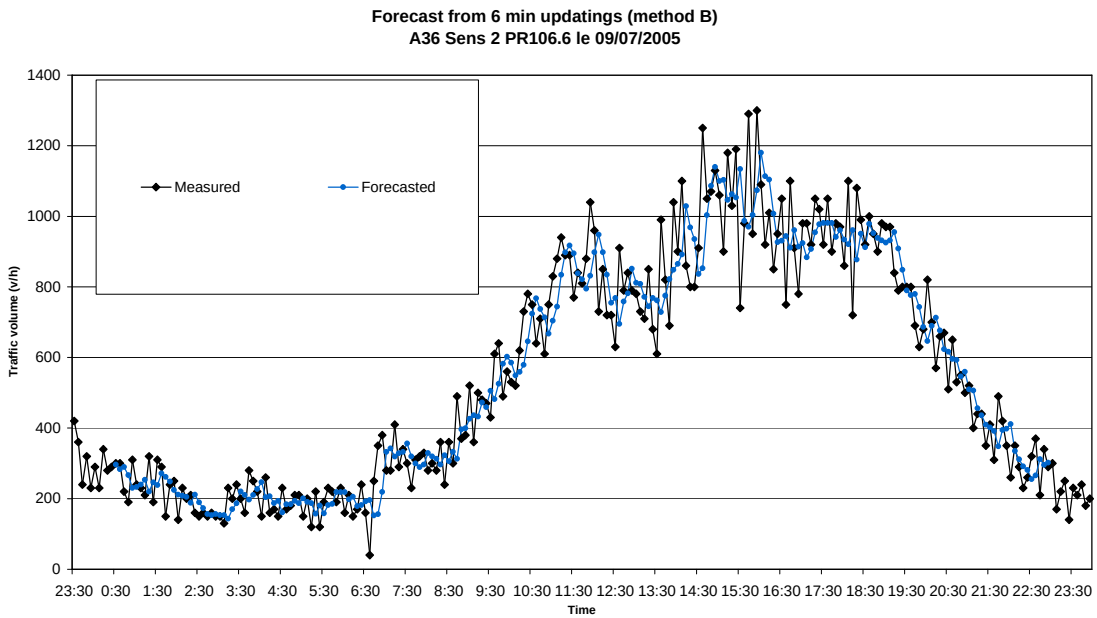
The first step is to transform the 1-hour based forecast traffic volume profile into a 6 min-based traffic volume profile. Several methods can be used: the first one, called constant breaking-up, simply consists to transform the 1-hour traffic volume into constant 6 min traffic volume slices, each one being equal to the this 1-hour traffic volume (when expressed in v/h). The second one, called linear breaking-up, consists in assuming that, at medium time of 1-hour time slice (say 10:30 am, 11:30 am,), the 6 min-based traffic volume (expressed in v/h) is equal to the traffic volume of the 1-hour based, linking all these points by straight lines and dividing each 1-hour slice into 10 6 min-slices accordingly.

Then, 2 different main ways for producing the 6 min-based traffic volume forecast are possible:

- method A : make the updating of the 1-hour historical demand only at changing hours (10:00 am, 11:00 am,) on the 3 next hours, transform it into 6 min profile and keep this 6 min profile until the next changing hour, simply reading the 6 min volumes on this profile along a sliding 2 hours window
- method B : use at any time the current 6 min volume measurement (averaged with the one before in order to smooth random variations) for updating the 6 min profile on the next 2 hours.

Several variants have been also tested.

The best results have been obtained with method B. Below is the example of 9 July 2005 where at any 6 min time t, the 1h forecast made at time t-1h has been plotted.



The comparison of the 2 methods is shown in the table below, where the AMSE of each 6 min. time slice of the 2 hour forecasts along the 7:00 am-7:00 pm period (calculated on 120 values for each of the 20 time slices) is given (in this particular case, method A appears slightly better than method B).

Absolute Mean Square Error
7:00 am-7:00 pm

Rank of the 6 min. time slice	Method A	Method B
00:06	121	119
00:12	121	125
00:18	121	125
00:24	121	130
00:30	121	122
00:36	121	120
00:42	121	125
00:48	121	127
00:54	121	132
01:00	121	138
01:06	138	136
01:12	138	141
01:18	138	139
01:24	138	145
01:30	138	145
01:36	138	143
01:42	138	143
01:48	138	143
01:54	138	142
02:00	138	142
Mean AMSE	**130**	**134**

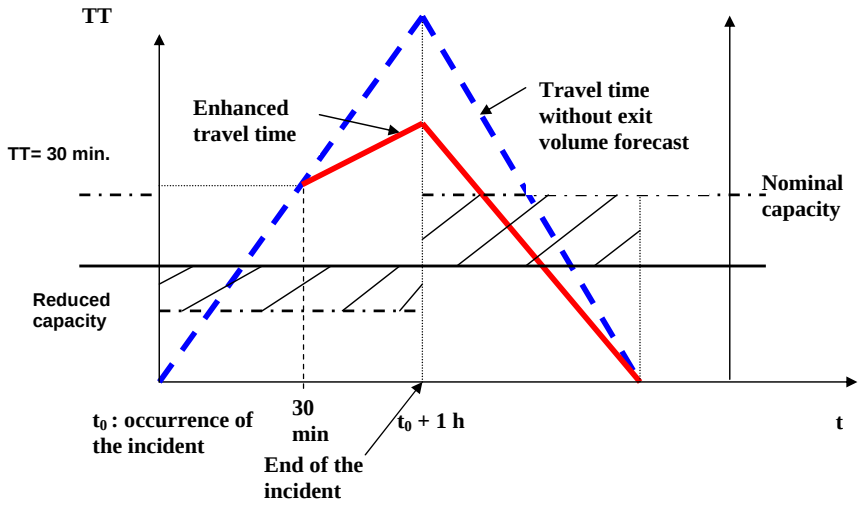
- ***Forecast of the exit volume***

When the cause of the congestion is geometric, for example a decrease of the number of lanes, then the exit volume is known, and constant from the beginning of the congestion until its end. In that case, the instantaneous travel time is also a forecasted travel time and there is no need to do any forecast of the exit volume. When the cause of the congestion is an accident, there are methods to predict the duration of the capacity restriction in relation with the characteristics of the accident: number and type of vehicles involved, presence of injured people, existence of spilled load, etc ISIS has developed for APRR such duration prediction tool, called PREDIRE, based on statistical relationships from 2 years of historical data (more than 1500 accidents in the log).

The principle is the following: knowing that the anticipated time when the carriageway is cleaned is T_1 (T_1 > current time t), then as soon as the travel time $TT_{Stock}(t)$ exceeds the time lasting until T_1, the exit volume must, during the period $TT(t) - T_1$, be taken different from the current exit volume and equal to the nominal capacity *Capa_nom*. The formula is then the following:

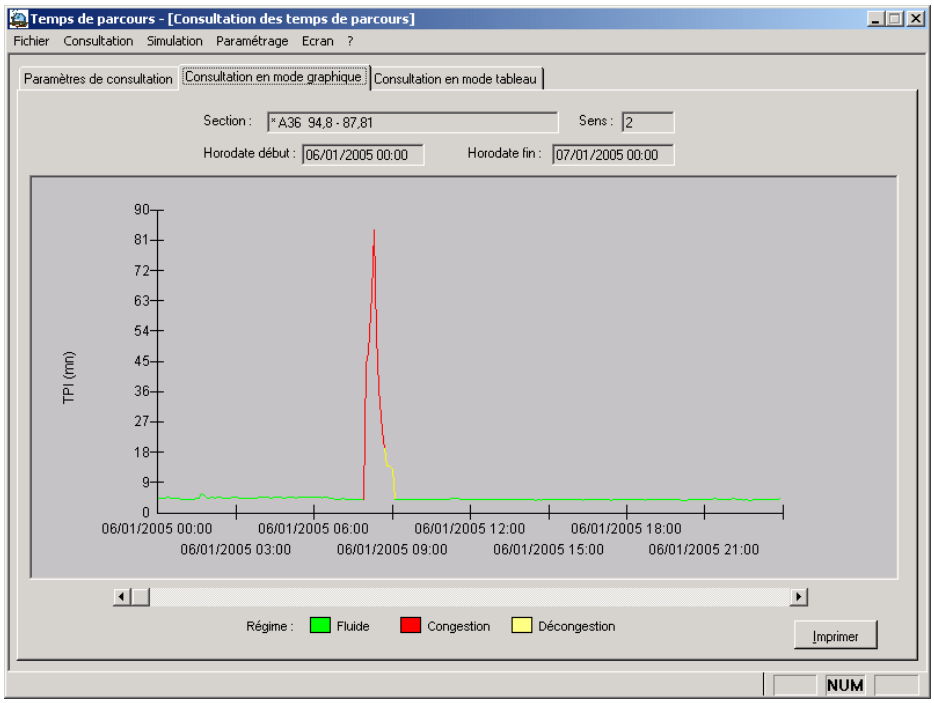
$$TT_{Stock}(t) = (T_1 - t) + [Stock(t) - (T_1 - t) \times Q_{downstream}(t)] / Capa_nom$$

The figure below shows the correction obtained, which leads to a more realistic estimation of the real travel time:



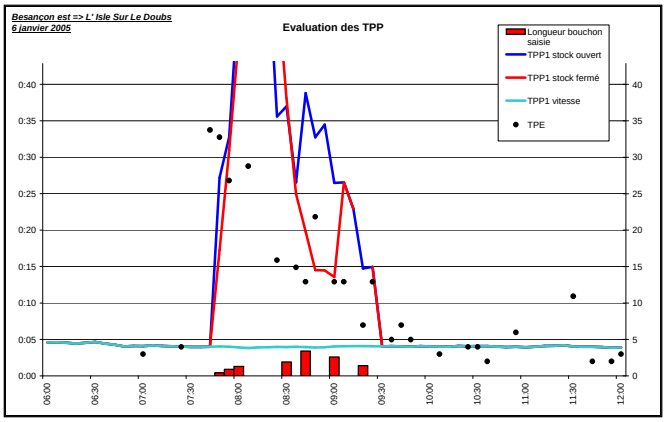
We give below an example of the enhancement produced on a real scenario: on the 6 January 2005, on A36 motorway, an accident occurred at 7:43 am; it was automatically detected by TPI at 8:00 am and ended at 9:25 am. The carriageway was cleaned at

The screen of the current TPI software version was the following:

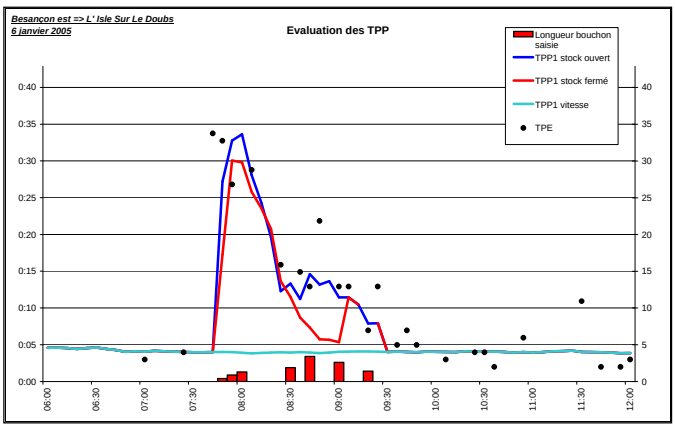


One can notice the sharp shape of TPI, with a somewhat unrealistic maximum instantaneous travel time of more than 80 minutes.

Below are the results of applying the forecast on exit traffic volume (NB: the forecast is not used in the stock calculation, which remains instantaneous):



Without exit volume forecast



With exit volume forecast

Some comments are necessary:

- The red curve is the automatic stock method on this closed section
- The dark blue curve is the open stock method applied to this section assuming that the upstream volume is unknown (in order to simulate an open section)
- The light blue curve is the speed method (no congestion on the traffic stations)
- The red histogram represents the queue lengths which have been logged by the operator at the PC; in this section, this data is used for the open section stock method, but for in the automatic stock method (no stock updating in this example)
- The black dots are the reference travel times obtained from the toll tickets.

One can observe that, after applying the exit volume forecast:

- The maximum travel time climbs down to 30 34 min.
- Curves are close to the reference travel times except at the end of the congestion for the automatic stock method
- The end of the congestion is very well modelled with the open stock method (as in general along the whole period) because of frequent stock updating thanks to the queue length data allowing for the correction of probably underestimated upstream traffic volumes.

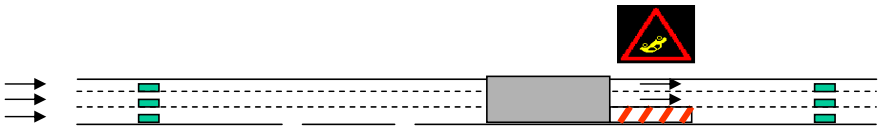
In conclusion of this section, the forecast of the exit traffic volume is able to greatly enhance the TPI method in smoothing the over evaluation produced by its instantaneous feature.

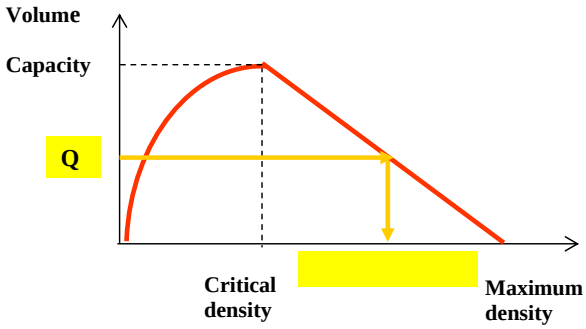
As a general conclusion, it is clear that the introduction of both stock and exit volume forecasts will produce significantly enhanced travel time estimations, alleviating the limits of the instantaneous approach of the current stock method.

C - The extension to open sections

The extension of the stock method to open sections (i.e. sections where some exit or entry volumes are not measured because of a lack of counting stations or their breakdown) is another important enhancement, presented below.

The idea is that the data of queue length observed on the field can be transformed into an estimation of the number of vehicles in the section, then, dividing by the exit volume gives the travel time.
The principle is illustrated in the figure below:





If Q denotes the exit volume, the density in the queue *Queue_density* is given by the fundamental diagram of the section as shown in the figure above.

If L denotes the length of the queue, then the number of the vehicles in the queue is given by the following formula:

$$N = \frac{L}{Queue_density}$$

The crossing time of the queue is given by:

$$TT = \frac{N}{Q}$$

The travel time of the section is then calculated by adding to TT the travel time on the upstream and downstream free parts of the section from their lengths (known from the kilometre points of the head and the tail of the queue) divided by the upstream and downstream speeds respectively.

The figures displayed above about the accident on A36 the 6 January 2005 show that the method is effective and can give far better results than the speed method when the congestion is not seen by the traffic data stations, provided that the queue lengths are measured accurately by the patrol man on the field. This not always the case because in some cases, this one can falsely deduce that the congestion is out: for example, because the traffic flow is free at his level, whereas this is an acceleration wave and an upstream congestion still remains.

This ends the considerations about the stock method and how to make the best use of traffic data and event data. We will see further that toll tickets data provide a very useful additional source to contribute to robust real-time travel time estimations on motorways.