

# **Multi-Sensor Data Fusion for Travel Time Estimation**

**Jiang Han**

A thesis submitted for the degree of Doctor of Philosophy of  
Imperial College London

Centre for Transport Studies  
Department of Civil and Environmental Engineering  
Imperial College London, United Kingdom

**March 2012**

# Abstract

The importance of travel time estimation has increased due to the central role it plays in a number of emerging intelligent transport systems and services including Advanced Traveller Information Systems (ATIS), Urban Traffic Control (UTC), Dynamic Route Guidance (DRG), Active Traffic Management (ATM), and network performance monitoring. Along with the emerging of new sensor technologies, the much greater volumes of near real time data provided by these new sensor systems create opportunities for significant improvement in travel time estimation. Data fusion as a recent technique leads to a promising solution to this problem. This thesis presents the development and testing of new methods of multi-sensor data fusion for the accurate, reliable and robust estimation of travel time.

This thesis reviews the state-of-art data fusion approaches and its application in transport domain, and discusses both of opportunities and challenging of applying data fusion into travel time estimation in a heterogeneous real time data environment. For a particular England highway scenario where ILDs and ANPR data are largely available, a simple but practical fusion method is proposed to estimate the travel time based on a novel relationship between space-mean-speed and time-mean-speed. In developing a general fusion framework which is able to fuse ILDs, GPS and ANPR data, the Kalman filter is identified as the most appropriate fundamental fusion technique upon which to construct the required framework. This is based both on the ability of the Kalman filter to flexibly accommodate well-established traffic flow models which describe the internal physical relation between the observed variables and objective estimates and on its ability to integrate and propagate in a consistent fashion the uncertainty associated with different data sources. Although the standard linear Kalman filter has been used for multi-sensor travel time estimation in the previous research, the novelty of this research is to develop a nonlinear Kalman filter (EKF and UKF) fusion framework which improves the estimation performance over those methods based on the linear Kalman filter. This proposed framework is validated by both of simulation and real-world scenarios, and is demonstrated the effectiveness of estimating travel time by fusing multi-sensor sources.

# Declaration of Contribution

At various stages during this PhD, I have been involved in collaborative efforts with both academic and industrial colleagues. In certain cases, the output of this collaboration is included in this thesis to better explain and support the research presented. In particular, my research has built upon collaborative work with my supervisors and other colleagues, working on several collaborative research papers that were presented at various conferences and submitted for journal publication. These are listed in the reference section and are all my own work.

I hereby declare that besides the collaboration referred to above I have personally carried out the work described in this dissertation.

.....  
Jiang Han

# Acknowledgements

Upon finishing this thesis that assembles the four years of work, I would like to express my deepest gratitude to my primary supervisor Prof. John Polak who guided me into this exciting research. This thesis could not be accomplished without his supportive knowledge, patient guidance, productive suggestions, and constant encouragement throughout the course of this PhD. I am also very grateful for his continuous financial support.

I would give my sincere appreciation to my second supervisor Dr. Javier Barria for this valuable advice around my work. He always kindly helps me out at the first time when I was stuck in a number of technique problems.

I would also thank Dr. Rajesh Krishnan for stimulating discussions and constructive suggestions about my research work, and especially for helping me with the sensor data and encouraging me to use Perl and Linux.

I was extremely grateful to Dr. Scott Le Vine, Dr. Jarlath Molloy and Dr. Shaojun Feng for being warm and understanding officemates. I would also thank Dr. Robin North and Dr. Jun Hu for the kind discussion about my research work. I would thank Fangce Guo and Guotao Hu for their supports, and also for all of the friends in CTS and department. I should express my particular thank to Jackie Sime for her administrative help.

Finally, I wish to express my heartfelt thanks to my parents and Keying Mao for their continuous support and encouragement.

# Contents

|  |           |
|--|-----------|
| <b>Abstract.....</b>   | <b>2</b>  |
| <b>Declaration of Contribution.....</b>  | <b>3</b>  |
| <b>Acknowledgements .....</b>  | <b>4</b>  |
| <b>Contents .....</b>  | <b>5</b>  |
| <b>Lists of Figures .....</b>  | <b>11</b> |
| <b>List of Tables .....</b>  | <b>14</b> |
| <b>Chapter 1 Introduction.....</b>   | <b>16</b> |
| 1.1 Background.....  | 16        |
| 1.1.1 Use of travel time .....   | 16        |
| 1.1.2 Measurement and estimation of travel time .....  | 17        |
| 1.2 Aims and Objectives .....  | 19        |
| 1.3 Thesis Outline .....   | 20        |
| <b>Chapter 2 Opportunities &amp; Challenges of Data Fusion in Travel Time Estimation .....</b> | <b>23</b> |
| 2.1 Heterogeneous Sensor Sources .....   | 23        |
| 2.1.1 Availability of the sensors.....   | 23        |
| 2.1.2 Problems with each sensor source used in travel time estimation.....                     | 25        |
| 2.1.3 Summary .....  | 26        |
| 2.2 Data Fusion .....  | 26        |
| 2.2.1 An overview of data fusion .....   | 26        |
| 2.2.2 Advantages of data fusion .....  | 27        |

|  |           |
|--|-----------|
| 2.2.3 Review of data fusion approaches within transport domain.....                    | 28        |
| 2.3 Previous work on data fusion in travel time estimation.....                        | 29        |
| 2.3.1 Related data fusion projects .....   | 30        |
| 2.3.2 Data fusion approaches .....   | 31        |
| 2.4 Discussion .....   | 32        |
| <b>Chapter 3 A Travel Time Estimation Model by ILDs Data with the Presence of ANPR</b> |           |
| <b>Observations.....</b>   | <b>34</b> |
| 3.1 Introduction.....  | 35        |
| 3.2 Background .....   | 35        |
| 3.2.1 Time-Mean-Speed vs. Space-Mean-Speed .....                                       | 35        |
| 3.2.2 Existing SMS estimation methods .....  | 38        |
| 3.3 Methodology .....  | 40        |
| 3.3.1 A further derivative on Wardrop’s formulation .....                              | 40        |
| 3.3.2 Finding the right solution for SMS.....  | 41        |
| 3.3.3 Approximating $\mathbb{E}[v_i^2]$ by using TMS .....                             | 43        |
| 3.3.4 Fusing ANPR observations into the estimation model.....                          | 45        |
| 3.4 Model Evaluation by England Highway Data .....                                     | 46        |
| 3.4.1 Data description .....   | 46        |
| 3.4.2 Evaluation results & analysis .....  | 48        |
| 3.5 Model Improvement by Introducing Traffic State Factor.....                         | 52        |
| 3.5.1 The effects of varying traffic states.....                                       | 52        |
| 3.5.2 Model refinement by segmenting traffic states .....                              | 52        |
| 3.5.3 Model refinement by clustering traffic states.....                               | 54        |
| 3.6 Conclusions.....   | 58        |
| <b>Chapter 4 Kalman Filter and Its Applications in Travel Time Estimation .....</b>    | <b>60</b> |
| 4.1 Introduction to the Basic Concept of Kalman Filter.....                            | 61        |
| 4.1.1 Historical background .....  | 61        |
| 4.1.2 Conceptual basis of Kalman filter .....  | 61        |
| 4.1.3 Assumptions.....   | 63        |
| 4.2 Linear Kalman Filter.....  | 64        |
| 4.2.1 The process to be estimated .....  | 64        |
| 4.2.2 Recursive solution to Kalman filter.....   | 66        |

|   |            |
|---|------------|
| 4.3 Nonlinear Kalman Filter .....   | 68         |
| 4.3.1 Nonlinear process to be estimated.....  | 68         |
| 4.3.2 Extended Kalman filter (EKF) algorithm.....                                     | 68         |
| 4.3.3 Unscented Kalman filter (UKF) algorithm .....                                   | 69         |
| 4.3.4 Comparison discussion between EKF and UKF .....                                 | 72         |
| 4.4 Review of Kalman Filter in the Traffic Estimation .....                           | 74         |
| 4.4.1 Introduction to the Kalman filter applications in the traffic estimation .....  | 74         |
| 4.4.2 Review of the applications based on single sensor source .....                  | 75         |
| 4.4.3 Review of the applications based on multiple sensor sources .....               | 77         |
| 4.4.3 Review of the applications based on nonlinear Kalman filter .....               | 80         |
| 4.5 Discussion & Summary .....  | 81         |
| <b>Chapter 5 A Fusion Framework Based on Kalman Filter .....</b>                      | <b>85</b>  |
| 5.1 Introduction.....   | 85         |
| 5.2 Macroscopic Traffic Flow Models.....  | 86         |
| 5.2.1 An overview of traffic flow models .....  | 86         |
| 5.2.2 Definition of variables.....  | 88         |
| 5.2.3 Conservation law.....   | 89         |
| 5.2.4 Static velocity-density relationship .....                                      | 91         |
| 5.2.5 Dynamic model .....   | 95         |
| 5.3 Sensor Observation Models .....   | 97         |
| 5.3.1 ILD model .....   | 97         |
| 5.3.2 GPS probe model .....   | 98         |
| 5.3.3 ANPR model .....  | 99         |
| 5.4 Nonlinear Kalman Filter Fusion Framework.....                                     | 100        |
| 5.4.1 Road link segmentation.....   | 100        |
| 5.4.2 Discretisation of macroscopic model .....                                       | 101        |
| 5.4.3 Discretisation of sensor observation model.....                                 | 103        |
| 5.4.4 Construction of Kalman filter fusion framework .....                            | 104        |
| 5.5 Issues and Solutions of Framework Implementation.....                             | 106        |
| 5.5.1 Model parameters.....   | 106        |
| 5.5.2 Biased noise .....  | 107        |
| 5.5.3 Missing observations.....   | 109        |
| 5.6 Summary .....   | 109        |
| <b>Chapter 6 Framework Implementations &amp; Evaluation Based on Simulations.....</b> | <b>111</b> |

|  |            |
|--|------------|
| 6.1 Introduction to the Simulation Experiment.....                                       | 111        |
| 6.1.1 Objectives of the simulation experiment.....                                       | 111        |
| 6.1.2 Simulation platform .....  | 112        |
| 6.1.3 Simulation Scenario .....  | 113        |
| 6.1.4 Observations from the simulation .....   | 114        |
| 6.2 Framework Implementation.....  | 119        |
| 6.2.1 The implementing structure for the simulation scenario .....                       | 119        |
| 6.2.2 The formulation of noise covariance matrices: <b>R</b> & <b>Q</b> .....            | 121        |
| 6.3 Fusion Performance .....   | 125        |
| 6.3.1 Parameters.....  | 125        |
| 6.3.2 Results of General fusion .....  | 126        |
| 6.3.3 Results of GPS and ILDs Fusion by UKF.....   | 129        |
| 6.4 Results of Sensitivity Analysis .....  | 130        |
| 6.4.1 Sensitivity to GPS probe .....   | 130        |
| 6.4.2 Sensitivity to ILD data .....  | 134        |
| 6.5 Performance Comparison with Other Methods .....                                      | 136        |
| 6.5.1 Comparison with neural networks.....   | 136        |
| 6.5.2 Comparison with linear Kalman filter based method .....                            | 141        |
| 6.6 Impact of Time-Varying Sensor Errors .....   | 143        |
| 6.6.1 The limitation of constant error structures .....                                  | 143        |
| 6.6.2 Theoretical impact.....  | 144        |
| 6.6.3 Impact analysis regarding the time-varying ILD error .....                         | 144        |
| 6.7 Summary .....  | 147        |
| <b>Chapter 7 Framework Implementation &amp; Evaluation Based on Real-World Data.....</b> | <b>149</b> |
| 7.1 The Experiment of Data Collection from Real-World .....                              | 149        |
| 7.1.1 Objectives of the experiment.....  | 149        |
| 7.1.2 Road link .....  | 150        |
| 7.1.3 Traffic volume surveys .....   | 151        |
| 7.2 ANPR Data .....  | 152        |
| 7.2.1 Errors in matched ANPR travel time observations .....                              | 152        |
| 7.2.2 ANPR data & its cleaning treatment .....   | 153        |
| 7.2.3 Comparative analysis by using different cleaning treatments .....                  | 155        |
| 7.3 GPS Data.....  | 163        |
| 7.3.1 Collection method .....  | 163        |

|  |            |
|--|------------|
| 7.3.2 Description of recorded GPS data .....   | 164        |
| 7.3.3 Error structure of the GPS data .....  | 165        |
| 7.3.4 Approximation & evaluation.....  | 167        |
| <b>7.4 Framework Implementation.....</b>   | <b>172</b> |
| 7.4.1 Abstract scenario.....   | 172        |
| 7.4.2 The implementing structure .....   | 172        |
| <b>7.5 Fusion Performance .....</b>  | <b>174</b> |
| 7.5.1 Parameters.....  | 174        |
| 7.5.2 Fusion results .....   | 175        |
| 7.5.3 The result of GPS sensitivity analysis with respect to the T-resolution .....      | 178        |
| 7.5.4 The result of comparison with linear Kalman filter .....                           | 179        |
| 7.5.5 Discussion on the minimum realistic levels of sensors configuration.....           | 181        |
| <b>7.6 Summary .....</b>   | <b>181</b> |
| <br>   |            |
| <b>Chapter 8 Conclusions &amp; Future Research Topics.....</b>                           | <b>183</b> |
| 8.1 Conclusions.....   | 183        |
| 8.2 Recommendations for further research.....  | 188        |
| <br>   |            |
| <b>Appendix A A New Method for Probabilistic Traffic State Identification Using ILDs</b> | <b>191</b> |
| A.1 Introduction.....  | 192        |
| A.2 Background .....   | 193        |
| A.3 Methodology .....  | 193        |
| A.3.1 Problem description.....   | 193        |
| A.3.2 Problem modelling .....  | 194        |
| A.3.3 Expectation-Maximisation (EM) algorithm .....                                      | 196        |
| A.3.4 Applying EM algorithm to the model .....   | 196        |
| A.3.5 Error handling .....   | 198        |
| A.4 Evaluation & Empirical Results.....  | 199        |
| A.5 Conclusion .....   | 201        |
| <br>   |            |
| <b>Appendix B A Complete Derivation of Wardrop (1952)'s Formula .....</b>                | <b>202</b> |
| B.1 Distribution of Speed in Time.....   | 202        |
| B.2 Distribution of Speed in Space.....  | 202        |
| B.3 Mean Speed.....  | 203        |

|   |            |
|---|------------|
| B.4 Derivative of the Formula .....   | 203        |
| <b>Appendix C More Evaluation Results for the Model proposed in Chapter 3 .....</b> | <b>205</b> |
| <b>Appendix D Hardware Specification of the GPS Logger .....</b>                    | <b>215</b> |
| <b>References .....</b>   | <b>218</b> |

## Lists of Figures

|   |    |
|---|----|
| Figure 3.1 Difference between TMS and SMS.....                                      | 38 |
| Figure 3.2 Relationship between TMS and $\mathbb{E}[v_i^2]$ .....                   | 44 |
| Figure 3.3 Flowchart of proposed fusion process .....                               | 45 |
| Figure 3.4 Location and topology of chosen links.....                               | 47 |
| Figure 3.5 Scatter plot between observed TMS and SMS from real-world data.....      | 48 |
| Figure 3.6 Result of polynomial fitting $\mathbb{E}[v_i^2]$ by TMS for Link-1 ..... | 49 |
| Figure 3.7 Scatter-plot of the estimation result for Link-1 .....                   | 49 |
| Figure 3.8 Comparison between errors from only using TMS and estimated SMS .....    | 50 |
| Figure 3.9 Scatter plot between observed TMS and SMS .....                          | 51 |
| Figure 3.10 $\mathbb{E}[v_i^2]$ vs. TMS under different traffic states .....        | 53 |
| Figure 3.11 Flow Chart of the segmentation based refined model .....                | 54 |
| Figure 3.12 Relationship between flow and occupancy .....                           | 55 |
| Figure 3.13 Flow Chart of the clustering based refined model.....                   | 56 |
| Figure 3.14 clustering result on scatter plot of flow vs. occupancy .....           | 57 |
| Figure 3.15 Fitting result on each clustered state .....                            | 57 |
| Figure 4.1 Illustration of dynamic system estimation.....                           | 62 |
| Figure 4.2 Summary of recursive solution to Kalman filter .....                     | 67 |
| Figure 4.3 Sketch of Kalman filter operation .....                                  | 67 |
| Figure 4.4 Comparison of mean and covariance propagation between UKF and EKF.....   | 73 |
| Figure 5.1 Traffic flow density vs. volume.....                                     | 90 |
| Figure 5.2 Speed-density-flow diagram.....  | 93 |
| Figure 5.3 Comparison among different velocity-density models.....                  | 94 |

|   |     |
|---|-----|
| Figure 5.4 Space-discrete of road link .....  | 100 |
| Figure 5.5 Fusion framework based on Kalman filter .....  | 106 |
| Figure 6.1 Simulation scenario .....  | 113 |
| Figure 6.2 Flow pattern from the simulation period .....  | 114 |
| Figure 6.3 True travel time vs. GPS probe observations .....  | 116 |
| Figure 6.4 Scatter plot of true travel time vs. GPS probes observations .....   | 116 |
| Figure 6.5 True travel time vs. ANPR observations .....   | 117 |
| Figure 6.6 Scatter plot of true travel time vs. ANPR observations.....  | 118 |
| Figure 6.7 EKF and UKF fusion results .....   | 127 |
| Figure 6.8 Comparison between UKF fusion and single sensor observations .....   | 128 |
| Figure 6.9 ILDs and GPS UKF fusion results .....  | 130 |
| Figure 6.10 Result of sensitivity analysis for GPS probes.....  | 132 |
| Figure 6.11 UKF fusion vs. GPS observations .....   | 134 |
| Figure 6.12 ILDs configurations for sensitivity analysis.....   | 135 |
| Figure 6.0.13 Result of sensitivity analysis for ILDs .....   | 135 |
| Figure 6.14 General Structure of (a) feed-forward networks and (b) recurrent networks .....                                 | 137 |
| Figure 6.15 Traffic flow pattern of estimation day vs. an example of training day .....                                     | 138 |
| Figure 6.0.16 Scatter plot of UKF fusion under C1 vs. neural networks fusion under T1 ...                                   | 139 |
| Figure 6.0.17 Scatter plot of UKF fusion under C1 vs. neural networks fusion under T2....                                   | 139 |
| Figure 6.18 Scatter plot of UKF fusion under the C2 calibration vs. neural networks fusion under the T1 training data ..... | 139 |
| Figure 6.19 Comparison results between LKF and UKF .....  | 142 |
| Figure 6.20 Flow error variance.....  | 145 |
| Figure 7.1 The Maidstone Road Link .....  | 150 |
| Figure 7.2 parallel pneumatic rubber tubes for automatic traffic volume surveys .....                                       | 151 |
| Figure 7.3 Result of cleaning treatment on ANPR travel time observation .....   | 155 |
| Figure 7.4 Cleaning results .....   | 159 |
| Figure 7.5 Cleaning results: F1M vs. F1L .....  | 160 |
| Figure 7.6 Overall comparison among different cleaning treatments.....  | 160 |
| Figure 7.7 An example of the detailed difference between Percentile and Kent treatment...                                   | 162 |
| Figure 7.8 The format of RMC output.....  | 164 |
| Figure 7.9 A part of sampled GPS log data .....   | 164 |
| Figure 7.10 Illustration of the GPS data collection.....  | 165 |
| Figure 7.11 Illustration of the effects of GPS errors regarding link length .....   | 171 |

|  |     |
|--|-----|
| Figure 7.12 The schematic representation of the experimental road link ..... | 172 |
| Figure 7.13 Fusion results from real-world data.....                         | 178 |
| Figure 7.14 Result of sensitivity analysis about the T-resolution.....       | 179 |
| Figure 7.15 Comparison results between LKF and UKF .....                     | 180 |

## List of Tables

|  |     |
|--|-----|
| Table 3.1 Estimate performance comparison.....   | 50  |
| Table 3.2 Estimate performance comparison between original model and refined model.....                        | 58  |
| Table 4.1 Summarised algorithm for EKF.....  | 69  |
| Table 4.2 Summarised algorithm for EKF.....  | 71  |
| Table 4.3 Summary of difference between EKF and UKF.....   | 73  |
| Table 4.4 Summary of the applications based on Kalman filter.....  | 84  |
| Table 6.1 Origin-Destination (OD) traffic demand for the simulation scenario.....                              | 114 |
| Table 6.2 Overtaking rule settings used in the simulation data.....  | 118 |
| Table 6.3 Performance comparison of ANPR observation.....  | 119 |
| Table 6.4 The parameters used for framework implementation.....  | 125 |
| Table 6.5 Bounds of the states and parameters for the UKF.....   | 125 |
| Table 6.6 The performance of general fusion.....   | 129 |
| Table 6.7 The performance of ILDs and GPS fusion.....  | 130 |
| Table 6.8 The settings for the parameters of GPS probes.....   | 131 |
| Table 6.9 Quantitative results of sensitivity analysis for GPS probes.....                                     | 133 |
| Table 6.10 Quantitative results of sensitivity analysis for ILDs.....  | 135 |
| Table 6.11 The performance comparison between UKF and neural networks.....                                     | 140 |
| Table 6.12 The performance comparison between UKF and LKF.....   | 141 |
| Table 6.13 The performance comparison between constant error and time-varying error....                        | 146 |
| Table 7.1 Quantitative Comparison results between Kent treatments and others.....                              | 161 |
| Table 7.2 The comparison between approximation and equal length evaluation regarding the GPS data quality..... | 170 |
| Table 7.3 The parameters used for framework implementation.....  | 175 |
| Table 7.4 Bounds of the states and parameters for the UKF.....   | 175 |

|   |     |
|---|-----|
| Table 7.5 The performance of ILDs and GPS fusion from real-world data .....         | 178 |
| Table 7.6 Quantitative results of sensitivity analysis about the T-resolution ..... | 179 |
| Table 7.7 The performance comparison between UKF and LKF based on real-world data . | 181 |

# Chapter 1

## Introduction

### 1.1 Background

#### 1.1.1 Use of travel time

The travel time experienced by users in a road network is one of the key elements in the traffic engineering, and is the fundamental variable to provide traffic information. Travel time is defined as the time necessary to traverse a route between any two points of interest, and it is of interest to both the road users and the authorities of transport management and operation. As a simple concept, travel time is a widely used indicator of the level of transport service to traffic engineers and planners, commuters, consumers, and businesses.

Palen et al. (1997) shows that travel time is the most useful information from the user point of view since it can enable people to decide the best time for starting a trip and the best routing option, or to modify the planning during the journey. The information about the travel time is also relevant to the people's daily life. Most of the travellers are concerned by the significant cost on the travel time (Noland & Small 1995, Noland & Polak 2002) and hence travellers will aim to reduce their overall cost by reducing the travel time of their journey. A report by the Federal Highways Administration in the U.S. Department of Transportation states that "*Commuters use a 'travel time budget,' theorized to be between 20 and 30 minutes per one-way commute, to locate their housing relative to work locations. The media report travel times and delays on urban freeways and streets with language like '... an accident on the northbound lanes of the Beltway has traffic delayed 10 to 15 minutes ...'*" (Turner et al.

1998). As a result, travellers are particularly interest in receiving reliable estimate about the travel time.

Driven by the desire of reducing travel time from road users, the transport authorities have been investing a colossal amount of effort and money to improve the network service in terms of the amount of congestion. Travel time is a typical used index to quantify the level of congestion. The UK Highways Agency uses the difference between the travel time under the free-flow conditions and the actual travel times to calculate the congestion indicator (Highways Agency 2011a). Transport for London (TfL) also uses a similar definition of congestion which is based on the knowledge of travel time information (TfL 2003a). TfL also uses travel time as one of the Network Performance Indices (NPIs) to indicate the performance of their networks and to evaluate the success of their traffic management and operation (TfL 2010). Due to the usefulness of the travel time as a metrics in quantifying road network performance, the public transport authorities are always keen to the models that can estimate the travel time as accurate as possible.

As the recent development of Intelligent Transport Systems (ITS), accurately estimating travel time with a certain device or method becomes a critical component of the system. The most relevant ITS application is Advanced Traveller Information Systems (ATIS) whose main service is to publish the current travel time information for the travellers. Dynamic Route Guidance (DRG) system as another ITS application uses current and predicted travel times to evaluate travel time costs on multiple routes, and recommend optimal routes to users. In the field of ITS in traffic management and operation, Urban Traffic Control (UTC) and Active Traffic Management (ATM) use the travel time information as one of the key inputs into their systems to adaptively optimise the network performance. All of these emerging ITS services and businesses require underlying travel time estimates which are sufficiently accurate and reliable to satisfy the requirement of applications.

### **1.1.2 Measurement and estimation of travel time**

The importance of travel time has been identified in the previous subsection. In order to obtain the travel time information, a number of technologies and methods have been developed since 1920s. Generally, the ways of obtaining travel time can be categorised into two groups: direct measurement and indirect estimation. The most conventional direct measurement method is the Moving Car Observer (MCO) (Taylor et al. 2000). This method involves driving a test vehicle along a route, and a passenger manually records the time when

a vehicle passes certain points on the road. The travel time between those points can then be determined. The main problem of this method is that the measured travel time is only from the test vehicles on the road, and it may not be sufficiently representative of the main traffic stream. This effect is more pronounced in urban areas due to the factors such as traffic lights, parked vehicles, and turning vehicles which delay some of the vehicles. Those factors lead to the short-term variability of travel times on a given route (Turner et al. 1998). Therefore, the MCO method suffers from the limited number of vehicles. Although increasing the number of test vehicles to have multiple measurements can improve this issue, the expensive cost of implementation makes it less attractive in practice. The recent development of Global Positioning System (GPS) technology allows MCO survey to operate in an automatic fashion. The test vehicle equipped with a GPS device can measure and record the travel time information based on the positioning data from the GPS. Zhao (1997) shows that it is very accurate to measure the travel time of the MCO vehicles by GPS.

Another type of GPS based travel measurement method is the probe vehicle (PV) technique. As the prevalence of GPS, a quite large number of the vehicles are equipped with GPS tracking devices for the fleet management business (Trakm8 2011). These fleet vehicles transmit their locations information on a periodic basis to a control centre. Travel time on the road can be obtained using the location information from a number of vehicles. Unlike the MCO method which is designed to measure the travel time only for a target route over a specific time period, the travel time measurement from probe vehicles provide much richer spatial and temporal information. The recent development of using mobile data to provide the vehicle location information enriches the availability of the data from probe vehicle (Alger et al. 2003). Similar as the GPS based measurement, mobile data from the cellular phone networks can also be used to locate the vehicles on the road which is able to be translated into travel time measurement. For example, ITIS Holdings plc consolidates GPS and mobile data from a number of different fleets, and these data are widely used for the purpose of obtaining a wide range of travel time information.

Automatic Number Plate Recognition (ANPR) system also measures the travel time directly. It automatically recognises the registration number of vehicles from video cameras by the image processing technique. By matching the registration numbers read about same vehicle passing by two camera sites, the travel time between those camera sites can be measured. Typically the recognition rate of ANPR is between 50% and 90% (van der Zijpp 1997), and Wiggins (1999) shows a recognition rate of 86% during a 2 hour survey in the UK.

Comparing to the MCO and PV methods, the travel time measurement from ANPR systems is more statistically confident due to the much larger sampled size. However, due to the high cost of system installation, the coverage of ANPR camera systems is very limited. The main use of this system is for the congestion charging (TfL 2003b)

Inductive Loop Detectors (ILDs) as the most widely deployed traffic detectors have been in use since 1960s (Klein et al. 2006). ILDs are typically placed beneath the surface of road, and are able to count the number of vehicles passing over them. As a fixed point based measurement, ILDs only output the traffic flow and occupancy data basically rather than the travel time over a stretch of road link. The desired travel time thus has to be estimated from those flow and occupancy measurement indirectly. During the past decades, a number of different methodologies have been proposed to estimate the travel time based on the output of ILDs such as regression modelling (Gault & Taylor 1981; Zhang & He 1998), correlation techniques (Dailey 1993; Petty et al. 1998), fuzzy-logic (Palacharla & Nelson 1999), neural networks (Anderson & Bell 1997; Cherrett et al. 2001; van Lint et. al 2002), k-nearest neighbour method (Robinson & Polak 2005), platoon identification (Lucas et al. 2004; Krishnan 2008a), Kalman filter (Dailey 1999; Wang & Papageorgiou 2005) and Bayesian inference (Sun et al. 2006). These developed methods show their effectiveness under different applications. However, the nature of fixed point measurement from ILDs largely limited the estimation performance.

## **1.2 Aims and Objectives**

Based on the above background introduction, the most of current travel time estimation methods are based on certain traffic measurement technologies, e.g. ILDs, ANPR camera and GPS positioning. In the past decades, both industry and academic researchers developed various methods to improve the estimation performance. Most of their works only focus on using single measurement source. Although some of them show relative accuracy and effectiveness, the inherent drawback of every measurement technology is still difficult to overcome. Currently, the research trend in this area is switching to make full use of all available traffic measurements to estimate travel time, which means applying the concept of data fusion to integrate the output of multiple traffic sensors. The aim of the proposed research is to use multi-sensor data fusion technique to integrate the observations from inductive loop detector, ANPR camera and GPS to perform better travel time estimation in both freeway and interurban environment.

The objectives of this thesis are summarised as follows:

- Identify the opportunities and challenges in travel time estimation based on multiple sensor sources, and determine the most appropriate fusion methodology based on the review of the state-of-the-art fusion methods
- Develop a new method for travel time estimation based on the ILDs and ANPR data which has less limitations and can be easily implemented in practice
- Review and address the gaps and issues of the selected fusion technique as the estimation methodology in travel time estimation
- Develop a novel data fusion framework for travel time estimation based on the selected fusion technique
- Implement and evaluate the proposed fusion framework by simulation
- Implement and evaluate the proposed fusion framework by real-world data

### **1.3 Thesis Outline**

To provide an overview and roadmap of this thesis, a summary of the main content of each chapter is given below.

*Chapter 1* introduces the background information about the importance of travel time and an overview of the travel time estimation. The aims and objectives of this research are also provided in this chapter.

*Chapter 2* introduces the availability of multiple sensor sources and discusses the drawbacks of each type of sensor data. The background information about the data fusion is provided to demonstrate the opportunities of data fusion in the problem of travel time estimation. An overview of the state-of-art fusion techniques is classified based on the application within the transport domain. This chapter also reviews the previous works related to this subject. Based on the review, the last section discusses the advantages of the Kalman filter as the fusion methodology in the field of travel time estimation.

*Chapter 3* develops a new model to estimate travel time (or space-mean-speed equivalently) based on the ILDs and ANPR data. The model is based on a further exploration of Wardrop (1952)'s formula which describes the relationship between space-mean-speed and time-mean-speed (output from ILDs). A novel approximate relationship between time-

mean-speed obtained from ILDs and space-mean-speed is proposed and used in the estimation model. In addition, a property of traffic state dependency is illustrated, and a more refined model is presented where the traffic states are segmented according to the flow and occupancy values obtained from the ILDs. The formulation of this relationship is based on the correlation between ANPR and ILDs observations. This model can be implemented in highway links which have plenty of ILDs data but very few ANPR paired observations. The model is evaluated by the real traffic data from England highways.

**Chapter 4** focuses on the more general fusion problem described in Chapter 2. Based on the reviews in Chapter 2, the Kalman filter technique is chosen as the fundamental fusion methodology. An introduction on the basic concept of Kalman filter and its advantages is presented on the first part of this chapter. The second part of the chapter describes both the linear Kalman filter and the more advanced nonlinear Kalman filter (Extended Kalman filter & Unscented Kalman filter), and a comparison discussion among them is also provided. The last part comprehensively reviews the most relevant works in the field of traffic estimation based on the Kalman filters.

**Chapter 5** developed a general fusion framework based on the nonlinear Kalman filter which is able to integrate the data from ILDs, GPS and ANPR. The main body of this chapter begins with an introduction on traffic flow models which are the basis of the Kaman filter implementation. The second part is to model the three available sensor sources, ILDs, GPS and ANPR, and illustrate how the data from these sensor sources is fed into the traffic flow models. The third part introduces the proposed fusion framework which uses a discretised space-time model to integrate both of the traffic flow models and observations from multiple sensors.

**Chapter 6** uses simulated based experiments to illustrate the framework implementations. The evaluation results based on the simulation scenario are comprehensively presented, including a general fusion based on ILDs, GPS and ANPR data, a more practical fusion based on only ILDs and GPS data, sensitivity analysis for the factors of GPS probes and the spatial availability of ILDs, and comparison analysis with neural networks and linear type Kalman filter.

**Chapter 7** presents the framework implementation and evaluation based on real-world traffic data. A designed experiment which aims to collect the real traffic data required for the evaluation is described. The framework performance is illustrated by a fusion between ILDs

and GPS data, since the observations ANPR is used as the comparison dataset. The sensitivity analysis with respect to the temporal resolution of GPS data and comparison analysis with linear Kalman filter is also provided.

*Chapter 8* concludes the research undertaken in this thesis. It also presents a number of topics for future work based on the ideas developed in this research.

*Appendix A* presents the work in the area of traffic state identification. The reason of presenting this part of work in the appendix is because its focuses on a different topic from the mainstream of this thesis (travel time estimation). It presents a novel new approach to traffic state estimation based on analysis of loop occupancy and flow data alone. The method in based on the assumption at the relationship between flow and occupancy displays distinct regimes according to whether the system is congested or uncongested. A probabilistic classifier is developed, based on the Expectation-Maximisation (EM) algorithm. This new algorithm is evaluated using data from the ILDs on highways and urban links. The results demonstrate that the proposed algorithm, which does not need site specific calibration, is able to identify the traffic state on both urban and highway satisfactorily. This traffic state estimation method is also relevant to the travel time estimation model proposed in Chapter 3 which is demonstrated to have a property of traffic state dependency.

## **Chapter 2**

# **Opportunities & Challenges of Data Fusion in Travel Time Estimation**

The current applications of traffic sensor technologies offer a wide spectrum of available data and heterogeneous sources of information. It provides the opportunities to make use of these sensor sources for the purpose of travel time estimation, which falls into a particular data fusion problem. This chapter firstly introduces the availability of those sensor sources and discusses the drawbacks of each type of sensor data. The second part of this chapter presents the background information about the data fusion and discusses the advantages of the data fusion. A classification about the common fusion techniques within the transport domain is described to provide an overview of the state-of-art. The third part reviews the previous works related to this subject. Based on the review, the last section discusses the advantages of the Kalman filter as the fusion methodology in the field of travel time estimation.

## **2.1 Heterogeneous Sensor Sources**

### **2.1.1 Availability of the sensors**

As the development of sensor technologies, the observations and measurements from all types of traffic sensors provide various real time data. Therefore, a wide spectrum of available data and heterogeneous sources of information are of potential use for estimating,

predicting and assessing traffic performance. According to the different natures of spatial observation, the sensor sources are classified into three groups:

- **Single point sensor:** is the most basic type of traffic sensor including: ILDs, magnetic detectors, fibre optic detectors, infrared detectors, microwave detectors, ultrasonic detectors and laser scanners (Klein et al 2006). Although different sensing technologies are used, all of these traffic sensors are based on a fixed location observation, and measure the temporal traffic characteristics. The research adopts the most commonly used single point sensor: ILDs as one of the fusion input. Robinson (2005) and Krishnan (2008a) provides a comprehensive introduction and discussion on the nature and properties of ILDs. Typically, the outputs from ILDs have flow and occupancy observations. For the double ILDs, the time-mean-speed can also be measured as another type of output (Mortimer 2002).
- **Paired monitoring:** this type of traffic sensing methods is section based measurement such as electronic toll collection (ETC), ANPR cameras and other forms of image processing technologies which normally belong to the class of automatic vehicle identification (AVI) (Klein 2001). In contrast to the single point sensor, the paired sensing method is able to monitor traffic along a section of road and obtain the travel time measurement directly. However, due to the high cost on the equipment installation and operation, the coverage of this type of sensor sources is limited to a relatively small proportion of road links. This research considers the ANPR sensor source from the type of paired monitoring. As introduced in Section 1.1.2, the output from ANPR camera system is direct travel time measurement, and Section 5.3.3 described the detailed modelling approach.
- **Probe vehicles:** according to this technique, vehicles on the road shift from a passive attitude to an active one and act as moving sensors, continuously recording or transmitting information about traffic conditions (Turner & Holdener, 1995; Turner et al 1998). In practice, some vehicle fleets are initiatively operated to drive along the road link and keep record of the travel time information, this is also called floating vehicle/car (Taylor et al, 2000). The sensor technologies used for data collection are normally base on GPS (Ochieng & Sauer 2002) or cellular phone tracking (Youngbin & Cayford 2000). The type of GPS based measurement is considered in this research, and the detailed sensor modelling is provided in Section 5.3.2.

### 2.1.2 Problems with each sensor source used in travel time estimation

Using single type of traffic sensor data to estimate travel time has been existed in both of academia and industry for decades. The research in Robinson (2005) and Krishnan (2008a) comprehensively reviewed a variety of travel time estimation techniques based on single sensor source. Considering the popularity and availability of sensor data, this research focuses on using ILDs, ANPR and GPS from each type of sensor sources to perform data fusion based travel time estimation. The problems existing in each of above sensor technologies are described as follows:

- **ILDs:** as the most common traffic sensor technology among those fixed location sensors, has been widely used since the 1960's (Klein et al 2006). The output of ILDs normally contains the parameters of flow, occupancy, and instantaneous speed at a given point (spot speed). Due to the nature of fixed location, the output observations of ILDs are only monitoring traffic at one point of road link. Therefore, they mainly fail in measuring spatial characteristics of traffic, which becomes the most inevitable shortcoming when the observed data is used to estimate travel time along a section of road. Besides the systematic errors due to the broken cables, electronic interference, communication noise, software error etc, the flow measurement from ILDs has the biased error resulting from the problem of lane-cross and low polling frequency (Krishnan 2008a).
- **ANPR:** the output from ANPR camera system is typically matched travel time between two locations where cameras deployed. Although ANPR cameras can output the travel time measurement explicitly, the low recognition and matching rate degrades the measurement performance and introduces bias. For example, the number plates of large vehicles such as bus and lorry are usually attached at different positions which are difficult to perform the process of the automatic recognition. It indicates that the observed travel time is aggregated from the class of relatively faster vehicles, since the class of large vehicles which has the low recognition rate tends to be slower. Nevertheless, it is reasonable to assume that ANPR technology is able to capture a far more representative travel time samples than the other sensing methods (Wiggins 1999).
- **GPS:** GPS based sensing system can output location and the corresponding time information. When more than one sample is obtained, it is straightforward to calculate

the travel time estimate between the two locations recoded by GPS samples. However, the GPS observations cannot represent overall average travel time of the road link, since it only account for very small proportion of whole vehicle population. The statistical feature of the travel time measurement error is largely affected by the type of probe vehicle and the driving behaviour. For example, trucks which are usually equipped with GPS tracking devices tend to be slower than private vehicles, and the travel time is likely to be underestimated. Moreover, the GPS positioning measurement has the error introduced by map-matching (Quddus 2006). Besides, the low temporal resolution and “canyon effect” in heavily built up urban environments may lead to a missing data problem (Ochieng et al. 2003).

### **2.1.3 Summary**

Based on the above description, each sensor technology has its specific advantages. However, each one also has some particular shortcomings, causing the device or technology to work improperly under certain circumstances. For example, ILDs have trouble measuring congested vehicle flow and only provide point-based time-mean speeds to estimate link travel times as a continuous stream value; GPS based technology cannot provide reliable traffic data collection due to the low density of devices usage and “canyon effect” in urban environment; ANPR matching is suffered from low matching rate. The inherent drawback of each measurement technology is still difficult to overcome, and it is not necessarily true that one type of system is superior to another one. In addition, the temporal and spatial availability of one system may not always satisfy the demand in the real applications. On the other hand, the much greater volumes of near real time data provided by these sensor systems create opportunities for significant improvement in travel time estimation but also raise important research challenges. To take advantages of the availability of heterogeneous sensor sources and meet the challenge of providing a better travel time estimation, the technique of data fusion provides a promising solution to this problem.

## **2.2 Data Fusion**

### **2.2.1 An overview of data fusion**

Multi-sensor data fusion is a technique by which the data from several heterogeneous sensors is combined through an integration/fusion process to provide comprehensive, accurate and

reliable information regarding to the fused objective. The advantages of this technology derives from its ability to track changing conditions and anticipate impacts more consistently than can be normally undertaken with a single data source (even a highly reliable one). Thus, multi-sensor data fusion makes it is possible for a process synergy in which the consolidation of individual data sources creates a more effective resource (Hackett & Shah 1990). An early definition of data fusion could also be found in White (1987) as:

*“a process dealing with the association, correlation, 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. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results.”*

The applications of data fusion have been existed in both military and non-military fields such as battlefield surveillance, guidance for autonomous vehicles, remote sensing, monitoring of manufacturing processes, robotics, and medical applications wildlife habitat monitoring, and detection of environment hazards (Hall 1992; Mnyika & Durrant-Whyte 1994; Dasarathy 1997). Several methodologies have been proposed in the literature for the purpose of multi-sensor fusion under heterogeneous data configurations. Different data fusion techniques were developed to fit into different applications and data set, due to the variety types of the sensor technologies and the heterogeneous nature of the information that need to be fused. These techniques spanned a wide range of areas including artificial intelligence, pattern recognition, statistical estimation and signal processing (Hall 1992).

### **2.2.2 Advantages of data fusion**

Fusing data from multi-sensor provides several advantages over data from one single sensor. Hall & Llinas (2001) gives comprehensive description and discussion about the superior features of data fusion approach in several applications. The advantages of data fusion approach in their works can be summarised as following:

- **Increased confidence:** more than one or one type of sensor can confirm the same object
- **Reduced ambiguity:** joint information from multiple sensors reduces the set of hypotheses about the object

- **Improved detection:** integration of multiple measurements of the same object improves signal-to-noise ratio, which increases the assurance of detection
- **Increased robustness:** one sensor can still contribute information when others are unavailable, inoperative or ineffective
- **Enhanced spatial and temporal coverage:** one sensor can work when or where another sensor can not

### 2.2.3 Review of data fusion approaches within transport domain

As discussed in last section, some of the data fusion based approaches draw largely from other disciplines and fields of studies and applications. The abundant literature and practices in those areas are naturally benefited to the field of traffic engineering. A brief classification of data fusion techniques which have been used within transport domain is summarised as follows:

- **Statistical approaches:** weighted combination and multivariate statistical analysis. This approach incorporates information accuracy and reliability via weights derived from statistical characteristics of estimation error. Among statistical techniques, the weighted mean approach is the simplest which is used for information combination. According to the discussion in El Faouzi (1997), this type of approaches is not applicable to the non-transferable sensor data or dissimilar performance of estimators.
- **Physical model based approaches:** this type of fusion methods takes the advantage of the physical models to describe the relationship between observed data and objective identities. For example, the traffic flow data from ILDs can be seen as a function of traffic link density (Daganzo 1997). Estimation processes, such as Kalman filtering (Chu et al 2005; Herrera & Bayen 2008), maximum likelihood estimation, and least squares approximation (Huang & Leung 2004; 2005), are representative methods and can be considered state-of the practice.
- **Feature based approaches:** feature-based inference techniques do not use physical models. Instead, correlation is performed by mapping the observed data into an identity declaration, which is also addressed as the challenge of identification. The typical feature based approaches includes Bayesian inference (El Faouzi 2006; Lindveld et al 2007), Dempster-Shafer inference (El Faouzi et al 2009), and generalized evidence processing (Klein 2001).

- **Cognitive based approaches:** this type of approaches attempt to imitate the inference processes of human analytics in recognizing object identity. The most widely used techniques include artificial neural networks (Xie 2004), genetic algorithms (Zhong 2004) and fuzzy set theory (Sroka 2004).

The application of data fusion techniques to solve complex estimation problems is not new. With the continuous emergence of a wide range of new traffic sensor technologies, the research trend in transport area is switching to make full use of all available traffic sensor observations to develop new applications. In the literature of transport domain, the interest for data fusion is quite new and it coincides with the development of ITS. The earliest research on the use of the data fusion in transport application was by Ben-Akiva & Morikawa in the late 1980s (Ben-Akiva & Morikawa 1987). In the following years, various applications of data fusion have been investigated and developed in a number of different fields of transport domain such as advanced traveller information systems (Sumner 1991; Tarko & Roupail 1993; Berka et al 1995; Dailey et al 1996; Choi & Chung 2001), automatic incident detection (Ivan 1995; Byun et al 1999; Cohen 2003), network control (Mueck 2002; Friedrich & Minciardi 2003), traffic demand estimation (Ben-Akiva & Morikawa 1987; Ashok & Ben-Akiva 1993; Lundgren et al 2003), accurate position estimation (Wei & Schwarz 1990; Li & Leung 2003; El-Sheimy et al 2006), and traffic monitoring and forecasting (Granger 1989; Cremer & Schrieber 1996; El Faouzi 1999; Nanthawichit et al 2003; van Hinsbergen & van Lint 2007). A comprehensive review in ITS related data fusion applications can be found at El Faouzi et al (2011). Based on the review above, it can be seen that data fusion has become a promising technique and involved in most of transport applications. The framework proposed in this paper focuses on applying data fusion technique into one of the most fundamental elements in traffic engineering: travel time estimation.

### **2.3 Previous work on data fusion in travel time estimation**

The applications of data fusion in traffic estimation started from early 1990s. The practical implementations of data fusion normally adopt a wide range of algorithms from mathematics, statistics, signal processing and artificial intelligence. To demonstrate the applicability of data fusion in TTE problem, this section gives a general review of the previous works which applied various techniques from data fusion aspects.

### 2.3.1 Related data fusion projects

Sumner (1991) utilised the technique of fuzzy logic in data fusion in two earlier IVHS systems, Pathfinder, in Los Angeles, and TravTek, in Orlando. In both projects, the data fusion process involves a fuzzy logic maximum height solution process. It assigns a value of quality and age to the data sources, and a score is produced by taking account of both value of quality and age. The data source with the highest score is considered the best estimator and its data is selected as the fusion output. Although this method is simple and successful as shown in the paper, it does not make full use of data from all sources, because the data from the rest of sources are completely ignored. Abandoning the remaining information neglects the possible interaction or interrelationship among the various data sources which could contribute to estimation performance.

Rouphail et al. (1993) applied several artificial neural networks (ANN) to model data fusion processes in the ADVANCE project. The nature and structure of neural networks are appropriate for solving rather complex problems, in which the exact relationships among variables and elements are not well understood and modelled. Hence, in their application, this ANN-based data fusion model is applied to recognize traffic flow patterns and estimate travel times for a major arterial corridor. The model fuses data from several sources on several road segments and is capable of extracting the correlations among different data sources. The preliminary results show that ANN-based data fusion models are able to offer relative correct travel time even when the raw data set has noise and errors.

Berka et al. (1995) proposed weighted averaging technique as data fusion method in the ADVANCE project. Their approach was applied to fuse travel time estimates from probe vehicles and fixed detector. The key of weighted averaging technique is to compute weights for the samples of each data source. In their application, several variables are used to perform weights calculation, including:

- the sum of weights comprising of a) reasonable probe vehicle data and b) weight assigned to fixed detector data
- standard deviations of a) probe vehicles travel time reports and b) fixed detector travel time reports
- fusion adjustment factors, which is used to control the contribution of each data source to the fused value

The advantage of their method is that weighted averaging has a complex structure and is able to take into account a large number of factors affecting the weight of each source. However, the determination of those variables is a difficult procedure and some of the model parameters have to be estimated in terms of historical data.

### **2.3.2 Data fusion approaches**

Based on the classification of data fusion approaches in the Section 2.2.3, this subsection briefly reviews the most relevant data fusion works which adopt the most representative technique from each of the categories: weighted averaging from statistical approaches, Artificial Neural Network (ANN) from cognitive based approaches, Bayesian from feature based approaches and Kalman filter from physical model based approaches.

Westerman et al. (1996) also applied weighted averaging fusion method for TTE and incident detection. Their TTE method primarily relies on ILD data with the enhancement and supplement from probe vehicle data. The method employs a parallel structure, in which different probe vehicle and loop detector algorithms perform with the support from each other. The estimates from each algorithm (i.e. hybrid loop detector algorithm, probe vehicle algorithm, and local-related and section-related comparison algorithm), are combined to make a final estimation, where the weighted averaging fusion method is applied. The weight factor of each component is determined by the numbers of the performed verification steps. The effectiveness of this data fusion method is not quantitatively approved and still questionable.

The most representative technique of cognitive based approaches is Artificial Neural Network (ANN), and it shows a great flexibility in TTE process. It because ANN as a data driven based methodology is able to make full use of these data sources and fit them into one general model. The advantage of ANN is that it can solve complex problems in which the precise interrelationships among elements are not well understood and defined. Roupail et. al (1993) applied several ANNs to fusing data from several sources on several road segments and they are capable of finding correlations between different data sources. This approach also has provisions to combine other types of data sources with different certainty degrees and attributes. Tao et al (2006) investigated and compared the performance of different types of neural networks. The finds from their work indicates that different types of neural network may specialize in different traffic context.

Bayesian based methods from the class of feature based approaches interpret the concept of probability as a measure of a state of knowledge. It provides a general formulation for reasoning about partial beliefs under conditions of uncertainty. Bayesian approach is suitable for multi-sensor sources fusion because the observations from all the sensor sources can be represented in the multivariate distribution function and processed within Bayesian framework. Choi & Chung (2002) used Bayesian linear pooling technique to construct a joint prior distribution of the two data sources, and processed by a fuzzy regression model on the support of historical data in this approach. Other work related to Bayesian based data fusion in data fusion can be found at Thomas (1998), EI Faouzi (2006) and Lindveld et al (2007).

Among physical model based approaches, Kalman filter technique is the most powerful tool to dynamically estimate the states of traffic from sensor observations. Nanthawichit et al (2003) proposed a Kalman filter based method for integrating probe vehicle data into fixed detector data to estimate traffic states on a freeway. Although this fusion method does not consider the reliability of data from each source and the dynamic varying of traffic condition in real time, the results presented in the paper still show a certain level of improvement when comparing with the estimate results only based on single sensor source. Chu et al (2005) uses the standard linear Kalman filter framework to fuse fixed loop detector data and probe vehicle data into the travel time estimation for the freeway. The gain of this approach is that it can take advantage of most reliable data of each sensor source, e.g. flow from ILDs and travel time from GPS (comparing with travel time estimation from fixed location sensors). The nonlinear version of Kalman filter (e.g. Extended Kalman filter and unscented Kalman filter) has also been developed for TTE in Gazis & Liu (2003), Wang & Papageorgiou (2005) and Ye et al. (2006), but all of them are all based on single sensor source. The use of nonlinear Kalman filter in data fusion has not been investigated in previous work.

## **2.4 Discussion**

According the above review, it can be seen that all of these approaches are able to perform the task of data fusion in travel time estimation. The approach chosen in this research as the fundamental methodology is the Kalman filter. Compared to the Kalman filter, the other popular methods such as ANN and Bayesian are less attractive due to their own inherent drawbacks. The key of ANN fusion method is to approximate the relationship between the travel time and input data by a learning/training process. However, it is a redundant process since this relationship has been well modelled in a number of traffic flow theories. The purely

data driven based approximation of ANN lacks robustness, especially when the data sources are highly unreliable. In addition, the learning process of ANN requires large amount of true historical travel time data, which makes ANN a less practical method in real world applications.

The Bayesian based method also needs historical data to build up its hypothesis system which is crucial for the estimation process. Furthermore, even if the historical data is available, it is still rather difficult to define multiple hypotheses and prior likelihood functions. From the review above, there are two approaches to address this. One requires the access to the validated historical data set with a certain level of statistical confidence to obtain the prior knowledge. The other one is to assume it has a certain form such as Gaussian, Mixture Gaussian or Poisson which is not always reliable and robust to reflect the condition of high level of traffic flow. Hence, both of them have their own limitation in practice. In addition, when there are a large number of hypotheses and conditionally-dependent multiple events involved, the complexity of disaggregating and estimating become very high.

The nature of the proposed Kalman filter based methodology is to estimate time-varying traffic state by a macroscopic traffic flow theory and optimally modify the estimates by the observations from multiple sensor sources. The Kalman filter technique becomes a promising and superior solution to the problems discussed above for three reasons. Firstly, Kalman filter can make use of well-developed traffic flow model to describe the physical relation between traffic states and observed traffic variables, which eliminates the dependence on historical data. Secondly, the data from all available sensor sources is treated as an update/correction to the state estimate. It makes the estimation more reliable than the other data-driven based methods because it combines the knowledge from traffic flow theory and information from multiple sensor sources in an optimal way. Thirdly, the Kalman filter is more flexible and robust in the real implementation since the filter can output the best estimate when the number of the sensor observations is changing or even there is no any sensor observation. A comprehensive review of using the Kalman filter to estimate travel time is given in Chapter 4 where more detailed description and discussion about the Kalman filter is also provided.

## Chapter 3

# A Travel Time Estimation Model by ILDs Data with the Presence of ANPR Observations

ILDs have been widely deployed and provide rich information about traffic in practice. A variety of TTE methods based on ILDs data have been reviewed in Chapter 2. The most fundamental problem in this field is how to translate fixed point based observations from ILDs into section based measurement, i.e. travel time or space-mean-speed. A well-known relationship between the time-mean-speed and the space-mean-speed was derived by Wardrop (1952). However, this relationship cannot be used in practice to estimate travel times as it requires knowledge of the variance of the space-mean-speed. The variance of the space-mean-speed is not measured by the ILDs and is normally not available in practice.

The contribution of this chapter is the development of a travel time estimation model based on the further exploration of Wardrop (1952)'s formula. An approximate relationship between time-mean-speed obtained from ILDs and space-mean-speed is proposed and used in the estimation model. In addition, a property of traffic state dependency is illustrated, and a more refined model is presented where the traffic states are segmented according to the flow and occupancy values obtained from the ILDs. The formulation of this relationship is based on the correlation between ANPR and ILDs observations. This model can be implemented in highway links which have plenty of ILDs data but very few ANPR paired observations. This model is attractive to the traffic engineering applications because the proposed relationship between space-mean-speed and time-mean-speed can be easily implemented in real-world scenario. From the perspective of data fusion, this model fuses the data from ILDs and ANPR,

and provides more accurate, reliable and robust travel time estimate, which is one of the objectives of this thesis.

## **3.1 Introduction**

ILDs provide a number of point based measurements of traffic variables such as spot speed or time-mean-speed, flow and occupancy. These output measurements are used to evaluate traffic performance for traffic operation and management. Besides such point based measurements, travel time is also an important indicator of traffic performance. However, ILDs cannot directly provide link travel times. Estimating travel time using data from widely deployed ILDs would be attractive to traffic management agencies, as models that use data from existing devices can minimise additional costs for obtaining more information.

For example, every link in England's highway network is equipped with ILDs to monitor and record the traffic flow, occupancy and time-mean-speed (Highways Agency 2009). Some roadway sections consisting of multiple links are installed with ANPR cameras to measure travel time. Due to the high cost of procuring and installing ANPR cameras, relatively few road sections have ANPR cameras in the highway network. Hence, there is no travel time information available for most of the links. Therefore, models that estimate travel time using data from ILDs is of particular interest for those links without ANPR cameras. Secondly, reliability and robust of travel time observations from ANPR cameras are diminished due to the problems of low matching rate and missing data (discussed in Chapter 2). Motivated by these practical problems, this chapter presents a novel travel time estimation model by fusing ILD and ANPR data. The proposed models are testing using the data obtained from the ILDs on the highways in England through the DATEX-II feed disseminated by the National Traffic Control Centre (NTCC). The travel time data used for calculating the accuracy of the proposed models are also obtained from the NTCC feed from the links equipped with ANPR cameras. The results demonstrate that the proposed formulation can estimate the space-mean-speed, and hence the travel time, accurately using real-world data.

## **3.2 Background**

### **3.2.1 Time-Mean-Speed vs. Space-Mean-Speed**

As the proposed method aims to model the relationship between Time-Mean-Speed (TMS) and Space-Mean-Speed (SMS), a number of concepts used in the formulation are outlined

here. A comprehensive description can be found in Daganzo (1997). In the domain of traffic theory, two distinct ways of calculating average speed are frequently used: TMS and SMS.

TMS is the average speed of all vehicles passing over a fix point over a period of time  $T$ , and is given by equation 3.1 (Roess et al. 1998)

$$TMS = \frac{1}{N} \sum_i v_i \quad (3.1)$$

where:

$v_i$  : spot-speed of the  $i$ -th vehicle at a fix point

$N$  : total number of vehicles measured during a given time period

From equation 3.1, TMS is defined as the arithmetic mean of the observations, spot-speed  $v_i$  of the vehicles at a given point. By the nature of ILDs measurement, this fix point type of observations can be obtained from output of ILD speed observations. It is can be seen that the ILDs based TMS only reflect the traffic condition at one specific point, rather than the spatial measurement unless the traffic flow is perfectly homogeneous over a length of road link.

In contrast to the fix point measurement of TMS, SMS measures the average vehicle speed spatially. Unfortunately, a variety of the definitions for SMS have existed in the literature (Hall 1996), and they are not necessarily same. There appears to be two main types of definitions. The first one is originally from Lighthill & Whitham (1955), which they attribute to Wardrop (1952). It is defined as the speed based on the average time taken to cross a given distance, or space,  $L$ :

$$SMS = \frac{NL}{\sum_i t_i} \quad (3.2)$$

When the spot-speed of individual vehicle  $v_i$  is measured, the SMS is given as the harmonic mean of these speeds (under the assumption of homogeneous speeds):

$$SMS = \frac{NL}{\sum_i \frac{L}{v_i}} = \frac{1}{\frac{1}{N} \sum_i \frac{1}{v_i}} \quad (3.3)$$

Then, the travel time is obtained by:

$$Travel\ Time = \frac{L}{SMS} \quad (3.4)$$

However, using the definition of equation 3.3 to represent SMS has an ambiguity about which spot speed is used to calculate the harmonic mean. Consequently, equation 3.3 gives an impression that the SMS could be calculated by taking the harmonic mean of spot speeds

measured at a fixed point. Although it contradicts the spatial nature of SMS measurements, Wardrop (1952), Lighthill & Whitham (1955) and Edie (1974) among other literature accepted this use of spot speeds at a fixed point to calculate SMS. This use of measurement is only valid when speeds maintain same with locations. If speeds vary over the length of road link, calculating SMS by using equation 3.3 will introduce a difference between the harmonic mean of spot speed at a point and the speed based on the average travel time over the length of the road link.

The second version of SMS calculation is more explicit defined in Haight (1963) and Leutzbach (1987). It is the average of spot speeds of all of the vehicles on a section of road at one instant of time, given by equation 3.5

$$SMS = \frac{1}{M} \sum_j v_j \quad (3.5)$$

where:

$v_j$  : spot-speed of the  $j$ -th vehicle along a section of road

$M$  : total number of vehicles measured on the section of road at one instant of time

From the above equation 3.5, the definition can be further interpreted as a scenario: taking a photograph of the road section; assuming each vehicle with a speedometer on its top and displaying the instantaneous speed when the photograph is taken; the defined SMS is the average value of all the instantaneous speeds. It is noticed that both of TMS definition (equation 3.1) and SMS definition (equation 3.5) are the arithmetic mean of the spot speed observations. The difference is that TMS is the average of spot speeds at one specific location over a period of time while SMS defined in equation 3.5 is the average of all the spot speeds over a distance of road link at one instant of time. Figure 3.1 illustrates the difference between the calculation of TMS and SMS defined in equation 3.5.

Although the second version of SMS (defined in equation 3.5) is more explicit than the first version (defined in equation 3.3), it is not suitable to calculate SMS in practice since it is impossible to obtain the spot speed of every vehicle on the road at any instant time. The first version of SMS definition required the assumption of homogeneous vehicle speeds along the road link, and it is approximately valid in freely flowing traffic such as highway. In addition, the first version SMS has a direct relationship with travel time (defined in equation 3.4) which makes it much more useful in both academic research and practice. The travel time estimation model presented in this chapter is also based on the first version of SMS definition.

Note that the first version of SMS definition will be used throughout this thesis such as the fusion model developed in Chapter 5.

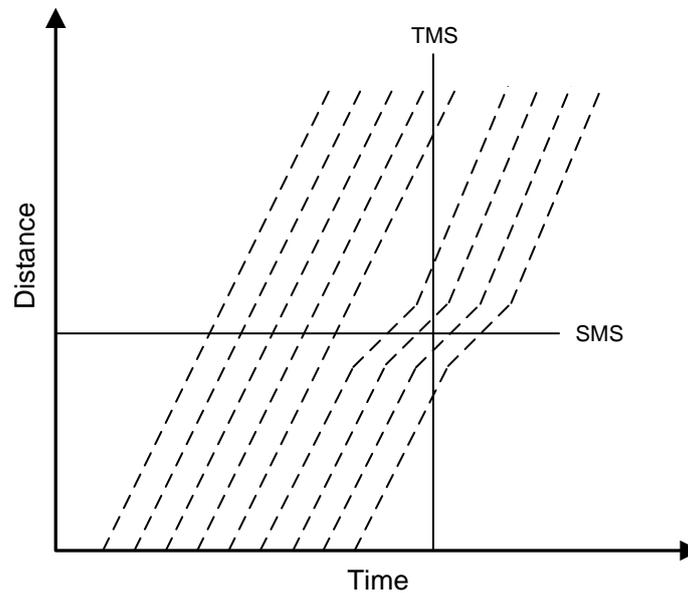


Figure 3.1 Difference between TMS and SMS

### 3.2.2 Existing SMS estimation methods

There exist a variety of travel time estimation methods in the literature, and a comprehensive review of those methods has been given in Chapter 2. This chapter focuses on developing a relationship between SMS and TMS, so that the TMS observations from ILDs can be used to estimate travel time. This section only reviews the related works which investigated the relationship between SMS and TMS.

The earliest research in this field is found in Wardrop (1952) who derived a relationship between TMS and SMS using the concepts of time-speed distribution and space-speed distribution. This well-known formulation is given as follows:

$$TMS = SMS + \frac{\sigma_{sms}^2}{SMS} \quad (3.6)$$

The introduced variable  $\sigma_{sms}^2$  is the variance in vehicle spot speeds about the SMS. A complete derivation of this formula is provided in the Appendix B. It can be seen that equation 3.6 has modelled a direct relationship between TMS and SMS in a simple form. With the knowledge of SMS and the variance of SMS,  $\sigma_{sms}^2$ , TMS can be calculated by this equation. On the other hand, in order to obtain SMS, TMS and  $\sigma_{sms}^2$  are required to be known. However, in most of practical cases,  $\sigma_{sms}^2$  cannot be obtained, which is why it is desirable to estimate SMS from the available TMS data. Due to the presence of the unknown (or even

immensurable) variable  $\sigma_{SMS}^2$ , in practice, Wardrop (1952)'s formulation is not able to be used to calculate SMS directly. This part of research explores the derivative of  $\sigma_{SMS}^2$ , and proposed a novel model to approximate  $\sigma_{SMS}^2$  by TMS from ILDs data, which makes Wardrop (1952)'s formulation an attractive method for travel time estimation.

Garber and Hoel (2001) developed a linear relationship between TMS and SMS, shown as:

$$TMS = 0.996SMS + 3.541 \quad (3.7)$$

Their model is based on a linear regression between observed TMS and SMS datasets from Interstate Highway 880 (I-880) in U.S. Although their method is simple and straightforward, the model parameters are subject to a specific road link and traffic stream characteristics. The model slope (0.966) and constant (3.541) are obtained by linear regression of the data from specific traffic data, which cannot be seen as the general cases. Rakha and Zhang (2005) examined this linear model by using traffic data also from Interstate Highway 880 (I-880) in U.S. The results showed that the TMS and SMS data is linearly fitted by different model parameters. Specifically, the optimal model constant was 2.389 rather than 3.541 in the model, and the slope was 0.986, as opposed to 0.966. According to the data (from England highways) used in this research, the optimal model slope is 0.7618, as opposed to 0.966, and the optimal constant is 33.1252, as opposed to 3.541. Hence, Garber and Hoel's model requires calibration for particular road link and could not be generalised in other scenarios. Another cause of inaccuracy is because linearity of the proposed relationship is not adequate in all circumstances to reach a desired level of accuracy.

Rakha and Zhang (2005) presented a formulation similar to Wardrop (1952)'s model to approximate SMS from TMS. The proposed formulation is given in equation 3.8.

$$SMS \approx TMS - \frac{\sigma_{TMS}^2}{TMS} \quad (3.8)$$

This model uses the statistics of the estimates to derive a modified relationship between SMS and TMS which computes SMS as a function of TMS. It can be seen from above equation that the variance of TMS,  $\sigma_{TMS}^2$  is required for the SMS calculation. However, the variance of TMS required by this model is not available from typical ILD installations in practice, e.g. the DATEX-II traffic data feed provided by the Highways Agency (HA) in England. Typically, only TMS, flow and occupancy are available from ILDs. Hence, their method is of limited practical relevance.

From the reviews on the existing literature above, although a number of methods have been proposed to model the relationship between SMS and TMS, lack of generalisation or requirement of observation which is normally not available in practice make these methods less attractive in real traffic applications. In order to overcome these drawbacks, a novel SMS estimation method is proposed in this chapter.

### 3.3 Methodology

#### 3.3.1 A further derivative on Wardrop's formulation

As discussed in 3.2.2., the main impediment to using Wardrop's formulation in equation 3.6 is that  $\sigma_{sms}^2$  is unknown. In order to make use of the relationship given in equation 3.6, a further formulation needs to be derived. Recalling  $\sigma_{sms}^2$  is the variance in vehicle instantaneous speeds  $v_i$  about the SMS, and defined in equation 3.9 as follows:

$$\sigma_{sms}^2 = \mathbb{E}[(v_i - SMS)^2] \quad (3.9)$$

where  $\mathbb{E}$  is the notation of expectation. Based on equation 3.9,  $\sigma_{sms}^2$  can be expanded as:

$$\sigma_{sms}^2 = \mathbb{E}[(v_i - SMS)^2] = \mathbb{E}[v_i^2] + SMS^2 - 2SMS\mathbb{E}[v_i] \quad (3.10)$$

The stationary traffic condition means that the traffic stream maintains same level of flow and speed along a stretch of road during the period of observation. Assuming stationary traffic conditions (Daganzo 1997), the instantaneous speed can be approximated to the spot speed measured by ILDs. According to this assumption, the vehicle instantaneous speed  $v_i$  in equation 3.10 is approximately equivalent to the spot speed of  $i$ -th vehicle measured by the ILD. Hence the instantaneous speed  $v_i$  in the following content is considered as spot speed, which leads to  $\mathbb{E}[v_i] = TMS$ . By this means, equation 3.10 can be re-written as follows:

$$\sigma_{sms}^2 = \mathbb{E}[v_i^2] + SMS^2 - 2TMS \cdot SMS \quad (3.11)$$

Replacing  $\sigma_{sms}^2$  in equation 3.6 using the expression in equation 3.11 and re-writing it:

$$2SMS^2 - 3TMS \cdot SMS + \mathbb{E}[v_i^2] = 0 \quad (3.12)$$

If it is assumed that  $\mathbb{E}[v_i]$  has a known value, then equation 3.12 is a quadratic equation with only SMS unknown. Using the quadratic formula, the solution to equation 3.12 is:

$$SMS = \frac{3TMS \pm \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]}}{4} \quad (3.13)$$

Equation 3.13 shows a direct way to calculate SMS by TMS under the assumption of known  $\mathbb{E}[v_i]$  value. The above quadratic solution provides two roots which are distinguished by “ $\pm$ ” in the equation 3.13, whereas, during any time interval, the SMS has only one value. It is obvious that one of the roots has to be omitted, i.e. either “+” or “-”. Next section proves that the root with “-” is invalid from different aspects.

### 3.3.2 Finding the right solution for SMS

From the equation 3.6, it is obvious that:

$$TMS = SMS + \frac{\sigma_{sms}^2}{SMS} \Rightarrow SMS \leq TMS \quad (3.14)$$

The above inequality can be considered as the constraint for the solution of SMS. Therefore, the problem of finding right solution is modelled as:

$$SMS = \frac{3TMS \pm \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]}}{4} \quad \text{subject to } SMS \leq TMS \quad (3.15)$$

The task becomes to validate which root (the one with “+” or “-”) can satisfy the constraint. To approach the roots validation, the constraint is divided into two parts as  $SMS = TMS$  and  $SMS < TMS$ , the rest of this section examines these two conditions separately and theoretically proves that only the root with “+” is valid.

#### **Constraint – 1: $SMS = TMS$**

From equation 3.15, only when  $\sigma_{sms}^2 = 0$ ,  $SMS = TMS$  is satisfied. Substituting  $\sigma_{sms}^2 = 0$  into 3.11,  $\mathbb{E}[v_i^2]$  is expressed as:

$$\begin{aligned} \sigma_{sms}^2 &= \mathbb{E}[v_i^2] + SMS^2 - 2TMS \cdot SMS \\ \Rightarrow \mathbb{E}[v_i^2] &= TMS^2 \quad \text{when } \sigma_{sms}^2 = 0 \text{ and } SMS = TMS \end{aligned} \quad (3.16)$$

Using equation 3.15 to replace  $\mathbb{E}[v_i^2]$  in the quadratic solution of SMS in equation 3.13:

$$SMS = \frac{3TMS \pm TMS}{4} \quad (3.17)$$

To satisfy the constraint of  $SMS = TMS$ , the symbol in equation 3.17 has to be “+” and the root with “-” is not valid and omitted. It means the solution for SMS has to be:

$$SMS = \frac{3TMS + \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]}}{4} \quad (3.18)$$

#### **Constraint – 2: $SMS < TMS$**

According to the derivative result from Constraint-1, the root with “-” has been eliminated. Then, the rest of the task is to prove the solution in equation 3.17 also satisfies Constraint-2.

**Definition -1:** 
$$\Delta = \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]}$$

**Definition -2:** 
$$f(x) = x^2$$

Recall =  $\mathbb{E}[v_i]$  , then

$$\Delta = \sqrt{9f(\mathbb{E}[v_i]) - 8\mathbb{E}[f(v_i)]} \quad (3.19)$$

Write:

$$v_i = \bar{v} + \epsilon \quad (3.20)$$

where  $\bar{v} = \mathbb{E}[v_i]$ , so  $\mathbb{E}[\epsilon] = 0$ . Then

$$\Delta = \sqrt{9f(\bar{v}) - 8\mathbb{E}[f(\bar{v} + \epsilon)]} \quad (3.21)$$

Apply Taylor series approximation around the mean of  $v_i$ , then

$$\Delta = \sqrt{9f(\bar{v}) - 8\mathbb{E}\left[f(\bar{v}) + \epsilon f'(\bar{v}) + \frac{1}{2}\epsilon^2 f''(\bar{v}) + \dots\right]} \quad (3.22)$$

Note that

$$f'(x) = 2x$$

$$f''(x) = 2$$

$$f^{(n)}(x) = 0, \text{ for } n \geq 3$$

Then

$$\begin{aligned} \Delta &= \sqrt{9f(\bar{v}) - 8\mathbb{E}[\bar{v}^2 + 2\epsilon\bar{v} + \epsilon^2]} \\ &= \sqrt{9f(\bar{v}) - 8\mathbb{E}[\bar{v}^2] - 8\mathbb{E}[\epsilon^2]} \quad (\text{using } \mathbb{E}[\epsilon] = 0) \end{aligned} \quad (3.23)$$

Replace  $\bar{v}$  with  $TMS$ , and apply Definition-2, equation 3.23 becomes

$$\Delta = \sqrt{TMS^2 - 8\mathbb{E}[\epsilon^2]} < TMS \quad (3.24)$$

Substitute above inequality into equation 3.18

$$\begin{aligned} SMS &= \frac{3TMS + \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]}}{4} \\ &< \frac{3TMS + TMS}{4} \\ &< TMS \end{aligned} \quad (3.25)$$

Therefore, equation 3.18 satisfies both of the two constraints, and is a valid solution for SMS in the proposed model.

A theoretical proof is given above to demonstrate that the root with “+” is the right solution for SMS. In addition, May (1990) shows that the difference between TMS and SMS estimates are of the order of 1% to 5% normally by investigating the real traffic data. The quadratic solution can be re-written as:

$$SMS = 75\%TMS \pm 25\% \sqrt{9TMS^2 - 8\mathbb{E}[v_i^2]} \quad (3.26)$$

If the root with “-” applies, the difference between TMS and SMS estimates will be more than 25%, which controverts May (1990)’s findings. On the other hand, the root with “+” could match the findings from May (1990)’s studies.

The proposed relationship between SMS and TMS is formulated in equation 3.18 by a further derivative on Wardrop’s formulation. In this novel formulation, SMS can be calculated from only two variables: TMS and  $\mathbb{E}[v_i^2]$ . TMS is available from ILDs data, while  $\mathbb{E}[v_i^2]$  is typically unknown. Hence, the problem becomes to construct a method to approximate  $\mathbb{E}[v_i^2]$  from available ILDs data, which will be addressed in the next section.

### 3.3.3 Approximating $\mathbb{E}[v_i^2]$ using TMS

According the finding in Section 3.3.2, a solution for calculate SMS is determined in equation 3.18. In the above derivation, it is assumed that  $\mathbb{E}[v_i^2]$  is known; this is not necessarily true in practice.  $\mathbb{E}[v_i^2]$  can be considered as the expected value of the squared spot-speeds measured by the ILD. Since the spot speed of each vehicle is not typically available from the ILD,  $\mathbb{E}[v_i^2]$  cannot be calculated directly using the ILD output. Hence,  $\mathbb{E}[v_i^2]$  needs to be estimated using the known value of TMS. This section introduces an approach which uses a nonlinear model to approximate  $\mathbb{E}[v_i^2]$  by TMS from ILDs data.

Using equation  $TMS = \mathbb{E}[v_i]$ , It can be reasonably assumed that the relationship between  $\mathbb{E}[v_i]$  and  $\mathbb{E}[v_i^2]$  is quadratic, which was supported by analysing a plot between the two quantities using ILD and ANPR data from the England highway network, shown in Figure 3.2.  $\mathbb{E}[v_i]$  is calculated by equation 3.18 with SMS from ANPR observations and TMS from ILD data. Note that the data is aggregated from six England highway links which have both of ANPR and ILD equipped, more detailed information about the data will be provided in 3.4.1. An obvious second-order nonlinear relationship is shown in the plot by visual inspection. Hence, the following quadratic equation was proposed:

$$\mathbb{E}[v_i^2] = aTMS^2 + bTMS + c \quad (3.27)$$

The above equation can be solved by finding the constant coefficients  $\{a, b, c\}$ . A nonlinear regression technique is used to perform the second-order polynomial fitting.

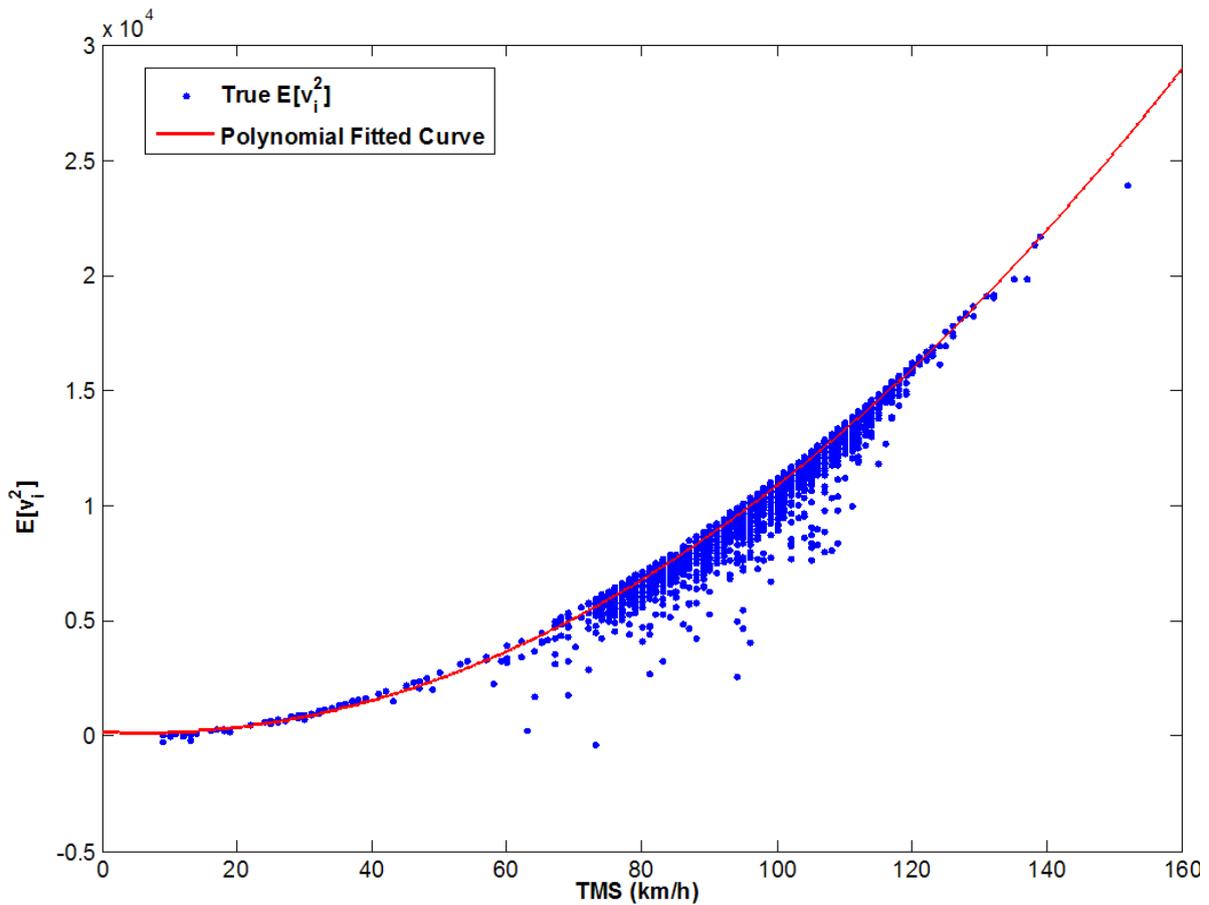


Figure 3.2 Relationship between TMS and  $\mathbb{E}[v_i^2]$

The fitting result is also shown in the Figure 3.2. It supports the earlier assumption that  $\mathbb{E}[v_i]$  is a second-order function of TMS. The constant coefficients  $\{a, b, c\}$  of the quadratic function were estimated to be  $\{a = 1.22, b = -15.21, c = 207.95\}$  with total 9,034 samples and  $R^2=0.9718$ . The  $p$ -value for each of the coefficient  $\{a, b, c\}$  is 0 with t-statistic  $\{-13.04, 13.85, 11.10\}$  respectively.

The data used to perform second-order polynomial fitting are from six highway links with different geographical topologies and traffic conditions, and the data are from one two weeks period. This empirical result demonstrates the generalisation of proposed modelling methodology. In summary, the finding of this novel second-order function suggests a solution to approximate  $\mathbb{E}[v_i]$  by using TMS from ILDs. Consequently, with known TMS and  $\mathbb{E}[v_i]$ , SMS can be estimate from equation 3.18.

### 3.3.4 Fusing ANPR observations into the estimation model

The coefficients of the quadratic function in equation 3.27 are obtained through nonlinear regression using ANPR travel time observations. The travel time observations from ANPR is relatively accurate, thus it seems to be unnecessary to employ ILD data to estimate travel time when ANPR observations are available. However, as discussed in Chapter 2, travel time observations from ANPR has the problems of missing data and low matching rate during a specific time period. It makes the application of only using ANPR to estimate travel time lack of robust and reliability. The introduced model in this chapter proposed a way to fuse ANPR data with ILDs data to perform a more robust and reliable travel time estimation. The fusion process is summarised as following steps:

- Step - 1:** prepare historical TMS observations from ILDs and SMS observations from ANPR
- Step - 2:** solve the coefficients  $\{a, b, c\}$  by regress the available TMS and SMS data in equation 3.27
- Step - 3:** compute  $\mathbb{E}[v_i^2]$  by equation 3.27 for the time period when ANPR observation is not available
- Step - 4:** calculate SMS by equation 3.18

A flowchart illustrates this fusion process in Figure 3.3

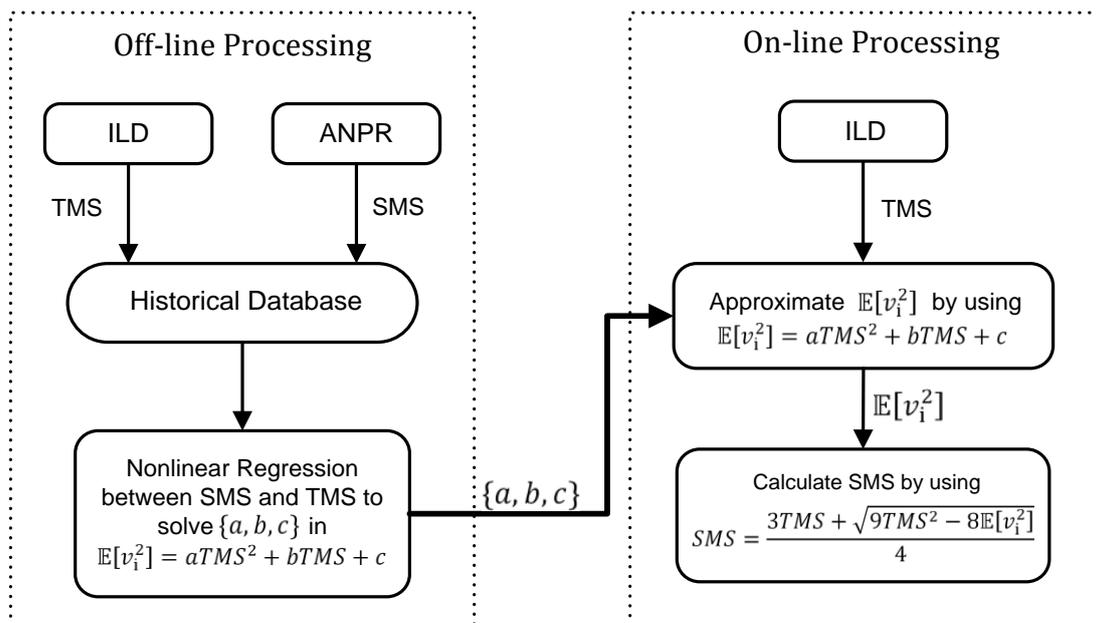


Figure 3.3 Flowchart of proposed fusion process

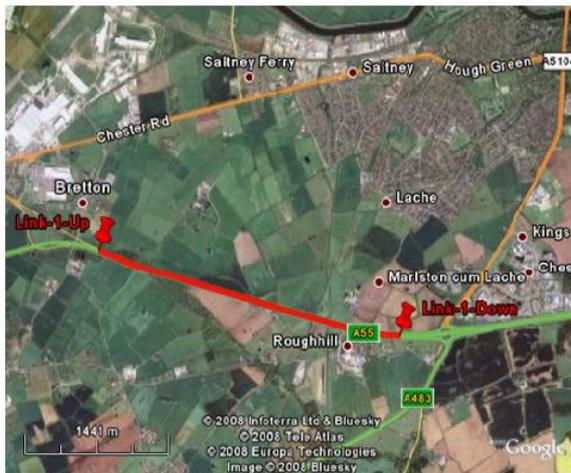
## 3.4 Model Evaluation by England Highway Data

### 3.4.1 Data description

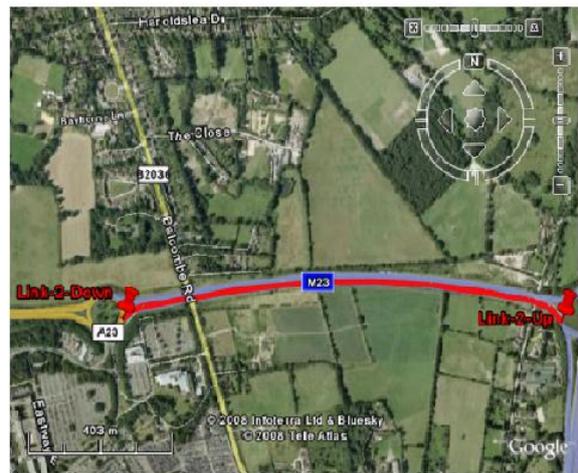
The proposed model is evaluated using ILD data from English highways obtained through the DATEX-II feed disseminated by the National Traffic Control Centre (NTCC). The feed contains ILD data consisting of flow, TMS and occupancy, and travel time data from ANPR links. Both ILD data and ANPR data are provided at 10-minute intervals. This means that every 5 minutes, there is one TMS data point from ILDs and one average travel time data point from ANPR links. The ILDs used in this research are per-lane double loops. The TMS data from ILDs is derived from the double loop using the time difference between the activations of the two loops and the distance between them. In order to test the proposed model, only links with both ILD data and ANPR data were used; it must be noted that most of links have no ANPR data available. In order for the evaluation of the model to be as generic as possible, the links were chosen from a variety of road links with different locations and topologies, and shown in Figure 3.4.

Data from the chosen links for a two week period between 15<sup>th</sup> July and 29<sup>th</sup> July 2008 was used in the evaluation. For the evaluation process, the TMS and SMS data samples are grouped pair-wise for each 10 minutes time interval. Due to sensor or communication failures, either ILD or ANPR camera data were missing for some time period. Time periods when either of the data points was missing were removed from the dataset. The remaining dataset had a total of 9,034 paired ILD (TMS) and ANPR (SMS) observations.

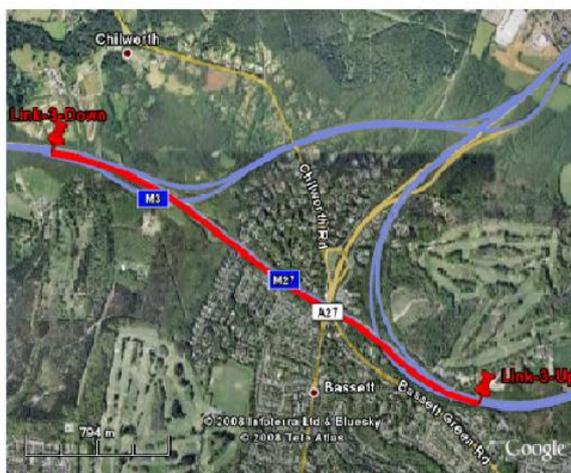
A scatter plot between observed TMS and SMS data is shown in Figure 3.5. Note that the data plotted in the figure is aggregated from above six different England highway links over 2 weeks time period. As discussed in Section 3.3.2, TMS is equal to or greater than SMS (equation 3.14), and the scatter plot clearly shows this relation. The objective of the proposed estimation model is to reduce this bias by using the developed underlying relationship between TMS and SMS. The rest of this section will provide estimation results based on the datasets described above.



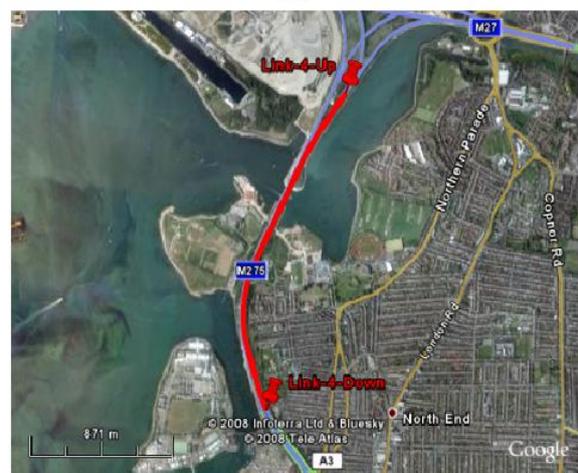
(a)



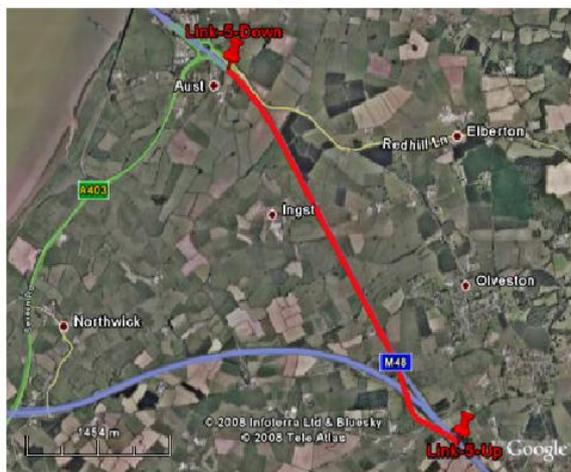
(b)



(c)



(d)



(e)



(f)

Figure 3.4 Location and topology of chosen links: (a) Link-1; (b) Link-2; (c) Link-3; (d) Link-4; (e) Link-5; and (f) Link-6 (Source: Google Earth)

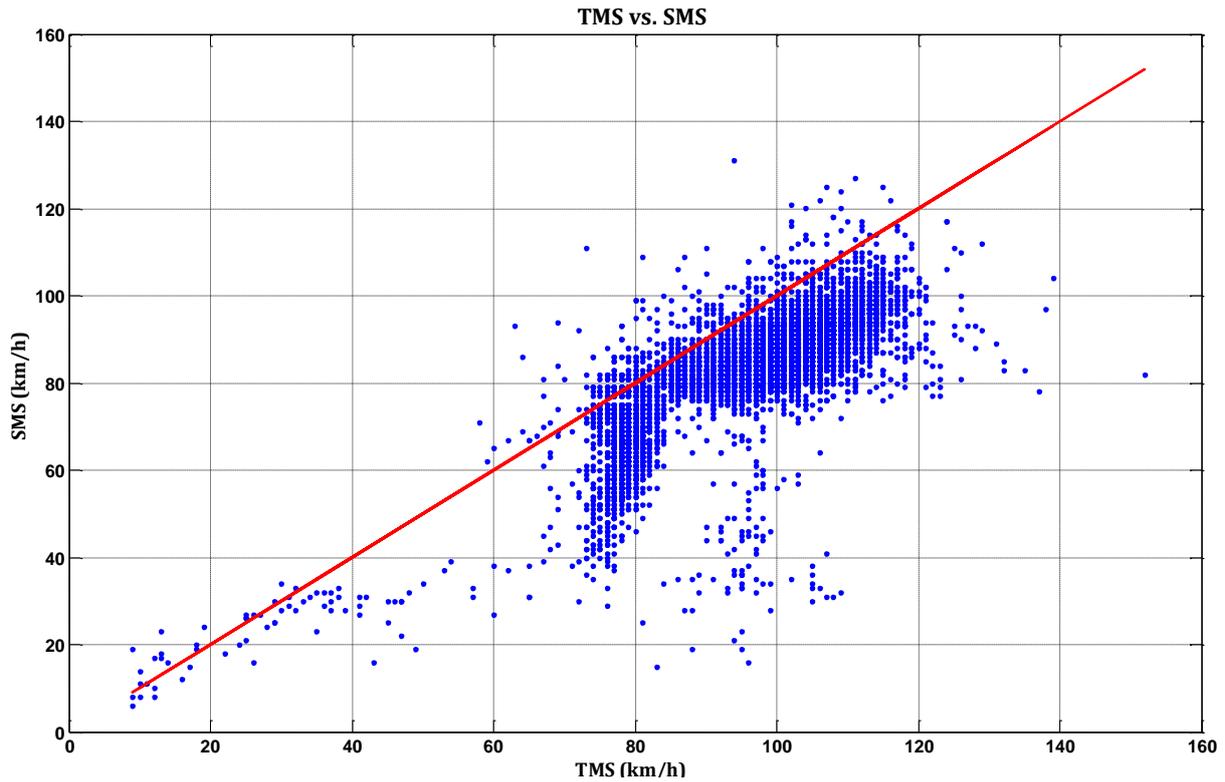


Figure 3.5 Scatter plot between observed TMS and SMS from real-world data

### 3.4.2 Evaluation results & analysis

To evaluate the proposed estimation model, the two weeks period dataset is divided into two parts: first one week and second one week. The first one week data is treated as the historical dataset to approximate coefficients  $\{a, b, c\}$  in equation 3.27. The TMS data from the second one week data is applied as the model input while SMS data from ANPR is used for performance evaluation. Figure 3.2 in the methodology section (Section 3.3) provides an overview of polynomial fitting result based on all the six different highway links. The purpose of that result is to show a general second-order function between TMS and  $\mathbb{E}[v_1^2]$  which is used to verify the proposed approximate formulation in equation 3.27. A separated polynomial fitting results for link-1 is shown in Figure 3.6. The model estimation results for link-1 are shown in Figure 3.7 and 3.8. The results for rest of the links are provided in Appendix C. The Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) of the model estimates are summarised in Table 3.1 for the quantitative evaluation results.

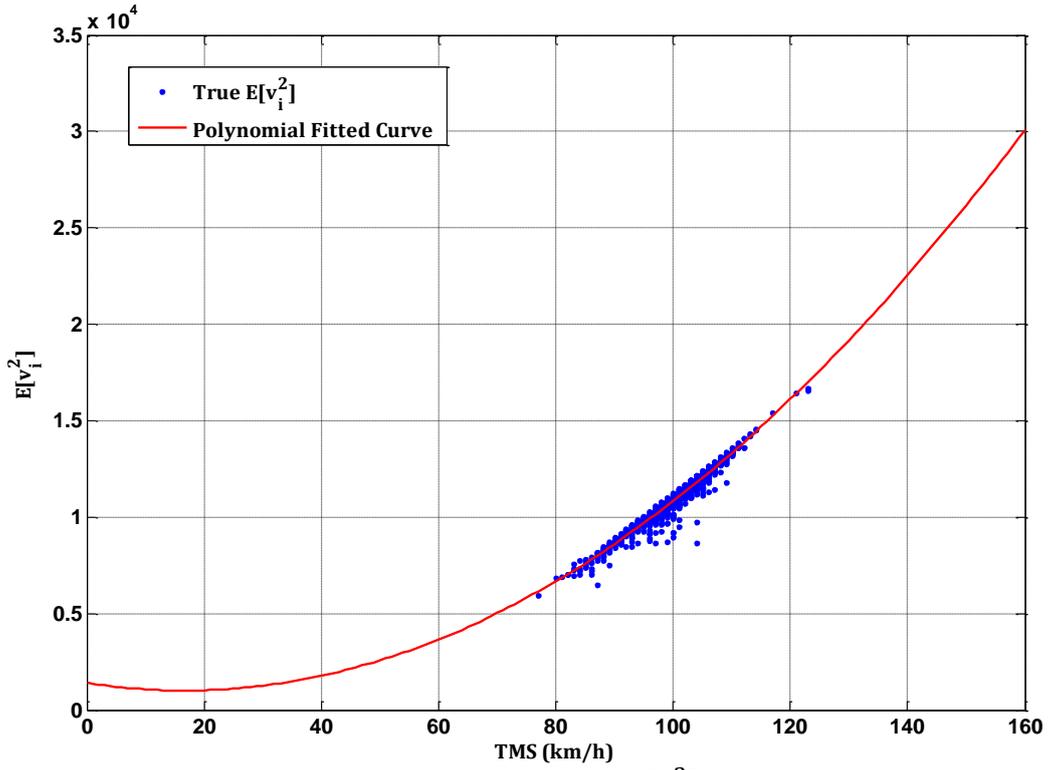


Figure 3.6 Result of polynomial fitting  $\mathbb{E}[v_i^2]$  by TMS for Link-1

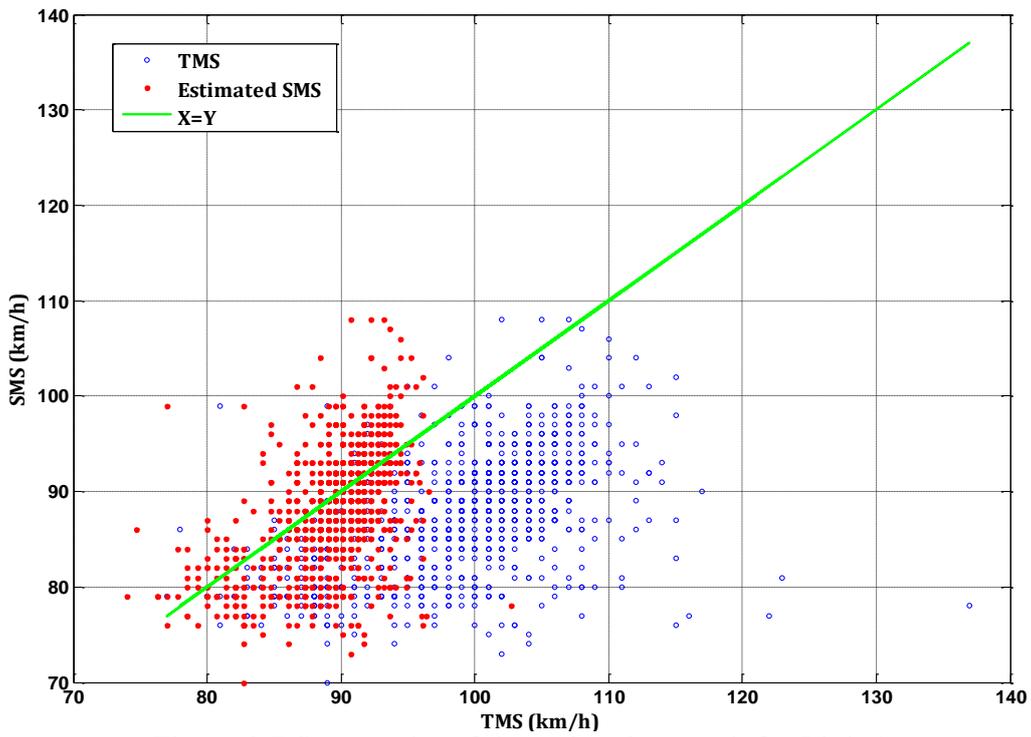


Figure 3.7 Scatter-plot of the estimation result for Link-1

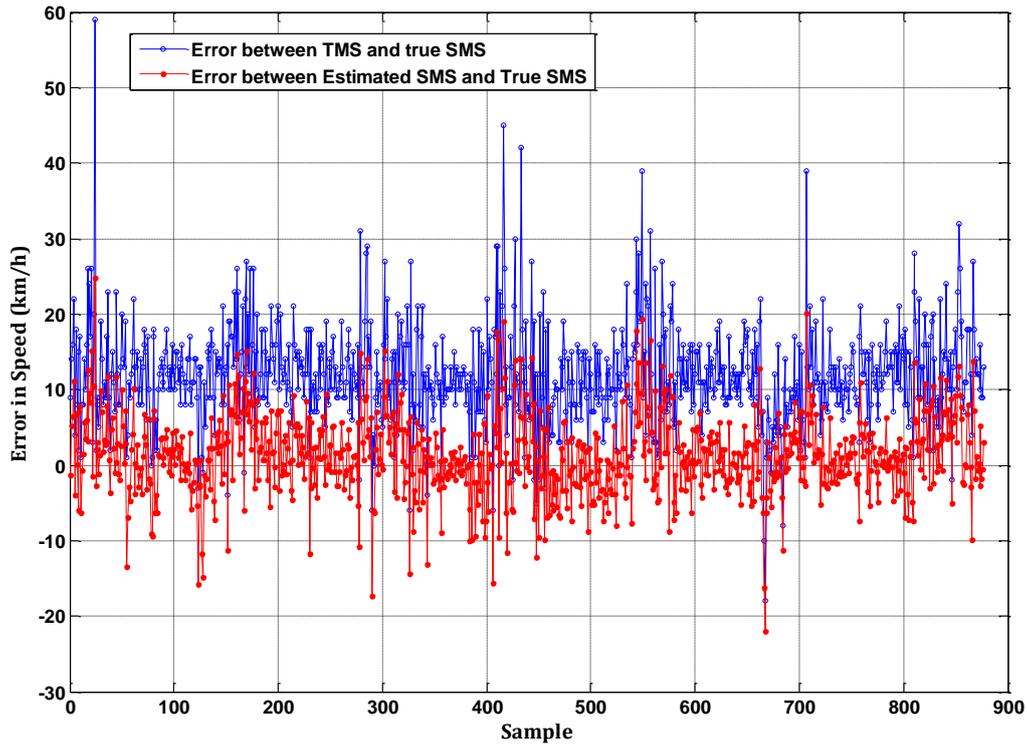


Figure 3.8 Comparison between errors from only using TMS and estimated SMS for Link-1

|                    | Estimation Model          | Link-1 | Link-2 | Link-3 | Link-4 | Link-5 | Link-6 |
|--------------------|---------------------------|--------|--------|--------|--------|--------|--------|
| <b>MAPE (%)</b>    | Only TMS                  | 14.13  | 11.57  | 16.70  | 16.82  | 21.55  | 26.21  |
|                    | Proposed Model            | 4.82   | 3.57   | 4.33   | 7.14   | 15.75  | 17.91  |
|                    | Percentage of Improvement | 65.89% | 69.14% | 74.07% | 57.55% | 26.91% | 31.67% |
| <b>RMSE (km/h)</b> | Only TMS                  | 13.62  | 10.66  | 16.08  | 16.71  | 14.93  | 20.16  |
|                    | Proposed Model            | 5.76   | 4.36   | 5.52   | 9.85   | 11.32  | 14.00  |
|                    | Percentage of Improvement | 57.71% | 59.10% | 65.67% | 41.05% | 24.18% | 30.56% |

Table 3.1 Estimate performance comparison

The scatter plots between SMS and TMS (blue circles) and SMS and estimated SMS (red dots) are shown in Figure 3.7. As expected, TMS is higher than SMS, while the estimated value of SMS is closer to the actual SMS. Figure 3.8 illustrates the comparison between the level of error before and after estimation (TMS and estimated SMS). The result shows that the proposed model reduced the error of speed measurement by approximately 10 km/h; Table 3.1 provides more detailed quantitative results in terms of MAPE and RMSE. From the table, a clear improvement over using TMS to estimate travel times can be seen, which demonstrates the effectiveness of the proposed model in this chapter.

However, the percentage of improvement across different links presents a fluctuation: from 31.67% to 74.07% for MAPE and from 24.18% to 65.67% for RMSE. According to the assumption of homogeneous traffic stream in the model development, changing in the traffic states could introduce error for the modelling process. Hence, the cause of the fluctuation of estimation performance across different links may be due to the varying degree of traffic states for those links. It means that a more stable traffic condition will have better estimation. To illustrate this analysis, Figure 3.9 shows the scatter plots of observations from Link-3 and Link-6 which have the largest and least improvement respectively. The observations within gray area in Figure 3.9 (b) (Link-6 result) are clearly under congested condition, since the TMS value is much higher than SMS. Comparing with the scatter plot from Link-3, Link-6 has much more observations under the congested condition, thus the estimation performance of Link-6 is worse than the performance of Link-3 accordingly. Therefore, the performance of the proposed model is affected by the traffic conditions. To make the proposed model less vulnerable to the varying of traffic condition, a refined model is introduced in next section.

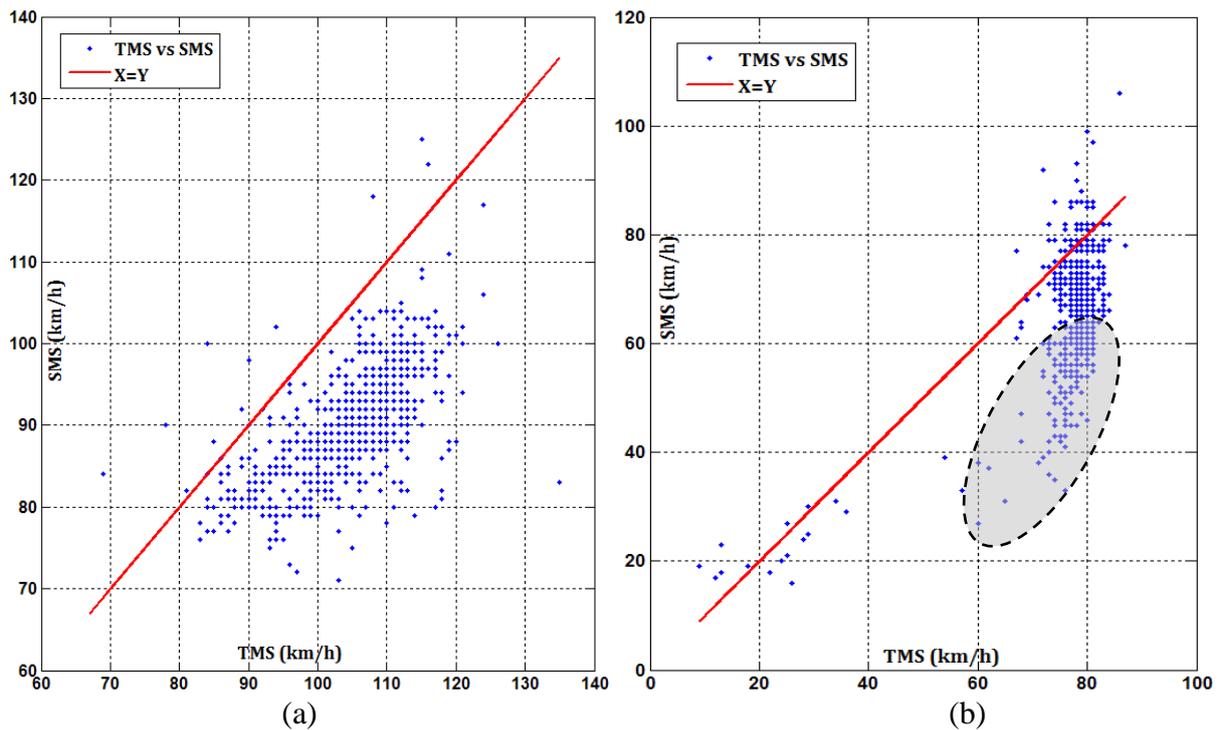


Figure 3.9 Scatter plot between observed TMS and SMS (a) Link-3 result; (b) Link-6 result

## 3.5 Model Improvement by Introducing Traffic State Factor

### 3.5.1 The effects of varying traffic states

According to the analysis and discussion at the end of last section, the traffic state affects the model performance. The proposed model is based on the assumption of homogeneous traffic streams, and the estimation process treats the traffic stream as a constant condition. It is not necessarily true in the real-world applications such as the illustrations in Figure 3.9. In order to take the traffic state into consideration, this section refines the proposed model to improve the estimation performance.

The use of constant coefficients  $\{a, b, c\}$  in equation 3.27 means that the approximation relationship between  $\mathbb{E}[v_i^2]$  and TMS holds good for all traffic conditions. Although the method based on this assumption shows better accuracy, the estimation performance can still be improved by refining the formulation. The calibrated relationship between  $\mathbb{E}[v_i^2]$  and TMS is valid for a given traffic state. However, when traffic states change, the coefficients in this relationship may vary. For example, consider two different traffic states at one link: (1) high occupancy and high flow; (2) high occupancy but low flow. According to this description, the second state represents congested state, while the first state represented uncongested state. It can be reasonably assumed that the variance of spot speed ( $\mathbb{E}[v_i^2]$  equivalently) is different between the first and second states. In the basic model, the same coefficients are used in Equation 3.27 to approximate  $\mathbb{E}[v_i^2]$  regardless of the traffic state. Therefore, the performance of model estimation will be degraded when the traffic state varies a lot.

### 3.5.2 Model refinement by segmenting traffic states

In order to investigate how the difference between traffic states affect the accuracy of the proposed model, a parameter  $\alpha$  is defined to represent the traffic state as follows:

$$\alpha = 1000 \cdot \frac{o}{q} \quad (3.28)$$

where

$o$ : Occupancy measured by ILD

$q$ : Flow measured by ILD

Occupancy and flow are readily available from the ILD output, which are used to calculate the defined parameter  $\alpha$ . Based on the value of the  $\alpha$ , traffic states can be categorised into

different groups. It is assumed that the coefficients of the relationship between  $\mathbb{E}[v_1^2]$  and TMS in equation 3.27 will be different within each traffic state group. In order to test the assumption, the dataset is segmented into 4 regimes to represent different traffic states based on the value of  $\alpha$ :  $\{\alpha \leq 3, 3 < \alpha \leq 5, 5 < \alpha \leq 10, \alpha > 10\}$ . Figure 3.10 shows the polynomial fitting results for these traffic states.

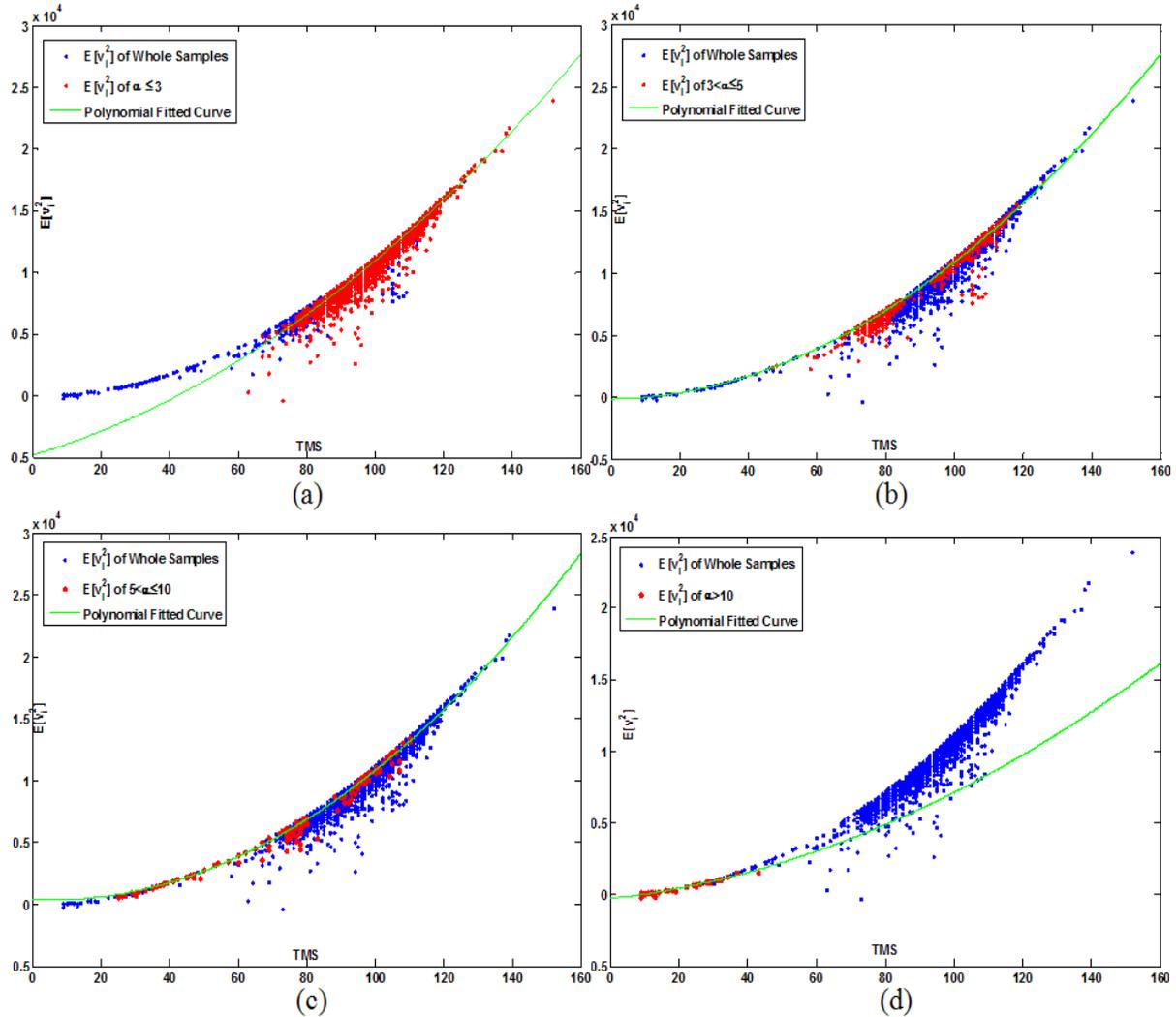


Figure 3.10  $\mathbb{E}[v_1^2]$  vs. TMS under different traffic states: (a) when  $\alpha \leq 3$ ; (b) when  $3 < \alpha \leq 5$ ; (c) when  $5 < \alpha \leq 10$ ; (d) when  $\alpha > 10$

The resulting plots are consistent with expectations. The fitted polynomial curves capture the relationship between  $\mathbb{E}[v_1^2]$  and TMS for a given traffic regime well. More specifically, for every regime defined by the range of  $\alpha$ , the fitted second-order curve has a different set of coefficients  $\{a, b, c\}$ , thus the corresponding  $\mathbb{E}[v_1^2]$  is approximated according to the different traffic states. By this means, for every incoming TMS observation from ILD, the traffic state is examined by the value of  $\alpha$  which is determined by flow and occupancy data, and then

SMS estimation is performed by using the most appropriated polynomial curves. The working process of this refined model is illustrated in Figure 3.11. The refined model based on the segmentation is validated by using MAPE and RMSE; results are shown in Table 3.2.

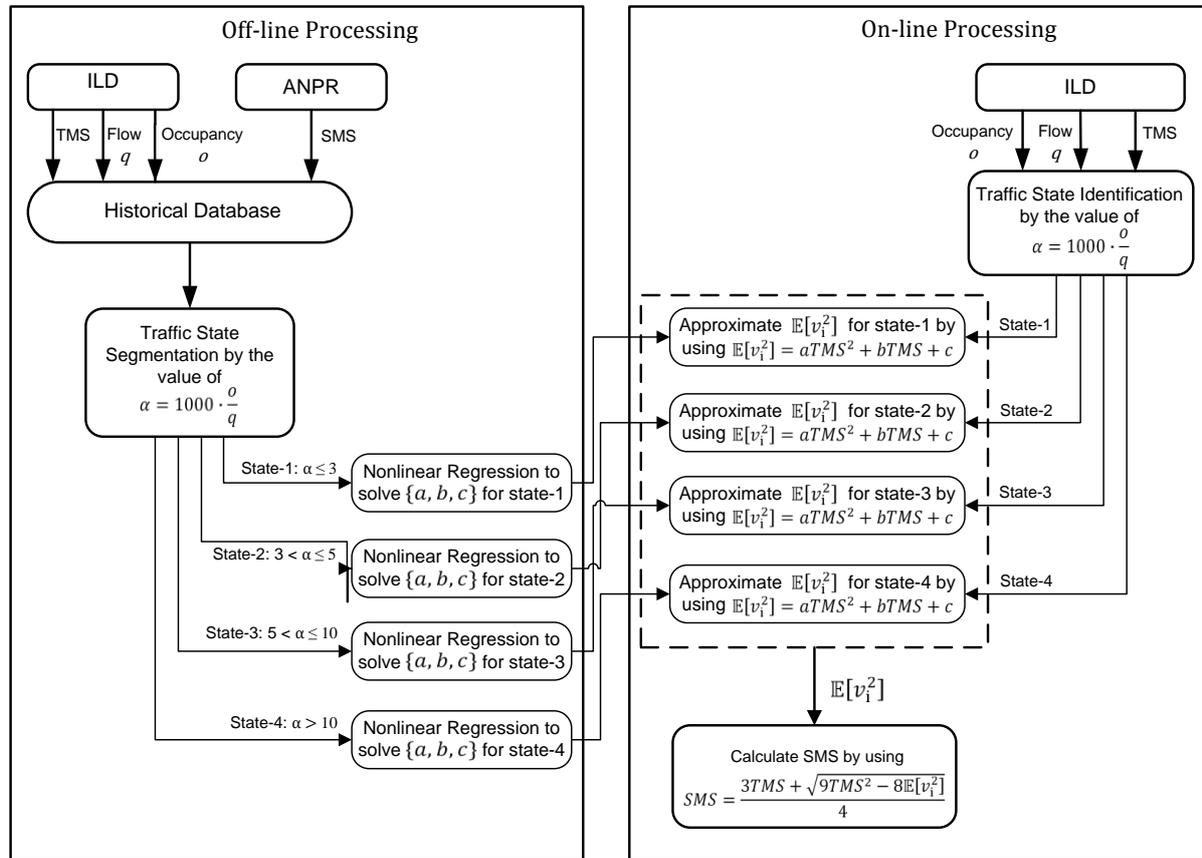


Figure 3.11 Flow Chart of the segmentation based refined model

### 3.5.3 Model refinement by clustering traffic states

The previous section explored how the traffic state affects the SMS estimation results. Although the state dependent method shows better estimation accuracy, the segmentation scheme used is arbitrary and may not be transferrable. The number and boundaries of the segments were determined based on prior analysis of data from the chosen links for which both TMS and SMS were available. For example, the occupancy value reported by an ILD will depend on its electromagnetic sensitivity. Hence the boundary values of  $\alpha$  between traffic states could vary between ILDs. Using fixed value of  $\alpha$  for segmenting traffic states is hence not a transferable approach.

In order to generalise the segmentation approach using flow and occupancy obtained from an arbitrary group of ILDs, traffic states are categorised into congested and uncongested regimes. The ideal relationship (Daganzo 1997) between flow and occupancy is given in

Figure 3.12 (a). The lower segment of the plot represents samples from the uncongested regime, while the upper segment represents the congested state. The scatter plot between real flow and occupancy data used in this study is given in Figure 3.12 (b). Occupancy increases as the flow increases during the uncongested regime, and occupancy decreases as the flow increases during the congested state. It is clear that it is easy to determine if a given data point (flow-occupancy point) is congested or not from the scatter plot by visual inspection, such as the observations within the grey area.

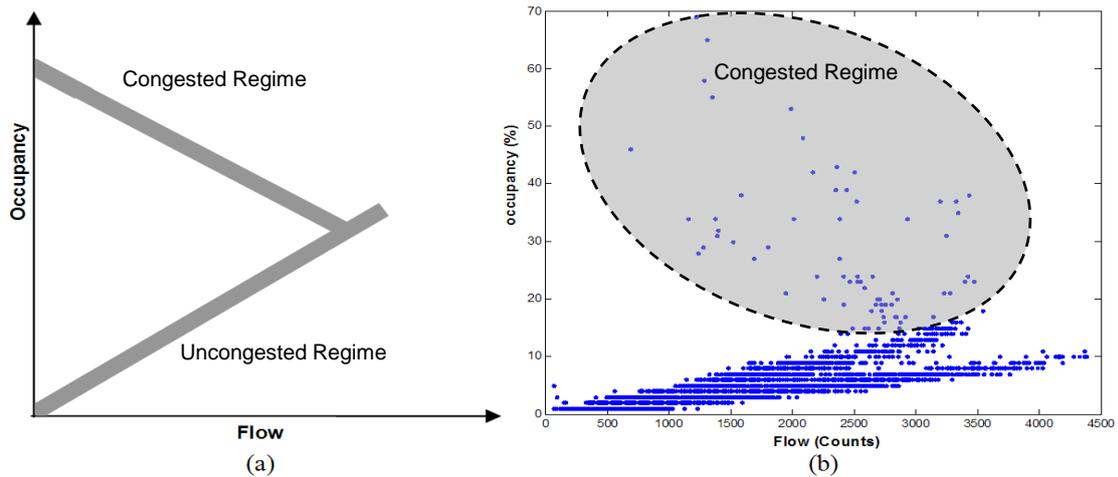


Figure 3.12 Relationship between flow and occupancy: (a) Theoretical relationship; (b) Real traffic data;

However, this differentiation between the traffic states needs to be carried out in an automated fashion for the proposed approach. Clustering techniques such as the k-means clustering (Weisstein 2005) can be used to partition a dataset into a number of different clusters. However, direct application of the k-means clustering technique to partition the data points into two clusters representing congested and uncongested regime did not yield satisfactory results. Specifically, the range of occupancies during congestion can be quite large, thus a number of data points in the congested regime were identified as uncongested. To solve this issue, a linear regression model is fitted to the points identified as uncongested. All data points that are identified as outliers are deemed as congested and are moved to the congested cluster. A two-step clustering approach is suggested as follows:

**Step - 1:** This first step is to cluster the data points into two clusters roughly representing congested and uncongested regimes using k-means clustering. The distance metric used is cosine, which uses the difference between the angles made by two different data points with the origin to determine cluster memberships. The use of the cosine distance metric takes advantage of the fact that the flow vs.

occupancy curve is linear in the congested regime, and most of the uncongested data points should be grouped in the same cluster. However, due to the range of occupancy values in the congested regime, some of the congested data points may be classified into the first cluster of uncongested data points.

**Step - 2:** The second step is to fit a linear regression model on the data points in the uncongested cluster identified in the first step. All the data points identified as outliers by the regression model are moved to the second cluster, representing the congested state.

Through this two-step clustering process, the traffic observations are categorised into congested and uncongested regimes. For each of regimes, the polynomial fitting process was applied to approximate the value of coefficients  $\{a, b, c\}$  which forms two distinct second-order functions as defined in equation 3.27. Consequently, according to the traffic state of incoming data (flow & occupancy from ILD),  $\mathbb{E}[v_1^2]$  is approximated based on one of these two fitted functions. The working process of this clustering based refined model is shown in Figure 3.13.

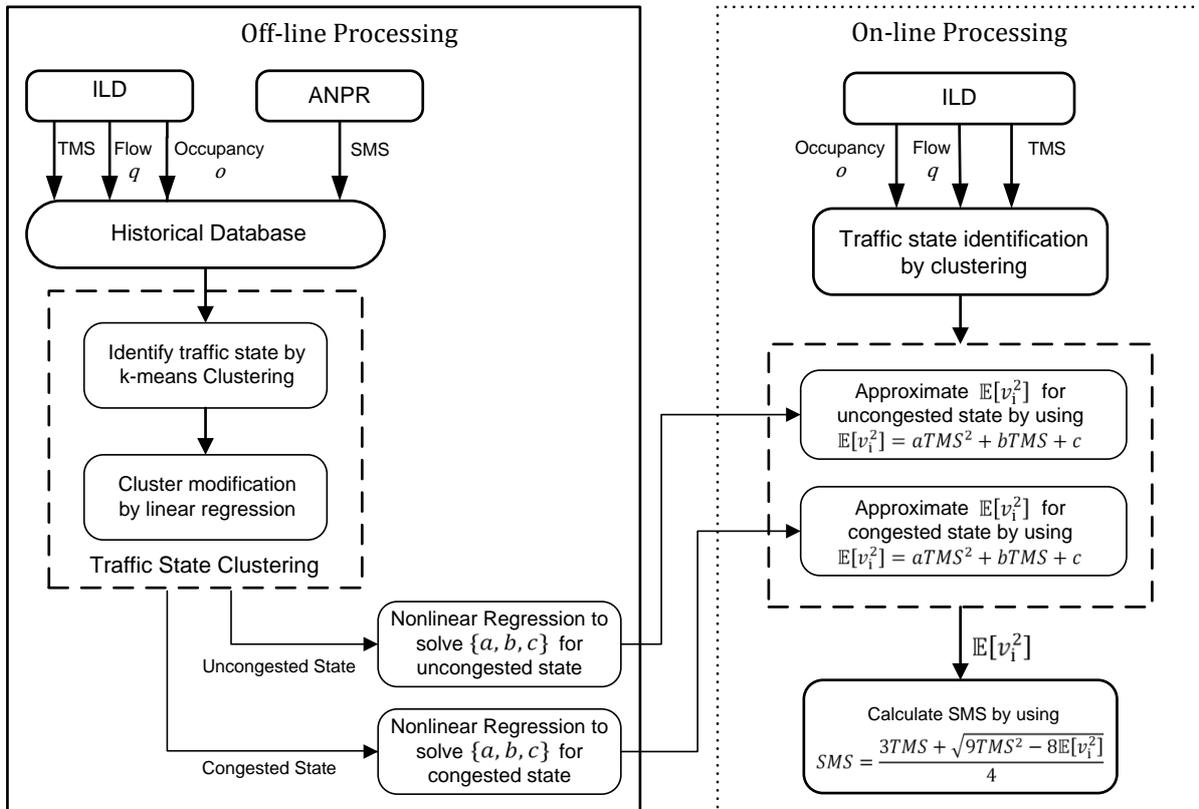


Figure 3.13 Flow Chart of the clustering based refined model

The result of this two-stage clustering is shown in Figure 3.14. It is shown that the proposed method produces satisfactory classification of traffic states from all the six ILDs used for evaluation. The result of polynomial curve fitting is shown in Figure 3.15. A clear difference is shown between the congested and uncongested traffic state. The SMS estimation results based on the two-step clustering method is also shown in Table 3.2.

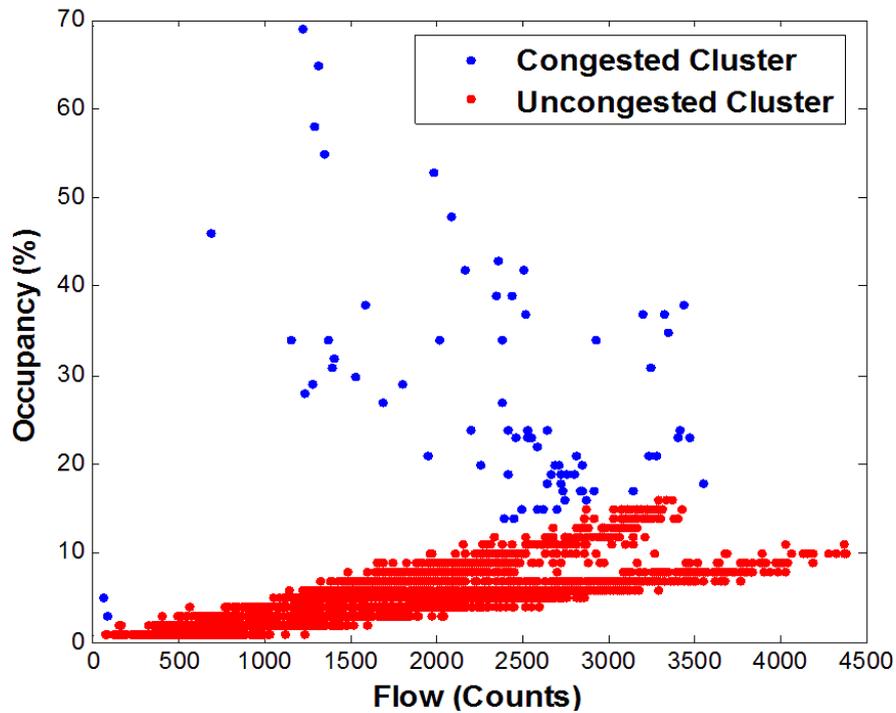


Figure 3.14 clustering result on scatter plot of flow vs. occupancy

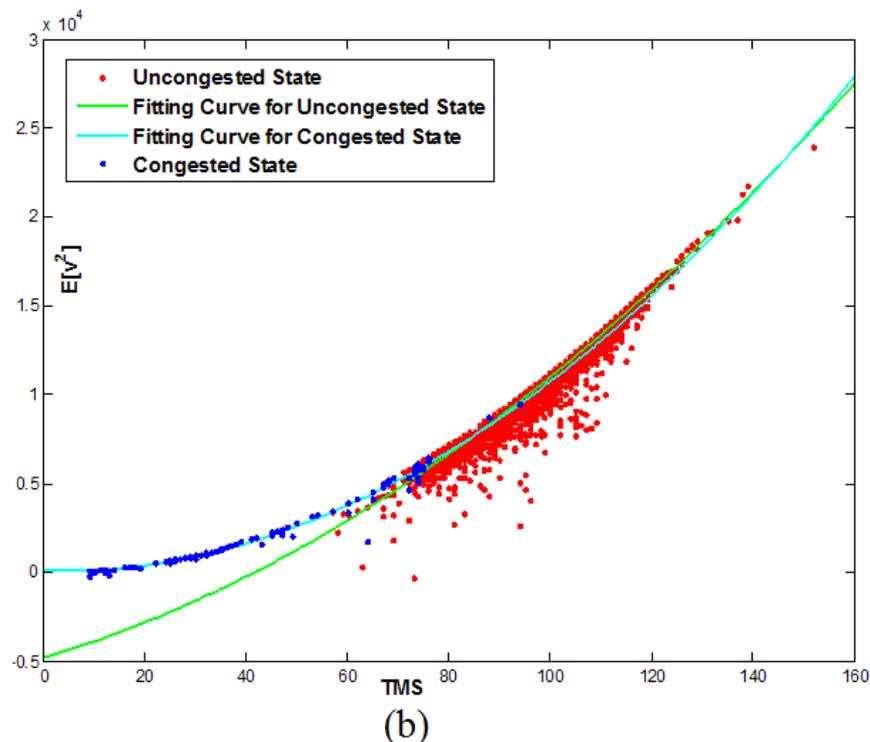


Figure 3.15 Fitting result on each clustered state

|                    | <b>Estimation Model</b>             | <b>Link-1</b> | <b>Link-2</b> | <b>Link-3</b> | <b>Link-4</b> | <b>Link-5</b> | <b>Link-6</b> |
|--------------------|-------------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| <b>MAPE (%)</b>    | Only TMS                            | 14.13         | 11.57         | 16.70         | 16.82         | 21.55         | 26.21         |
|                    | Original Estimation Model           | 4.82          | 3.57          | 4.33          | 7.14          | 15.75         | 17.91         |
|                    | Segmentation based Estimation Model | 3.43          | 2.76          | 2.09          | 6.76          | 13.64         | 15.95         |
|                    | Clustering based Estimation Model   | 3.42          | 2.76          | 2.95          | 6.52          | 14.58         | 16.01         |
| <b>RMSE (km/h)</b> | Only TMS                            | 13.62         | 10.66         | 16.08         | 16.71         | 14.93         | 20.16         |
|                    | Original Estimation Model           | 5.76          | 4.36          | 5.52          | 9.85          | 11.32         | 14.00         |
|                    | Segmentation based Estimation Model | 4.25          | 3.16          | 3.42          | 9.12          | 9.91          | 12.96         |
|                    | Clustering based Estimation Model   | 4.24          | 3.26          | 3.40          | 9.01          | 10.61         | 12.97         |

Table 3.2 Estimate performance comparison between original model and refined model

Table 3.2 shows that SMS estimated using the models presented in this paper, and a clear improvement over using TMS to estimate travel times can be seen. When data points from Links 1, 2, 3 and 4 are largely uncongested, the performance of the refined models (segmented and clustering models) is only marginally better than the originally proposed model. In contrast, Links 5 and 6 have more data points from the congested state. Therefore, the segmentation and clustering based models provide higher estimation accuracy than the originally proposed model. The MAPE values of Link 5 and Link 6 are 17.29% and 17.51% based on the proposed clustering model during congestion period, while the values are 23.24% and 34.84% for the traditional estimation method based on TMS. Although the overall performance of the segmentation based model is better than that of the clustering based model, the generality of the clustering based model makes it transferrable to other links. Moreover, the segmentation based approach requires knowledge about SMS to accurately define the segments.

### 3.6 Conclusions

This chapter presented a simple transferrable method to model the relationship between TMS and SMS. It was found that a quadratic function is able to model the relationship between TMS and the variance of SMS, which is normally an unknown quantity. Comparing with other works in this area, the proposed model has better accuracy than Garber and Hoel (2001)'s linear relationship between TMS and SMS. It is also easier and has fewer limitations to implement than ANN and k-NN based methods from the aspect of traffic engineering.

Rakha and Zhang (2005)'s research is the most relevant in this area. Although it presents a new formulation to link TMS and SMS at second moments, it requires the knowledge of variance of TMS which is not available from typical ILD installations; e.g. the DATEX-II traffic data set. The proposed model in this chapter overcomes this limitation and only TMS value from ILDs is used to estimate SMS. It was also shown that differences in traffic state affect the parameters of this relationship. A location specific segmentation based approach was proposed to partition the data into four different traffic states. In addition, a generic two-step clustering method based on the k-means clustering technique was used to partition the data into congested and uncongested states. The quadratic relationship was calibrated for each partition separately. Based on the real-world data from English highways, it was shown that the SMS estimated using all the above models provide more accurate estimates of travel times compared to travel time estimation using TMS directly. The segmentation based models provided more accurate estimates of SMS compared to the original model. The original and clustering based models can be used for real-world ITS applications easily.

## **Chapter 4**

# **Kalman Filter and Its Applications in Travel Time Estimation**

A variety of data fusion techniques have been reviewed and discussed in Chapter 2. Compared to the Kalman filter, the other popular methods such as ANN and Bayesian inference are less attractive due to their own inherent drawbacks. The key of ANN fusion method is to approximate the relationship between the travel time and input data by a learning/training process. However, it is a redundant process since this relationship has been well modelled in a number of traffic flow theories. The purely data driven based approximation of ANN lacks robustness, especially when the data sources are highly unreliable. In addition, the learning process of ANN requires large amount of true historical travel time data, which makes ANN a less practical method in real world applications. The Bayesian based method also needs historical data to build up its hypothesis system which is crucial for the estimation process. Furthermore, even the historical data is available, it is still rather difficult to define multiple hypotheses and prior likelihood functions.

The Kalman filter technique becomes a promising and superior solution to the problems discussed above for two main reasons. Firstly, Kalman filter can make use of well-developed traffic flow model to describe the physical relation between traffic states and observed traffic variables, which eliminates the dependence on historical data. Secondly, the data from all available sensor sources is treated as an update/correction to the state estimate. It makes the estimation more reliable and robust because it combines the knowledge from traffic flow theory and information from multiple sensor sources in an optimal way. Therefore, Kalman

filter is chosen as the fundamental framework for the proposed travel time estimation fusion model. The objectives of this chapter are: 1) to provide a comprehensive introduction and discussion on the theory of Kalman filter including nonlinear version such as Extended Kalman filter (EKF) and Unscented Kalman filter (UKF); 2) to present an in-depth review of the existing travel time estimation methods based on Kalman filter; 3) discuss the gaps between the state-of-art applications and challenges in this field.

## **4.1 Introduction to the Basic Concept of Kalman Filter**

### **4.1.1 Historical background**

The Kalman filter was created by Rudolf E. Kalman in his famous paper Kalman (1960), although it has its roots as far back as Karl Gauss in 1795. This original paper described a recursive solution to the discrete data linear filtering problem. The similar methods had been proposed a little earlier than Kalman (1960) in Swerling (1958) and Stratonovich (1959 & 1960). Schmidt was the pioneer who implemented the Kalman filter into navigation and guidance systems for space vehicle systems, and his work was recognised as the first successful application of the Kalman filter and published in Schmidt (1981). Along with the development of sensor technology, the application of Kalman filter in the field of multi-sensor fusion becomes attractive to both of academia and industry. Willner et al. (1976) firstly examined several Kalman filter algorithms that can be used for state estimation within a multi-sensor system. Since then, the Kalman filter has been studied and applied in a number of diverse subjects such as process control, tracking and navigation, and had a dramatic impact on the data fusion related applications.

### **4.1.2 Conceptual basis of Kalman filter**

Theoretically the Kalman filter is an estimator for the problem of estimating the instantaneous state of a dynamic system by using measurements related to the system state. For a physical system such as the density of road traffic, the first attempt to describe the behaviours/states of the system is to develop a mathematical model which could represent the physical laws of the system. On the other hand, in order to observe the actual system state, some measurement technologies (sensors) are used to output observations regarding to certain variables of interest. Although both of the physical model and sensors could provide information about the system states, neither of them is perfect or adequate to depict the

varying of system state precisely. Firstly, no mathematical model is perfect, and there exists a certain level of errors within the modelling. In addition, the dynamic system is normally driven by controlled inputs which cannot be determined absolutely, thus some forms of uncertainty would be introduced into the model. Secondly, the measurement instruments or sensors do not provide perfect and complete data about the system. The observations from sensors are inevitably corrupted by noise to some extent, and it makes the variables that are observed from sensors are always different from its true value. Besides, the observation data from sensors does not generally provide all the desired information due to the inherent drawbacks or functional failure of the sensor devices.

The problem discussed above widely exists in the engineering practice especially in a multi-sensor system. In summary, there are two ways to approach the state of a dynamic system, one is through the mathematical model based on the physical law of the system, and the other one is through the observations which are provided by sensors. In Figure 4.1,  $\mathbf{x}_k$  is the underlying state of dynamic system which is also the variable to be estimated. Direct observations about this state variable are not available in the system, but the behaviour of this system evolving over time  $k, k+1, k+2 \dots$  is known as a mathematical model.  $\mathbf{y}_k$  is the observation from sensor output, and it may be the sensing result regarding to the state itself or the other variable related to the state. Neither of these two ways is perfect, and the problem becomes how to combine these two and obtain the best estimate. Kalman filter provides an optimal solution to this type of estimation problem.

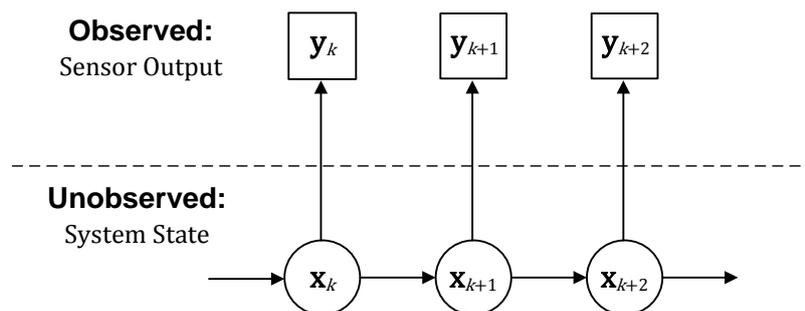


Figure 4.1 Illustration of dynamic system estimation

Maybeck (1979) gave a concise definition: *a Kalman filter is simply an optimal recursive data processing algorithm*. In his definition, the “filter” actually refers to *data processing algorithm*. The word recursive means that the estimation of the state is updated at each time step which is based on the previous estimate (one time step before) and the new input data. The key word in the definition is *optimal*. Although there are different aspects of defining *optimal* dependent on the different criteria for the evaluation, it has been shown that, under

the assumptions which will be introduced in next section, Kalman filter is optimal with respect to any reasonable criterion (Sorenson 1970, Maybeck 1979, Grewal & Andrews 2001). The most important aspect of optimality is that Kalman filter combines and associates all information which is available to be used in the estimation, i.e.

- the knowledge of the physical law of the dynamic system
- the observations from sensor devices and the knowledge of relationship between the observations and system state
- the statistical description of the system modelling error, observations noise and uncertainty in the dynamics system
- any available information about the initial conditions of the variables of interest

### **4.1.3 Assumptions**

The above section states that Kalman filter is an optimal estimator under some particular assumptions. There are three assumptions among the standard Kalman filter theory: 1) both of dynamic system and measurement are linear; 2) both of system and measurement noises are white; 3) both of system and measurement noises are Gaussian distributed. Each of the assumption is discussed as follows:

#### ***Assumption – 1: linear system***

The original development of Kalman filter started from linear system, since a linear system model is more easily manipulated with other engineering tools, and linear system theory is also much more complete and practical than nonlinear which is shown in Section 4.2. When the objective system is nonlinear, this assumption can be relaxed by a model linearisation process which forms the nonlinear version of Kalman filter: Extended Kalman filter (EKF) and Unscented Kalman filter (UKF). The details of EKF and UKF will be presented in Section 4.2.

#### ***Assumption – 2: white noise***

The noise terms of system modelling and measurement are assumed to be white noise. It implies that the noise value is not correlated over time. In other words, if a noise value is known at current time, it cannot be used to predict the noise value at any other time. Although perfect white noise does not exist in real world, this assumption is feasible to be approximated in practice. Any biased noise/error term can be obtained or approximated by off-line analysis or through other approaches. As long as the biased noise is removed, the

remaining noise term behaves as unbiased randomness, which is able to satisfy the assumption of whiteness.

### ***Assumption – 3: Gaussian noise***

The assumption of white noise concerns the time relationship of the noise. This assumption of Gaussian noise is related to the amplitude of the noise. It assumes that the probability density of noise amplitude follows the shape of normal curve (Gaussian/normal distribution). This assumption can be justified as two parts. Firstly, in a practical engineering application, the first and second order statistics (mean and variance) of a noise process will be known at best. In a situation when any higher order statistics is absent, there is no better type of distribution to assume than Gaussian density. Secondly, a system or measurement noise is normally caused by a number of small sources. When a number of independent random variables are added together, it can be demonstrated that this summation of independent random variables can be approximated closely by a Gaussian probability density, regardless of the forms of the individual densities. If the assumption of Gaussian noise is removed, Kalman filter can be shown as the best (minimum error variance) filter out of the class of unbiased filters (Trees 2001).

The particular three assumptions discussed above are made for the convenience of applying tractable mathematics which is used to developing Kalman filter theory. Maybeck (1979) indicated that it should not be confined by the assumption of linear system with white Gaussian noise when designing a Kalman filter based estimation system in real applications. Next section will briefly describe the derivation of Kalman filter based on above three assumptions. The proposed data fusion based Kalman filter estimation framework in this thesis takes these three assumptions into consideration. A justification analysis is provided in Chapter 5 to discuss the feasibility of applying Kalman filter into travel time estimation under these assumptions.

## **4.2 Linear Kalman Filter**

### **4.2.1 The process to be estimated**

Section 4.1.2 introduced the conceptual basis of the Kalman filter, which consists of a time-varying model describing the dynamic behaviour of the system and a measurement model describing the relationship between system state and observed variables. If the system state

which is to be estimated is defined as the notation  $\mathbf{x}_k \in \mathfrak{R}^n$  where  $k$  stands for the time step, and the state could be expressed as an  $n$  dimension vector:

$$\text{State vector:} \quad \mathbf{x}_k = [x_k^{(1)}, x_k^{(2)}, x_k^{(3)}, \dots, x_k^{(n)}]^T \quad (4.1)$$

where  $x_k^{(i)}$  is the  $i$ -th dimension state element. Equation 4.2 and 4.3 mathematically formulates the system model and measurement model state equation and observation equation:

$$\text{State equation:} \quad \mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \quad (4.2)$$

$$\text{Observation equation:} \quad \mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \quad (4.3)$$

where,

$\mathbf{x}_k$ :  $n \times 1$  system state vector at the time step  $k$ , and  $\mathbf{x}_k \in \mathfrak{R}^n$

$\mathbf{y}_k$ :  $m \times 1$  observation vector from sensor output at time step  $k$ , and  $\mathbf{y}_k \in \mathfrak{R}^m$

$\mathbf{u}_k$ :  $l \times 1$  optional control input for the system at time step  $k$ , and  $\mathbf{u}_k \in \mathfrak{R}^l$

$\mathbf{A}$ :  $n \times n$  matrix relates the state  $\mathbf{x}_{k-1}$  to the state  $\mathbf{x}_k$

$\mathbf{B}$ :  $n \times l$  matrix relates the optional control input  $\mathbf{u}_{k-1}$  to the state  $\mathbf{x}_k$

$\mathbf{H}$ :  $m \times n$  matrix relates the state  $\mathbf{x}_k$  to the measurement  $\mathbf{y}_k$

$\mathbf{w}_k$ :  $n \times 1$  system model noise

$\mathbf{v}_k$ :  $m \times 1$  observation noise

The dynamic state system is governed by a linear stochastic difference equation shown in equation 4.2, and the observation about the system is a linear function of the system state shown in equation 4.3. The transition matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{H}$  in both of equations are assumed to be constant here, however, in practice, they can change over time. Both of the noise terms  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are white and normal distributed with zero mean and known covariance matrices  $\mathbf{Q}_k$  and  $\mathbf{R}_k$ :

$$\begin{aligned} \mathbf{w}_k &\sim \mathcal{N}(0, \mathbf{Q}_k) \\ \mathbf{v}_k &\sim \mathcal{N}(0, \mathbf{R}_k) \\ \mathbb{E}[\mathbf{w}_k \mathbf{w}_j^T] &= \mathbf{Q}_k \delta_{k-j} \\ \mathbb{E}[\mathbf{v}_k \mathbf{v}_j^T] &= \mathbf{R}_k \delta_{k-j} \\ \mathbb{E}[\mathbf{w}_k \mathbf{v}_j^T] &= 0 \end{aligned} \quad (4.4)$$

where  $\delta_{k-j}$  is Kronecker delta function, i.e.

$$\delta_{k-j} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad (4.5)$$

The goal is to estimate the state  $\mathbf{x}_k$  based on the knowledge of the system dynamics described in state equation (equation 4.2) and the availability of noisy measurement  $\mathbf{y}_k$  described in observation equation (equation 4.3), as well as the covariance matrices  $\mathbf{Q}_k$  and  $\mathbf{R}_k$  of the noise terms  $\mathbf{w}_k$  and  $\mathbf{v}_k$ .

## 4.2.2 Recursive solution to Kalman filter

The details of deriving the solution of Kalman filter can be found in a number of text books such as Grewal & Andrews (2001), Haykin (2001) and Maybeck (1990). This section only provides the Kalman filter solutions to the estimation process described in 4.2.1. In order to form the Kalman filter solution, priori and posteriori state estimates at time step  $k$  are defined as:

$\hat{\mathbf{x}}_k^-$ : the priori estimate

$\hat{\mathbf{x}}_k$ : the posteriori estimate

Based on the above two definition about the state estimates, the state equation can be rewritten as a priori states predict format as well as priori states covariance:

**Predict:**

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} \quad (4.6)$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}_{k-1} \quad (4.7)$$

where  $\mathbf{P}_k$  is estimate error covariance, defined as:

$$\mathbf{P}_k = \mathbb{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T] \quad (4.8)$$

At each time step, the posteriori state estimate  $\hat{\mathbf{x}}_k$  is updated according to the weight difference (*Kalman Gain*, defined by  $\mathbf{K}$  in equation 4.9) between actual measurement  $\mathbf{y}_k$  and priori estimate:

**Correct:**

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R}_k)^{-1} \quad (4.9)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^-) \quad (4.10)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H})\mathbf{P}_k^- \quad (4.11)$$

where  $\mathbf{K}_k$  is the *Kalman Gain* at the time step  $k$ , and it is  $n \times m$  matrix that is chosen to minimise the posteriori error covariance  $\mathbf{P}_k$  defined in equation 4.8,  $\mathbf{I}$  is a diagonal matrix with same dimension as  $\mathbf{P}_k$ .

At each time step, the first task of Kalman filter can be thought of *predict*, which is to estimate the priori estimate  $\hat{\mathbf{x}}_k^-$  by using equation 4.6 and the priori states covariance  $\mathbf{P}_k^-$  by

using equation 4.7. The second task is considered as *correct*, which incorporates the incoming observations information into the priori estimate  $\hat{\mathbf{x}}_k^-$  to obtain an improved posteriori estimate  $\hat{\mathbf{x}}_k$ . The Kalman filter iterates the predict-correct task to approach the recursive solution for the estimation process. A complete picture of high level Kalman filter process is shown in Figure 4.2. Detailed operational sketch of Kalman filter is illustrated in Figure 4.3.

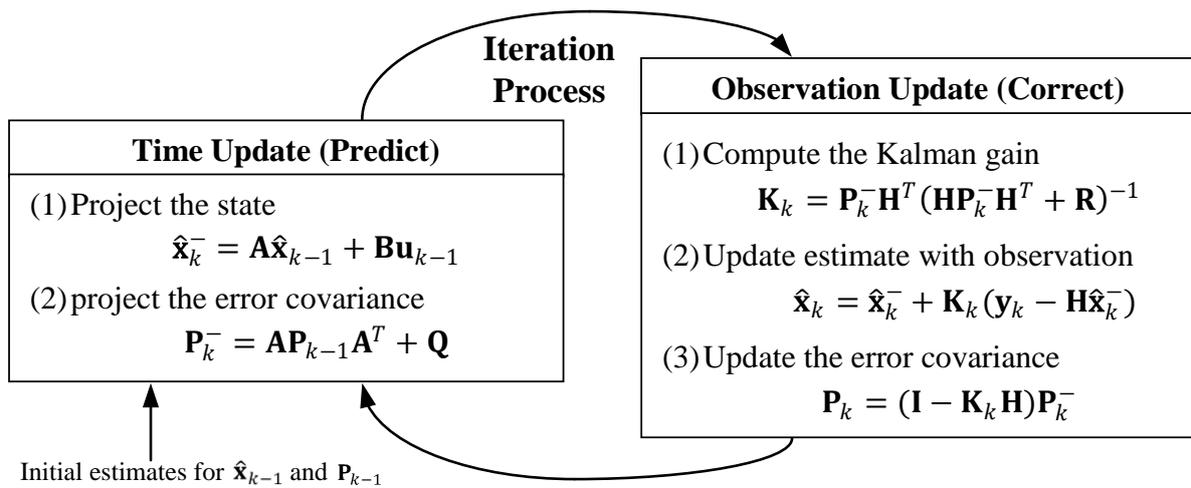


Figure 4.2 Summary of recursive solution to Kalman filter

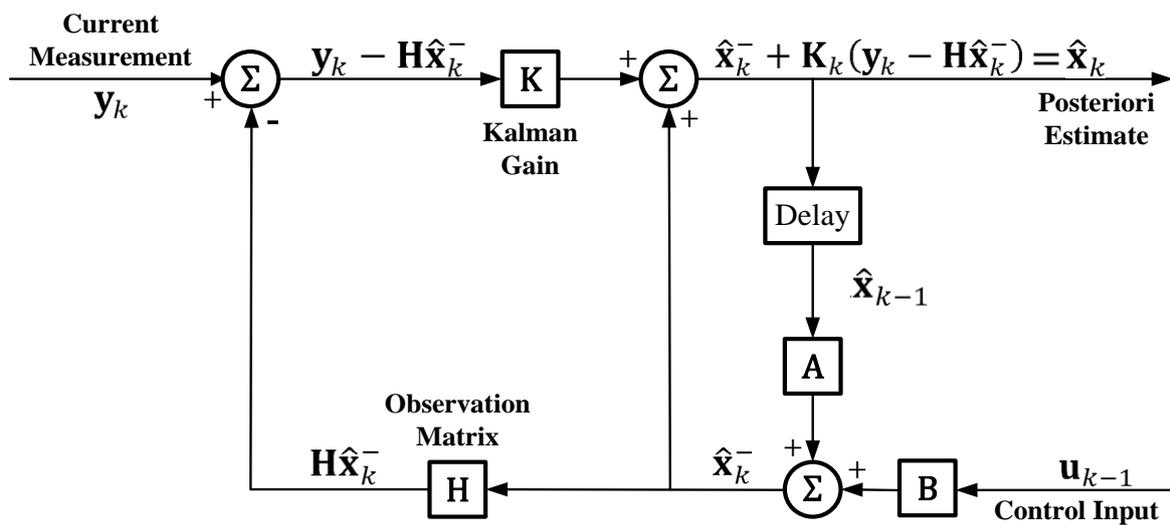


Figure 4.3 Sketch of Kalman filter operation

## 4.3 Nonlinear Kalman Filter

### 4.3.1 Nonlinear process to be estimated

As described in section 4.2.1, the state process addressed by the Kalman filter is governed by a *linear* stochastic difference equation. For this linear case, section 4.2.2 provides a general estimation solution. However, some of the dynamic systems are nonlinear, such as the relationship between traffic density and corresponding travel time. Similar as the state and observation equations introduced in section 4.2.1, the nonlinear process is described by the following equations:

$$\text{State equation:} \quad \mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (4.12)$$

$$\text{Observation equation:} \quad \mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k) \quad (4.13)$$

where the function  $f$  in the state equation (4.12) is nonlinear, and it relates the state  $\mathbf{x}$  at the previous time step  $k-1$  to the state at the current time step  $k$ . The nonlinear function also contains the control input  $\mathbf{u}$  and zero-mean process noise  $\mathbf{w}$  as its independent variables. The function  $h$  in the observation equation could be nonlinear as well, and it relates the state  $\mathbf{x}_k$  to the observation  $\mathbf{y}_k$ . The noise term  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are same as the definition in the linear version.

It is obvious that the solution of linear Kalman filter cannot be applied into this nonlinear version process, since the state transition matrix  $\mathbf{A}$  and observation matrix  $\mathbf{B}$  no longer exist. In order to make use of the recursive solution of the linear Kalman filter described in section 4.2.2, the nonlinear function  $f$  and  $h$  need to be linearised first. Extended and Unscented Kalman filter are the most popular approaches to the nonlinear Kalman filtering. The next two sections will introduce the algorithm of each one separately.

### 4.3.2 Extended Kalman filter (EKF) algorithm

Extended Kalman Filter (EKF), as the most straightforward extension of linear version, is optimally approximated via first order Taylor series expansion (linearisation) of the appropriate nonlinear functions. The assumptions for the EKF are the same as for the Kalman filter, except that the state and observation functions may be nonlinear. To solve the nonlinear filtering problem, the system is linearised at the estimated state for each time step. The linearisation process is based on a first-order Taylor series expansion, and uses the partial derivatives of the process and measurement functions to compute estimates. To form the nonlinear state difference and measurement relationships, both of state and observation

equations that linearise the estimation process about the equation 4.2 and 4.3 are given as follows:

$$\mathbf{x}_k \approx \tilde{\mathbf{x}}_k + \mathbf{A}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \mathbf{W}_{k-1}\mathbf{w}_{k-1} \quad (4.14)$$

$$\mathbf{y}_k \approx \tilde{\mathbf{y}}_k + \mathbf{H}_k(\mathbf{x}_k - \tilde{\mathbf{x}}_k) + \mathbf{V}_k\mathbf{v}_k \quad (4.15)$$

where,

- $\mathbf{x}_k$  and  $\mathbf{y}_k$  are the actual state and observation vectors
- $\tilde{\mathbf{x}}_k$  and  $\tilde{\mathbf{y}}_k$  are the approximate state and observation vectors from the Taylor expansion
- $\hat{\mathbf{x}}_k$  is the posteriori estimate of the state
- $\mathbf{w}_k$  and  $\mathbf{v}_k$  are the noise terms as in equation 4.2 and 4.3
- $\mathbf{A}_k$ ,  $\mathbf{W}_k$ ,  $\mathbf{H}_k$ , and  $\mathbf{V}_k$  are the Jacobian matrix of partial derivatives, and will be shown later

The main objective of this thesis is to apply Kalman filter as a tool to perform data fusion in travel time estimation. The details of EKF deriving are skipped, and the Table 4.1 summaries the process of EKF algorithm.

|  |   |        |
|--|---|--------|
| <b>Process:</b>  | $\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})$  | (4.16) |
|  | $\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k)$  | (4.17) |
| <b>Initialisation:</b>                                 | $\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0]$   | (4.18) |
|  | $\mathbf{P}_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$   | (4.19) |
| <i>For <math>k \in \{1, 2, \dots, \infty\}</math>:</i> |   |        |
| <b>Predict:</b>  | $\hat{\mathbf{x}}_k^- = f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \bar{\mathbf{w}})$  | (4.20) |
|  | $\mathbf{P}_k^- = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{W}_{k-1}\mathbf{Q}_{k-1}\mathbf{W}_{k-1}^T$  | (4.21) |
| <b>Correct:</b>  | $\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T)^{-1}$   | (4.22) |
|  | $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - h(\hat{\mathbf{x}}_k^-, \bar{\mathbf{v}}))$   | (4.23) |
|  | $\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-$  | (4.24) |
| <b>Where,</b>  | $\mathbf{A}_k \triangleq \left. \frac{\partial f(\mathbf{x}, \mathbf{u}_k, \bar{\mathbf{w}})}{\partial \mathbf{x}} \right _{\hat{\mathbf{x}}_k} \quad \mathbf{W}_k \triangleq \left. \frac{\partial f(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{w})}{\partial \mathbf{w}} \right _{\bar{\mathbf{w}}} \quad \mathbf{H}_k \triangleq \left. \frac{\partial h(\mathbf{x}, \bar{\mathbf{v}})}{\partial \mathbf{x}} \right _{\hat{\mathbf{x}}_k^-} \quad \mathbf{V}_k \triangleq \left. \frac{\partial h(\hat{\mathbf{x}}_k^-, \mathbf{v})}{\partial \mathbf{v}} \right _{\bar{\mathbf{v}}}$ |        |

Table 4.1 Summarised algorithm for EKF

### 4.3.3 Unscented Kalman filter (UKF) algorithm

The Unscented Kalman Filter (UKF) as an alternative nonlinear filter does not require a linearisation step as EKF. Instead of performing linearisation to analytically propagate the

statistics of state variable in EKF, UKF uses a deterministic sampling technique, called Unscented Transformation (UT), to represent the state distribution by a number of carefully chosen sample points (Julier et al. 1995, Wan & Merwe 2000). These points are typically called “sigma points”, and then propagated through the nonlinear transformation to obtain a set of transformed sigma points which can capture the true mean and covariance of the random variables. The posterior mean and covariance are approximated accurately with second order Taylor series expansion in contrast to the first order Taylor series of EKF. A more detailed discussion about the difference between EKF and UFK will be presented in the next section, and this section only provides an introduction on the UFK algorithm generally.

The fundamental part of UFK is the Unscented Transformation (UT), and it is a method for calculating the statistics of a random variable which is undertaken throughout a nonlinear transformation. Considering a  $L$ -dimension random variable  $\mathbf{x}$  with mean  $\bar{\mathbf{x}}$  and covariance  $\mathbf{P}_x$ , and it is propagated through a nonlinear function  $\mathbf{y} = f(\mathbf{x})$ . To calculate the statistics of  $\mathbf{y}$ , a matrix is formed by  $2L+1$  sigma vectors  $i$  according to the following formulas:

$$\mathbf{x}_0 = \bar{\mathbf{x}} \quad (4.25)$$

$$\mathbf{x}_i = \bar{\mathbf{x}} + (\sqrt{(L + \lambda)\mathbf{P}_x})_i \quad i = 1, 2, \dots, L \quad (4.26)$$

$$\mathbf{x}_i = \bar{\mathbf{x}} - (\sqrt{(L + \lambda)\mathbf{P}_x})_{i-L} \quad i = L + 1, L + 2, \dots, 2L \quad (4.27)$$

$$\lambda = \alpha^2(L + \kappa) - L \quad (4.28)$$

where,

$\alpha$ : a constant determines the spread of the sigma points around  $\bar{\mathbf{x}}$  and usually set to a small positive value, e.g.,  $1e - 4 \leq \alpha \leq 1$  (refer to Julier et al. 1995)

$\kappa$ : a constant scaling parameter which is usually set to 0 or  $3 - L$  (refer to Julier et al. 1995)

$(\sqrt{(L + \lambda)\mathbf{P}_x})_i$ : the  $i$ th column of the matrix square root

These sigma vectors are then propagated through the nonlinear function,

$$\Psi_i = f(\mathbf{x}_i) \quad i = 0, 1, \dots, 2L \quad (4.29)$$

and the mean and covariance of  $\mathbf{y}$  are approximated using a weighted sample mean and covariance of the posterior sigma points,

$$\bar{\mathbf{y}} \approx \sum_{i=0}^{2L} W_i^{(m)} \Psi_i \quad (4.30)$$

$$\mathbf{P}_y \approx \sum_{i=0}^{2L} W_i^{(c)} \{\Psi_i - \bar{\mathbf{y}}\} \{\Psi_i - \bar{\mathbf{y}}\}^T \quad (4.31)$$

with weights  $W_i$  given by

$$W_0^{(m)} = \lambda / (L + \lambda)$$

$$W_0^{(c)} = \frac{\lambda}{L + \lambda} + (1 - \alpha^2 + \beta)$$

$$W_i^{(m)} = W_i^{(c)} = \frac{1}{\{2(L+\lambda)\}} \quad i = 1, 2, \dots, 2L \quad (4.32)$$

The UKF algorithm is a straightforward application of the UT to the recursive estimation process as the standard Kalman filter in section 4.2.2. The state variable is refined to the concatenation of the original state and the noise variables:  $\mathbf{x}_k^a = [\mathbf{x}_k, \mathbf{w}_k, \mathbf{v}_k]$ , and the UT is applied to this new formed state variables  $\mathbf{x}_k^a$  to calculate the corresponding sigma matrix  $\mathbf{P}_k^a$ . The complete equations for the algorithm are summarised in Table 4.2.

**State-observation model:**

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (4.33)$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k) \quad (4.34)$$

**Initialisation:**

$$\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0] \quad (4.35)$$

$$\mathbf{P}_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T] \quad (4.36)$$

$$\hat{\mathbf{x}}_0^a = \mathbb{E}[\mathbf{x}_k^a] = [\hat{\mathbf{x}}_0^T, \mathbf{0}, \mathbf{0}]^T \quad (4.37)$$

$$\mathbf{P}_0^a = \mathbb{E}[(\mathbf{x}_0^a - \hat{\mathbf{x}}_0^a)(\mathbf{x}_0^a - \hat{\mathbf{x}}_0^a)^T] = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \quad (4.38)$$

For  $k \in \{1, 2, \dots, \infty\}$ :

**Sigma points Calculation:**

$$\mathbf{x}_{k-1}^a = \left[ \hat{\mathbf{x}}_{k-1}^a, \hat{\mathbf{x}}_{k-1}^a + \gamma \sqrt{\mathbf{P}_{k-1}^a}, \hat{\mathbf{x}}_{k-1}^a - \gamma \sqrt{\mathbf{P}_{k-1}^a} \right] \quad (4.39)$$

**Predict:**

$$\mathbf{x}_{k|k-1}^x = f(\mathbf{x}_{k-1}^x, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}^w) \quad (4.40)$$

$$\hat{\mathbf{x}}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathbf{x}_{i,k|k-1}^x \quad (4.41)$$

$$\mathbf{P}_k^- = \sum_{i=0}^{2L} W_i^{(c)} [\mathbf{x}_{i,k|k-1}^x - \hat{\mathbf{x}}_k^-][\mathbf{x}_{i,k|k-1}^x - \hat{\mathbf{x}}_k^-]^T \quad (4.42)$$

$$\mathbf{\Psi}_{k|k-1} = h(\mathbf{x}_{k|k-1}^x, \mathbf{v}_{k-1}^v) \quad (4.43)$$

$$\hat{\mathbf{y}}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathbf{\Psi}_{i,k|k-1} \quad (4.44)$$

**Correct:**

$$\mathbf{P}_{y_k y_k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathbf{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k^-][\mathbf{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k^-]^T \quad (4.45)$$

$$\mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathbf{x}_{i,k|k-1}^x - \hat{\mathbf{x}}_k^-][\mathbf{\Psi}_{i,k|k-1} - \hat{\mathbf{y}}_k^-]^T \quad (4.46)$$

$$\mathbf{K}_k = \mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} \mathbf{P}_{y_k y_k}^{-1} \quad (4.47)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \hat{\mathbf{y}}_k^-) \quad (4.48)$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{P}_{y_k y_k} \mathbf{K}_k^T \quad (4.49)$$

**Where,**  $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{w}^T, \mathbf{v}^T]$ ,  $\mathbf{a} = [(x)^T, (w)^T, (v)^T]$ ,  $\gamma = \sqrt{(L + \lambda)}$ ,  $L$  is the dimension of the state variables,  $\lambda$  is defined in equation 4.28, and  $W_i$  is the weights defined in equation 4.32

Table 4.2 Summarised algorithm for EKF

#### 4.3.4 Comparison discussion between EKF and UKF

The classic Kalman filter (linear) was developed for application to the estimation of linear processes. However, in practice, many real estimation problems are based on nonlinear processes such as the travel time estimation in the traffic domain. To relax the assumption of linearity in the original Kalman filter theory, EKF and UKF are developed to deal with nonlinear system within the classic Kalman filter estimation framework. The main difference between them is the techniques used in the linearisation step. As described in the above section, EKF uses partial derivatives of the process and measurement functions to obtain Jacobians matrix to linearise the process, while UKF uses a deterministic sampling technique, Unscented Transformation (UT), to capture the mean and covariance estimates with a minimal set of sample points. Due to the difference of nature of linearisation steps, UKF is superior to EKF in several aspects which are discussed below.

As the description in the Kalman filter theory, the propagation of mean and covariance of the (conditional) state pdf is directly related to the estimate accuracy. For EKF, the pdf is propagated through a linear approximation of the system around the operating point at each time instant. By the linearisation step in EKF operation, all pdf are Gaussian, while the real pdf passing through the nonlinear system is obviously non-Gaussian. The consequence of these approximations is that convergence cannot be guaranteed, or more specifically, this linearisation can lead to filter instability if the time intervals are not sufficiently small (Julier et al. 1995). In contrast to EKF, UKF does not use a linearization of the system and the noises are not assumed to be Gaussian. For UKF, the mean and the covariance of the (conditional) state pdf is represented by so-called sigma points which are defined with appropriate weights attached to each point. The sigma points and the weights are carefully chosen such that the weighted mean and covariance approximate the true mean and covariance of the pdf. Julier et al. (1995) & (2000) and Wan & Merwe (2000) have shown the better accuracy of UKF than EKF theoretically. An illustrated example of the difference of mean and covariance propagation between UKF and EKF is shown in Figure 4.4.

From the view of Taylor series expansion, the posterior mean and covariance of UKF are approximated accurately with second order (Taylor series expansion) for any nonlinearity, whereas EKF only provides first order approximation to the nonlinear system. It means that UKF can retain more information through the linearisation process than EKF. The first order approximation of EKF could introduce large errors in the true posterior mean and covariance of the transformed state, which may lead to divergence of the filter estimation.

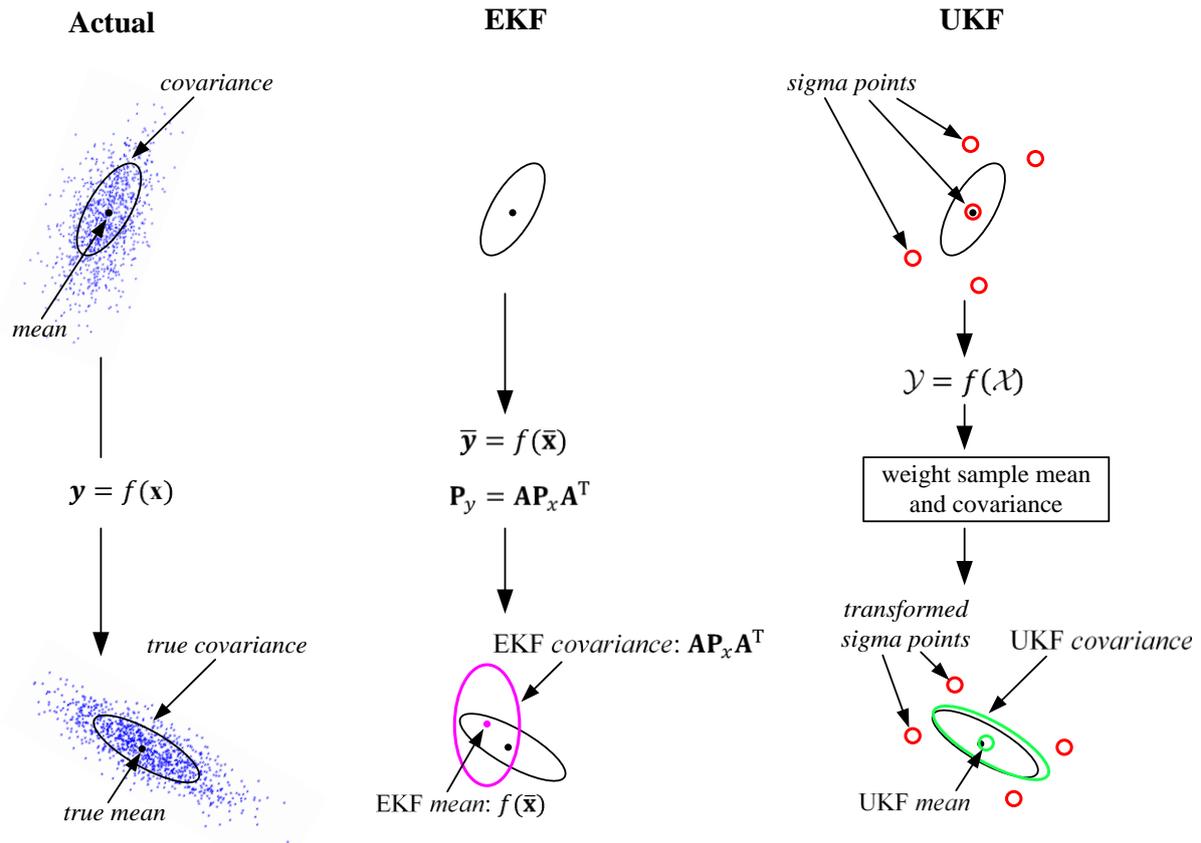


Figure 4.4 Comparison of mean and covariance propagation between UKF and EKF (adapted from Wan & Merwe 2000)

Another drawback of the EKF linearisation is that it required the explicit Jacobian calculations based on the partial derivative of the nonlinear function. The derivations of Jacobian matrices are nontrivial in most realistic applications, and it usually has significant difficulties of implementation. Unlike EKF, UKF is not necessary to calculate Jacobian matrices, thus the algorithm is easy to implement in practice. In addition, the computation complexity of UKF is same order as EKF. A summarised comparison between EKF and UKF is listed in the Table 4.3.

| Filter Type | Linearisation Method     | Assumption                     | Optimal Type   | Accuracy     | Required Calculation |
|-------------|--------------------------|--------------------------------|----------------|--------------|----------------------|
| EKF         | Partial derivatives      | Gaussian noise                 | Local optimal  | First order  | Jacobian matrices    |
| UKF         | Unscented transformation | Not necessarily Gaussian noise | Global optimal | Second order | Sigma Points         |

Table 4.3 Summary of difference between EKF and UKF

## **4.4 Review of Kalman Filter in the Traffic Estimation**

### **4.4.1 Introduction to the Kalman filter applications in the traffic estimation**

By the mid 1960s, Kalman filter based estimation methodology has been widely applied into a number of areas such as tracking, system control and navigation. Along with the development of sensor technology, the advantages of Kalman filter make it become one of the most successful data fusion methods in multi-sensor environment. The Kalman filter entered the view of transport research from the late 1990s (Dailey 1999), and started to be used in the travel time/traffic flow estimation. In recent years, the Kalman filter has attracted more and more attention in the application in the traffic estimation, since the data from heterogeneous traffic sensors is generally available.

An overview of travel time estimation has been provided in Chapter 2. It discussed that the Kalman filter, as one of the physical model based estimation approaches, is a promising solution to the problems in the field of travel time estimation, especially in a multi-sensor environment. From the descriptions of the Kalman filter algorithms above, it can be seen that Kalman filter as an optimal estimation technique has a very general but practical framework for the implementation in the field of travel time estimation. Firstly, Kalman filter can make use of well-developed traffic flow model to describe the physical relation between traffic states and observed traffic variables, which eliminates the dependence on historical data. Secondly, the data from all available sensor sources is treated as an update/correction to the state estimate. It makes the estimation more reliable and robust because it combines the knowledge from traffic flow theory and information from multiple sensor sources in an optimal way.

This section provides an in-depth review of the Kalman filter applications into traffic estimation. According to the difference of sensor availability and traffic model employed, the literature is categorised into three parts: linear Kalman filter for single sensor source, linear Kalman filter for multiple sensor sources and nonlinear Kalman filter for single sensor source. To the best of author's knowledge, the nonlinear Kalman filter for multiple sensor sources has not been studied, and the work in this field is one of the contributions of this research.

## 4.4.2 Review of the applications based on single sensor source

### *Dailey (1999)'s Approach*

Dailey presented a Kalman filter based algorithm to estimate the time-mean-speed by using single loop detector data. The basic filter model is expressed as:

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1} \quad (4.50)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad (4.51)$$

where,

$$\mathbf{x}_k = \begin{bmatrix} \bar{s}_k \\ \bar{s}_{k-2} \end{bmatrix}, \quad \mathbf{A}_k = \begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}, \quad \mathbf{y}_k = \begin{bmatrix} \frac{O_k}{N_k} \\ \frac{O_{k-1}}{N_{k-1}} \end{bmatrix}, \quad \mathbf{H}_k = \begin{bmatrix} -\frac{3\bar{l}}{T} \left[ \frac{\bar{s}_{k-1}^2 + \sigma_s^2}{\bar{s}_{k-1}^4} \right] & 0 \\ 0 & -\frac{3\bar{l}}{T} \left[ \frac{\bar{s}_{k-2}^2 + \sigma_s^2}{\bar{s}_{k-2}^4} \right] \end{bmatrix}$$

$\bar{s}_k$  is the TMS measured by reliable devices (in this paper, the dual-loop detectors).  $\mathbf{A}_k$  is the state transition matrix,  $a$  and  $b$  in  $\mathbf{A}$  are constant which are selected using forward/backward least squares estimates of the AR (Auto Regression) coefficients for the experimentally measured speed.  $T$  is the sampling period,  $N_i$  is the volume in the  $i$ th sampling interval,  $O_i$  is the lane occupancy in the  $i$ th sampling interval. The noise contributions are:

$$\mathbf{Q} = \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \sigma_{O/N}^2 & 0 \\ 0 & \sigma_{O/N}^2 \end{bmatrix}$$

where  $\sigma_s^2$  is the variance of  $\bar{s}$ , and  $\sigma_{O/N}^2$  is the variance of  $O/N$ .

In this Kalman filter modelling, the state variable  $\mathbf{x}_k$  (in equation 4.50) represents the estimate of TMS. The novelty of this approach is to approximate a relationship (shows as observation transition matrix  $\mathbf{H}_k$  in equation 4.51) b empirical value between  $\mathbf{x}_k$  (estimate of TMS) and  $\mathbf{y}_k$  (the ratio of occupancy and volume from ILDs) based on the following equation:

$$O_i = \frac{1}{T} \sum_{j=1}^{N_i} \frac{l_{ij}}{s_{ij}} \quad (4.52)$$

where  $s_{ij}$  is the  $j$ th vehicle's instant speed in the  $i$ th sampling interval,  $l_{ij}$  is the  $j$ th vehicle's effective vehicle length (EVL) in the  $i$ th sampling interval. The detailed derivative from Equation (4.50) to  $\mathbf{H}_k$  can refer to the paper of Dailey (1999). Although the objective of this research is to estimate TMS rather than SMS, the pioneering approach shows a methodological example of applying Kalman filter into traffic estimation for the subsequent studies.

### ***Chien et al. (2003)'s Approach***

Chien et al. proposed a simple method to apply Kalman filter to estimate and predict travel time. The context of their application is for freeway applications and based on the travel time observation from probe vehicles. Their work first appeared in Chen & Chien (2001), and follow-up work was presented in Chien et al. (2003). The basic methodologies applied are exactly same in both papers. The estimate objective of their approach is the travel time which is represented as the state variables in the state difference equation as:

$$x_k = \phi_{k-1}x_{k-1} + w_{k-1} \quad (4.53)$$

where  $x_k$  denotes the travel time which is to be estimated,  $\phi_k$  denotes the state transition parameter, and  $w_k$  denotes the noise term which has a normal distribution as:

$$w \sim \mathcal{N}(0, Q) \quad (4.54)$$

The observation equation associated with state variable  $x_k$  is given by

$$y_k = x_k + v_k \quad (4.55)$$

where  $y_k$  denotes the observation of travel time and  $v_k$  denotes the measurement error which also has a normal distribution as:

$$v \sim \mathcal{N}(0, R) \quad (4.56)$$

The advantage of this approach is that it shows a rather simple and straightforward example of Kalman filter application. Compared to Dailey (1999)'s approach, it does not require further approximation on relating observation with estimated state (equation 4.55), because both of observation and state estimate are configured as travel time. Besides, the state variable (travel time) defined in both of the state and observation models is a scalar variable rather than a vector, thus, the covariance matrix  $\mathbf{Q}$  and  $\mathbf{R}$  become variance  $Q$  and  $R$ . Therefore, it reduced the computational complexity and easy to implement in practice.

### ***Qiu & Ran (2007)'s Approach***

Qiu & Ran applied Kalman filter to estimate travel time (or SMS equivalently) in freeway network. The data source they used is cellular phone data which contains the time and location information. The use of cellular phone data is similar to the application of GPS devices. Both of them is to record the time and location data of one vehicle and calculate the average travel speed between two locations. The initiative of their research is that current network-based cellular probe traffic monitoring system has much lower cost than the traditional GPS probe method, and the accuracy of mobile localization technology has been remarkably improved (Zhao, 2000).

The basic Kalman filter configuration applied in this work is similar to Chien et al. (2003)'s approach. The state difference and observation equations are shown as follows:

$$x_{j,k} = g_{j,k}^1 x_{j-1,k-1} + g_{j,k}^2 x_{j,k-1} + g_{j,k}^3 x_{j+1,k-1} + w_{j,k-1} \quad (4.57)$$

$$y_{j,k} = x_{j,k} + v_{j,k} \quad (4.58)$$

Where,  $x_{j,k}$  represents the travel speed for the traffic flow in link  $j$  during the  $k$ th time interval,  $j+1$  and  $j-1$  represents the upstream and downstream links which are adjacent to link  $j$ , and  $g_{j,k}^*$  (“\*” may be one of {1, 2, 3}) denotes the transition parameters which associate the speed state of last time interval ( $k-1$ ) at upstream, current or downstream link ( $j-1$ ,  $j$ , or  $j+1$ ) with current speed state at link  $j$ .  $y_{j,k}$  is the observed variable from cellular phone measurement. The noise terms  $w_{j,k-1}$  and  $v_{j,k}$  are also normal distribution as equation 4.45 and 4.46. In their practical implementation, Equation (4.57) is simplified as:

$$x_{j,k} = g_{j,k}^2 x_{j,k-1} + w_{j,k-1} \quad (4.59)$$

It indicates that the impact of speed states of upstream and downstream has been eliminated which means they assume that the traffic is in a stable pattern along the links. Although the paper did not mention how the state transition parameter  $g_{j,k}^*$  is calculated, it is assumed that the parameter is approximated and calibrated by historical dataset from other measurement equipments (e.g. ILDs).

#### 4.4.3 Review of the applications based on multiple sensor sources

Last section reviewed three of most representative applications of Kalman filter based on the single sensor source. Although their results show a certain level of effectiveness, due to the limitation of single sensor observations (discussed in Chapter 2), research in this field has switched to investigate TTE problem based on multiple sensor sources, which is also the aim and initiative of this research.

##### *Nanthawichit et al (2003)'s Approach*

Nanthawichit et al (2003) proposed a method for integrating probe vehicle data into fixed detector data to estimate traffic states on a freeway. Their approach selected Payne (1971)'s macroscopic traffic flow model as state difference equation. Payne's macroscopic model (shown as equation 4.60-4.62) is rather suitable for Kalman filter application because it formulates the varying of traffic density as time series which has similar form as the state difference equation in Kalman filter.

$$\rho_{j,k} = \rho_{j,k-1} + \frac{\Delta t}{\Delta L_j} (q_{j-1,k-1} - q_{j,k-1} + r_{j,k-1} - s_{j,k-1}) \quad (4.60)$$

$$v_{j,k} = v_{j,k-1} + \frac{\Delta t}{\tau} \{v_e[\rho_{j,k-1}] - v_{j,k-1}\} + \frac{\Delta t}{\Delta L_j} v_{j,k-1} [v_{j-1,k-1} - v_{j,k-1}] - \frac{\eta \Delta t}{\tau \Delta L_j} \cdot \frac{\rho_{j+1,k-1} \cdot \frac{\lambda_j}{\lambda_{j+1}} - \rho_{j,k-1}}{\rho_{j,k-1} + \kappa} \quad (4.61)$$

$$q_{j,k} = \alpha [v_{j,k} \rho_{j,k}] + (1 - \alpha) [v_{j+1,k} \rho_{j+1,k}] \quad (4.62)$$

The macroscopic traffic variables in the model were defined as follows:

$\rho_{j,k}$ : density of segment  $j$  at time  $k$

$v_{j,k}$ : space mean speed of segment  $j$  at time  $k$

$q_{j,k}$ : flow rate at the boundary point between segments  $j$  and  $j+1$  at time  $k$

$r_{j,k}$ : entry flow rate at ramp of segment  $j$  at time  $k$

$s_{j,k}$ : exit flow rate at ramp of segment  $j$  at time  $k$

$\Delta t$  is the time increment;  $\lambda_j$  is the number of lanes in segment  $j$ ;  $v_e(\cdot)$  is the speed at equilibrium state, which can be obtained from the density-speed relationship; and  $\tau$ ,  $\eta$ , and  $\kappa$  are model parameters.  $\alpha$  is a weighting parameter ranging from 0 to 1.

When applied this macroscopic traffic flow model into Kalman filter, traffic density ( $\rho$ ) and space mean speed ( $v$ ) are treated as state variables:  $\mathbf{x}_k = (\boldsymbol{\rho}_k, \mathbf{v}_k)$ , whereas traffic flow ( $q$ ) spot speed ( $w$ ) are treated as the observation variables:  $\mathbf{y}_k = (\mathbf{q}_k)$ , Both of Equation (4.60) and equation (4.61) are treated as state equations, while observation equation consists of a relationship between traffic flow and state variables, as in equation 4.62. The solution of this filter configuration is same as the approach in Chien et al. (2003). The fusion method in this approach is simply from a weighted average of speed from two data sources. The weighting factor of 0.5 is assigned to both sources of data and then obtaining the observations. Although this fusion method does not consider the reliability of data from each source and the dynamic varying of traffic condition in real time, the results presented in the paper still show a certain level of improvement when comparing with the estimate results only based on single sensor source.

### ***Chu et al (2005)'s Approach***

Chu et al (2005) uses the standard linear Kalman filter framework to fuse fixed loop detector data and probe vehicle data for the travel time estimation for the freeway. The model

proposed in their application is based on the conservation equation in traffic flow which was first proposed by Lighthill & Whitham (1955) and Richards (1957), i.e.,

$$\frac{\partial q}{\partial l} + \frac{\partial x}{\partial t} = 0 \quad (4.63)$$

where  $q$  is flow (vehicles/hour),  $x$  is density (vehicles/km),  $l$  is location, and  $t$  is time. If the speed of such traffic fluids is  $v$ , the following basic identity is obtained:

$$q = x \cdot v \quad (4.64)$$

Similar to the approach in Nanthawichit et al (2003), the traffic density is selected as the state variable in the state equation of Kalman filter and represented as time series:

$$x_k = x_{k-1} + \frac{1}{\lambda \cdot L} \{ [q_{u,k} + q_{on,k}] - [q_{d,k} + q_{off,k}] \} \quad (4.65)$$

Where  $\lambda$  is the number of lanes on the mainline freeway,  $L$  is length of the section between upstream and downstream detectors.  $q_{u,k}$  and  $q_{d,k}$  are traffic flows of the upstream and downstream boundaries at time interval  $k$ ,  $q_{on,k}$  and  $q_{off,k}$  are total on-ramp and off-ramp traffic flows at time interval  $k$ . They assumed that the traffic inside of the section is homogeneous, an estimation of the section travel time is:

$$TT_k = \frac{L}{v_k} = \frac{L}{q_k} \cdot x_k \quad (4.66)$$

where  $TT_k$  is the travel time which is treated as the observations estimated from GPS based probe vehicles,  $q_k$  as the traffic flow measurement can be obtained from output of ILDs directly. Therefore, the state and observation equations of Kalman filter are formulated as follows:

$$x_k = x_{k-1} + u_k + w_{k-1} \quad (4.67)$$

$$TT_k = H_k x_k + v_k \quad (4.68)$$

where  $u_k$  and  $H_k$  are estimated by:

$$u_k = \frac{1}{\lambda \cdot L} \{ [q_{u,k} + q_{on,k}] - [q_{d,k} + q_{off,k}] \} \quad (4.69)$$

$$H_k = \frac{L}{\alpha [q_{u,k} + q_{on,k}] + (1 - \alpha) [q_{d,k} + q_{off,k}]} \quad (4.70)$$

The fusion process in this approach is to treat the data (travel time) from GPS based probe vehicle as the observation variables and treat the data (flow count) from ILDs as parameter for state input ( $u_k$ ) and observation transition coefficient ( $H_k$ ) calculation. Compared to the approach in Nanthawichit et al (2003), this method fused two different outputs from each sensor source (flow from ILDs and travel time from GPS based probe vehicles) rather than

simply weighting two travel time or speed measurements. The gain of this approach is that it can take advantage of most reliable data of each sensor source, e.g. flow from ILDs and travel time from GPS (comparing with travel time estimation from fixed location sensors). In addition, according to the description in Kalman filter algorithm (Section 4.2.2), the state and observation noise terms are crucial to the performance of estimate result. This approach approximates the state and observation noise based on the measurement difference of multiple data sources, which is also benefit from the data fusion aspect.

### 4.4.3 Review of the applications based on nonlinear Kalman filter

#### *Gazis & Liu (2003)'s Approach Based on EKF*

In Gazis & Liu (2003)'s approach, EKF is applied for estimating vehicle counts for two roadway sections in tandem. The basic formulation of the Kalman filter is the following. The state variables ( $x$ ) for a multi-section roadway are chosen as the numbers of vehicles in each section. The state difference equation for the state variables at the  $k$ th time step are given by

$$x_{j,k} = x_{j,k-1} + u_{j,k-1} - u_{j+1,k-1} + w_{j,k-1} - w_{j+1,k-1} \quad (4.71)$$

where the quantities  $u_{j,k}$  and  $u_{j+1,k}$  are the numbers of vehicles entering and leaving section  $j$ , at the  $k$ th time step, respectively. The quantity  $w_{j,k}$  represents the counting error for the quantity  $u_{j,k}$ . The observations are determined by the phenomenological relationship between speed and density:

$$y_{j,k} = y_{j,f} \exp\left[-(x_{j,k}/n_{j,0}L_j)^2/2\right] + v_{j,k} \quad (4.72)$$

where the subscript  $j$  labels the sections, and the velocity  $y_{j,k}$  is the observation input for EKF using the real time speed data. The parameters  $y_{j,f}$ ,  $L_j$ ,  $v_{j,k}$  represent the free flow speed, the section length, and the error associated with the speed estimation, respectively. The quantity  $n_{j,0}$  is the density corresponding to the maximum flow in the  $j$ th section. It is obvious that the observation equation (equation 4.72) associates the observation variable ( $y_{j,k}$ ) with state variable ( $x_{j,k}$ ) by a nonlinear formulation. To solve this problem, EKF is applied here to linearise equation 4.72. The detailed process is exactly same as the standards equations in Table 4.1. Besides developing a model to apply EKF technique in traffic estimate problem in this approach, this research also demonstrate an estimation improvement by coupling the detector counts with independent density estimates, subject to uncorrelated errors. It is because observation errors at the joint of the two sections are shared by both sections.

### ***Ye et al. (2006)'s Approach Based on UKF***

Ye et al. (2006) presented a new method of speed estimation for freeway by using an UKF method. The data input is based on occupancy and count observations from single ILDs. This paper first discussed the drawback of the application of EKF in traffic estimation. They took the Dailey (1999)'s approach as the example, and described that linearization in the EKF will produce highly unstable filters if assumptions are not met, and the derivation of the Jacobian matrices often lead to significant implementation difficulties. To overcome this problem, they applied UKF estimate speed without the linearization steps required by the EKF. The basic filter configuration in this approach is same as Dailey (1999)'s approach. To use the UKF solution equations in Table 4.2, the sigma points are created by:

$$\left. \begin{aligned} \sigma_0 &= \bar{\mathbf{x}} \\ \sigma_i &= \bar{\mathbf{x}} + (\sqrt{(n + \kappa)\mathbf{P}_x})_i \\ \sigma_{i+n} &= \bar{\mathbf{x}} - (\sqrt{(n + \kappa)\mathbf{P}_x})_i \end{aligned} \right\} \quad (4.73)$$

$$\left. \begin{aligned} W_0 &= \kappa/(n + \kappa) \\ W_i &= 1/[2(n + \kappa)] \\ W_{i+n} &= 1/[2(n + \kappa)] \end{aligned} \right\} \quad (4.74)$$

The predict and update process is same as the standard solution. The only difference is that  $\sigma_0$  is set to  $\bar{\mathbf{x}}$ , the average of previous state, which is different from using AR to find appropriate coefficients in Dailey (1999), shown as follows:

$$\sigma_0 = \bar{\mathbf{x}} = \frac{\hat{\mathbf{x}}_{k-1} + \hat{\mathbf{x}}_{k-2}}{2} \quad (4.75)$$

In the implementation of the UKF, they treated speed variance as constant, and value of standard deviation of speed is experimentally determined and used. In addition to speed variance (state noise), the current observation noise  $\sigma_{O/N,k}^2$  is determined recursively by the variance of observation based on last noise value  $\sigma_{O/N,k-1}^2$ .

## **4.5 Discussion & Summary**

Sections 4.1 to 4.3 introduced the basic concept and algorithms of standard Kalman filter, EKF and UKF. The related applications of in the field of traffic estimation are provided in section 4.4. Table 4.4 summarises these applications based on different approaches. From the literature review in this field, three issues need to be addressed: 1) how to choose the appropriate traffic model for the Kalman filter equations; 2) how to determine the process and measurement noise/errors; 3) how to choose the types of Kalman filter.

As described above, one of the advantages of Kalman filter is that it can make use of the internal physical relation between the observed variables and objective estimates. According to the review, it is found that different approaches use different traffic flow models to associate the observed variables with the objective travel time such as the model in Dailey (1999), macroscopic flow model in Nanthawichit et al (2003) and Lighthill & Whitham (1955) conservation equation in Chu et al (2005). The degree of precision of the adopted traffic model will affect the overall estimation performance for Kalman filter application. However the traffic model chosen in those applications is limited by the observations obtained from sensors, which means the precision of the applied traffic model is compromised for the available observations. The data fusion based Kalman filter application have richer available observations, which provides more options to choose more appropriate traffic model to interpret the relationship between the objective travel time and all types of available sensor observations.

The second problem in implementing Kalman filter is that it assumes that the noise covariance of observations and process are known which is not realistic in our applications. Although the reviewed researches provide several methods to approximate or assume the noise covariance, an improvement could still be achieved by taking advantages of multi-sensor data fusion. For example, the data from GPS probe vehicles provide more direct and accurate travel time observations for certain road segments. This type of observations could be used to approximate the noise covariance such as the approach in Chu et al (2005).

From the review of the nonlinear Kalman filters (EKF and UKF), it can be learned that either EKF or UKF can overcome the limitation on the linearity of standard Kalman filter. Hence, the filter design is not necessary to be constrained by the choosing linear traffic model and more advanced nonlinear traffic model can be implemented directly. According to the review and comparison discussion in Section 4.3.4, UKF is more suitable than EKF for the traffic estimation problem in the following ways: 1) higher accuracy than EKF; 2) better convergence properties by dual filtering; 3) more stable than Extended Kalman filter; 4) easy to implement.

In summary, the Kalman filter is one of the most suitable data fusion techniques for the problem of travel time estimation in multi-sensor environment. Its application in this field has been studied and verified in a number of research works. However, only linear Kalman filter has been applied into the multi-sensor based travel times estimation, and the research on the nonlinear Kalman filter only focus on the estimation from single sensor sources. One of main

research objective is to fill this gap, and to develop a nonlinear Kalman filter based travel time estimation framework, which is able to take the advantage of both of more advanced traffic flow theory and the availability of data from multi-sensor. The details of the proposed Kalman filter are introduced in the next chapter.

| Approach                         | State Variable  | Observation Variable          | Traffic Related Model                                      | Filter Type | Noise Estimation  | Data Source                                       |
|----------------------------------|-----------------|-------------------------------|--|-------------|---|---|
| <i>Dailey (1999)</i>             | TMS             | ratio of occupancy and volume | $O_i = \frac{1}{T} \sum_{j=1}^{N_i} \frac{l_{ij}}{s_{ij}}$ | linear      | empirical value and off-line extraction                               | occupancy and count observations from single ILDs |
| <i>Chien et al.(2003)</i>        | travel time     | travel time                   | none   | linear      | historical data   | GPS based probe vehicle data                      |
| <i>Qiu &amp; Ran (2007)</i>      | SMS             | SMS                           | none   | linear      | historical data   | cellular phone based probe vehicle data           |
| <i>Nanthawichit et al (2003)</i> | density and SMS | flow                          | Payne (1971)'s macroscopic traffic flow model              | linear      | not mention   | fusion based on probe data and fixed loop data    |
| <i>Chu et al (2005)</i>          | density         | travel time                   | traffic flow conservation equation                         | linear      | based on the difference of multiple data sources                      | fusion based on probe data and fixed loop data    |
| <i>Gazis &amp; Liu (2003)</i>    | vehicle count   | flow speed                    | $y_i = y_f \exp[-(x_i/n_{i,0}L_i)^2/2]$                    | extended    | based on the uncorrelated errors of state and observation variables   | spot speed and flow data from ILDs                |
| <i>Ye et al. (2006)</i>          | TMS             | ratio of occupancy and volume | $O_i = \frac{1}{T} \sum_{j=1}^{N_i} \frac{l_{ij}}{s_{ij}}$ | unscented   | observation noise is determined recursively based on last noise value | occupancy and count observations from single ILDs |

Table 4.4 Summary of the applications based on Kalman filter

## **Chapter 5**

# **A Fusion Framework Based on Kalman Filter**

### **5.1 Introduction**

According to the review and discussion in previous chapter, it is clear that the applications of Kalman filter demonstrate the superiority in the field of multi-sensor travel time estimation. Although a number of studies based on Kalman filter have been undertaken in this area, all of the existing fusion approaches are based on the standard linear Kalman filter. As described in previous chapter, the performance of a Kalman filter based estimation method is largely depended on how accurate the selected state/observation process could describe the dynamics of the system. For the case of travel time estimation, these state/observation processes are from the area of traffic flow theory. Due to limitation of the standard Kalman filter, the selection of traffic process is constrained to linear traffic flow models which have worse performance in describing the dynamics of traffic flow than nonlinear models. On the other hand, the existed works on using nonlinear Kalman filter for travel time estimation are merely based on a single sensor source, and their approaches demonstrate the effectiveness of using nonlinear traffic flow models in Kalman filter based solutions. Therefore, the gap exposed is in using nonlinear Kalman filter to fuse multi-sensor data where more advanced nonlinear traffic models can be applied.

In order to fill the gaps within the application of Kalman filter in this area, a general fusion framework based on nonlinear Kalman filter is developed in this chapter. The main body of this chapter begins with an introduction on traffic flow models which are the basis of the Kaman filter implementation. The second part is to model the three available sensor sources,

ILDs, GPS and ANPR, and illustrate how the data from these sensor sources is fed into the traffic flow models. The third part introduces the proposed fusion framework which uses a discretised space-time model to integrate both the traffic flow models and observations from multiple sensors. The framework introduced in this chapter is kept as general as possible, and the detailed implementation and fusion results are presented in next chapter.

## **5.2 Macroscopic Traffic Flow Models**

### **5.2.1 An overview of traffic flow models**

The subject of traffic flow modelling has existed in academia for more than fifty years. The earliest research is found in Lighthill and Whitham (1955) where they presented a model to describe the analogy of vehicles in traffic flow. Followed by their work, a broad scope of models has been developed to mathematically formulate different aspects of traffic flow behaviours. These models can be largely categorised into three groups based on the level of detail: microscopic, mesoscopic and macroscopic. Although this research uses the macroscopic models as the fundamental process for Kalman filter fusion framework, a brief review on all of three types of models is given below for the purpose of completion.

#### **Microscopic model**

A microscopic model is based on the behaviours of vehicles and drivers as well as their interaction at a high level of details. It means that in microscopic model, each vehicle is examined separately. For example, the behaviour of lane-change of each vehicle in the traffic stream is described as a detailed chain of drivers' decisions (Hoogendoorn & Bovy 2000a). The most famous type of microscopic model is *Car-Following* models which originally studied by Chandler et al. (1958) and Forbe et al. (1958) and the recent reviews can be found in Bando (1995) and Kachroo & Ozbay (1999). Car-following models assume the driver adjusts his or her acceleration according to the traffic conditions in front. The formulations of these models basically use a continuous function to describe the vehicle position, and the motion of each vehicle is governed by an ordinary differential equation (ODE) which depends on the velocity and location of the vehicle in front. Another type of microscopic models is the *Cellular Automata* or *particle hopping* models (Nagel 1996 & 1998). It is different from *Car-Following* models in that it is a fully discrete model. The road link is segmented into a string of cells which are either empty or occupied by one vehicle. However, microscopic model has the drawback of high computationally expense. Each car has an ODE

to be solved at each time step, and the size of entire system to be solved increased dramatically as the number of vehicles increase.

### **Mesoscopic model**

Mesoscopic flow model formulates traffic flow at a medium level of details. Unlike microscopic model, the behaviours of vehicles and drivers are not distinguished nor described individually in mesoscopic model. It uses more aggregated terms such as probability distribution functions to describe the traffic flows. In contrast to the macroscopic model (introduced in next paragraph), the behaviour rules in mesoscopic model are expressed at an individual level. For example, the velocity distributions are described at specific locations and time instants. The dynamics of these distributions are generally governed by a variety of processes such as acceleration, interaction between vehicles and lane-changing to describe the individual driver's behaviour. There are three well known examples of mesoscopic models, i.e. *headway distribution* models (Buckley 1968, Branston 1976, Hoogendoorn & Bovy 1998), cluster models (Kerner & Konhauser 1993), and gas-kinetic continuum models (Prigogine & Herman 1971, Klar & Wegener 1999, Hoogendoorn & Bovy 2000b).

### **Macroscopic model**

Macroscopic traffic flow model is analogous to the theories of fluid dynamics. It represents the behaviours of traffic flow in terms of aggregated measures such as volume/flow, average speed and traffic density. Unlike microscopic model, macroscopic model sacrifices the high level of traffic details, but benefits from the efficiency of dealing with much larger scope. Normally, the model consists of a limited number of equations which are relatively easy to perform necessary calculation. In addition, the number of parameters in macroscopic model is small and, more importantly, easy to observe and measure. Thus, the calibration and validation of macroscopic models require less effort than those of microscopic and mesoscopic models. Since the travel time as an aggregated variable usually describes the high level of traffic conditions, macroscopic model is more appropriate to be applied into the problem of travel time estimation. In the proposed Kalman filter fusion framework, the macroscopic model is then selected to model the traffic dynamic behaviours accordingly.

Since the invention of mathematically modelling traffic flow, the area of macroscopic model is always an active research subject, and a large number of different types of macroscopic models have been developed. This research is about applying those macroscopic models into the Kalman filter fusion framework for the purpose of travel time estimation,

thus only the most relevant models will be introduced. The macroscopic models could be categorised into three groups: conservation law, static velocity-density relationship and dynamic model. Section 5.2.3 to 5.2.5 will provide a brief review about each of them, and the discussion about the opportunities of applying them into the Kalman filter framework will be provided in Section 5.4.

## 5.2.2 Definition of variables

Before the details of macroscopic model and proposed Kalman filter framework are presented, some of the common used variables are defined in this section. The purpose of this section is to provide a clarified description on the variables which may be used frequently throughout the rest of thesis, rather than a comprehensive review of basic traffic flow theory. Therefore, the description is kept brief, and more detailed context could be referred to text books such as Leutzbach (1988), Daganzo (1997), and Roess et al. (1998).

### Mean Speed

In Chapter 3, the concepts of Time-Mean-Speed and Space-Mean-Speed have been introduced and defined in equation 3.1 and 3.2. A time-space version of these definitions will be simply provided as follows:

$$v_{\text{TMS}}(x, t) = \frac{1}{M} \sum_i v_i \quad (5.1)$$

$$v_{\text{SMS}}(x, t) = \frac{NL}{\sum_i \frac{L}{v_i}} = \frac{1}{\frac{1}{N} \sum_i \frac{1}{v_i}} \quad (5.2)$$

where  $v_{\text{TMS}}(x, t)$  is the TMS measured at location  $x$  around time  $t$ , and  $v_{\text{SMS}}(x, t)$  is the SMS measured around location  $x$  at time  $t$ .  $v_i$  is the spot speed of  $i$ th vehicle passing by location  $x$  over a time period and  $M$  is the corresponding total number of measured vehicles.  $N$  is the total number of measured vehicles traversing the road link with a length of  $L$ .

### Density

Density is a typical variable from physics that was adopted by traffic science. Density  $\rho$  reflects the number of vehicles per unit length of road link, and defined as follows:

$$\rho(x, t) = \frac{N}{L} \quad (5.3)$$

where  $\rho(x, t)$  is the density measured around location  $x$  at time  $t$ ,  $N$  is the total number of vehicles along the measured section of road which is the actually same as the  $N$  in the definition of SMS in equation 5.3, and  $L$  is the length of measured road section.

### Flow

The flow or flow rate represents the number of vehicles that passes a certain cross-section per time unit. It is thus a point measurement measured over a period of time. For a time period  $\Delta T$  at given location  $x$  and time instant  $t$ , the flow is defined as follows:

$$q(x, t) = \frac{M}{\Delta T} \quad (5.4)$$

where  $M$  is number of vehicles passing location  $x$  around time  $t$  during  $\Delta T$ .

One important relationship between flow, SMS and density is shown as:

$$q(\rho(x, t), v(x, t)) = \rho(x, t) \cdot v(x, t) \quad (5.5)$$

where the flow  $q(\rho, v)$  can be seen as a function of density  $\rho$  and SMS  $v$ . Note that the notation of SMS is denoted as  $v$  rather than  $v_{\text{SMS}}$ , and it will be used throughout the rest of thesis for the purposes of simplicity.

### 5.2.3 Conservation law

The name of conservation law is from the physical principle of conservation. When physical quantities remain the same during some process, these quantities are said to be conserved. For the case of traffic flow, the number of vehicles in a section of a road link is regarded as the physical quantity. The process of traffic flow is to keep it fixed, i.e., the number of vehicles coming in is equal to the number of vehicles coming out of the section. Consider a stretch of freeway where the vehicles moving from left to right as shown in Figure 5.1. Assuming there is no in and off ramps, the number of vehicles within  $[x_1, x_2]$  at a given time  $t$  is the integral of the traffic density:

$$N = \int_{x_1}^{x_2} \rho(x, t) dx \quad (5.6)$$

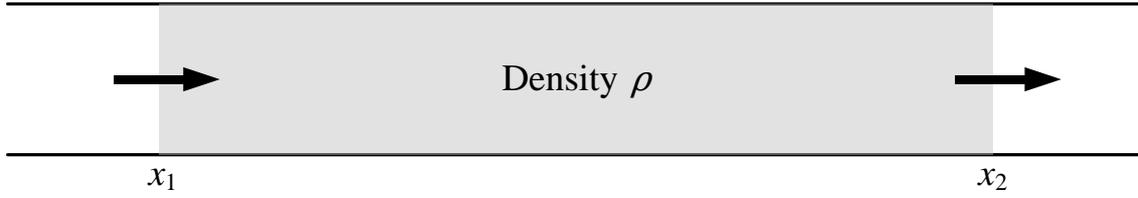


Figure 5.1 Traffic flow density vs. volume

The change of the number of vehicles within  $[x_1, x_2]$  is due to the change at the boundaries only. Therefore, the following equation holds:

$$\frac{dN}{dt} = q_{in}(\rho, v) - q_{out}(\rho, v) \quad (5.7)$$

Substituting  $N$  with equation 5.6, the above equation 5.7 becomes:

$$\frac{\partial}{\partial t} \int_{x_1}^{x_2} \rho(x, t) dx = q_{in}(\rho, v) - q_{out}(\rho, v) \quad (5.8)$$

On the other hand, the change in the number of vehicles with respect to distance is given by:

$$q_{in}(\rho, v) - q_{out}(\rho, v) = - \int_{x_1}^{x_2} \frac{\partial q}{\partial x}(\rho, v) dx \quad (5.9)$$

Combine equation 5.8 and equation 5.9,

$$\int_{x_1}^{x_2} \left[ \frac{\partial \rho}{\partial t}(x, t) + \frac{\partial q}{\partial x}(\rho, v) \right] dx = 0 \quad (5.10)$$

Therefore, the **conservation law** is established as:

$$\frac{\partial \rho}{\partial t}(x, t) + \frac{\partial q}{\partial x}(\rho, v) = 0 \quad (5.11)$$

The equation 5.11 regarding as the conservation law describes the number of vehicles in section  $x$  changes according to the difference of the flow rate at the boundaries  $x$  and  $x + dx$  of the section.

The flow-density-speed relationship in equation 5.5 and the conservation law in equation 5.11 are two independent equations, and constitute the traffic flow system. Note that these two equations hold exactly theoretically. Several model specifications which are all based on approximation will be introduced in the following subsections.

### 5.2.4 Static velocity-density relationship

The traffic density and vehicle velocity are related by the conservation of vehicles as in equation 5.5 and 5.11. Hence, there are two equations but three variables in the system, and it is desired to have another independent function fitted into the traffic flow model. The velocity function about the density  $V(\rho)$  is the choice to fill this gap. The first steady-state velocity-density linear model was introduced by Greenshields (1935), and various models were developed followed by Greenshields' model such as logarithmic model (Greenberg 1959), exponential model (Underwood 1961), the one-parameter family polynomial model (Zhang 1999), the modified exponential model (Papageorgiou & Kotsialos 2002), the multiregime model (Edie 1961).

#### Greenshields (1935)'s linear model

This model is simple and widely used, and given by

$$V(\rho) = v_f \left(1 - \frac{\rho}{\rho_j}\right) \quad (5.12)$$

where  $v_f$  is the free flow speed and  $\rho_j$  is the jammed density or maximum density. The traffic propagates at free flow speed  $v_f$  at zero density, while the traffic stops when the density reach the maximum density  $\rho_j$ .

#### Greenberg (1959)'s logarithmic model

The Greenshields' linear model generally describes the trend of the velocity as density varying, i.e., velocity decreases as the density increases. However, according to the real data observations, this relationship is nonlinearly decreased rather than a linear type (Daganzo, 1997). Greenberg firstly introduced a logarithmic type model to represent this nonlinearity as follows:

$$V(\rho) = v_f \ln \left(\frac{\rho_j}{\rho}\right) \quad (5.13)$$

Although the Greenberg's logarithmic model as a nonlinear function is more appropriate than Greenshields' linear model in practice, it have an obvious drawback: when the density tends to zero, the velocity tends to infinity. This shows the inability of the model to describe the speeds at lower densities.

#### Underwood (1961)'s exponential model

To improve the drawback of Greenberg's logarithmic model, Underwood developed an exponential based model, give as follows:

$$V(\rho) = v_f \exp\left(-\frac{\rho}{\rho_j}\right) \quad (5.14)$$

In this model, speed becomes zero only when density reaches infinity which is not necessarily true in real world. Hence, this model is not suitable to describe the speeds at high densities.

### **Edie (1961)'s Multiregime model**

The drawbacks of both of Greenberg and Underwood's models are due to the assumption of single-regime, i.e., it is assumed the same speed-density relation is valid for the entire range of densities seen in traffic streams. However, the traffic behaviour is different at different densities, and it is verified by with field observations. Therefore, the speed-density relation will also be different in different range of densities. Based on this concept, Edie (1961) proposed a two-regime model which is the combination of Greenberg model and the Underwood model, shown as:

$$V(\rho) = \begin{cases} \exp\left(-\alpha_1 \frac{\rho}{\rho_j} + \alpha_2\right), & \rho \leq \rho_j \\ \beta_1 + \beta_2 \ln\left(\frac{\rho_j}{\rho}\right), & \rho \geq \rho_j \end{cases} \quad (5.15)$$

It is obvious that this model removed the shortcomings such as the violation of boundary conditions in Greenberg and Underwood's models. However, this type of model has 4 extra model parameters  $\{\alpha_1, \alpha_2, \beta_1, \beta_2\}$  to be pre-determined. The task of parameter calibration makes it less appealing in practice.

### **Polynomial model (Zhang 1999)**

Zhang (1999) used a polynomial model which also eliminates the problem of boundary condition in Greenberg and Underwood's models, shown as:

$$V(\rho) = v_f \left(1 - \left(\frac{\rho}{\rho_j}\right)^n\right) \quad (5.16)$$

Although this model does not assume a two-regime velocity-density relation, it still well enough to capture the trend of speed when the density is either small or large. Comparing with Edie's two-regime model, it is easier to implement since only one parameter is required

to calibrate. However, the nonlinearity of velocity-density relation shows a concave shape while this polynomial model is a convex function, which may lead to unsatisfied modelling accuracy.

### Modified exponential model (Papageorgiou & Kotsialos 2002)

The modified exponential model can be seen as an extended version of Underwood (1961)'s exponential model, and it is given by:

$$V(\rho) = v_f \exp\left(-\frac{1}{a}\left(\frac{\rho}{\rho_{cr}}\right)^b\right) \quad (5.17)$$

where  $\rho_{cr}$  is the critical density. Unlike the jammed density  $\rho_j$ , the critical density  $\rho_{cr}$  is the density at which free flow traffic is to breakdown. More specifically,  $\rho_{cr}$  indicates the density value when the traffic is changing from homogenous and stationary state to heterogeneous and unstationary, or from uncongested state to congested state. This phenomenon is simply illustrated by a fundamental diagram in Figure 5.2, and more details can refer to Daganzo (1997).

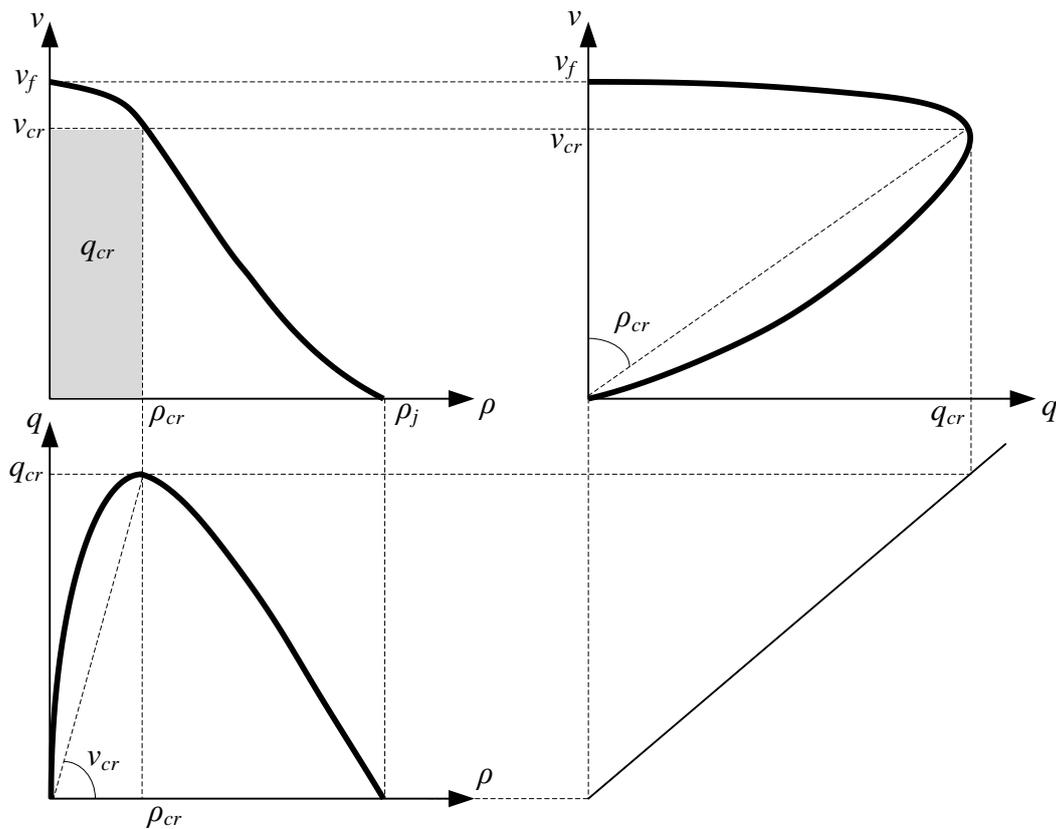


Figure 5.2 Speed-density-flow diagram

By introducing the parameter of critical density, the modified exponential model also distinguish this velocity-density relation into two different regimes as the Edie (1961)'s Multiregime model. The advantages of this model over the Multiregime model is that it uses one complete formula rather than the two segmented functions, and it required less model parameters to be calibrated. It has better concavity-convexity feature than the polynomial model in terms of fitting accuracy. In addition, it is common to set the model parameter  $a = b$  in practice, which makes the number of model parameter equal to polynomial model.

Based on the above discussion about the different types of static velocity-density models, it can be seen that the modified exponential model is superior to other typical models. Therefore, the modified exponential model in equation 5.17 is employed in the proposed Kalman filter fusion framework, and the more details of implementation will be introduced in Section 5.4. Figure 5.3 shows an example of modelling the velocity-density relation with different types of models by real traffic data, the data source used here is same as the one in Section 3.4.1, Chapter 3. Note that the objective of this thesis is not to find the *best* velocity-density model, and quantitative based comparison analysis is out of this thesis's scope.

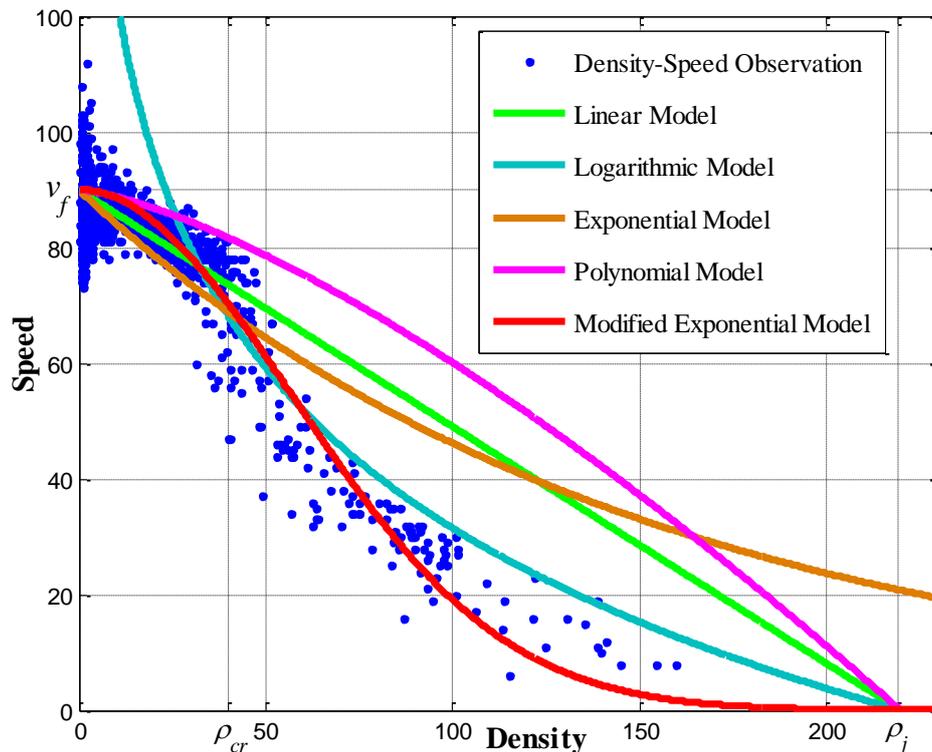


Figure 5.3 Comparison among different velocity-density models

### 5.2.5 Dynamic model

The aim of the models presented in above section is to describe the static relationship between traffic speed and density. As known, the traffic varies according to the changes of a number of certain factors. The underlying cause which drives the traffic state to vary is formulated into another type of macroscopic models, named dynamic model here. Since the traffic variables are usually spatial and temporal dependent such as  $q(x, t)$ ,  $\rho(x, t)$  and  $v(x, t)$ , the dynamics of these variables is modelled in the form of partial derivative equation (PDE) with respect to location and time. There are two prototypes of dynamic model: first one is called the LWR model which is a nonlinear, first-order hyperbolic PDE based on conservation law. The second one is a nonlinear, second-order model known as the PW model, which is based on two coupled PDEs, one is given by the conservation law and a second equation that imitate traffic flow. The rest of this section will introduce these two models briefly.

#### LWR model (Lighthill & Whitham 1955, Richards 1956)

The LWR model is the first model used in describing the dynamics problem of traffic flow, name after the authors in Lighthill & Whitham (1955) and Richards (1956). The LWR model is a scalar, time-varying, non-linear, hyperbolic partial differential equation. The model governing equation is the conservation law in equation 5.11, and the traffic density is the conserved quantity shown as:

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial [V(\rho(x, t))\rho(x, t)]}{\partial x} = 0 \quad (5.18)$$

with the flow  $q(x, t)$  replaced by:

$$q(\rho, v) = \rho(x, t) \cdot V(\rho(x, t)) \quad (5.19)$$

and  $V(\rho(x, t))$  is the velocity function presented in previous section.

The basic assumption in the LWR model about the velocity is that it is only dependent on the traffic density, i.e., any changes to the velocity are due to the changes of density. The drawback of this assumption is that the traffic is in equilibrium when only a velocity-density function is used. Daganzo (1995) stated that given a particular density, the velocity of the traffic stream will be fixed, and the LWR model does not recognise that there is a distribution of desired velocities across the vehicles. Therefore, the LWR model is not able to capture all of the complex interactions for a realistic traffic environment. For this reason, modifications

to the LWR model have been suggested. One approach is to use more complicated velocity-density functions such as the diffusion model (Musha & Higuchi 1978, Burns & Kang 1991). The second way is to couple the conservation law with a second equation that tries to imitate the traffic motion.

### **PW model (Payne 1971, Whitham 1974)**

PW model was proposed in the 1970s independently in Payne (1971) and Whitham (1974). This model was the first model to couple the speed dynamics as a second equation, and has been widely used for decades. Same as the LWR model, the first equation of PW model is the conservation law as discussed in previous section:

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial [v(x, t)\rho(x, t)]}{\partial x} = 0 \quad (5.20)$$

In the LWR model, a particular form of  $v$  is assumed as a function of density  $\rho$ . However, in higher order models,  $v$  and  $\rho$  are assumed to be independent and a second equation is formed to link them. In PW model, this second equation is derived from a simple car-following rule (Montroll 1961), given by:

$$\frac{\partial v(x, t)}{\partial t} + v(x, t) \frac{\partial v(x, t)}{\partial x} = \frac{V(\rho(x, t)) - v(x, t)}{\tau} - \frac{\psi}{\rho(x, t)} \frac{\partial \rho(x, t)}{\partial x} \quad (5.21)$$

where  $\tau$  is a relaxation time,  $V(\rho(x, t))$  is the velocity-density function given in Section 5.2.4  $\psi = \bar{\omega}/\tau$  is the anticipation constant with  $\bar{\omega} = -dV(\rho)/\rho$  describing the decrease rate in the equilibrium velocity with increasing density. PW model consists of both of equation 5.20 and 5.12, and a more compact form is given as follows:

$$\begin{cases} \rho_t + (\rho v)_x = 0 \\ v_t + vv_x = \frac{V(\rho) - v}{\tau} - \frac{\psi}{\rho} \rho_x \end{cases} \quad (5.22)$$

where independent variables  $\{x, t\}$  are removed and the notation  $(\cdot)_\phi = \partial(\cdot)/\partial\phi$ . It can be seen that the two-equation PW model does not only describe the dynamics of traffic density but also the dynamics of the velocity of traffic stream in a more complete way:

- the term  $vv_x$  describes changes in the mean velocity due to inflow and outflow vehicles
- the term  $(V(\rho) - v)/\tau$  describes the tendency of traffic flow to relax to an equilibrium velocity

- the term  $(\psi/\rho)\rho_x$  describes the anticipation of macroscopic drivers on spatially changing traffic conditions downstream

It is obvious that the PW model is more advanced than the LWR model since more underlying factors are formulated into the model which makes it more realistic in practice. Therefore, this model is chosen as the state process model in the developed Kalman filter fusion framework which will be presented in Section 5.4.

## 5.3 Sensor Observation Models

The previous section introduced the most popular macroscopic traffic models which could be used as the state process equations in Kalman filter fusion framework. On the other hand, to make use of these traffic models and perform the data fusion, the observations from multiple sensor sources as the input for the fusion framework need to be modelled. This section describes the fundamental modelling approaches for the sensor sources used in this research: ILDs, GPS probes and ANPR cameras. The main discussion on these three types of sensors has been given in Chapter 2. This section only focuses on the mathematically formulating the outputs of these sensors rather than the working principles and sensing performance of the sensors.

### 5.3.1 ILD model

Normally, the most directly observations from ILDs are flow and occupancy according to the working principle of ILD (Klein et al. 2006). For the double-ILDs, the TMS can be observed locally. Although other traffic parameters can be calculated or approximated from ILD outputs (Robinson 2005), only flow will be considered here since the macroscopic models only contains this parameters. The errors associated with the flow observations can be mainly categorised into two parts. One part is regarded as systematic random error which normally includes the broken cables, electronic interference, communication noise, software error etc. Robinson & Polak (2006) provides a comprehensive description on the sources of these errors in more details. The other part is the bias error which is generally due to the problem of lane-cross and low polling frequency (Krishnan 2008a). Thus, the flow observation from ILDs  $q_{ILD}$  is given by:

$$q_{ILD}(x, t) = q(x, t) + \varpi + \omega \quad (5.23)$$

where  $q$  denotes actual flow value,  $\varpi$  and  $\omega$  represent the terms of random error and bias errors respectively. The forms of  $\varpi$  and  $\omega$  are shown as:

$$\varpi \sim \mathcal{N}(0, \Phi) \quad (5.24)$$

$$\omega \sim \mathcal{L}(\mu, \Psi) \quad (5.25)$$

The random error  $\varpi$  is modelled by a normal distribution with 0 mean and  $\Phi$  variance. As discussed in Section 4.1.3, Chapter 4, when adding a number of independent random variables together, the normal distribution is the most reasonable approximation about the distribution of this summation. For the bias errors term  $\omega$  caused by lane-cross and low polling frequency, the bias usually introduces an underestimate result. Therefore, the mean of the distribution  $\mu < 0$ , and the type of distribution is unknown, indicated by  $\mathcal{L}(\cdot)$ .

### 5.3.2 GPS probe model

The observation of GPS probes is naturally a geographic coordinate (latitude, and longitude) and the corresponding time. More detailed description can refer to Zhao (1997). In order to make use of this data source in the travel time estimation, the observed coordinate and corresponding time have to be converted into the mean speed of probe vehicle  $v_{\text{GPS}}$  along the position  $x$  at time  $t$ , which is given by:

$$v_{\text{GPS}}(x, t) = \frac{L(c_1, c_2)}{t_2 - t_1} + \varepsilon \quad (5.26)$$

where  $L(c_1, c_2)$  is the length between two observed coordinate  $c_1$  and  $c_2$ ,  $t_1$  and  $t_2$  are the time when the  $c_1$  and  $c_2$  are observed.  $\varepsilon$  is noise term which represents the errors of GPS observations. The sources of errors have been discussed in Chapter 2. The noise term  $\varepsilon$  in this model only considers the systematic error and the error introduced by map-matching (Quddus 2006). The time and space resolution of  $v_{\text{GPS}}(x, t)$  in this research is observed in a scale of 30 seconds and 400 meters, thus the added error effect (from both of systematic error and map-matching error) is reasonably assumed by a normal distributed noise term  $\varepsilon$ :

$$\varepsilon \sim \mathcal{N}(0, \Omega) \quad (5.27)$$

As presented in Section 5.2, the speed variable in the macroscopic model is SMS, whereas the observation from GPS probe is only a sample of the whole vehicle population. Since the size of the samples is relatively small comparing to the total vehicles in the traffic stream,

there exists an error between the true SMS and  $v_{\text{GPS}}$ . This type of error has been discussed earlier in Chapter 2 as well, and simply indicated represented by:

$$v_{\text{GPS}}(x, t) = v(x, t) + \epsilon \quad (5.28)$$

The statistical feature of  $\epsilon$  is largely affected by the type of probe vehicle and the driving behaviour. For example, trucks which are usually equipped with GPS tracking devices tend to be slower than private vehicles, and the observation  $v_{\text{GPS}}$  is likely to be underestimated. Without losing any generality,  $\epsilon$  is assumed to be a non zero-mean normal distributed as follows:

$$\epsilon \sim \mathcal{N}(\nu, \theta) \quad (5.29)$$

In the real application, the type of probe vehicles is normally known, and this bias could be eliminated by empirical analysis.

### 5.3.3 ANPR model

The observation from ANPR camera system is the travel time averaged from matched vehicles. As discussed in Chapter 2, the ANPR also has the problem of sampling error, since the matched vehicles are only part of the entire vehicle population which travels through the locations of the paired ANPR cameras. Note that the percentage of the sampled vehicle is much larger than the percentage of GPS probe. Typically the recognition rate of ANPR is between 50% and 90% (van der Zijpp 1997), and Wiggins (1999) shows a recognition rate of 86% during a 2 hour survey in the UK. Robinson (2005) states that the recognition rate depends primarily on the vehicle characteristics, ANPR system used, quality of installation, and weather condition, then a vehicle that has its license plate successfully read at an upstream point will most probably be successfully detected at a downstream point, and the matching rate is likely to be very similar to the recognition rate.

Similar to GPS probe model, the ANPR observation model is given by

$$v_{\text{ANPR}}(x, t) = v(x, t) + \varphi \quad (5.30)$$

$$\varphi \sim \mathcal{N}(\zeta, Y) \quad (5.31)$$

The noise term  $\varphi$  is approximated by a normal distribution with mean  $\zeta$  and variance  $Y$ . This introduced sampling noise is normally biased, i.e.  $\zeta \neq 0$ . The number plates of large vehicles such as bus and lorry are usually attached at different positions which are difficult to perform the process of the automatic recognition. It indicates that the observed travel time is

aggregated from the class of relatively faster vehicles, since the class of large vehicles which has the low recognition rate tends to be slower. In other words, the observation  $v_{ANPR}(x, t)$  is likely to be an overestimate.

## 5.4 Nonlinear Kalman Filter Fusion Framework

### 5.4.1 Road link segmentation

The PW macroscopic model introduced in Section 5.2 assumes the traffic stream is homogeneous. However, the traffic conditions or states will not be consistent especially when there exists on/off ramps, junctions or congestion at a specific location along the road link. Therefore, the assumption of homogeneity cannot be satisfied for a practical road link. To overcome this problem, Papageorgiou et al. (1990) proposed a segmented based method to subdivide a stretch of road link into several segments with length of 500 metres (or less). In making use of the PW macroscopic model in the proposed fusion framework, the road link was segmented based on Papageorgiou et al. (1990)'s method.

In the discrete space form, the lengths of  $N$  segments are denoted with  $L_i, i = 1, 2, \dots, N$ . The aggregated traffic variables, i.e. traffic density, SMS and traffic flow are described at each of the segment as shown in following figure:

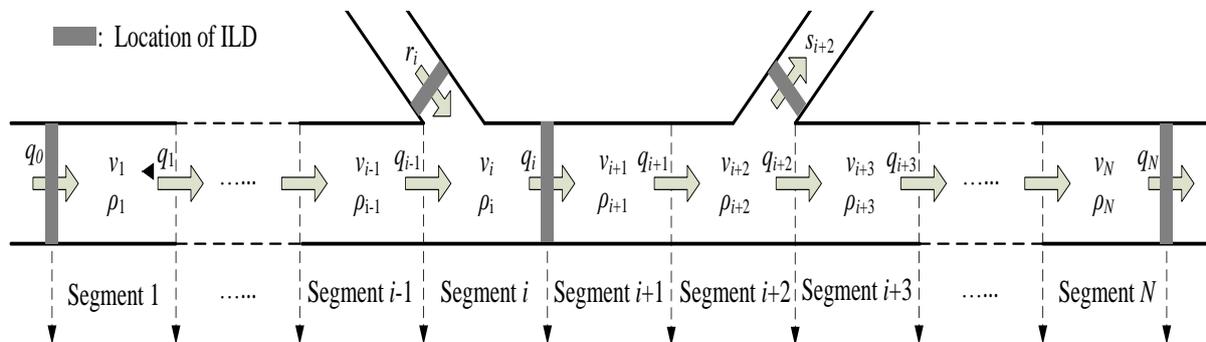


Figure 5.4 Space-discrete of road link

- $\rho_i$  is the traffic density, the number of vehicles in segment  $i$  divided by the segment length  $L_i$
- $v_i$  is the SMS, the average speed of all vehicles included in segment  $i$
- $q_i$  is the traffic flow, the number of vehicles leaving segment  $i$  divided by the observed time period
- $r_i/s_i$  is the inflow/outflow in segment  $i$  (if any)

While subdividing the road link into segments, the boundaries of each segment should be carefully chosen, any geometric inhomogeneities (e.g. lane drops, on/off-ramps, junctions) and locations of the installed traffic detectors are normally determined as the segments boundaries.

### 5.4.2 Discretisation of macroscopic model

As presented in Section 5.2, the original derivative of the macroscopic models is continuous. Although the continuous PDE describes the underlying dynamics of the traffic behaviours, it is too difficult to apply directly in practice. The observations from sensors are all in the discrete form, thus it is necessary to discretise the continuous PDE of macroscopic model into time-space difference equations which can be fitted into the scenario of segmented link in previous section.

#### Discretisation of traffic variables

The traffic variables used in the macroscopic models are in the form of  $(x, t)$  which describe the variables in continuous space  $x$  and time  $t$ . To approach the discretisation of the traffic variables,  $x$  is subdivided by the length of each of segment, and  $t$  is subdivided by the time period of observations  $T$ , shown as:

$$x \rightarrow 0, L_1, L_2, \dots, L_i, \dots, L_N \quad (5.32)$$

$$t \rightarrow 0, T, 2T, \dots, kT, \dots, Nk \quad (5.33)$$

and the traffic variables are represented in discrete form by

$$q(x, t) \rightarrow q_i(k) \quad (5.34)$$

$$v(x, t) \rightarrow v_i(k) \quad (5.35)$$

$$\rho(x, t) \rightarrow \rho_i(k) \quad (5.36)$$

where  $i$  denotes the  $i$ th segment and  $k$  denotes the  $k$ th time interval.

#### Discretisation of velocity-density model

The velocity-density model is straightforwardly discretised by replacing the continuous variables into the discrete form in equation 5.34-5.36, given by

$$V(\rho_i(k)) = v_f \exp\left(-\frac{1}{a}\left(\frac{\rho_i(k)}{\rho_{cr}}\right)^b\right) \quad (5.37)$$

### Discretisation of PW model

The PDE of PW model in Section 5.2.5 is given by:

$$\begin{cases} \frac{\partial \rho(x, t)}{\partial t} + \frac{\partial v(x, t)\rho(x, t)}{\partial x} = 0 \\ \frac{\partial v(x, t)}{\partial t} + v(x, t)\frac{\partial v(x, t)}{\partial x} = \frac{V(\rho(x, t)) - v(x, t)}{\tau} - \frac{\psi}{\rho(x, t)} \frac{\partial \rho(x, t)}{\partial x} \end{cases} \quad (5.38)$$

According to the temporal and spatial discretisation in equation 5.32 and 5.33, the partial derivative  $\partial f(x, t)/\partial t$  and  $\partial f(x, t)/\partial x$  can be expressed as

$$\frac{\partial f(x, t)}{\partial t} \rightarrow \frac{f_i(k) - f_i(k-1)}{T} \quad (5.39)$$

$$\frac{\partial f(x, t)}{\partial x} \rightarrow \frac{f_i(k) - f_{i-1}(k)}{L_i} \quad (5.40)$$

Thus, the first equation of PW model is discretised as

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [v_{i-1}(k)\rho_{i-1}(k) - v_i(k)\rho_i(k)] \quad (5.41)$$

and replacing the term  $v(x, t)\rho(x, t)$  by  $q(x, t)$ :

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k)] \quad (5.42)$$

This equation is also known as the discrete conservation law. The physical meaning of this equation is that the density dynamics of the segment is due to the different between inflow and outflow. Considering the ramp inflow/outflow in the scenario of space-discrete road link in Figure 5.4, the equation for discrete conservation law becomes:

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (5.43)$$

For the second equation of PW model, the velocity dynamics, the discretisation of partial derivative formulation in equation 5.39 and 5.40 can be also applied except for the term  $\psi \partial \rho(x, t)/\rho(x, t) \partial x$ . According to the description in Section 5.2.5, this term is assumed to reflect drivers' anticipatory behaviour. It is approximated by using a forward discretisation scheme (Smulders 1990)

$$\frac{\psi}{\rho(x,t)} \frac{\partial \rho(x,t)}{\partial x} \rightarrow \frac{\psi}{L_i} \frac{(\rho_{i+1}(k) - \rho_i(k))}{\rho_i(k) + c} + \xi_i^v \quad (5.44)$$

$\xi_i^v$  denotes the noise introduced by the approximation. Apply the above equation and equation 5.39 and 5.40 into the velocity dynamics equation, and combine the discrete conservation law, the discretised formulation of PW model is given by

$$\begin{cases} \rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \\ v_i(k+1) = v_i(k) + \frac{T}{\tau} [V(\rho_i(k)) - v_i(k)] + \frac{T}{L_i} v_i(k) [v_{i-1}(k) - v_i(k)] \\ - \frac{\psi T}{\tau L_i} \frac{(\rho_{i+1}(k) - \rho_i(k))}{\rho_i(k) + c} + \xi_i^v(k) \end{cases} \quad (5.45)$$

Through the discretisation process described above, the macroscopic model is converted into the discrete domain, and is suitable to be applied into the scenario of segmented road link. The dynamic equations of density  $\rho_i(k)$  and speed  $v_i(k)$  clearly show a form of time difference functions which can be constructed as the state process of Kalman filter. The next section will present the Kalman filter fusion framework by combining the velocity-density model and dynamic PW model.

### 5.4.3 Discretisation of sensor observation model

The discrete versions of sensor observation models are presented in this section. Unlike the macroscopic model, the sensor models are simply discretised on the time domain since the locations of the sensor are determined.

#### Discretisation of ILD model

$$q_{\text{ILD}}(x,t) \rightarrow q_{\text{ILD}}^j(k) \quad (5.46)$$

where  $q_{\text{ILD}}^j(k)$  denotes the flow observation from the  $j$ th ILD at the time interval  $k$

#### Discretisation of GPS model

$$v_{\text{GPS}}(x,t) \rightarrow v_{\text{GPS}}^i(k) \quad (5.47)$$

where  $v_{\text{GPS}}^i(k)$  denotes the space speed observation from the  $i$ th segment at the time interval  $k$

#### Discretisation of ANPR model

$$v_{\text{ANPR}}(x,t) \rightarrow v_{\text{ANPR}}(k) \quad (5.48)$$

The discretisation of ANPR model is simply to aggregate the observation at the time period of  $T$ , and then apply the equation 5.33.

#### 5.4.4 Construction of Kalman filter fusion framework

From the description in Chapter 4, Kalman filter estimation framework mainly consists of three key components: state to be estimated, equation(s) of state process, and equation(s) of observation process. The previous sections in this chapter have prepared the fundamentals of each component. This section generally introduces the how these elements are fitted into the Kalman filter based fusion framework.

##### State

The system state in the proposed framework is defined as the dynamic density, SMS and flow, and formed by a three-dimension vector, given by

$$\mathbb{X} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{v} \\ \boldsymbol{q} \end{bmatrix} = \begin{bmatrix} (\rho_1, \rho_2, \dots, \rho_N)^T \\ (v_1, v_2, \dots, v_N)^T \\ (q_1, q_2, \dots, q_N)^T \end{bmatrix} \quad (5.49)$$

##### State process

The underlying function(s) of the state process is to describe the dynamics of the state. The nonlinear, discretised PW model (equation 5.44) used to model the first two state elements,  $\boldsymbol{\rho}$  and  $\boldsymbol{v}$ . The third element flow  $\boldsymbol{q}$  is modelled by an auxiliary random-walk equation as

$$\boldsymbol{q}(k+1) = \boldsymbol{q}(k) + \boldsymbol{\xi}^q(k) \quad (5.50)$$

Combining equation 5.44 and 5.46, a compact form of state process is shown as follows

$$\mathbb{X}(k+1) = \mathcal{F}[\mathbb{X}(k), \mathbb{u}(k), \mathbb{w}(k)] \quad (5.51)$$

where  $\mathcal{F}$  denotes the nonlinear three-dimension function which describes the dynamics of the state  $\mathbb{X}$ .  $\mathbb{u}(k)$  as the control input represents the ramp inflow/outflow of each segment and the model parameters as follows

$$\mathbb{u} = [\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{z}]^T \quad (5.52)$$

where  $\mathbf{z}$  is the vector containing all the model parameters  $\{v_f, \rho_{cr}, a, b, \psi, \tau, c\}$ .  $\mathbf{w}(k)$  is the noise vector of which each element denotes the corresponding state noise for each state variable, shown as

$$\mathbf{w} = [\xi^\rho, \xi^v, \xi^q]^T \quad (5.53)$$

### Observation process

The multi-sensor observations are flow from ILDs, speed from GPS and ANPR. To associate these available observations with the system state, the observation models introduced in Section 5.3 and 5.4 are employed as:

$$\mathbf{q}_{ILD}(k) = \boldsymbol{\rho}(k)\mathbf{v}(k) + \boldsymbol{\omega}(k) + \boldsymbol{\omega}(k) \quad (5.54)$$

$$\mathbf{v}_{GPS}(k) = \mathbf{v}(k) + \boldsymbol{\epsilon}(k) \quad (5.55)$$

$$v_{ANPR}(k) = \sum_{i=1}^N v_i(k) + \varphi(k) \quad (5.56)$$

The observation vector is

$$\mathbf{Y}(k) = [\mathbf{q}_{ILD}, \mathbf{v}_{GPS}, v_{ANPR}]^T \quad (5.57)$$

These equations can be written as a compact form as

$$\mathbf{Y}(k) = \mathcal{H}[\mathbf{X}(k), \mathbf{v}(k)] \quad (5.58)$$

where  $\mathcal{H}$  is the multi-dimension transition function which consists of the equation 5.54-5.56.  $\mathbf{v}(k)$  is the noise vector given by

$$\mathbf{v}(k) = [\boldsymbol{\omega}, \boldsymbol{\omega}, \boldsymbol{\epsilon}, \boldsymbol{\epsilon}, \varphi]^T \quad (5.59)$$

### Complete structure

The three components of Kalman filter have been presented above, thus the complete fusion framework is constituted in a compact form as

$$\begin{aligned} \mathbf{X}(k+1) &= \mathcal{F}[\mathbf{X}(k), \mathbf{u}(k), \mathbf{w}(k)] \\ \mathbf{\Gamma}(\mathbf{X}, \mathbf{Y}, \mathbf{u}, \mathbf{w}, \mathbf{v}): & \quad \updownarrow \\ & \mathbf{Y}(k) = \mathcal{H}[\mathbf{X}(k), \mathbf{v}(k)] \end{aligned} \quad (5.60)$$

It can be seen that the equation 5.60 is a typical structure of nonlinear Kalman filter which contains all the available observations from multi-sensor sources. By applying EKF or UKF algorithm introduced in Chapter 4, the state can be directly estimated which leads to the

solution of the developed Kalman filter based fusion framework. Figure 5.5 illustrates the complete framework.

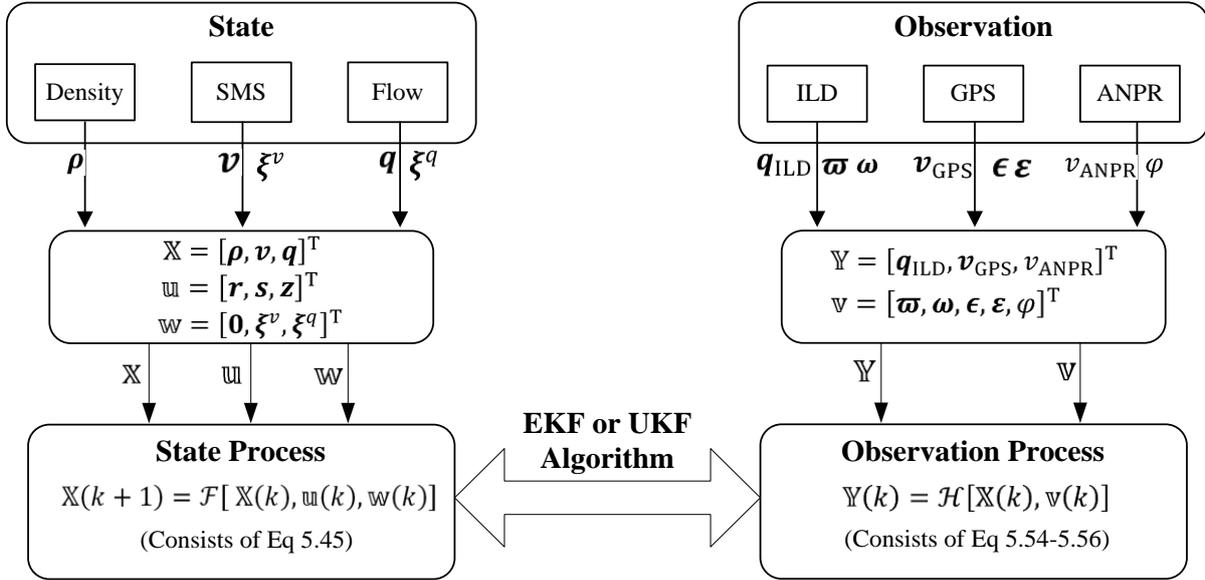


Figure 5.5 Fusion framework based on Kalman filter

## 5.5 Issues and Solutions of Framework Implementation

### 5.5.1 Model parameters

For the Kalman filter framework developed in Chapter 5, the model parameters  $\{v_f, \rho_{cr}, a, b, \psi, \tau, c\}$  are normally unknown and may vary with environmental conditions. However, the model results are known to be most sensitive to variations of the free speed  $v_f$ , critical density  $\rho_{cr}$ , and exponents  $\{a, b\}$  (Papageorgiou et al., 1990). Therefore, only these four parameters will be considered as the unknown model parameters in this research, and the rest of them will be assumed to be obtained by off-line model calibration which has been studied in Cremer & Papageorgiou (1981). Their calibration used the data from 13 daytime hours, while the neural networks normally need at least 10 days data for the purpose of training (Park & Rilett 1999; Huisken & van Berkum 2003; Liu et al. 2006 and Li et al. 2008). This superiority of Kalman filter to neural networks is due to the nature of reliance on historical data. Neural network as a pure data-driven based method requires large amount historical data to extract as much patterns as possible. In contrast, the Kalman filter based approach developed in this chapter has a well-defined model to describe the dynamics of traffic flow and the relation with the sensor observations. The reliance on the historical data is

to calibrate three model parameters  $\{\psi, \tau, c\}$ . The studies in Cremer & Papageorgiou (1981), Papageorgiou et al. (1989) and Papageorgiou et al. (1990) uses different real world scenarios to illustrate that these three parameters are not sensitive to the modelling performance, and normally one day data which contains uncongested, congested and transition period is adequate for the calibration.

The conventional method of selecting these parameters  $\{v_f, \rho_{cr}, a, b\}$  was assumed they were constant and could be performed by off-line calibration (Gazis & Liu 2003, Hegyi et al. 2006, Pueboobpaphan et al. 2007). As discussed above, the assumption of constant parameters may lead to degraded model performance. In Wang & Papageorgiou (2005) which uses EKF to estimate traffic state based on ILD data, they proposed an on-line scheme to estimate these parameters by introducing an auxiliary random walk process. Denoting the model parameters  $\{v_f, \rho_{cr}, a, b\}$  as  $\mathbf{p}$ , and the random walk process is given as follows

$$\mathbf{p}(k+1) = \mathbf{p}(k) + \xi^p(k) \quad (5.61)$$

where  $\xi^p(k)$  represents a vector of zero-mean Gaussian white noise. By integrating this random walk process into the state process of Kalman filter, this approach provides a solution to the on-line estimation of model parameters. However, in their paper, it is not mentioned that how to determine the variance of this introduced Gaussian white noise  $\xi^p(k)$ . This research extends this scheme to introduce a clear way to obtain this variance. By taking advantage of multiple sensor sources, the observations from GPS probes is employed to approximate the variance of  $\xi^p(k)$  shown as

$$\mathbf{Q}^p(k) = \mathbb{E}[\xi^p(k)\xi^p(k)^T] = \mathbb{E}[[\mathbf{p}(k) - \mathbf{p}_{\text{GPS}}(k)][\mathbf{p}(k) - \mathbf{p}_{\text{GPS}}(k)]^T] \quad (5.62)$$

where  $\mathbf{p}_{\text{GPS}}(k)$  is the model parameters approximated by the GPS observations. By using equation 5.62, the variance of introduced random walk is obtained straightforwardly. Note that the exponent model parameters  $\{a, b\}$  are usually assumed to be same ( $a = b$ ), and it will reduce the complexity of approximation.

### 5.5.2 Biased noise

In the previous section, the observations from ILDs, GPS probes and ANPR are modelled in equations 5.23, 5.28 and 5.30 respectively. All of these observations have a biased or non-zero mean noise term. As introduced in Section 4.1.3, Chapter 4, the Kalman filter achieves optimal estimation result when the state and observation noises are zero mean. Therefore, this

issue needs to be considered in implementation. From the discussion in Section 5.3.2 and 5.3.3 in Chapter 5, the cause of the biases in GPS and ANPR is that the observed samples only from a particular type of vehicles, and there is certain difference between the travel behaviour of observed samples and overall traffic stream. In real applications, the type of these observation samples is known, thus the biased mean could be approximated by the analysis based on empirical data.

For the biased noise of the ILDs, a simple treatment is used in the implementation. From the discussion in Section 5.3.1, Chapter 5, the main causes of the flow observations from ILDs are lane-cross and low polling frequency. Since the ILDs on a section of road link measure the same traffic stream, the mean of the biased noise is assumed to be constant over all of the ILDs within the objective road link. It indicates that the mean of difference between two adjacent flow observations from ILDs is approximately zero. For the density dynamic equation in the state process,

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (5.63)$$

the second term in the RHS of the equation is actually the flow increment/decrement, and it can be rewritten as

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} \Delta q_i(k) + \xi^{\Delta q}(k) \quad (5.64)$$

where  $\Delta q_i$  denotes the difference between inflow and outflow observations for the segment  $i$ . Based on the discussion above, the variable  $\xi^{\Delta q}$  calculated from ILDs observations can be assumed as normal distributed with zero mean:

$$\xi^q \sim \mathcal{N}(0, \mathbf{Q}^{\Delta q}) \quad (5.65)$$

The flow equation in observation process can also be converted into a difference form as

$$\Delta \mathbf{q}_{\text{ILD}}(k) = \Delta[\boldsymbol{\rho}(k)\mathbf{v}(k)] + \boldsymbol{\omega}(k) \quad (5.66)$$

where

$$\boldsymbol{\omega} \sim \mathcal{N}(0, \mathbf{R}^q) \quad (5.67)$$

By using this difference based treatment on the flow observations from ILDs, the mean of observation noise is approximated to zero which satisfies the assumption of Kalman filter for optimal estimation.

### 5.5.3 Missing observations

In the general fusion framework in Chapter 5, the observations are assumed to be available for every segment and time interval. In practice, this assumption is not realistic since the observations from GPS probes are limited by the spatial coverage and percentage rate. There may have missing observations for some of the segments and time intervals. One of the advantages of Kalman filter is that it is able to perform the estimation process without the incoming observations. For example, if the GPS observation is not available at time step  $k$ , the observation vector  $\mathbb{Y}(k)$  is reduced as

$$\mathbb{Y}(k) = [\mathbf{q}_{\text{ILD}}, \mathbf{v}_{\text{GPS}}, \mathbf{v}_{\text{ANPR}}]^T \Rightarrow \tilde{\mathbb{Y}}(k) = [\mathbf{q}_{\text{ILD}}, \mathbf{v}_{\text{ANPR}}]^T \quad (5.68)$$

and the observation process becomes

$$\tilde{\mathbb{Y}}(k) = \tilde{\mathcal{H}}[\mathbb{X}(k), \tilde{\mathbf{v}}(k)] \quad (5.69)$$

where  $\tilde{\mathcal{H}}$  is the reduced functions given by

$$\Delta \mathbf{q}_{\text{ILD}}(k) = \Delta[\boldsymbol{\rho}(k)\mathbf{v}(k)] + \boldsymbol{\omega}(k) \quad (5.70)$$

$$\mathbf{v}_{\text{ANPR}}(k) = \sum_{i=1}^N v_i(k) + \boldsymbol{\varphi}(k) \quad (5.71)$$

The noise term  $\tilde{\mathbf{v}}(k)$  is also reduced as

$$\tilde{\mathbf{v}}(k) = [\boldsymbol{\omega}, \boldsymbol{\varphi}]^T \quad (5.72)$$

It can be seen that the issue of missing observations only has effect on the dimension of observation vectors and equations. For the extreme case when there is no observation during the time interval, the observation process becomes null, and the task of the estimation is only based on the state dynamic process. Although the estimation performance will be degraded under the condition of missing observations, the developed fusion framework is still functional and outputs the estimate optimally. It is one of the advantages of Kalman filter based estimation over other data driven based methods such as neural networks.

## 5.6 Summary

A general Kalman filter based fusion framework is established in this Chapter. In the problem of designing Kalman filter estimation system, the state process and observation process are most important component to be determined. A number of macroscopic traffic models have been examined, and the PW dynamic model and modified exponential velocity-density model

are chosen as the key equations for these two processes respectively. On the other hand, the observations from multiple sources are also carefully modelled, and fitted into the observation process. To make the fusion framework available for real-world application, a discretisation process is utilised to convert the theoretically continuous traffic models into discrete temporal-spatial scenario. Based on this scenario, the Kalman filter based fusion framework is constructed. It can be seen that this framework makes use of the well-developed macroscopic traffic model and the observations from multi-sensor sources, i.e. all the knowledge related to the estimation problem, which the superiority of Kalman filter is as discussed in Chapter 4. Some of the issues of framework implementations have been addressed, such as such as unknown model parameters, biased noise and missing observations. The proposed solutions to these issues are provided accordingly.

The aim of this chapter is to identify and develop all the necessary elements of the fusion framework. The detailed framework implementations are not universal due to the difference of scenarios. The next chapter will present the implementation results based on simulation experiment. A real traffic data based experiment is carried out to demonstrate the fusion performance in chapter 6.

## Chapter 6

# Framework Implementations & Evaluation Based on Simulations

A general Kalman filter based fusion framework was presented in the previous chapter. It demonstrated that using a well-known macroscopic flow model as the state and observation processes is possible to appropriately establish the structure of the Kalman filter. This chapter uses simulation based experiments to illustrate the implementation of the framework. The results based on the simulation scenario are comprehensively presented, including a general fusion based on ILDs, GPS and ANPR data, a more practical fusion based on only ILDs and GPS data, sensitivity analysis for the factors of GPS probes and the spatial availability of ILDs, and comparison analysis with neural networks and linear type Kalman filter.

### 6.1 Introduction to the Simulation Experiment

#### 6.1.1 Objectives of the simulation experiment

The purpose of the simulation experiment presented in this chapter is to generate the data required for evaluating the performance of the fusion framework presented in last chapter. Generally, there are several factors that might affect the fusion performance, such as coverage and configuration of ILDs, sampling percentage and rate of GPS probes, calibration of model parameters and measurement noise. This experiment investigates how these factors will have an effect on the estimation performance. The objectives of this experiment are summarised as follows:

1. Evaluate and compare the performance of proposed fusion framework by both of EKF and UKF algorithm

2. Analyse the sensitivity of the proposed framework with respect to the factors of GPS probes
3. Analyse the sensitivity of the proposed framework with respect to different ILDs configurations
4. Compare the fusion performance with other methods such as neural networks and linear Kalman filter

The criterion used for quantitatively evaluating the performance is defined in four metrics, i.e. Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Root Mean Squared Percentage Error (RMSPE) which are defined as follows:

$$\text{MPE:} \quad J_{MPE}^{(x)} = 100 \cdot \frac{1}{N} \sum_{n=1}^N \frac{(x_n - \hat{x}_n)}{x_n} \quad (6.1)$$

$$\text{MAPE:} \quad J_{MAPE}^{(x)} = 100 \cdot \frac{1}{N} \sum_{n=1}^N \frac{|x_n - \hat{x}_n|}{x_n} \quad (6.2)$$

$$\text{RMSE:} \quad J_{RMSE}^{(x)} = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \hat{x}_n)^2} \quad (6.3)$$

$$\text{RMSPE:} \quad J_{RMSPE}^{(x)} = \sqrt{\frac{1}{N} \sum_{n=1}^N \frac{(x_n - \hat{x}_n)^2}{x_n^2}} \quad (6.4)$$

where,

$N$  is the total of valid states during the evaluation time period

$x_n$  is the actual value of the  $n$ th state during the evaluation time period

$\hat{x}_n$  is the estimated value of the  $n$ th state during the evaluation time period

### 6.1.2 Simulation platform

The platform of the experiment is based on PARAMICS (PARAllel MICROscopic Simulation, version 2007.1) which is a package of microscopic simulation tools used to model the movement and behaviour of individual vehicles on urban and highway road networks. It provides the functions of detailed modelling for the traffic system conceivably. The characteristics of drivers, vehicles and the interactions between vehicles are simulated in a real time fashion. PARAMICS is fit to ITS studies due to its high performance, scalability and the ability of modelling the emerging ITS infrastructures, such as ILDs.

In the designed experiment, the travel time can be obtained by analysing the recorded log file which contains timing for each simulated vehicle coming in/out each link. This travel time value is used to compare against with the estimate results from the fusion framework, and then to calculate the metrics for performance evaluation. The detailed simulation scenario and sensor sources simulation will be introduced in the next two subsections.

### 6.1.3 Simulation Scenario

A stretch of road section is considered in the simulation. The whole section is divided into 8 segments with lengths of 400m, and is shown in Figure 6.1. Each segment is formed as link in the simulation. The main links are set to the type of major urban road with 2 lanes of 7.3 m wide while the ramp links are set to the type of minor urban road with 1 lane of 3.7 m wide. The speed limit for all of the links is 30 miles/hour. The entry and exit loops are placed to monitor in/out flow of the section, internal loops are placed at the each boundary between segments. All the loops are configured at 4 Hz sampling frequency which is the standard for UK urban ILD implementations. The targeted road section for travel time estimation is defined between entry loop and exit loop.

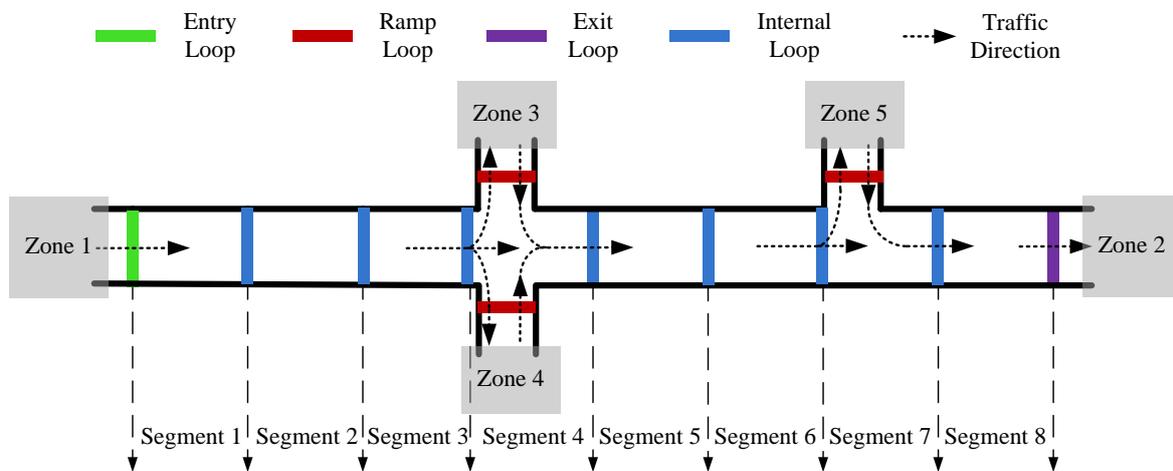


Figure 6.1 Simulation scenario

The reason of adopting the above road scenario (such as the junction and ILD configurations, speed limit and road link features) is to make it as similar as the real-world scenario described in the next chapter. To make the simulation more realistic, three junctions are also included in the simulation scenario and installed with ramp loops to record the turn in/out flow. There are 5 traffic demand zones, and the Origin-Destination (OD) matrix is summarised in Table 6.1. The simulation was configured to run for 24-hour period and the demand for peak and off-peak time period is multiplied by different scaling factors to reflect

the real traffic flow situation. The simulation is operated in the *stochastic demand mode*, resulting in the actual simulated demand fluctuating around the specific demand. The reason of adopting this configuration is to make the output from the simulation more realistic since the traffic demand in real-world is not perfectly deterministic. Figure 6.2 shows the generated flow pattern from entry and exit loops over the whole simulation period, and the result is aggregated at 5-min interval.

|        |        | Destination |        |        |        |        |       |
|--------|--------|-------------|--------|--------|--------|--------|-------|
|        |        | Zone 1      | Zone 2 | Zone 3 | Zone 4 | Zone 5 | Total |
| Origin | Zone 1 | -           | 1500   | 400    | 400    | 600    | 2900  |
|        | Zone 2 | -           | -      | -      | -      | -      | -     |
|        | Zone 3 | -           | 850    | -      | 400    | 500    | 1750  |
|        | Zone 4 | -           | 850    | 200    | -      | 600    | 1650  |
|        | Zone 5 | -           | 600    | -      | -      | -      | 600   |
|        | Total  | -           | 3800   | 600    | 800    | 1700   | 6900  |

Table 6.1 Origin-Destination (OD) traffic demand for the simulation scenario

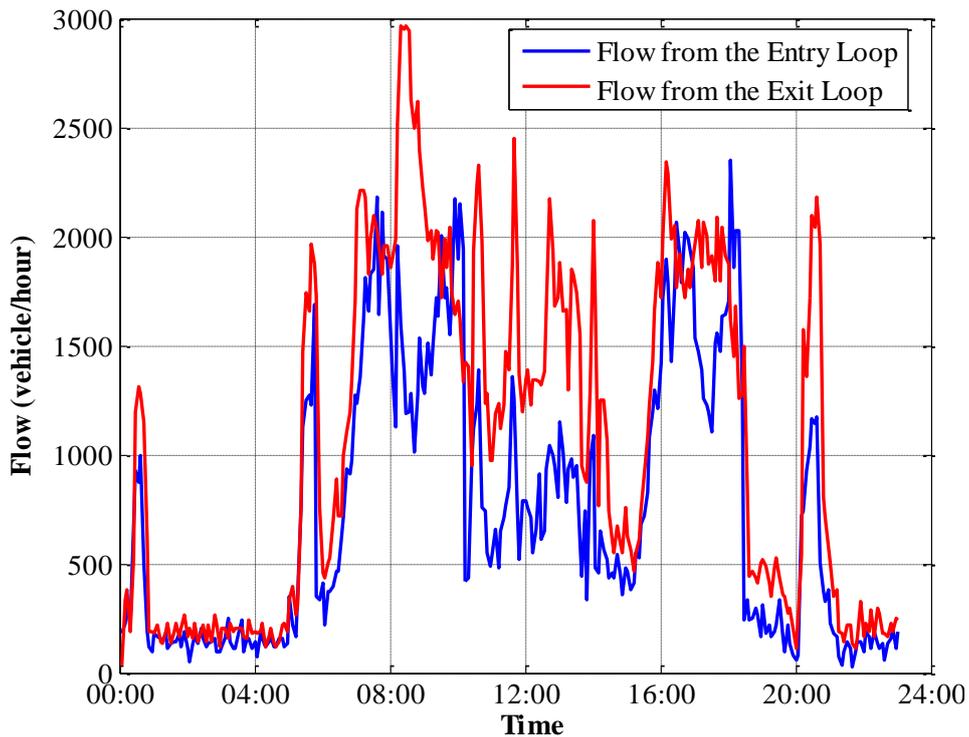


Figure 6.2 Flow pattern from the simulation period

### 6.1.4 Observations from the simulation

The traffic data such as travel time, density and flow is obtained by analysing the log file recorded based on the simulation scenario introduced above. In order to implement the fusion framework, the observations from ILDs, GPS probes and ANPR have to be generated, and

the noise associated with each of the sensor sources also needs to be modelled. This section introduces how the observations are generated from the sensor sources.

### **ILD**

In the PARAMICS simulation, the function of ILD is well developed as one of infrastructure component, and the flow observation from ILDs can be obtained directly from the loop log file. An example of these ILD observations is given in Figure 6.2. The entry, exit and ramp loops in the simulation scenario are considered as the key loops which are compulsory inputs to the fusion framework, while the internal loops are optionally selected to investigate the sensitivity of ILDs coverage and configuration. As discussed in Section 5.3.1, the noise term of ILD normally consists of two parts, unbiased systematic noise and biased noise due to lane-cross and low polling frequency. The mean of systematic noise is simply generated as the percentage of actual flow value, and its percentage value is set to 10%. For the biased noise, Krishnan (2008a) developed a custom program based on Visual Basic (VB) language to log the necessarily noised data automatically. His program interrogates the SNMP (Simple Network Management Protocol) interface of the PARAMICS model to record the cross-lane output of ILDs in the simulation at 4Hz. In this simulation, his program is used to introduce the biased noise into the flow observations which reflects the error resulting from cross lane effect.

### **GPS probe**

The position and corresponding timing information of each vehicle is available in PARAMICS simulation. Therefore, the GPS probe observations are simply modelled from a selection of the vehicle population. The selection is performed at a constant sampling rate in terms of the percentage of the whole vehicle population. In order to evaluate the sensitivity with respect to different sampling rate, the simulation is configured at different sampling rates from 0% to 30%. Another factor of GPS probe modelling is the GPS updating frequency, regarded as temporal resolution or T-resolution. The effect of this T-resolution on the fusion performance is also evaluated through sensitivity analysis by setting T-resolution from 1 second to 1 minute. An added Gaussian white noise is then added into the travel time observations from sampled vehicles to represent the GPS measurement noise. The standard derivation of this added noise is defined as the percentage of actual observed travel time. A range from 1% to 20% is used for the purposed of sensitivity evaluation. Figure 6.3-6.4 shows an example of the travel time observations from GPS probes at 15% sampling rate, 10 seconds T-resolution and 10% added noise. The result is aggregated at 5-min interval. It can

be seen that the observations from GPS probes reflect the overall trend of travel time but with a high fluctuation at each aggregated point, and the GPS observations at some of intervals are missed.

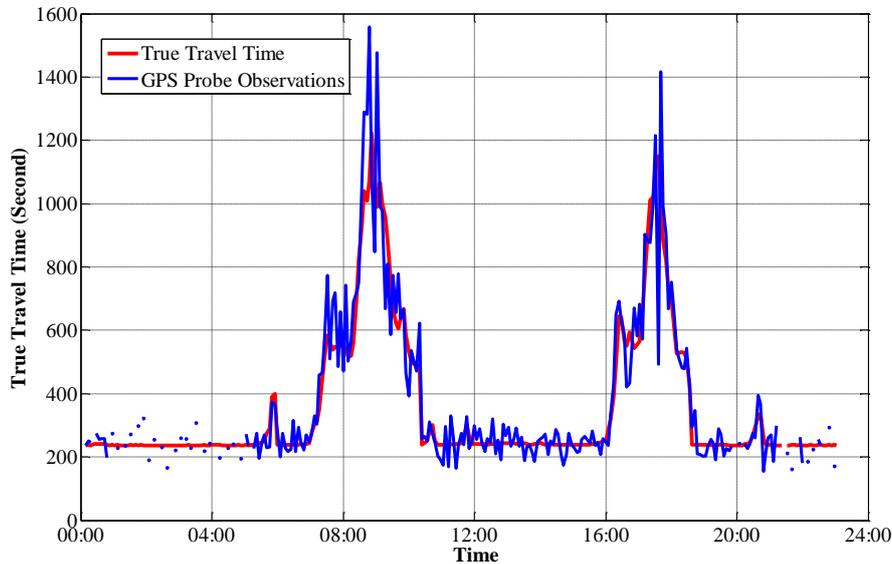


Figure 6.3 True travel time vs. GPS probe observations

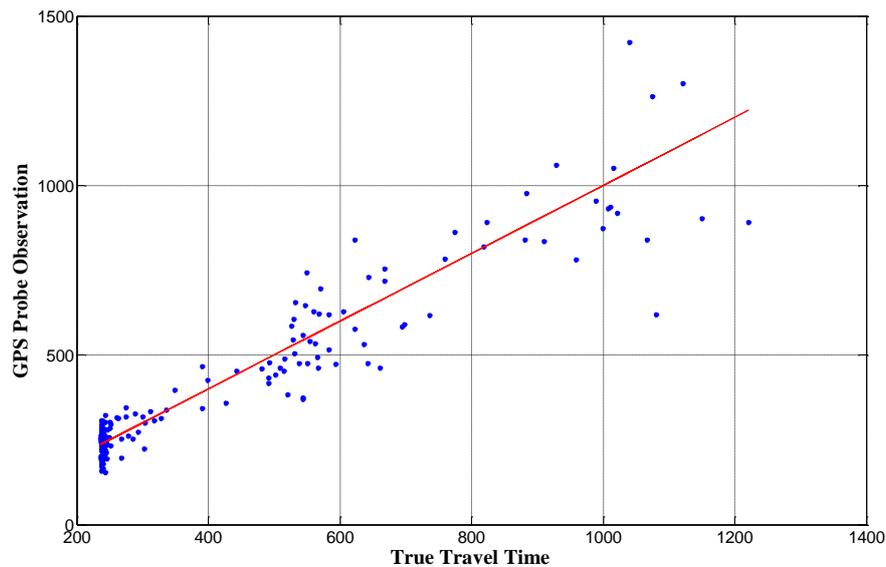


Figure 6.4 Scatter plot of true travel time vs. GPS probes observations

## ANPR

ANPR directly measures the travel time from a pair of camera locations. In the designed simulation scenario, the objective link for travel time estimation is defined as the road section between entry loop and exit loop. As mention above, the detailed location and timing information about each vehicle is available to the simulation, the travel time observations from ANPR are obtained by averaging the time which vehicles passing through the locations of entry and exit loop. The average recognition rate of the whole vehicle population is

determined by assigning different recognition rates to the long vehicle group (50%) and normal vehicle group (90%). By taking account of the proportion of each vehicle group (long: 12.4%; normal: 87.6%), the theoretical average recognition rate is 85.04% ( $12.4\% \times 50\% + 87.6\% \times 90\%$ ), and the actual recognition rate from the simulation data is 86.6%. This recognition rate is selected according to the studies by Wiggins (1999) which shows a recognition rate of 86% during a 2 hour survey in the UK. Regarding the level of accuracy in travel time observations from ANPR, there is very limited research in this area probably because of the difficulty of obtaining the data of ground truth travel time. To my best knowledge, the most relevant study about the accuracy of the travel time observations from ANPR is reported in Bertini et al. (2005). They showed an 11% relative error based on the Oregon Route 18 in west coast of U.S. while my simulated ANPR travel time has a relative error of 8.22%. It can be used to support that the baseline of ANPR accuracy in my comparison result is high enough.

Figure 6.5-6.6 shows the ANPR travel time observation at 86.6% recognition rate, and the result is aggregated at 5-min as same as the example of ILDs and GPS probes above. From these two figures, ANPR observations are much more accurate than the observations from GPS probes which is due to more samples of travel time are directly measured. It can be seen that there exists a slightly biased noise which is due to the sampling from different types of vehicles as discussed above and in Section 5.3.3. Although the ANPR observations show high accuracy in terms of travel time measurement, the coverage of ANPR camera system is very typically limited and is not available for wide spread application. The main part of performance evaluation will investigate the case for which only fuses ILDs and GPS data into the developed estimation framework.

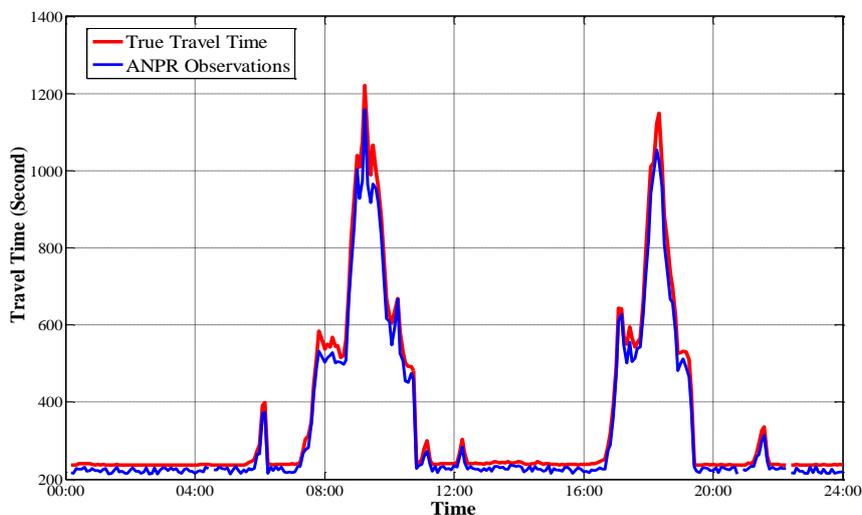


Figure 6.5 True travel time vs. ANPR observations

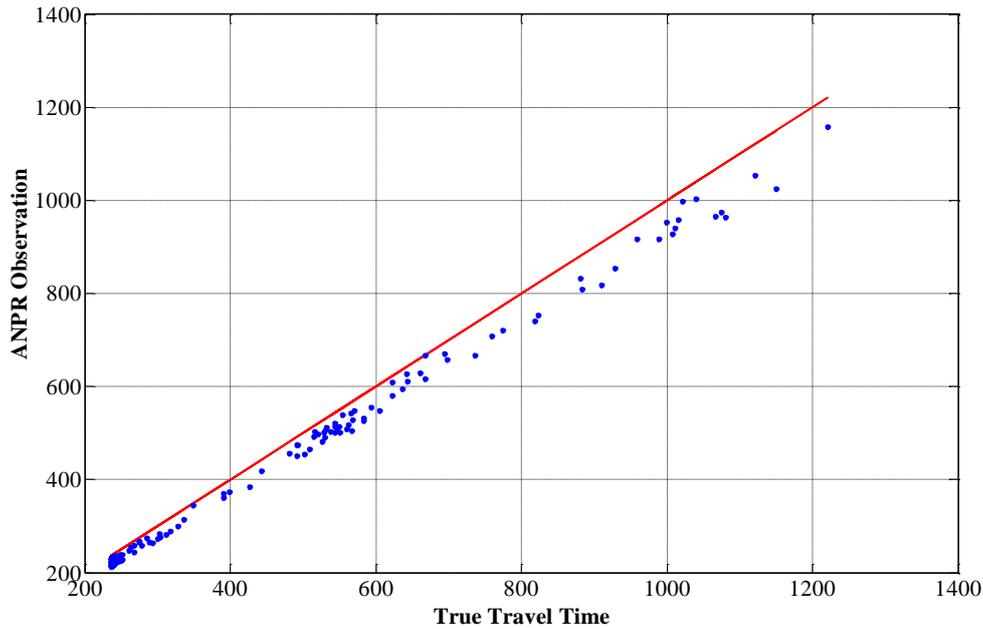


Figure 6.6 Scatter plot of true travel time vs. ANPR observations

In order to obtain the best achievable performance of ANPR travel time observations, the matched data is tested by a cleaning process which aims to remove the outlier observations. Although a number of methods have been developed to clean matched plate license data, Robinson & Polak (2006) demonstrated that Overtaking Rule (OR) approach is more efficient than other statistical based cleaning method. This section shows the result of applying the pre-filtering OR approaches to process the simulated ANPR data. The details about the theory of the pre-filtering OR approach refer to Robinson (2005). The parameters of the overtaking rule are configured as Table 6.2.

|  |     |
|--|-----|
| Number of following vehicles to include  | 50  |
| Tolerance time: uncongested (secs)   | 80  |
| Tolerance time: congested (secs)   | 800 |
| Travel time threshold for congested / uncongested                                      | 600 |
| Pre-filtering: percentage of fastest vehicles which are not allowed to exclude others. | 5%  |

Table 6.2 Overtaking rule settings used in the simulation data

The result of applying the OR approach to the ANPR data based on above parameters is given in Table 6.3. It can be seen that there is no significant improvement of applying cleaning approach, and the MPE and MAPE are even slightly worse. The limited improvement of RMSE and RMSPE is more relevant to the variance of the data. Filtering out very slow and fast vehicles reduces the error variance and makes the shape of the distribution

closer to the true travel time. However, the main error is introduced by excluding the large slow vehicles which have a low recognition rate rather than the outliers. Disallowing “abnormally” slow vehicles by OR even increases this type of bias of the ANPR data. The main purpose of using a cleaning treatment such as OR is to remove the vehicles which may take an alternative route, stop en-route or not attempt to travel in a prompt fashion. In the simulation scenario, none of these is included, thus it is not surprise that the cleaning treatment does not improve the result reasonably. Therefore, the evaluation based on the simulation still uses the ANPR data without cleaning. More analysis about applying different cleaning treatments into real ANPR data will be provided in Section 7.2.

|                                  | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|----------------------------------|-----------|------------|------------|-------------|
| <b>No Cleaning</b>               | 6.47%     | 8.22%      | 40.16      | 9.92%       |
| <b>Pre-filtering OR Approach</b> | 6.58%     | 8.29%      | 39.21      | 9.73%       |

Table 6.3 Performance comparison of ANPR observation between no cleaning and pre-filtering OR cleaning

## 6.2 Framework Implementation

### 6.2.1 The implementing structure for the simulation scenario

Section 5.4.3 introduced the general structure of the developed fusion framework. It provides the theoretical fundamentals for the applications of the Kalman filters. In order to implement the fusion framework into the simulation scenario, this general structure has to be specified in terms of state vector, state transition matrix and observation transition matrix.

#### State & observation vectors

The compact forms of state and observation vectors have been given in equation 5.45 and 5.53. For this case of implementation, these vectors are specified as

$$\mathbb{X} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{v} \\ \boldsymbol{q} \end{bmatrix} = [\rho_1 \ \cdots \ \rho_8 \ v_1 \ \cdots \ v_8 \ q_0 \ \cdots \ q_8 \ r_4 \ s_4 \ r_7 \ s_7]^T \quad (6.5)$$

$$\mathbb{Y} = \begin{bmatrix} \boldsymbol{q}_{ILD} \\ \boldsymbol{v}_{GPS} \\ \boldsymbol{v}_{ANPR} \end{bmatrix} = [l_1 \ \cdots \ l_L \ v_1^{GPS} \ \cdots \ v_M^{GPS} \ v_{ANPR}]^T \quad (6.6)$$

where

$\rho_i$ : the state of density at  $i$ th segment,  $i = 1 \cdots 8$

- $v_i$ : the state of speed at  $i$ th segment,  $i = 1 \cdots 8$
- $q_n$ : the state of flow at  $n$ th boundary,  $i = 0 \cdots 8$
- $r_4/r_7$ : the state of inflow at segment 4/7
- $s_4/s_7$ : the state of outflow at segment 4/7
- $l_j$ : the flow observation of  $j$ th ILD,  $i = 1 \cdots L$
- $v_m^{\text{GPS}}$ : the speed observation of the GPS probe for  $m$ th segment
- $v_{\text{ANPR}}$ : the speed observation of the ANPR for the whole section

The number of GPS probe observations  $M$  is not necessarily equal to the total number of segments (8 segments in this case), and more specifically:  $M \leq 8$ . It because the GPS probe observation may be not available for some of segments, i.e. the issue of missing observations (discussed in Section 5.5.3).  $L$  is the number of ILDs along the section. Note that the entry, exit and ramp loops are compulsory in this case, thus  $L \geq 5$ . If any of the boundaries is placed by an ILD, the observation from that ILD is added into the observation vector accordingly. For example, if the internal loops are available at boundary  $q_3$  and  $q_6$ , the total number of ILD observations is  $L = 7$ . The observation vectors then becomes

$$\mathbb{Y} = [l_1 \cdots l_7 \quad v_1^{\text{GPS}} \cdots v_M^{\text{GPS}} \quad v_{\text{ANPR}}]^T \quad (6.7)$$

In the on-line fusion process, the number of observation may change due to the issue of missing observation. This leads to a varying-dimension of observation vector while the dimension of state vector keeps constant.

In the rest of this section, this type of ILD configuration will be used to illustrate how the framework is implemented. The sensitivity analysis about the number and configuration of ILDs will be presented in the following section.

### State & observation process

The state process is introduced in Section 5.4.3, and summarised in equation 5.44. A compact form is recalled from equation 5.47 as follows

$$\mathbb{X}(k+1) = \mathcal{F}[\mathbb{X}(k), \mathbb{u}(k), \mathbb{w}(k)] \quad (6.9)$$

where state noise vector  $\mathbb{w}$  has same dimension as state vector  $\mathbb{X}$ , is given by

$$\mathbb{w} = \begin{bmatrix} \xi^\rho \\ \xi^v \\ \xi^q \end{bmatrix} = [\xi^{\rho_1} \cdots \xi^{\rho_8} \quad \xi^{v_1} \cdots \xi^{v_8} \quad \xi^{q_0} \cdots \xi^{q_8} \quad \xi^{r_4} \quad \xi^{s_4} \quad \xi^{r_7} \quad \xi^{s_7}]^T \quad (6.10)$$

$\mathbb{u}$  stands for the input control vector which contains the model parameters. In Section 5.5.1, it is discussed that the parameters  $\{v_f, \rho_{cr}, a\}$  are sensitive to the model accuracy while the rest of the model parameters  $\{\psi, \tau, c\}$  will be assumed known from an off-line calibration. In this case of implementation, two different schemes regarding the input control vector  $\mathbb{u}(k)$  will be used to testify the fusion performance. One is to calibrate all of the model parameters, and then  $\mathbb{u}(k)$  is given by

$$\mathbb{u} = [v_f, \rho_{cr}, a, \psi, \tau, c]^T \quad (6.11)$$

The second scheme is to consider the model sensitive parameters as system state, and the treatment introduced in Section 5.5.1 will be employed. By this means, state vector  $\mathbb{X}$  and  $\mathbb{u}$  become

$$\mathbb{X} = \begin{bmatrix} \rho \\ v \\ q \\ p \end{bmatrix} = [\rho_1 \cdots \rho_8 \quad v_1 \cdots v_8 \quad q_0 \cdots q_8 \quad r_4 \quad s_4 \quad r_7 \quad s_7 \quad v_f \quad \rho_{cr} \quad a]^T \quad (6.12)$$

$$\mathbb{u} = [\psi, \tau, c]^T \quad (6.13)$$

If the model parameters are considered as the system state such as equation 6.11, the noise vector  $\mathbb{w}$  is expanded by adding the noise terms of each parameter

$$\mathbb{w} = \begin{bmatrix} \xi^\rho \\ \xi^v \\ \xi^q \\ \xi^p \end{bmatrix} = [\xi^{\rho_1} \cdots \xi^{\rho_8} \quad \xi^{v_1} \cdots \xi^{v_8} \quad \xi^{q_0} \cdots \xi^{q_8} \quad \xi^{r_4} \quad \xi^{s_4} \quad \xi^{r_7} \quad \xi^{s_7} \quad \xi^{v_f} \quad \xi^{\rho_{cr}} \quad \xi^a]^T \quad (6.14)$$

The observation process has been modelled in equation 5.50 - 5.53, the compact form is recalled from equation 5.55 as follows

$$\mathbb{Y}(k) = \mathcal{H}[\mathbb{X}(k), \mathbb{v}(k)] \quad (6.15)$$

The noise vector of observation is specified by

$$\mathbb{v} = \begin{bmatrix} \gamma^{\text{ILD}} \\ \gamma^{\text{GPS}} \\ \gamma^{\text{ANPR}} \end{bmatrix} = [\gamma_1^{\text{ILD}} \cdots \gamma_7^{\text{ILD}} \quad \gamma_1^{\text{GPS}} \cdots \gamma_M^{\text{GPS}} \quad \gamma^{\text{ANPR}}] \quad (6.16)$$

## 6.2.2 The formulation of noise covariance matrices: R & Q

In order to implement the EKF and UKF algorithm, the noise covariance matrices R & Q (introduced in Section 4.2.1) need to be determined. Moreover, the selection of these two matrices is a key factor which affects the performance of filtering results. Although R & Q are important to the Kalman filter implementation, according to the review in this field, none of the existed literature investigates the form of these two matrices explicitly (Wang &

Papageorgiou 2005, Hegyi et al. 2006). This subsection will provide analytical forms regarding the structure of  $\mathbf{R}$  &  $\mathbf{Q}$  based on the simulation scenario.

The definition of  $\mathbf{R}$  is given by

$$\mathbf{R} = \mathbb{E}[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T] \quad (6.17)$$

For the case of  $\mathbf{w}$  in equation 6.15, the covariance matrix  $\mathbf{R}$  has the dimension of  $28 \times 28$ . Among the state variables in  $\mathbf{X}$ , the noise term of speed  $\mathbf{v}$  is due to the modelling inaccuracy and assumed to be white Gaussian distribution (see Section 5.4.2), and the noise term of flow  $\mathbf{q}$  and model parameters  $\mathbf{p}$  are constructed by auxiliary random walk processes whose distribution is zero-mean Gaussian white noise as well. Thus,

$$\xi^{q_i} = \xi^{s_i} = \xi^{r_i} \sim \mathcal{N}(0, \sigma_q^2) \quad (6.18)$$

$$\xi^{v_i} \sim \mathcal{N}(0, \sigma_v^2) \quad (6.19)$$

$$\xi^{v_f} \sim \mathcal{N}(0, \sigma_{v_f}^2) \quad (6.20)$$

$$\xi^{\rho_{cr}} \sim \mathcal{N}(0, \sigma_{\rho_{cr}}^2) \quad (6.21)$$

$$\xi^a \sim \mathcal{N}(0, \sigma_a^2) \quad (6.22)$$

Recalling the equation 5.60,

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (6.23)$$

the noise term of density  $\rho_i$  is associated with the noise of flow. In the simulation scenario, the  $r_i$  and  $s_i$  are only present at  $i = 4, 7$ . According to the property of the summation of normally distributed random variables:

$$\sigma_{X \pm Y}^2 = \sigma_X^2 + \sigma_Y^2 \quad (6.24)$$

the statistics of the density is given by

$$\xi^{\rho_i} \sim \begin{cases} \mathcal{N}\left(0, \frac{4T^2}{L_i^2} \sigma_q^2\right) & i = 4, 7 \\ \mathcal{N}\left(0, \frac{2T^2}{L_i^2} \sigma_q^2\right) & \text{otherwise} \end{cases} \quad (6.25)$$

Since each of the noise terms has zero mean, thus the covariance matrix  $\mathbf{R}$  is simplified as

$$\mathbf{R} = \mathbb{E}[\mathbf{w}\mathbf{w}^T] \quad (6.26)$$

According to the independency,

$$\mathbb{E}[\xi^{v_i} \xi^*] = \mathbb{E}[\xi^{\rho_i} \xi^*] = 0 \quad (6.27)$$

where  $*$  stands for the all the other elements in the noise vector  $\mathbf{w}$ . It can be seen from equation 6.23, that dependency exists between the density noise  $\xi^{\rho_i}$  and the noise terms from the boundaries of the  $i$ th segment, the on/off-ramps of the  $i$ th segment and the density of the  $(i-1)$ th segment:  $\{\xi^{q_{i-1}}, \xi^{q_i}, \xi^{s_i}, \xi^{r_i}, \xi^{\rho_{i-1}}\}$ . The derivations of them are given as follows

for  $i \neq 4, 7$

$$\mathbb{E}[\xi^{\rho_i} \xi^{q_{i-1}}] = \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i}) \right) \xi^{q_{i-1}} \right] = \frac{T}{L_i} \mathbb{E}[(\xi^{q_{i-1}})^2 - \xi^{q_{i-1}} \xi^{q_i}] = \frac{T}{L_i} \sigma_q^2 \quad (6.28)$$

$$\mathbb{E}[\xi^{\rho_i} \xi^{q_i}] = \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i}) \right) \xi^{q_i} \right] = \frac{T}{L_i} \mathbb{E}[\xi^{q_{i-1}} \xi^{q_i} - (\xi^{q_i})^2] = -\frac{T}{L_i} \sigma_q^2 \quad (6.29)$$

$$\mathbb{E}[\xi^{\rho_i} \xi^{s_i}] = \mathbb{E}[\xi^{\rho_i} \xi^{r_i}] = 0 \quad (6.30)$$

$$\begin{aligned} \mathbb{E}[\xi^{\rho_i} \xi^{\rho_{i-1}}] &= \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i}) \right) \left( \frac{T}{L_{i-1}} (\xi^{q_{i-2}} - \xi^{q_{i-1}}) \right) \right] \\ &= -\frac{T^2}{L_i L_{i-1}} \mathbb{E}[\xi^{q_{i-1}} \xi^{q_{i-2}} - (\xi^{q_{i-1}})^2 - \xi^{q_i} \xi^{q_{i-2}} + \xi^{q_i} \xi^{q_{i-1}}] \\ &= -\frac{T^2}{L_i L_{i-1}} \sigma_q^2 \end{aligned} \quad (6.31)$$

for  $i = 4, 7$

$$\begin{aligned} \mathbb{E}[\xi^{\rho_i} \xi^{q_{i-1}}] &= \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i} + \xi^{s_i} - \xi^{r_i}) \right) \xi^{q_{i-1}} \right] \\ &= \frac{T}{L_i} \mathbb{E}[(\xi^{q_{i-1}})^2 - \xi^{q_{i-1}} \xi^{q_i} + \xi^{s_i} \xi^{q_{i-1}} - \xi^{r_i} \xi^{q_{i-1}}] = \frac{T}{L_i} \sigma_q^2 \end{aligned} \quad (6.32)$$

$$\begin{aligned} \mathbb{E}[\xi^{\rho_i} \xi^{q_i}] &= \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i} + \xi^{s_i} - \xi^{r_i}) \right) \xi^{q_i} \right] \\ &= \frac{T}{L_i} \mathbb{E}[\xi^{q_{i-1}} \xi^{q_i} - (\xi^{q_i})^2 + \xi^{s_i} \xi^{q_i} - \xi^{r_i} \xi^{q_i}] = -\frac{T}{L_i} \sigma_q^2 \end{aligned} \quad (6.33)$$

$$\begin{aligned} \mathbb{E}[\xi^{\rho_i} \xi^{s_i}] &= \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i} + \xi^{s_i} - \xi^{r_i}) \right) \xi^{s_i} \right] \\ &= \frac{T}{L_i} \mathbb{E}[\xi^{q_{i-1}} \xi^{s_i} - \xi^{q_i} \xi^{s_i} + (\xi^{s_i})^2 - \xi^{r_i} \xi^{s_i}] = \frac{T}{L_i} \sigma_q^2 \end{aligned} \quad (6.34)$$

$$\begin{aligned} \mathbb{E}[\xi^{\rho_i} \xi^{r_i}] &= \mathbb{E} \left[ \left( \frac{T}{L_i} (\xi^{q_{i-1}} - \xi^{q_i} + \xi^{s_i} - \xi^{r_i}) \right) \xi^{r_i} \right] \\ &= \frac{T}{L_i} \mathbb{E}[\xi^{q_{i-1}} \xi^{r_i} - \xi^{q_i} \xi^{r_i} + \xi^{s_i} \xi^{r_i} - (\xi^{r_i})^2] \\ &= -\frac{T}{L_i} \sigma_q^2 \end{aligned} \quad (6.35)$$

Same as the case of  $i \neq 4, 7$

$$\mathbb{E}[\xi^{\rho_i} \xi^{\rho_{i-1}}] = -\frac{T^2}{L_i L_{i-1}} \sigma_q^2 \quad (6.36)$$

Therefore, a complete form of noise covariance matrix  $\mathbf{R}$  is given by

$$\mathbf{R} = \mathbb{E}[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T] = \mathbb{E}[\mathbf{w}\mathbf{w}^T] \quad (6.37)$$

where each position of the matrix is summarised as follows

$$\begin{cases} \mathbb{E}[\xi^{\rho_i} \xi^{\rho_{i-1}}] = -\frac{T^2}{L_i L_{i-1}} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{q_{i-1}}] = \frac{T}{L_i} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{q_i}] = -\frac{T}{L_i} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{s_i}] = \begin{cases} \frac{T}{L_i} \sigma_q^2 & i = 4, 7 \\ 0 & i \neq 4, 7 \end{cases} \\ \mathbb{E}[\xi^{\rho_i} \xi^{r_i}] = \begin{cases} \frac{-T}{L_i} \sigma_q^2 & i = 4, 7 \\ 0 & i \neq 4, 7 \end{cases} \end{cases} \quad \begin{cases} \mathbb{E}[(\xi^{\rho_i})^2] = \begin{cases} \frac{4T^2}{L_i^2} \sigma_q^2 & i = 4, 7 \\ \frac{2T^2}{L_i^2} \sigma_q^2 & i \neq 4, 7 \end{cases} \\ \mathbb{E}[(\xi^{q_i})^2] = \sigma_q^2 \\ \mathbb{E}[(\xi^{v_i})^2] = \sigma_v^2 \\ \mathbb{E}[(\xi^{v_f})^2] = \sigma_{v_f}^2 \\ \mathbb{E}[(\xi^{\rho_{cr}})^2] = \sigma_{\rho_{cr}}^2 \\ \mathbb{E}[(\xi^a)^2] = \sigma_a^2 \\ \text{others} = 0 \end{cases} \end{cases} \quad (6.38)$$

By this means, the explicit form of state noise covariance matrix  $\mathbf{R}$  can be determined. It can be seen that the correlations among state variables are carefully considered.

On the other hand, the definition of  $\mathbf{Q}$  is given by

$$\mathbf{Q} = \mathbb{E}[(\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^T] \quad (6.39)$$

The GPS probe and ANPR observations are zero mean white normally distributed. Using the treatment of biased noise in Section 5.5.2, the observations from ILDs are also considered as zero mean. The observation covariance  $\mathbf{Q}$  is then simplified as

$$\mathbf{Q} = \mathbb{E}[\mathbf{v}\mathbf{v}^T] \quad (6.39)$$

The statistical properties of observation noise are

$$\gamma_i^{\text{ILD}} \sim \mathcal{N}(0, \sigma_{\text{ILD}}^2) \quad (6.40)$$

$$\gamma_i^{\text{GPS}} \sim \mathcal{N}(0, \sigma_{\text{GPS}}^2) \quad (6.41)$$

$$\gamma^{\text{ANPR}} \sim \mathcal{N}(0, \sigma_{\text{ANPR}}^2) \quad (6.42)$$

Since all of the noise terms of observations are pairwise independent, the form of the  $\mathbf{Q}$  is shown as

$$\mathbf{Q} = \begin{pmatrix} \gamma_1^{\text{ILD}} & \dots & \gamma_7^{\text{ILD}} & \gamma_1^{\text{GPS}} & \dots & \gamma_M^{\text{GPS}} & \gamma^{\text{ANPR}} \\ \sigma_{\text{ILD}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\text{ILD}}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\text{GPS}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\text{GPS}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{ANPR}}^2 \end{pmatrix} \begin{cases} \gamma_1^{\text{ILD}} \\ \dots \\ \gamma_7^{\text{ILD}} \\ \gamma_1^{\text{GPS}} \\ \dots \\ \gamma_M^{\text{GPS}} \\ \gamma^{\text{ANPR}} \end{cases} \quad (6.43)$$

## 6.3 Fusion Performance

### 6.3.1 Parameters

The structure and all the necessary variables have been well formulated in the previous section. It is quite straightforward to apply the EKF & UKF algorithms (equation 4.16-4.24 & equation 4.33-4.49) into the constructed framework. The parameters used are summarised in Table 6.4.

|  |
|--|
| $\tau = 10 \text{ (s)}, \psi = 18 \text{ (km}^2\text{/h)}, c = 5 \text{ (veh/km)}$ $\text{(calibrated by one day data)}$ |
| $\sigma_q^2 = 100 \text{ (veh/h)}^2$   |
| $\sigma_v^2 = 3 \text{ (km/h)}^2$  |
| $\sigma_{v_f}^2 = 0.3 \text{ (km/h)}^2$  |
| $\sigma_{\rho_{cr}}^2 = 10^{-2} \text{ (veh/km)}^2$  |
| $\sigma_a^2 = 10^{-4}$   |

Table 6.4 The parameters used for framework implementation

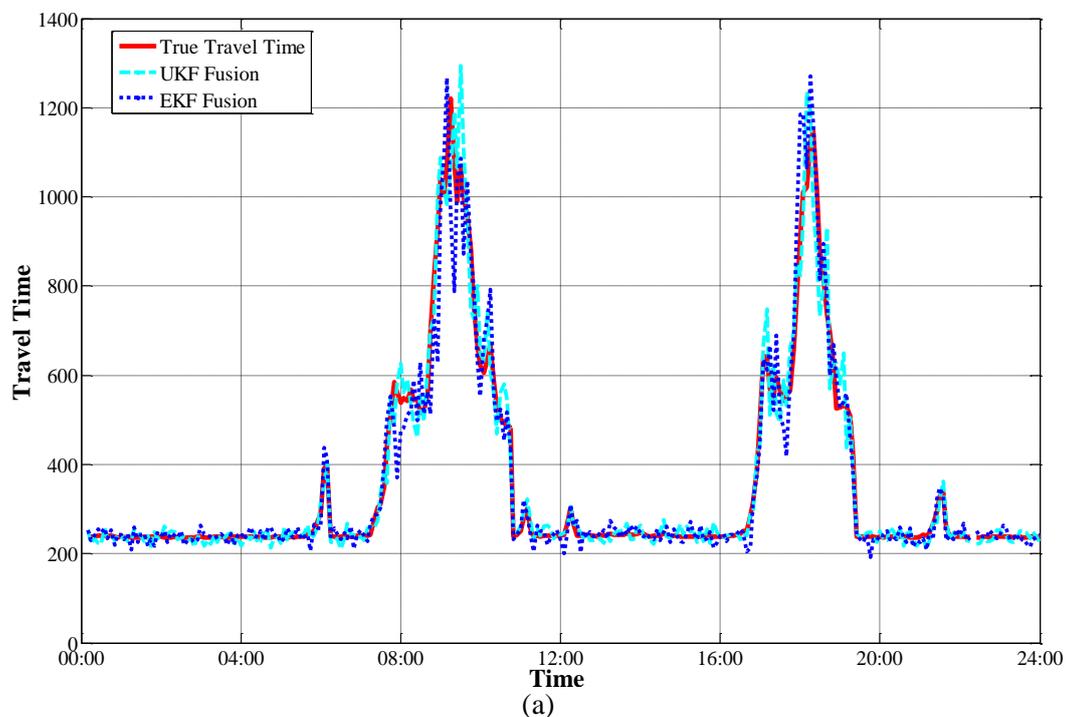
In the application of the UKF, the sigma points generated by the algorithms may not be physically meaningful, such as negative density. In order to eliminate this occurrence, limits for the upper and lower bound of states and parameters are imposed, see Table 6.5.

|   |
|---|
| $0 \text{ (veh/km)} \leq \rho_i \leq 160 \text{ (veh/km)}$    |
| $8 \text{ (km/h)} \leq v_i \leq 80 \text{ (km/h)}$            |
| $35 \text{ (km/h)} \leq v_f \leq 65 \text{ (km/h)}$           |
| $25 \text{ (veh/km)} \leq \rho_{cr} \leq 70 \text{ (veh/km)}$ |
| $1 \leq a \leq 3$   |

Table 6.5 Bounds of the states and parameters for the UKF

### 6.3.2 Results of General fusion

This subsection shows the fusion results from ILD, GPS and ANPR. The observations are from 7 ILDs including the on/off ramps, and are corrupted with 10% noise. The GPS probes are sampled at 15% rate with T-resolution of 10 seconds and 10% Gaussian white noise, and ANPR has a 70% matching rate. Figure 6.7 (a)-(b) gives the overall results of the EKF and UKF fusion against the true travel time, and the scatter plots of each of them are given in Figure 6.7 (c)-(d). It can be seen that during the uncongested/light traffic period, the UKF and EKF show a quite similar fusion result. For the congestion period, the UKF fusion outperforms the EKF. More specifically, the error variance of the UKF is smaller than the EKF. It is clear from the scatter plot and error plot figures that the UKF fusion points of large travel time (congestion period) are less dispersive than the EKF results. This finding is coherent with the discussion in Section 4.3.4. During the congested period, the state process tends to be highly nonlinear. The linearisation process of the EKF (Jacobians matrix) and the UKF (sigma points) becomes the key to the filtering performance. By the comparison discussion in Section 4.3.4, the UKF is able to capture more nonlinear information than the EKF, which leads to better accuracy of the UKF than EKF. The superiority of the UKF over the EKF is also verified in the metrics of fusion performance (provided in Table 6.4). Therefore, UKF is selected as the optimal fusion algorithm for the rest of analysis.



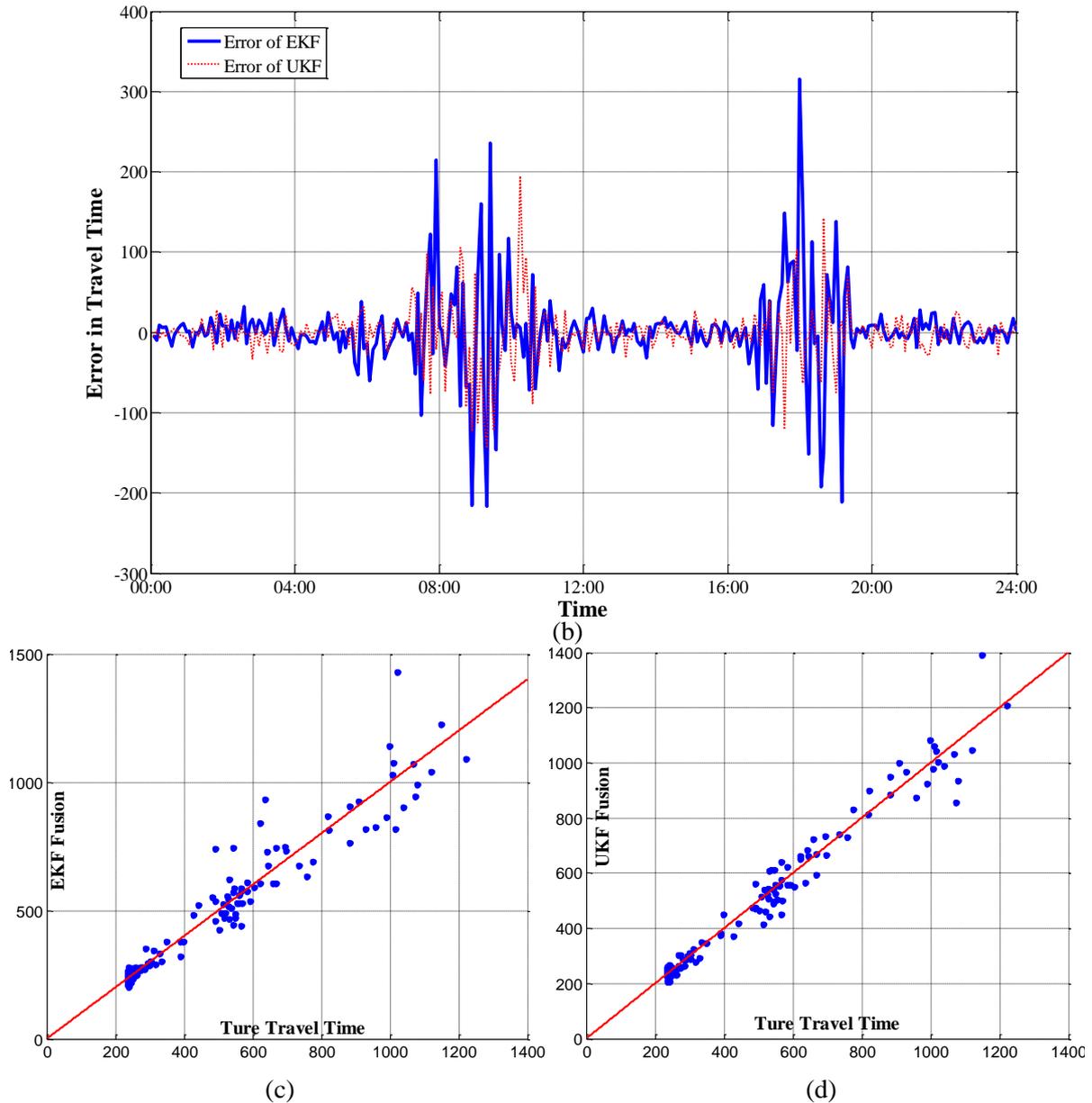


Figure 6.7 EKF and UKF fusion results: (a) time plot; (b) error plot; (c) EKF scatter plot; (d) UKF scatter plot

Figure 6.8 illustrates the comparison between using UKF fusion and only GPS probe or ANPR observations. The fusion result is better than any of the single type of the observations without doubt. It can be seen that the biased error of ANPR is reduced as expected, i.e. the MPE is reduced from 6.47% to 0.57%. However, the improvement of fusion over ANPR observation is relatively limited. It is because the ANPR observations are simulated at 70% matching rate which observes the travel time accurate enough. The result presented in this section is to show the performance of general fusion from all three sensor types. In practice, the ANPR observation will not be available widely, the evaluation on only fusing ILDs and GPS probes data will provide in next section.

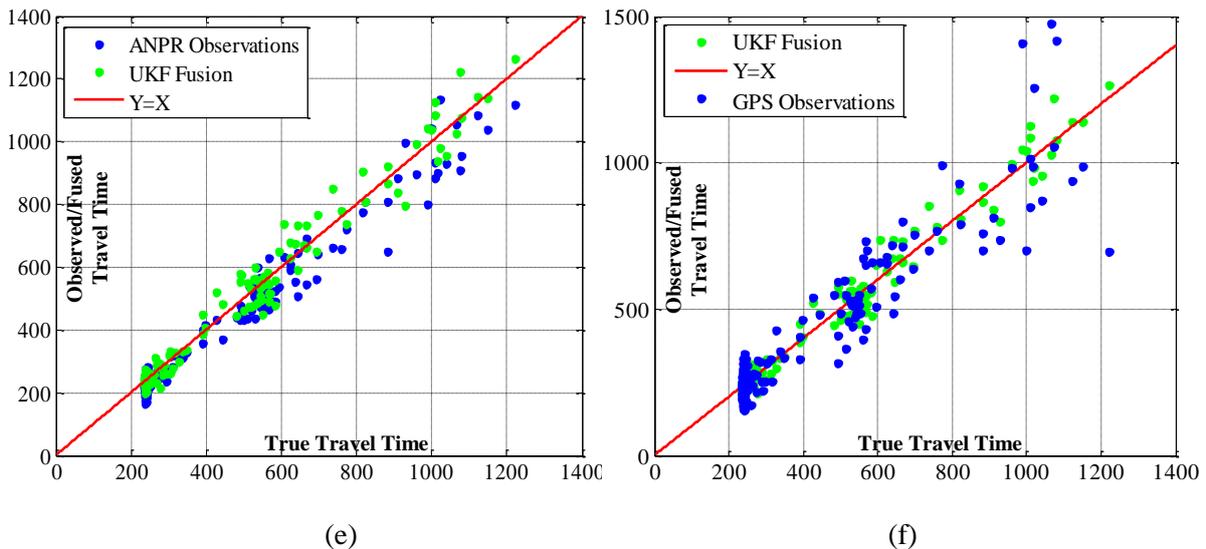
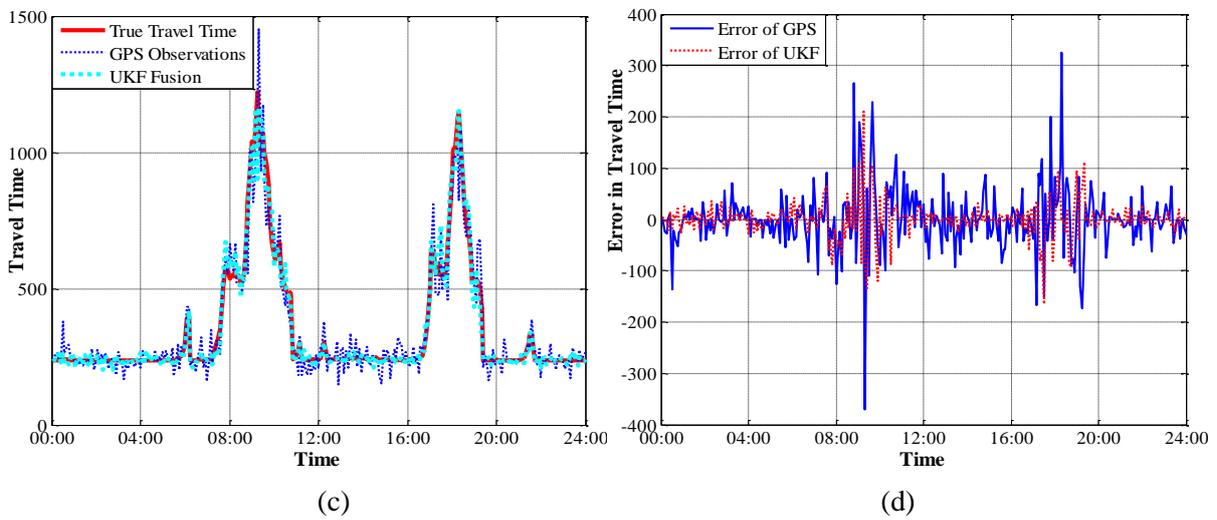
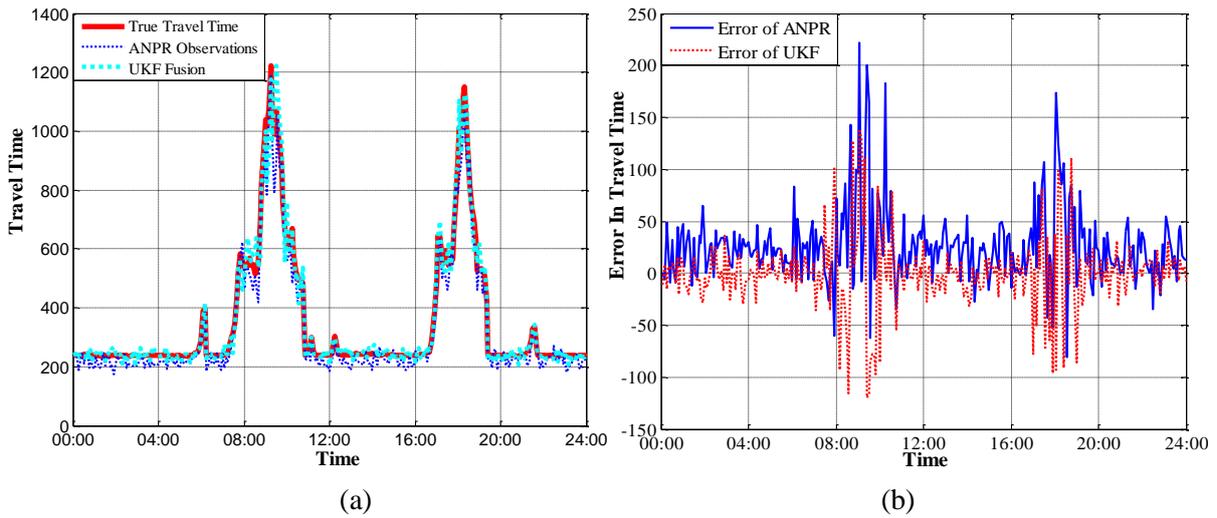


Figure 6.8 Comparison between UKF fusion and single sensor observations: (a) time plot of UKF and ANPR; (b) error plot of UKF and ANPR; (c) time plot of UKF and GPS; (d) error plot of UKF and GPS; (e) scatter plot of UKF and ANPR; (f) scatter plot of UKF and GPS

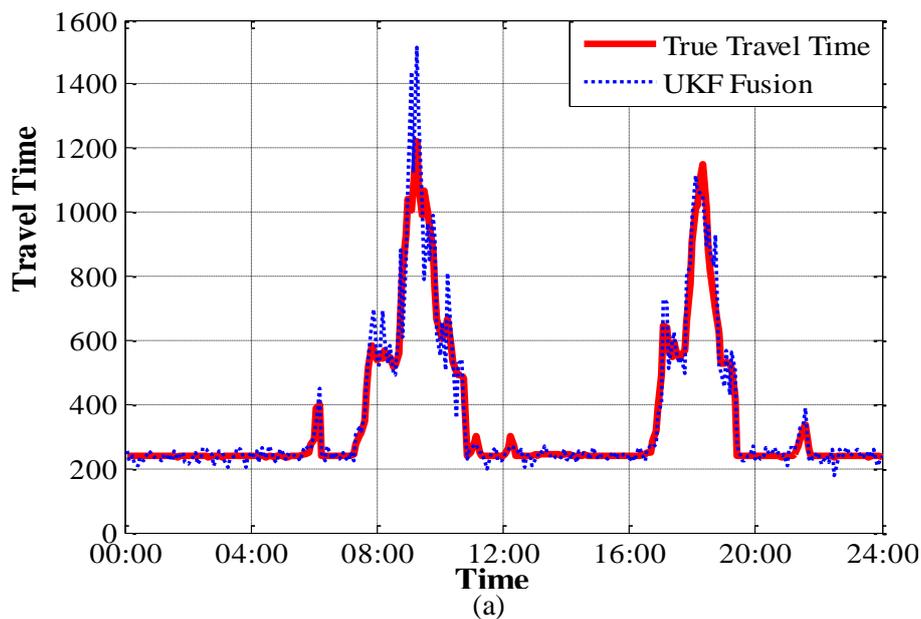
|                              | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|------------------------------|-----------|------------|------------|-------------|
| <b>EKF Fusion</b>            | 0.11%     | 6.57%      | 53.61      | 9.14%       |
| <b>UKF Fusion</b>            | 0.57%     | 4.69%      | 33.81      | 6.13%       |
| <b>ANPR Observation</b>      | 6.47%     | 8.22%      | 40.16      | 9.92%       |
| <b>GPS probe Observation</b> | -0.42%    | 11.46%     | 72.74      | 14.85%      |

7 ILDs with 10% additive noise; ANPR with 86.6% matching rate; 15% GPS probes rate with 10 second T-resolution and 10% additive noise

Table 6.6 The performance of general fusion

### 6.3.3 Results of GPS and ILDs Fusion by UKF

This section shows the fusion performance by only considering the GPS and ILDs data. The variable of ANPR is removed from the observation vector (equation 6.6), and the observation process and noise vector are also altered accordingly. The configurations of ILDs and GPS probes are same as the previous section. The fusion result is illustrated in Figure 6.9, and Table 6.5 provides the metrics for quantitative evaluation. This result indicates that the fusion performance is degraded without ANPR observations. However, there is only limited degradation comparing to the case when ANPR is available, i.e. the MAPE increases from 4.69% to 7.09%. Note that the GPS probes are simulated in an ideal way in terms of sampling rate, T-resolution and noise level. The sensitivity analysis with respect to the GPS errors will be provided in the Section 6.4.1.



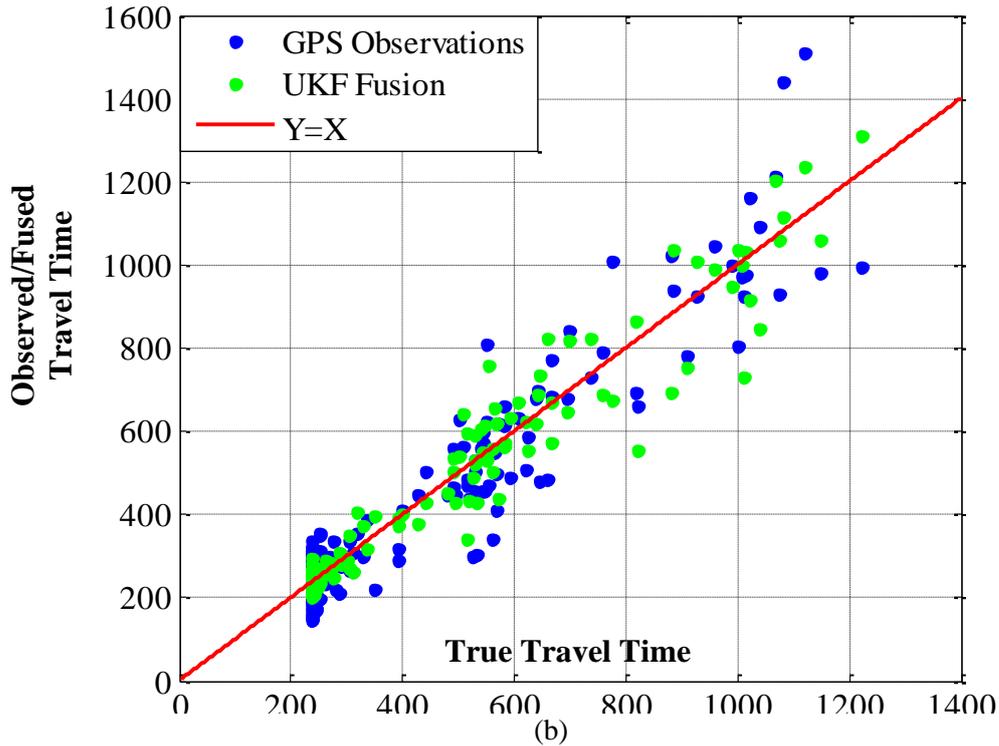


Figure 6.9 ILDs and GPS UKF fusion results

|  | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|--|-----------|------------|------------|-------------|
| <b>UKF Fusion from all three sensor types</b>  | 0.57%     | 4.69%      | 33.81      | 6.13%       |
| <b>UKF Fusion from ILDs and GPS</b>  | -0.92%    | 7.09%      | 59.76      | 9.44%       |
| <b>GPS probe observation</b>   | -0.42%    | 11.46%     | 72.74      | 14.85%      |
| 7 ILDs with 10% additive noise; 15% GPS probes rate with 10 second T-resolution and 10% additive noise |           |            |            |             |

Table 6.7 The performance of ILDs and GPS fusion

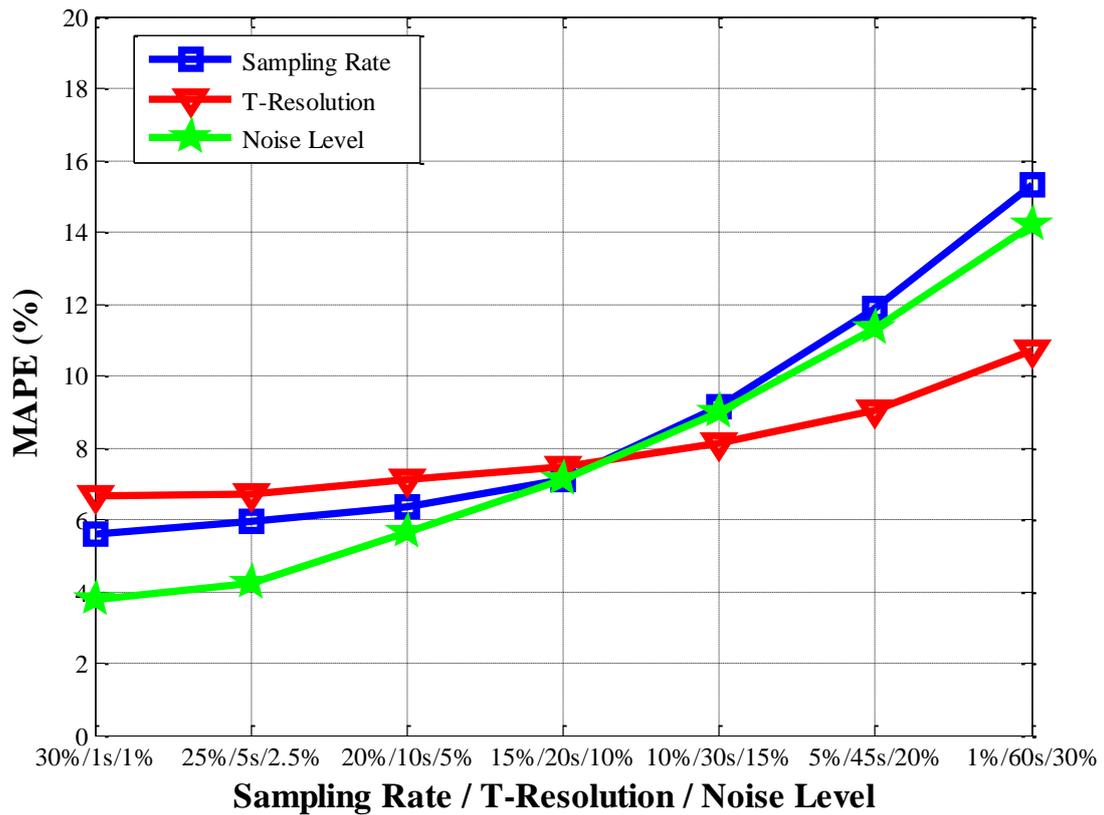
## 6.4 Results of Sensitivity Analysis

### 6.4.1 Sensitivity to GPS probe

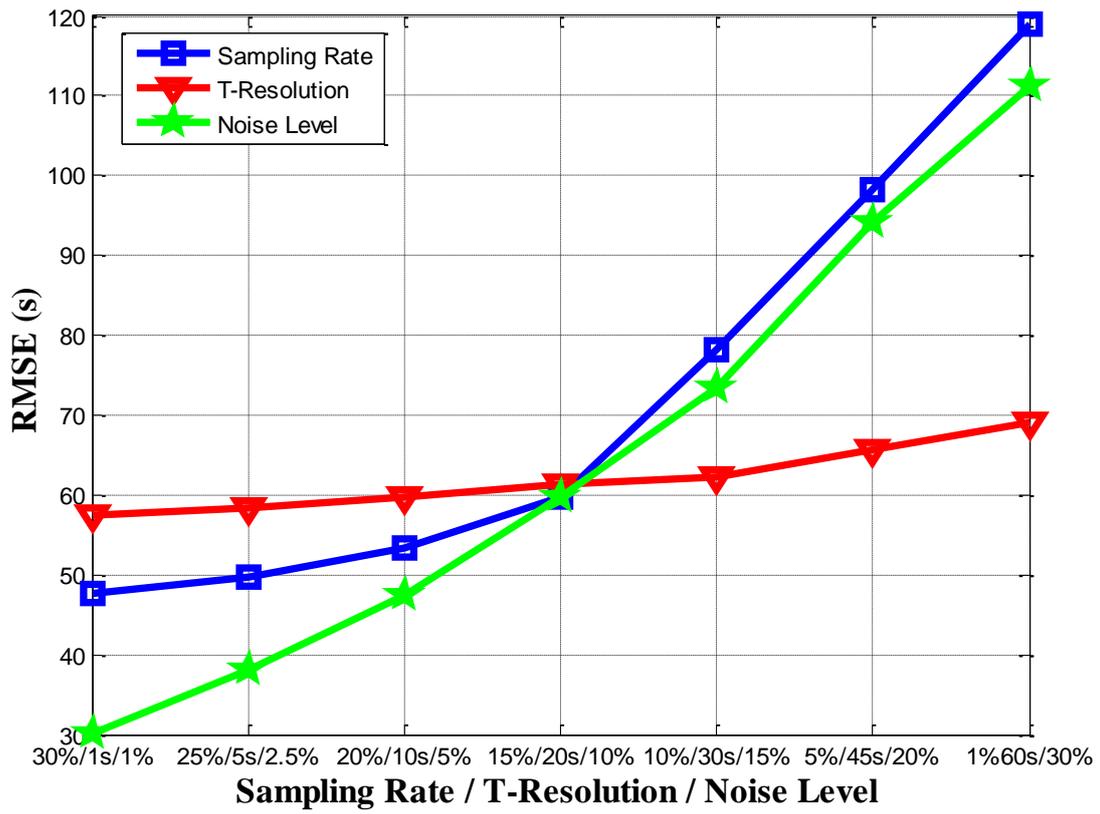
The fusion results presented in the previous section are based on determined features of GPS observations, i.e., 15% sampling rate and 10 seconds T-resolution with 10% added noise. This subsection will analyse the sensitivity of the fusion performance with respect to these three factors. A setting for the values of each factor is given in Table 6.6, and the results of the sensitivity analysis are shown in Figure 6.10 and Table 6.7. Note that this sensitivity analysis is based on the fusion from ILDs and GPS, and the configuration of ILDs is same as above.

|                      |      |      |       |       |       |       |       |
|----------------------|------|------|-------|-------|-------|-------|-------|
| <b>Sampling rate</b> | 1%   | 5%   | 10%   | 15%   | 20%   | 25%   | 30%   |
| <b>T-resolution</b>  | 1(s) | 5(s) | 10(s) | 20(s) | 30(s) | 45(s) | 60(s) |
| <b>Noise level</b>   | 1%   | 2.5% | 5%    | 10%   | 15%   | 20%   | 30%   |

Table 6.8 The settings for the parameters of GPS probes



(a)



(b)

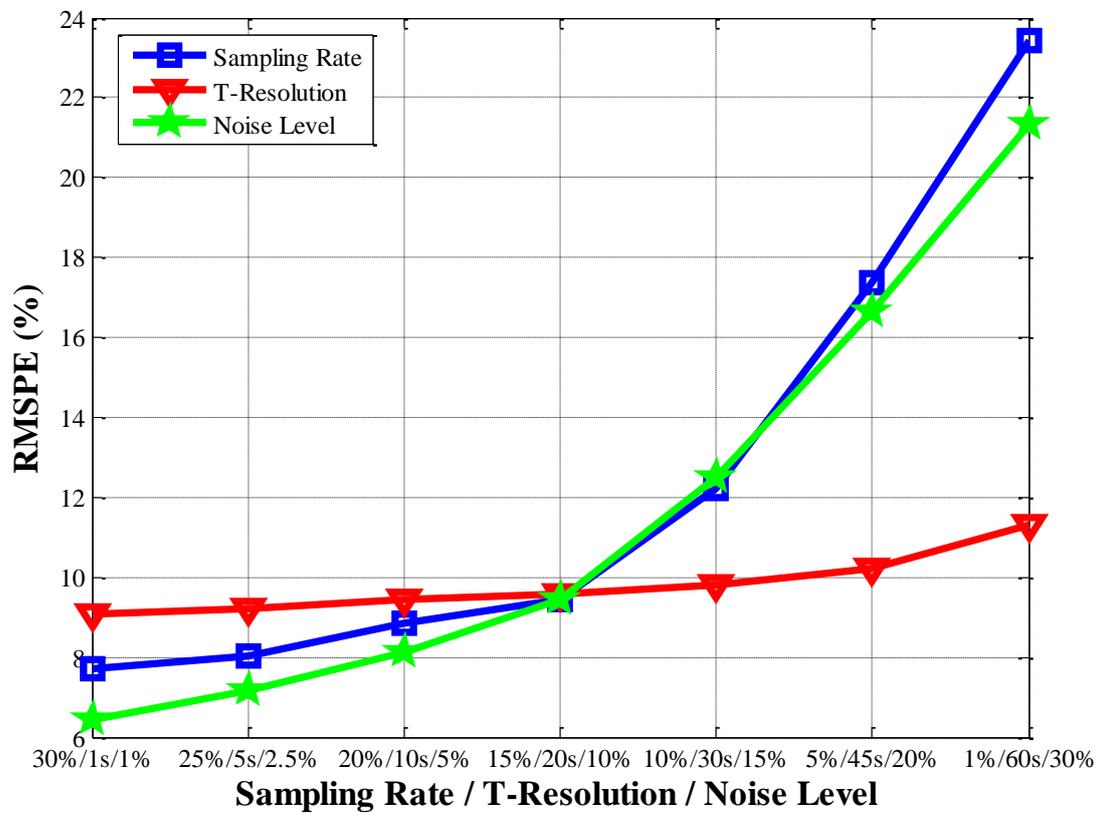


Figure 6.10 Result of sensitivity analysis for GPS probes: (a) MAPE result; (b) RMSE result; (c) RMSPE result

| <b>Sampling Rate</b>  | <b>1%</b> | <b>5%</b>   | <b>10%</b> | <b>15%</b> | <b>20%</b> | <b>25%</b> | <b>30%</b> |
|---|-----------|-------------|------------|------------|------------|------------|------------|
| $J_{MPE}$   | 3.30%     | 1.98%       | -1.83%     | -0.92%     | -1.30%     | 1.27%      | 0.82%      |
| $J_{MAPE}$  | 15.34%    | 11.87%      | 9.15%      | 7.09%      | 6.33%      | 5.94%      | 5.60%      |
| $J_{RMSE}$  | 119.03s   | 98.32s      | 78.14s     | 59.76s     | 53.45s     | 49.66s     | 47.75s     |
| $J_{RMSPE}$   | 23.43%    | 17.39%      | 13.23%     | 9.44%      | 8.86%      | 8.02%      | 7.71%      |
| 7 ILDs with 10% additive noise; 10 second T-resolution and 10% additive noise |           |             |            |            |            |            |            |
| <b>T-resolution (s)</b>   | <b>1</b>  | <b>5</b>    | <b>10</b>  | <b>20</b>  | <b>30</b>  | <b>45</b>  | <b>60</b>  |
| $J_{MPE}$   | 0.73%     | -0.81%      | -0.92%     | 1.45%      | -1.96%     | 2.33%      | 2.93%      |
| $J_{MAPE}$  | 6.68%     | 6.71%       | 7.09%      | 7.47%      | 8.13%      | 9.02%      | 10.69%     |
| $J_{RMSE}$  | 57.56s    | 58.31s      | 59.76s     | 61.34s     | 62.21s     | 65.62s     | 69.00s     |
| $J_{RMSPE}$   | 9.09%     | 9.23%       | 9.44%      | 9.56%      | 9.82%      | 10.22%     | 11.32%     |
| 7 ILDs with 10% additive noise; 15% sampling rate and 10% additive noise      |           |             |            |            |            |            |            |
| <b>Noise Level</b>  | <b>1%</b> | <b>2.5%</b> | <b>5%</b>  | <b>10%</b> | <b>15%</b> | <b>20%</b> | <b>30%</b> |
| $J_{MPE}$   | 0.12%     | -0.34%      | -0.62%     | -0.92%     | 1.42%      | -2.13%     | -2.52%     |
| $J_{MAPE}$  | 3.77%     | 4.25%       | 5.63%      | 7.09%      | 8.98%      | 11.32%     | 14.20%     |
| $J_{RMSE}$  | 30.02s    | 38.11s      | 47.33s     | 59.76s     | 73.32s     | 94.18s     | 111.30s    |
| $J_{RMSPE}$   | 6.42%     | 7.16%       | 8.14%      | 9.44%      | 12.48%     | 16.63%     | 21.32%     |
| 7 ILDs with 15% sampling rate and 10s T-resolution                            |           |             |            |            |            |            |            |

Table 6.9 Quantitative results of sensitivity analysis for GPS probes

From the results of Figure 6.10, it can be seen that the fusion performance is not sensitive to the T-resolution, while the other two factors, sampling rate and noise level, have obvious impact on the fusion performance generally. The T-resolution is actually the time interval of GPS samples. The average travel time for each of segment is around 45 seconds in the simulation scenario. The resolution which is smaller than 45 seconds will not improve the fusion performance significantly. For the sensitivity to the sampling rate, although the result shows a relatively large difference from a rate of 1% to 30%, the performance tends to be converged when the sampling rate increases to around 15-20%. It indicates that continue increasing the percentage of GPS probes only achieves limited improvement. From the above figure, the level of noise affects the fusion performance linearly. One example of UKF fusion result with “worse” GPS probe observations (sampling rate at 5%, 30s T-resolution and 20%

added noise) is shown in Figure 6.11. Comparing with the result in Figure 6.9, a much clearer improvement can be seen.

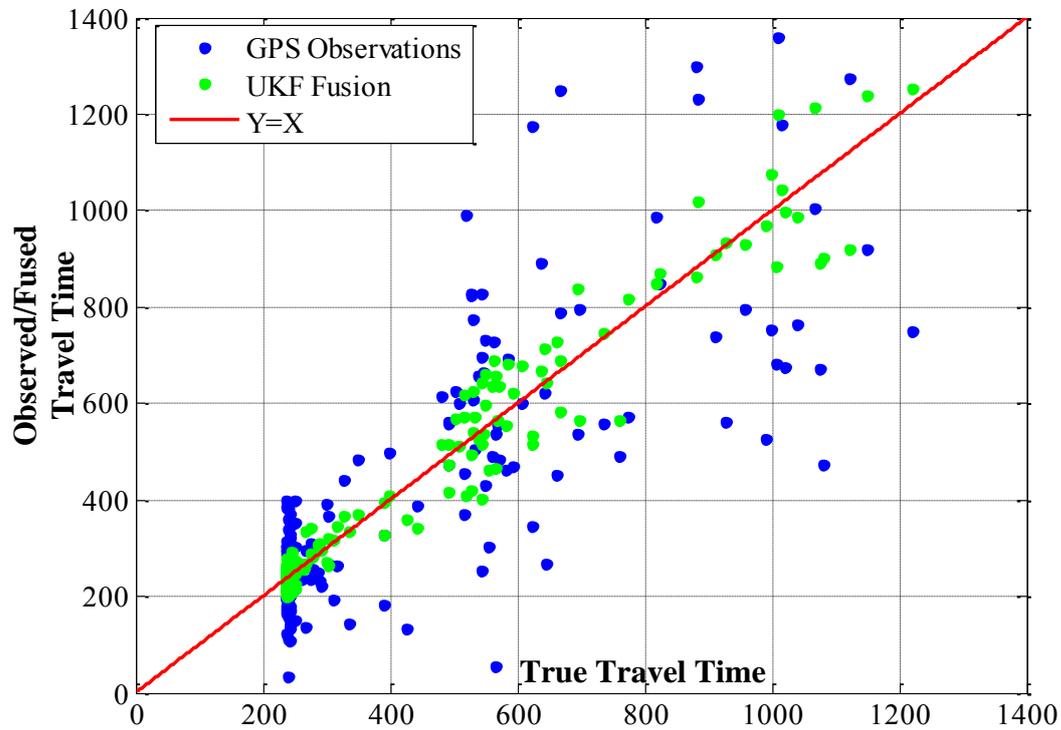


Figure 6.11 UKF fusion vs. GPS observations at GPS probe sampling rate at 5%, 30s T-resolution and 20% added noise

### 6.4.2 Sensitivity to ILD data

The above subsection illustrated how the factors of GPS probes affect the fusion performance. These computations are all based on the traffic flow observations from 7 ILDs which are placed at the entry, exit, 3 ramps and 2 upstream loops of junctions. To analyse the fusion performance with respect to the number and locations of the internal ILDs, the following ILDs configurations are evaluated by UKF fusion:

- (1) C-1: all the internal loops are available
- (2) C-2: both of junction upstream and downstream ILDs are available (L3, L4, L5, L6)
- (3) C-3: only junction upstream ILDs are available (L3, L6)
- (4) C-4: only junction downstream ILDs are available (L4, L7)
- (5) C-5: the non-junction ILDs are available while no junction ILDs (L1, L2, L5)

The notations  $\{L1, L2, \dots, L7\}$  indicates the locations of the loops which are shown in Figure 6.12. The factors of GPS probe are set at 10% sampling rate, 30 seconds T-resolution and 10% noise level. The evaluation results are presented in Figure 6.13 and Table 6.8.

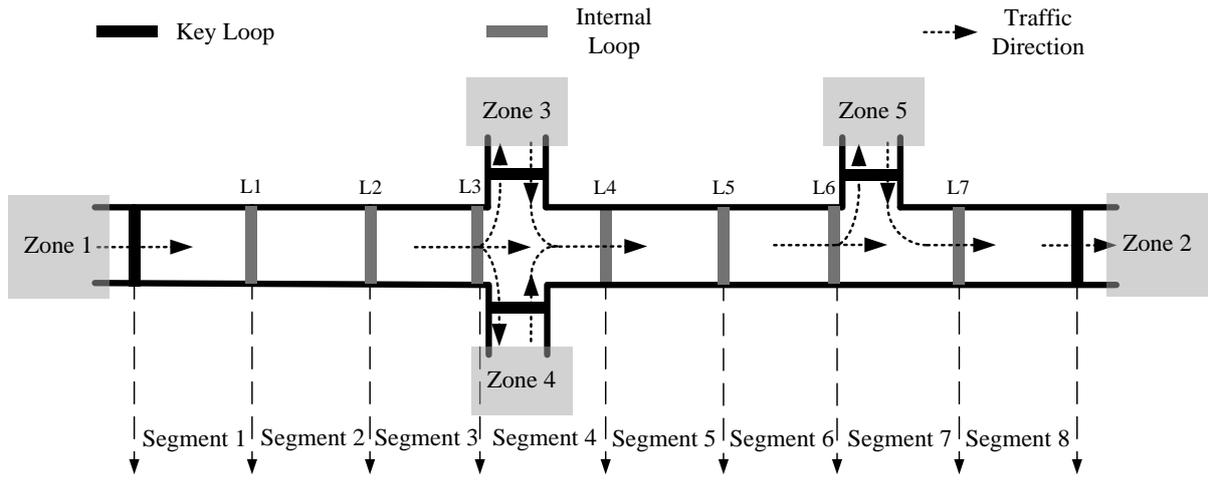


Figure 6.12 ILDs configurations for sensitivity analysis

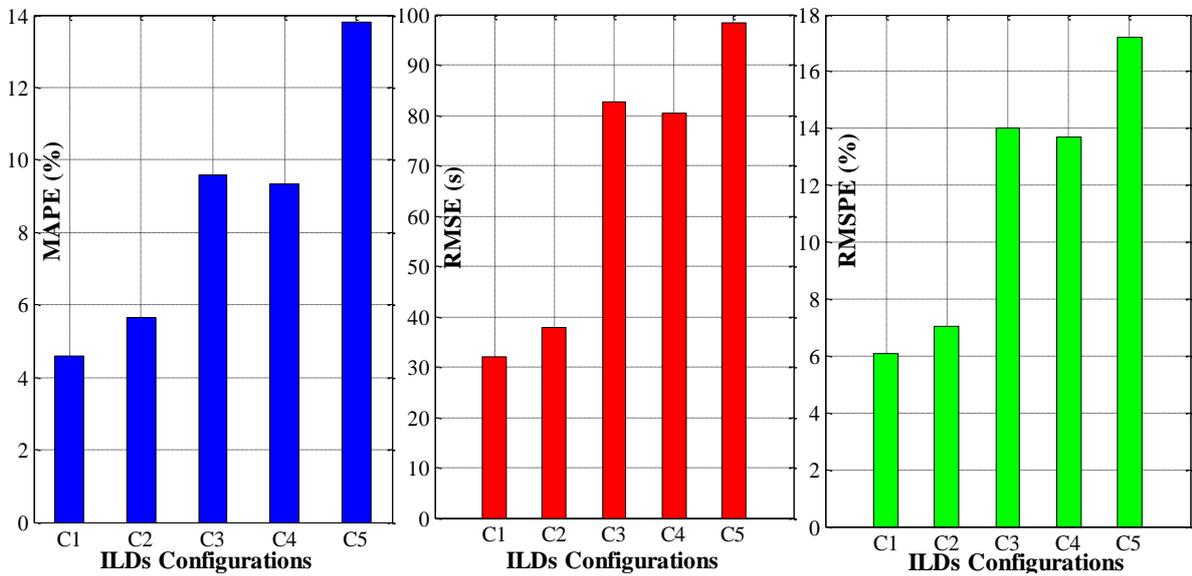


Figure 6.0.13 Result of sensitivity analysis for ILDs

| ILDs Configuration | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|--------------------|-----------|------------|------------|-------------|
| C1                 | -0.44%    | 4.58%      | 32.02      | 6.09%       |
| C2                 | -0.73%    | 5.67%      | 37.85      | 7.03%       |
| C3                 | -1.87%    | 9.59%      | 82.74      | 14.02%      |
| C4                 | 1.32%     | 9.33%      | 80.48      | 13.71%      |
| C5                 | 2.65%     | 13.79%     | 98.30      | 17.22%      |

10% GPS probes rate with 30 second T-resolution and 10% additive noise; 10% ILDs noise

Table 6.10 Quantitative results of sensitivity analysis for ILDs

From the results of sensitivity evaluation, it can be seen that the loops near the junctions are key to the fusion performance. The difference between C1 and C2 are relatively small, and it means that the impact of the absent internal loops which are far from the junctions is limited. On the other hand, there is a significant degradation of the fusion performance when either upstream or downstream of the junction loops are not observed, shown as the difference between C2 and C3/C4. It is because the dynamics of traffic flow is much more complex for the segments which contain the junctions, and the accuracy of macroscopic traffic model is degraded comparing to the case of non-junction segments. In addition, the fusion performance of C3 and C4 are relatively similar. It indicates that the effect of either upstream or downstream junction loops is roughly equivalent.

## 6.5 Performance Comparison with Other Methods

### 6.5.1 Comparison with neural networks

The method of neural network has been introduced in Chapter 2. It is one of most popular methods in the area of traffic estimation and prediction. In practical terms, neural network is a non-linear statistical data modelling tool. It can be used to model complex relationships between inputs and outputs or to find patterns in data. Unlike the Kalman filter which used the known physical relationship as the modelling approach, neural network is purely a data-driven system. It requires certain amount of historical data to perform the modelling process, which is referred as *learning* or *training*. It means the dataset about the observations and estimated quantity is compulsory for the implementation of neural networks. More discussion about the advantages and disadvantages of neural networks have been presented in Chapter 2, this section will compare the estimation performance between two typical types of neural networks and the developed UKF based fusion method by using the simulation scenario.

A number of different types of neural networks have been applied in the area of traffic estimation/prediction. This thesis only selects the two most widely used neural networks in terms of difference of network topologies: *feed-forward networks* (FF-N) and *recurrent networks* (R-N). A brief about the difference between these two types of networks is provided as follows:

- 1) Feed-forward networks (**FF-N**): the data flow from input to output units is strictly feed-forward. The data processing can extend over multiple layers of units but no feedback connections are present, i.e., connections extending from outputs of units to

inputs of units in the same layer or previous layers. There are no cycles or loops in the network. The current use of feed-forward networks has a multi-layers structure, which is shown in Figure 6.14 (a).

- 2) Recurrent networks (**R-N**): it contains feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units suffer from a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. The general topology of recurrent neural network is shown in Figure 6.14 (b).

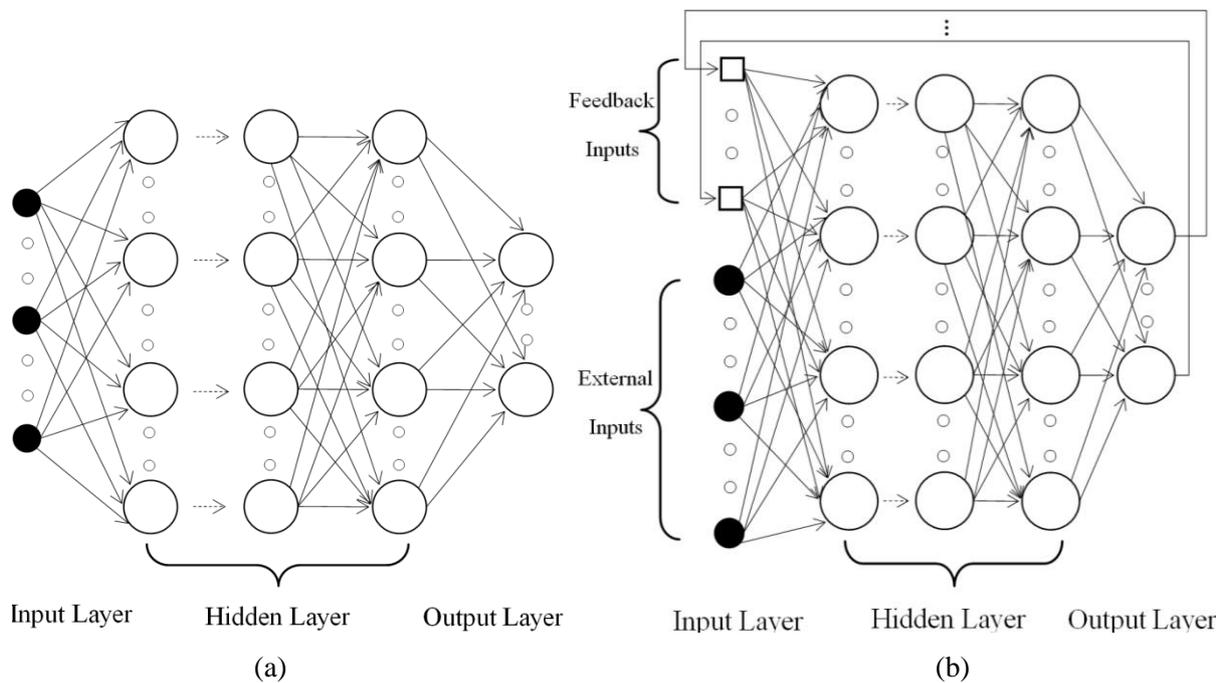


Figure 6.14 General Structure of (a) feed-forward networks and (b) recurrent networks

The neural network toolbox in Matlab is used to implement these two types of networks, and the Levenberg-Marquardt backpropagation algorithm (Zurada 1992, Haykin 1999) is adopted for the neural network training. MSE (mean square error) is used as the error criterion. The similar applications of these two approaches in the field of travel time estimation can be found in Park & Rilett (1999), Huisken & van Berkum (2003), Liu et al. (2006) and Li et al. (2008). The training data is from 10 days simulation providing 2880 samples. According to the nature of neural network, the estimation performance is supposed to be relevant to the similarity between the patterns in the training data and the estimation period. In order to evaluate the impact from the pattern difference of the training data, 10 days training data is generated by two schemes:

- 1) T1: the traffic demand is configured at a similar pattern as the estimation day. The average difference between the training day and the estimation day is less than 30%

- 2) T2: the traffic demand is configured at a very different way to the estimation day. Therefore, the traffic flow patterns of the training days are relatively non-recurrent to the estimation day. The average difference is larger than 50%. Figure 6.15 illustrates the flow pattern of the estimation day vs. an example of training day under the configuration of this scheme.

On the other hand, the developed UKF method does not necessarily require the historical data as the neural networks dose. In order to fairly compare these two types of methods, the follows two schemes are designed for the UKF to make use of historical data:

- 1) C1: no historical data used.
- 2) C2: the historical data is used to calibrate the model parameters  $\{v_f, \rho_{cr}, a\}$ . By this means, these parameters no longer need to be estimated in the Kalman filter process. The state vector and process are reduced accordingly.

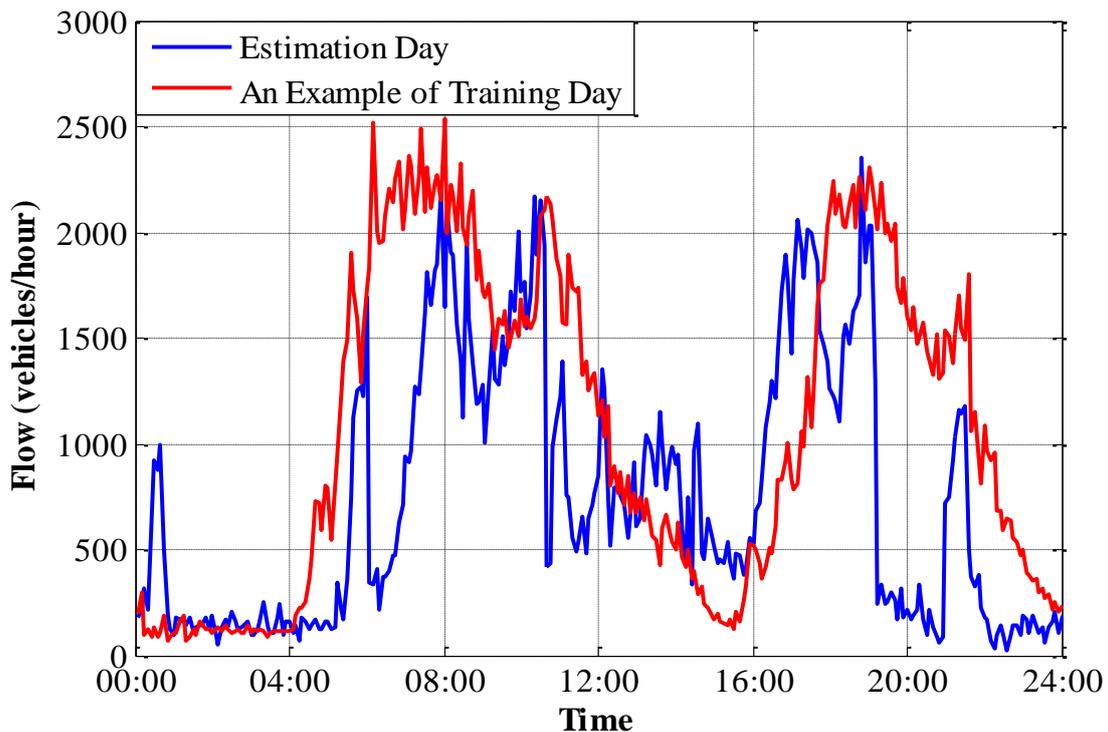


Figure 6.15 Traffic flow pattern of estimation day vs. an example of training day

The sensors settings are 7 ILDs with 10% additive noise, 10% GPS sampling rate with 30 seconds resolution and 10% noise. The comparison results are shown in Figure 6.16-6.18 and Table 6.9.

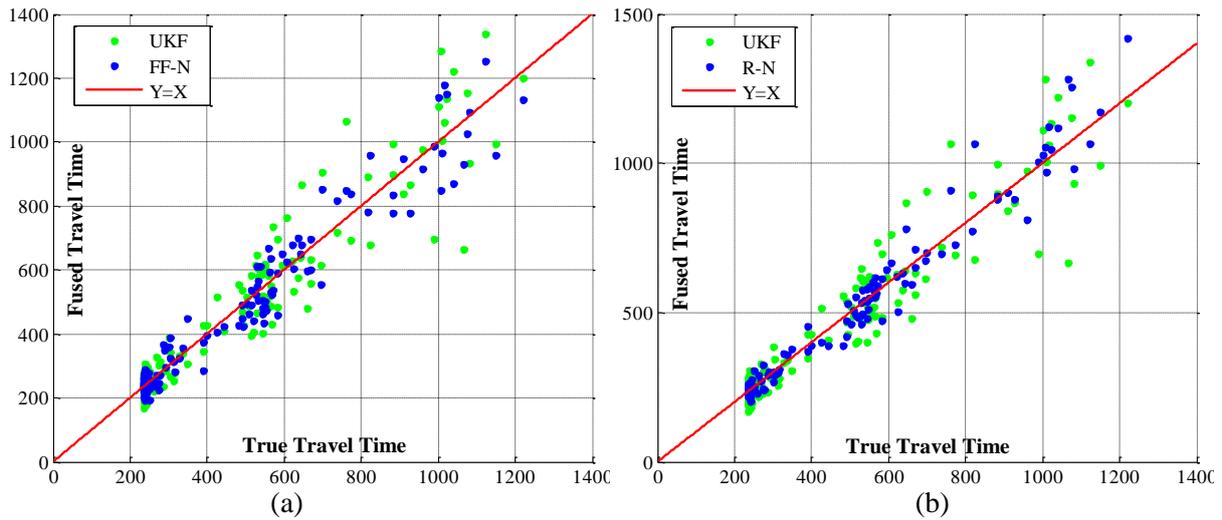


Figure 6.0.16 Scatter plot of UKF fusion under the C1 calibration vs. neural networks fusion under the T1 training data: (a) feed-forward networks; (b) recurrent networks

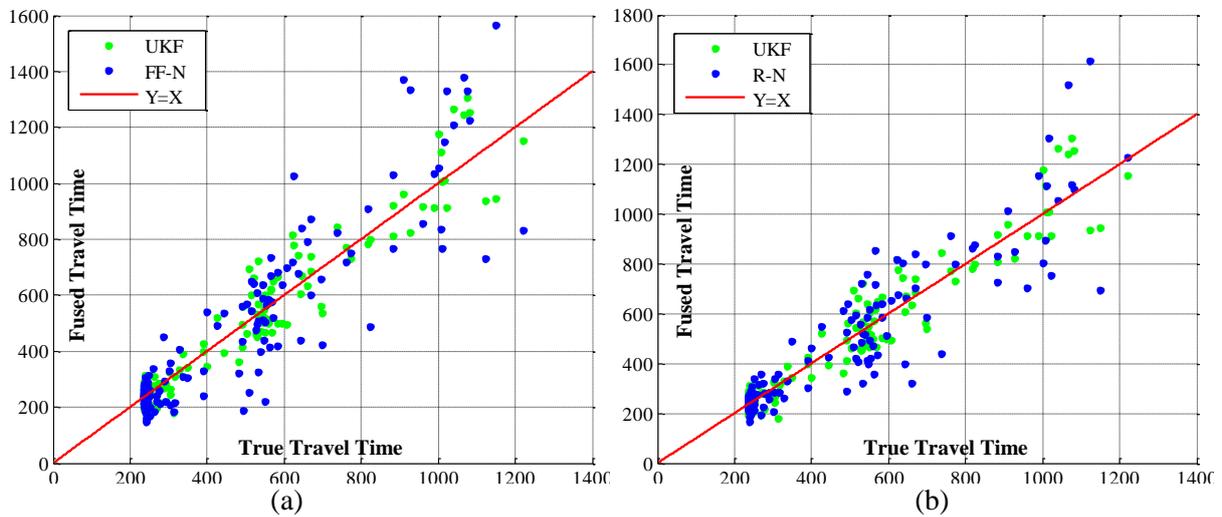


Figure 6.0.17 Scatter plot of UKF fusion under the C1 calibration vs. neural networks fusion under the T2 training data: (a) feed-forward networks; (b) recurrent networks

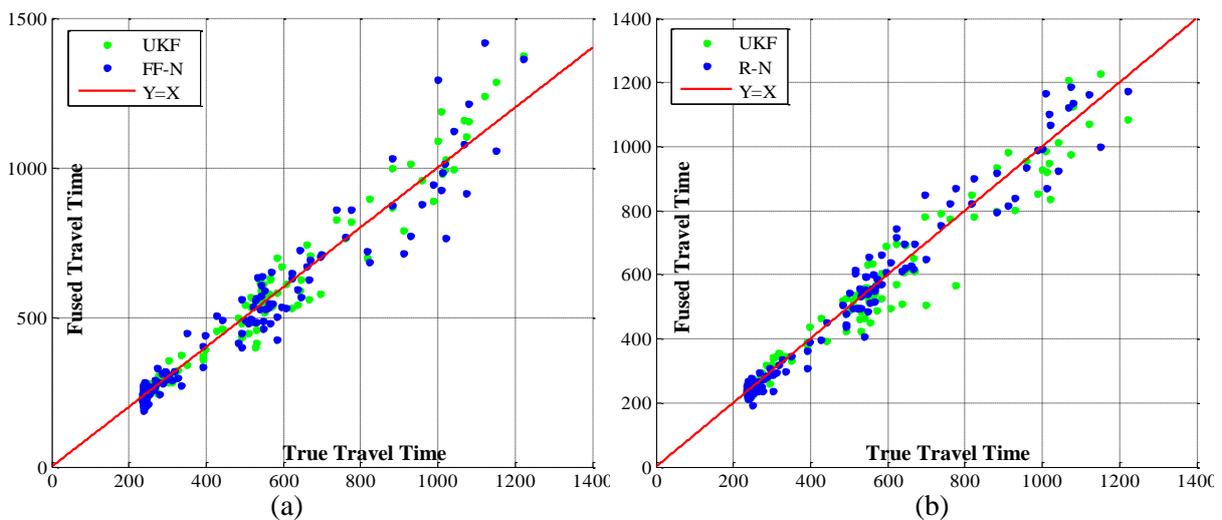


Figure 6.18 Scatter plot of UKF fusion under the C2 calibration vs. neural networks fusion under the T1 training data: (a) feed-forward networks; (b) recurrent networks

| <b>Fusion Method</b> | <b>Training Data / Model Calibration</b> | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|----------------------|--|-----------|------------|------------|-------------|
| <b>FF-N</b>          | <b>T1</b>                                | 0.82%     | 6.87%      | 65.06      | 8.91%       |
|                      | <b>T2</b>                                | 2.12%     | 11.73%     | 98.45      | 17.96%      |
| <b>R-N</b>           | <b>T1</b>                                | 0.52%     | 5.79%      | 57.85      | 7.60%       |
|                      | <b>T2</b>                                | 1.90%     | 10.16%     | 89.33      | 16.27%      |
| <b>UKF</b>           | <b>C1</b>                                | -1.87%    | 9.59%      | 82.74      | 14.02%      |
|                      | <b>C2</b>                                | 0.69%     | 6.02%      | 60.25      | 7.98%       |

10% GPS probes rate with 30 second T-resolution and 10% additive noise; 10% ILDs noise

Table 6.11 The performance comparison between UKF and neural networks

From the above comparison results, it can be seen that the neural networks (both of feed forward and recurrent types) with an appropriate training dataset have better fusion performance than the UKF method without the calibration of model parameters. It is not surprise since the network networks make use of the knowledge of the historical data which has similar patterns of the observations and traffic flow. The performance of neural networks is largely dependent on the quality of training data. As the results from T2 training data which contains the data from relatively different traffic conditions, the fusion performance of neural networks is degraded dramatically, and is worse than UKF fusion. It means that the model about the relationship between the observations and the estimated travel time is not accurately established by neural networks for this case. By contrast, the traffic model used in Kalman filter is determinedly built-in, and the fusion performance of UKF is not restricted to the historical data. It is one of the advantages of the Kalman filter discussed in Section 4.5 and 5.6. On the other hand, the fusion performance of UKF with the calibration by historical data shows slightly worse (less than 1% worse in terms of MAPE and RMSPE) than recurrent type of neural network with an appropriate training dataset. However, the requirement for the “good” historical data is difficult to be satisfied in practice. Therefore, comparing to neural networks, the developed UKF can achieve similar fusion performance when the historical data is correctly available and is more reliable when the historical data is inappropriate or unavailable for the training process, and is also computationally much more efficient.

## 6.5.2 Comparison with linear Kalman filter based method

One of the contributions of this thesis is to apply the nonlinear Kalman filter (EKF and UKF) into the problem of travel time estimation by fusing multiple sensor sources. The previous approaches in this field are all based on using the linear type of Kalman filter (introduced in Section 4.2) to fuse the data from ILDs and GPS probes. The most relevant works are presented at Chu et al (2005) and van Lint (2007) which are reviewed in Section 4.4. The traffic flow model used in these approaches is the LWR model which has been introduced in Section 5.2.5. It has been discussed that comparing to the nonlinear PW model, the LWR model as a linear formulation is not able to capture all of the complex interactions for a realistic traffic environment. This subsection compares the fusion performance between LWR model based linear Kalman filter (LKF) and the PW model based UKF.

In the implementation of LKF, the state and observation process is simply reduced as

**State**

$$\mathbb{X} = \boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_N)^T \quad (6.44)$$

**State process**

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i} [q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)] \quad (6.45)$$

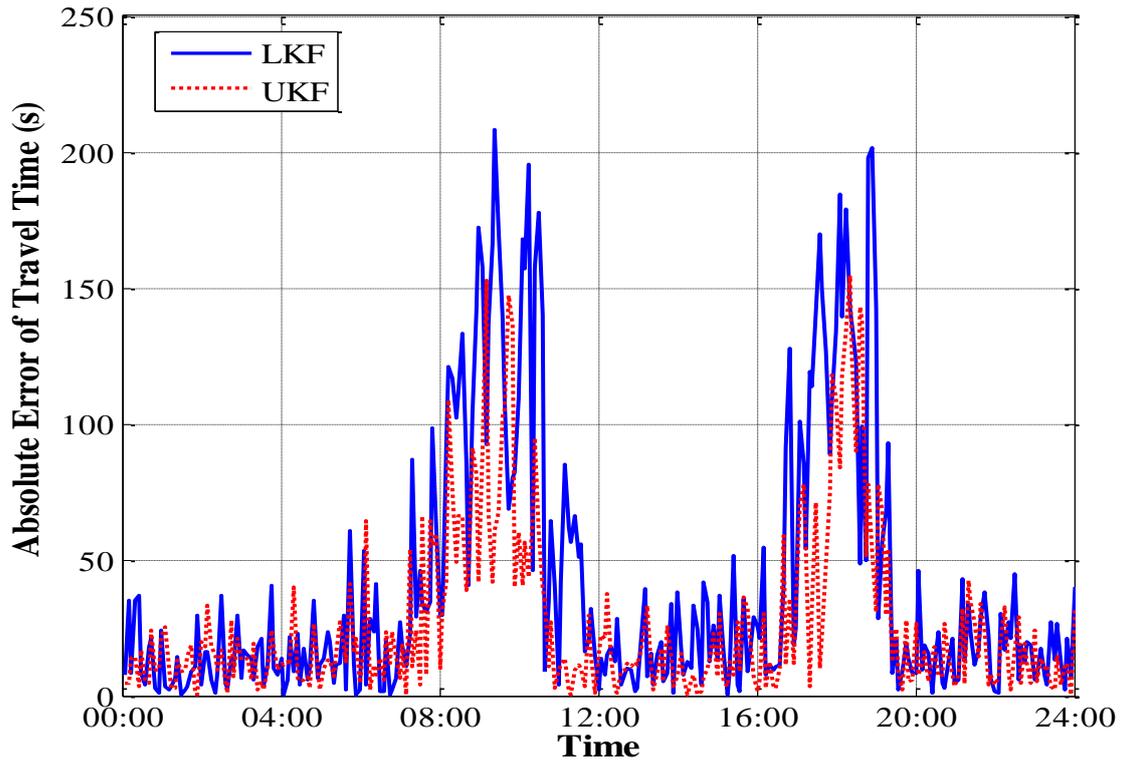
**Observation process**

$$\frac{1}{v_{\text{GPS}}(k)} = \frac{1}{\mathbf{q}(k)} \boldsymbol{\rho}(k) + \boldsymbol{\epsilon}(k) \quad (6.46)$$

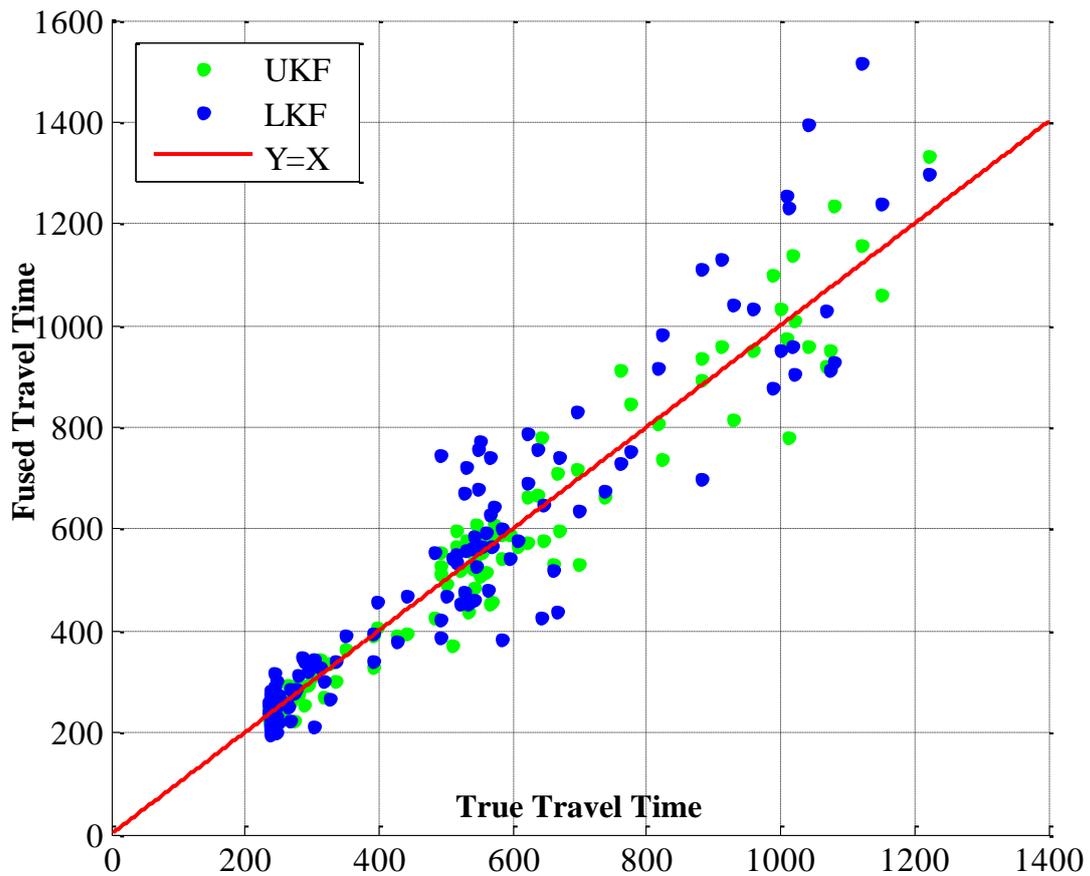
where the flow variable  $\mathbf{q}$  is obtained from ILDs. By using the solution to LKF (introduced in Section 4.2.2), the fusion from ILDs and GPS data is performed adaptively. In the performance evaluation, the ILDs and GPS probes configurations are set to 7 ILDs with 10% additive noise, 10% GPS sampling rate with 30 seconds resolution and 10% noise. The comparison results are shown in Figure 6.19 and Table 6.10.

|  | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|--|-----------|------------|------------|-------------|
| <b>UKF</b>   | -1.87%    | 9.59%      | 82.74      | 14.02%      |
| <b>LKF</b>   | 1.92%     | 11.83%     | 96.68      | 17.43%      |
| <b>GPS probe observation</b>   | -0.42%    | 14.46%     | 115.07     | 22.31%      |
| 7 ILDs with 10% additive noise; 10% GPS probes rate with 30 second T-resolution and 10% additive noise |           |            |            |             |

Table 6.12 The performance comparison between UKF and LKF



(a)



(b)

Figure 6.19 Comparison results between LKF and UKF: (a) absolute error plot (b) scatter plot

The comparison result in Table 6.10 shows a clear overall improvement of the UKF over the LKF, i.e. around 20% reduce in terms of MAPE, RMSE and RMSPE. From the Figure 6.19 (a), the error of LKF is similar as the error of UKF during the light traffic or the uncongested period, while the UKF shows an obvious smaller error level than LKF during the congestion period. It because the heavy traffic tends to be much more complexly interacted than light traffic, which cannot be appropriately described by the linear type of LWR model. The developed UKF which is based on the nonlinear PW model is able to capture more detailed dynamic behaviours of heavy traffic, and has the better fusion performance than LKF especially under the heavy traffic condition. Table 6.10 also lists the metrics for the observations from GPS probes. The improvement of LKF over only the GPS observation is around 25%, which is the gain from fusing the ILDs data.

## **6.6 Impact of Time-Varying Sensor Errors**

### **6.6.1 The limitation of constant error structures**

The error variance of each sensor source in the simulation processing is assumed to be constant (in terms of percentage of readings) over the time, which means that the error terms of sensor sources are assumed to be independent of the current traffic conditions and independent of the surrounding physical environment. However, in reality, these assumptions about the error variance may not be satisfied. For example, the ILDs flow observations may have larger error variance during the congestion period and the GPS positioning observations may be affected by the “canyon effect”.

In order to overcome this limitation, the relations between the error variance of the sensor sources and the current traffic conditions and the surrounding physical environment need to be determined. To establish these relations, the ground truth observations for each sensor source have to be available for the calibration process. For example, the relation between the ILDs flow error variance and the traffic conditions requires 1) the true traffic flow value at each location of ILD site; 2) a model which can describe the current traffic conditions. For the second requirement, the probabilistic traffic state identification method proposed in Appendix A shows a promising solution to this problem. During the period of this research, lack of the ground truth flow data constrains the development of this relation. In addition, due to the difference of the ILD sensitivities and the geographical properties of the sensor sites,

there might not be a general model to describe the relation explicitly. The problem of lacking of true values also applies to the GPS and ANPR sources.

In the case of time-varying sensors error, the assumption of constant error structure inevitably has impact on the fusion results. This section will discuss the theoretical impact and then use a scenario to provide quantitative results of this impact.

### **6.6.2 Theoretical impact**

From the introduction on the properties of Kalman filter in 4.1, the filtering result will achieve optimal estimate when the error variance of observations is perfectly known at each time step. When the current true error variance is different from the one used for Kalman filter calculation, the result is no longer to be optimal but the best estimate (minimum error variance) out of the class of unbiased filters (Trees 2001). More specifically, the imperfect error variance of the sensor source will cause the calculated Kalman gain to differ from its optimal value. As described in 4.4.2, the Kalman gain adjusts the weight of each source according to its error variance, i.e. assigns larger weights to those sources which have smaller error variances and smaller weight to those having larger error variances. For example, when the GPS observations are from a stretch of road which is surrounded by dense high buildings, those observations have relatively large error variance due to the “canyon effect”. In order to achieve the optimal estimate, the Kalman gain has to adaptively assign smaller weight to the GPS source. However, without the knowledge about the value of varying error variance, the filter still operates based on the pre-defined error variances. It makes the error of the fusion result dominated by the error of GPS observations. On the other hand, during some period of time, the error variance of one sensor source may be smaller than its defined value. To achieve the optimal filtering performance, the Kalman gain should assign larger weight to this sensor source. Similarly, failing to adapt this information will make the fusion error largely from the other sensor sources.

### **6.6.3 Impact analysis regarding the time-varying ILD error**

This subsection designs a scenario where the ILD flow percentage error is no longer constant over the time period. As the discussion above, the ILD flow observations experience relatively larger error variance during the congestion period than the light traffic condition due to the effect of stop-go vehicles from slow moving traffic flow. The original constant error percentage is shown in Figure 6.20 (a) while the structure of the time-varying error is

shown in Figure 6.20 (b) comparatively. In order to reflect the increase of error variance level during the congestion period, the time-varying pattern is modelled by a double triangular function with the form:

$$p_i^{TV}(t) = \begin{cases} (20 - 10|t - 8|)\%, & |t - 8| < 1 \\ (20 - 10|t - 18|)\%, & |t - 18| < 1 \\ 5\%, & 0 \leq t \leq 7 \text{ or} \\ & 9 \leq t \leq 17 \text{ or} \\ & 19 \leq t \leq 24 \end{cases} \quad (6.47)$$

where  $p_i^*(t)$  is the error percentage factor for the flow observation from of  $i$ th ILD at time  $t$ .

Accordingly, the form of the original constant error is given as follows:

$$p_i^C(t) = 10\%, \quad 0 \leq t \leq 24 \quad (6.48)$$

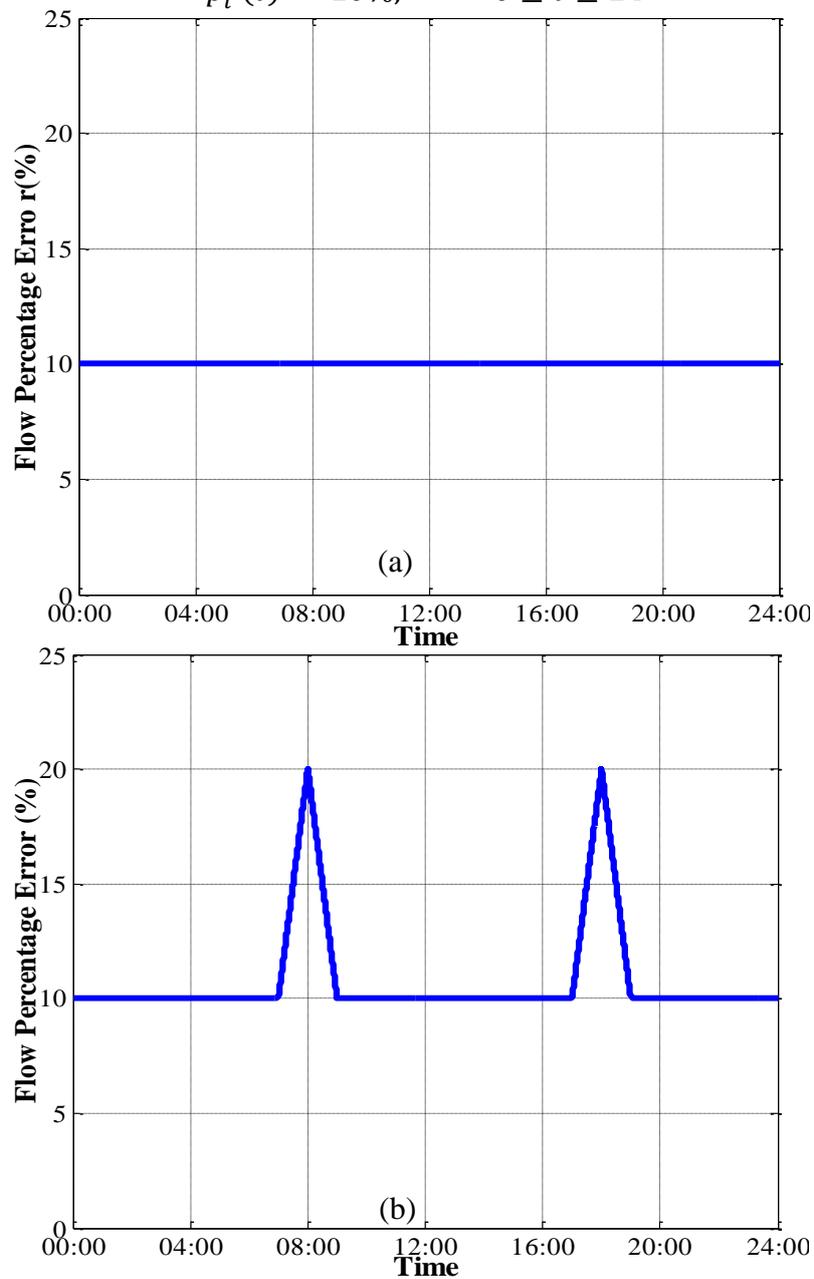


Figure 6.20 Flow error variance: (a) constant (b) time-varying

It can be seen that the designed time-varying error structure reflects the pattern of two rush hour period: 7:00 – 9:00 and 17:00 – 19:00. The percentage error linearly increases from 10% to 20% when the congestion starts, and decreases back to 10% at the end of the congestion period.

To evaluate the impact of this time-varying error structure on the fusion performance, the proposed UKF fusion framework is implemented in the following two ways:

1. The time-varying error structure is fully known to the fusion framework, it means the error variance of the ILD flow observations is perfectly calculated for the whole time period.
2. The fusion framework assumes a constant percentage error value: 10%. It leads to an imperfect calculation of the error variance.

The comparatively results of these two ways of framework implementation is given in Table 6.4. To highlight the impact of the time-varying error, the results only during the congestion period are separately presented in the Table 6.4 as well.

|                             |   | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|-----------------------------|---|-----------|------------|------------|-------------|
| Whole time period           | Fusion with known time-varying error    | -1.23%    | 7.53%      | 65.45      | 10.11%      |
|                             | Fusion without known time-varying error | -2.85%    | 9.72%      | 88.41      | 14.43%      |
| Only congestion Time Period | Fusion with known time-varying error    | -2.37%    | 8.69%      | 74.85      | 12.97%      |
|                             | Fusion without known time-varying error | -4.24%    | 14.27%     | 123.63     | 22.56%      |

Table 6.13 The performance comparison between constant error and time-varying error

From the above table, clear performance degradation is shown when the time-varying error is not known to the framework. The MAPE and RMSE increase 30% and 43% during the whole time period, and increase 64% and 65% during the congestion period. On the other hand, when the time-varying error structure is known to the fusion framework, the MAPE and RMSE only degrade 6% and 9% comparing to the case of constant error variance. Therefore, a well modelled relationship between the traffic conditions and the time-varying error structure can improve this problem, and Section 8.2 recommends that the integration of traffic state estimation and wavelet denoising could be a promising solution.

## 6.7 Summary

This chapter used a simulation experiment to evaluate the performance of the developed fusion framework. The simulation is based on a segmented urban road link with junctions, which can be fitted into the space-time fusion framework perfectly. The multiple sensor sources have also been well-modelled with the consideration of noise terms. The detailed approaches of framework implementation were presented in Section 6.2. The formulations of state/observation vectors and state/observation processes were developed for the use of nonlinear Kalman filter algorithms, i.e. EKF and UKF. The noise covariance matrices of state and observation, as one of most important but difficult components, were carefully studied and explicitly derived. Comparing to other approaches, this research provided the analytical forms of the state and observation noise covariance matrices which take the correlation properties among the state variables into consideration.

The results of framework evaluations were presented in Section 6.3-6.5. Firstly, the results from general fusion showed that the UKF outperforms the EKF. This is consistent with the comparison discussion between the accuracy of the EKF and the UKF in Section 4.3.4. Secondly, the sensitivity analysis about impacts of GPS probe factors on the fusion performance demonstrated that 1) increasing the sampling rate of GPS probe within the range of 0-20% leads to noticeable improvement, and the performance tends to converge beyond 20%; 2) the noise level of GPS observations affects the fusion performance linearly; 3) the fusion performance is relatively insensible to the T-resolution (the updating frequency of GPS probes) which is at the same scale as the estimated travel time. Thirdly, the results in Section 6.4.2 showed that the loops near the junctions play a more important role than the loops placed in the middle of two junctions, and the fusion with either junction upstream loops or downstream loops has equivalent performance. Fourthly, through the comparison analysis with the methods based neural networks and linear Kalman filter, the results shows 1) with proper training datasets, the neural networks are more accurate than the proposed UKF method. If the historical datasets are used to calibrate the model parameters in the developed fusion framework, the UKF method is slightly worse than the recurrent neural network and slightly better than the feed forward neural network; 2) Comparing to the methods based on the linear Kalman filter, the proposed UKF method has an obvious improvement (around 20% in terms of MAPE, RMSPE).

The framework implementations and evaluations described in this chapter are only based on the designed simulation scenario. The availability and noise terms of road sensors

supplement may not be same as assumed in the scenario. In addition, the traffic behaviour could be more complex than the one from the simulation. The next chapter will evaluate the performance of the developed fusion framework by implementing it into real traffic environment.

## **Chapter 7**

# **Framework Implementation & Evaluation Based on Real-World Data**

The previous chapter described the framework implementation and evaluation based on simulation data. In order to demonstrate the effectiveness of the proposed fusion framework in more practical environment, a real-world experiment was planned and carried out to collect the real traffic data. The fusion performance is evaluated based on the data collected from the real-world experiment, including the sensitivity analysis with respect to the T-resolution of GPS data and comparison analysis with the linear Kalman filter. The implementation details are basically same as those described in Chapter 6, thus this chapter will mainly focus on the evaluation results.

### **7.1 The Experiment of Data Collection from Real-World**

#### **7.1.1 Objectives of the experiment**

The previous chapter has demonstrated the effectiveness of the proposed fusion framework by simulation based evaluation. To verify its fusion performance in real-world environment, this chapter introduces an experiment which is designed to collect real traffic data required for the implementation of the framework. Unlike the simulated scenario, the true travel time is normally unavailable in practice. In order to perform the evaluation without knowing the travel time from ground truth, the observations from ANPR system are employed as the

compared travel time to the fusion results. This is because the direct travel time measurement from ANPR camera system is the most accurate and reliable data source in this case as discussed in Section 5.3.3. Therefore, the objective of this experiment is to evaluate the fusion performance based on the ILD and GPS data by comparing with the ANPR data. The metrics used for quantitative evaluation was same as the four variables introduced in Chapter 6: MPE, MAPE, RMSE and RMSPE (equation 6.1-6.4).

### 7.1.2 Road link

As mentioned above, the selected road link has to be monitored by a pair of ANPR cameras. In addition, the observations from ILDs should be generally available along the road link for the framework implementation. More specifically, two ILDs flow observations are required to be available at the entry and exit point of the link, and the in/out flow around the junctions are also required. Considering the availability of those sensors, an interurban road link with the length of 3.2 km (A229 Loose Road to A229 Hayle Road) in the area of Maidstone, UK is selected for the data collection. The topology of this road link is shown in Figure 7.1.



Figure 7.1 The Maidstone Road Link

There are 3 ILDs along the selected link as shown in above figure. Obviously, this ILD coverage is not adequate for the framework implementation since there is no flow observation around the entry/exit (start/end) point of the link and the main junctions such as the merging junction with Sutton Road (A273) and the Boughton Lane. In order to observe these missing flow observations, this experiment sets up two types of traffic volume surveys: 4 automatic surveys and 2 manual surveys which will be introduced in the follow subsection.

### 7.1.3 Traffic volume surveys

As the request of this experiment, 4 Automatic Traffic Surveys (ATS) were carried out in the Loose Road Maidstone area, at the following locations:

- A229 Loose Road Maidstone near the junction with Walnut Tree Avenue
- Boughton Lane Maidstone near the junction with Loose Road
- Postley Road Maidstone near the junction with Sheal's Crescent
- A229 Hayle Road near the junction with College Road

The surveys were all carried out between Tuesday 19th October and Saturday 30th October 2010, in order to cover the scheduled manual count (will be introduced following) and journey time survey date of Thursday 21st October 2010. All installed counters were synchronised to the UTMC data collection software clock, in order to count times as best as possible with those recorded time stamps from UTMC ANPR and ILDs data. The traffic volume is collected by installing two parallel pneumatic rubber tubes at each site 1 metre apart, stretched across the road surface and connected to individual data recorders. Figure 7.2 shows an example of the installed rubber tubes which are installed near the junction with Walnut Tree Avenue in this experiment. The outputs from this rubber tubes are similar as those from typical ILDs: flow and speed (if the ILDs are double loops).



Figure 7.2 parallel pneumatic rubber tubes for automatic traffic volume surveys

For the locations where the rubber tubes are not suitable to be set up (due to the safety and road geography consideration), 2 Manual Traffic Surveys (MTS) were undertaken at the following locations:

- A229/A274 Wheatsheaf Junction - Loose Road/Sutton Road Maidstone.
- A229 Upper Stone St & Sheals Crescent Junction Maidstone.

The survey was undertaken on Thursday 21<sup>st</sup> October 2010 between the hours of 07:00 and 19:00 and there was no reported problem with the surveys during that time period. The output from manual surveys is the traffic volume counts aggregated at 5 minutes interval.

## **7.2 ANPR Data**

### **7.2.1 Errors in matched ANPR travel time observations**

Section 2.1.2 generally described one of the error sources in ANPR travel time observations. It is due to the bias of sampling, i.e., large vehicle or vehicles immediately behind large vehicles. The simulation based evaluation in last chapter took this type of error into consideration. In real world, the travel time observations also have the error from unreasonable vehicle trajectories. More specifically, for a matched travel time observation, it is not possible to explicitly check if this vehicle travelled reasonably between the two ANPR sites. From Robinson & Polak (2006), the examples of the unreasonable vehicle trajectories may from the following cases:

- Vehicles which travel by an alternative route or detour rather than the direct route between two ANPR sites
- Vehicles which stop en-route, e.g., dropping off passengers, stopping by a local shop or filling up at a gas station
- Vehicles which do not attempt to travel in a prompt fashion, but rather travelled in an irrational manner, e.g., do not attempt to move into a faster moving lane
- Vehicles which are not restricted to normal traffic regulations, e.g., emergency vehicles

Therefore, it is necessary to process the raw matched travel time data remove the observations from those unreasonable vehicles. This process is normally named as ANPR data cleaning method. The simulation based evaluation in last chapter applied the pre-filtering overtaking rule approach (Robinson & Polak 2006) to demonstrate that the cleaning process did not improve the data quality since the simulation did not include the unreasonable

vehicles. However, this designed experiment used the real world ANPR data which did not have the error related to the invalid travel time observations from unreasonable vehicles. Next subsection will describe the real world ANPR data used in the evaluation and its cleaning mechanism.

### 7.2.2 ANPR data & its cleaning treatment

The aggregated 5-min ANPR travel time data from the selected road link in Maidstone (Figure 7.1) was provided by Kent County Council. In practice, due to the issue of privacy concern, the traffic authorities who manage the ANPR operation only provide the aggregated travel time data for traffic monitoring and other ITS applications. The records of individual vehicle travel time are normally treated as sensitive or confidential data. For the purpose of scientific research, the raw ANPR recognition records from a period of 12 hours were restrictedly provided by Kent County Council. Each of the number plate readings was replaced by a unique reference number, which was adequate to analyse the matched travel time data while no privacy/security controlled protocol was violated. The following paragraph describes the cleaning method used by Kent County Council to generate the aggregated travel time data.

Different traffic management operators may apply different cleaning methods to filter the raw individual travel time observations from their ANPR systems. There a number factors which may affect the choices of the selected cleaning method, e.g., the requirement of accuracy, the efficiency of the processing algorithm and the available resource of performing such a process. The cleaning mechanism adopted by Kent County Council (referred as Kent method) is in an offline fashion and described as follows:

- 1) **Extracting:** the travel times are based on arrivals at the *downstream* end of the link, and the raw data (observations of number plates) is taken from the camera feeds. The travel times for individual vehicles are determined by the time difference of the matched number plates recorded at upstream and downstream sites of the ANPR link.
- 2) **Coarse cleaning:** This process eliminates the matched travel time which would imply exceedingly long journeys (over 30 minutes) and those where the downstream readings preceded the matched upstream one.
- 3) **Fine cleaning:** After the coarse cleaning, each measured travel time then undergoes a validation process based on the average of the travel times for the five vehicles before and after the target vehicle arriving at the end point of the link. The target

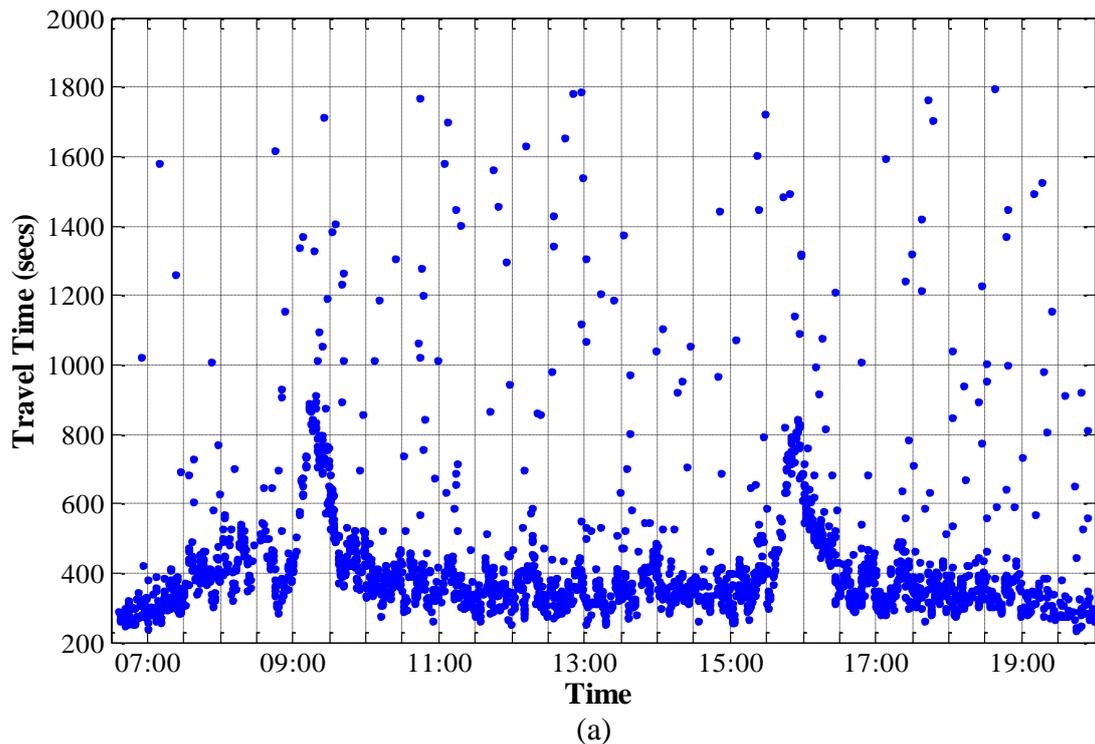
vehicle being validated is determined as invalid and discarded if its travel time exceeds twice that the average. Mathematically, the travel time of the  $i$ th vehicle is identified as a valid vehicle observation if its travel time  $TT_i$  meets the criteria formulated in the following equation:

$$TT_i < 2 \cdot \mathbb{E}[Q_i^{(5)}], \text{ where } Q_i^{(5)} = \sum_{\substack{j=i-5 \\ j \neq i}}^{i+5} TT_j \quad (7.1)$$

$\mathbb{E}[\cdot]$  is the expectation operation.

- 4) Aggregating:** The output travel time is then aggregated based on the average of the valid travel time observations during every 5-minute period.

Figure 7.3 (a) shows the ANPR observations after coarse cleaning by which the exceedingly long travel time (over 30 minutes) observations are removed. The unreasonable travel time observations are clearly illustrated over the whole time period. Figure 7.3 (b) shows the result of aggregated travel time after applying the fine cleaning treatment. By visual inspection, it can be seen that most of the unreasonably travel time observations are identified as invalid. For those probable “outliers” among valid observations, it is difficult to determine if it belongs to the group of valid vehicles or not without knowing the ground truth. Therefore, it is not possible to explicitly evaluate the performance of this cleaning treatment. Next subsection applies a number of the state-of-the-art cleaning approaches to comparatively show the effect of different treatments.



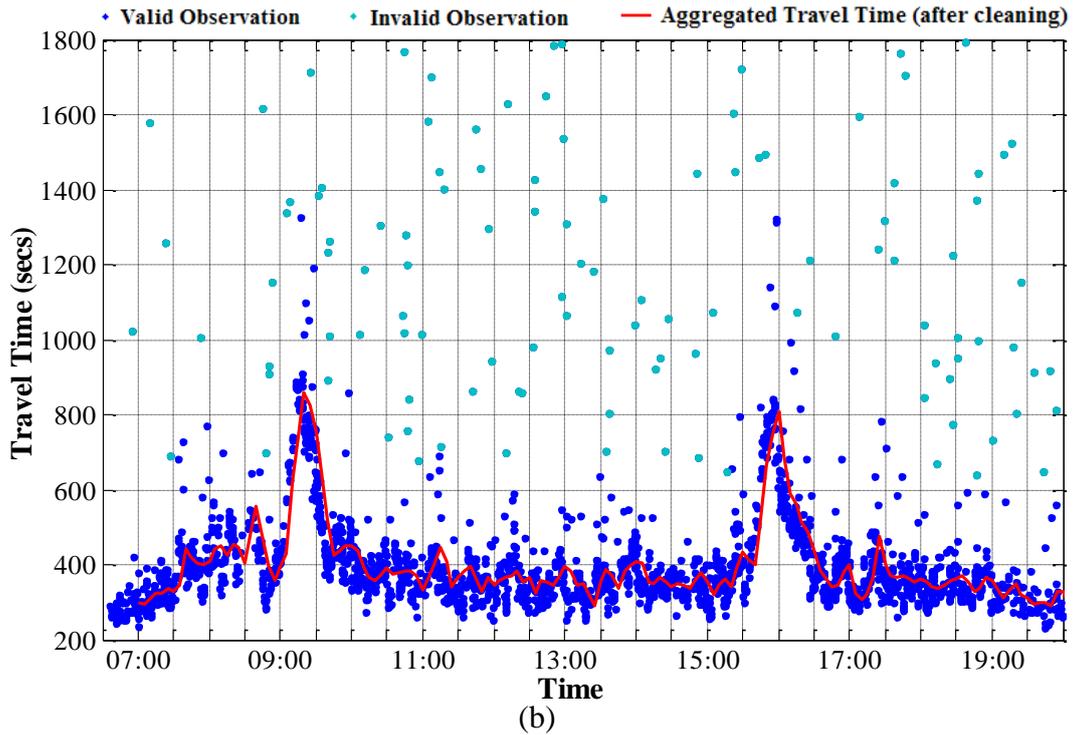


Figure 7.3 Result of cleaning treatment on ANPR travel time observation (a) before fine cleaning (b) after fine cleaning

### 7.2.3 Comparative analysis by using different cleaning treatments

#### Different cleaning treatments

According to the literature review in this field, the existing cleaning treatments are categorized as the following three types:

- **Statistical approach:** this group of approaches focuses on removing the outliers based on statistical analysis. In general, outlier detection purely depends on hard thresholds in terms of percentile or degree of spread, which were examined in (Clark et al. 2002).
  - 1) **Percentile test:** outliers are defined as observations falling outside a range of the 10th and the 90th percentile.
  - 2) **Mean Absolute Deviation (MAD) test:** the degree of spread is calculated as:

$$MAD = \frac{\sum_i^n |TT_i - TT_{median}|}{n} \quad (7.2)$$

where,  $TT_{median}$  is the median travel time of the total  $n$  vehicles during the aggregation time period (5 minutes in this evaluation). Outliers are then defined as observations outside the following criteria:

$$TT_{median} - 3 \cdot MAD \leq TT_i \leq TT_{median} + 3 \cdot MAD \quad (7.3)$$

- **Fuzzy clustering approach:** this type of cleaning does not actually remove the outliers, but assigns different weights to each of the observations according to its degrees of membership to the fuzzy clusters. Zheng & McDonald (2008) applied the fuzzy 1-means and fuzzy 1-line algorithms to estimate the travel time based on the raw ANPR matched data. The two fuzzy clustering based approaches are summarised as follows:

3) **Fuzzy 1-mean (F1M):** a point prototype cluster centre is adopted for the data set  $A$  which contains all the matched data within the aggregation time period. Denote  $c$  as the centre for the fuzzy set  $A$ , and  $d(TT_i, c)$  is the distance between the  $i$ th matched data and the cluster centre  $c$ . Correspondingly, a null cluster  $\bar{A}$  is hypothetically defined with constant distance  $D$  to each of the matched data  $TT_i$ . The criterion function for the F1M to be minimised becomes

$$J_{F1M} = \sum_{i=1}^n [w(TT_i)]^2 d^2(TT_i, c) + \sum_{i=1}^n [1 - w(TT_i)]^2 D^2 \quad (7.4)$$

where  $w(TT_i)$  is the degree of membership (or weights) of  $TT_i$  belonging to the fuzzy set  $A$  while  $[1 - w(TT_i)]$  is the degree of membership belonging to the hypothetical null set  $\bar{A}$ .

4) **Fuzzy 1-line (F1L):** In contrast to F1M, the prototype cluster of F1L uses a line instead of a point. The point-to-point distance  $d(TT_i, c)$  then becomes the point-to-line distance  $D[TT_i, L(\alpha, \beta)]$  where the line  $L(\alpha, \beta)$  is conventionally expressed in a two-dimensional space as

$$L(\alpha, \beta): y = \alpha x + \beta \quad (7.5)$$

The criterion function for F1L is accordingly becomes

$$J_{F1L} = \sum_{i=1}^n [w(TT_i)]^2 D^2 [TT_i, L(\alpha, \beta)] + \sum_{i=1}^n [1 - w(TT_i)]^2 D^2 \quad (7.6)$$

The solution of minimising  $J_{F1M}$  and  $J_{F1L}$  in terms of the weight associated to each observation  $w(TT_i)$  is based on the classic optimisation approach: equalising the partial derivative (with respect to each weight) to zero. The explicit formulations were originally developed in Dunn (1974) for F1M and Bezdek et al. (1981) for F1L. Zheng & McDonald (2008) also summarised these formulated solutions and applied into the case of ANPR data cleaning. Matlab Statistics Toolbox provides the in-built fuzzy clustering functions based on the above solutions which were used to for the following evaluation.

- **Traffic behaviour based approach:** the traffic behaviour based approach examines the actual cause of outliers from the traffic engineering point of view, which has been discussed in Section 7.2.1. In summary, it is because some of the matched vehicles have abnormal behaviours such as taking an alternative route, stopping en-route or not attempting to travel in a prompt fashion, and thus result in unreasonable long travel time observations. The traffic behaviour based approaches uses the serial structure of the data set to help detect those erroneous or inappropriate data records.

5) **Overtaking Rule (OR):** by considering the normal traffic behaviour, Robinson & Polak (2006) proposed this OR cleaning treatment based on the following assumptions:

- “▪ *On a single-lane road, in the urban environment, it would not be expected for valid vehicles to be overtaken by following vehicles.*
- *On a multi-lane road, in the urban environment, valid vehicles may be overtaken. However, it would be expected that the difference between the travel time of the overtaken slower vehicle and the faster vehicle would not be great, due to the small link length and slow speed limits in urban areas.*”

According to above assumptions, the target vehicle is regarded as outlier if the following rules are met:

- “▪ *The target vehicle has been overtaken by another vehicle.*
- *The target vehicle arrives at end point more than a certain time (the tolerance time) after any vehicle which has overtaken it.*”

Mathematically, the travel time of the  $i$ th vehicle is identified as a valid vehicle observation if its travel time  $TT_i$  meets the criteria formulated in the following equation:

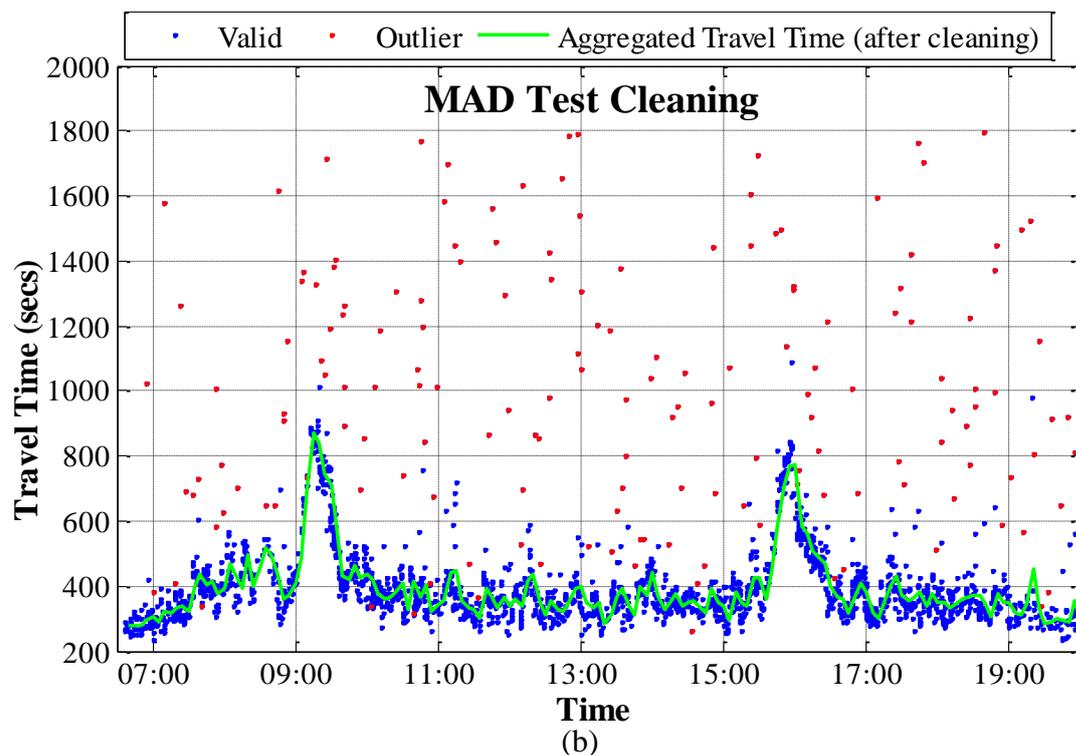
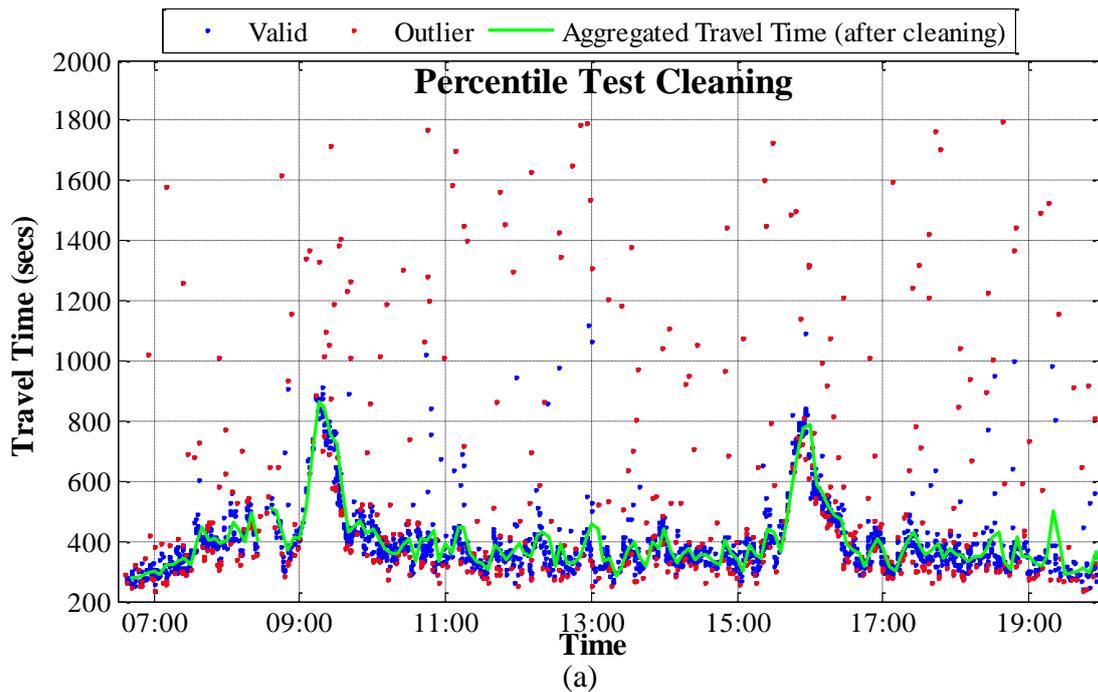
$$TT_i \leq TT_j + TD_{ij} + T_c \quad \forall j \in \{i - N, i - N + 1, \dots, i - 1\} \quad (7.8)$$

where  $TT_j$  is the travel time of the following vehicle,  $TD_{ij}$  is the time between the target and following vehicle passing start point and  $T_c$  is the tolerance time.  $N$  is the number of following vehicles took into consideration.

6) **Kent Treatment (KT):** this cleaning treatment has been described in Section 7.2.2. Note that the provider of ANPR data used in this evaluation is currently implementing this cleaning method.

## The cleaning results based on above six different treatments

Figure 7.4 (a) – (d) shows the results of applying Percentile, MAD, OR and Kent cleaning treatments to the raw individual matched data, especially the effects of outlier detection. Since fuzzy 1-mean and fuzzy 1-line treatments do not explicitly identify the outliers, Figure 7.5 only shows the aggregated travel time results by applying these two cleaning treatments. An overall comparison among these six treatments is presented in Figure 7.6.



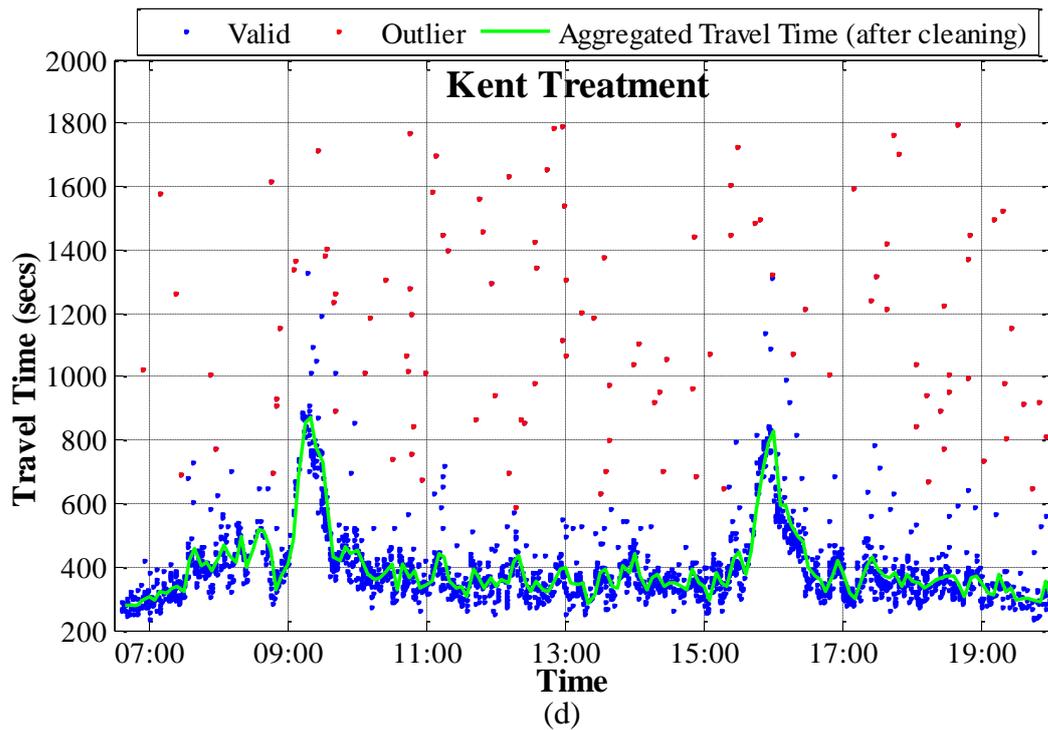
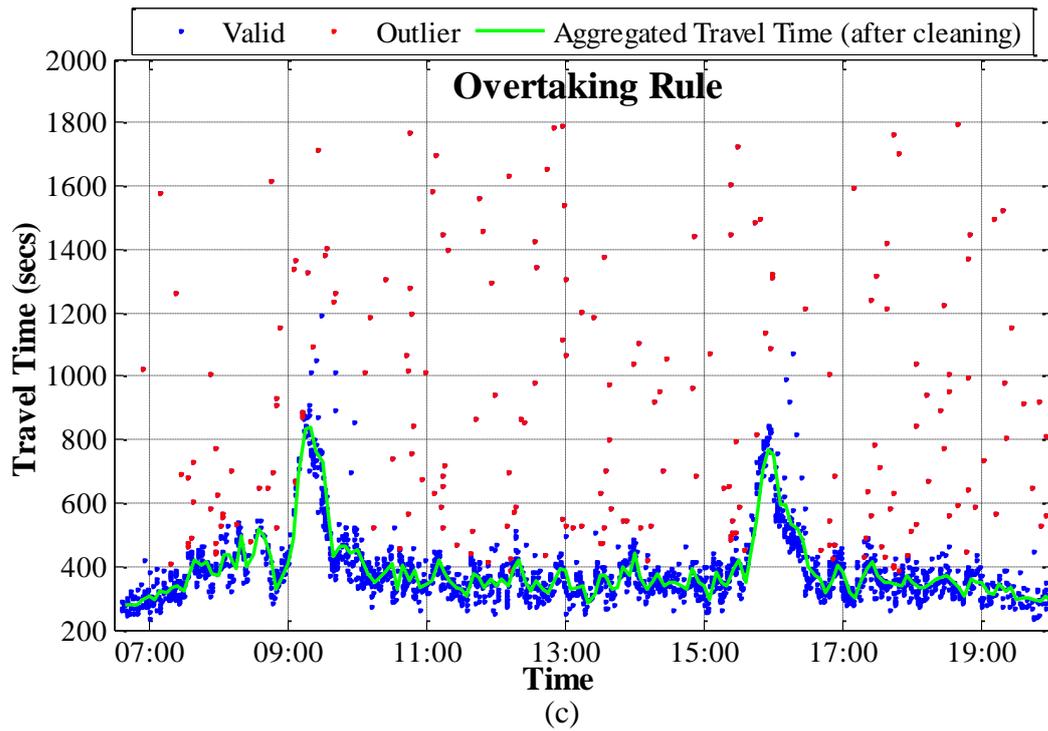


Figure 7.4 Cleaning results (a) percentile test; (b) MAD test; (c) overtaking rule (d) Kent treatment

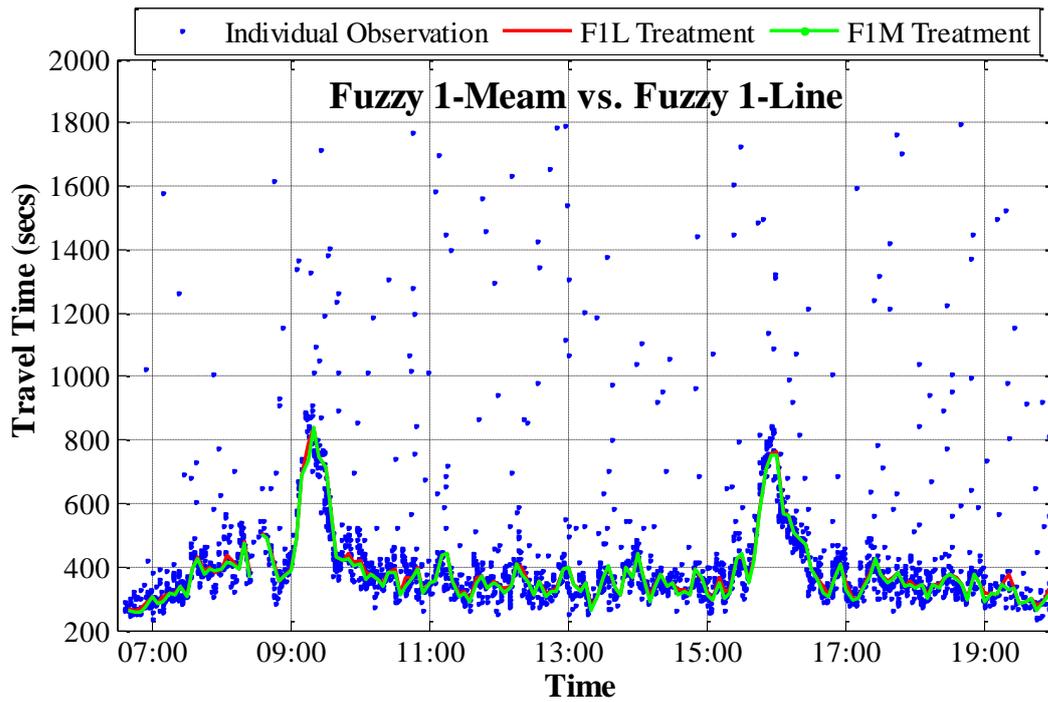


Figure 7.5 Cleaning results: F1M vs. F1L

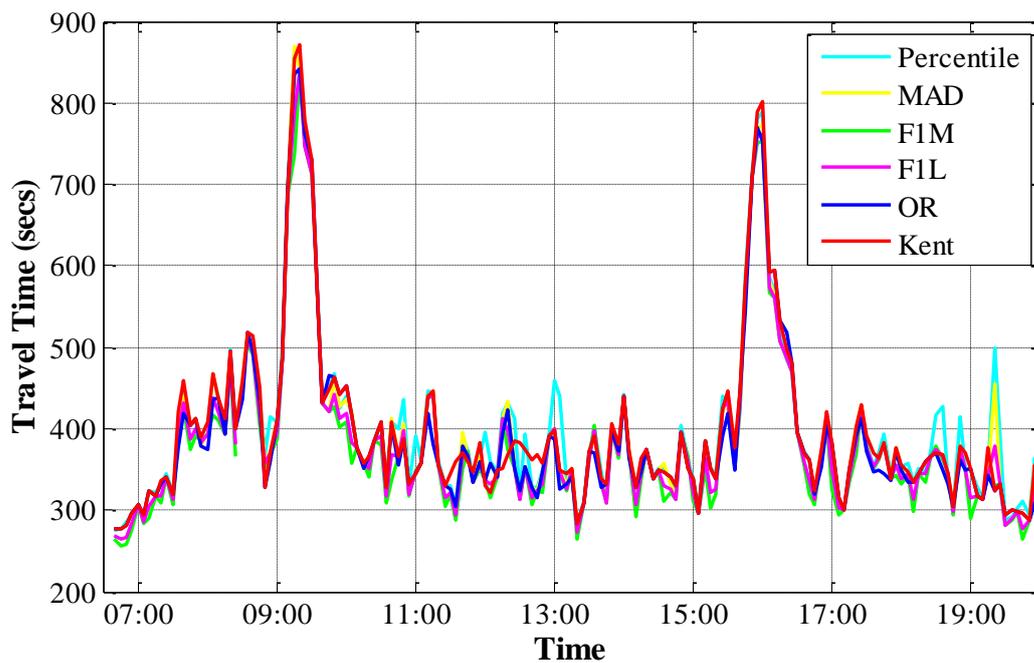


Figure 7.6 Overall comparison among different cleaning treatments

It should be noted that the true travel time (ground truth) is actually not known and has to be estimated from the data. An explicit evaluation over these treatments is not possible without knowing the ground truth. However, the results show the difference between the Kent method (the one used for the framework evaluation) and other methods. Table 7.1 gives the quantitative results by comparing the Kent treatment to others, and the calculation of the metrics uses the Kent treatment as baseline.

|            | MPE    | MAPE  | RMSE  | RMSPE |
|------------|--------|-------|-------|-------|
| Percentile | 1.11%  | 3.52% | 24.84 | 7.09% |
| MAD        | -0.77% | 2.26% | 18.77 | 5.14% |
| F1M        | -4.13% | 4.72% | 26.40 | 6.50% |
| F1L        | -2.78% | 3.64% | 20.33 | 5.14% |
| OR         | -1.82% | 2.47% | 17.93 | 4.44% |

Table 7.1 Quantitative Comparison results between Kent treatments and others

## Discussion

From the outlier plots (Figure 7.4 (a)-(d)), it can be seen that the statistical approaches detect more outliers than the traffic behaviour based approaches. It is not surprise since the percentile method classed 20% of all the observations as outliers which is a rather arbitrary treatment. The MAD method also uses a hard threshold to detect the outliers, and it risks excluding some good observations that could be a true reflection of the wide variability of the individual travel time. It can be seen that the percentile cleaning treatment shows relatively noticeable differences around 12:45 - 13:15 and 18:30 – 19:30. More specifically, the estimated travel time from the percentile treatment is higher than the other methods. Figure 7.7 illustrates this detailed difference during 19:10 – 19:30. The reason is because both of percentile and MAD potentially assumes the observations are symmetrically distributed about the median and expect to work reasonably well for an approximated normal distribution. However, when the sample size is small, the presence of invalid observations makes the distribution on longer symmetrical about the median. The cleaning based on a hard threshold therefore causes this difference. By visual inspection, the long travel time observations (the purple points in Figure 7.7) are very likely from the unreadable vehicle trajectories while identified as the valid observations by percentile treatment. In addition, during some traffic conditions such as the transition period between congested and uncongested, the traffic flow is no longer stationary, thus the distribution of those time period can be significantly different from normal.

Fuzzy clustering based approaches are expected to perform better than statistical approaches since they use the degree of membership to determine the weights of each observation rather than the hard threshold used in statistical approaches. The F1M algorithm normally assumes the observed travel time data is stationary which is not true for some traffic conditions as discussed in above paragraph. The F1L algorithm improves this limitation by

using a line as the cluster centre instead of a point in F1M algorithm. However, when the sample size of the observations is small and most of the valid observations are stationary, the presence of outliers has strong impact on the regressed line cluster which results in assigning larger weights to those outliers.

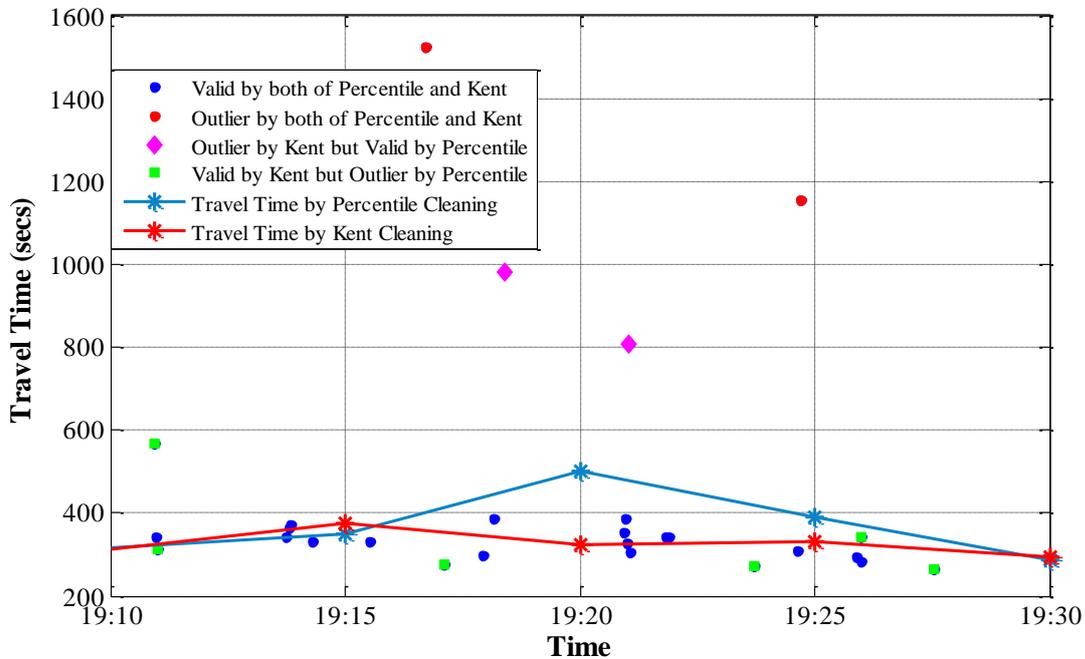


Figure 7.7 An example of the detailed difference between Percentile and Kent treatment

According to the above discussion, both of the statistical and fuzzy clustering approaches are based on using pure mathematical language to describe the cleaning problem. The underlying indication of the outlier observations is not fully considered in those cleaning treatments. In contrast, the traffic behaviour based approaches analyse the actual cause of the outliers by considering the fundamental rules of the traffic behaviour. They examine each observation based on the time serial structure of the data set from the traffic engineering point of view. Therefore, this type of cleaning treatment is theoretically more appropriate than other approaches.

The evaluation of the proposed fusion framework will use the estimated travel time based on Kent treatment for two reasons. Firstly, the ground truth travel time is not available to explicitly evaluate the performance of above cleaning treatment. From the overall comparison result (Figure 7.6), there is no significant difference between the Kent treatment and other ones generally. Particularly, according to the discussion, Kent treatment as one of the traffic behaviour approaches theoretically outperforms than statistical and fuzzy clustering approaches. In addition, the quantitative metrics (Table 7.1) shows that the Kent cleaning

result is closest to the other traffic behaviour based method - OR treatment in terms of RMSE and RMSPE. Secondly, in practice, the individual matched travel time data is not widely available from the organisations which operate the ANPR system, due to the issue of privacy concern. For the selected road link in Maidstone area, the ANPR travel time data is processed by Kent treatment and then provided for other practical applications. Therefore, it is fairly reasonable to assume the Kent treatment is the best possible way to clean the ANPR matched data and used their aggregated 5-minute data as the baseline for the evaluation.

## **7.3 GPS Data**

### **7.3.1 Collection method**

Moving car observer (MCO) technique or floating vehicle (Taylor et al. 2000) was employed in order to collect the GPS data for the framework implementation, In contrast to the conventional MCO, the travel time observations were automatically recorded by GPS logger which was placed in the floating vehicle. In this experiment, two vehicles were driven repeatedly around the road link for 12 hours which is the same time period as the manual survey. The process of vehicle floating was designed as follows:

- vehicle A starts to travel while vehicle B holds until A finishes the journey
- vehicle B starts to travel while vehicle A is returning back to start point
- if vehicle A returns to start point earlier than vehicle B finish the journey, vehicle A holds until B finishes
- if vehicle A returns to start point later than vehicle B finish the journey, vehicle A starts the journey directly
- repeat continuously until end

The purpose of this data collection process is to make sure that the one-way traffic from start point to end point is observed by at least one vehicle. It largely eliminates the situation that both of the two vehicles are congested on the returning journey so that no data is collected from the targeted one-way link. As stated in Section 5.3.2, the travel time observations from limited samples have associated with sampling error. In order to reduce sampling error, the following driving protocols were observed:

- not over speeding than speed limit of the road
- catch up the main traffic stream
- make the number of overtaking and overtaken largely equal

### 7.3.2 Description of recorded GPS data

The GPS logger used in this experiment was a “AMOD GPS Photo Tracker AGL3080” and outputted the data in NMEA0183 (National Marine Electronics Association) format at one record per second rate. The hardware specification of this device is provided in Appendix D. During the experiment time period, the logging mode is set to RMC (Recommended Minimum Navigation Information) whose output format is given in Figure 7.3. A part of sampled data logged by the device is shown in Figure 7.4

```

1          2 3          4 5          6 7  8  9  10 11| 12
|          | |          | |          | |  |  |  |  |  |
$--RMC,hhmmss.ss,A,llll.ll,a,yyyy.yy,a,x.x,x.x,xxxx,x.x,a*hh

```

- 1) Time (UTC)
- 2) Status, V = Navigation receiver warning
- 3) Latitude
- 4) N or S
- 5) Longitude
- 6) E or W
- 7) Speed over ground, knots
- 8) Track made good, degrees true
- 9) Date, ddmmyy
- 10) Magnetic Variation, degrees
- 11) E or W
- 12) Checksum

Figure 7.8 The format of RMC output

```

1 $ADVER,3080,2.2
2 $GPRMC,060458.431,A,5114.6584,N,00031.3780,E,0.11,15.99,211010,,,A*50
3 $GPRMC,060459.431,A,5114.6590,N,00031.3779,E,0.19,355.38,211010,,,A*66
4 $GPRMC,060500.431,A,5114.6590,N,00031.3779,E,0.13,353.50,211010,,,A*69
5 $GPRMC,060501.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6C
6 $GPRMC,060502.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6F
7 $GPRMC,060503.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6E
8 $GPRMC,060504.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*69
9 $GPRMC,060505.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*68
10 $GPRMC,060506.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6B
11 $GPRMC,060507.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6A
12 $GPRMC,060508.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*65
13 $GPRMC,060509.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*64
14 $GPRMC,060510.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6C
15 $GPRMC,060511.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6D
16 $GPRMC,060512.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6E
17 $GPRMC,060513.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6F
18 $GPRMC,060514.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*68
19 $GPRMC,060515.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*69
20 $GPRMC,060516.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6A
21 $GPRMC,060517.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*6B
22 $GPRMC,060518.000,A,5114.6590,N,00031.3779,E,0.00,353.50,211010,,,A*64

```

Figure 7.9 A part of sampled GPS log data

### 7.3.3 Error structure of the GPS data

For this source of GPS observations, although the technology used for the data collection is GPS, the nature of this method is originally from the Moving Car Observer (MCO) which has been discussed in Section 1.1.2. The reason of adopting the GPS-assisted MCO data source is because of the very limited number of monitored GPS traces available around the study area. To extract the travel time from the GPS log data, the required information is the time stamps of the GPS observations which are closest to the entry and exit point of the segments, and it is illustrated in Figure 7.5.

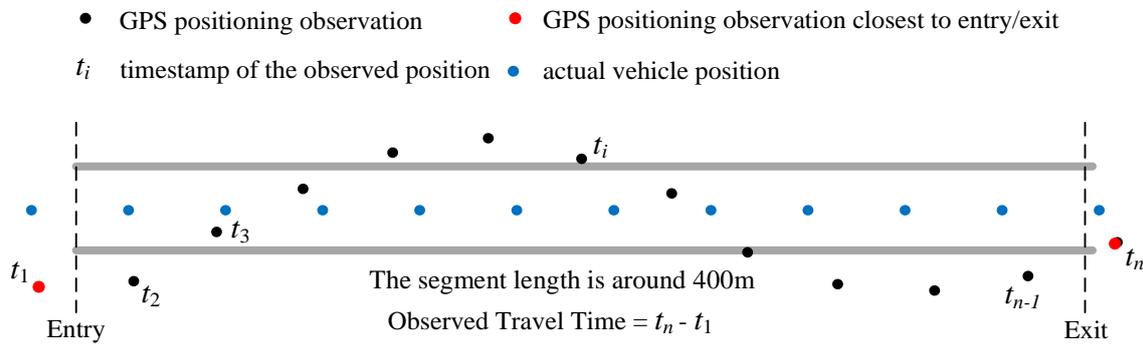


Figure 7.10 Illustration of the GPS data collection

It is obvious that this source of GPS observations does not involve a conventional map-matching process since it is known that the cars were driving along the selected link. Therefore, the error of this GPS assisted MCO is largely from two factors: 1) MCO sampling error; 2) GPS positioning error. For the first factor, the problem of limited vehicle samples leads that the observed travel time inadequately represents the main traffic stream. For the available experiment resource (not increase the number of floating vehicles), the designed driving protocol described at beginning of this section will make the floating vehicles approximately cruise at the average speed of the whole vehicle population. For the error variance of travel time observations caused by the second factor, an approximated calculation is developed to obtain a simple expression for the error term. The observed space mean speed from GPS is formulated as

$$\tilde{v}_{\text{GPS}} = \frac{\tilde{l}_2 - \tilde{l}_1}{\tilde{T}_2 - \tilde{T}_1} \quad (7.8)$$

where  $\tilde{l}_1$  and  $\tilde{l}_2$  are the observed locations (entry and exit of the segment) while  $\tilde{T}_1$  and  $\tilde{T}_2$  are the corresponding timestamps of the observations. According to IGS Products (IGS, 2009), the accuracy of the timestamps (satellite clocks) is 5 ns which is small enough to be ignored

in the application of travel time data collection, i.e.,  $\tilde{T}_1 \approx T_1$  and  $\tilde{T}_2 \approx T_2$  where  $T_1$  and  $T_2$  denote the true timestamps. Assuming the error term of positioning is normally distributed with zero mean and  $\sigma$  meters standard deviation, then

$$\tilde{v}_{\text{GPS}} = \frac{l_2 - l_1}{T_2 - T_1} + \frac{\varepsilon_2 - \varepsilon_1}{T_2 - T_1} = v_{\text{GPS}} + \gamma^{\text{GPS}} \quad (7.9)$$

where  $v_{\text{GPS}}$  is the true GPS space mean speed, and  $\gamma^{\text{GPS}}$  is the error associated with the actual observation  $\tilde{v}_{\text{GPS}}$ .  $\varepsilon_1$  and  $\varepsilon_2$  is expressed as:

$$\varepsilon_1 \sim \mathcal{N}(0, \sigma^2) \quad (7.10)$$

$$\varepsilon_2 \sim \mathcal{N}(0, \sigma^2) \quad (7.11)$$

The error term of the GPS space mean speed observation can be modelled as

$$\gamma^{\text{GPS}} = \frac{\varepsilon_2 - \varepsilon_1}{T_2 - T_1} \sim \mathcal{N}\left(0, \frac{2\sigma^2}{(T_2 - T_1)^2}\right) \quad (7.12)$$

i.e.

$$\sigma_{\text{GPS}}^2 = \frac{2\sigma^2}{(T_2 - T_1)^2} = \frac{2\sigma^2}{TT_{\text{GPS}}^2} \quad (7.13)$$

From the above equation, the variance of the observed space mean speed by GPS is related to two parameters: the observed travel time  $TT_{\text{GPS}}$  and the variance of GPS positioning error  $\sigma^2$ .  $TT_{\text{GPS}}$  is directly obtained by finding the timestamp of the observations which are closest to the entry and exit points. As mentioned above, the error of the observed timestamp is at a level of  $10^{-9}$  meters which is much less than the order of magnitude of the other variables and observations. It is then fairly assumed to be error free in the applications. The main issue is to estimate the positioning error variance  $\sigma^2$ . There are a number of factors can affect the performance of GPS positioning, e.g., geographical environment, non-line-of-sight, multipath effect, ionosphere and the systematic noise. The combination of these factors makes the estimation of the error variance  $\sigma^2$  difficult in practice as the error structure can be time-varying and dependent on the surrounding environment, and the impact will be discussed at the end of next section.

The GPS data logger recorded the direct instantaneous speed known as Doppler speed. This speed is measured based on the Doppler effect which is the change in frequency of a wave for an observer moving relative to the source of the wave (Roess 1998). The frequency difference between the received frequency and the satellite transmitted frequency can be then used to measure the speed of GPS receiver (Seeber 2009). In this experiment, the GPS data is recorded at 1 sec rate which generates the trackpoint Doppler speed of the vehicle. An

estimate of space-mean-speed can be obtained by taking the average of those Doppler speed measurements. However, the use of Doppler speed did not considered because the positioning error still exists in this case. The desired space-mean-speed is measured for a specific road segment. The errors of the GPS observed locations regarding segments boundaries have same effect on the estimated space-mean-speed.

### 7.3.4 Approximation & evaluation

#### Approximation

The output from the GPS logger has been shown in Section 7.3.2. It is nearly impossible to exactly determine the actual positioning error variance only based on the available output data without more detailed GPS data. In order to make use of the GPS data in the proposed fusion framework, the positioning error variance is approximated in a simple way and the related limitations will be discussed in the last part of this section. The typical positioning accuracy of the GPS with a SiRF Star III chip for a moving car is around 10 meters (Wormley 2010), thus the error standard deviation is assumed to be  $\sigma = 10$ . According to Equation 7.13, the error variance of the observed GPS space mean speed is then expressed as:

$$\sigma_{\text{GPS}}^2 = \frac{2\sigma^2}{TT_{\text{GPS}}^2} \approx \frac{200}{TT_{\text{GPS}}^2} \quad (7.14)$$

In order to demonstrate the impact of this approximation, next part addresses the difficulty to perform a direct evaluation for the travel time observations. The following section then uses an indirect method to analyse the effects.

#### Difficult to directly evaluate it based on the available data

In order to evaluate the accuracy of this approximation, intuitively it can be performed by examining the variance of the error between the travel time observed by GPS and the travel time observed by ANPR. The link travel time from the GPS observations can be obtained by

$$TT_{\text{GPS}}(k) = \sum_{i=1}^8 tt_{\text{GPS},i}(k) = \sum_{i=1}^8 \frac{L_i}{\tilde{v}_{\text{GPS}}(k)} \quad (7.15)$$

where  $tt_{\text{GPS},i}$  is the travel time observed by GPS for segment  $i = \{1, 2, \dots, 8\}$ ,  $\tilde{v}_{\text{GPS}}$  is the corresponding space mean speed and  $L_i$  is the length of the  $i$ th segment. According to Equation 7.9 and 7.12,

$$\tilde{v}_{\text{GPS}}(k) \sim \mathcal{N}\left(v_{\text{GPS}}, \frac{2\sigma^2}{TT_{\text{GPS},i}^2}\right) \quad (7.16)$$

Denote  $\varepsilon_{\text{ANPR,GPS}}$  as the error of the GPS travel time observation (compared to ANPR),

$$\varepsilon_{\text{ANPR,GPS}}(k) = \text{TT}_{\text{ANPR}}(k) - \text{TT}_{\text{GPS}}(k) \quad (7.17)$$

The mean and variance of  $\varepsilon_{\text{ANPR,GPS}}$  can be obtained from the actual observed values of ANPR and GPS at each time interval. On the other hand, according to Equation 7.15, and 7.16, the GPS travel time observations have an error term introduced by the positioning error, and is expressed as:

$$\text{TT}_{\text{GPS}}(k) \sim \mathcal{F}(U(\sigma), V(\sigma)) \quad (7.18)$$

where  $\mathcal{F}(\cdot)$  is a probability distribution which is resulted from the equation 7.16. The mean  $U(\sigma)$  and variance  $V(\sigma)$  is as functions of variance of positioning error  $\sigma$ . The theoretical formulation of  $\varepsilon_{\text{ANPR,GPS}}$  then becomes

$$\varepsilon_{\text{ANPR,GPS}}(k) \sim \mathcal{F}((\text{TT}_{\text{ANPR}}(k) - U(\sigma)), V(\sigma)) \quad (7.19)$$

Based on equation 7.17, the sampling mean and variance values of  $\varepsilon_{\text{ANPR,GPS}}$  are calculated by the available ANPR and GPS observations. If the mean and variance of distribution  $\mathcal{F}$  can be formulated as functions of  $\sigma$ , the theoretical value of  $\sigma$  can be solved with the knowledge of ANPR observations. According to equation 7.15,

$$\text{TT}_{\text{GPS}}(k) \sim \frac{1}{\tilde{v}_{\text{GPS}}(k)} \quad \text{where } \tilde{v}_{\text{GPS}}(k) \sim \mathcal{N}\left(v_{\text{GPS}}, \frac{2\sigma^2}{\text{TT}_{\text{GPS},i}^2}\right) \quad (7.20)$$

which indicates that  $\text{TT}_{\text{GPS}}$  is related to the reciprocal of normally distributed Random Variables (RVs). It is known that the mean and higher moments of a reciprocal normal distribution do not exist. More specifically, according to the rule of change of variables, i.e.,

$$f_Y(y) = \left| \frac{d}{dy}(g^{-1}(y)) \right| \cdot f_X(g^{-1}(y)) \quad (7.21)$$

where RV  $x$  has pdf  $f_X(x)$ , RV  $y$  has pdf  $f_Y(y)$  and  $y = g(x)$ . In this case,

$$g(x) = \frac{1}{x} \quad (7.22)$$

By using equation 7.21 and 7.22, and set  $\frac{1}{\tilde{v}_{\text{GPS}}(k)} = y$ , the probability density function (pdf) of  $y$  can be derived as:

$$pdf(y) = \frac{y^{-1} - \mu_y}{\sqrt{2\pi\sigma_y^3}y^2} e^{-\frac{(y^{-1}-\mu_y)^2}{2\sigma_y^2}} \quad (7.23)$$

where  $\mu_y = v_{\text{GPS}}$  and  $\sigma_y^2 = \frac{2\sigma^2}{\text{TT}_{\text{GPS},i}^2}$ . The mean and variance of  $pdf(y)$  is expressed as:

$$\mathbb{E}[y] = \int y \cdot pdf(y) dy = \int \frac{y^{-1} - \mu_y}{\sqrt{2\pi\sigma_y^3}y^2} e^{-\frac{(y^{-1}-\mu_y)^2}{2\sigma_y^2}} dy \quad (7.24)$$

$$\mathbb{V}[y] = \int (y - \mathbb{E}[y])^2 \cdot pdf(y) dy \quad (7.25)$$

It can be demonstrated that both of  $\mathbb{E}[y]$  and  $\mathbb{V}[y]$  are not integrable (Barndorff et al. 1982).

### Link length based indirect evaluation

Since a direct estimate of the moments of the GPS travel time is not available, in this section we propose an alternative indirectly approach. This is based on observing that since the length of the road link is always known, we can compare the known quantity with an estimate of road length based on the observed space mean speed. This estimate of road length will embody uncertainty due to errors in the measurement of speed. Thus, by comparing the link length estimated in this way to the true length, the quantity of the estimated speed can be assessed, albeit in a relative fashion. The total link length  $L$  consists of 8 segments with length  $L_i$   $i = 1, 2, \dots, 8$ . The length of each segment can be calculated as:

$$\tilde{L}_i = \tilde{v}_{\text{GPS},i} \cdot TT_i \quad (7.26)$$

where  $\tilde{L}_i$  is the calculated length which contains the error terms, and  $TT_i$  is the true segment travel time. By adding all of the segment length together, the total link length is

$$\tilde{L} = \sum_{i=1}^8 \tilde{L}_i = \sum_{i=1}^8 \tilde{v}_{\text{GPS},i} \cdot TT_i \quad (7.27)$$

According to Equation 7.16, the above equation can be represented as:

$$\tilde{L} = \sum_{i=1}^8 \tilde{v}_{\text{GPS},i} \cdot TT_i \sim \sum_{i=1}^8 TT_i \cdot \mathcal{N}\left(v_{\text{GPS},i}, \frac{2\sigma^2}{TT_{\text{GPS},i}^2}\right) \quad (7.28)$$

Assuming  $TT_i \approx TT_{\text{GPS}}$  (a strong assumption, which will be discussed later), then

$$\tilde{L} \sim \sum_{i=1}^8 \mathcal{N}(TT_i \cdot v_{\text{GPS},i}, 2\sigma^2) = \mathcal{N}(L, 128\sigma^2) \quad (7.29)$$

i.e.,

$$\tilde{L} - L \sim \mathcal{N}(0, 128\sigma^2) \quad (7.30)$$

The LHS of the above formulation is the error of GPS observed link length, RHS is a simple normal distribution with zeros mean and  $128\sigma^2$  variance. This method avoids the difficulty of solving reciprocal normal distribution by propagating the error terms to the link length calculation. Both of the actual link length  $L$  and the observed value  $\tilde{L}$  are available, thus the error sampling mean and variance can be calculated. Therefore, it provides an aspect to justify the GPS experiment data.  $\sigma^2$  is then estimated by

$$\tilde{\sigma}^2 = \frac{1}{128} \mathbb{V}[\tilde{L}] \quad (7.31)$$

### Results & Discussion

Table 7.2 summarised the results for the equal length evaluation (Equation 7.30) and Figure 7.6 illustrates the calculated link length error in a time series plot. The difference in percentage between the approximation and the actual data calculation is remarkable: 20.51%

in standard deviation and 36.81% in variance. One reason is because the approximation about GPS positioning error is based on the technical specifications of the hardware, which assumes the device works in a *normal* condition which has not been defined clearly. On the other hand, the results calculated from the observations are based on the real world environment. The collected data can be corrupted by a number of sources such as bad satellite geometry, non-line-of-sight, multipath effect, ionosphere effects, satellite clock error and the systematic noise (Hofmann 2001). The combination of these factors makes the estimation of the error variance difficult and increases the uncertainty in practice as the error structure can be time-varying and dependent on the surrounding environment. Due to the limitation of the GPS devices used in this data collection experiment, lack of more detailed data and only considering the location and timing data leads to relatively large uncertainty about the GPS observations. To fully analyse those noise sources and uncertainties, extra data needs to be obtained from the GPS receivers such as signal strength, level of signal-to-noise-ratio, information about the satellites, and dilution of precision. By having these more detailed data about the working condition of the GPS devices, the error structure can be modelled in a time-varying fashion which will be addressed as one of the future research areas in Chapter 8.

|  |                                  | <b>Mean</b> | <b>Standard Deviation</b> | <b>Variance</b>         |
|--|----------------------------------|-------------|---------------------------|-------------------------|
| <b>Calculated based on equal length</b>                              | Absolute value                   | -31.69 (m)  | 142.32(m)                 | 20256 (m <sup>2</sup> ) |
|  | Percentage in total link length  | -1.00 %     | 4.48%                     | 0.22%**                 |
| <b>Expected by approximation</b>                                     | Absolute value                   | 0           | 113.1(m)                  | 12800 (m <sup>2</sup> ) |
|  | Percentage in total link length* | —           | 3.56%                     | 0.13%**                 |
| <b>Difference in %</b>   | Absolute value                   | —           | 20.51%                    | 36.81%                  |
|  | Percentage in total link length* | —           | 0.92%                     | 0.09%                   |
| *Total link length is 3176 meters                                    |                                  |             |                           |                         |
| **Calculated as the variance value divided by square of total length |                                  |             |                           |                         |

Table 7.2 The comparison between approximation and equal length evaluation regarding the GPS data quality

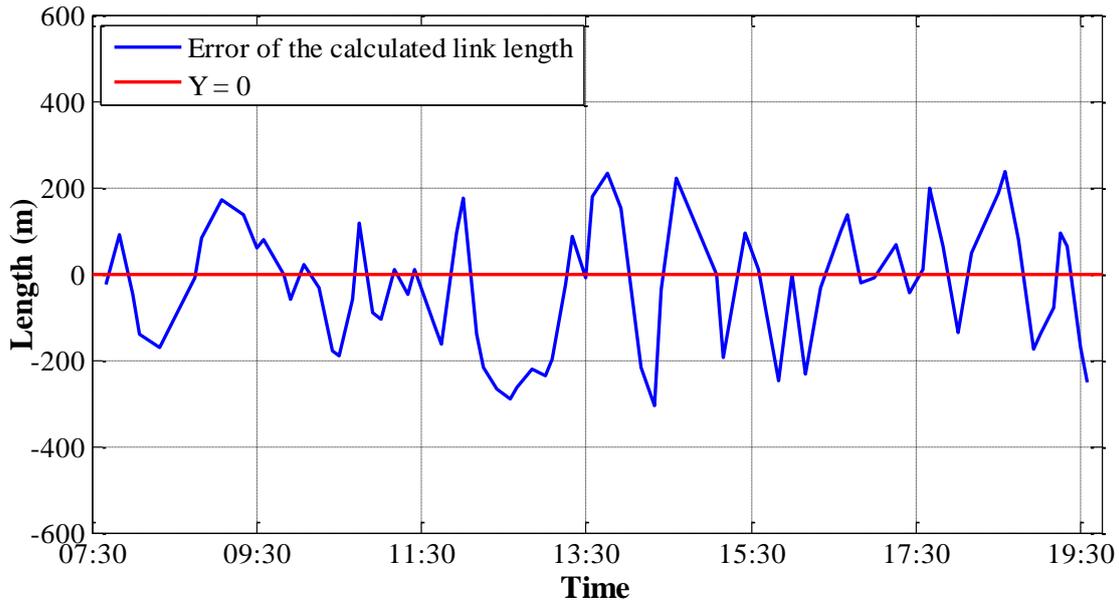


Figure 7.11 Illustration of the effects of GPS errors regarding link length

The other reason resulting in the difference between the approximate error variance and the actual sampling variance is because the real segment travel time is assumed by the GPS observations in the evaluation. The vehicles equipped with GPS are only samples from the whole vehicle population, which leads to the sampling error in the travel time observations as discussed in Section 7.3.1 and 7.3.2. Therefore, the difference between the actual travel time and the GPS observed travel time is the other cause of link length calculation. Therefore, the actual difference associating to GPS poisoning error is expected to be smaller than the one in the evaluation. In addition, the error propagation to large scale measurement (link length) has the effect of increasing magnitude.

In summary, this evaluation is performed in an indirect fashion, i.e. propagating the GPS observation error to the link length. By comparing the calculated link length to the true value, a number of the findings are obtained for the implementation of fusion framework. Firstly, although the absolute difference of the link length estimation between the approximate variance and the observed data is noticeable, the relative difference in percentage of total measurement is much smaller. Therefore the use of the approximate error variance is we believe reasonable in this case. Secondly, the actual error pattern shown in Figure 7.6 fluctuates around zero which indicates that the assumption of non-time-varying error is largely satisfied in this experiment. Thirdly, the difficulty of direct evaluation and the limitations of the assumption made in the indirect evaluation indicate the potential value of obtaining and analysing more detail GPS data.

## 7.4 Framework Implementation

### 7.4.1 Abstract scenario

In order to apply the PW macroscopic traffic flow model (introduced in Section 5.2.5) into the segmented link framework (introduced in Section 5.4.1), a schematic representation of the experimental road link is given in Figure 7.5. The locations of the ANPR, ILDs and survey points are shown in this figure as well.

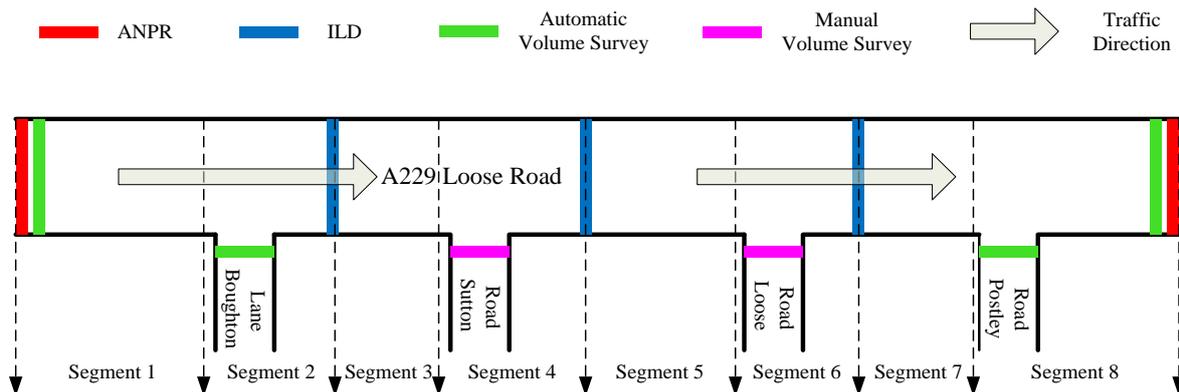


Figure 7.12 The schematic representation of the experimental road link

Similar as the implementation in the simulation scenario described in the previous chapter, the target link is divided into 8 segments according to the locations of ILDs and junctions. The in/out flows from the junctions are measured by either automatic or manual traffic volume survey. A pair of ANPR cameras is located in the entry and exit point of the link to monitoring the average travel time.

### 7.4.2 The implementing structure

Section 6.2 uses a simulation based scenario to illustrate how the proposed fusion framework is implemented. The fusion structure for the real-world scenario is same as the simulated one except the travel time observation from ANPR is considered as the true travel time used to evaluate the fusion performance. State/observation vector, state/observation process and noise covariance matrices as the key parts of the fusion framework are modelled as the same approach in Section 6.2. A brief description is provided as follows.

#### State vector

$$\mathbb{X} = \begin{bmatrix} \rho \\ v \\ q \\ p \end{bmatrix} \quad (7.1)$$

where

$\rho$ : the traffic density of each segment

$v$ : the mean speed of each segment

$q$ : the flow at the boundaries of each segment

$p$ : the model parameters  $v_f, \rho_{cr}, a$  to be estimated on-line

### Observation vector

$$\mathbb{Y} = \begin{bmatrix} \mathbf{q}_{ILD} \\ \mathbf{q}_{ATS} \\ \mathbf{q}_{MTS} \\ \mathbf{v}_{GPS} \end{bmatrix} \quad (7.2)$$

where

$\mathbf{q}_{ILD}$ : the flow observations from ILDs

$\mathbf{q}_{ATS}$ : the flow observations from automatic traffic survey

$\mathbf{q}_{MTS}$ : the flow observations from manual traffic survey

$\mathbf{v}_{GPS}$ : the travel time observations from GPS probes

### State process

$$\mathbb{X}(k+1) = \mathcal{F}[\mathbb{X}(k), \mathbb{u}(k), \mathbb{w}(k)] \quad (7.3)$$

where

$$\mathbb{u} = [\psi, \tau, c]^T \quad (7.4)$$

$$\mathbb{w} = \begin{bmatrix} \xi^\rho \\ \xi^v \\ \xi^q \\ \xi^p \end{bmatrix} \quad (7.5)$$

$\mathbb{u}$  is the model parameter,  $\mathbb{w}$  is the noise terms associating with the corresponding state vector and  $\mathcal{F}$  is the macroscopic traffic model in the equation 5.44.

### Observation process

$$\mathbb{Y}(k) = \mathcal{H}[\mathbb{X}(k), \mathbb{v}(k)] \quad (7.6)$$

where

$$\mathbb{v} = \begin{bmatrix} \gamma^{ILD} \\ \gamma^{ATS} \\ \gamma^{MTS} \\ \gamma^{GPS} \end{bmatrix} \quad (7.7)$$

$\mathbb{v}$  is the noise term associating with each sensor observation and  $\mathcal{H}$  is the model described in equation 5.50-5.51.

## State noise covariance

$$\mathbf{R} = \mathbb{E}[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T] \quad (7.8)$$

Each position of the matrix is in the same form as derived in equation 6.38:

$$\left\{ \begin{array}{l} \mathbb{E}[\xi^{\rho_i} \xi^{\rho_{i-1}}] = -\frac{T^2}{L_i L_{i-1}} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{q_{i-1}}] = \frac{T}{L_i} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{q_i}] = -\frac{T}{L_i} \sigma_q^2 \\ \mathbb{E}[\xi^{\rho_i} \xi^{s_i}] = \begin{cases} \frac{T}{L_i} \sigma_q^2 & i = 2, 4, 6, 8 \\ 0 & i \neq 2, 4, 6, 8 \end{cases} \\ \mathbb{E}[\xi^{\rho_i} \xi^{r_i}] = \begin{cases} -\frac{T}{L_i} \sigma_q^2 & i = 2, 4, 6, 8 \\ 0 & i \neq 2, 4, 6, 8 \end{cases} \end{array} \right. \quad \left\{ \begin{array}{l} \mathbb{E}[(\xi^{\rho_i})^2] = \begin{cases} \frac{4T^2}{L_i^2} \sigma_q^2 & i = 2, 4, 6, 8 \\ \frac{2T^2}{L_i^2} \sigma_q^2 & i \neq 2, 4, 6, 8 \end{cases} \\ \mathbb{E}[(\xi^{q_i})^2] = \sigma_q^2 \\ \mathbb{E}[(\xi^{v_i})^2] = \sigma_v^2 \\ \mathbb{E}[(\xi^{v_f})^2] = \sigma_{v_f}^2 \\ \mathbb{E}[(\xi^{\rho_{cr}})^2] = \sigma_{\rho_{cr}}^2 \\ \mathbb{E}[(\xi^a)^2] = \sigma_a^2 \\ \text{others} = 0 \end{array} \right. \quad (7.9)$$

## Observation noise covariance

$$\mathbf{Q} = \mathbb{E}[(\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^T] \quad (7.10)$$

Comparing to the simulation scenario, the observation from ANPR is removed and two types of observations from automatic and manual traffic surveys are added. Therefore, the structure of observation noise covariance is amended accordingly as:

$$\mathbf{Q} = \begin{pmatrix} \gamma_1^{\text{ILD}} & \dots & \gamma_7^{\text{ILD}} & \gamma_1^{\text{ATS}} & \dots & \gamma_4^{\text{ATS}} & \gamma_1^{\text{MTS}} & \gamma_2^{\text{MTS}} & \gamma_1^{\text{GPS}} & \dots & \gamma_M^{\text{GPS}} \\ \sigma_{\text{ILD}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\text{ILD}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\text{ATS}}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\text{ATS}}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{MTS}}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{MTS}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{GPS}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{GPS}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\text{GPS}}^2 \end{pmatrix} \left\{ \begin{array}{l} \gamma_1^{\text{ILD}} \\ \dots \\ \gamma_7^{\text{ILD}} \\ \gamma_1^{\text{ATS}} \\ \dots \\ \gamma_4^{\text{ATS}} \\ \gamma_1^{\text{MTS}} \\ \gamma_2^{\text{MTS}} \\ \gamma_1^{\text{GPS}} \\ \dots \\ \gamma_M^{\text{GPS}} \end{array} \right. \quad (7.11)$$

## 7.5 Fusion Performance

### 7.5.1 Parameters

The structure and all the necessary variables have been formulated in the previous section. It is quite straightforward to apply the EKF & UKF algorithms (equation 4.16-4.24 & equation 4.33-4.49) into the constructed framework. The parameters used are summarised in Table 7.1.

---


$$\tau = 10 \text{ (s)}, \psi = 18 \text{ (km}^2\text{/h)}, c = 5 \text{ (veh/km)}$$

(calibrated by one day data)

---


$$\sigma_q^2 = \sigma_{ILD}^2 = \sigma_{ATS}^2 = 100 \text{ (veh/h)}^2$$

$$\sigma_v^2 = 3 \text{ (km/h)}^2$$

$$\sigma_{v_f}^2 = 0.3 \text{ (km/h)}^2$$

$$\sigma_{\rho_{cr}}^2 = 10^{-2} \text{ (veh/km)}^2$$

$$\sigma_a^2 = 10^{-4}$$

$$\sigma_{MTS}^2 = 50 \text{ (veh/h)}^2$$

$$\sigma_{GPS}^2 = 1.25\% v_{GPS}$$


---

Table 7.3 The parameters used for framework implementation

In the application of the UKF, the sigma points generated by the algorithms may not be physically meaningful, such as negative density. In order to eliminate this occurrence, limits for the upper and lower bound of states and parameters are imposed, see Table 7.2.

---


$$0 \text{ (veh/km)} \leq \rho_i \leq 160 \text{ (veh/km)}$$

$$8 \text{ (km/h)} \leq v_i \leq 80 \text{ (km/h)}$$

$$35 \text{ (km/h)} \leq v_f \leq 65 \text{ (km/h)}$$

$$25 \text{ (veh/km)} \leq \rho_{cr} \leq 70 \text{ (veh/km)}$$

$$1 \leq a \leq 3$$


---

Table 7.4 Bounds of the states and parameters for the UKF

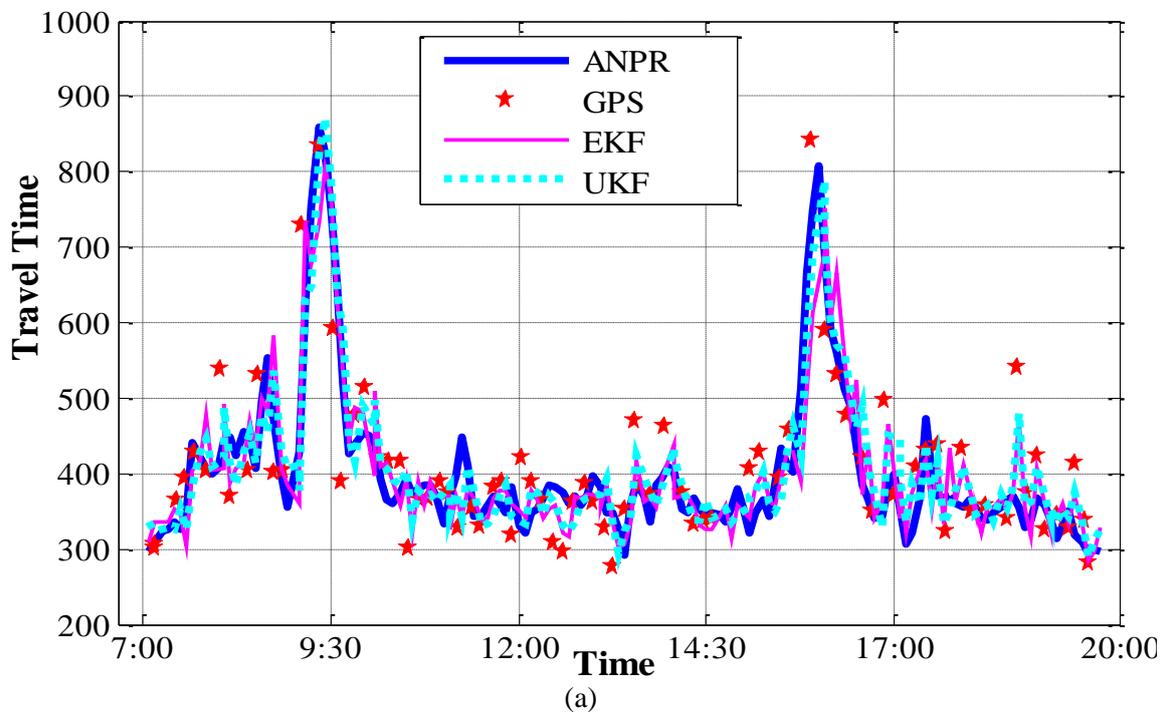
## 7.5.2 Fusion results

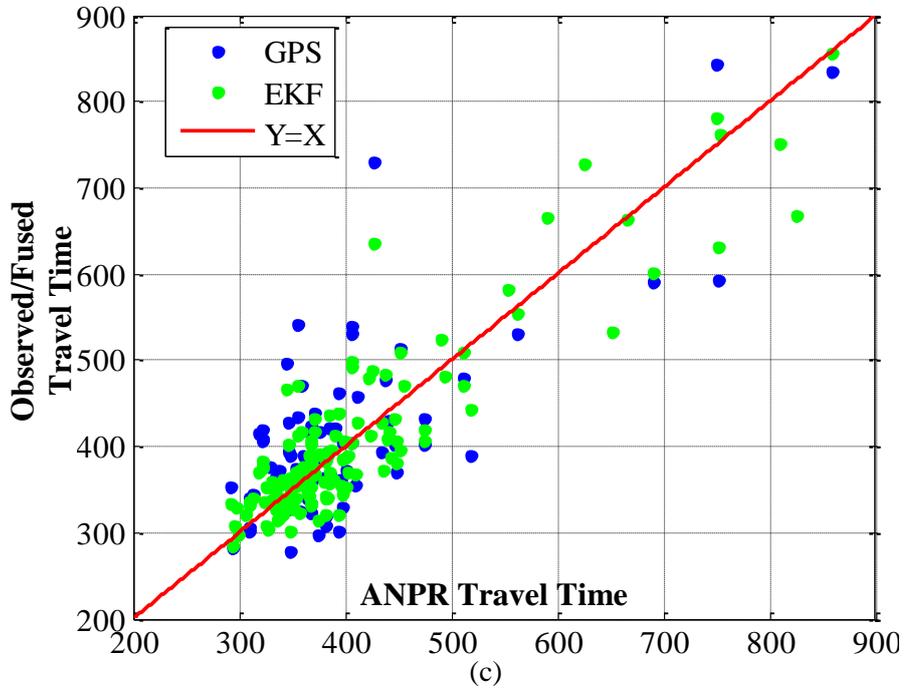
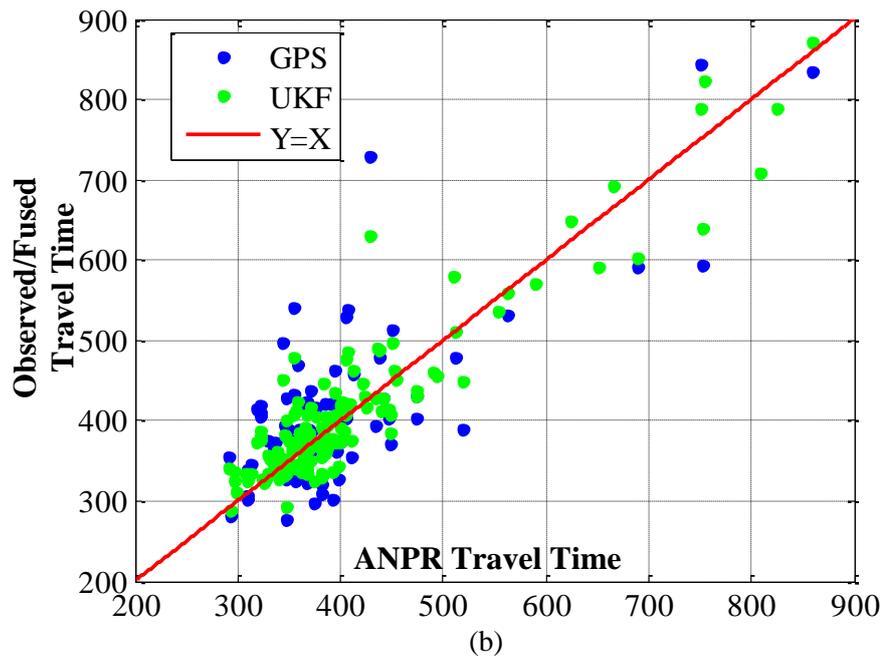
This subsection shows the fusion results from flow observations from ILDs, ATS, MTS and travel time observations from GPS. Although the GPS data is recorded at 1 second resolution, the T-resolution used in this evaluation is set to 10s. The result of sensitivity analysis with respect to other T-resolution values will be provided in next subsection.

For the source of the GPS observations, although the technology used for the data collection is GPS, the nature of this method is originally from the Moving Car Observer (MCO) which has been discussed in Section 1.1.2. Among the literature about MCO methods, Hellinga & Fu (2002) developed a method of estimating population mean travel times even when bias exists in the arrival time distribution. Based on the simulation, a 14% relative error

was shown in their paper while the relative errors from my simulation and real-world data are 11.46% and 13.35% respectively, which both are better than their improved results. Ferman et al. (2007) provided a comprehensive analytical evaluation of traffic information system based on GPS probe vehicles. Their simulated results showed that the probability that the relative error is less than 10% is around 71%, which is also close to the results used in this research.

Figure 7.6 (a) gives the overall results of the EKF and UKF fusion against the travel time observed from ANPR, and the scatter plots of each of them are given in Figure 7.6 (b)-(d). It can be seen that during the uncongested/light traffic period, the UKF and EKF show a quite similar fusion result. As the increasing of the congestion level, the performance of UKF fusion is better than EKF. The overall improvement of UKF over EKF is around 12% in terms of MAPE and RMSE. More specifically, the error variance of the UKF is smaller than the EKF. It is clear from the scatter plot figures that the UKF fusion points of large travel time (congestion period) are less dispersive than the EKF results. This finding is consistent with the discussion in Section 4.3.4 and the results from simulation in Section 6.3.2. Comparing with the GPS observations, the UKF achieves roughly 50% improvement in terms of MAPE and RMSPE.





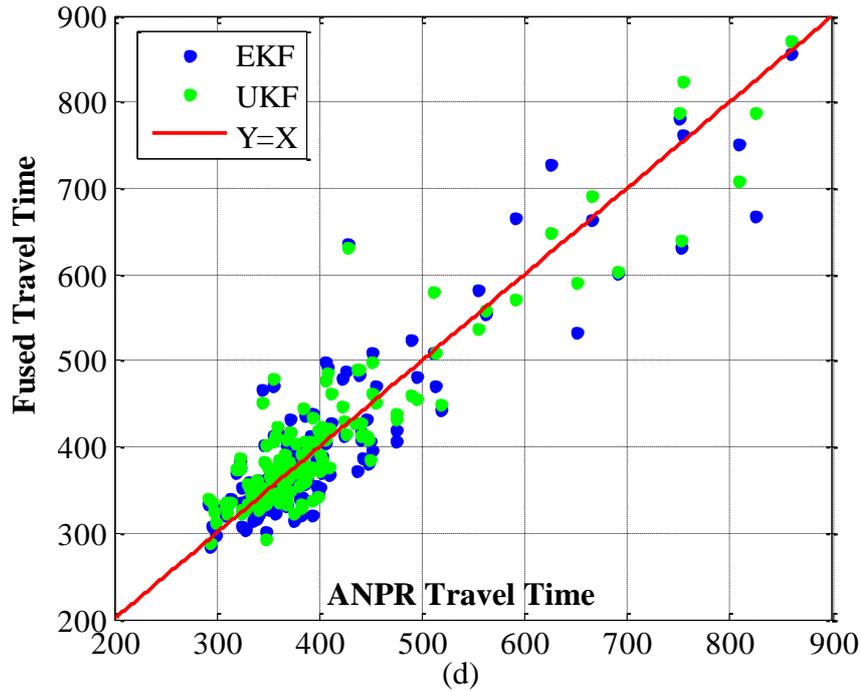


Figure 7.13 Fusion results from real-world data: (a) time plot between the EKF, UKF and GPS; (b) scatter plot between the UKF and GPS; (c) scatter plot between the EKF and GPS; (d) scatter plot between the UKF and EKF

|                   | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|-------------------|-----------|------------|------------|-------------|
| <b>UKF Fusion</b> | 1.93%     | 6.99%      | 39.78      | 9.68%       |
| <b>EKF Fusion</b> | -0.98%    | 7.85%      | 45.34      | 10.58%      |
| <b>GPS</b>        | -0.20%    | 13.35%     | 71.55      | 17.94%      |

Table 7.5 The performance of ILDs and GPS fusion from real-world data

From the result of MPE comparison, it can be seen that the GPS observation is unbiased (MPE is -0.20%) while the UKF and EKF has a slightly bias (1.93% and -0.98%). There are two sources which lead to this biased estimation results: 1) the bias introduced by ILDs and ATS due to the problem of lane-cross; 2) the bias introduced by model parameters since a perfect model calibration is difficult to be carried out in real-world scenario.

### 7.5.3 The result of GPS sensitivity analysis with respect to the T-resolution

A comprehensive sensitivity analysis has been provided in the simulation based evaluation. Due to the limitations of real-world scenario, only the sensitivity of GPS T-resolution is appropriate to be evaluated. This section evaluates the impact of the different T-resolutions on the fusion performance based on the real-world data. As mentioned in Section 7.1.4, the observation of GPS positioning is recorded every 1 second. Similar as the sensitivity analysis

in the simulation scenario, the range for the choices of T-resolution is set to {1, 5, 10, 20, 30, 45, 60(s)}. The results are shown in Figure 7.7 and Table 7.4.

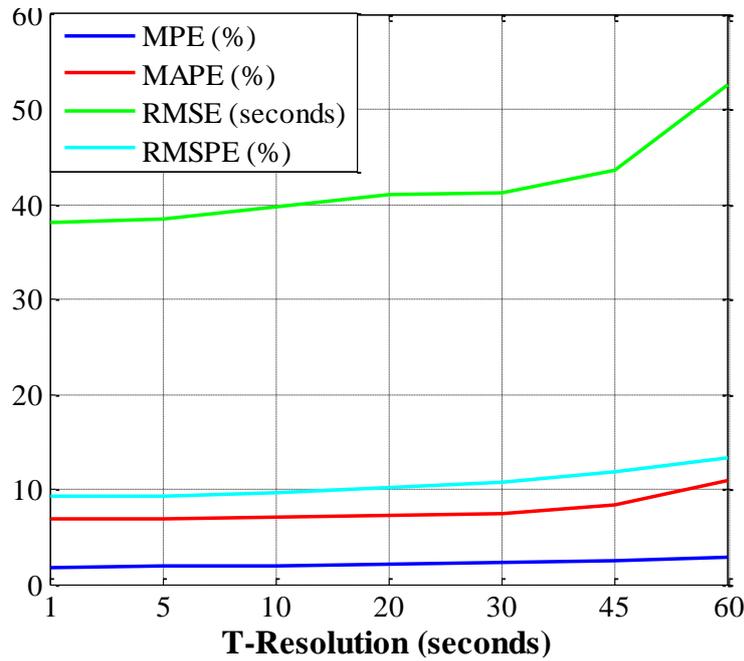


Figure 7.14 Result of sensitivity analysis about the T-resolution

| T-resolution (s) | 1      | 5      | 10     | 20     | 30     | 45     | 60     |
|------------------|--------|--------|--------|--------|--------|--------|--------|
| $J_{MPE}$        | 1.73%  | 1.85%  | 1.93%  | 2.11%  | 2.26%  | 2.47%  | 2.83%  |
| $J_{MAPE}$       | 6.88%  | 6.92%  | 6.99%  | 7.28%  | 7.45%  | 8.42%  | 10.84% |
| $J_{RMSE}$       | 38.02s | 38.43s | 39.78s | 41.09s | 41.27s | 43.59s | 52.63s |
| $J_{RMSPE}$      | 9.19%  | 9.34%  | 9.68%  | 10.12% | 10.73% | 11.81% | 13.74% |

Table 7.6 Quantitative results of sensitivity analysis about the T-resolution

From the above results, it can be seen that the fusion performance is not sensitive to the T-resolution when the resolution is higher than 60 seconds. When the value of T-resolution increased to 60s which is relatively larger than the average travel time for each segment (40s), the RMSE has a noticeable increment. This pattern of performance sensitivity about the T-resolution is same as the results from simulation, and more discussion has been given in Section 6.4.1.

### 7.5.4 The result of comparison with linear Kalman filter

This section illustrates the fusion results from both of UKF and LKF to demonstrate the improvement of the UKF over LKF. The implementation of the LKF is same as the one presented in the simulation comparison (Section 6.5.2). The comparison results are shown in Figure 7.8 and Table 7.5.

A clear overall improvement of the UKF over the LKF is shown in above results, i.e. around 30% reduce in terms of MAPE and RMSE, and 24% reduce in terms of RMSPE. More specifically, the improvement during the congestion period is more remarkable than the uncongested period. This finding is consistent with the simulation evaluation, and more discussion is given in Section 6.5.2. The MPE of the UKF is slightly larger than the LKF, which indicates that the fusion result based on the UKF has larger biased error than the LKF. It is because the inaccuracy of model parameters for the UKF has larger impact on the fusion performance than the LKF.

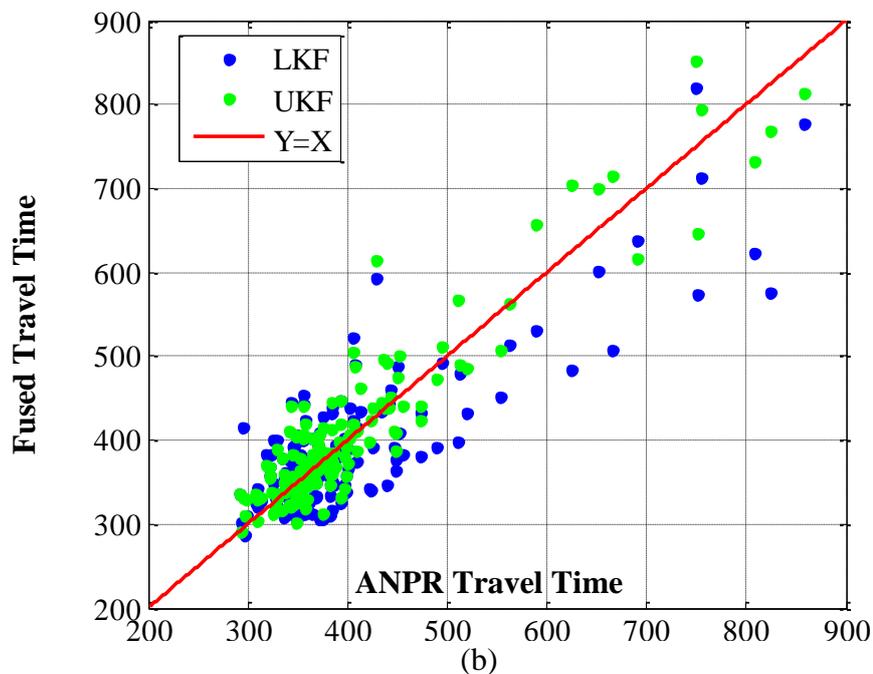
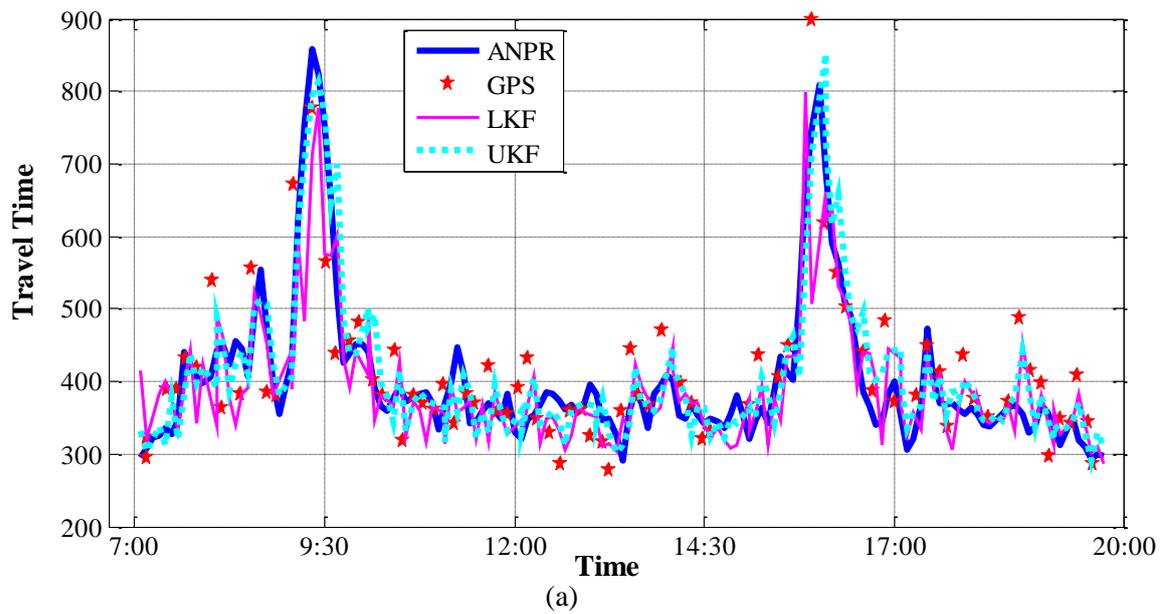


Figure 7.15 Comparison results between LKF and UKF: (a) time plot (b) scatter plot

|                              | $J_{MPE}$ | $J_{MAPE}$ | $J_{RMSE}$ | $J_{RMSPE}$ |
|------------------------------|-----------|------------|------------|-------------|
| <b>UKF</b>                   | 1.93%     | 6.99%      | 39.78      | 9.68%       |
| <b>LKF</b>                   | 1.21%     | 10.05%     | 58.41      | 12.81%      |
| <b>GPS probe observation</b> | -0.20%    | 13.35%     | 71.55      | 17.94%      |

Table 7.7 The performance comparison between UKF and LKF based on real-world data

### 7.5.5 Discussion on the minimum realistic levels of sensors configuration

The developed fusion framework is evaluated by a designed world scenario in this chapter. Both of the Automatic Traffic Surveys (ATS) and Manual Traffic Surveys (MTS) are carried out to collect the necessary traffic data which is the theoretical input for the fusion framework. However, this sensor configuration is not realistic enough to be implemented and maintained for daily use. There are two potential ways to obtain or approximate the required flow difference values by existing infrastructures:

- The signal cycles: from the discretised macroscopic traffic model which is used as the dynamic state equation in the Kalman filter, the actual required input is the difference between the inflow and outflow for each segment. For a signalised urban road link, this difference of the traffic flow is related to the controlling signal cycles. Therefore, the dynamics of the flow can be approximated by the signal cycle periods of the adjacent segments. By this means, the requirement of the upstream and downstream ILDs flow observations can be relaxed.
- The ILDs from SCOOT system: there are single SCOOT ILDs installed around each junction monitoring the turning flow. However, these flow measurements of these existing SCOOT ILDs were not available to this research. During the data collection period, the ATS and MTS are configured to collect those turning traffic volumes. The use of existed SCOOT ILDs data will eliminate the requirement of ATS and MTS as settled in the experiment

## 7.6 Summary

This chapter used the traffic data collected from real-world to evaluate the performance of the proposed fusion framework. The task of real data collection is based on a planned experiment in an interurban road link in the area of Maidstone, UK. Besides the existed ANPR camera system and three ILDs along the road link, 4 automatic and 2 manual traffic surveys are

carried out to collect the required flow data for the framework implementation. The GPS data is collected by two floating vehicles which are equipped with GPS logger and driven along the road link repeatedly for 12 hours. The framework implementation is based on the same approach as the one used in the simulation scenario. Unlike the simulation, the observation noise covariance from real-world is unknown. Section 7.3.1 proposed an approximation method to determine the level of noise variance of GPS observations.

The results in Section 7.3.2 clearly meet the expected level of fusion performance. Roughly 50% overall improvement of the UKF fusion over GPS observation can be seen, and there is 12% improvement of UKF over EKF in terms of MAPE and RMSE. The results from sensitivity analysis with respect to the T-resolution of GPS data are shown to be consistent with the findings from the simulation scenario. For the comparison between UKF and LKF, clear overall improvement of the UKF over the LKF is shown as expected, i.e., around 30% reduce in terms of MAPE and RMSE, and 24% reduce in terms of RMSPE. Through the analysis of the MPE result, the UKF has a slightly larger biased error comparing to EKF, LKF or GPS observations. It is mainly due to imperfect model calibration, which leads to the inaccurate model parameters. This problem can be reasonably improved through a more carefully model calibration by the historical data which is not available in this case.

## Chapter 8

# Conclusions & Future Research Topics

This chapter concludes the research undertaken in this thesis. It also presents a number of topics for future work based on the ideas developed in this research.

### 8.1 Conclusions

The aim of this research is to use multi-sensor data fusion technique to integrate the observations from inductive ILDs, ANPR cameras and GPS to perform better travel time estimation in both freeway and interurban environment. To achieve this aim, the research objectives are defined accordingly in Chapter 1. The rest of this thesis successfully addressed each of the objectives. This section revisits these objectives and summarises the relevant works and contributions.

**Identify the opportunities and challenges in travel time estimation based on multiple sensor sources, and determine the most appropriate fusion methodology based on the review of the state-of-the-art fusion methods**

- 1) The recent development of traffic sensor technologies provides a wide spectrum of available data and heterogeneous sources of information that are of potential use for estimating travel time. Section 2.1.1 classified the most popular sensor sources based

on the nature of measurement into three types: single point sensor, paired monitoring and probe vehicles. The features of each type of sensor were reviewed and discussed.

- 2) ILDs, ANPR and GPS were selected as the most representative sensor sources from each of the sensing groups. The issues of using each of them for travel time estimation were discussed in Section 2.1.2. ILDs have trouble measuring congested vehicle flow and only provide point-based time-mean speeds to estimate link travel times as a continuous stream value; GPS based technology cannot provide reliable traffic data collection due to the low density of devices usage and errors from both of “canyon effect” and map-matching in urban environment; ANPR matching is suffered from low matching rate.
- 3) The suitability and advantages of the data fusion technique in the multi-sensor environment was discussed in Section 2.2. The most popular data fusion techniques and their applications in the transport domain were classified and reviewed. The Kalman filter was chosen as the fusion methodology in this research due to a number of advantages over other methods which were discussed in Section 2.4.

**Develop a new method for travel time estimation based on the ILDs and ANPR data which has less limitations and can be easily implemented in practice**

- 1) A further exploration of Wardrop (1952)’s formula which describes the relationship between space-mean-speed and time-mean-speed (output from ILDs) was studied in Chapter 3. A novel approximate relationship between time-mean-speed obtained from ILDs and space-mean-speed from ANPR was proposed and used in the estimation model.
- 2) A property of traffic state dependency was illustrated, and a more refined model was presented where the traffic states were segmented according to the traffic state which was based only flow and occupancy from the ILDs data. The evaluation results by real traffic data from England Highways showed an average 54% improvement in terms of MAPE and RMSE.
- 3) The discussion in Chapter 3 showed that the proposed method has better transferability than the most relevant conventional methods such as Garber and Hoel (2001) and Rakha and Zhang (2005)’s approaches. It was also easier and has fewer limitations to implement than ANN and k-NN based methods from the aspect of traffic engineering.

### **Review and address the gaps and issues of the selected fusion technique as the estimation methodology in travel time estimation**

- 1) Chapter 2 stated the reasons of adopting the Kalman filter as the proposed fusion methodology. Chapter 4 described the details of Kalman filter theory including the nonlinear version: EKF and UKF, and discussed its relevance to the data fusion problem in this research.
- 2) A comprehensive review of the existing applications of Kalman filter in the area of traffic estimation was provided in second part of Chapter 4. The key findings from the review were twofold: a) the performance of applying linear type of Kalman filter in this field were limited by the traffic models used in the filter framework; b) the nonlinear version of Kalman filter was only applied based on the single sensor source, more specifically ILDs. The application of the nonlinear Kalman filters (EKF and UKF) for the multi-sensor fusion has not been studied.
- 3) The issues of using nonlinear Kalman filter were addressed as: 1) how to choose the appropriate traffic model for the Kalman filter structure; 2) how to determine the process and measurement noise/errors; 3) how to choose the types of Kalman filter. The model development and implementations in Chapter 5-7 fully studied these issues.

### **Develop a novel data fusion framework for travel time estimation based on the selected fusion technique**

- 1) A general Kalman filter based fusion framework was developed in Chapter 5 to achieve this objective. In choosing the appropriate traffic model for the Kalman filter structure, a number of macroscopic traffic models have been examined, and the PW dynamic model and modified exponential velocity-density model were chosen as the key equations for the state process. The observations from multiple sources were also carefully modelled, and fitted into the observation process.
- 2) To make the fusion framework available for real-world application, a discretisation process was utilised to convert the theoretically continuous traffic models into discrete temporal-spatial scenario. It can be seen that this framework makes use of the well-developed macroscopic traffic model and the observations from multi-sensor sources, i.e. all the knowledge related to the estimation problem, which the superiority of Kalman filter is as discussed in Chapter 4. Some of the issues of framework implementations have been addressed, such as such as unknown model parameters,

biased noise and missing observations. The solutions to these issues were proposed in Section 5.5

### **Implement and evaluate the proposed fusion framework by simulation**

- 1) A simulation scenario was designed to evaluate the performance of the developed fusion framework in Chapter 6. The simulation scenario is based on a segmented urban road link with junctions, which can be fitted into the space-time fusion framework perfectly. The multiple sensor sources have also been well-modelled with the consideration of noise terms.
- 2) The details of framework implementation were presented in Section 6.2. The formulations of state/observation vectors and state/observation processes were developed for the use of nonlinear Kalman filter algorithms, i.e. EKF and UKF. The noise covariance matrices of state and observation, as one of most important but difficult components, were carefully studied and explicitly derived. Comparing to other approaches about the noise covariance matrices, this research provided the analytical forms of these two matrices which take the correlation properties among the state variables into consideration.
- 3) The results of framework evaluations based on simulation were presented in Section 6.3-6.5. The key findings were summarised as follows:
  - The results from general fusion showed that the UKF outperforms the EKF which is consistent with the comparison discussion between the accuracy of the EKF and the UKF in Section 4.3.4.
  - The sensitivity analysis about impacts of GPS probe factors on the fusion performance demonstrated that 1) increasing the sampling rate within the range of 0-20% leads to noticeable improvement, and the performance tends to converge beyond 20%; 2) the noise level of GPS observations affects the fusion performance linearly; 3) the fusion performance is relatively insensible to the T-resolution (the updating frequency of GPS probes) which is at the same scale as the estimated travel time.
  - The results in Section 6.4.2 showed that the ILDs near the junctions play a more important role than the ILDs placed in the middle of two junctions, and the fusion with either junction upstream ILDs or downstream ILDs has equivalent performance.

- Through the comparison analysis with the methods based neural networks and linear Kalman filter, the results showed 1) with proper training datasets, the neural networks are more accurate than the proposed UKF method. If the historical datasets were used to calibrate the model parameters in the developed fusion framework, the UKF method is slightly worse than the recurrent neural network and slightly better than the feed forward neural network; 2) Comparing to the methods based on the linear Kalman filter, the proposed UKF method has an obvious improvement (around 20% in terms of MAPE, RMSPE).

### **Implement and evaluate the proposed fusion framework by real-world data**

- 1) An experiment of real traffic data collection was carried out to evaluate the performance of the proposed fusion framework. The task of real data collection was based on a planned experiment in an interurban road link in the area of Maidstone, UK. Besides the existed ANPR camera system and three ILDs along the road link, 4 automatic and 2 manual traffic surveys were designed and implemented to collect the required flow data for the framework implementation. The GPS data was collected by two floating vehicles which were equipped with GPS logger and driven along the road link repeatedly for 12 hours. More detailed description of the experiment was provided in Chapter 7.
- 2) The framework implementation for the real-world data was based on the same approach as the one used in the simulation scenario. Unlike the simulation, the observation noise covariance from real-world was unknown. Section 7.3.1 proposed an approximation method to determine the level of noise variance of GPS observations.
- 3) The results of framework evaluations based on real traffic data were presented in Section 7.3. The key findings were summarised as follows:
  - The results in Section 7.3.2 clearly meet the expected level of fusion performance. Roughly 50% overall improvement of the UKF fusion over GPS observation can be seen, and there is 12% improvement of UKF over EKF in terms of MAPE and RMSE.
  - The results from sensitivity analysis with respect to the T-resolution of GPS data are shown to be consistent with the findings from the simulation scenario.

- For the comparison between UKF and LKF, clear overall improvement of the UKF over the LKF is shown as expected, i.e., around 30% reduce in terms of MAPE and RMSE, and 24% reduce in terms of RMSPE.
- Through the analysis of the MPE result, the UKF has a slightly larger biased error comparing to EKF, LKF or GPS observations. It is mainly due to imperfect model calibration, which leads to the inaccurate model parameters. This problem can be reasonably improved through a more carefully model calibration by the historical data which is not available in this case.

## **8.2 Recommendations for further research**

This section identifies the limitations within the current work, and outlines the possible approaches for the improvement. Meanwhile, a number of potential research topics related to the work in this PhD work are also recommended.

### **Integrating the traffic state estimates into the proposed travel time estimation method**

Chapter 3 demonstrates that the proposed estimation method has a property of traffic state dependency. The approach to this problem in Chapter 3 is based on considering the traffic state as a determined factor which is either a congested state or an uncongested state. It is obvious that this traffic state identification cannot fully interpret those states falling into the transition zone between uncongested and congested conditions. During the course of this PhD research, a probabilistic traffic state identification method has been developed and presented in Appendix A. This method is able to assign a fair value of congestion probability to each observation. Hence, there is potential in integrating this probabilistic state identification method into the proposed travel time estimation model and improving the modelling inaccuracy causing by the varying traffic state.

### **Investigating the possibilities of implementing other traffic models into the fusion framework**

The traffic model used for the Kalman filter fusion is PW model. This research has not investigating the use of other types of traffic models in the fusion framework. For example, in a busy signalised urban road link, the traffic behaviour cannot be well described in the PW model. It requires the aids from other traffic model such as queueing model to make the Kalman filter based fusion framework capture the dynamics of traffic flow in more detailed level. The main difficulty existing in this research direction is that the other traffic models

may not have a suitable time-difference formulation which can be directly applied into the state process of Kalman filter.

### **Applying the other more advanced filtering techniques**

The core technique used in the proposed fusion framework is nonlinear Kalman filter: EKF and UKF. Within the domain of emerging statistics and signal processing, there exist other types of more advanced estimation techniques such as wavelet and particle filter.

- **The wavelet** as an advanced signal processing technique has already been used in the transport area (Ghosh & Adeli 2003; Qiao, et al. 2003; Chen, et al. 2004; Venkatanarayana, et al. 2005). The application of using wavelet in the travel time estimation based on multi-sensor fusion is still a gap. The recommended research direction of applying wavelet into this field is from two aspects: sensor data denoising and traffic pattern extraction.
- **The particle filters** are normally used as an alternative to the EKF and UKF in the estimation process. With sufficient samples, they approach the Bayesian optimal estimate, and the particle filters can be more accurate than either EKF or UKF in theory (Doucet & Johansen 2008). Thus, there is an opportunity to apply particle filter to improve the fusion performance in the area of travel time estimation.

### **Making use of the more position information**

- **GPS observations:** the use of GPS probes as one of sensor sources is evaluated based on a designed experiment. The collection of GPS data is basically from the MCO method. A number of issues such as map-matching error and low spatial and temporal coverage have not been fully examined in this research. On the other hand, more detailed data from advanced GPS receivers can be benefit for fully analysing the observation error sources and uncertainties, e.g., signal strength, level of signal-to-noise-ratio, information about the satellites, and dilution of precision. By having these extra data about the working condition of the GPS devices, the error structure can be modelled in a time-varying and fashion to improve the fusion performance.
- **Mobile phones as another sensor source:** the data from GPS is used as one of the sensor input for the purpose of fusion based travel time estimation. Comparing to the GPS data, the positioning data from pervasive mobile phones has similar nature of observations. Although the observations errors from mobile phones are relatively larger than GPS, the much huger amount of mobile phone data leads to a promising

source of sensor data for the travel time estimation. Besides, the data retrieving from mobile phones contains much richer information about the people's travel behaviours than the data from the vehicle based GPS devices.

## **Appendix A**

### **A New Method for Probabilistic Traffic State Identification Using ILDs**

Identifying relevant traffic states is an important problem in a number of areas of traffic operations and control, especially those involving real time ITS systems. A simple but particularly important version of the state estimation problem is to determine whether or not traffic is congested, based on data from inductive loop detectors (ILDs). This problem arises in various forms in both urban and interurban contexts. For example, Chapter 3 explored how traffic state could affect estimation model, and a k-means based data clustering method is developed to improve model performance. This chapter investigates the problem of traffic state identification further, and proposes a probabilistic based identification method, i.e. outputting a probability of congestion for each observation.

A number of methods to automatically determine the traffic state are available in the academic literature including simple classifiers based on thresholds and more sophisticated approaches based on fuzzy logic. Most of the methods however, either require auxiliary data inputs that are not always readily available in real time or depend on loop-specific parameters that are not readily transferrable across sites. Moreover, most methods encounter difficulties in dealing with traffic states that fall into the transition zone between uncongested and congested conditions.

The contribution of this chapter is to develop a novel new approach to state estimation based on analysis of loop occupancy and flow data alone. The method is based on the assumption that the relationship between flow and occupancy displays distinct regimes according to whether the system is congested or uncongested. A probabilistic classifier is

developed, based on the Expectation- Maximisation (EM) algorithm. This new algorithm is evaluated using data from the ILDs on highways and urban links. The results demonstrate that the proposed algorithm, which does not need site specific calibration, is able to identify the traffic state on both urban and highway satisfactorily.

## **A.1 Introduction**

Identifying the traffic state is an important problem that has practical applications in a number of ITS areas. Stated simply, the problem of traffic state identification is to determine if the traffic is congested or uncongested. The information about traffic state can be treated as an input into traffic management and control system. For example, if current traffic state is determined as congested, a corresponding traffic control action could be carried out to ease the traffic flow. In the EPSRC and TSB funded FREEFLOW project in the UK, a learning based system is being developed to recommend traffic control interventions if traffic problems are detected. One of the inputs required to identify traffic problems is information regarding whether a given link is congested or uncongested. In addition, traffic state as an indicator of level of congestion can also be used in travel time estimation. Han et al. (2009) investigated the relationship between Time-Mean-Speed (TMS) and Space-Mean-Speed (SMS) under different traffic states. Although the authors treated the traffic states as a parameter in their travel time estimation model, the states classification method is based on pre-determined segments and lack of transferability.

The process of traffic state identification is dependent on the observations from available traffic detectors. Inductive loop detectors (ILDs) as the most widely used type of traffic detectors are chosen as the data source for developing the proposed identification method. The output from urban ILDs normally contains flow and occupancy information, while highway ILDs can additionally output speed data. In order to make the proposed method compatible to both urban and interurban context, our approach is based on ILD flow and occupancy data alone. Typically, classifying traffic states into congested and uncongested by using flow and occupancy data from ILDs has two difficulties. The first one is to make the method generalised and transferrable. Due to the differences among sensitivities and configurations of various types of ILDs, the congested/uncongested flow-occupancy features do not always constrain to identical pattern among different ILD sites. The second one is that it is not reasonable to simply determine the traffic state as either congested or uncongested, especially for those flow-occupancy observations which fall into the transition zone.

Regarding to the above two difficulties, the aim of this paper is to develop a general and transferrable method to identify traffic state in a probabilistic fashion.

## **A.2 Background**

A number of attempts to automatically determine the traffic state are available in the academic literature. Lao et al. (2007) use Fuzzy Logic to classify the traffic state into uncongested, “crowded” and congested using Fuzzy Logic; however, they used driver inputs rather than traffic sensor data. Narayanan et al. (2003) also used Fuzzy Logic to classify traffic using speed and inter-vehicle distance as input variables, using fixed thresholds in their classification method. Threshold based methods are generally not transferrable since the occupancy values reported by each ILD will depend on its electromagnetic sensitivity, and the thresholds could be different for different ILDs. Jiang et al. (2003) used Fuzzy Clustering of traffic sensor data consisting of flow, occupancy and spot-speed to cluster traffic into four states representing increasing levels of congestion. Of the above models, only the method presented in Jiang et al. (2003) provide a method to automatically identify the traffic state using traffic sensor data. However, the spot-speed data required in this work is not available from urban ILDs in UK. Moreover, the traffic states do not correspond to known traffic states in traffic engineering, though this criticism can be addressed by reducing the number of clusters in the proposed method. However, it is not clear if the modified method will correctly classify traffic into congested and uncongested states.

## **A.3 Methodology**

### **A.3.1 Problem description**

Based on the objective of this research, flow and occupancy data from ILDs is the only input for state identification. According to Daganzo (1997), the ideal relationship between flow and occupancy is given in Figure 1 (a). The lower segment of the plot represents samples from the uncongested regime, while the upper segment represents the congested state. The scatter plot between real flow and occupancy data used in this study is given in Figure 5(b)-(d). It is clear that it is easy to determine if a given data point (flow-occupancy point) is congested or not from the scatter plot by visual inspection.

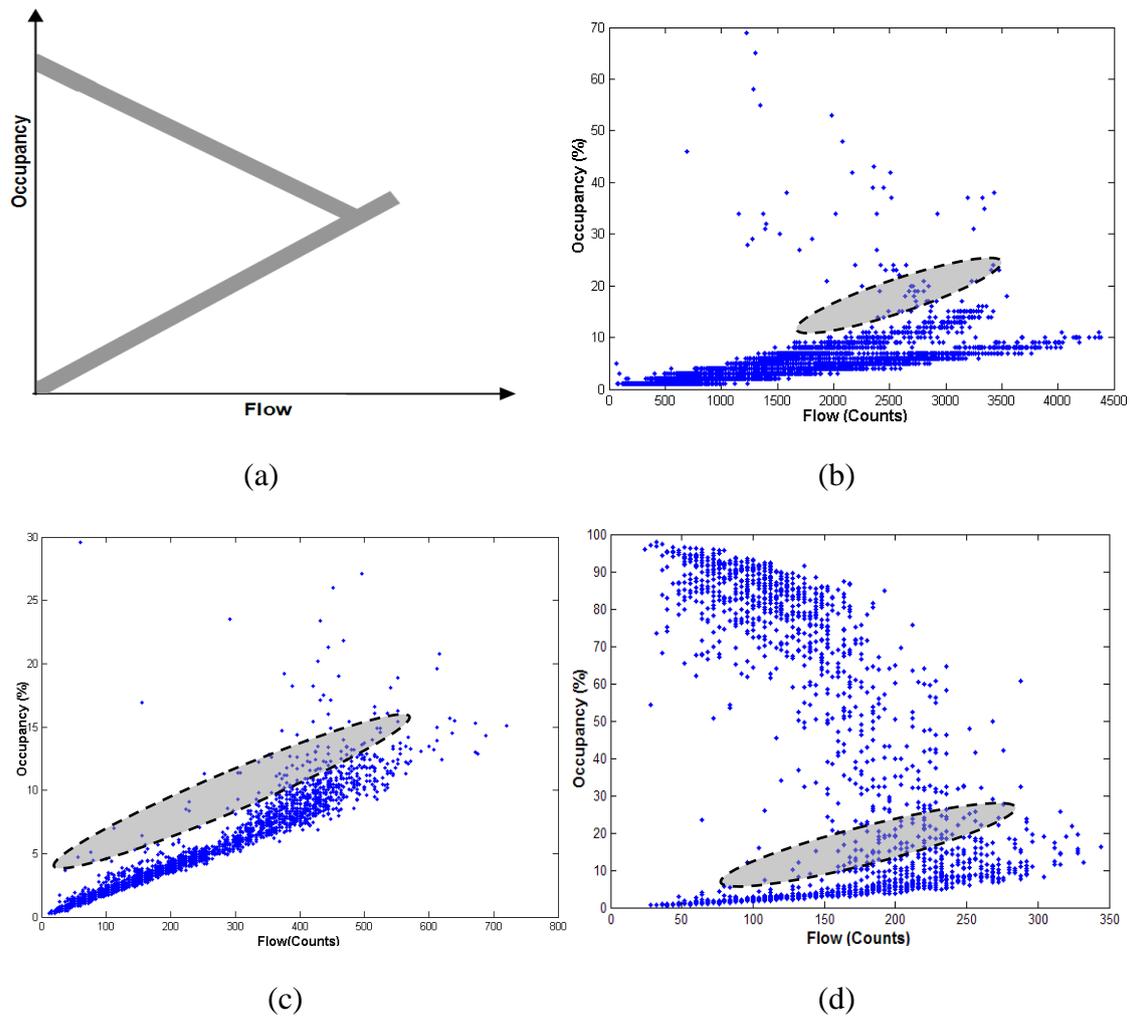


Figure A.1 Relationship between flow and occupancy

From the flow-occupancy scatter plots in Figure 5(b)-(d), the two main difficulties discussed in the section of Introduction can be further explained. These three plots are from three different ILD sites and have distinct traffic patterns. Firstly, it can be seen that the ratios between flow and occupancy from uncongested regime have great difference. A simply pre-determined threshold can not be applied for every case. Secondly, the observations from the dark areas belong to the transition regime. Identifying those observations as absolutely congested or uncongested is not fair because there is no explicit and unique definition on congested/uncongested traffic state through flow and occupancy data. Hence, it is more reasonable to assign a probability of congested for each observation.

### A.3.2 Problem modelling

According to the description in last section, the traffic state is related to the ratio between flow and occupancy which is defined as follows:

$$\alpha_i = \frac{o_i}{q_i} \quad (\text{A.1})$$

where  $o_i$  and  $q_i$  are the occupancy and flow of the  $i$ th observation measured by ILD, respectively. The traffic state of  $i$ th observation is represented by notation  $Z_i$  with  $Z_i = \{0, 1\}$ . It assumes that the traffic states are generated from two distinguishable regimes which are either congested ( $Z_i = 1$ ) or uncongested ( $Z_i = 0$ ). The second assumption made in this modelling is that the probability distributions of both congested and uncongested states are Gaussian distribution, shown as follows:

$$p(\boldsymbol{\alpha})|_{Z=0} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2) \quad (\text{A.2})$$

$$p(\boldsymbol{\alpha})|_{Z=1} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2) \quad (\text{A.3})$$

where  $p(\boldsymbol{\alpha})|_Z$  is the probability density function of  $\boldsymbol{\alpha}$  given by a traffic state, and  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  is Gaussian distribution with mean  $\boldsymbol{\mu}$  and variance  $\boldsymbol{\sigma}^2$ . Then, the problem is modelled as solving the probability of  $P(Z_i=1|\boldsymbol{\alpha}=\alpha_i)$ , from Bayesian theory:

$$P(Z_i=1|\boldsymbol{\alpha}=\alpha_i) = \frac{P(Z_i=1, \alpha_i)}{P(\alpha_i)} = \frac{P(\alpha_i | Z_i=1)P(Z_i=1)}{P(\alpha_i)} \quad (\text{A.4})$$

According to Equation (2) and (3):

$$P(\alpha_i) = \gamma_0 \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2) + \gamma_1 \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2) \quad (\text{A.5})$$

where  $\gamma_0=P(Z_i=0)$  and  $\gamma_1=P(Z_i=1)$  are the mixture factors, and  $\gamma_0 + \gamma_1=1$ . The Equation (5) is a typical Gaussian mixture model (GMM) with unknown parameters  $\Theta = (\gamma_0, \gamma_1, \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2, \boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$ . In this case, a probabilistic model is defined as:

$$p(\boldsymbol{\alpha} | \Theta) = \sum_k \gamma_k p_k(\boldsymbol{\alpha} | \theta_k) \quad (\text{A.6})$$

Each  $p_k$  is a Gaussian distribution function parameterised by  $\theta_k$ , where  $\theta_k = (\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)$  and  $k = \{0, 1\}$ . The expression in Equation (4) can be rewritten as:

$$P(Z=1|\boldsymbol{\alpha}=\alpha_i) = \frac{\gamma_1 p_1(\boldsymbol{\alpha}=\alpha_i | \theta_1)}{p(\boldsymbol{\alpha}=\alpha_i | \Theta)} \quad (\text{A.7})$$

From Equation (7), it can be seen that only the parameters  $\Theta$  of GMM is unknown to our model. Solving unknown parameters for a probabilistic model can be considered as well-known Maximum likelihood estimation (MLE) problem:

$$\mathcal{L}(\Theta | \boldsymbol{\alpha}) = p(\boldsymbol{\alpha} | \Theta) \quad (\text{A.8})$$

Based on MLE theory, the goal is to find the  $\Theta$  that maximises  $\mathcal{L}$ , i.e.

$$\Theta^* = \arg \max_{\Theta} \mathcal{L}(\Theta | \alpha) \quad (\text{A.9})$$

Expectation-Maximization (EM) algorithm is an elaborate technique to find the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set (McLachlan & Krishnan 2008). Hence, our approach adopts EM algorithm to resolve the above problem.

### A.3.3 Expectation-Maximisation (EM) algorithm

In statistics, expectation-maximisation (EM) algorithm is used to find maximum likelihood estimates (MLE) of parameters in probabilistic models, where the model depends on unobserved latent variables. EM algorithm as an iterative method alternates between performing an expectation (E) step and a maximisation (M) step. The E step computes the expectation of the log likelihood with respect to the current estimate of the distribution, and the (M) step computes the parameters which maximize the expected log likelihood found on the E step. Given a likelihood function  $\mathcal{L}(\Theta | \alpha)$ , the E and M steps are generally expressed as follows:

$$\text{E step:} \quad Q(\Theta, \Theta^{(l-1)}) = E[\log \mathcal{L}(\Theta | \alpha) | \alpha, \Theta^{(l-1)}] \quad (\text{A.10})$$

$$\text{M step:} \quad \Theta^{(l)} = \arg \max_{\Theta} Q(\Theta, \Theta^{(l-1)}) \quad (\text{A.11})$$

where  $\Theta^{(l-1)}$  are the current parameters estimates that is used to evaluate the expectation and  $\Theta$  are new parameters that needs to be adjusted at each iteration, and ultimately will be optimized in an attempt to maximize the likelihood. These two steps are repeated as necessary. Each iteration is guaranteed to increase the log likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function. The details and theory behind the EM algorithm can refer to Dempster (1977), Redner & Walker (1984) and McLachlan & Krishnan (2008).

### A.3.4 Applying EM algorithm to the model

Prior to elaborating the process of applying EM algorithm, another issue is necessarily addressed. According to the problem modelling in section 3.2, the goal is to estimate the parameters  $\Theta$  of GMM by using flow-occupancy observations  $\alpha$ . However,  $\alpha$  as the historical

data set recoded by ILDs, is only part of the complete data set and may have missing data. Hence,  $\alpha$  can not represent the whole Gaussian mixture distribution. It is assumed that a complete data set exists  $\mathbf{w}=(\alpha, \mathbf{B})$ ,  $\alpha$  is called incomplete data and  $\mathbf{B}$  is called missing data. Taking the missing data  $\beta$  into consideration, the  $Q$  function in EM algorithm (Equation 10) becomes:

$$Q(\Theta, \Theta^{(l-1)})=E[\log \mathcal{L}(\Theta | \alpha, \beta) | \alpha, \Theta^{(l-1)}] \quad (\text{A.12})$$

Note that  $\beta$  is unknown, random, and presumably governed by the underlying Gaussian mixture distribution. Since both  $\alpha$  and  $\beta$  are independent variables for the whole data set,  $Q(\Theta, \Theta^{(l-1)})$  in E step can be expanded as follows (Bilmes 1998):

$$\begin{aligned} Q(\Theta, \Theta^{(l-1)}) &= \sum_{\beta} \log(\mathcal{L}(\Theta | \mathbf{a}, \beta)) p(\beta | \mathbf{a}, \Theta^{(l-1)}) \\ &= \sum_k \sum_i \log(\gamma_k p_k(\alpha_i | \theta_k)) p(k | \alpha_i, \Theta^{(l-1)}) \\ &= \sum_k \sum_i \log(\gamma_k) p(k | \alpha_i, \Theta^{(l-1)}) + \sum_k \sum_i \log(p_k(\alpha_i | \theta_k)) p(k | \alpha_i, \Theta^{(l-1)}) \quad (\text{A.13}) \end{aligned}$$

To maximise the above expression, only the term containing  $\gamma_k$  and the term containing  $\theta_k$  are required to be maximised since they are not related. The expression for  $\gamma_k$  can be found by solving following equation:

$$\frac{\partial}{\partial \gamma_k} \left[ \sum_k \sum_i \log(\gamma_k) p(k | \alpha_i, \Theta^{(l-1)}) + \lambda (\sum_k \gamma_k - 1) \right] = 0 \quad (\text{A.14})$$

or

$$\sum_i \frac{1}{\gamma_k} p(k | \alpha_i, \Theta^{(l-1)}) + \lambda = 0 \quad (\text{A.15})$$

where  $\lambda$  is the Lagrange multiplier with the constraint that  $\sum_k \gamma_k = 1$ . Summing both sides over  $k$ , it can be obtained that  $\lambda = -N$ ,  $N$  is the number of observations  $\alpha$ , resulting in:

$$\gamma_k = \frac{1}{N} \sum_i p(k | \alpha_i, \Theta^{(l-1)}) \quad (\text{A.16})$$

Since

$$p_k(\alpha | \theta_k) = \frac{1}{2\pi\sigma_k} \exp \left[ -\frac{1}{2} \left( \frac{\|\alpha - \mu_k\|}{\sigma_k} \right)^2 \right] \quad (\text{A.17})$$

the term containing  $\theta_k$  is reformed as follows:

$$\sum_k \sum_i \left( -\frac{1}{2} \log \sigma_k - \frac{1}{2\sigma_k} (\alpha_i - \mu_k)^2 \right) p(k | \alpha_i, \Theta^{(l-1)}) \quad (\text{A.18})$$

Taking the derivative of above term with respect to  $\mu_k$  and  $\sigma_k$ , then setting to zero:

$$\sum_i \frac{1}{\sigma_k} (\alpha_i - \mu_k) p(k | \alpha_i, \Theta^{(l-1)}) = 0 \quad (\text{A.19})$$

$$\sum_i [\sigma_k - (\alpha_i - \mu_k)^2] p(k | \alpha_i, \Theta^{(l-1)}) = 0 \quad (\text{A.20})$$

The expression for  $\mu_k$  and  $\sigma_k$  can be obtained by solve above two equations:

$$\mu_k = \frac{\sum_i \alpha_i p(k | \alpha_i, \Theta^{(l-1)})}{\sum_i p(k | \alpha_i, \Theta^{(l-1)})} \quad (\text{A.21})$$

$$\sigma_k = \frac{\sum_i (\alpha_i - \mu_k)^2 p(k | \alpha_i, \Theta^{(l-1)})}{\sum_i p(k | \alpha_i, \Theta^{(l-1)})} \quad (\text{A.22})$$

The Equations (16), (21) and (22) are the estimates for unknown parameters  $\Theta$  of assumed two dimensions Gaussian mixture distribution. They perform both the expectation step and the maximization step simultaneously. The process of the EM algorithm is to use the current derived parameters as the updates for next iteration. The likelihood of Gaussian mixture distribution parameterized by current estimate  $\Theta$  is guaranteed to be larger than the likelihood from last iteration.

### A.3.5 Error handling

The problem modelling and proposed solution are based on the assumption of Gaussian distribution for data points in congested and uncongested states. This assumption is valid for most of cases. However, when ILD data is corrupt or during extreme traffic conditions, the ratio  $\alpha$  from ILD observations may be exceptionally low (very high flow but low occupancy), which results in the model identifying that the traffic is in congested state. This phenomenon of low  $\alpha$  value happens very rarely, based on our exploration of real ILD data. Statistically, this can be expressed as follows:

$$\alpha^* < \mu_0 - k\sigma_0 \quad (k=3) \quad (\text{A.23})$$

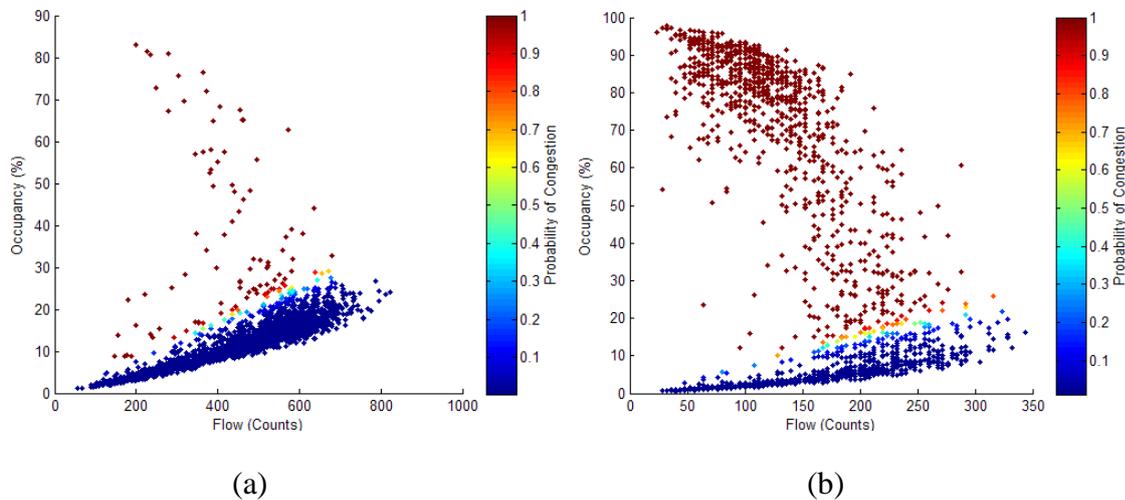
$$N^* \ll N \quad (\text{A.24})$$

where  $N^*$  is the number of abnormally low  $\alpha$  value observations, and  $N$  is the number of all uncongested observations. The EM algorithm is based on the assumption of Gaussian distribution of data sets. For such abnormal and extreme cases, a high probability of congestion may be assigned to that observation. In order to avoid this incorrect state identification, a further error handling module is added to the model. This module is summarised as follows.

- 1) According to the value of  $\mu_0$  and  $\sigma_0^2$  estimated by the EM algorithm, calculate the lower bound for  $\alpha$  using Equation (23).
- 2) If the value of  $i$ th observation  $\alpha_i$  is smaller than  $\mu_0 - k\sigma_0$ , then assign 0 as the probability of congestion for that observation.

## A.4 Evaluation & Empirical Results

The proposed traffic state identification method is evaluated using real traffic data from both urban and highway ILD sites. In order to evaluate the method under complex traffic conditions, the urban scenario is chosen from Russell Square corridor which is one of the arteries leading into central London from north London. The more details about this arterial and its ILD configurations can be found in Krishnan (2008b). One month of ILD data consisting of flow and occupancy readings at 15-minute intervals from this corridor were used for testing. For the inter-urban scenario, ILDs from English motorways were used. Data for these ILDs were obtained from the DATEX-II feed disseminated by the National Traffic Control Centre (NTCC) (National Traffic Control Centre 2009). One month of ILD data consisting of flow and occupancy readings at 5-minute intervals were used for testing.



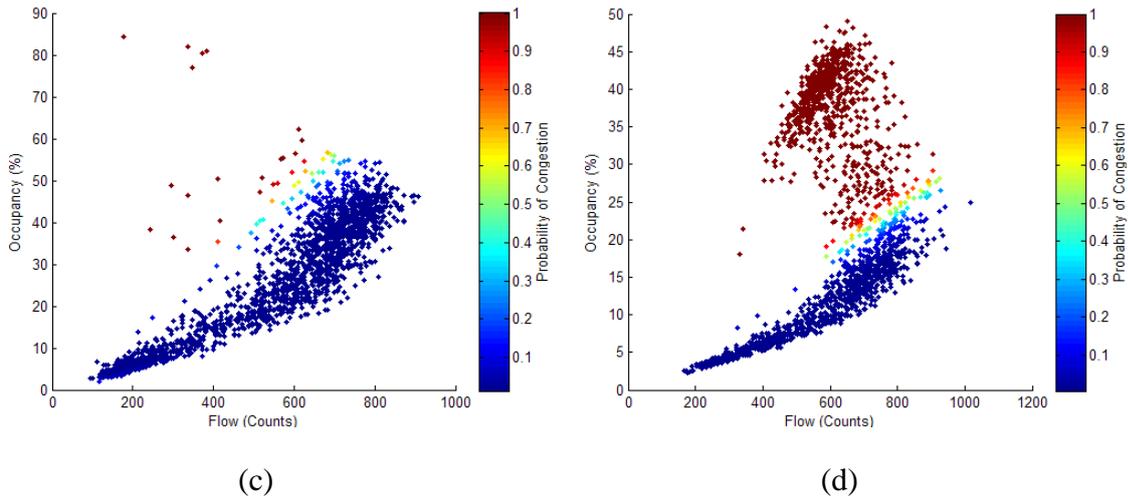


Figure A.2 Identification results of four different urban ILD outputs

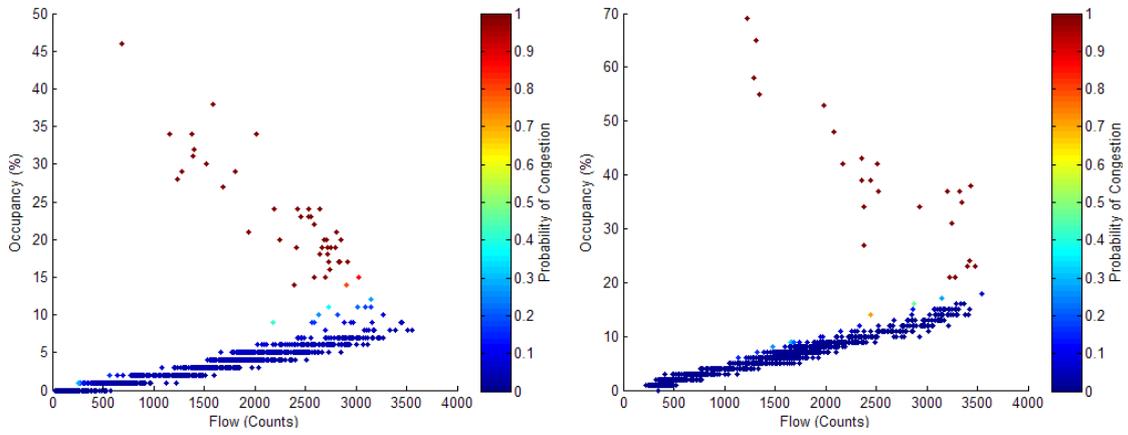


Figure A.3 Identification results of two different highway ILD outputs

From the flow-occupancy scatter plot in Figure 2 and 3, it can be seen that the proposed method effectively identifies the traffic states and assigns reasonable probability of congestion for the observations. Due to the difference of ILD sensitivities and the context of observing traffic, these six figures show outstandingly different patterns of traffic states. The results of state identification illustrate the effectiveness of the proposed method for a variety of traffic scenarios, which demonstrates the generality and transferability of the proposed EM based method. By visual inspection, the estimated probability of congested state is credible as expected. Especially when the observations fall into the transition regime, the proposed method solves the problem that identifies the traffic states either congested or uncongested successfully.

## **A.5 Conclusion**

This paper presented a novel method using EM algorithm to identify traffic state as the probability of congested/uncongested. According to the objective of the proposed research, the problem was modelled as finding the unknown parameters for a two-dimension Gaussian mixture distribution. Based on MLE theory, EM algorithm was applied to estimate these unknown parameters optimally by maximising the likelihood of the underlying Gaussian mixture distribution. The new method was evaluated using flow-occupancy data from the ILDs on urban and highways links. The empirical results demonstrated that the proposed method, which does not need site specific calibration, is satisfactorily general and transferable and provides a probabilistic oriented solution to traffic state identification.

## Appendix B

### A Complete Derivation of Wardrop (1952)'s Formula

This appendix provides step by step derivation of Wardrop's formula introduced in Chapter 3.

#### B.1 Distribution of Speed in Time

Suppose that there are subsidiary stream with flow  $q_1, q_2, \dots, q_c$  and speed  $v_1, v_2, \dots, v_c$ . Let the total flow be given by

$$Q = q_1 + q_2 + \dots + q_c = \sum_{i=1}^c q_i \quad (\text{B. 1})$$

and let

$$f_1 = \frac{q_1}{Q}; f_2 = \frac{q_2}{Q}; \dots; f_c = \frac{q_c}{Q} \quad (\text{B. 2})$$

then,  $f_1, f_2, \dots, f_c$  are the frequencies in time of vehicles whose speed are  $v_1, v_2, \dots, v_c$ , and

$$\sum_{i=1}^c f_i = 1 \quad (\text{B. 3})$$

#### B.2 Distribution of Speed in Space

Consider the subsidiary stream with flow  $q_i$  and speed  $v_i$ . The average time-interval between its vehicles is evidently  $1/q_i$ , and the distance travelled in this time is  $v_i/q_i$ . It follows that

the density of this stream in space the number of vehicles per unit length of road at any instant (the concentration) is given by

$$k_i = \frac{q_i}{v_i}, \quad i = 1, 2, \dots, C \quad (\text{B. 4})$$

Total concentration is given by

$$K = \sum_{i=1}^c k_i \quad (\text{B. 5})$$

Then,

$$f'_i = \frac{k_i}{K} \quad (\text{B. 6})$$

gives the frequencies  $f'_1, f'_2, \dots, f'_c$  of  $v_1, v_2, \dots, v_c$  in space.

### B.3 Mean Speed

With each of above two distribution of speed, there is associated a mean value, given by

**Time-mean-speed:**

$$\bar{v}_t = \sum_{i=1}^c \frac{q_i}{Q} v_i = \sum_{i=1}^c f_i v_i \quad (\text{B. 7})$$

**Space-mean-speed:**

$$\bar{v}_s = \sum_{i=1}^c \frac{k_i}{K} v_i = \sum_{i=1}^c f'_i v_i \quad (\text{B. 7})$$

Note:

$$k_i v_i = q_i \quad (\text{B. 8})$$

$$\bar{v}_s = \sum_{i=1}^c \frac{q_i}{K} = \frac{Q}{K} \quad (\text{B. 9})$$

$$Q = K \bar{v}_s \quad (\text{B. 10})$$

### B.4 Derivative of the Formula

$$\begin{aligned} \bar{v}_t &= \sum_{i=1}^c \frac{q_i}{Q} v_i \quad (\text{accordind to B. 7}) \\ &= \sum_{i=1}^c \frac{k_i v_i^2}{Q} \quad (\text{according to B. 4}) \end{aligned}$$

$$\begin{aligned}
&= K \sum_{i=1}^c \frac{f'_i v_i^2}{Q} \quad (\text{according to B. 6}) \\
&= \sum_{i=1}^c \frac{f'_i v_i^2}{\bar{v}_s} \quad (\text{according to B. 10}) \\
&= \sum_{i=1}^c \frac{f'_i [\bar{v}_s + (v_i - \bar{v}_s)]^2}{\bar{v}_s} \\
&= \bar{v}_s \sum_{i=1}^c f'_i + \frac{1}{\bar{v}_s} \sum_{i=1}^c f'_i (v_i - \bar{v}_s)^2 + 2 \sum_{i=1}^c f'_i (v_i - \bar{v}_s) \quad (\text{B. 11})
\end{aligned}$$

Since,

$$\begin{aligned}
\sum_{i=1}^c f'_i (v_i - \bar{v}_s) &= \sum_{i=1}^c \frac{k_i}{K} \left( v_i - \frac{Q}{K} \right) \\
&= \sum_{i=1}^c \frac{k_i v_i}{K} - \sum_{i=1}^c \frac{k_i Q}{K^2} \\
&= \sum_{i=1}^c \frac{k_i v_i}{K} - \sum_{i=1}^c k_i \left( \sum_{j=1}^c \frac{q_j}{K^2} \right) \\
&= \sum_{i=1}^c \frac{k_i v_i}{K} - \sum_{j=1}^c \frac{q_j}{K^2} \left( \sum_{i=1}^c k_i \right) \\
&= \sum_{i=1}^c \frac{k_i v_i}{K} - \sum_{j=1}^c \frac{q_j}{K^2} K \\
&= \sum_{i=1}^c \frac{k_i v_i}{K} - \sum_{j=1}^c \frac{q_j}{K} \\
&= \sum_{i=1}^c \frac{k_i v_i - q_j}{K} \\
&= 0 \quad (\text{B. 12})
\end{aligned}$$

Substituting the equation B.11 with equation B.12,

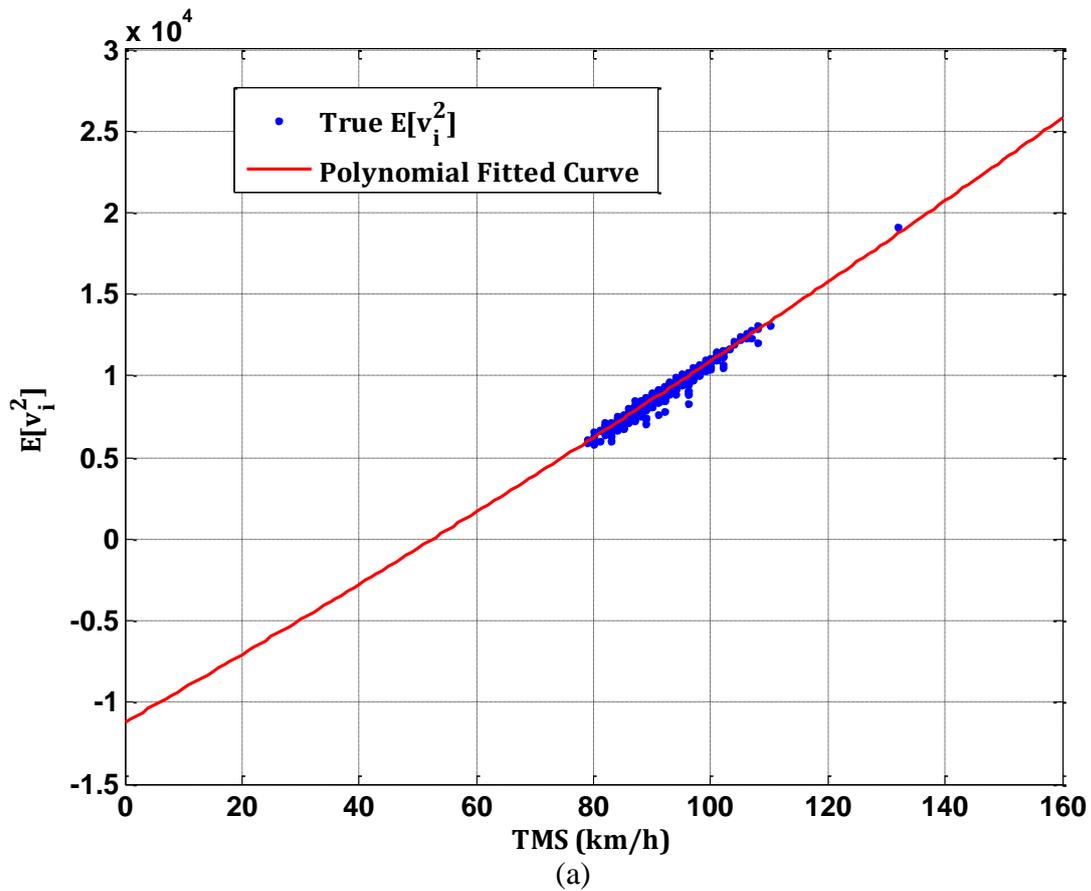
$$\begin{aligned}
\bar{v}_t &= \bar{v}_s \sum_{i=1}^c f'_i + \frac{1}{\bar{v}_s} \sum_{i=1}^c f'_i (v_i - \bar{v}_s)^2 \\
&= \bar{v}_s + \frac{\sigma^2}{\bar{v}_s} \quad (\text{B. 13})
\end{aligned}$$

where

$$\sigma^2 = \sum_{i=1}^c f'_i (v_i - \bar{v}_s)^2 \quad (\text{B. 14})$$

## Appendix C

### More Evaluation Results for the Model proposed in Chapter 3



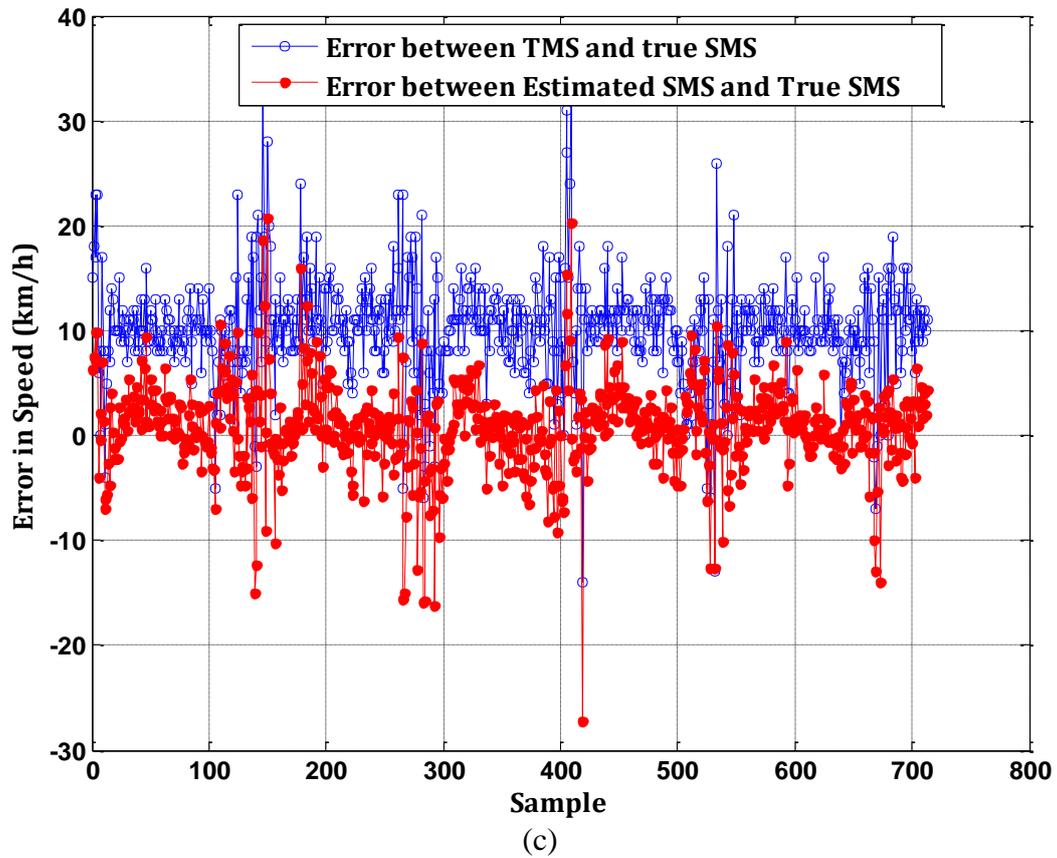
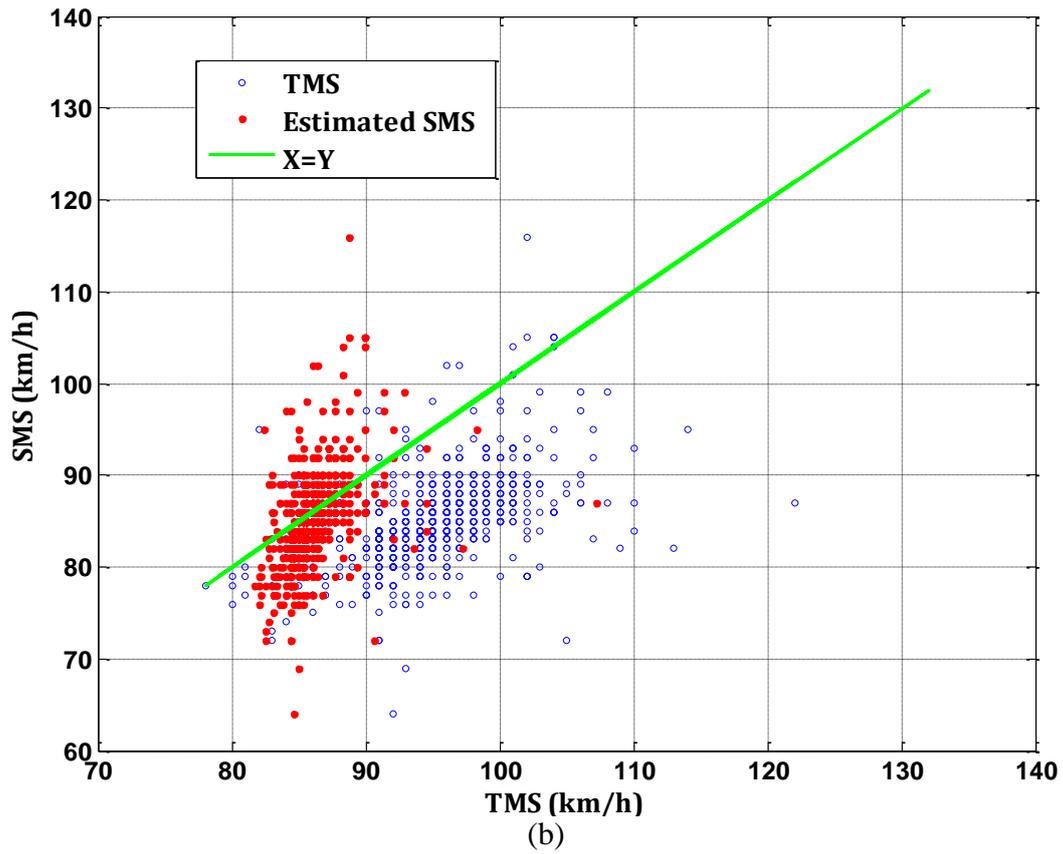
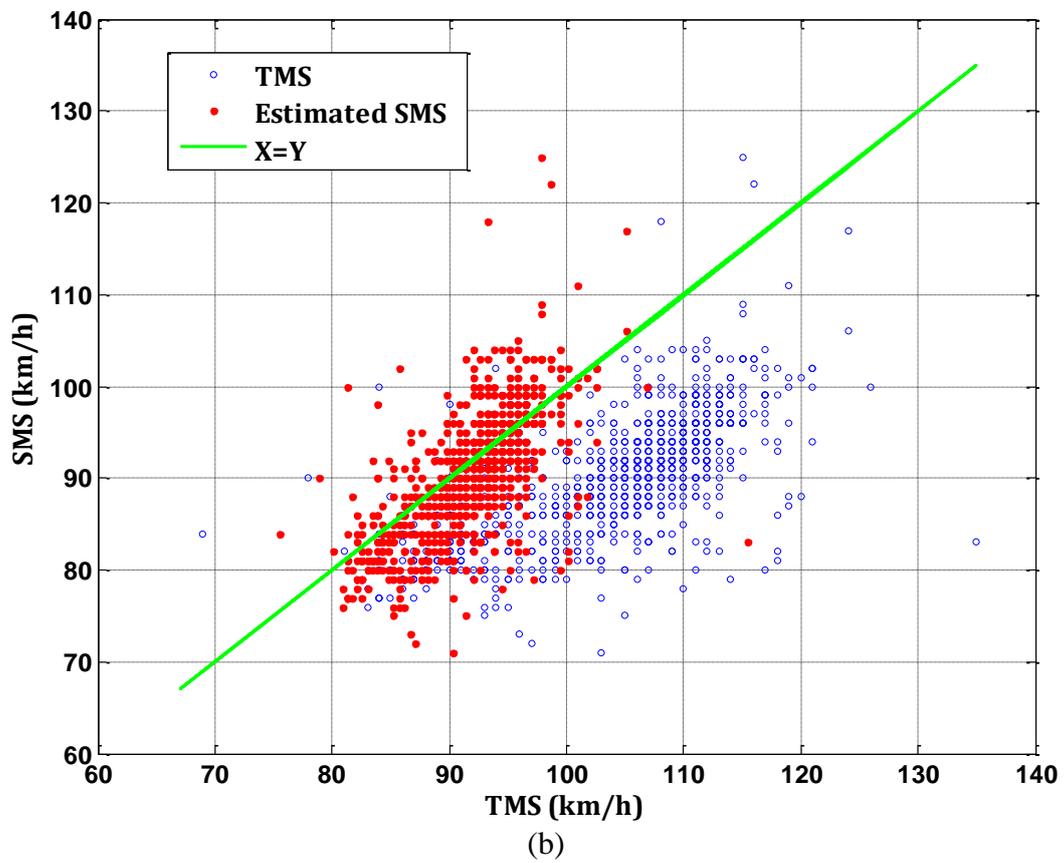
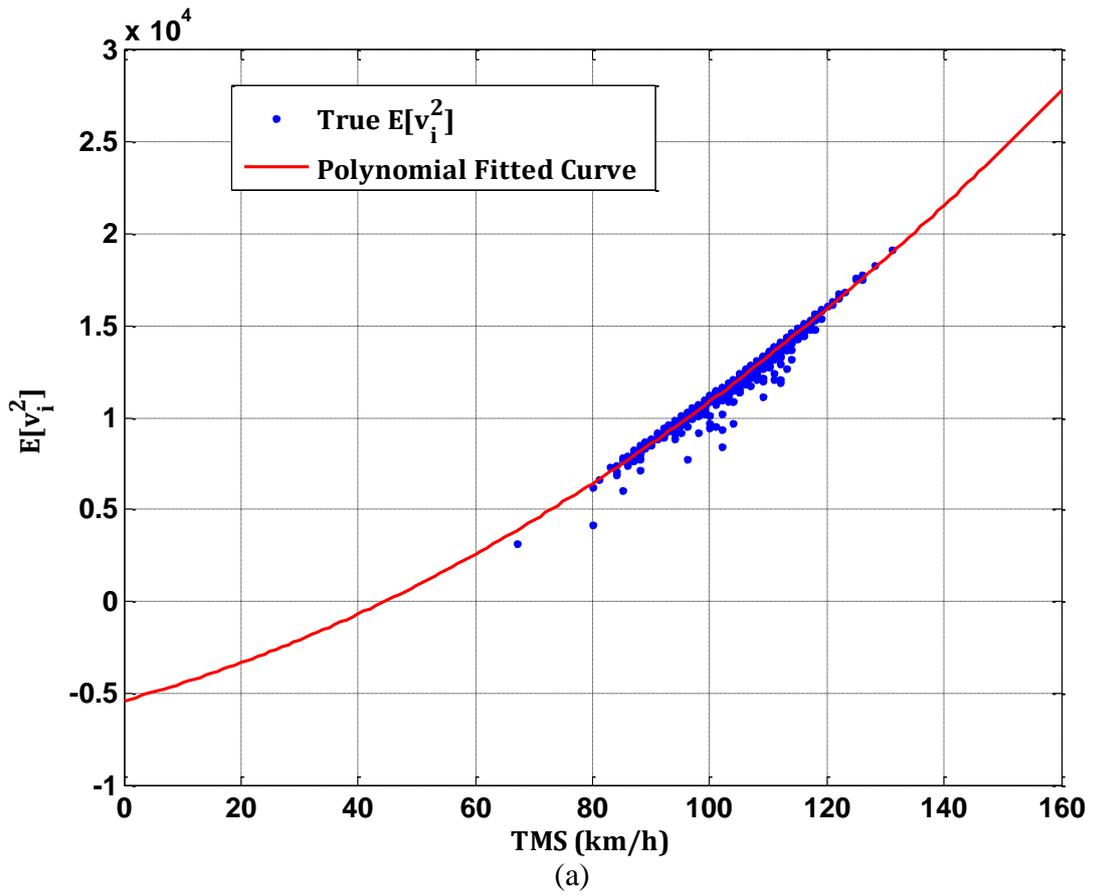


Figure C.1 Estimation Results for Link-2. (a) Result of polynomial fitting  $E[v_1^2]$  by TMS; (b) Scatter-plot of the estimation result; (c) Comparison between errors from only using TMS and estimated SMS



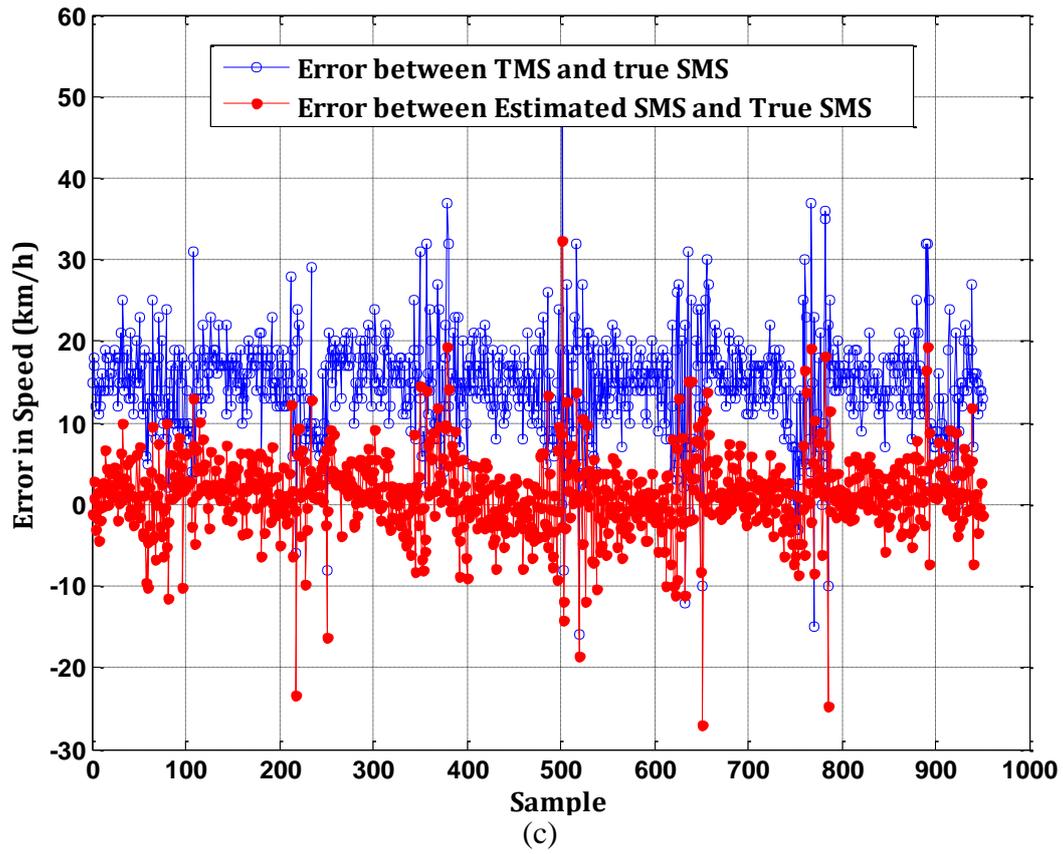
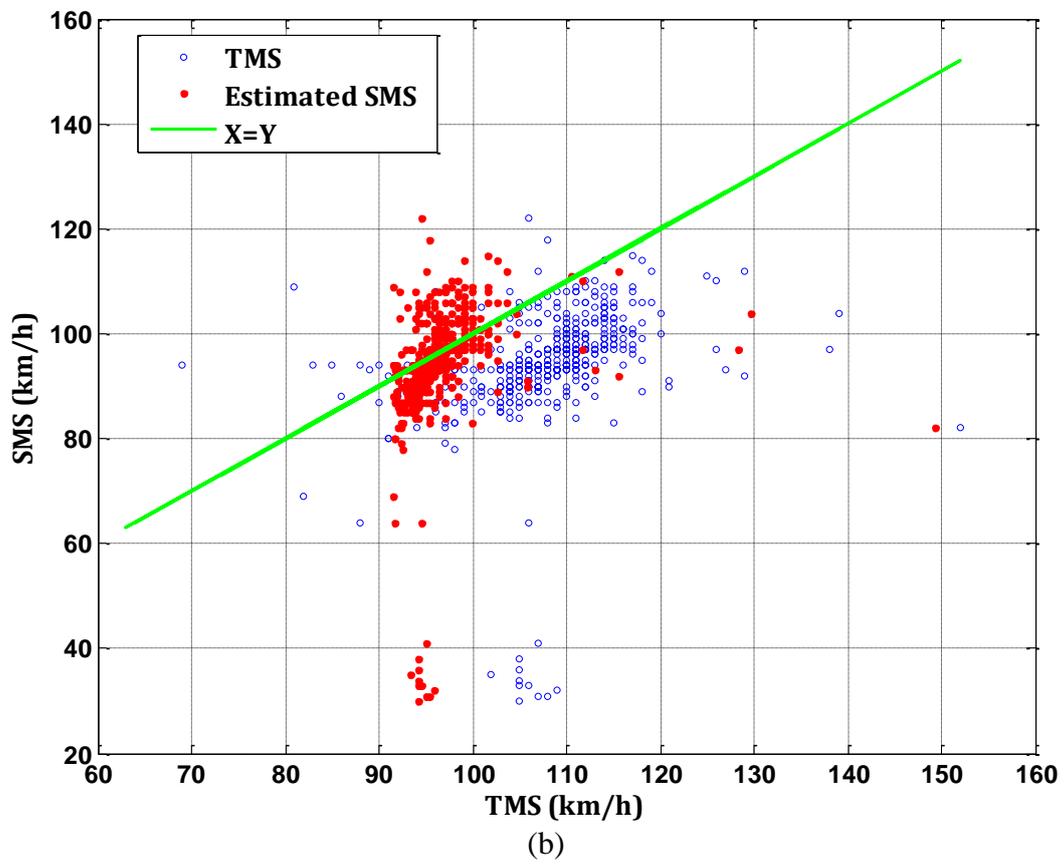
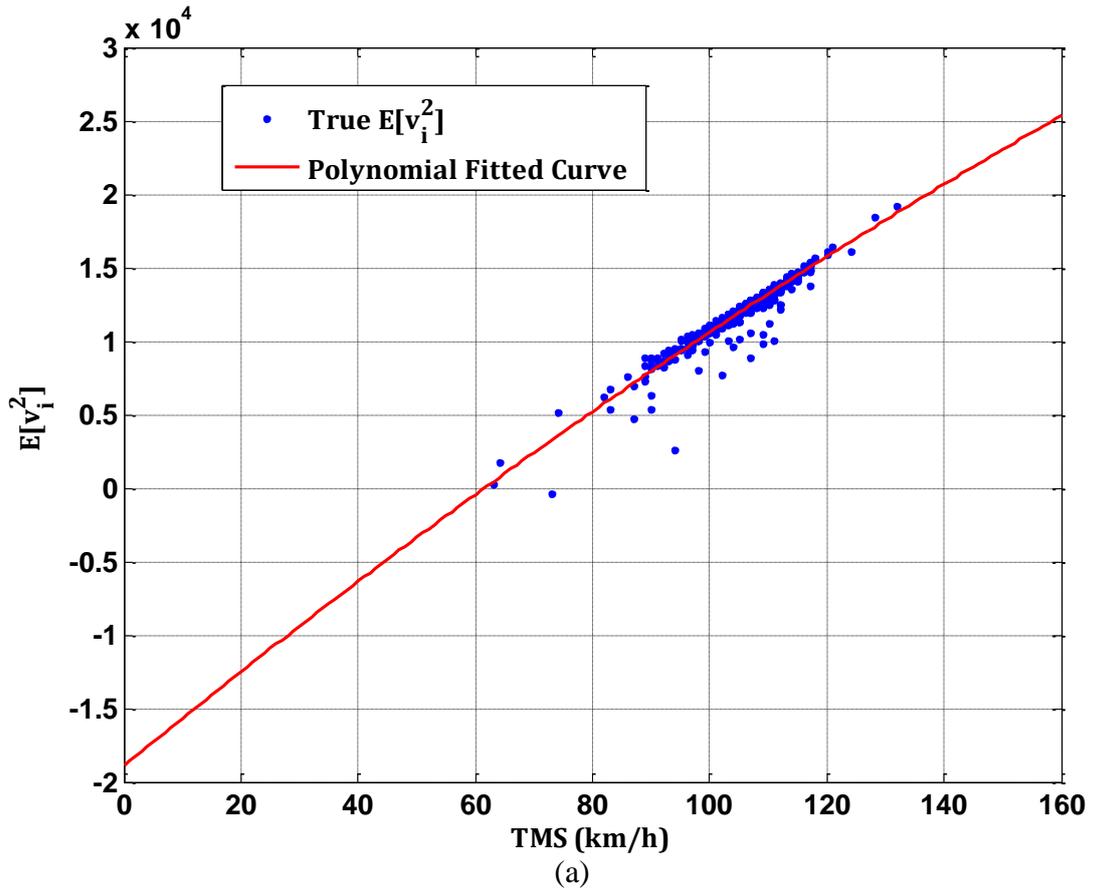


Figure C.2 Estimation Results for Link-3. (a) Result of polynomial fitting  $E[v_1^2]$  by TMS; (b) Scatter-plot of the estimation result; (c) Comparison between errors from only using TMS and estimated SMS



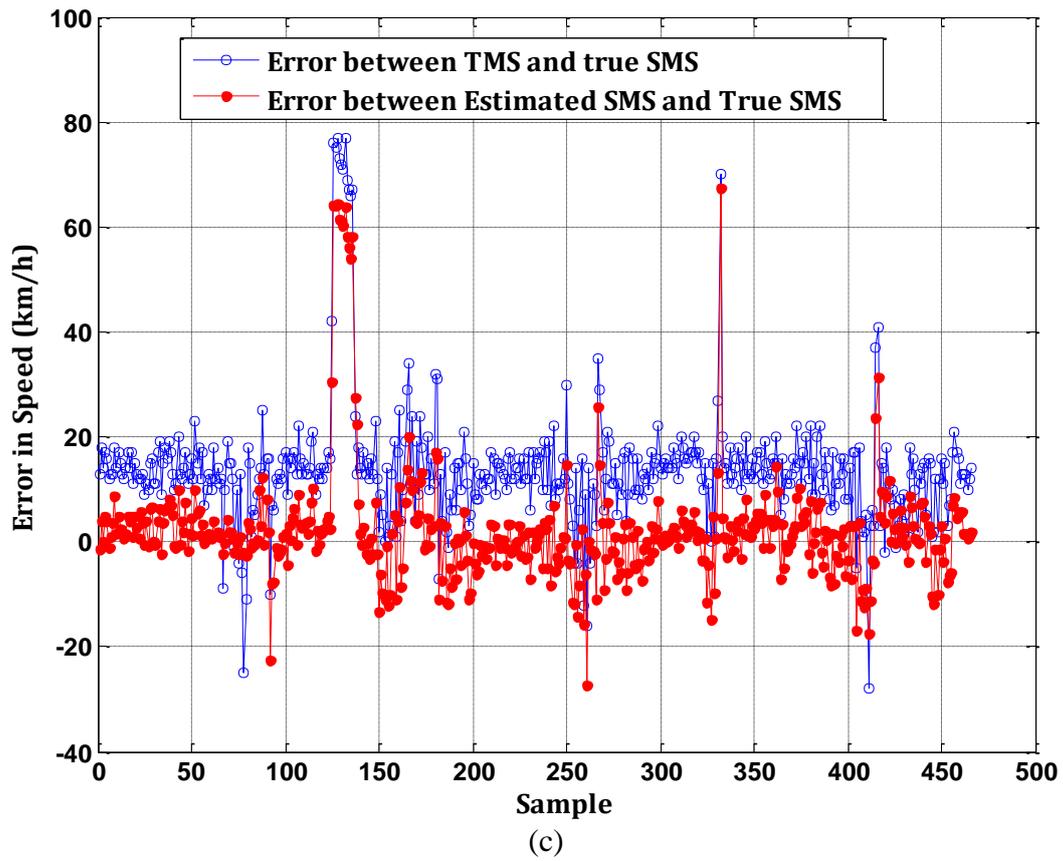
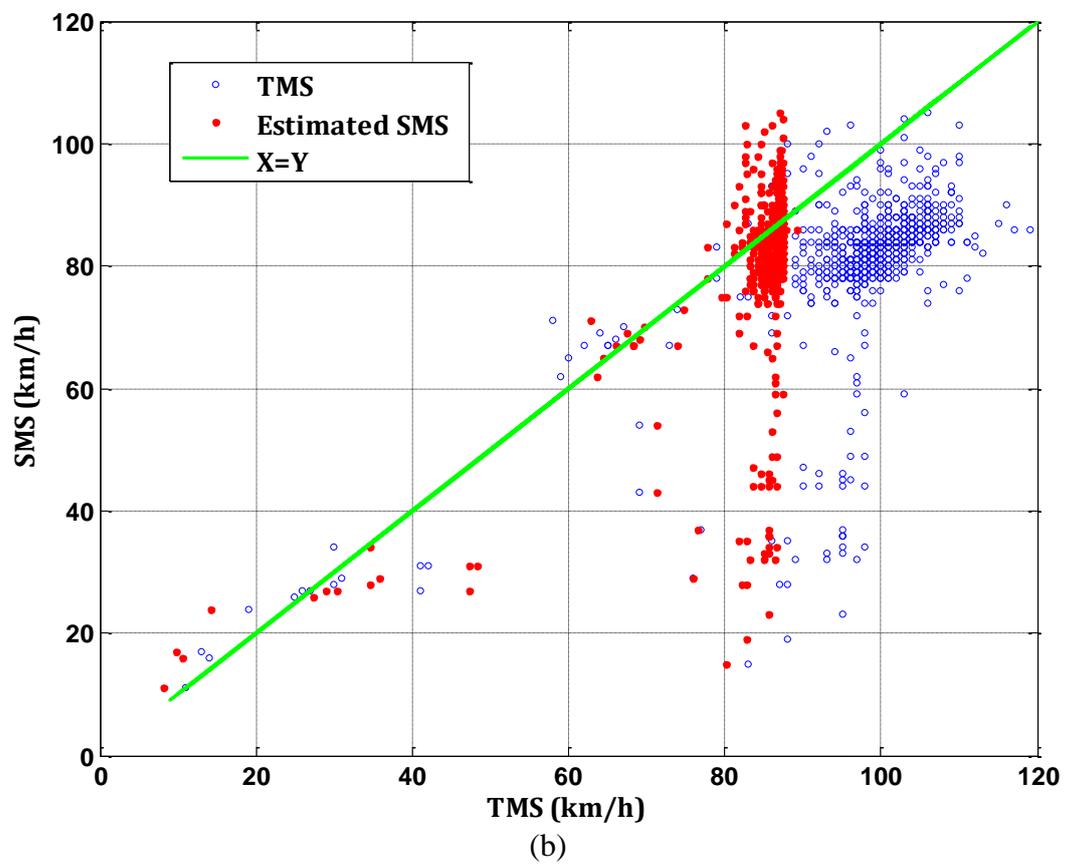
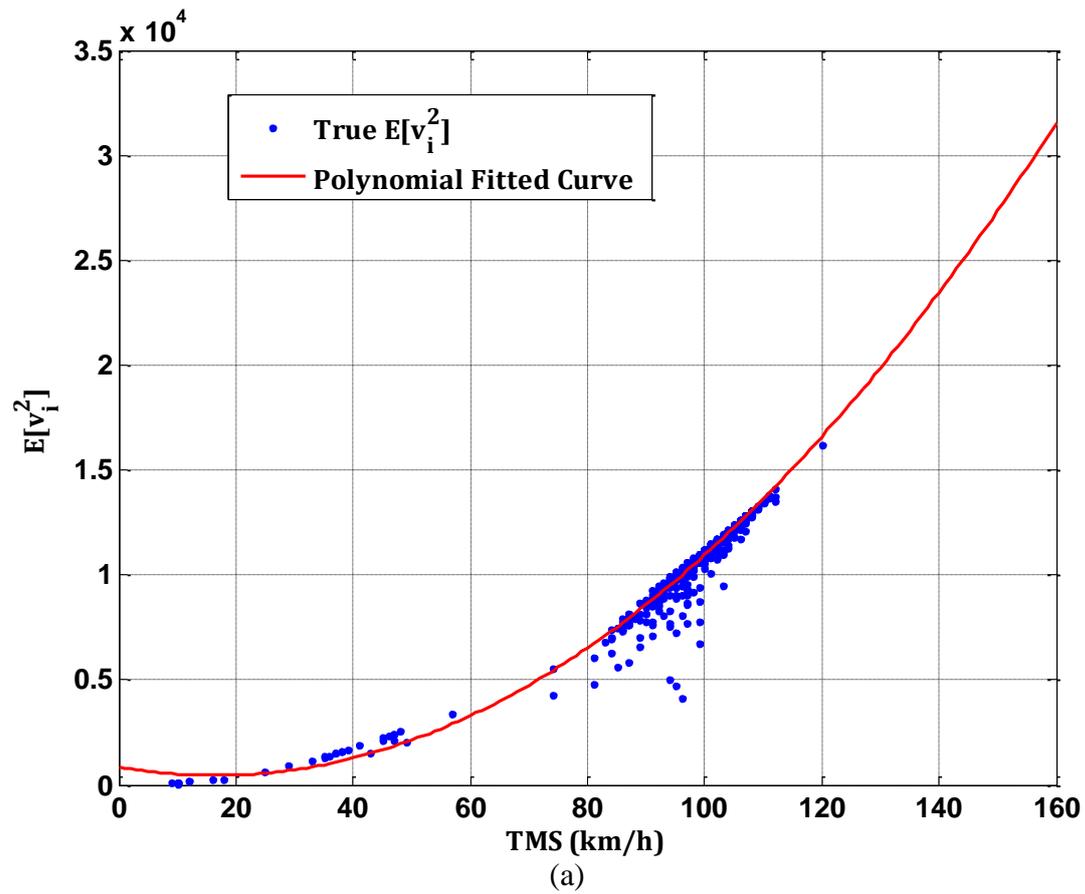


Figure C.3 Estimation Results for Link-4. (a) Result of polynomial fitting  $E[v_i^2]$  by TMS; (b) Scatter-plot of the estimation result; (c) Comparison between errors from only using TMS and estimated SMS



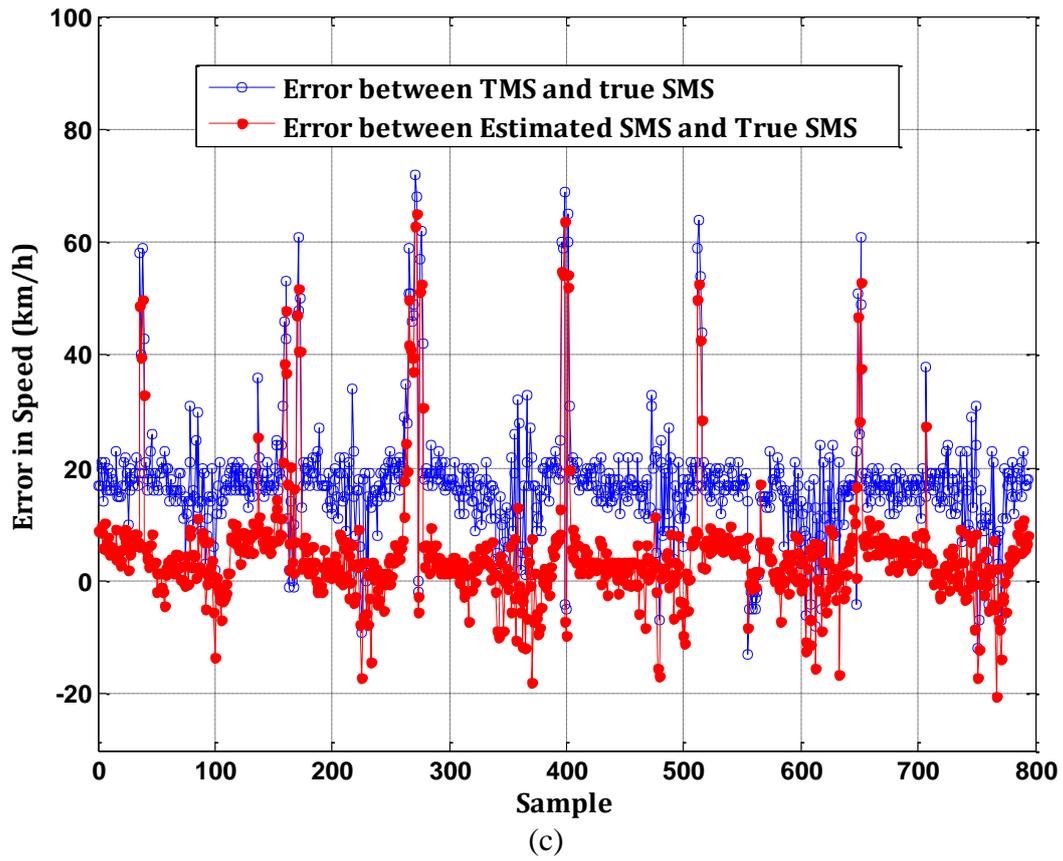
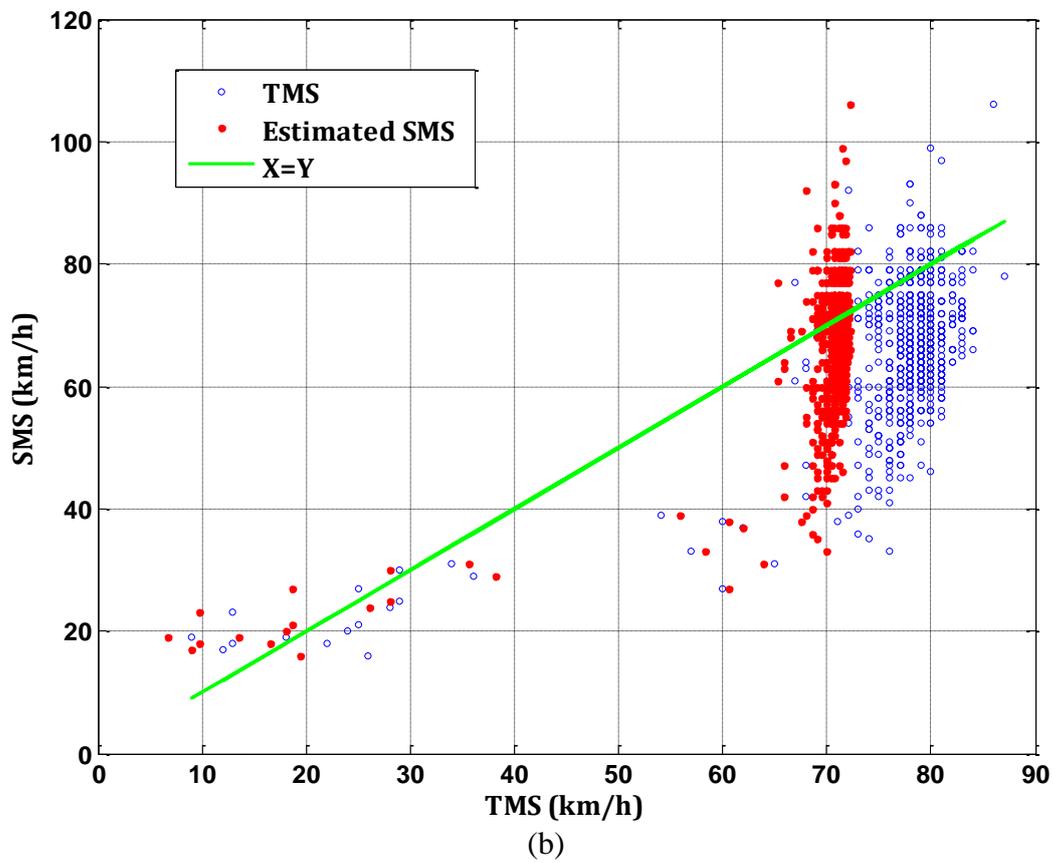
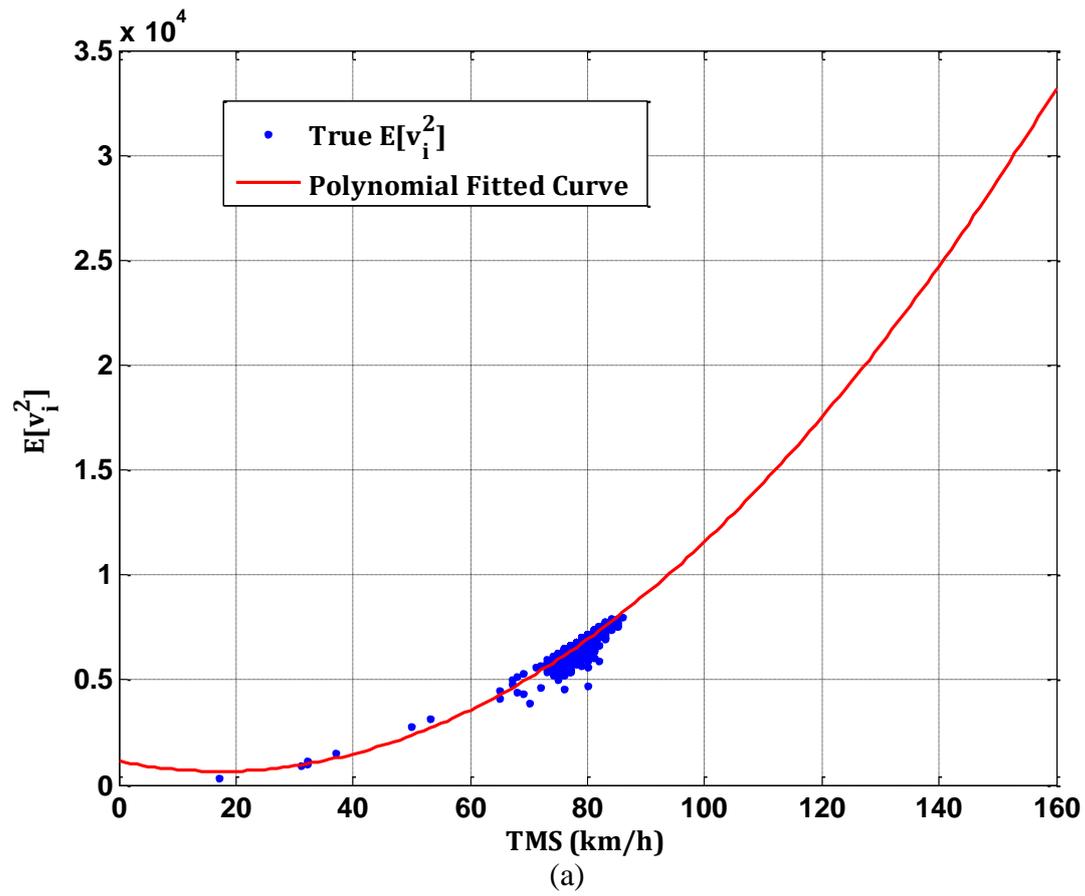


Figure C.4 Estimation Results for Link-5. (a) Result of polynomial fitting  $E[v_1^2]$  by TMS; (b) Scatter-plot of the estimation result; (c) Comparison between errors from only using TMS and estimated SMS



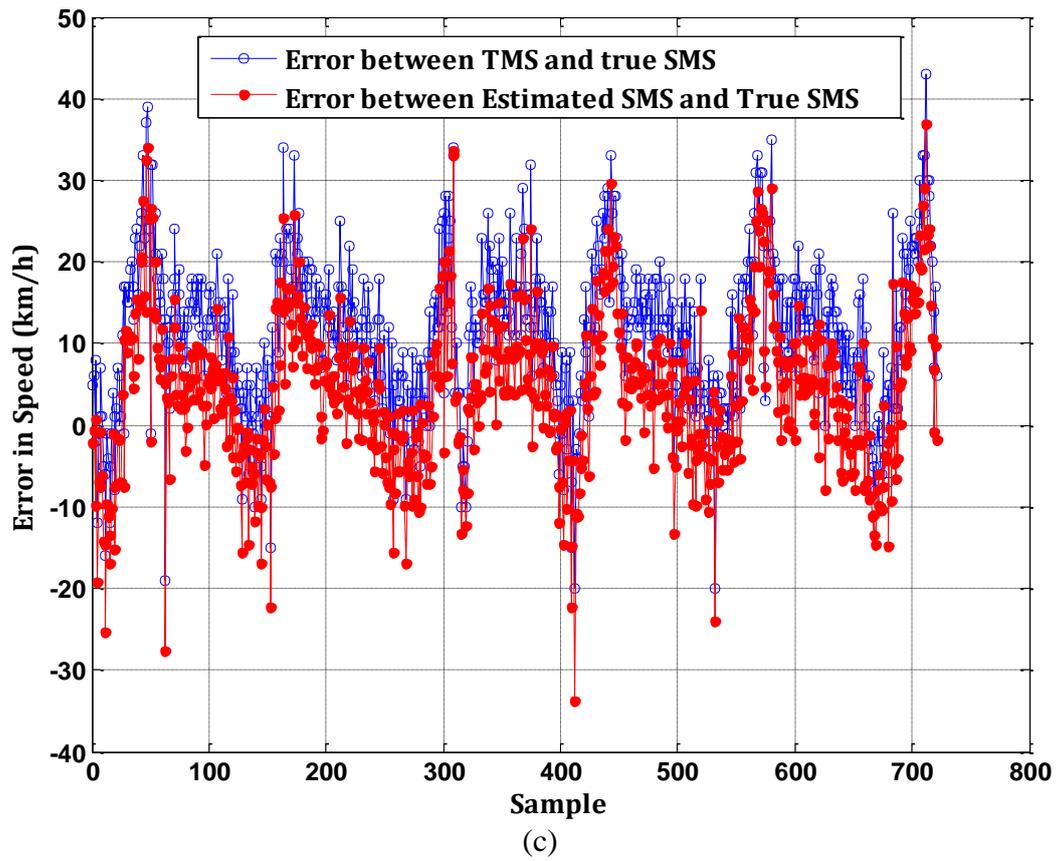


Figure C.5 Estimation Results for Link-6. (a) Result of polynomial fitting  $E[v_i^2]$  by TMS; (b) Scatter-plot of the estimation result; (c) Comparison between errors from only using TMS and estimated SMS

## Appendix D

# Hardware Specification of the GPS Logger

*The information is from AGL3080 GPS Photo Tracker User Manual V2.2*

## AMOD GPS Photo Tracker AGL3080

**We Add More Fun to Photo Sharing!**



© Copyright

AMOD Technology Co.,  
Ltd. 8F, 46, Lane 10, Jihu Road, Neihu  
Taipei, Taiwan, 114, R. O. C.  
E-mail: sales@amod.com.tw

## Hardware Specification

|                          |   |
|--------------------------|---|
| <b>GPS</b>               |   |
| Chipset                  | SiRF Star III high performance low power GPS receiver IC                    |
| Frequency                | L1 1,575.42 MHz   |
| Channels                 | 20 parallel tracking channels   |
| GPS Tracking Sensitivity | - 158 dBm   |
| <b>LED Indicators</b>    |   |
| Power On/Off             | Amber   |
| GPS Fix                  | Green   |
| Memory Full              | Red   |
| <b>General</b>           |   |
| Storage Capacity         | 128 Mbytes (1 Gbit NAND flash memory)                                       |
| Interface                | USB 2.0 full speed  |
| Battery                  | 3 AAA standard size batteries   |
| Operating Time           | > 15 hours continuous operation (3 AAA 900 mAh rechargeable battery @ 25°C) |
| Operating Temperature    | - 20 ~ 50 °C  |
| Storage Temperature      | - 30 ~ 80 °C  |
| Humidity                 | 95 % non-condensing   |
| Dimension                | 90 mm × 45 mm × 23 mm   |
| Weight                   | ~ 50 g (not including battery)  |

## Package Contents

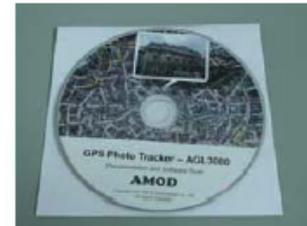
Each box contains the following items:



AGL3080 Hanging Strip & USB Cable



CD-ROM (Manual, Software Tools)



## Certifications

CE, FCC, VCCI certified, RoHS Compliant

**FCC Notices** This Device complies with Part 15 of the FCC rules; operation is subject to the following two conditions.

- .(1) This device may not cause harmful interference and,
- .(2) This device must accept any interference received.

## Trademarks and Copyrights

Windows is registered trademark of Microsoft Corporation.

Mac is registered trademark of Apple Inc.

JetPhoto is registered trademark and copyrighted by Atomix Technologies Corp.

GPSVisualizer is registered and copyrighted by Adam Schneider, Schneider@pobox.com

The AGL3080 and AMOD is registered trademark of AMOD Technologies Co., Ltd.



**AMOD Technology Co., Ltd. 8F, 46, Lane 10, Jihu Road, Neihu Taipei, Taiwan, 114, R.**

**O. C. E-mail: [sales@amod.com.tw](mailto:sales@amod.com.tw)**

## References

- Alger, M., Quartier, F. & Lang, T. (2003), 'Generating road traffic information from GSM signaling data', 10th World Congress on Intelligent Transport Systems, Madrid, Spain.
- Anderson, J. & Bell, M. (1997), 'Travel time estimation in urban road networks', IEEE Conference on Intelligent Transportation System, pp.924-929.
- Ashok, K. & Ben-Akiva, M. (1993), 'Dynamic O–D matrix estimation and prediction for real-time management systems', Proceedings of the 12th ISTTT, pp. 465-484.
- Avrekh, M., Han, J., Wilkinson, L. & Savage, S. (2011), 'Road sage website', <http://groups.sims.berkeley.edu/roadsage>, accessed on 19/09/2011.
- Bando, M. (1995), 'Dynamical model of traffic congestion and numerical simulation', *Physical Review E*, **51** pp. 1035-1042.
- Ben-Akiva, M. & Morikawa, T. (1987), 'Methods to combine different data sources and estimates OD matrices', Proceedings of the 10th ISTTT, Cambridge, MA, pp. 459-481.
- Berka, S., Tian, X. & Tarko, A. (1995) 'Data fusion algorithm for ADVANCE', Advance Working Paper 48.
- Bertini, R.L., Lasky, M., Monsere, C.M. (2005), 'Validating predicted rural corridor travel times from an automated license plate recognition system: Oregon's frontier project', Proceedings of the 2005 IEEE Conference on Intelligent Transportation Systems, pp.296-301.
- Bezdek, J.C., Coray, C., Gunderson R., & Watson, J. (1981), 'Detection and characterisation of cluster sub-structure I, linear structure: FCL', *Journal of Applied Mathematics*, **40**, pp.339-357.
- Bilmes, A. J. (1998), 'A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden markov models', Technical report, University of Berkeley.

- Branston, D. (1976), 'Models of single lane time headway distributions', *Transportation Science*, **10**, pp.125-148.
- Buckley, D.J. (1968), 'A semi-Poisson model of traffic flow', *Transportation Science*, **2**, pp.107-132
- Burns, J.A. & Kang, S. (1991), 'A control problem for Burgers equation with bounded input/output', *Nonlinear Dynamics*, **2**, pp. 235-262.
- Byun, S. C., Choi, D. B., Ahn, B. H. & Ko, H. (1999), 'Traffic incident detection using evidential reasoning-based data fusion', Proceedings of the 6th World Congress on ITS, Toronto, Canada.
- Chandler, R. E., Herman, R. & Montroll, E. W. (1958), 'Traffic dynamics: studies in car following', *Operations Research* **6**, pp. 165-184.
- Chen, M. & Chien, S.I.J. (2001), 'Dynamic freeway travel time prediction using probe vehicle data: link-based vs. path-based', 80th Annual Meeting of the Transportation Research Board, CD-ROM.
- Chen, S., Wang, W. & Qu, G. (2004), 'Combining wavelet transform and Markov model to forecast traffic volume', Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai.
- Cherrett, T., Bell, H. A. & McDonald, M. (2001), 'Estimating Vehicle Speed using single inductive loop detectors', Proceeding of the Institution of Civil Engineers Transport 147, pp.23-32.
- Chien, S.I.J., Liu, X. & Ozbay, K. (2003), 'Predicting travel time for the south Jersey real time motorist information system', Transportation Research Record: Journal of the Transportation Research Board (1855), pp. 32-40.
- Choi, K. Y. & Chung, Y. (2001), 'Travel time estimation algorithm using GPS probe and loop detector data fusion', 80th Annual Meeting, Transportation Research Board CD-ROM.
- Chu, L., Oh, J.-S. & Recker, W. (2005), 'Adaptive Kalman filter based freeway travel time estimation', 84th Annual Meeting, Transportation Research Board, CD-ROM.
- Clark, S., Grant-Muller, S. & Chen, H. (2002), 'Cleaning of matched license plate data', *Transportation Research Record*, **1804**, pp.1-7.
- Cohen, S. (2003), 'Fusion of incident detection algorithms', Proceedings of the Multisource Data Fusion in Traffic Engineering Workshop, 87, pp. 109-117.

- Cremer, M. & Papageorgiou, M. (1981), 'Parameter identification for a traffic flow model', *Automatica* **17**, pp. 837-843.
- Cremer, M. & Schrieber, S. (1996), 'Monitoring traffic load profiles with heterogeneous data sources configurations', in: Lyons, J.-B. Lesort (Ed.), *Proceedings of the 16th International Symposium on Transportation and Traffic theory (ISTTT)*, Pergamon, Elsevier, pp. 41-54.
- Daganzo, C. (1995), 'Requiem for second-order fluid approximation to traffic flow', *Transportation Research Part B*, **29**, pp. 277-286.
- Daganzo, C. F. (1997), 'Fundamentals of transportation and traffic operations', **Pergamon. ISBN: 0-08-042785-5.**
- Dailey, D. J. (1993), 'Travel Time Estimation using cross correlation techniques', *Transportation Research B* 27(2), pp. 97-107.
- Dailey, D. J., Harn, P. & Linet, P. J. (1996), 'ITS data fusion', Research project T9903, Task 9: ATIS/ATMS Regional IVHS Demonstration, Report# WA-RD 410.1, Washington State, WADOT.
- Dasarathy, B. V. (1997), 'Sensor fusion, potential exploitation: innovative architectures and illustrative applications', *Proceedings of IEEE* **85** pp.24-38.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) 'Maximum-likelihood from incomplete data via the EM algorithm', *Journal of Royal Statistical Society*, 39, pp.1-38.
- Doucet, A. & Johansen, A. M (2008), 'A tutorial on particle filtering and smoothing: fifteen years later', Technical report, Department of Statistics, University of British Columbia.
- Dunn, J. C. (1973), 'A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters', *Journal of Cybernetics*, 3, pp.32-57.
- Eddie, L.C. (1961), 'Car-following and steady-state theory for non-congested traffic', *Operations Research*, **9**, pp. 66-75.
- El Faouzi, N.-E. (1997), 'Heterogeneous data source fusion for impedance indicators', *IFAC Symposium on Transportation Systems*, Chania, Greece, Vol. 3, pp.1375-1380.
- El Faouzi, N.-E. (1999), 'Combining predictive schemes in short-term traffic forecasting', in: A. Ceder (Ed.), *Proceedings of the 17th International Symposium on Transportation and Traffic theory (ISTTT)*, Jerusalem, Pergamon, Elsevier, pp. 471-487.

- EI Faouzi, N.-E. (2006), 'Bayesian and evidential approaches for traffic data fusion: methodological issues and case study', 85th Annual Meeting, Transportation Research Board, CD-ROM.
- EI Faouzi, N.-E., Klein, L., & Mouzon, O. D. (2009), 'Improving travel time estimates from loop and toll collection data with Dempster-Shafer data fusion', 88th Annual Meeting, Transportation Research Board, CD-ROM.
- EI Faouzi, N.-E., Leung, H. & Kurian, A. (2011), 'Data fusion in intelligent transportation systems: progress and challenges – a survey', *Information Fusion*, **12**, pp. 4–10.
- El-Sheimy, N., Chiang, K. & Noureldin, A. (2006), 'The utilization of artificial neural networks for multisensory system integration in navigation and positioning instruments', *IEEE Transactions on Instrumentation and Measurement* **55**, pp.1606-1615.
- Ferman, A. M., Blumenfeld, E. D. & Dai, X. W. (2007), 'An analytical evaluation of a real-time traffic information system using probe vehicles', *Journal of Intelligent Transportation Systems*, **9**, pp.23–24.
- Forbes, T.W., H.J. Zagorski, E.L. Holshouser, & W.A. Deterline (1958), 'Measurement of driver reactions to tunnel conditions', *Proceedings of Highway Research Board*, **37**, pp. 345-357.
- Friedrich, B. & Minciardi, R. (2003), 'Data fusion techniques for adaptive traffic signal control', IFAC Symposium on Transportation Systems, Tokyo, Japan, pp. 53–55.
- Gault, H. E. & Taylor, I. G. (1981), 'The use of the output from vehicle detectors to assess delay in computer-controlled area traffic control systems.', University of Newcastle, Transport Operations Research Group, Research Report no.37.
- Garber, N. & Hoel, L. (2001), 'Traffic and highway engineering', 3rd ed, **CL-Engineering. ISBN: 0534387438.**
- Gazis, D. & Liu, C. (2003), 'Kalman filtering estimation of traffic counts for two network links in tandem', *Transportation Research B* (37), pp. 737-745.
- Ghosh, D. S. & Adeli, H. (2003), 'Wavelet-clustering-neural network model for freeway incident detection', *Computer-Aided Civil and Infrastructure Engineering* **18**(5), pp.325-338.
- Granger, W. J. (1989), 'Combining forecasts: twenty years later', *Journal of Forecasting* **8**, pp. 167-173.
- Greenberg, H. (1959), 'An analysis of traffic flow', *Operations Research*, **7**, pp. 79-85.

- Greenshields, B.D. (1935), 'A study in highway capacity', *Highway Research Board*, **14**, pp. 458-474.
- Grewal, M. S. & Andrews, A. P. (2001), 'Kalman filtering: theory and practice using MATLAB', 2nd Edition, John Wiley & Sons, Inc., ISBN: 0-471-39254-5.
- Hackett, J. K. & Shah, M. (1990), 'Multi-sensor fusion: a perspective', Proceedings of 1990 IEEE International Conference on Robotics and Automation, pp.1324-1330.
- Haight, F. A. (1963), 'Mathematical theories of traffic flow', **Academic Press, New York, ISBN: 0-10-012315250X.**
- Hall, D. L. (1992), 'Mathematical techniques in multisensor data fusion', **Artech House, Norwood, MA, ISBN: 1580533353.**
- Hall, F. L. (1996), 'Traffic stream characteristics in traffic flow theory', **Washington, D.C.: US Federal Highway Administration.**
- Han, J., Krishnan, R., Polak, J. & Barria, J. (2009), 'On the estimation of space mean speed from inductive loop detector data', Presented at the 41th Annual Meeting of the Universities Transport Studies Group, London, UK.
- Haykin, S. (1999), 'Neural networks: a comprehensive foundation, 2<sup>nd</sup> Edition', **Prentice Hall, ISBN: 0139083855.**
- Haykin, S. (2001), 'Kalman filtering and neural networks', Wiley & Sons, ISBN: 0471369985.
- Hegyí, A., Girimonte, D., Babuska, R. & De Schutter, B. (2006), 'A comparison of filter configurations for freeway traffic state estimation', *Proceedings of the IEEE Intelligent Transportation Systems Conference*, Toronto, Canada, pp.17-20.
- Hellinga, B. & Fu, L. (2002), 'Reducing bias in probe based arterial link travel time estimates', *Transportation Research C* **10**, pp.257-273.
- Herrera, J. C. & Bayen, A. M. (2008), 'Traffic flow reconstruction using mobile sensors and loop detector data', 87th Annual Meeting, Transportation Research Board, CD-ROM.
- Highways Agency (2009), 'National traffic control centre (NTCC) road network', **Technical Report, Publication code PR237/06, Highways Agency Publications Group.** Available from [http://www.tih.org.uk/images/2/24/D1\\_NTCC\\_Road\\_Network.pdf](http://www.tih.org.uk/images/2/24/D1_NTCC_Road_Network.pdf) [accessed 29 May 2011].
- Highways Agency (2011a), 'Media briefing: traffic congestion', Highways Agency website, <http://www.highways.gov.uk/news/8579.aspx>, accessed on 15/09/2011.

- Highways Agency (2011b), 'Traffic England website', <http://www.traffic-england.com>, accessed on 19/09/2011.
- Hofmann-Wellenhof, B. (2001), 'Global positioning system : theory and practice, 5<sup>th</sup> edition', **Springer-Verlag, ISBN: 3211835342.**
- Hoogendoorn, S.P. & Bovy, P.H.L. (1998), 'A new estimation technique for vehicle-type specific headway distributions', *Transportation Research Record* **1646**, pp. 18-28.
- Hoogendoorn, S.P. & Bovy, P.H.L. (2000a), 'State-of-the-art of vehicular traffic flow modelling', *Journal of Systems and Control Engineering – Proceedings of the Institution of Mechanical Engineers, Part I*, vol.215, pp. 283-303.
- Hoogendoorn, S.P. & Bovy, P.H.L. (2000b), 'Modelling multiple user-class traffic flow', *Transportation Research B* **34**, pp. 123-146.
- Houston Star (2011), 'Houston star website', <http://traffic.houstontranstar.org/>, accessed on 19/09/2011.
- Huisken, G. & van Berkum, E. C. (2003), 'A comparative analysis of short-range travel time prediction methods', 82nd Annual Meeting of the Transportation Research Board, CD-ROM.
- The International GNSS Service (IGS) (2009), 'IGS Products Table', <http://igsceb.jpl.nasa.gov/components/prods.html>, accessed on 25/02/2012.
- Ivan, J. N., Schofer, J. L., Coppelman, F. S. & Massone, L. L. E. (1995), 'Real-time data fusion for arterial street incident detection using neural networks', *Transportation Research Record* 1497 pp. 27-35.
- Jiang, G., Wang, J., Zhang, X. and Gang, L. (2003), 'The study on the application of Fuzzy Clustering Analysis in the dynamic identification of road traffic state', The IEEE 6th International Conference on Intelligent Transportation Systems, Shanghai, China.
- Julier, S.J., Uhlmann, J.K. & Durrant-White, H. (1995), 'A new approach for filtering nonlinear systems', In *Proceedings of the American Control Conference*, pp. 1628–1632.
- Julier, S.J., Uhlmann, J.K. & Durrant-White, H. (2000), 'A new method for nonlinear transformation of means and covariances in filters and estimators', *IEEE Transactions on Automatic Control*, **45**, pp.477–482.
- Kachroo, P. & Ozbay, K. (1999), 'Feedback control theory for dynamic traffic assignment', **Springer-Verlag, ISBN: 1852330597.**

- Kalman, R.E. (1960), 'A new approach to linear filtering and prediction problems', *Transaction of the ASME's Journal of Basic Engineering*, pp. 35-45.
- Kerner, B.S. & Konhauser, P. (1993), 'Cluster effect in initially homogeneous traffic flow', *Physical Review E*, **48**, pp. 2335-2338.
- Klar, A. & Wegener (1999), 'A hierarchy of models for multilane vehicular traffic I & II: modelling', *SIAM Journal of Applied Mathematics*, **59**, pp.1002-1011.
- Klein, L. (2001), 'Sensor technologies and data requirements for ITS', **Artech House ITS Library, ISBN: 0-58053-077-X**.
- Klein, L.A., Mills, M.K. & Gibson, D.R. (2006), 'Traffic Detector Handbook', third edition volume I, Federal Highway Administration, USA.
- Krishnan, R. K. (2008a), 'Travel time estimation and forecasting on urban roads', PhD thesis, Centre for Transport Studies, Imperial College London.
- Krishnan, R. K. (2008b), 'A deductive travel time estimation method for signalised corridors using single loop detector data, Presented at the 40th Annual Meeting of the Universities Transport Studies Group, Portsmouth, UK.
- Lao, Y., Yun, M., Tang, S., Wang, C., Yang, X. & Chu, H. (2007), 'Evaluation method for accuracy of road traffic state information', Proceedings of the First International Conference on Transportation Engineering 2007 (ICTE 2007). Chengdu, China.
- Leutzbach, W. (1987), 'Introduction to the theory of traffic flow', **Springer-Verlag, New York, ISBN: 0387171134**.
- Li, M., Zou, Z., Bu, F. & Zhang, W. (2008), 'Application of vehicle-infrastructure integration (VII) data on real-time arterial performance measurements', 87th Annual Meeting of the Transportation Research Board, CD-ROM.
- Li, W. & Leung, H. (2003), 'Constrained unscented Kalman filter based fusion of GPS/INS/digital map for vehicle localization', IEEE International Conference on Intelligent Transportation Systems, Shanghai, China, pp. 1362-1367.
- Lighthill, M. J. & G. B. Whitham (1955), 'On kinematic waves: a theory of traffic flow on long crowded roads', *Proceedings of the Royal Society A229* **1178**, pp. 317-145.
- Lindveld, C., Logie, M. & Polak, J. W. (2007), 'A Bayesian framework for the integration of multiple transport information sources', 86th Annual Meeting, Transportation Research Board, CD-ROM.

- Liu, H., van Zuylen, H., van Lint, H. & Salomons, M. (2006), 'Urban arterial travel time prediction with state-space neural networks and Kalman filters', 85th Annual Meeting of the Transportation Research Board, CD-ROM.
- Lucas, D. E., Mirchandani, P. B. & Verma, N. (2004), 'Online travel time estimation without vehicle identification', *Transportation Research Record* 1867, pp.193-201.
- Lundgren, J.T., Peterson, A. & Tengroth, S. (2003), 'Methods for pre-adjusting time dependent origin destination matrices', *Proceedings of 10th World Congress and Exhibition on Intelligent Transportation Systems and Services*, Madrid, Spain, pp. 16-20.
- McLachlan, G. & Krishnan, T. (2008), 'The EM algorithm and extensions', **Wiley, New York, ISBN: 0471201707.**
- May, A.D. (1990), 'Traffic flow fundamentals', **Prentice Hall, New Jersey, ISBN: 0139260722.**
- Maybeck, P.S. (1979), 'Stochastic models, estimation, and control, Volume 1', **Academic Press, Inc., ISBN: 0-12-480701-1.**
- Maybeck, P.S. (1990), 'The Kalman filter: an introduction to concepts', **Springer-Verlag, ISBN: 0-387-97240-4.**
- Mnyika, J. & Durrant-Whyte, H. (1994), 'Data fusion and sensor management: a decentralized information theoretic approach', **Ellis Horwood, London, ISBN: 0133031322.**
- Montroll, E.W. (1961), 'Acceleration noise and clustering tendency of vehicular traffic', in *Theory of Traffic Flow*, R. Herman, Ed, **Elsevier, Amsterdam** pp.147-157.
- Mortimer, K. (2002), 'Enhancing technology', *Traffic Engineering & Control* 43(5), pp.197-197.
- Mueck, J. (2002), 'Estimation methods for the state of traffic at traffic signals using detectors near the stop-line', *Traffic Engineering and Control*.
- Musha, T. & Higuchi, H. (1978), 'Traffic current fluctuation and the burgers', *Japanese Journal of Applied Physics*, **17**, pp. 811-816.
- Nagel, K. (1996), 'Particle hopping models and traffic flow theory', *Physical Review E*, **53**, pp. 4655-4672.
- Nagel, K. (1998), 'From particle hopping models to traffic flow theory', *Transportation Research Record*, **1644**, pp. 1-9.

- Nanthawichit, C., Nakatsuji, T. & Suzuki, H. (2003), 'Application of probe vehicle data for real-time Traffic state estimation and short-term travel time prediction on a freeway', 82th Annual Meeting, Transportation Research Board, CD-ROM.
- Narayanan, R., Udayakumar, R., Kumar, K. & Subbaraj, L. (2003), 'Quantification of congestion using fuzzy logic and network analysis using GIS', Proceedings of Map India Conference 2003, New Delhi, India.
- National Traffic Control Centre (NTCC) Road Network (2009), Technical Report, Publication code PR237/06, Highways Agency Publications Group, Available from [http://www.tih.org.uk/images/2/24/D1\\_NTCC\\_Road\\_Network.pdf](http://www.tih.org.uk/images/2/24/D1_NTCC_Road_Network.pdf), accessed on 15/11/2009.
- Noland, R. B. & Polak, J. W. (2002), 'Travel time variability: a review of theoretical and empirical issues', *Transport Reviews* 22(1), pp.39-54.
- Noland, R. B. & Small, K. A. (1995), 'Travel time uncertainty, departure time choice, and the cost of morning commutes', *Transportation Research Record* 1493, pp.150-158.
- Ochieng, W. & Sauer, K. (2002), 'Urban road transport navigation: performance of the global positioning system after selective availability', *Transportation Research C* 10, pp.171-187.
- Ochieng, W., Polak, J. W., Noland, R. B., Park, J. Y., Zhao, L., Briggs, D., Gulliver, J., Crookell, A., Evans, R., Walker, M. & Randolph, W. (2003), 'Integration of GPS and dead reckoning for real-time vehicle performance and emissions monitoring', *GPS Solutions* 6(4), pp. 229–241.
- Palacharla, P. & Nelson, P. (1999), 'Application of fuzzy logic and neural networks for dynamic travel time estimation', *International transactions in operational research* 6(1), pp.145-160.
- Palen, J. (1997), 'The need for surveillance in intelligent transportation systems', *Intellimotion* 6(1), pp.1-3.
- Papageorgiou, M., Blosseville, J.M., & Haj-Salem, H. (1989), 'Macroscopic Modelling of traffic flow on the Boulevard Peripherique in Paris', *Transportation Research B*, **23**, pp. 29-47.
- Papageorgiou, M., Blosseville, J.M., & Haj-Salem, H. (1990), 'Modeling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris—Part I: modeling', *Transportation Research A*, **24**, pp. 345-359.

- Papageorgiou, M. & Kotsialos, A. (2002), 'Freeway ramp metering: an overview', *IEEE Transactions on Intelligent Transportation Systems*, **3**, pp. 271-281.
- Park, D. & Rilett, L. (1999), 'Forecasting freeway link travel times with a multilayer feed-forward neural network', *Computer Aided Civil and Infrastructure Engineering*, **14**, pp. 357-367.
- Payne, H.J. (1971), 'Models of freeway traffic and control', Simulation Councils Proceedings Series: Mathematical Models of Public Systems, vol.1, pp.51-61.
- Petty, K. F., Bickel, P., Ostland, M., Rice, J., Schoenberg, F. & Jiang, J. (1998), 'Accurate estimation of travel times from single loop detectors', *Transportation Research A* 32(1), pp.1-17.
- Prigogine, I. & Herman, R. (1971), 'Kinetic theory of vehicular traffic', **American Elsevier, New York, ISBN: 0444000828.**
- Pueboobpaphan, R., Nakatsuji, T. & Suzuki, H. (2007), 'Unscented Kalman filter-based real-time traffic state estimation' 86th Annual Meeting of the Transportation Research Board, CD-ROM.
- Qiao, F., Wang, X. & Yu, L. (2003), 'Optimizing aggregation level for ITS data based on wavelet decomposition', 82th Annual Meeting of Transportation Research Board, CD-ROM.
- Qiu, Z. & Ran, B. (2007), 'Kalman filtering applied to network-based cellular probe traffic monitoring', 86th Annual Meeting of the Transportation Research Board, CD-ROM.
- Quddus, A.M. (2006), 'High integrity map matching algorithms for advanced transport telematics applications', PhD thesis, Centre for Transport Studies, Imperial College London.
- Rakha H. & Zhang W. (2005), 'Estimating traffic stream space-mean speed and reliability from dual and single loop detectors', *Transportation Research Record: Journal of the Transportation Research Board*, **1925**, pp. 38-47.
- Redner, R. A. & Walker, H. F. (1984), 'Mixture densities, maximum likelihood and the EM algorithm', *SIAM Review*, **26**, pp. 195-239.
- Richards, P.I. (1956), 'Shockwaves on the Highway', *Operation Research*, vol.4, pp. 42-51.
- Robinson, S. (2005), 'The development and application of an urban link travel time model using data derived from inductive loop detectors', PhD thesis, Centre for Transport Studies, Imperial College London.

- Robinson, S. & Polak, J. (2006), 'Overtaking Rule approach for the cleaning of matched license plate data', *Journal of Transportation Engineering*, **132**, pp. 609-617.
- Roess, R. P., McShane, W. R. & Prassas, E. S. (1998), 'Traffic engineering', **ISBN: 0-13-461336-8**.
- Rosen, J. & Gothard, L. Q. (2009), 'Encyclopedia of Physical Science', **Infobase Publishing, ISBN 0-816-07011-3**.
- Rouphail, N.M., Targo, A., Nelson, P.C. & Palacharla P. (1993), Travel time data fusion in ADVANCE - a preliminary design concept, ADVANCE Working Paper Series, No. 21, University of Illinois at Chicago, Chicago, IL USA.
- Schmidt, S.F. (1981), 'The Kalman filter: recognition and development for aerospace applications', *Journal of Guidance and Control*, **4**, pp.4.
- Seeber, G. (2003), 'Satellite geodesy: foundations, methods, and applications, 2<sup>nd</sup> edition', **Walter de Gruyter & Co, ISBN: 3110175495**.
- Walter de Gruyter. ISBN: 3-11-012753-9.
- Smulders, S. (1990), 'Control of freeway traffic flow by variable speed signs', *Transportation Research B*, **24**, pp. 111-132.
- Sorenson, H. W. (1970), 'Least-squares estimation: from Gauss to Kalman', *IEEE Spectrum*, vol. 7, pp. 63-68.
- Sroka, R. (2004), 'Data fusion methods based on fuzzy measures in vehicle classification process', *Instrumentation and Measurement Technology Conference*, Como, Italy, pp.18-20.
- Stratonovich, R.L. (1959), 'On the theory of optimal non-linear filtering of random functions', *Theory of Probability and its Applications* 4, pp. 223-225.
- Stratonovich, R.L. (1960), 'Application of the Markov Processes Theory to Optimal Filtering', *Radio Engineering and Electronic Physics* 5 (11), pp. 1-19.
- Sumner, R. (1991), 'Data fusion in Pathfinder and TravTek', *Proceedings of 1991 Vehicle Navigation and Information Systems Conference*, Vol.1, Dearborn, MI, pp. 71-75.
- Sun, S., Zhang, S. & Yu, G. (2006), 'A Bayesian network approach to traffic flow forecasting', *IEEE Transactions on Intelligent Transportation System* 7(1), pp.124-132.
- Swerling, P. (1958), 'A proposed stagewise differential correction procedure for satellite tracking and prediction', *RAND*, Santa Monica, California, USA.

- Tarko, A. & Rouphail, N. (1993), 'Travel time fusion in ADVANCE', Pacific Rim TransTech Conference: A Ride into the Future, Seattle, USA, pp. 25–28.
- Taylor, M. A. P., Woolley, J. E. & Zito, R. (2000), 'Integration of the global positioning system and geographical information systems for traffic congestion studies', *Transportation Research C* **8**, pp. 257-285.
- TfL (2003a), 'Impacts monitoring - first annual report', Transport for London, <http://www.tfl.gov.uk/assets/downloads/Impacts-monitoring-report1.pdf>, accessed on 15/09/2011.
- TfL (2003b), 'Congestion Charging Fact sheets: Camera Enforcement', <http://www.transportforlondon.gov.uk>.
- TfL (2010), 'Network performance indices (NPI) for traffic directorate', Technical Report NPI for TD V0.02, London Streets, Traffic Directorate, Transport for London.
- Thomas, N. E. (1998), 'Multi-State and Multi-Sensor Incident Detection Systems for Arterial Streets'. *Transportation Research Part C* **6** (5/6), pp. 337-357.
- Transport Scotland (2007), 'Traffic Scotland website', <http://www.nadics.org.uk/>, accessed on 19/09/2011.
- Trees, H.L.V. (2001), 'Detection, estimation and modulation theory', Wiley-Interscience, ISBN: 0471095176.
- Turner, S. M. & Holdener, D. J. (1995), 'Probe vehicle sample sizes for real-time information: the Houston experience, Proceedings of Vehicle Navigation and Information Systems Conference, Seattle, WA , USA, pp. 3–10.
- Turner, S. M., Eisele, W. L., Benz, R. J. & Holdener, D. J. (1998), 'Travel time data collection handbook', Technical Report FHWA-PL-98-035, Federal Highway Administration, USA.
- Underwood, R.T. (1961), 'Speed, volume and density relationships', *Quality and Theory of Traffic Flow, Yale Bureau of Highway Traffic*, pp.141-88.
- van der Zijpp, N. (1997), 'Dynamic OD-Matrix estimation from Traffic Counts and automated vehicle identification data', *Transportation Research Record* **1607**, pp.87–94.
- van Hinsbergen, C. P. I J. & van Lint, J.W.C. (2007), 'Bayesian combination of travel time prediction models', 86th Annual Meeting, Transportation Research Board, CD-ROM.

- van Lint, J.W.C. & Hoogendoorn, S.P. (2007), 'The technical and economic benefits of data fusion for real-time monitoring of freeway traffic', Proceedings of the 11th World Congress on Transport Research, Berkeley, CA, USA.
- Venkatanarayana, R., Smith, B. L. and Demetsky, M. J. (2005), 'Traffic pattern identification using wavelet transforms', 84th Annual Meeting of Transportation Research Board, CD-ROM.
- Wan, E. A. & Merwe, R. V. (2000), 'The unscented Kalman filter for nonlinear estimation', In IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control, pp.153–158.
- Wang, P.& Papageorgiou, M. (2005), 'Real-time freeway traffic state estimation based on extended Kalman filter: a general approach', *Transportation Research Part B*, **39**, pp. 141–167.
- Wardrop, J. G. (1952), 'Some theoretical aspects of road traffic research', *Proceedings of the Institute of Civil Engineers* **1(2)**, pp. 325-378.
- Wei, M. & Schwarz, K. P. (1990), 'Testing a decentralized filter for GPS/INS integration', Proceedings of IEEE Position Location and Navigation Symposium, Las Vegas, NV, pp. 429-435.
- Weisstein, E.W. (2005), 'K-means clustering algorithm', **MathWorld – A Wolfram Web Resource**, Available from <http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html> [accessed 05 March 2009].
- Welsh Assembly Government (2011), 'Traffic Wales website', <http://www.traffic-wales.com/traffic>, accessed on 19/09/2011.
- Westman, M., Litjens, R. & Linnartz, J. P. (1996), 'Integration of probe vehicle and induction loop data - estimation of travel times and automatic incident detection', California PATH Research Report, UCB-ITS-PRR-96-13, Institute of Transportation Studies, University of California, Berkeley, CA.
- White, F. E. (1987), 'Data fusion lexicon, Joint Directors of Laboratories', Technical Panel for C3, Data Fusion Sub-Panel, Naval Ocean Systems Centre, San Diego.
- Whitham, G. B. (1974), 'Linear and nonlinear waves', **John Wiley, New York, ISBN: 0471940909**.
- Wiggins, A. E. (1999), 'Helsinki journey time monitoring system', **IEE Seminar on CCTV and Road Surveillance (Ref. No. 1999/126)**.

- Willner, D., Chang, C.B. & Dunn, K.P. (1976), 'Kalman filter algorithm for a multi sensor system', In Proceedings of IEEE conference on decision and control, Clearwater, Florida, pp. 570-574.
- Wormley, S. J. (2010), 'GPS errors & estimating your receiver's accuracy', available online at [http://edu-observatory.org/gps/gps\\_accuracy.html](http://edu-observatory.org/gps/gps_accuracy.html), accessed on 22/09/2011. accessed on 19/09/2011
- Xie, C., Cheu, R. L. & Lee, D. H. (2004), 'Improving arterial link travel time estimation by data fusion', 83th Annual Meeting, Transportation Research Board, CD-ROM.
- Ye, Z., Zhang, Y. & Middleton, D.R. (2006), 'An unscented Kalman filter method for speed estimation using single loop detector data', 85th Annual Meeting of the Transportation Research Board, CD-ROM.
- Youngbin, Y. & Cayford, R. (2000), 'Investigation of vehicles as probes using global positioning system and cellular phone tracking', PATH Working Paper.
- Zhang, H.M. (1999), 'A mathematical theory of traffic hysteresis', *Transportation Research B*, **33**, p1-23.
- Zhang, M. & He, J. C. (1998), 'Estimating Arterial Travel Time Using Loop Data', Public Policy Center, University of Iowa.
- Zhao, Y.L. (1997), 'Vehicle location and navigation systems', **Artech House, ISBN: 0890068615**.
- Zhao, Y. (2000), 'Mobile Phone Location Determination and Its Impact on Intelligent Transportation Systems', *IEEE Transactions on Intelligent Transportation Systems* **1**, pp.55-64.
- Zheng, P. & McDonald, M. (2008), 'Estimation of travel time using fuzzy clustering method', *IET Intelligent Transport Systems*, **3**, pp.77-86.
- Zhong, M., Lingras, P. & Sharma, S. (2004), 'Estimation of missing traffic counts using factor, genetic, neural, and regression techniques', *Transportation Research Part C* **12**, pp.139-166
- Zurada, J. M. (1992), 'Introduction to artificial neural systems', **West Publishing Company, Boston, ISBN: 0314933913**.