Statistical modelling and analysis of traffic: a dynamic approach

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/9421](https://dspace.lboro.ac.uk/2134/9421)

Publisher: © Karandeep Singh

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
STATISTICAL MODELLING AND ANALYSIS OF TRAFFIC:
A DYNAMIC APPROACH

By

Karandeep Singh

A Doctoral Thesis
Submitted in partial fulfilment of the requirements
for the award of
Doctor of Philosophy of Loughborough University

March 2012

© by Karandeep Singh 2012
ABSTRACT

In both developed and emerging-economies, major cities continue to experience increasing traffic congestion. To address this issue, complex Traffic Management Systems (TMS) are employed in recent years to help manage traffic. These systems fuse traffic-surveillance-related information from a variety of sensors deployed across traffic networks. A TMS requires real-time information to make effective control decisions and to deliver trustworthy information to users, such as travel time, congestion level, etc. There are three fundamental inputs required by TMS, namely, traffic volume, vehicular speed, and traffic density. Using conventional traffic loop detectors one can directly measure flow and velocity. However, traffic density is more difficult to measure. The situation becomes more difficult for multi-lane motorways due to drivers’ lane-change behaviour.

This research investigates statistical modelling and analysis of traffic flow. It contributes to the literature of transportation and traffic management and research in several aspects. First, it takes into account lane-changes in traffic modelling through incorporating a Markov chain model to describe the drivers’ lane-change behaviour. Secondly, the lane change probabilities between two adjacent lanes are not assumed to be fixed but rather they depend on the current traffic condition. A discrete choice model is used to capture drivers’ lane choice behaviour. The drivers’ choice probabilities are modelled by several traffic-condition related attributes such as vehicle time headway, traffic density and speed. This results in a highly nonlinear state equation for traffic density.

To address the issue of high nonlinearity of the state space model, the EKF and UKF is used to estimate the traffic density recursively. In addition, a new transformation approach has been proposed to transform the observation equation from a nonlinear form to a linear one so that the
potential approximation in the EKF & UKF can be avoided.

Numerical studies have been conducted to investigate the performance of the developed method. The proposed method outperformed the existing methods for traffic density estimation in simulation studies. Furthermore, it is shown that the computational cost for updating the estimate of traffic densities for a multi-lane motorway is kept at a minimum so that online applications are feasible in practice. Consequently, the traffic densities can be monitored and the relevant information can be fed into the traffic management system of interest.

Keywords: Density Estimation, Discrete Choice Model, Kalman Filters, Lane-Change Behaviour, Logit Model, Markov Chain, Statistical Analysis and Modelling, Transport and Traffic Management.
ACKNOWLEDGEMENTS

This thesis was completed, first and foremost with the support of my family especially great-grandfather Baba Thakur Singh and grandfather Baba Ajeet Singh, my spouse who has always been a source of inspiration to me and helped me in every possible way during the journey of my Ph.D. I owe a unique debt of gratitude and appreciation to Prof Baibing Li for his support and guidance whilst giving me encouragement throughout the years at Loughborough University. It is true that without his support and motivation, this thesis would never have been done. My thanks also go to Prof Xu Ling, my external examiner from Manchester Business School, Prof Jiyin Liu, my Independent Assessor, and to Dr Grammatoula Papaioannou, of my Independent Panel Member, at the School of Business and Economics, Loughborough University.

In addition, I extend my special thanks to my friend Mandeep Singh Senior Researcher at the BAE systems, for undertaking to proof read my thesis.

A special thanks is due to the Intelligent Transport Systems, University of Washington, which kindly supported and provided the required data for this research through their website.

Last but not least, I am indebted to my wife, son and parents and in-laws for their devotion, perseverance, and moral support throughout the research and writing of this thesis.
DEDICATION

This thesis is dedicated to my father Dhan Dhan Sri Guru Gobind Singh Ji and mother Dhan Dhan Sri Mata Sahib Kaur Ji, who have continually given me the strength and encouragement to rise to new challenges in life especially through the challenge of completing this PhD.
CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

Karandeep Singh

6 - Mar - 2012
LIST OF ABBREVIATIONS

Active Traffic Management (ATM)
Advanced Incident Detection (AID)
Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks (AIMSUN2)
Advanced Traffic Management System (ATMS)
Artificial Neural Networks (ANN)
Auto Regressive Moving Averages (ARMA)
Average Density Ratio (ADR)
Average Headway (AHD)
Average Lane Speed (AS)
Automatic Vehicle Location (AVL)
Central Processing Unit (CPU)
Closed Circuit Television (CCTV)
Coordinated Ramp Metering (CRM)
Department for Traffic (DfT)
Discretionary Lane Change (DLC)
Driver Vehicle Operator (DVO)
Extended Kalman Filter (EKF)
Gaussian Random Variable (GRV)
Generalized Likelihood Ratio (GLR)
Independent and Identically Distributed (iid)
Inductive Loop Detectors (ILDs)
List of Abbreviations

Intelligent Transportation Systems (ITS)
Jump Markov Linear Model (JMLM)
Kalman Filter (KF)
Kalman Filtering Technique (KFT)
Linear-Quadratic Regulator (LQR)
Linear-Quadratic-Gaussian (LQG)
Mandatory Lane Change (MLC)
Markov Chain Monte Carlo (MCMC)
Microscopic Traffic Simulator - (MITSIM)
Minimum Mean Square Error (MMSE)
National Instruments Manufacturers Association (NEMA)
Origin Destination (OD)
Particle Filter (PF)
Root Mean Square Error (RMSE)
Root Mean Square Error (RMSE)
Sequential Monte Carlo methods (SMC)
Simulation of Intelligent Transport Systems (SITRAS)
Single-Input and Single-Output System (SISO)
Traffic Data Acquisition and Distribution (TDAD)
Traffic Management Systems (TMS)
United Kingdom (UK)
United States of America (USA)
Unscented Kalman Filter (UKF)
Unscented Transformation (UT)
List of Abbreviations

Variable Speed Limit (VSL)
<table>
<thead>
<tr>
<th>Symbol(s)</th>
<th>Terms</th>
<th>Unit(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>occ</td>
<td>occupancy lies within $[0, 1]$</td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>traffic density</td>
<td>veh/km</td>
</tr>
<tr>
<td>$vll$</td>
<td>valid vehicle length</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>vehicular speed</td>
<td>km/h</td>
</tr>
<tr>
<td></td>
<td>free flow speed (maximum permitted speed in the section)</td>
<td>km/h</td>
</tr>
<tr>
<td></td>
<td>maximum permitted traffic density in the section</td>
<td>veh/km</td>
</tr>
<tr>
<td></td>
<td>value of speed at which maximum flow or capacity is reached</td>
<td>km/h</td>
</tr>
<tr>
<td></td>
<td>state vector at $k$th ($t$th) time step</td>
<td></td>
</tr>
<tr>
<td></td>
<td>updated state</td>
<td></td>
</tr>
<tr>
<td></td>
<td>predicted state</td>
<td></td>
</tr>
<tr>
<td></td>
<td>measurement at $k$th ($t$th) time step</td>
<td></td>
</tr>
<tr>
<td></td>
<td>control input</td>
<td></td>
</tr>
<tr>
<td></td>
<td>process noise</td>
<td></td>
</tr>
<tr>
<td></td>
<td>state transition matrix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>control-input matrix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>observation matrix which maps the true state space into the observed space</td>
<td></td>
</tr>
<tr>
<td></td>
<td>measurement noise</td>
<td></td>
</tr>
<tr>
<td>$Q$ and $R$</td>
<td>process noise covariance and the measurement noise covariance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>free flow speed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>density corresponding to the maximum flow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>length of segment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>density corresponding to the maximum flow in the $i$th section</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>vehicle count error (from state equation)</td>
<td>km/h</td>
</tr>
<tr>
<td>$\tau$</td>
<td>vehicle speed error (from observation error)</td>
<td>km/h</td>
</tr>
<tr>
<td></td>
<td>innovation sequence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>normalised innovation sequence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean of normalised innovation sequence</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>average number of vehicles entering/leaving the section</td>
<td></td>
</tr>
<tr>
<td></td>
<td>transition probability that a vehicle moves from state $j$ to state $k$</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF IMPORTANT CONCEPTS

**Speed** - Speed in traffic flow is defined as the distance covered per unit time.

**Time Mean Speed** - If speed is measured by keeping time as reference it is called time mean speed. Time mean speed is measured by taking a reference area on the roadway over a fixed period of time. In practice, it is measured by the use of loop detectors.

**Space Mean Speed** - If speed is measured by space reference it is called space mean speed. Space mean speed is the speed measured by taking the whole roadway segment into account. Consecutive pictures or video of a roadway segment track the speed of individual vehicles, and then the average speed is calculated.

**Traffic Count/Density** – It is defined as the number of vehicles per unit area of the roadway.

**Traffic Flow** – It is the number of vehicles passing a reference point per unit of time, and is measured in vehicles per hour.

**Headway** – It is a measurement of the distance/time between vehicles in a transit system. The precise definition varies depending on the application, but it is most commonly measured as the distance from the tip of one vehicle to the tip of the next one behind it, expressed as the time headway or space headway it will take for the trailing vehicle to cover that distance.

**Traffic Congestion** – It is a condition on road networks that occurs as use increases, and is characterized by slower speeds, longer trip times, and increased vehicular queueing.

**Traffic Collision** - It is also known as a traffic accident, motor vehicle collision, motor vehicle accident, car accident, automobile accident, Road
Traffic Collision (RTC) or car crash, occurs when a vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstruction, such as a tree or utility pole. Traffic collisions may result in injury, death and property damage.

**Travel Delay** - The Highway Capacity Manual defines delay as "The additional travel time experienced by a driver, passenger, or pedestrian." Delay is thus the difference between an "ideal" travel time and the actual travel time. Since the definition of delay depends on a hypothetical "ideal travel time," delay is not always directly measurable in the field.

**Variable Message Sign** - A variable message sign, often abbreviated VMS, CMS, or DMS, and in the UK known as a matrix sign is an electronic traffic sign used on roadways to give travelers information about special events. Such signs warn of traffic congestion, accidents, incidents, roadwork zones, or speed limits on a specific highway segment.

**Magnetometer** – It is an instrument used to detect the presence of a metallic object or to measure the intensity of a magnetic field.

**Vehicle Count Error** ($\sigma$) – It is the error in the number of vehicles (entering/leaving) gathered from the ILDs, incorporated in the state equation of the state space model.

**Vehicle Speed Error** ($\tau$) - It is the error in the speed of vehicles (entering/leaving) gathered from the ILDs, incorporated in the measurement equation of the state space model.

**Transition probability** – It is defined as the probability with which a vehicle moves from state (lane) $j$ to state (lane) $k$. 
List of Important Concepts

**Signalised Links** – These are termed as those roadway sections/segments which have traffic signals in them. A traffic signal may or may not join two or more sections, linking them together and hence termed as signalised links.

**Probe Vehicle** – The probe vehicle are unique in that they are typically intelligent transportation system (ITS) applications designed primarily for collecting data in real-time. Their primary applications are for a specific purpose other than travel time data collection, such as real-time traffic operations monitoring, incident detection, and route guidance applications.

**Probe Data** – Data collected by the probe vehicles is termed as probe data.

**Lane Blocking Incidents** – The incidents which cause the blocking of one or more lanes of a motorway are called as lane blocking incidents.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>CERTIFICATE OF ORIGINALITY</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF NOTATIONS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF IMPORTANT CONCEPTS</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xviii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xxii</td>
</tr>
<tr>
<td>CHAPTER I - INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Statement of Research</td>
<td>4</td>
</tr>
<tr>
<td>1.2 Research Objectives</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Research Methodology</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 Literature Review</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2 Study Design and Data Collection</td>
<td>8</td>
</tr>
<tr>
<td>1.3.3 Preliminary Analysis of the Data</td>
<td>8</td>
</tr>
<tr>
<td>1.3.4 Developing the Traffic Density State Space Model</td>
<td>9</td>
</tr>
<tr>
<td>1.3.5 Estimation of the Traffic Density</td>
<td>9</td>
</tr>
<tr>
<td>1.3.6 Computational Analysis</td>
<td>10</td>
</tr>
<tr>
<td>1.3.7 Statistical Analysis and Comparison of Results</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Contribution of the Research</td>
<td>11</td>
</tr>
<tr>
<td>1.5 Organisation of the Thesis</td>
<td>12</td>
</tr>
<tr>
<td>CHAPTER II - LITERATURE REVIEW</td>
<td>15</td>
</tr>
<tr>
<td>A. TRAFFIC STUDIES</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Sensor Technologies</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Inductive Loop Detector (ILD)</td>
<td>22</td>
</tr>
<tr>
<td>2.2.1 Introduction</td>
<td>22</td>
</tr>
<tr>
<td>2.2.2 How They Work</td>
<td>24</td>
</tr>
<tr>
<td>2.2.3 Advantages of ILDs</td>
<td>25</td>
</tr>
<tr>
<td>2.2.4 Disadvantages of ILDs</td>
<td>25</td>
</tr>
<tr>
<td>2.2.5 Application of ILDs for Estimating Traffic Variables</td>
<td>28</td>
</tr>
<tr>
<td>2.3 Traffic Density Estimation</td>
<td>29</td>
</tr>
</tbody>
</table>
Table of Contents

2.3.1 Introduction ........................................................................................................................................... 29
2.3.2 Different Ways to Obtain Traffic Densities ......................................................................................... 30
2.3.4 Traffic Densities From Estimation Methods ......................................................................................... 30
2.4 Lane Change Manoeuvre .......................................................................................................................... 36
2.4.1 History of Lane Change Theory - Gipps Model (1986) ...................................................................... 37
2.4.2 Limitations of Gipps Model .................................................................................................................. 37
2.4.3 Different Types of Lane Changes .......................................................................................................... 38
2.4.4 Gap Acceptance Theory ...................................................................................................................... 39
2.4.5 Different types of Lane-Change Models ............................................................................................... 41
2.5 Traffic Density and Lane-Change Behaviour ............................................................................................. 46
2.6 Speed-Density Relationship .................................................................................................................... 49
2.6.1 History .................................................................................................................................................. 49
2.6.2 Different Approaches for defining the Speed-Density Relationship ................................................. 49
2.6.3 The Properties of Speed-Density Relationships ................................................................................. 52
2.6.4 Our Choice of Speed–Density Model ................................................................................................. 54
B. KALMAN FILTERING TECHNIQUE .............................................................................................................. 54
2.7 Introduction .............................................................................................................................................. 54
2.7.1 General (Discrete) Kalman Filter ........................................................................................................ 55
2.7.2 Kalman Filter – A State Space System ............................................................................................... 57
2.7.3 Assumptions Behind the Formulation of The KF ............................................................................. 60
2.8 Extended Kalman Filter .......................................................................................................................... 61
2.8.1 Definition ............................................................................................................................................. 61
2.8.2 EKF – A State Space System ............................................................................................................. 61
2.9 Unscented Kalman Filter ........................................................................................................................ 63
2.9.1 Definition ............................................................................................................................................. 63
2.9.2 The Unscented Transformation ......................................................................................................... 64
2.10 Kalman Filter and Traffic Management .................................................................................................. 68
2.10.1 Applications of General KF in Traffic Management ....................................................................... 69
2.10.2 Applications of EKF in Traffic Management ...................................................................................... 70
2.10.3 Applications of UKF in Traffic Management ................................................................................... 72
2.11 Summary .................................................................................................................................................. 74

CHAPTER III - A SPEED-DENSITY TRANSFORMATION APPROACH .......................................................... 75
3.1 Background of the Problem ....................................................................................................................... 75
Table of Contents

4.8.3 Results .............................................................................................................................. 113
4.9 Summary .............................................................................................................................. 113

CHAPTER V - DISCRETE CHOICE BASED LANE-CHANGE MODEL ........................................ 115
5.1 Shortcomings of the Markov chain model in Chapter 4 ..................................................... 115
5.2 Discrete Choice Analysis ..................................................................................................... 116
  5.2.1 Introduction ..................................................................................................................... 116
  5.2.2 Logit Model ..................................................................................................................... 117
5.3 Discrete Choice Based Lane-Change Behaviour .................................................................. 118
5.4 State Space Model .............................................................................................................. 121
5.5 Density Estimation via the Extended Kalman Filter ............................................................ 122
5.6 Density Estimation via the Unscented Kalman Filter .......................................................... 123
  5.6.1 Traffic Data Simulation ................................................................................................. 126
  5.6.2 Addressed Research Issues ........................................................................................... 128
5.7 Simulation Results .............................................................................................................. 128
  5.7.1 Simulation Analysis Using EKF Approach ................................................................. 129
  5.7.2 Simulation Analysis Using UKF Approach ............................................................... 130
  5.7.3 Comparison between EKF and UKF Based Approaches ............................................ 136
  5.7.4 Comparison with Existing Method ................................................................................ 138
5.8 Summary ............................................................................................................................ 141

CHAPTER VI - REAL TRAFFIC DATA ANALYSIS ................................................................. 142
6.1 Data Collection ..................................................................................................................... 142
6.2 Data Screening ..................................................................................................................... 143
6.3 System Identification ......................................................................................................... 145
6.4 Applications ......................................................................................................................... 150
  6.4.1 Peak Traffic Time Analysis (8:00 am – 10:00 am) ....................................................... 151
  6.4.2 Non-Peak Traffic Time Analysis (10:00 am – 12:00 noon) ......................................... 159
6.5 Numerical Comparison ......................................................................................................... 168
6.6 Summary ............................................................................................................................. 168

CHAPTER VII - SUMMARY AND CONCLUSIONS ................................................................. 169
7.1 Summary ............................................................................................................................. 169
7.2 Conclusions ......................................................................................................................... 171
7.3 Future Work ......................................................................................................................... 172
  7.3.1 Possible Extensions Of The Current Developed Model ................................................ 173
Table of Contents

7.3.2 The Estimation/Prediction Of Other Traffic Variables .............................................174
7.3.3 Use Of Other Filtering Techniques ........................................................................175
BIBLIOGRAPHY ...........................................................................................................176
**LIST OF FIGURES**

**Figure 1.1** Schematic Diagram of a Motorway ........................................... 8  
**Figure 2.1** Schematic representation of Literature Review ............................. 14  
**Figure 2.2** Integrated Transportation Systems (ITS) ........................................ 15  
**Figure 2.3** Inductive-loop detector system (Traffic Detector Handbook, 1991) .......... 23  
**Figure 2.4** Motorway segment showing Lane-change ..................................... 35  
**Figure 2.5** Structure of lane changing models (Ben-Akiva et al., 2007) ............... 39  
**Figure 2.6** Dynamic system of KF (Ye, 2007) .................................................. 56  
**Figure 2.7** Operation of KF (Welch & Bishop, 2001) ....................................... 57  
**Figure 2.8** Operation of EKF (Welch & Bishop, 2006) ..................................... 61  
**Figure 2.9** Unscented transformation of the UKF (Ye, 2007) ............................. 64  
**Figure 2.10** Operation of the UKF ............................................................... 66  
**Figure 3.1** A schematic plot showing two sections of a motorway (Gazis & Liu, 2003) 74  
**Figure 3.2** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=2$, $\tau=2$ and $\lambda=8$ ................................................................. 84  
**Figure 3.3** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=1$ and $\lambda=10$ ................................................................. 85  
**Figure 3.3** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=2$ and $\lambda=3$ ................................................................. 86  
**Figure 3.5** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis & Liu, (2003) method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=1$ and $\lambda=1$ ................. 88  
**Figure 3.4** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis & Liu, (2003) method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=2$ and $\lambda=3$ ................. 89
List of Figures

Figure 4.1 A schematic plot showing an N-segment and M-lane motorway section .......................... 96
Figure 4.2 Markov Chain state change diagram .............................................................................. 97
Figure 4.3 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single section motorway with $\sigma=1$, $\tau=1$ and $\lambda=1$ .................................................................................................................... 104
Figure 4.4 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single section motorway with $\sigma=1$, $\tau=1$ and $\lambda=3$ .................................................................................................................... 105
Figure 4.5 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single section motorway with $\sigma=1$, $\tau=1$ and $\lambda=5$ .................................................................................................................... 106
Figure 4.6 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis and Liu, (2003) method for three lane single section motorway with $\sigma=1$, $\tau=1$ and $\lambda=10$ ........................................................................................................... 108
Figure 4.7 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis and Liu, (2003) method for three lane single section motorway with $\sigma=1$, $\tau=1$ and $\lambda=5$ .................................................................................................................... 108
Figure 5.1 Lane change probability representation using the developed method .......................... 124
Figure 5.2 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with EKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=8$ ........................................................................................................... 126
Figure 5.3 Lane change probability representation using the developed method with EKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=8$ ........................................................................................................... 126
Figure 5.4 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with EKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=2$ and $\lambda=4$ ........................................................................................................... 128
Figure 5.5 Lane change probability representation using the developed method with EKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=2$ and $\lambda=4$ ........................................................................................................... 128
Figure 5.6 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=8$ ........................................................................................................... 129
Figure 5.7 Lane change probability representation using the developed method with UKF
for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=8$…………………. 130

**Figure 5.8** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=4$…………………………………… 131

**Figure 5.9** Lane change probability representation using the developed method with UKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=4$…………………………………… 135

**Figure 5.10** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the existing method (Gazis & Liu, 2003) for lane 1 (left) and lane 2 (right) with $\sigma=2$, $\tau=2$ and $\lambda=8$…………………………………… 136

**Figure 5.11** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=2$, $\tau=2$ and $\lambda=8$…………………………………… 136

**Figure 6.1** A section of study from Motorway SR – 99…………………………………… 148

**Figure 6.2** The traffic counts using the developed method for 2-lane motorway using peak time real traffic data…………………………………… 148

**Figure 6.3** The traffic counts using the Gazis & Liu (2003) method for 2-lane motorway using peak time real traffic data…………………………………… 149

**Figure 6.4** A section of study from Motorway I – 5…………………………………… 147

**Figure 6.5** The traffic counts using the developed method for 3-lane motorway using peak time real traffic data…………………………………… 150

**Figure 6.6** The traffic counts using the Gazis & Liu (2003) method for 3-lane motorway using peak time real traffic data…………………………………… 154

**Figure 6.7** A section of study from Motorway I – 90…………………………………… 150

**Figure 6.8** The traffic counts using the developed method for 4-lane motorway using peak time real traffic data…………………………………… 152

**Figure 6.9** The traffic counts using the Gazis & Liu (2003) method for 4-lane motorway using peak time real traffic data…………………………………… 150

**Figure 6.10** The traffic counts using the developed method for 2-lane motorway using non-peak time real traffic data…………………………………… 154

**Figure 6.11** The traffic counts using the Gazis & Liu (2003) method for 2-lane motorway using peak time real traffic data…………………………………… 154

**Figure 6.12** The traffic counts using the developed method for 3-lane motorway using peak time real traffic data…………………………………… 156
List of Figures

Figure 6.13 The traffic counts using the Gazis & Liu (2003) method for 3-lane motorway using peak time real traffic data…………………………………………………… 156

Figure 6.14 The traffic counts using the developed method for 4-lane motorway using peak time real traffic data…………………………………………………… 158

Figure 6.15 The traffic counts using the Gazis & Liu (2003) method for 4-lane motorway using peak time real traffic data…………………………………….. 158
LIST OF TABLES

Table 2.1 Strengths and weaknesses of commercially available sensor technologies (Mimbela and Klein, 2007) ................................................................. 18

Table 2.2 Common faults of ILD data, and tests for these faults (Robinson and Polak, 2004) ........................................................................................................... 26

Table 2.3 Deterministic Single-regime Speed-Density Models (Wang et al., 2009) .... 51

Table 2.4 Deterministic Multi-regime Speed-Density Models (Wang et al., 2009) .... 52

Table 3.1 Deterministic Single-regime Speed-Density Models with Transformations ... 78

Table 3.2 Average RMSEs for two motorway sections over 100 simulation runs using the developed method with $\lambda=8$ ................................................................. 84

Table 3.3 Average RMSEs for two motorway sections over 100 simulation runs using the developed method with $\lambda=10$ ................................................................. 85

Table 3.4 Average RMSEs for two motorway sections over 100 simulation runs using the developed method with $\lambda=3$ ................................................................. 87

Table 3.5 Average RMSEs for two motorway sections over 100 simulation runs using the developed method with $\lambda=8$ ................................................................. 89

Table 3.6 Average RMSEs for two motorway sections over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=10$ ....................................................... 90

Table 3.7 Average RMSEs for two motorway sections over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=3$ ....................................................... 90

Table 4.1 Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=10$ ......................................................... 104

Table 4.2 Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=3$ ......................................................... 106

Table 4.3 Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=5$ ......................................................... 107

Table 4.4 Average RMSEs for three lane motorway section over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=10$ ............................................. 108

Table 4.5 Average RMSEs for three lane motorway section over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=5$ ............................................. 109

Table 5.1 Average RMSEs for a two-lane motorway over 100 simulation runs using the
List of Tables

developed state equation with $\lambda=8$ with EKF........................................ 127

Table 5.2 Average RMSEs for a two-lane motorway over 100 simulation runs using the
developed state equation with $\lambda=4$ with EKF........................................ 129

Table 5.3 Average RMSEs for a two-lane motorway over 100 simulation runs using the
developed state equation with $\lambda=8$ with UKF........................................ 130

Table 5.4 Average RMSEs for a two-lane motorway over 100 simulation runs using the
developed state equation with $\lambda=4$ with UKF........................................ 132

Table 5.5 Average RMSEs for a two-lane motorway over 100 simulation runs using the
existing method (Gazis & Liu, 2003) state equation with $\lambda=4$....................... 136

Table 5.6 Average RMSEs for a two-lane motorway over 100 simulation runs using the
existing method (Gazis & Liu, 2003) state equation with $\lambda=8$....................... 137

Table 6.1 Raw data values i.e. counts and speeds from two ILDs at the two ends of the
chosen segment for the peak traffic time analysis................................................ 144

Table 6.2 Rearranged data values i.e. counts and speeds from two ILDs at the two ends
of the chosen segment for the peak traffic time analysis...................................... 145

Table 6.3 Mean of normalised innovation sequence values considering 2-lane motorways
for developed and Gazis & Liu, (2003) method................................................. 167

Table 6.4 Mean of normalised innovation sequence values considering 4-lane motorways
for developed and Gazis & Liu, (2003) method................................................. 167
CHAPTER I

INTRODUCTION

The evolution of human civilisation has been underpinned by advancements in transportation and communication. Like many socio-economic systems, roads and transport infrastructure are key elements to the economic success of developed and third world nations. To cope with excessive traffic, transportation systems are developed including 6-8 lane motorways and toll roads such as the M6 in UK. The purpose of developing well linked road networks is not only to allow effective transportation of goods and resources but to minimize the traffic congestion in the network and to reduce incidents that may lead to heavy loss of life and property.

With the advent of CPU and the associated processing power of computers, there is significant progress in the field of transportation and traffic management. Better computing systems coupled with advanced programming techniques allow the end user to be able to receive much more information. With the development of Advanced Traffic Management System (ATMS), in the USA, Intelligent Transportation Systems (ITS), Active Traffic Management (ATM), in the UK, the traffic industry has changed completely. Various ways to collect real-time data such as loop detectors, probe vehicles, microwave (radar), laser, infrared, ultrasonic, magnetometer, video image processing provide enhanced means of dealing with traffic related problems. The most recent example is the navigation systems which are widely used in countries like the USA and UK. These systems are capable of providing directions using navigation maps from origin to destination along with the expected travel time and distance travelled.
Application of statistics in traffic studies has engrossed many researchers to study this rapidly developing area. A number of new techniques are emerging for study such as neural networks, Bayesian analysis, and multi-level modelling for the application of vehicular speed estimation, travel times, density estimation, congestion times, incident detection, accident prediction, and lane changing behaviour of commuters. These different traffic variables are related to one another and various statistical models have been introduced in the past to represent these relationships. The relationships that have been modelled are complex and the variables are often categorical in nature. In addition, a large proportion of transport studies are observational and not amenable to the design of the experiments. Consequently, the commonly used statistical methods and its applications in transportation reflect the unique characteristics of its subjects (Spiegelman et al., 2010). Virtually all statistical traffic models are based on the relationships between flow, density, and velocity. Using conventional traffic detectors one can only measure, but is more difficult to measure. The focus of this research is the statistical modelling and analysis of vehicular density estimation, which is one of the most critical variables in the study of traffic management (Coifman, 2003).

Intuitively, it is expected that the number of vehicles in a certain section of the road is somehow related to the mean speed of the vehicles in the section. In general, the higher the velocity, the lower the density and vice-versa. Thus, the density estimation has to be consistent with the velocities. In this research, a detailed study shall be carried out with the emphasis on estimation and control aspects of traffic densities. The estimation procedure is analogous to that of Gazis & Liu (2003). Gazis & Liu (2003) presented an approach to estimating density using a speed-density relationship using the KF. Their model does not take into account the lane change behaviour of drivers. However, in this research, we have taken into consideration the
effect of lane-change behaviour in a multi-lane multi-section motorway. Thus, we have proposed a recursive density estimation method which includes drivers’ lane-change behaviour modelled statistically. The estimation procedure uses the phenomenological relationship between speed and density given by Drake et al., (1967), and is based on the Kalman filter technique (KFT) which is explained later in the thesis. One important element of this approach is that the parameters of the phenomenological relation are considered as state variables of the state space system.

As stated earlier, there are various ways to collect real time velocity and count data. Loop Detectors have become one of most important sources of real time traffic data. ILDs are a cheap source to collect traffic data and thus are widely used in traffic studies (Cheung et al., 2005). In this research it is assumed that the data is collected from ILDs. Applications of the data collected from ILDs are not only limited to speed or travel time estimation but can also be utilised to estimate other traffic variables such as incident detection, density estimation, lane-change behaviour, etc when coupled with other local information such as the surrounding area, buildings, weather & time of day.

Traffic flow counts and their approximate speeds are the by-products of the surveillance sensors such as magnetic loop detectors (embedded in the pavement). Initializing with a reasonable vehicle count in a motorway section, the vehicle count in a section can be updated using counts at the entrance and exit of the respective section. After estimating the number of vehicles in different sections on a real time basis, the travel delay along the road can be attained and further used in the implementation of control systems. Thus to manage and control magnanimous traffic systems, estimation of traffic densities can be very useful.
1.1 Statement of Research

Traffic-responsive control systems require real-time information to make sensible control decisions. Effective surveillance of the systems is key to the effective control of traffic systems. Intelligent Transport Systems (ITS) are equally reliant upon the effective surveillance for the control parameters of the systems appropriately and delivering reliable information to the end users.

Transportation planning and analysis including preparing control systems needs a detailed study of transport variables such as traffic flow, vehicle speed, travel time, congestion level, queue lengths, lane-changes, etc. These traffic parameters are usually considered as the key decision variables to establish traffic control and management strategies. Sometimes they are used to characterise congestion for future use in Advanced Traveller Information System (ATIS). One of the most commonly used forms of data for transportation planning is traffic flows on the motorway segments. Hence, the problem of modelling and estimating the number of vehicles traversing a motorway segment is an important issue necessitating a great amount of attention and research from many transportation agencies like the Department for Traffic (DfT) UK and the UK Highway Agency.

Generally, it is believed that traffic density modelling and estimation is more difficult than the estimation of distance mean speed and flow. The reason behind this is partially because the most popularly used sensors are ILDs which are point sensors while density is a range concept. Location, weather and vehicle types, etc. are some of the disturbances to density estimation. Further, drivers’ lane choice and lane changing behaviour affects lane-wise density significantly. As a matter of fact, both lane-wise distance mean speed and density are very difficult to be measured and/or be estimated accurately in real-time although they are required in Coordinated Ramp Metering (CRM) and Variable Speed Limit (VSL) control (Qiu et al., 2009).
Certainly, traffic flow is affected the most by drivers’ lane-change behaviour. During incidents lane-changes from blocked lanes may cause irregular traffic movements among parallel lanes which may lead to congestion or delayed spillback events. If the situation is not managed appropriately and in a timely manner, it may lead to significant traffic problems such as congestion, gridlock, secondary accidents, etc. (Sheu & Ritchie, 2001). Hence, investigation of important traffic behaviours such as lane-change is vital and may be especially helpful in the development of advanced models and technologies for traffic control and management systems.

On motorways drivers’ lane-choice behavior and lane-changing activity has received increasing attention in the literature during the last decade (Laval & Daganzo, 2004). Research has produced both conjectures (Brackstone et al. 1998; Wei et al. 2000) and empirical evidence (Ahn, 2004; Cassidy & Bertini, 1999) suggesting how drivers’ lane choices or behaviour affect the traffic stream such as its capacity. It is to be noted that along with changing speed, lane-change is the only operation a driver can exercise but it has received very little attention compared to other traffic variables such as speed, travel times, etc. Due to limited construction funds and ever-increasing travel demand there is a need to understand complex traffic behaviour so as to provide the optimum traffic management solution.

Considerable efforts have been made to model the combination of density and lane-changing behaviour (Chang and Gazis, 1975; Sheu, 1999; Coifman, 2003) to produce improved density estimates. However, many lane-change models assume accurate measurement of density. All of the models are limited by the small amount of spatial data available for validation (Coifman, 2003). In most cases lane-change behaviour was modelled as a function of flow but it is expected that other traffic factors such as location, time of day and vehicle mix are significant.
The purpose of this research is to model drivers’ lane change statistically and using this model to estimate the traffic density between parallel lanes on a multi-lane multi-section motorway. Unlike previous studies, we consider that drivers change lanes and that lane-change effect cannot be ignored. The methodology proposed in the past is used for the density estimation and is extended to incorporate explicit consideration of lane-changing on a multi-lane motorway. Also, to counter effect the non-linearity in the complex traffic model, we propose a transformation to deal with the issue of approximation due to linearisation to avoid the possible bias in the estimation in results.

1.2 Research Objectives

This thesis presents statistical modelling and estimation of traffic density for multi-lane motorways using the information provided from ILDs. We incorporate a Markov chain model to characterise drivers’ lane-change behaviour where the transition probabilities of the Markov chain are further modelled using a discrete choice model. This lane-change model is then combined with a state space model to capture the dynamics of traffic flow. To address the highly non-linear nature of the traffic model, a recently developed powerful approach, the UKF and EKF is used. In addition, a new transformation approach has been proposed to transform the observation equation from a nonlinear form to a linear one so that the potential approximation in the UKF and EKF can be avoided.

The estimation of traffic density is based on the research framework developed by Gazis & Liu (2003) where the traffic density over time is modelled using a dynamic state space equation and the measurements on traffic flow are provided by the phenomenological relationship between speed and density (Drake et al., 1967). The existing studies mainly focus on the scenario that vehicles’ lane-change movements are not common in a
motorway and can thus be ignored. This research, however, takes into account of lane-change movements in the traffic modelling and incorporates a Markov chain model to describe the lane-change behaviour. Therefore, the proposed research objectives can be summarized as follows:

- To develop a mathematical transformation from non-linear to linear observation equation (Gazis & Liu 2003) so that the approximation can be avoided when using KF for traffic density estimation.
- To model drivers’ lane-change behaviour using Markov chain theory where the states of the Markov chain represent different lanes and transition probabilities of the Markov chain explain drivers’ lane-change behaviour respectively.
- To further extend the above Markov chain model to incorporate the dynamic nature of lane-change behaviour by modelling the transition probabilities using discrete choice model.
- To build a state space model for estimating the vehicular densities via the KF, EKF and UKF.
- To set up simulation experiment and carry out numerical studies using MATLAB to validate the estimation results.
- To compare the results of the developed model with some of the existing methodologies for traffic density estimation multi-lane multi-segment motorways.
- To gather real time ILD traffic data and carry out data analysis for validating the estimation results.

1.3 Research Methodology

1.3.1 Literature Review

A comprehensive review of the literature was conducted covering many aspects of traffic studies. Literature exploring sensor technologies,
specifically ILDs and their applications, was researched and reviewed. Traffic density and the lane-change behaviour of the drivers were studied and existing literature related to speed-density relationship was researched. For estimation of densities, the KFT including EKF, UKF and their applications in the traffic studies was researched & applied in this study.

1.3.2 Study Design and Data Collection

Study design included the selection of the study corridors from the US based motorways for data collection. A typical motorway scene is shown in the Figure 1.1. The data was collected from TDAD (Traffic Data Acquisition and Distribution) which is a project, developed with the sponsorship of the Washington State Department of Transportation (TDAD, 2011). For the present research, the volume and speed data are required from ILDs and hence corridors equipped with both loops and speed traps are selected. As per TDAD “Each night at midnight, a program analyzes the previous day’s data to sum the volume measurements from each of the "station" sensors. These values represent the total number of vehicles, in all mainline lanes, that have passed a particular location within a given 20-second step. The resulting daily counts are then placed in a database table.....”

Details about the data collection sites will be discussed later in the thesis. MATLAB is used for simulation purposes and testing the accuracy of the techniques employed in the work which is explained later.

1.3.3 Preliminary Analysis of the Data

After obtaining the field data, extensive data reduction and quality control were carried out to identify and correct any discrepancies in the data. Analysis included checking for any missing data values (speed and volume) individually, as well as in the combination at different locations. As mentioned in the previous sub-section, polling cycle of the detector system
during data collection was 20 seconds, but the cycle occasionally skipped to 60 or 80 seconds, leading to missing data (for details about handling of missing data, see chapter 6, section 6.2).

![Aerial Camera](image)

**Figure 1.1** Schematic Diagram of a Motorway

### 1.3.4 Developing the Traffic Density State Space Model

The present study incorporates the approach of state space modelling to examine the problem of recursive estimation of traffic density using the information provided by ILDs embedded on motorways.

As mentioned in Section 1.2, a Markov chain model is used to describe the lane-change behaviour where states of the chain represent different lanes of the motorway and the lane-change probabilities between two adjacent lanes explains the lane-change behaviour. Moreover, these probabilities are not fixed but rather depend on the current traffic condition. A discrete choice model (Ben-Akiva and Lerman, 1985; Train, 2003) is used to capture drivers’ lane choice behaviour. The drivers’ choice probabilities are modelled by several traffic related attributes such as vehicle time headway, traffic density and speed (Chang and Kao, 1991).

### 1.3.5 Estimation of the Traffic Density

The highly non-linear measurement equation (from the state space model)
is used for traffic density estimation. The traffic densities are estimated using the measurements from the observation equation based on the relationship between vehicular speed and traffic density. To address the issue of high non-linearity of the state space model, the KFT and its various versions (such as EKF and UKF) are used to estimate the traffic density recursively.

1.3.6 Computational Analysis

To test the validity of the developed model numerical studies were carried out using the widely employed modelling software package called MATLAB. The proposed model is based on the traffic flow theory and uses volume and speed data from ILDs as input. Lane wise traffic density during a particular time interval between two selected locations was calculated as the model output. The validity of the model was first checked using simulated data with MATLAB. Validation of the results whilst using field data was also carried out using loop and speed trap data.

1.3.7 Statistical Analysis and Comparison of Results

Statistical analysis was performed, to check the validity of the results. A performance measure of Root Mean Square Error (RMSE) was used for this purpose. The validity checks were carried out independently at each stage of this thesis. In all three versions of the proposed model the validation checks are performed and the methods are compared with existing methods. Towards the end of the thesis real data analysis was carried out and the traffic density was estimated using KF. Furthermore, it has been shown that the developed model performs better than the existing methods under variable traffic flow conditions.
1.4 Contribution of the Research

This research study contributes to the literature of traffic studies in several aspects.

- Firstly, it takes into account drivers’ lane choice via lane-change movements in the traffic modelling and incorporates a Markov chain model to describe the lane-change behaviour. The existing studies mainly focus on the scenario that vehicles’ lane-change movements are not common in a motorway and can thus be ignored. This research, however, considers the Markov chain processes to describe lane-change behaviour stochastically, so that the state equation can better reflect the movements of real traffic flow.

- Secondly, the lane-change probabilities between two adjacent lanes are not assumed to be static (not changing with time) but rather they depend on the current traffic condition. We use a discrete choice model (Ben-Akiva and Lerman, 1985; Train, 2003) to capture drivers’ lane choice behaviour. The drivers’ choice probabilities are modelled by several traffic-condition related attributes such as vehicle time headway, traffic density and speed (Chang and Kao, 1991). Hence, there is greater probability that the model is applicable in practice since it uses real time information rather than just fixed parameters as found in the existing literature (Chang and Kao, 1991).

- To address the issue of high non-linearity of the state space model, UKF is used to estimate the traffic density recursively. Other versions of KF have also been employed to validate the developed methodology.

- This study incorporates a suitable transformation to deal with the non-linear observation equation of Gazis and Liu (2003) so that the approximation is avoided when using KF to estimate the traffic
density. Hence the results obtained are free from effect of linearization.

- Accuracy and Robustness – The numerical results presented in this research study show the accuracy and efficiency of the proposed models as compared to the other existing methodologies. The numerical studies here proved the robustness of the developed models as they are applicable under more realistic, variable road traffic conditions.

This research has led to three research papers as follows.


Moreover, the developed models are applicable to a wider range of traffic infrastructure including multi-lane multi-section motorways unlike previous studies which were focused mainly on single lane roads (Szeto and Gazis, 1972; Papageorgiou and Vigos, 2008; Vigos, 2008; Stankova & Schutter, 2010).

1.5 Organisation of the Thesis

This thesis is organised into seven chapters. Chapter I introduce the thesis with a detailed explanation of the research challenge under consideration followed by the main research objectives aimed to be achieved during this
research study. The research methodology section briefly explains the various tools and methods applied in this study followed by the explanation of the main contributions of the study, closing the chapter by presenting the structure and organisation of the thesis.

Chapter II presents a critique of the literature review in traffic studies including sensor technologies, ILDs, traffic density estimation, lane-change behaviour and speed-density relationship. Literature was also studied relating to modern filtering techniques of control theory including KF, EKF and UKF along with the applications of KF in the traffic studies.

Chapter III proposes the first approach developed for density estimation based on a speed-density relationship using data from ILDs. The developed approach is based on the model by Gazis and Liu (2003) but with important modifications via a suitable transformation.

The above developed model is further extended and presented in Chapter IV introducing the Markov chain lane-change model where it is assumed that drivers do change lanes whilst on multi-lane motorways. Chapter V then considers the scenario that the lane-change probabilities between two adjacent lanes are not fixed but rather they depend on the current traffic condition. A discrete choice model is used to capture drivers’ lane choice behaviour. The drivers’ choice probabilities are modelled using various road traffic attributes such as vehicle time headway, traffic density and speed.

Chapter VI presents real data analysis and validates the applicability of the above developed models. The real time data is collected from ILDs and speed traps (double inductive loops) from 2, 3 and 4-lane motorways. The System identification technique is used to identify the system and KF is employed to estimate traffic densities on real time basis. A comparative study with the existing method is also presented in this chapter.

Chapter VII consists of a summary of the research study and conclusions. This chapter also lists some further possible extensions of the developed
Chapter I - INTRODUCTION

models, recommendations for the future research and presents a brief study on how the currently estimated density can be used to estimate or predict some other traffic variables in real time followed by the bibliography towards the end of this thesis.
CHAPTER II

LITERATURE REVIEW

An efficient transportation system of any nation is one of the most important pre-requisite for its growth and advancement. To sustain & increase these levels of development, magnanimous transport infrastructures (roads, motorways, flyovers, roundabouts) are introduced to provide smooth movement of goods and commuters. Consequently, problems such as traffic jams, congestion, accidents and collisions also occur. Currently many researchers are working to handle congestion scenarios by estimating section denseness and respective speeds. In this chapter we present a detailed literature review carried out for this research which can be broadly divided into two categories namely traffic Studies and KFT and their application in traffic studies.

The following diagram explains the structure of the literature review carried out in this research study.

![Figure 2.1 Schematic representation of Literature Review](image-url)
A. TRAFFIC STUDIES

Intelligent transportation systems (ITS) promise more efficient transportation networks. In order to achieve greater network efficiency these systems use advanced information processing and communication technologies to manage traffic systems (such as signal setting, ramp metering, incident detection, verification and clearance) and to control the flow of vehicles. A schematic representation of ITS is shown in Figure 2.2.

The Advanced Traffic Management System (ATMS) is central to an ITS, which employs advanced technology to assist the operator by providing tools for:

- managing data coming into the control centre;
- generating information that depicts the operational status of the motorway network from various diverse resources;
• detecting and resolving incidents;
• controlling the traffic network.

An Integrated Transportation System (ITS) basically comprises of three main sub-systems namely Advanced Traffic Management System (ATMS), Traffic Management Centre (TMC) and ATIS. Initially the traffic situation is monitored and data is assembled via Closed Circuit Television (CCTV) cameras, AID systems and Automatic Vehicle Location (AVL) systems. Gathered data is thus passed on to the TMC which in turn uses the information to provide knowledge in terms of congestion and incident maps to the ATIS. The ATIS then release the knowledge for the end users in the form Variable Message Signs (VMS), call centres and via the internet, hence completing data transformation loop through usable information to actionable knowledge.

The use of advanced control techniques and automated devices in an ITS composed of motorways or urban avenues and streets is becoming a very common approach to help alleviate the daily impact of congested traffic. Adaptive cruise control, advanced traction and braking schemes, driver alert systems, variable signal systems, and real-time on-ramp metering control, are among these new technologies Varaiya (1993), Horowitz (2000). According to Mimbela and Klein (2007), the goals of an ITS include the following:

• Enhance public safety;
• Reduce congestion;
• Improve access to travel and transit information;
• Generate cost savings to motor carriers, transit operators, toll authorities, and government agencies; and
• Reduce detrimental environmental impacts.
In this chapter we will discuss the different traffic variables and related research found in the literature. We will present the review of literature involving traffic studies and filtering techniques, this will discuss various traffic related areas including theory about ILDs, speed-density relationship, lane-change behaviour and various forms of KFs including EKF and UKF.

2.1 Sensor Technologies

Sensors used for vehicle detection and surveillance by ITS may be described as consisting of three components; the transducer, a signal processing device and a data processing device (Mimbela and Klein, 2007). The transducer detects the presence of a vehicle and provides input to the signal-processing device which typically converts the transducer’s output into an electrical signal. The data-processing device then converts the electrical signal into traffic parameters. Typical traffic parameters include vehicle presence, count, speed, class, gap, headway, occupancy, weight and link travel time. As per the Traffic Detector Handbook published by the U.S. Federal Highway Administration, the two types of sensors are In-Road sensors and Over-Road sensors.

In-motorway sensors are those which may be embedded in the pavement of the motorway, in the subgrade of the motorway or may be taped or attached to the surface of the motorway.

As an example in-motorway sensors include (Mimbela and Klein, 2007)

- ILDs, which usually require saw-cutting in the pavement;
- weigh-in-motion sensors, which are embedded in the pavement;
- magnetometers, which may be embedded or placed underneath a paved motorway or bridge structure.

Over-roadway sensors are mounted above the surface of the roadway either above the roadway or alongside the roadway, away at some distance from the nearest traffic lane.
A Few examples may include (Mimbela and Klein, 2007)

- video image processors that utilize cameras mounted on poles adjacent to the roadway, on structures that span the roadway, or on traffic signal mast arms over the roadway;
- microwave radar sensors placed adjacent to the roadway or right over the lanes to be monitored;
- ultrasonic, passive infrared and laser radar sensors usually mounted over the lanes to be monitored.

Strengths and weaknesses of commonly known and commercially available sensors and sensor technologies are presented in Table 2.1 as follows.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive Loop</td>
<td>- Flexible design to satisfy large variety of applications.</td>
<td>- Installation requires pavement cut.</td>
</tr>
<tr>
<td></td>
<td>- Mature, well understood technology.</td>
<td>- Decreases pavement life.</td>
</tr>
<tr>
<td></td>
<td>- Large experience base.</td>
<td>- Installation and maintenance require lane closure.</td>
</tr>
<tr>
<td></td>
<td>- Provides basic traffic parameters (e.g., volume, presence, occupancy,</td>
<td>- Wire loops subject to stresses of traffic and temperature.</td>
</tr>
<tr>
<td></td>
<td>speed, headway, and gap).</td>
<td>- Multiple detectors usually required to monitor a location.</td>
</tr>
<tr>
<td></td>
<td>- Insensitive to inclement weather such as rain, fog, and snow.</td>
<td>- Detection accuracy may decrease when design requires detection of a</td>
</tr>
<tr>
<td></td>
<td>- Provides best accuracy for count data as compared with other commonly</td>
<td>large variety of vehicle classes.</td>
</tr>
<tr>
<td></td>
<td>used techniques.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Common standard for obtaining accurate occupancy measurements.</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Magnetometer (Two-axis fluxgate magnetometer)</td>
<td>- High frequency excitation models provide classification data.</td>
<td>- Installation requires pavement cut.</td>
</tr>
<tr>
<td></td>
<td>- Less susceptible than loops to stresses of traffic.</td>
<td>- Decreases pavement life.</td>
</tr>
<tr>
<td></td>
<td>- Insensitive to inclement weather such as snow, rain, and fog.</td>
<td>- Installation and maintenance require lane closure.</td>
</tr>
<tr>
<td></td>
<td>- Some models transmit data over wireless RF link.</td>
<td>- Models with small detection zones require multiple units for full lane detection.</td>
</tr>
<tr>
<td>Magnetic (Induction or search coil magnetometer)</td>
<td>- Can be used where loops are not feasible (e.g., bridge decks).</td>
<td>- Installation requires pavement cut or boring under roadway.</td>
</tr>
<tr>
<td></td>
<td>- Some models are installed under roadway without need for pavement cuts.</td>
<td>- Cannot detect stopped vehicles unless special sensor layouts and signal processing software are used.</td>
</tr>
<tr>
<td></td>
<td>- Insensitive to inclement weather such as snow, rain, and fog.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Less susceptible than loops to stresses of traffic.</td>
<td></td>
</tr>
<tr>
<td>Microwave Radar</td>
<td>- Typically insensitive to inclement weather at the relatively short ranges encountered in traffic management applications.</td>
<td>- CW Doppler sensors cannot detect stopped vehicles.</td>
</tr>
<tr>
<td></td>
<td>- Direct measurement of speed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Multiple lane operation available.</td>
<td></td>
</tr>
<tr>
<td>Active Infrared (Laser radar)</td>
<td>- Transmits multiple beams for accurate measurement of vehicle position, speed, and class.</td>
<td>- Operation may be affected by fog when visibility is less than ≈20 ft (6 m) or blowing snow is present.</td>
</tr>
<tr>
<td></td>
<td>- Multiple lane operation available.</td>
<td>- Installation and maintenance, including periodic lens cleaning, require lane closure.</td>
</tr>
<tr>
<td>Passive Infrared</td>
<td>- Multizone passive sensors measure speed.</td>
<td>- Passive sensor may have reduced sensitivity to vehicles.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter II – LITERATURE REVIEW

| Ultrasonic                        | Multiple lane operation available. |
|                                  | Capable of overheight vehicle detection. |
|                                  | Large Japanese experience base. |

- Some models not recommended for presence detection.
- Environmental conditions such as temperature change and extreme air turbulence can affect performance. Temperature compensation is built into some models.
- Large pulse repetition periods may degrade occupancy measurement on motorways with vehicles traveling at moderate to high speeds.
- Cold temperatures may affect vehicle count accuracy.
- Specific models are not recommended with slow moving vehicles in stop-and-go traffic.
- Installation and maintenance, including periodic lens cleaning, require lane closure when camera is mounted over roadway (lane closure may not be required when camera is mounted at side of roadway).
- Performance affected by inclement weather such as fog, rain, and snow; vehicle shadows; sun glint; vehicle projection into adjacent lanes; occlusion; day-to-night

| Acoustic                          | Passive detection. |
|                                  | Insensitive to precipitation. |
|                                  | Multiple lane operation available in some models. |

| Video Image Processor            | Monitors multiple lanes and multiple detection zones/lane. |
|                                  | Easy to add and modify detection zones. |
|                                  | Rich array of data available. |
|                                  | Provides wide-area detection when information gathered at one camera location can be linked to another. |

- Performance affected by inclement weather such as fog, rain, and snow; vehicle shadows; sun glint; vehicle projection into adjacent lanes; occlusion; day-to-night
transition; vehicle/road contrast; and water, salt grime, icicles, and cobwebs on camera lens.
• Requires 30- to 50-ft (9- to 15-m) camera mounting height (in a side-mounting configuration) for optimum presence detection and speed measurement.
• Some models susceptible to camera motion caused by strong winds or vibration of camera mounting structure.
• Generally cost-effective when many detection zones within the field-of-view of the camera or specialized data are required.
• Reliable nighttime signal actuation requires street lighting.

As shown in the table, there are various ways to collect real time velocity and count data such as loop detectors, probe vehicles, microwave (radar), laser, infrared, ultrasonic, magnetometer and video image processing. However, ILDs are a cheap source to collect traffic data and thus are widely installed (Cheung et al., 2005). So the main focus of this research would be to tap the rich data available from commonly used ILDs. ILDs are explained in detail in the next section.

2.2 Inductive Loop Detector (ILD)

2.2.1 Introduction
ILD systems were first applied to the study of transportation in the 1960s
(Gillmann, 2002). A vehicle detector system is defined as "....a system for indicating the presence or passage of vehicles." - National Instruments Manufacturers Association (NEMA). These systems hold an important place in the branch of traffic management and are considered as the backbone of traffic management.

In simple words, ILDs are wires laid in 2m long by 1.5 m wide loops, several centimetres under the road (Robinson 2005). They provide input for traffic-actuated control, motorway surveillance, traffic responsive control and data collection systems (NEMA, 1983). For more than 50 years, these detectors have been used for motorway traffic counts, surveillance and control (Labell et al. 1989). Out of all the detector systems, the most widely used is the ILD system (Raj & Rathi 1994; Traffic detector Handbook 1991).

The inductance loop detectors supply vehicle presence, vehicle count and occupancy in real-time. Although speed cannot be measured directly, it can be estimated by using a two-loop speed trap or a single loop along with the knowledge of vehicle length and effective loop length (Klein, 2001). Loop data are typically relayed to a centralized TMC for analysis.

ILDs are the most commonly used sensors in traffic studies. Although there are other sensors for data collection, the high cost/benefit performance of ILD makes them the most extensively used sensors (Liu and Ma, 2006). The rich source of the ILD data is the reason for selecting them as the basis of this study for data collection. Traffic algorithms are data intensive, and require accurate information about prevailing traffic conditions. ILDs have achieved more than a 95% detection rate on vehicle counts and have a margin of error of less than 5mph on speed measurements (Ahmed, 1986). Although the loops are reading data many times a second, the data is typically reported back to the TMC at intervals of 20 or 30 seconds.
2.2.2 How They Work

An ILD senses the presence of a conductive metal object by inducing currents in the object, which reduce the loop inductance. If the current decreases beyond a certain threshold then the presence of the metallic object or vehicle is signaled by the output of a binary ‘1’ bit. If no vehicle is detected then an output of binary ‘0’ bit is signaled (Robinson 2005). The elements of a detector include the following:

- an inductive loop (of one or more turns of wire embedded in the roadway pavement)
- a pull box
- a lead-in wire (running from the wire loop to a pull box)
- a lead-in cable (connecting the lead-in wire at the pull box to the controller unit)
- a loop controller, which normally consists of a tuning network, a signal amplifier, a data accumulator and other detector electronics

A basic loop is an insulated electrical wire, usually several meters to a side, with several turns. The electromagnetic field depends upon the shape of the loop so it may be square, rectangle, diamond, circular or octagonal. For instance, diamond loops reduce the probability of detecting vehicles in adjacent lanes. The pull-box, usually located adjacent to the road, houses the splices between the lead-in cable from the controller and the lead-in wires from the loop. Lead-in wires are usually shielded and twisted to eliminate...
disturbances from external electromagnetic fields, such as adjacent loops. The controller electronics, usually housed in a rugged cabinet in a safer more accessible location, detect, amplify, and process the loop signals. The controller provides power & excitation for the loop. The controller unit also supports other functions such as selection of loop sensitivity and pulse or presence mode operation to detect vehicles that pass over the detection zone of the loop. ILDs are placed either on or up to twenty inches below the road surface (Traffic Detector Handbook, 1991).

2.2.3 Advantages of ILDs

The advantages of ILDs can be summed up as follows:

- ILDs continue to be the best in all weather, all light condition sensors for many applications, when properly installed and maintained.
- It is one of the most consistently accurate detector technologies in terms of vehicle counts as per the Hughes Aircraft Study (research evaluating non-intrusive detection accuracy).
- ILDs perform well in both high and low volume traffic and in different weather conditions.
- ILDs meet even the most stringent vehicle flow error specifications required by some ITS application.

2.2.4 Disadvantages of ILDs

The drawbacks of ILDs are as follows:

- Due to improper connections made in the pull boxes and in the application of sealants over the sawcut, the loop detector system may suffer from poor reliability. These problems are accentuated when loops are installed in poor quality pavements or in areas where utilities frequently dig up the roadbed.
• The inability to measure the speed directly is also one of the disadvantages. If the speed is required, then a two-loop speed trap (dual loop detector) is employed or in the case of a single loop detector, an algorithm involving loop length, average vehicle length, time over the detector, and number of vehicles counted is used (Klein, 2001).

• In some locations ILDs are not the most appropriate detectors. Such locations may include where pavement conditions are unfavourable on structures or where detection is needed across railway tracks.

ILDs are one of the first choices for data collection because of their high reliability (Klein and Kelley, 1996). However, if they are not maintained then ILDs are subject to failure. Table 2.2 lists common faults of ILDs. Problems can be encountered when aggregating the raw data. For example a vehicle may be present over a detector at the end of one time period and the beginning of the next time period. However the vehicle will only be counted in the first time period. In literature some of the earlier research studies considered loop detector errors, its causes and effects. These include the work by Dudek et al. (1974), Courage et al. (1976), Chen and May (1987), Bender & Nihan (1988), Gibson et al. (1998), Berka (1998).

Sometimes, before using ILD data in any of the traffic models, some researchers verified the quality of the ILD data. Most data quality tests are univariate and revolve around checking each attribute independently of other attributes (flow, occupancy, pulse length, etc.) against a minimum and maximum threshold (Chen and May, 1987). Data quality tests based on multivariate analysis have been proposed by Jacobson et al (1990) and Cleghorn et al (1991), but are not in wide use in Traffic Management Centres (TMCs).
Table 2.2 Common faults of ILD data, and tests for these faults (Robinson and Polak, 2004)

<table>
<thead>
<tr>
<th>Error Name</th>
<th>Description and Characteristics of error</th>
<th>Test for Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILD broken cable</td>
<td>ILD not outputting any data. All zero output recorded</td>
<td>• All zero output</td>
</tr>
<tr>
<td>ILD chattering</td>
<td>An ILD incorporates an analogue to digital converter. If the analogue signal varies slightly around the</td>
<td>• Minimum on pulse time</td>
</tr>
<tr>
<td></td>
<td>threshold by which a vehicle presence is indicated, then the ILD will rapidly change from an ‘on’ to an</td>
<td>• Maximum flow</td>
</tr>
<tr>
<td></td>
<td>‘off’ state. This is often caused by crosstalk.</td>
<td></td>
</tr>
<tr>
<td>ILD Crosstalk</td>
<td>Interaction between ILD and other electrical devices. Will output erroneous ‘1’s.</td>
<td>• Minimum on pulse time</td>
</tr>
<tr>
<td>ILD desensitised</td>
<td>ILD gets out of tune, particularly when the wire becomes wet, reducing the threshold at which a vehicle</td>
<td>• Check against historical data (preferably data following recalibration)</td>
</tr>
<tr>
<td></td>
<td>presence is determined. The occupancy recorded is more sensitive to this than the flow recorded.</td>
<td></td>
</tr>
<tr>
<td>ILD hanging</td>
<td>Sensors can get stuck in either an ‘on’ or ‘off’ position.</td>
<td>• Maximum on and off pulse time</td>
</tr>
<tr>
<td>ILD pulsing</td>
<td>Changes in a vehicle’s inductance over its length may lead to two pulses being recorded for the same</td>
<td>• Minimum entropy</td>
</tr>
<tr>
<td></td>
<td>vehicles</td>
<td>None available</td>
</tr>
<tr>
<td>ILD recovery time</td>
<td>After a period of sustained occupancy the ILD takes time to return to its original sensitivity. This</td>
<td>None available</td>
</tr>
<tr>
<td></td>
<td>recovery time may lead it to not being able to detect a separate vehicle following another vehicle.</td>
<td></td>
</tr>
<tr>
<td>Communications</td>
<td>Communication link between ILD and traffic management centre broken. No raw data obtained.</td>
<td>• All zero output.</td>
</tr>
<tr>
<td>failure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Following the detailed discussion about ILDs, we will now discuss the applications of ILDs in traffic studies.
2.2.5 Application of ILDs for Estimating Traffic Variables

For a given motorway section, if inductive loops are installed with the knowledge of entry and exit flow, the vehicle count or volume can be determined between the two adjacent loop stations. Using this information, along with the vehicle count change and the distance between the two stations, an average density can be estimated. Due to the widespread deployment of ILDs, it’s easy to use the available infrastructure to measure the respective traffic variables (Qiu et al. 2009).

As mentioned already, ILDs have been employed to estimate vehicle counts, densities, speeds, travel time and to detect incidents on motorways. Gazis & Knapp (1971) used ILDs at the entrance and exit point of the section to estimate the number of vehicles using speed and flow measurements. Flows and occupancies from ILDs have been used by Andrew (1980) to estimate segment density in all types of traffic conditions. Coifman (2003) used ILD data to estimate lane in-flow and lane density using vehicle re-identification techniques.

ILDs have been used by Ni (2007), Sun et al. (2004), Kurkjian et al. (1980), Papageorgiou and Vigos (2008), Vigos (2008), Knapp (1973), Szeto and Gazis (1972) to estimate densities on motorways.

Nam and Drew (1996, 1998, 1999) used ILDs to present a macroscopic model to estimate motorway travel times over motorways. Dailey (1993) measured travel times using cross-correlation techniques with inductance loop data. A linear input-output auto regressive moving average (ARMA) model was proposed by Van Arem et al. (1997) to estimate travel time from ILD data. Oh et al. (2003) used cumulative counts at two detector points to estimate travel time based on fluid model relations. Other researchers who used ILDs to estimate various traffic variables include Coifman (1998), Coifman and Cassidy (2002), Kuhne et al. (1997), Sun et al. (1999, 2004), Palacharla and Nelson (1999), Hoogendoorn (2000).
Traffic flow (vehicle counts) is one of the by-products of ILDs. Initializing with a reasonable vehicle count in a roadway section, a vehicle count in the section can be updated using counts at the entrance and exit of the respective section. After estimating the number of vehicles in different sections on a real-time basis, the travel delay along the road can be attained and further used in the implementation of control systems. The next section discusses one of the important traffic variables, the traffic density and its estimation, in detail.

2.3 Traffic Density Estimation

2.3.1 Introduction

Vehicular density plays a vital role for managing and controlling traffic operations in urban networks. It is important to have the knowledge of traffic density in real-time especially in large cities for signal control and effective traffic management (Ozkurt and Camci, 2009). The Highway Capacity Manual (2000) states that the traffic density is a critical criterion to classify the service level of a road.

As an instantaneous and range concept, traffic density is primarily defined by looking into a snapshot photo of the traffic by an aerial camera along a stretch of motorway (Greenshields, 1933; Greenshields, 1935). The density in simple words is the number of vehicles divided by the length of stretch of the motorway. Average density over lanes is naturally defined by further dividing the number of lanes.

Note that this concept can be described as continuous in space but discrete in time. However, practical traffic network systems, particularly motorway networks, do not have aerial cameras to continuously monitor the traffic in real-time. Although dense point sensor system (inductive or magnetic loop detectors) could approximate continuous measurement in space, the cost in general is prohibitive. In practice, point sensors such as
ILDs are popularly used for traffic detection, which could continuously count vehicle number in traffic stream in real-time at the sensor locations (Qiu et al., 2009).

2.3.2 Different Ways to Obtain Traffic Densities

In general, traffic density is obtained in two ways:

- the traffic survey and
- the estimation method based on the information which may be derived from raw counts obtained from one of the detection devices.

Traffic surveys can be classified into in-out traffic flow survey and high altitude photographs, but due to high costs involved, the latter is rarely used (Hu and Yang, 2008). This means that the counts from detectors are the cheapest available option for the researchers to estimate densities and therefore monitor the traffic flow state. However, these counts are subject to errors, which can degrade the density estimates substantially. The focus of this research is limited to estimation methods for traffic densities.

2.3.4 Traffic Densities From Estimation Methods

2.3.4.1 Filtering Based Estimation Methods

To improve count estimates, Szeto and Gazis (1972) in the early 1970s used filtering technique applying extended version of the KF. Their algorithm assumes a discrete-time control system which re-linearises the dynamics about each new estimate as they become available. As a consequence of re-linearization, large initial estimation errors are not allowed to propagate through time and therefore the linearity assumptions are less likely to be compromised. In this estimation process, the authors assumed that all the constant parameters were the same for all the road sections; however this is not true especially in the 21st century, even under mild conditions on motorways. It is also assumed that the traffic density
outside the exit of the section was constant in time which also is not possible in real situations except if there is Stop-Go traffic or a Traffic jam. Therefore the proposed algorithm was ‘model-limited’.

EKF has been used by Sheu (1999) for presenting a stochastic system modelling approach to extract real-time information of section-wide inter-lane as well as intra-lane-traffic densities utilizing lane traffic counts and occupancies data collected from conventional point detectors. A stochastic system can be regarded as a time varying system since its components change with time (Santina et al., 1994). A discrete-time, nonlinear stochastic model is then proposed using defined state variables and a set of assumptions to formulate the specified stochastic system. The authors’ assumption about this stochastic system, that each type of state variable is mutually independent of the other types of state variables, is very difficult to achieve in practice. Due to the limitations of field data used for modelling and unavailability/less availability of time-varying lane density data, this limited the evaluation of the proposed method.

To incorporate missing traffic density measurements, Sun et al., (2003), used the linearized Cell transmission Model based KF to provide the estimation of spatiotemporal density considering traffic states as the Markov chain. Three methods are evaluated based on traffic data from I-210 (in USA). The criterion used is the relative error at a fixed distance point where the measurement is available. For a stretch of motorway, the real-time estimation of the density at those measurement points was incorporated. However, due to the assumption of single-wave front (only one entry and exit), their models are applicable to short motorway segments rather longer segments with many entries and exits, so this model needs modification. It was also assumed that vehicles travelling at the maximum speed may not cross multiple cells (segments) in one step (section).
In line with filtering techniques Stankova & Schutter (2010) proposed the particle filtering based traffic density estimation method using the sampling importance resampling method (Arulampalam, 2002, 2004) as a selection step. They introduced Jump Markov linear model (JMLM) (Tugnait, 1982) for motorway traffic density estimation. Their effort of contributing to the traffic literature is worthwhile but the research is based on the assumption that the motorways are free from entries and exits.

2.3.4.2 Density Estimation on Multi-lane Motorways

Moving to a multi-lane scenario, Knapp (1973) developed an algorithm which can be used for multi-lane traffic without restriction on lane change where vehicle counts were previously obtained by matching the sequences of vehicle lengths detected by the detectors at the input and output of the section. He developed this algorithm to provide information for improving traffic flow through bottlenecks by input regulation. The drawback with this model was that it was working correctly only for a very short range of time i.e., just between 20-40 sec and performance was unfavourable for any smaller or larger value than this range. Also the assumption of constant vehicle length may prove more problematic while discussing multi-lane motorways (as motorways are usually use by variety of vehicles).

2.3.4.3 Occupancy Measurements Based Methods

Time and space occupancy relationships have been taken into consideration for estimating densities by Papageorgiou & Vigos (2008) and Vigos (2008). Again, the KF was used here but to deliver real time estimates of the vehicle-count in signalised links based on measurements of (at least) three loop detector stations at both extreme points and at the middle of the segment respectively. However, there are not many sections of road like this with the appropriately embedded loop detectors and therefore the
applicability of model was very limited. Moreover, this model can be applied only when the effective vehicle lengths equal to the corresponding physical vehicle lengths. In other words, the respective loop detector effectively shrinks to a line and the time-occupancy signal was to be activated for as long as that line was covered by a passing vehicle. For getting reliable results, one has to use more internal detectors within the section which may or may not need to be equally distanced.

Vehicle length based estimation of traffic densities using occupancies was studied by Gerlaugh and Huber (1983) using the valid vehicle length (which equals the sum of the length of vehicle and the width of the detector). Mathematically this relation can be represented as

\[ \text{where density} (\text{veh/km}) \text{ refers to time occupancy ([0,1]) and vll stands for valid vehicle length (unit: m). The calculation of valid vehicle length is one of the main concerns with this model along with the assumption of homogenous traffic flow.} \]

The loop detector measurements based estimation method by Kurkjian et al., (1980), says that occupancy of a detector during a particular time interval is the percentage of that interval during which the detector signals the presence of vehicles. From measurements at neighbouring detector stations, the number of vehicles and therefore the density, on the segment between stations can be tracked. The estimate is calculated using the KFT. The detection of spatially inhomogeneous conditions is performed using a Generalized Likelihood Ratio (GLR) event detection scheme. Both the filter and the GLR algorithm are simple discrete-time scalar equations. It is claimed that the unique point of the study is that it estimates the segment density accurately in all types of traffic conditions, and it does this by detecting spatially inhomogeneous traffic conditions and compensating the
density estimate appropriately for the adverse effects of the in-homogeneities. However, the overall estimation method was quite multifaceted. (Hu and Yang, 2008).

2.3.4.4 Miscellaneous Methods

(a) Vehicle Re-identification Based Method - For estimating traffic densities and lane inflows (net number of vehicles to enter (or leave) the lane) in a motorway lane between detector stations, Coifman (2003) employed information from a sparse vehicle re-identification system and used vehicle arrivals at each detector station. Although the methodology is based on the basic principle of conservation of vehicles, it is the measured travel time that allows for density estimation and upstream offset that allows for lane inflow measurement.

The approach is unique but it assumes that inflow and outflow should be balanced to get good results, i.e., the density estimation would be degraded if they are unbalanced and the estimates will drift apart. Again the algorithm is condition specific and it fails to work over mixed traffic conditions on the link; free flow at the upstream station and congested at the downstream station. Apparently, when the tail of the queue is within the link, drivers exhibit different lane changing behaviour. So, practical application is rather limited. Data used for this work was from the ILDs on the 5-lane motorway from I-880 in USA.

(b) Time-stamped Traffic Counts Based Method – Time stamped data was used as an input in the formulation of the n-t method (cumulative curves of vehicle counts in the cumulative count-time) given by Ni (2007), who uses the definition (of traffic flow) to determine traffic flow characteristics directly, including traffic density. To meet their defined criteria (for density estimation) it states: it involves only one type of sensor that is capable of
determining density, flow and distance mean speed; the measurement should be accurate; it should preserve the fundamental relationship among density, flow and distance mean speed; and it should be compatible with ITS application. Although there were great attempts to develop a new method, there were a lot of issues that hindered successful application in practice. Firstly, the method requires that no vehicle is lost or gained within any given section. Therefore, the method is ideal for basic motorway sections but not for sections containing ramps where traffic can flow in and out at middle of the sections.

Secondly, the method requires that traffic should be counted and time-stamped accurately, i.e., without miss-counts or duplication. This requirement poses challenges to some types of sensors. For example, loop detectors are prone to malfunction (see Table 2.2) and video cameras have difficulty in coping with night time and bad weather. Thirdly, the method is ideally initiated when the roadway section under analysis is empty, such as at midnight. Otherwise, one needs to choose a reference car properly so that the car does not travel too slowly or too quickly. The sensors at both ends start counting traffic at the passage of the reference car. Finally, a mechanism is needed to eliminate error accumulation due to the use of cumulative vehicle counts. An easy (but possibly not the best) solution is to reset traffic counts with the passage of a new reference car.

(c) Data from Mobiles Phones - Alessandri et al., (2003) suggested that the presence of mobile phones on a motorway can be exploited to provide information on the traffic behaviour and to estimate traffic variables. Their model combines the knowledge of traffic variables and information from the wireless network. It describes the evolution of mean speed and density for all vehicles as well as the percentage of cars with on-board cellular activity. EKF is used to estimate traffic variables such as density using the
measurements of mobile phone densities in the various cells of the wireless network. The motorway segment is divided into number of sections which are covered by a cell of the wireless network. Though it is proved to be a stepping stone towards using high-tech telecommunication channel for traffic research, the application is rather limited as the research only includes cars/vehicles with active mobile phones (switched on and in working condition) in them.

2.4 Lane Change Manoeuvre

Virtually all traffic models are based on the relationship between flow, density, and velocity. Using conventional traffic detectors one can only measure , but is more difficult to quantify (Coifman, 2003). In addition to the difficulties of capturing simple spatial measurements such as , it has long been recognized that lane change manoeuvres can influence the relationships underlying traffic flow theory or even disrupt the relationships if lane-change manoeuvres are not accounted for (Coifman, 2003).
Figure 2.4 Motorway segment showing Lane-change

2.4.1 History of Lane Change Theory - Gipps Model (1986)

Lane-changing plays an important role in traffic studies and impact significantly on traffic flow. Lane-changing models first came into existence in the late 1980’s. Gipps (1986) presented the first lane-changing model intended for micro-simulation tools and developed a framework with lane-changing decisions in urban driving scenarios. Influence of traffic signals, obstructions, transit lanes and different vehicle types such as heavy vehicles and on drivers’ lane selection choice were also modelled. Assuming logical driver behaviour, this model focuses on the decision-making process considering the potential conflicting goals.

In Gipps model, the following factors (Hidas, 2002) influence whether drivers decide to change lanes or not:

- whether it is physically possible and safe to change lanes without an unacceptable risk of collision,
- the location of permanent obstructions,
- the presence of special purpose lanes such as transit lanes,
- the driver’s intended turning movement,
- the presence of heavy vehicles, and
- the possibility of gaining a speed advantage.

The model also considers the urgency of the lane-changing manoeuvre through the driver’s gap acceptance and braking behaviour. However, it was assumed that a gap of sufficient size should be available in the target lane so that the lane-change manoeuvre can take place. That was when a lane change manoeuvre was classed as ‘safe’.

2.4.2 Limitations of Gipps Model

For the implementation of the lane-changing algorithms in SITRAS (Simulation of Intelligent TRAnsport Systems), Gipps model provided an
opportune starting point but this assumption was found to be a serious limitation in congested and incident-affected conditions which needed further consideration. This study modelled driver behaviour deterministically and did not estimate model parameters formally. One more drawback of Gipps model was that the check of the feasibility of lane-changing needs to be done for every vehicle during the vehicle update process. The reason behind this was that the check was performed before actually checking whether the vehicle needs to change lane. This appeared to be illogical, however, in order to minimise the number of vehicles for which further checks were needed; it was beneficial to perform the fastest check first from a computational efficiency point of view. In Gipps model the feasibility of a lane-change was based on relatively simple conditions, which may justify the selected order. However, if we need to apply more complex procedures for a necessary lane-change, it may be better to establish the need for a lane-change before dealing with the feasibility of the manoeuvre.

2.4.3 Different Types of Lane Changes

Based on Gipps’ model, lane-changing behaviours were implemented by several micro-simulators. In CORSIM, (which is microscopic traffic simulator that uses FREESIM to simulate motorways and NETSIM to simulate urban streets, (Halati et al., 1997)) lane-changes are classified as follows:

- Mandatory (MLC) Lane Change
- Discretionary (DLC) Lane Change

MLC are performed when the driver must leave the current lane (e.g. in order to avoid a lane blockage or to use an exit or to follow their path or to comply with lane use regulations) and will change to the nearest acceptable lane.
DLC are performed when the driver perceives that driving conditions in the target lane are better, but a lane-change is not imperative (see the Figure 2.5 for details).

Usually, drivers do not change lanes unless they feel the need to do so. To reach the final destinations, drivers will need to do mandatory lane changes when the current lane is not available. On the other hand, drivers will do discretionary lane changing, to alter and adjust the vehicle speed. It should be noted that even under a mandatory lane-changing situation, the driver does not need to change lanes immediately.

Merging is a representative case of mandatory lane changing which can be stated as the means of entering a vehicle from a ramp into the main lane. Merging occurs regularly at ramps, it is easy to note merging situations. As merging significantly affects the main lanes’ volume and traffic flows, the gap acceptance phenomenon that occurs at merging sections is a research topic in its own right. “And it is possible to present precise lane changing via gap acceptance” (Hwang and Park, 2005). In the next sub-section, we will present the gap acceptance theory and its applications for modelling lane changes.

### 2.4.4 Gap Acceptance Theory

Gap acceptance models are used to model the execution of lane-changes. The available gaps are compared to the smallest acceptable gap (critical gap) and if the available gaps are greater, a lane-change is executed. Gaps may be defined either in terms of time gap or free space. Most models require that both gaps are acceptable, namely the lead gap and the lag gap. The lead gap is the gap between the subject vehicle and the vehicle ahead of it in the lane to which it is changing. The lag gap is the gap relative to the vehicle behind in that lane.
Gap acceptance is considered to be an important component of the lane changing process and has thus received much attention in the literature. Earlier efforts were based on the distribution of the critical gap with no attempt to describe the fundamental behaviour. Critical gap can be defined as the unobservable minimum gap a driver is willing to accept in order to merge (Ahmed et al., 1996).

Ahmed (1999), in outlining the lane-changing model, assumed that the driver considers the lead gap and the lag gap separately. In order to execute the lane-change, both gaps must be acceptable. Critical gaps are assumed to follow a lognormal distribution in order to guarantee that they are non-negative. Parameters of the target lane and gap acceptance models have been jointly estimated.

Daganzo (1981) used the probit model approach to estimate the parameters of a normal distribution of critical gaps at intersections and to determine the heterogeneity of driver’s behaviour. This scheme accounted for both within driver variation and across driver variation. However, some estimation problems came to notice when later on Heckman (1981) revealed

![Figure 2.5 Structure of lane-changing models (Ben-Akiva et al., 2007)](image-url)
that to identify the model, one of the above-mentioned components of stochastic variation needed to be normalized.

Factors which affect gap acceptance behaviour at intersections are the delay (up to the occurrence of the gap under consideration) and first gap indicator. Incorporating these factors, Kita (1993) also used the binary logit model to formulate the problem of gap acceptance at merging points between entries and motorways. Gap length and relative velocity are used as explanatory variables where gap length holds for the remaining distance to the end of the acceleration lane. However, in this case, the probable serial correlation in a sequence of rejected gaps until one is accepted, was not considered in the model formulation and estimation.

After discussing lane-change modelling via gap acceptance theory, we now present different lane-change models.

2.4.5 Different types of Lane-Change Models

A number of microscopic and macroscopic traffic simulation models were developed in the last few decades which incorporated some form of a lane-changing model. Mostly the implemented lane changing models are based on a set of rules, but the description of the rules is usually superficial and incomplete. So we are presenting the two broadly classified methods as follows:

- Rule Based Lane-Change Methods
- Non-rule Based Lane-Change Methods

a. Rule Based Methods - Yang and Koutsopoulos (1996) developed a rule-based lane-changing model that is applicable only for motorways. Their model is implemented in Microscopic Traffic Simulator (MITSIM). Unlike Gipps (1986), to model drivers' lane-change behaviour they used a probabilistic framework when they faced conflicting goals. When the speed
of the lead vehicle is below a desired speed, the trailing driver considers a discretionary lane-change and checks neighbouring lanes for opportunities for increasing speed. To determine whether the current speed is low enough and the speeds of the other lanes are high enough to consider a DLC, two parameters, the impatience factor and the speed indifference factor were used.

Yang and Koutsopoulos (1996) also developed a gap acceptance model that captures the fact that the critical gap length under an MLC situation is lower than that under a DLC situation. However, no formal parameter estimation was performed and a framework to do so was never developed.

In order to capture both MLC and DLC situations, Ahmed et al. (1996) and Ahmed (1999), developed and estimated the parameters of a lane-changing model using discrete choice framework as a sequence of four steps. These include the decision to consider a lane-change, the choice of a target lane, acceptance of gaps in the target lane, and performing the lane-change manoeuvre. As far as model estimation is concerned, the utilities capturing the first and the fourth steps cannot be uniquely identified in the absence of any indicator available to the analyst differentiating these two steps. He estimated parameters of the model only for a special case: merging from a motorway entry.

Ahmed used second-by-second vehicle trajectory data to estimate the parameters of this model. The model failed to explain the conditions that trigger MLC situations. Therefore, parameters of the MLC and DLC components of the model were estimated separately. Gap acceptance models were estimated jointly with the target lane model in each case. However, the gap acceptance model cannot be applied to a case of forced merging or merging through courtesy yielding. In this scenario, due to high congestion levels, gaps of acceptable lengths may not exist and in order to merge, gaps have to be created.
Considering normal flow conditions (i.e. without incidents) Yousif and Hunt (1995) studied the relationship between lane utilization and traffic flow in dual carriageway roads. They proposed a microscopic simulation model for investigating lane-change behaviour on multi-lane unidirectional roadways. Similar logic, as in the case of Gipps (1986), was applied to establish the rules of the desire and the possibility to change lane on roadways. Again, the assumption of the model is that if the available gap in the target lane is smaller than a given acceptable limit, no lane changing will take place. The Model was acceptable for normal flow conditions but it failed to produce realistic results when traffic flow was affected by incidents and lane closures.

Wagner et al. (1997), following in the steps of Yousif and Hunt described a ‘minimal microscopic’ traffic model developed to reproduce macroscopic characteristics of traffic flow on multi-lane roads over a wide range of flow levels under normal (incident-free) conditions. The target of this study is to define ‘realistic’ traffic rules for the modelling of lane usage on multi-lane roads. They define a set of rules for lane-changing that describe when a car wants to change lane and a ‘security constraint’ rule which specifies that a car which wants to change lane is not allowed to hinder the car behind on the other lane. Under normal conditions the model was found to be able to reproduce satisfactory lane usage characteristics on multi-lane roads. However, for other traffic flow conditions such as stop-go and under incident-like situations, the model has not provided satisfactory results.

Barcelo et al. (1996) described the microscopic traffic simulator for incident free situations, Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks (AIMSUN2), which developed for modelling real-time traffic management and information systems. The lane changing model was based on Gipps’ model (1986). Throughout the
simulation time period, a vehicle on the network is continuously modelled according to several driver behaviour models (car following, lane changing, gap acceptance). Although it is stated that AIMSUN2 can also model incidents, no information is given on how the model deals with lane changing under incident scenarios.

Fritzsche (1994) describes a microscopic traffic simulation model to overcome bottleneck scenarios where one lane of a multi-lane road is temporarily closed. This is a typical situation where vehicles trapped behind the lane closure during congested flow conditions cannot move into the unblocked lane without the active cooperation of drivers in that lane. The weak part of the study was very brief explanation of the lane-changing rules in the paper. Further, they are not considered for forced or cooperative lane-changing behaviour which might be one of the options for the drivers trapped in the blocked lane.

b. Non Rule Based Methods - As compared to the commonly used rule based methods, Hunt and Lyons (1994) applied somewhat different approach by developing a driver decision-making model for lane changing using artificial neural networks (ANN). Their model works by assessing simple visual pattern-based input describing the driving environment around the vehicle about to change lane. Their model does not consider possible cooperation between drivers during lane-changes.

Cooperative lane changing is explicitly addressed in MITSIM (Yang and Koutsopoulos, 1996), in which a courtesy yielding function is used to make space for a vehicle moving into the neighbouring lane. Although no details of the process are described in the paper, the concept appears to be similar to that implemented in SITRAS.

Coming to forced lane changing, not much information was found in the literature but a similar behaviour was discussed in a study (Troutbeck and
Kako, 1997) of gap acceptance behaviour at roundabout entries. They found that ‘gap-forcing’ or ‘priority-sharing’ behaviour exists to a small extent even at low saturation levels, and at high saturation levels the ratio of forced gaps to all the merged gaps may be up to 6–12% (Hidas, 2002). It was found that as the level of congestion increases drivers attempting to enter the roundabout are becoming more insistent and/or drivers in the circulating carriageway are becoming more eager to share their right-of-way by slowing down to create an acceptable gap for the incoming vehicles. Common driving experience indicated by this type of priority sharing behaviour is even more prevalent among vehicles trapped behind a lane blockage during incident situations, where the ratio of forced gaps is likely to be much higher than the ratios observed by Troutbeck and Kako (1997).

The findings of this study were used to develop the forced and cooperative lane-changing models in SITRAS (Hidas, 2002; Hidas and Behbahanizadeh, 1998). They introduced microscopic traffic network simulation models and presented the details of the lane-changing models including both normal and forced or cooperative lane-changing. In this model depending on the distance to the point of lane-changing, downstream turning movements and lane blockages may trigger either MLC or DLC, where the lane-change must be completed. In order to obey lane-use regulations, MLCs are also performed. In an attempt to obtain speed or queue advantage, DLCs are performed. Initial results showed that without forced lane-changing procedures, incident scenarios could not be simulated realistically. Due to limited perception and recognition of traffic situations and with over simplified decision-making of the Driver Vehicle Operators (DVOs), performance of the model had shown some weaknesses at motorway merge sections. In order for a traffic model to be useful for the intended purpose, it must satisfy both microscopic and macroscopic traffic
management criteria at the same time; autonomous agent techniques proved to be helpful in achieving this goal.

We have seen how lane-change theory and lane-change methods in general have been employed in traffic studies. Next we present the specific studies combining lane-change behaviour with the traffic density estimation. In these studies the two concepts have been combined to study the various aspects of these two traffic variables.

2.5 Traffic Density and Lane-Change Behaviour

Real time inter-lane variables such as traffic density and lane-changing are fundamental to traffic control and management in urban areas. On many occasions they are used as decision variables to determine strategies for traffic control and management in real time. They can also be studied to characterise road traffic congestion for further use in Advanced Traveller Information Systems (ATIS). Considerable effort has been made to estimate and model density along with lane-change behaviour.

Density was estimated considering lane-change as infrequent by Gazis & Knapp (1971). They applied the Kalman filtering technique to estimate vehicle counts on a roadway section, using the noisy measurements of speed and flow by sensors. They considered lane-changes in small numbers as part of system noise, with known statistical properties. It was then questioned; what if one considered extensive lane-changes in the existing model? It was thought that the system equations may have significant error and this error may persist even after implementation of the KF.

To modify the algorithm so that lane-change can be taken into account to improve estimation accuracy, Chang and Gazis (1975), designed a model which used detector data along with aerial data of vehicle counts. They ignored lane-changes between the next nearest lane as it was insignificant. They presented linear and non-linear lane changing models and showed that
consideration of lane-changes, using either of the methods measurably improves the accuracy of the density estimation using discrete KFT. However, their model was applicable to very short section no more than ¾ mile. Also they considered that data from the loop detectors was error free which is a rarity in real world scenarios and the availability of aerial data is also limited due to cost factors.

In the late 1990’s it was thought that information about intra-lane and inter-lane traffic variables such as queuing, lane-changing and lane density can play an important role in finding the magnitude of road traffic stability, which is helpful in further assessing the capability and the performance of traffic infrastructure. Sheu (1999) utilized lane traffic counts detected from loop detectors and proposed a stochastic system to extract real-time information of section-wide inter-lane as well as intra-lane traffic such as lane-changing fractions and lane densities. He undertook an extensive and detailed study and designed an estimator which provided real time information using the EKF. Limited data provided very preliminary results and further data in terms of time-varying lane density and lane-changing fractions would be necessary to adequately evaluate the proposed method. Conclusions were not drawn about lane-blocking incident scenarios and short term queue overflow occurrences, hence limiting the model’s applicability.

Many of the lane-changing models assume accurate measurement of density; however one should accept the range of uncertainties due to the conventional operation of vehicle loop detectors. The small amount of available spatial data for validation limits the applicability of almost all the models. For example, in most cases a lane-change was modelled as a function of flow, but it is likely that excluded factors such as location, time of day, and vehicle mix are significant. Coifman (2003) used lane inflow (the net number of vehicles to enter (or leave) the lane) for estimating traffic
densities. He employed information from sparse vehicle re-identification systems and uses vehicle arrivals at each detector station.

Although the methodology is based on the basic principle of conservation of vehicles, it is the measured travel time that allows for density estimation and upstream offset that allows for the lane inflow measurement. The approach is unique but it assumes that inflow and outflow should be balanced to get good results, i.e., the density estimation would be degraded if they were unbalanced and the estimates would drift apart. Again, the algorithm is condition specific and it fails to work over mixed traffic conditions on the link: free flow at the upstream station and congested at the downstream station. Apparently, when the tail of the queue is within the segment, drivers exhibit different lane-changing patterns. It was also found that as vehicle speed increases, dual loop detectors’ measurement resolution degrades. So, the practical application is rather limited.

In the next section we will discuss the well-known traffic relationship called speed-density relationship (Traffic Flow Theory, 2003). The relationship between speed and density are easily observable in the real world. The speed-density relationship serves as the basis to understand system dynamics in various disciplines. It has been used to model moving objects (or particles) in many scientific areas: pedestrians (Smith & Modeling, 1989), conveyors, network information packages (Gabor, 2002), crowd dynamics (Helbing, 2001), molecular motors and biological systems (Chowdhary et. al, 2005). In this research, we focus on the speed density relationship of vehicular traffic flow to estimate traffic densities or vehicle counts in a multi-lane motorway scenario. This relationship is discussed in detail in the next section.
2.6 Speed-Density Relationship

2.6.1 History

It has been over 80 years since Greenshields’ (1935) seminal paper, “A Study of Traffic Capacity” was published in 1935. It started a new era of transportation science and management by attaching empirically derived curves to a fitted linear model of the speed-density relationship. Being strongly empirical, the efforts to find a perfect theory to explain these particular shapes mathematically never cease, but they always achieved limited success (Wang et al., 2009). From the beginning of this type of research, there was fairly large amount of effort devoted for revising or improving such an over-simplified relationship. In recent decades, researchers developed many theories on speed-density relationship. For specific traffic conditions, there have been a large number of classic speed–density models presented in the traffic management literature. However, propagation of traffic flow is complex with a degree of randomness.

2.6.2 Different Approaches for defining the Speed-Density Relationship

In literature two approaches have been found stating speed-density relationships, namely

- the classical approach (purely mathematical)
- the phenomenological or behavioural approach.

In the classical approach, an analytical expression containing different parameters is proposed and then these parameters are estimated by fitting expressions to traffic data. The models of Greenshield, Greenberg, Underwood and Drake have been derived in this manner. In these models, an interpretation of the traffic variables in terms of traffic flow properties is sought to produce an analytical expression with a phenomenological meaning.
However, the behavioural models are based on assumptions about the driver behaviour with respect to some traffic variables. The models derived from car-following theory belong to this approach. While these relationships provide useful insights into the traffic dynamics problem, they may be restrictive and inadequate for such applications. They describe the relationships among average speed, density and flow under steady-state conditions. We will discuss these approaches in detail.

a. The Classical Approach - A purely mathematical model for the speed density relationship, going back to the early 1930s was proposed by Greenshields’ (1935). In his linear model relating speed and density no assumptions were made as to the behaviour of the drivers. The most interesting aspect of this particular model was that its empirical basis consisted of half a dozen points in one cluster near free-flow speed, and a single observation under congested conditions (Traffic Flow Theory, 2010). The linear relationship comes from connecting the cluster with a single point. It was surprising that such simple analytical methods were used in 1935, but their results (the linear speed-density model) have continued to be so widely accepted for so long. Greenshield (1935) presented the relationship:

\[ \frac{s}{k} = \frac{q}{k} \]

where \( k \) is the traffic density, its maximum value and the maximum speed. For two decades this relationship was considered very useful until in 1959, a logarithmic relationship model was put forward by Greenberg (1959) given as

\[ q = k \ln s \]

where \( q \) is the value of speed at which maximum flow or capacity is reached.
Greenberg’s paper had two data sets. The first data set was derived from speed and headway data on individual vehicles, which “was then separated into speed classes and the average headway was calculated for each speed class”. In other words, the vehicles which appear at one data point (speed class) may not even have been travelling together. Whereas, a density can always be calculated as the reciprocal of average headway, when that average is taken over vehicles (that may not have been travelling together), it is not clear what that density is meant to represent.

Duncan (1976, 1979) proposed a tree-step procedure which includes calculating density from speed and flow data, fitting a speed-density function to that data, and then transforming the speed-density function into a speed-flow function. He further showed that the tree step procedure results in a curve that did not fit the original speed-flow data particularly well. Later in 1979 Duncan showed that minor changes in the speed-density function led to major changes in the speed-flow function.

b. **The Behavioural Approach** – The Second approach, sometimes named as the behavioural one, includes car-following models which gave rise to a new class of the speed-density models tested by Drake *et al.* (1967) defined as

As mentioned in the Chapter 2, pp. 20-26 of Special Report on Traffic Flow Theory (2010), the results of their testing suggest that the speed-density models are not particularly good. Logic says that if the consequences of a set of premises are shown to be false, then at least one of the premises is not valid. It is possible, then, that the car-following models are not valid for motorways. This is not surprising, as they were not developed for this context.
In addition to the above models, kinetic theory of traffic flow was applied to calculate speed-density curves numerically by Prigogine and Herman (1971).

A further higher order model was proposed by Pipes (1967) as:

\[- \quad , \]

which was then generalised as

\[- \quad , \]

where

\[ m, n \] were simply exponents.

After defining and presenting the speed-density relationship, it is important to understand some of the main characteristics of these relationships.

### 2.6.3 The Properties of Speed-Density Relationships

This section reviews the mathematical properties (Castillo & Benitez, 1995) of the speed-density curves. A first group of properties are those that may be regarded as trivial or obvious:

- the values of the speed go from zero to a maximum referred to as free flow speed, \( \), the speed of a vehicle travelling alone;
- the values of traffic density lie between zero (free flow) to a maximum called jam density ;
- the free flow speed is the limit of the desired speed when the spacing tends to infinity, or inversely, when the density approaches zero, thus ;
• vehicles stop at jam density, therefore;
• speed decreases with density, that is, 
  \[ \text{where } (') \text{ stands for the derivative with respect to density.} \]

Another property should be added to account for the fact that as traffic flow becomes lighter, the dependence of speed on density reduces significantly, since the interaction between drivers also reduces to zero. Thus, the following equation must hold.

Speed-density models are also classified as single or multi regime models, such as free flow regime and congested regime. Next we present Table 2.3 & 2.4 which summarise some of the well-known Single and Multi regime Speed-Density Models.

**Table 2.3** Deterministic Single-regime Speed-Density Models (Wang *et al.*, 2009)

<table>
<thead>
<tr>
<th>Deterministic Models</th>
<th>Function</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenshields’ Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenberg Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underwood Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwestern Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drew Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drake’s Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipes-Munjal Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.4 Deterministic Multi-regime Speed-Density Models (Wang et al., 2009)

<table>
<thead>
<tr>
<th>Multi-regime Model</th>
<th>Free Flow Regime</th>
<th>Congested Regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edie</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>Two-regime Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Greenberg</td>
<td></td>
<td>_____</td>
</tr>
</tbody>
</table>

2.6.4 Our Choice of Speed–Density Model

For this research, following Gazis & Liu (2003), we have selected the Drake et al (1967) model. We wish to perform suitable transformations on a speed-density relationship. After researching the literature we found that this is one of the best available relationships and we can perform the required transformation on it. The reason for this choice was simple that among the other curves of this family, Drake et al. (1967) was the one which holds all of the properties (Castillo & Benitez, 1994) mentioned in Section 2.6.3

After discussing the detailed literature review in the area of traffic studies, we now switch on to the second part of the chapter. This part would be focused on the KF and its applications in the field of traffic management.

B. KALMAN FILTERING TECHNIQUE

2.7 Introduction

The KF is an efficient recursive filter that estimates the state of linear dynamic system from a series of noisy measurements. It is used in a wide range of engineering applications from radar (signal processing) to computer vision, and it is an important topic in control theory and control systems engineering. Together with the linear-quadratic regulator (LQR), the KF
solves the linear-quadratic-Gaussian control problem (LQG). They are capable of estimating states that are normally not measurable or that are too complicated to measure directly, using costly sensors. KFs use a priori knowledge about the system and noise acting on the system and they have proven to be robust to parameter changes. The priori knowledge on the process and measurement noise enables the design engineers to build and tune the KF in such a way that optional estimation of system state variables goes together with optimal rejection of measurement noise.

Since the KF was developed, many updated or improved versions have also been introduced to deal with basic underlying assumptions (discussed later). In this research we will discuss discrete KF, EKF and UKF.

2.7.1 General (Discrete) Kalman Filter

The KF proposed by Kalman (1960), is one of the most advanced methods in modern control theory.

Theoretically it is an estimator for what is called the linear-quadratic problem which is the problem of estimating the instantaneous state of a linear dynamic system perturbed by white noise---by using measurements linearly related to the state but corrupted by white noise, the resulting estimator is statistically optimal with respect to any quadratic function of estimation error.

Practically it is certainly one of the greater discoveries in the history of statistical estimation theory. Its most immediate applications have been for the control of complex dynamic systems. To control a system, you must first know what it is doing. For these applications, it is not always possible or desirable to measure every variable that you want to control and the KF provides a means for inferring the information from indirect (and noisy) measurements. For prediction scenarios, it allows to predict the likely future courses of dynamic systems that people are not likely to control.

(Grewal and Andrew, 2001)
More explicitly, KF can be defined as an optimal recursive data processing algorithm. For a better understanding of the definition, the meanings for optimal, recursive and data processing algorithm (Ye, 2007) are described as follows:

- Optimal means that the KF incorporates all information that can be provided to it. The KF uses (a) knowledge of the system and measurement device dynamics, (b) the statistical description of the system noises, measurement errors and uncertainty in the dynamics models, and (c) any available information about initial conditions on the variables of interest. Along with the above information, it processes all measurements to estimate the current value of the variables of interest.

- Recursive means that the KF does not require previous data to be stored and reprocessed every time a new measurement is taken. Only the estimated state from the previous time step and current measurement are required to obtain the estimate of the current state. This is a very important feature for the practicality of filter implementation.

- In many practical applications, the filter acts as a data processing algorithm and is just a computer program in a central processor.

![Dynamic system of KF](Ye, 2007)

**Figure: 2.6** Dynamic system of KF (Ye, 2007)
2.7.2 Kalman Filter – A State Space System

The KF is a dynamic system. A dynamic system can be defined as a system which varies with time. Figure 2.6 explains a brief structure of KF consisting of two parts. In the first part, a process equation (2.1(a)) predicts the new state. The equation uses the information of the previous state. Once a new state is predicted, the measurement can be predicted via a measurement equation (2.1(b)). It can be seen that the dynamic system uses prior knowledge for prediction. As mentioned already, the operation of the KF includes two steps, and the prediction belongs to the first step. The second step is the correction, in which the predicted state is updated based on the difference (innovation/residual) of the true and predicted measurements. The process in the form of equations, a state-transition equation and a measurement equation (Bozic, 1994), can be explained as follows:

\begin{align}
\mathbf{x}_t &= \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{w}_t, \\
\mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t,
\end{align}

(2.1(a))

(2.1(b))

where

- is the state vector at \( t \)th time step from previous time step;
- is the measurement at \( t \)th time step;
- is the state transition model which is applied to the previous state;
- is the observation model which maps the true state space into the observed space;
- is the control input;
- is the process noise;
- is the measurement noise.
We would follow the above definition of a dynamic model throughout the thesis. However, it is important to mention that the above state space model or dynamic model have been defined (in the literature) in another form which is explained as follows:

\begin{align}
(2.2(a)) \\
(2.2(b))
\end{align}

where

- is the state vector at \( k \)th time step from previous time step;
- is the measurement at \( k \)th time step;
- is the state transition model which is applied to the previous state;
- is the control-input model which is applied to the control vector ;
- is the observation model which maps the true state space into the observed space;
- is the control input;
- is the process noise;
- is the measurement noise.

The KF method operates with two phases per time step: the time update phase to “predict” the new state, and the measurement update phase to “correct” the new state. The operation of the KF related to equation (2.2 (a & b)) is shown in Figure 2.7.
The process noise covariance and the measurement noise covariance are represented by $Q$ and $R$ respectively. The process noise (in equation 2.2(a)) and the measurement noise (in equation 2.2(b)) are assumed to be white (zero-mean) and Gaussian: $N(0, R)$. Supplying the initial values of the previous state and the covariance, the KF projects the state and error covariance ahead in the time update step. The second step starts with the computation of the Kalman gain, which is the one that yields Minimum Mean Square Error (MMSE) estimates. Next, the second step is to update the state by incorporating the measurement. The updated state is called *a posteriori* state. Correspondingly, the predicted state is *a priori* state. The final task in the measurement update is to compute *a posteriori* error covariance.

An important thing to note here is that the KF is a Minimum Mean Square Error (MMSE) estimator. If the error in the posterior state estimation is then the KF seeks to minimize . This is
equivalent to minimising the trace of the posterior error covariance; hence by minimising the trace, the optimal Kalman gain can be determined.

2.7.3 Assumptions Behind the Formulation of The KF

The three basic assumptions used for the formulation of KF are:

- Firstly, the system is assumed to be linear, which means that the KF can only be applied to linear problems but not for any type of non-linear problems. In fact this limits the usage of KF to a great extent.

- Secondly, the KF assumes that random variable (RV) such as state and noise have Gaussian distribution. The probability of a Gaussian RV has the shape of a normal curve. This also limits the applicability of the filter significantly.

- Finally, both the process noise and the measurement are white. In other words, noise value is not correlated in time. Thus, the knowledge of the current noise does not contribute towards noise prediction at other time steps.

These assumptions restricted the applicability of KFs and thus the need of more accurate and flexible KF led to the development of other forms of KF such as EKF, UKF, Mixture KF, Adaptive KF, etc. In first of half of this chapter, we discussed the estimation of vehicle densities using KF and in this thesis we present the estimation of vehicle densities considering the lane-change behaviour on motorways using EKF and UKF. On the basis that the underlying speed-density (traffic) relation for this research is non-linear (Drake et al., 1967), the choice of EKF and UKF is justified. We will now explain EKF and UKF in detail.
2.8 Extended Kalman Filter

2.8.1 Definition

As per existing research, the conditional probability density functions (that provide the minimum mean-square error estimate) no longer remain Gaussian when either the system state dynamics or the observation dynamics is non-linear. The optimal non-linear filter will propagate these non-Gaussian functions and evaluate their mean, which may result in a high computational burden. A non-optimal approach to solve the problem, in the frame of linear filters, is the EKF. The EKF implements a KF for a system dynamics that results from the linearization of the original non-linear filter dynamics around the previous state estimates (Ribeiro, 2004).

2.8.2 EKF – A State Space System

Let us assume that our process has a state vector , that the process is governed by the non-linear stochastic difference equation

\[ w_i \]

with a measurement that is

\[ z_i \]

where and represent the process and measurement noise. Here, the non-linear function in the difference equation (2.3) and relates the state at the previous time step to the state at the current time step including as any driving function and the zero-mean process noise. Similarly, a non-linear function in the measurement equation (2.4) relates the state to the measurement.

It is important to note here that after undergoing their respective non-linear transformations, the distributions (or densities in the continuous case)
of the random variables are no longer normal. To estimate a process with non-linear difference and measurement relationships, we consider

\begin{align}
\text{(2.5)} \\
\text{(2.6)}
\end{align}

where

- and \( \mathbf{x} \) are the actual state and measurement vectors,
- and \( \hat{\mathbf{x}} \) are the approximate state and measurement vectors,
- \( \hat{\mathbf{x}} \) is an a posterior estimate of the state at step \( k \),
- \( \mathbf{f} \) is the Jacobian matrix of partial derivatives of \( f \) with respect to \( \mathbf{x} \),
- \( \mathbf{h} \) is the Jacobian matrix of partial derivatives of \( h \) with respect to \( \mathbf{x} \),
- \( \mathbf{u} \) is the Jacobian matrix of partial derivatives of \( u \) with respect to \( \mathbf{x} \),
- \( \mathbf{v} \) is the Jacobian matrix of partial derivatives of \( v \) with respect to \( \mathbf{x} \).
Figure 2.8 Operation of EKF (Welch & Bishop, 2006)

Figure 2.8 shows the complete picture of the operation of EKF. It is to be noted that we have substituted $\mathbf{F}$ for $\mathbf{A}$ to remain consistent with the earlier “super minus” a priori notation and the subscript $k$ is involved with the Jacobians $\mathbf{A}, \mathbf{W}, \mathbf{H}, \mathbf{V}$. Also, an important note here is that Jacobian in the equation for the Kalman gain serves to correctly propagate only the relevant component of the measurement information (see Welch & Bishop, 2006).

2.9 Unscented Kalman Filter

2.9.1 Definition

The UKF addresses the linearization and approximation issues of the EKF. This problem is tackled by using a deterministic sampling approach known as the unscented transformation. The state distribution is
approximated by a Gaussian Random Variable (GRV), but is now specified using a minimal set of carefully chosen sample points (called sigma points) around the mean. *These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true non-linear system, capture the posterior mean and covariance accurately to the 3rd order (Taylor series expansion) for any non-linearity* (Wan & Merwe, 2000). One of the advantages of this technique is that it does not require the explicit calculation of Jacobians, which can be a difficult task for complex functions.

### 2.9.2 The Unscented Transformation

The fundamental part of UKF is unscented transformation (UT) which actually calculates the statistics of a random variable undergoing non-linear transformation. The UT builds on the principle that it is easier to approximate a Gaussian distribution than it is to approximate an arbitrary non-linear function or transformation (Julier & Uhlmann, 1997). Before moving further we first explain the technique of an UT.

Let \( x \) be a \( n \)-dimensional random variable with mean \( \mu \) and covariance \( \Sigma \). Consider,

\[
\begin{align*}
\psi(x) &= \psi(x_1, x_2, \ldots, x_n) \\
\end{align*}
\]

which is the non-linear function through which the variable \( x \) is propagated. To calculate the statistics of \( \psi(x) \), we form a matrix \( \Sigma^{\psi} \) of \( 2n+1 \) sigma points which are selected deterministically so that their sample mean and covariance are the same as those of \( x \), i.e. \( \mu_{\psi} = \mu \) and \( \Sigma_{\psi} = \Sigma \) respectively. The weighted points or sigma points are chosen in the following manner.
where

and provides an extra degree of freedom to fine-tune the higher order moments of the approximation (which acts as a scaling parameter);

= a scaling factor that determines the spread of sigma points around

= the th column of the matrix square root of ;

and

the weight which is associated with th point.

The matrix square root can be calculated using Cholesky factorisation method (Press et al., 1992). Once the above weighted sigma points are selected, they are propagated through the non-linear function to obtain the set of transformed sigma points,

and the mean, covariance and cross covariance for are approximated using a weighted sample mean and covariance of the posterior sigma points as follows (Juleir & Uhlmann, 1997):

1) The mean is calculated by the weighted average of the transformed sigma points,
2) The covariance and cross-covariance are given by the weighted outer product of the sigma points and/or transformed sigma points,

\[ \mathbf{P}_{11} \quad \text{(2.10)} \]

\[ \mathbf{P}_{12} \quad \text{(2.11)} \]

where \( \alpha \) and \( \kappa \) are scalar weights of mean and covariance respectively. It should be noted that all weights should be equal or greater than zero. An UT can be represented as a schematic diagram as shown in Figure 2.9

![Figure 2.9 Unscented transformation of the UKF (Ye, 2007)](image)

It is to be noted that the radius of the sphere that bounds all the sigma points increases with the increase of the dimension of the state space. Under such situations, the sigma points are possible to sample non-local effects,
although they still capture the sample mean and covariance correctly (Merwe et al., 2004). To address this problem, the sigma points can be scaled away or from the mean of the prior distribution by a proper choice of $\lambda$:

1. The distance of the $i$th sigma point from $\mu$ is proportional to $\lambda$;
2. The sigma points are scaled away from $\mu$;
3. The sigma points are scaled towards $\mu$.

To solve this problem Juleir (2002) developed a Scaled SUT which replaces the original set of sigma points with a transformed set by

$$
\begin{align*}
\text{(2.13)}
\end{align*}
$$

where $\lambda$ is a positive scaling parameter which is aimed at minimizing the possible higher order effects. The respective weights of sigma points can be transformed by

$$
\begin{align*}
\text{(2.14)}
\end{align*}
$$

Let

$$
\begin{align*}
\text{(2.15)}
\end{align*}
$$

Then the combined step showing sigma point selection and scaling can be written as:
where

is a parameter to integrate prior knowledge of the distribution of

represents the mean weight,

denotes the covariance weight of the \( \text{th} \) sigma point respectively.

Figure 2.10 shows the operation UKF. It is to be noted that apart from the calculation of sigma points, UKF is similar to the KF for time and measurement update steps.

Figure 2.10 Operation of the UKF

2.10 Kalman Filter and Traffic Management

The KF has been applied to many fields such as robotics (Wen and Durrant-Whyte, 1992), image processing (Durrant-Whyte et al., 1990) and economics (LeRoy and Roger, 1977). The KF also has many applications in transportation. For example, Okutani and Stephanedes (1984) used the KF for forecasting short-term motorway traffic flow. Kessaci et al. (1989) presented the KF to estimate traffic-turning movement ratios based on loop detector data. The KF was used to construct an autonomous driving system employed on public roads (Behringer et al., 1992) and to improve the accuracy and reliability of an Omege-GPS (Global Positioning System) aircraft navigation system (Schlachta and Studenny, 1990).

KF has been applied to many traffic studies, such as the dynamic estimation of traffic density (Gazis and Knapp, 1997; Szeto and Gazis,
1972), motorway origin-destination demand matrices (Hu et al. 2001) and the prediction of traffic volume and travel time (Okutani and Stephanedes, 1984; Gazis and Liu, 2003). However, these papers did not describe how to estimate the covariance matrices of the state and observation noise sequences although it is a key issue in the Kalman filtering technique (KFT).

2.10.1 Applications of General KF in Traffic Management

The KFT has been integrated into macroscopic models for the real-time estimation of traffic states. Payne et al. (1987), Pourmoallem et al. (1997) and Suzuki and Nakatsuji (2002) applied the KFT as a feedback method to update the estimated traffic states on a motorway. Nahi and Trivedi (1973) proposed the relationship between traffic counts and traffic states and updated the states using the KFT. In these studies, observed traffic data was only taken from fixed vehicle detectors, however with long separations between successive detectors, estimation results would probably be less accurate.

In order to estimate traffic states more accurately, Nanthawichit, Nakatsuji and Suzuki (2003) suggested using traffic information from additional sources such as probe vehicles. In their view, with its ability to cover a road network, the probe vehicle technique has great potential in this respect. The probe data was integrated into the observation equation of the KF, in which state equations are represented by a macroscopic traffic flow model. Estimated states were updated with information from both stationary detectors and probe vehicles. However, they assumed that the probe data could be obtained accurately (i.e., there is no consideration on the unreliability of data such as the data error from communication devices, or unavailable data when probe vehicles pass the signal obstructed area), and the effect of the data being biased due to individual willingness of probe drivers was neglected, so the study was not supported by the real world data.
Chu, Oh and Recker, (2005) continued this work within the KFT framework, using an adaptive KF that incorporated two data sources, i.e. point detector data and area-wide probe data. In their proposed method, the traffic system was regarded as a discrete-time dynamic system. They proposed a study which addresses statistics on system model noises derived from both model errors and detector errors, and presented an algorithm to estimate section travel times with on-line estimation of such error statistics. However, they assumed that the traffic inside the respective section was homogenous which is not true in practical sense. Moreover, they too, like Nanthawichit, Nakatsuji and Suzuki, (2003) assumed that the probe data and double loop speed data are error free.

Probe vehicles in the form of transit vehicles were treated as an alternative source of traffic data by Cathey and Dailey (2002) who estimated the time series of vehicle locations and speeds. They also used fixed detectors as another data source, not in a combination but alternatively. They proposed a method where they updated the vehicle position, speed, and acceleration using the KFT. The system equations they used were taken from individual vehicle motion. In addition, apart from average speed, no interest has been paid to the estimation of the other fundamental traffic flow variables of traffic density and volume. Chen and Chien (2001) used the KFT to update the predicted travel time using an autoregressive function. They also used only probe data for predicting future travel time.

2.10.2 Applications of EKF in Traffic Management

EKF was subsequently used to improve the estimation of a number of traffic parameters, as discussed by Isaksson and Gustafsson (1995). Szeto and Gazis (1972) used the EKF in order to improve the count estimates. Szeto and Gazis (1972) used EKF in the early 1970s to improve count estimates. Their algorithm assumes a discrete-time control system which re-
linearises the dynamics of new estimate as they become available. As a consequence of re-linearisation, large initial estimation errors are not allowed to propagate through time and therefore the linearity assumptions are less likely to be violated. In this estimation process, the authors assumed that all the constant parameters were the same for all the sections of road which is not true even under mild motorway conditions. Secondly, the authors assumed that the traffic density outside the exit of the section of road was constant in time which also is not possible in real situations except if there is Stop-Go traffic or a traffic jam. So overall, this algorithm was ‘model-limiting’.

EKF has been used by Sheu (1999) for presenting a stochastic system modelling approach to extract real-time information of section-wide inter-lane as well as intra-lane traffic densities utilising lane traffic counts and occupancy data collected from conventional point detectors. A stochastic system can be regarded as a time varying system since its components change with time (Santina et al., 1994). This system is specified utilising six types of states variables. A discrete-time, non-linear stochastic model is then proposed using defined state variables and a set of assumptions to formulate the specified stochastic system. The authors’ assumption about this stochastic system that each type of state variable is mutually independent of the others was not a realistic one and is very difficult to achieve in practice. Due to the limitations of field data used for the model and the unavailability or limited availability of time-varying lane density data, the evaluation of the proposed method was inadequate.

The EKF approach is used for the real-time estimation of the complete traffic state in motorway stretches by Wang and Papageorghiou (2005). This is a similar study where a specific macroscopic traffic flow model (Papageorghiou et al., 1990) is considered for the traffic state estimator design. However, the developed approach to the traffic state estimation may
employ any validated macroscopic traffic flow model that can be formulated in the respective canonical state-space form. It was assumed that Gaussian white noise is involved in the model and in the measurements. However, a series of other tests demonstrated that the estimator is not confined to this assumption, that is, even if the noise is not strictly Gaussian white, the designed EKF can still work at a very satisfactory level.

Dailey (1999) also used the EKF approach to estimate the population speed parameter. This approach produces a smooth estimate of vehicular speed by pooling information over successive time intervals. One consequence of these approximations, however, is that the resulting estimate is not robust in the sense that it greatly depends on the linearisation (Ye et al., 2006).

For estimating traffic variables over a non-Gaussian platform, Li (2009) extended the Gaussian KF to a non-Gaussian form in which observations have a Reciprocal Inverse Gaussian (RIG) distribution, resulting in a non-Gaussian KF. He investigated the distributions of vehicular speed measurements and showed that RIG distributions provide a suitable model for space-mean speed measurements. He formulated the problem of online estimation and forecasting of vehicular speed as a dynamic generalised linear model (DGLM) with an RIG observational distribution. Due to the absence of real time data of vehicular speeds, estimates of vehicular speed obtained by the speed trap were used. However, the speed estimates provided by a speed trap are not entirely error-free (Hazelton, 2004), but they provide a reference point for evaluation.

2.10.3 Applications of UKF in Traffic Management

UKF is a newly developed technique not employed extensively. However, a few traffic related studies have employed UKF including Ye et al. (2006) and Mihaylova et al. (2007).
Considering the fact that the speed estimation is a non-linear problem, Ye et al. (2006) presented an UKF is used for the non-linear traffic speed estimation problem. They showed that UKF performs better than the EKF, and the traditional g-estimator method has poor performance. However, they did the analysis considering only one lane of the motorways i.e., the right most lane. The right lane was selected because speed variation and the percentages of long vehicles in these lanes tend to be higher, which presents a more critical scenario for speed estimation (Ye et al. 2006). However this may not be true under certain conditions such as congestion or stop-go traffic.

Mihaylova et al. (2007) presented an UKF for the motorway traffic flow estimation. The UKF is developed using traffic and observation models with aggregated variables. The traffic is modelled by a recently developed stochastic compositional traffic model (Boel & Mihaylova, 2004) with interconnected states of neighbour segments. Furthermore, the UKF performance is compared with Particle Filter (PF) and performance is investigated and validated by simulated data and by real traffic data from a Belgian motorway. It has been shown that an advantage of the UKF compared to the PF is that it is less computationally expensive. Both the results with simulated and real traffic data confirm that the UKF provides accurate tracking performance, however, slightly less accurate than the PF.

KFs application is not limited to these traffic variables (speed and traffic flow) but it has been employed to estimate densities by various researchers in the literature. Different versions of KFs including EKF (Wang, 2003), (Gazis & Liu, 2003); Mixture KF (Sun et al., 2004), ordinary KF (Vigos et al., 2008; uses occupancies as input data) have been used to estimate traffic densities over motorways. Our work is based on the model discussed by Gazis & Liu (2003) where the authors have applied the EKF for obtaining density estimates by coupling the detector counts with independent density
estimates, subject to uncorrelated errors. Later, in the thesis, we have also applied the UKF for analysing Markov chain lane-change model discussed in the Chapter 5.

2.11 Summary

In this chapter we have presented the detailed literature review discussing various traffic variables of interest in this research and the filtering techniques especially, the KF. We presented the review in two main sections i.e.

- traffic studies and
- the KFTs

We discussed key topics related to this research study including ILD studies, speed-density relationship theory and lane-change behaviour and defined various forms of KFs including EKF and UKF.

Next, we will present our developed models (methodologies) one by one in the upcoming chapters (3, 4, 5) starting with the very first contribution in the form of the ‘Transformation Model’.
CHAPTER III

A SPEED-DENSITY TRANSFORMATION APPROACH

In this chapter we will first discuss the problem formulation. We will explain the availability of data and how we consider a section of motorway. We then summarise the existing methodology of Gazis and Liu (2003) defining their state space model and underlying assumptions. Then we will discuss shortcomings of the existing methodology. Finally we will introduce a newly developed transformation method and the resulting state space model. We will also discuss how a transformation approach deals with the non-linearity issue of the observation equation within the Gazis & Liu work. Finally, we will use KF to estimate the vehicle densities on multi-segment motorways.

3.1 Background of the Problem

Effective real time control and management of motorway traffic relies on information regarding the number of vehicles in different segments (in other words, the denseness of the section) of a multi-lane motorway. This chapter considers a motorway section with $N$ different segments and $M$ different lanes where each segment of the multi-lane motorway is defined to be a detection zone with one upstream and one downstream embedded ILD loop detector, plus a speed sensor (such as speed cameras), as schematically shown in Figure 3.1. This means a segment is enclosed by the ILDs at both the ends within a particular section which is made up of $N$ such segments. It is to be remembered that ILDs provide traffic data in the form of traffic counts and occupancy per 20-30 seconds. Speed measurements can be
obtained from speed sensors (such as double loop detectors, speed cameras, etc).

In practice, the length of each motorway segment is not long where ILD stations are usually deployed about 500 meters to 1,000 meters apart for most strategic motorway networks. Also see the discussion in Gazis & Liu (2003). In this chapter, we will investigate the estimation of traffic density in each segment of the section via the KF using the speed-density relationship and traffic information from ILDs.

3.2 Summary of Existing Methodologies - Gazis & Liu’s Model

In this subsection, we briefly summarize a typical state space model used in the literature for the estimation of traffic densities; more specifically the model used by Gazis & Liu (2003). The idea behind their model simply states that the estimate can be improved by considering the fact that the counting error for vehicles leaving a given segment is the same as the error for the vehicles entering the very next segment.

Let denote the state vector, where is the number of vehicles in lane of segment in time step . Gazis & Liu (2003) consider a 1-lane 2-segment problem and the state equation is formulated on the basis of a traffic conservation
equation, where the number of vehicles in each lane $j$ of a segment $i$ is equal to the number of vehicles in the previous time step, plus the net gain of vehicles entering and leaving the upstream and downstream loop detectors (see, e.g., Gazis & Liu, 2003):

$$
N_j^i(t) = N_j^i(t-1) + (C_j^i(t) - D_j^i(t))
$$

(3.1)

where $C_j^i(t)$ and $D_j^i(t)$ denote the counts of vehicle entering and leaving the upstream and downstream detectors in lane $j$ of segment $i$ in time step $t$. $\epsilon_j^i(t)$ and $\delta_j^i(t)$ denote the corresponding counting errors.

In traffic management, the phenomenological relationship between traffic speed ($v$) and density ($K$) is given by: (see e.g., Drake et al., 1967)

$$
v = v_0 - \frac{K}{K_m}
$$

(3.2)

where $v_0$ is the free flow speed and $K_m$ is the density corresponding to the maximum flow in a lane of a roadway segment. This relationship is treated as the observation equation in Gazis & Liu (2003). Specifically, by taking into account noise, the observation collected in lane $j$ of segment $i$ is assumed to satisfy the following equation:

$$
N_j^i(t) = v_0 - \frac{K_j^i(t)}{K_m} + \epsilon_j^i(t) + \delta_j^i(t)
$$

(3.3)

where $L_i$ is the length of segment $i$. $\epsilon_j^i(t)$ and $\delta_j^i(t)$ are the corresponding noises. They are assumed to be independent of each other with a common variance (see Gazis & Liu, 2003). If the parameters $v_0$ and $K_m$ are not available
in practice, they can be treated as parameters (Gazis & Liu, 2003). Using the state space model (3.1 – 3.3) which is non-linear in nature, Gazis & Liu applied EKF for estimating densities. Similar state space models were also incorporated in other previous studies such as Wang & Papageorgiou, (2005).

After summarising the Gazis & Liu’s model, we now discuss the shortcomings of this approach which led to the need of developing a more accurate methodology. And finally, we present a newly developed density estimation model which is based on a transformation approach.

3.3 A Transformation Approach
3.3.1 Shortcomings of the Gazis & Liu’s Model

The KF is an efficient recursive filter that estimates the state of the linear dynamic system from a series of noisy measurements. So the equation (3.3) needed to be linearised which in Gazis and Liu (2003) is done by using an extended version of KF i.e. EKF. However, this linearisation can sometimes lead to incorrect results and this is the reason for the occasional failure of the Gazis and Liu (2003) approach. This approximation may lead to inaccurate estimation results as shown later in this chapter. The usual drawback to the implementation of any EKF is the need to heuristically derive optimum noise covariance matrices (Boulter 2000). Another is the trade-off between the need for computing power and the need for filter bandwidth (i.e. update time). Most implementations usually result in such long update times that the filter estimates are of little practical use in real time systems (Boulter, 2000).

In order to overcome this problem, we applied an appropriate transformation to the equation (3.2) and yielded a linear form so as to apply general (discrete) KF to the resulting observation equation (linear in nature). The main advantage of the transformation approach is that it leads to a linear
observation equation which further minimises the risk of approximation. So instead of linearising (or approximating) the model at the estimation stage (or filtering stage), we have dealt with it at an earlier stage. The KF is provided with a linear system and thus the resulting model obtained is robust and accurate as results show later.

### 3.3.2 The Proposed Transformation

Our main aim is to remove non-linearity from the state space model. So, we propose a transformation to transform the observation equation from a non-linear form to a linear one so that the approximation is completely avoided.

Specifically, for the relationship between vehicular speed and density, equation (3.2), we define a transformed speed observation as follows:

\[ y_{\text{trans}} = \frac{\mu}{\rho} \left( \frac{\sigma}{\rho} \right) \]

From equation (3.2), it is straightforward to obtain

\[ y = \frac{\mu}{\rho} \left( \frac{\sigma}{\rho} \right) \]

By taking into account of noise, the transformed observation equation can be written as:

\[ y_{\text{trans}} = y + \epsilon \]

where the noise is assumed to be independent of each other, having a zero mean and a constant variance denoted as .
It is important to mention that this transformation approach can be generalised to any of the speed-density relationship developed so far in the literature. Table 3.1 explains the respective transformation with respect to the list of speed-density relationships presented in Table 2.3 (see Section 2.6.2 (b)).

<table>
<thead>
<tr>
<th>Deterministic Models</th>
<th>Function</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenshields’ Model</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>Greenberg Model</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>Underwood Model</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>Northwestern Model</td>
<td>_ _</td>
<td>_</td>
</tr>
<tr>
<td>Drew Model</td>
<td>_ _</td>
<td>_</td>
</tr>
<tr>
<td>Drake’s Model</td>
<td>_ _</td>
<td>_</td>
</tr>
<tr>
<td>Pipes-Munjal Model</td>
<td>_</td>
<td></td>
</tr>
</tbody>
</table>

Similarly, we can show the transformations for the multi-regime speed-density models.

### 3.3.3 The New State Space Model

After transforming the observation equation, the resultant state space model is linear in nature. First of all, the state equation (equation 3.1) can be extended from two-segment model to the model that can accommodate $N$ roadway segments and can be rewritten in a matrix form:
where each entry of \( a \) is the net gain and each entry of \( b \) is \( b_i \). Following Gazis & Liu (2003), the counting errors and \( \epsilon \) are assumed to be independent of each other, having an identical distribution with zero mean and variance \( \sigma^2 \).

Let \( \Sigma \) denote the covariance matrix of \( b \). The entries of \( \Sigma \) can be worked out as follows. For any \( i \) and \( j \),

\[
\Sigma_{ij} = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } \epsilon, b_i \text{ are independent}, \\ 0 & \text{otherwise}. \end{cases}
\]

In addition, for any lane \( j \), since the number of vehicles leaving a segment \( i \) is equal to the number of vehicles entering the neighbouring segment \( i+1 \), we have

\[
\Sigma_{ij} = 0 \quad \text{for all } j.
\]

All the other entries of \( \Sigma \) are zero due to the independence.

Let \( H \) be a diagonal matrix with each diagonal entry equal to \( H_i \). The observation equation can thus be written in a matrix form:

\[
\begin{align*}
(3.6) \quad y &= Hx + \epsilon, \\
\text{where} \quad y &= \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \\
x &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \\
H &= \begin{bmatrix} H_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_n \end{bmatrix}.
\end{align*}
\]

where \( y \) and \( x \) are two vectors with entries \( y_i \) and \( x_i \) respectively. The covariance matrix of \( y \) is \( \Sigma_y \).

Since both the state equation and observation are linear now, the ordinary KF (or the discrete KF) can be applied. The main advantage of this approach is that the approximation in equation (3.2) caused by linearisation is avoided. It is to be noted here that \( a \) and \( b \) act as parameters (as in the case of Gazis & Liu, 2003).

### 3.4 Density Estimation - Kalman Filter Formulation

Let \( \hat{x} \) and \( \hat{x}_{f} \) denote the one-step forecast of the state
vector and the corresponding covariance matrix in each time step $t$. In addition, let $P_t$ denote the covariance matrix of the estimated state vector in each time step $t$. Applying the KF, the traffic density vector of a multi-lane multi-segment motorway can be estimated recursively as follows (see, e.g. Gazis & Liu, 2003; Simon, 2006):

**Step 1.** Initialization - To initialise the KF we set the state variables and covariance matrix as:

$$X_0, P_0, \text{ and } t=0.$$

**Step 2.** One-step forecast of the state vector (Predict the state vector and the covariance matrix):

$$\hat{X}_t, \hat{P}_t.$$

**Step 3.** Computing the Kalman gain matrix

$$K_t,$$

where $K_t$ can be defined using the transformation approach as shown above.

**Step 4.** Updating the estimate of the state vector and its covariance matrix

$$X_{t+1}, P_{t+1}.$$

**Step 5.** Let $t=t+1$ and return to Step 2.

As explained earlier, the above Kalman filtered estimates of vehicular densities obtained are free from linearisation errors unlike the case of Gazis and Liu (2003).
3.5 Simulation Studies

3.5.1 Traffic Data Simulation

The developed KF based estimator will be tested in this section by the use of microscopic simulation. One major advantage of carrying out simulation studies is that ‘true’ values of vehicular densities are known a priori so that it is straightforward to assess the performance of an estimation method in terms of accuracy (Li, 2009). A self-developed microscopic simulator was used to describe the traffic phenomena on a selected segment of single lane motorway having multiple segments. The two segments under consideration are of 400 and 500 meters length respectively. The simulation scenario with the duration of 2 hours (estimation time step 20s) will be used to execute the simulation investigations. All simulation start with an initial traffic count values setting as $\rho_i$; here the quantity $\rho_i$ is the density corresponding to the maximum flow in the $i$th section; $L_i$ is the section length; and ratio $\alpha$ stands for the denseness ratio of the segment and we assume that the segment is 50% dense as compared to the maximum density value which is taken as 32 (Gazis and Liu, 2003). This value is used to truncate the KF estimates. Following Gazis and Liu (2003), free speeds for the two segments are set equal to 104.76 km/h.

The real time traffic density in each lane was simulated using the state equation (3.5) where counts of vehicles entering and leaving each segment, and , were simulated as Poisson variates with a mean of $\lambda$. The error terms and were simulated as normal variates with zero mean and variance . The speed measurements were simulated using the equation (3.6) where the noise were simulated as normal variates with zero mean and variance . The setting of the parameters $\lambda$, and , varied from experiment to experiment in the simulation below to reflect different
scenarios. The number of input vehicles was chosen as 3, 5, 8, 10 respectively and the values for the parameters were set as discussed later in the coming sections.

### 3.5.2 Assessment Method

We assess the developed method via simulation experiments. In total, 100 experiments were conducted. The evaluation of all scenarios was based on the following Root Mean Square Error (RMSE) criterion:

\[
\text{RMSE} = \sqrt{\frac{\sum_{k} (\hat{N}_k - N_k)^2}{T}}
\]

where \(\hat{N}_k\) is real and \(N_k\) the estimated vehicle counts in a segment. Note that the estimates were produced in every time step \(k\); the produced estimate was then compared using the RMSE criterion to the corresponding simulated value.

### 3.5.3 Addressed Research Issues

In this subsection we list various research issues which are of interest. These issues are carefully selected so that they can explain the real traffic scenarios when applied to a motorway.

(i) **Effect of moderate number of input vehicles** – In the first issue, we investigated the performance of the developed method under the normal traffic condition. We considered a moderate number (8-10 vehicles per time step) of input vehicles flowing through the segment.

(ii) **Effect of light traffic conditions** – Then we discussed if we had less input vehicles, how it would affect the performance of the newly developed model. Here we consider that every 20 seconds, 3-5 vehicles are entering a segment.
(iii) **The impact of the choice of parameters** – We have three main parameters; namely the count error, the speed error and the number of vehicles entering every 20 seconds. So, we discuss how the different values of these parameters affect the model estimation results.

(iv) **Comparison with Existing Methodologies** – Lastly, we compare the developed model with the existing methodologies, more specifically with Gazis & Liu’s model.

### 3.6 Simulation Results

In this section we discuss the above listed research issues one by one considering the simulation settings discussed in section 3.4.1

#### 3.6.1 Effect Of Moderate Number Of Input Vehicles

First, we considered a scenario where a moderate range of vehicles enter the motorway segment. Moderate here stands for any number between 8 to 10 vehicles per 20 second of time step as ILDs provide or update the count values every 20 seconds. It is to be noted that we are assuming that a segment can have at most 32 vehicles per time step (see Gazis & Liu, (2003) for details).

Figure 3.2 displays the ‘actual’ and KF estimated vehicle counts respectively in one run of the simulation experiment, where on average there were 8 vehicles per 20 seconds entering into the segment. Note that the estimates were updated every 20 seconds with the count error (σ) as 2 vehicles per time step and a speed error (τ) as 2 km/hr. In the plots (Figure 3.2), the dotted line (red) represents the observed values and real line (blue) represents the estimated values obtained using the KF. It is clear from the
graphs that the developed transformation method performed well: the estimated traffic density was fairly close to the simulated traffic density.

![Figure 3.2](image)

**Figure 3.2** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=2$, $\tau=2$ and $\lambda=8$.

To assess the developed approach quantitatively, the experiment was repeated 100 times. Table 3.2 displays the corresponding RMSE values for the developed method averaged over the 100 runs of the experiment.

**Table 3.2** Average RMSEs for two roadway segments over 100 simulation runs using the developed method with $\lambda=8$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\tau$ = 1</th>
<th>$\tau$ = 2</th>
<th>$\tau$ = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$ = 1</td>
<td>0.6514 ; 0.8310</td>
<td>0.6324 ; 0.7683</td>
<td>0.6060 ; 0.7343</td>
</tr>
<tr>
<td>$\sigma$ = 2</td>
<td>0.9936 ; 1.2316</td>
<td>0.9823 ; 1.1966</td>
<td>0.9830 ; 1.1761</td>
</tr>
<tr>
<td>$\sigma$ = 3</td>
<td>1.2947 ; 1.5815</td>
<td>1.3054 ; 1.6075</td>
<td>1.2950 ; 1.5608</td>
</tr>
<tr>
<td>$\sigma$ = 5</td>
<td>1.8191 ; 2.2449</td>
<td>1.8488 ; 2.2567</td>
<td>1.8262 ; 2.2416</td>
</tr>
<tr>
<td>$\sigma$ = 8</td>
<td>2.5121 ; 3.0898</td>
<td>2.5530 ; 3.1434</td>
<td>2.5433 ; 3.1068</td>
</tr>
</tbody>
</table>

Two values in every column represent the RMSE values in segment 1 and 2 respectively. Table 3.2 clearly indicates that the error of just 0.6 to 1.5 vehicles occur under acceptable conditions. However, the error may increase.
to 3 if the situation changes further (error in traffic count increases) as can be seen from the Table 3.2

Further, Figure 3.3 represents a similar kind of situation with $\lambda=10$, where every 20 seconds of time step, 10 vehicles enter the motorway section.

![Figure 3.3](image)

**Figure 3.3** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=1$ and $\lambda=10$

In this experiment, the count error was considered to be of 1 vehicle per time step and speed error to be of 1 km/hr. Further Table 3.3 displays the corresponding RMSE values for the developed method averaged over the 100 runs of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 1$</th>
<th>$\tau = 2$</th>
<th>$\tau = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 1$</td>
<td>0.6410 ; 0.7826</td>
<td>0.6148 ; 0.7439</td>
<td>0.6104 ; 0.7309</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>0.9742 ; 1.1959</td>
<td>0.9777 ; 1.1704</td>
<td>0.9663 ; 1.1639</td>
</tr>
<tr>
<td>$\sigma = 3$</td>
<td>1.2684 ; 1.5394</td>
<td>1.2805 ; 1.5597</td>
<td>1.2717 ; 1.5495</td>
</tr>
<tr>
<td>$\sigma = 5$</td>
<td>1.7792 ; 2.1775</td>
<td>1.8221 ; 2.2382</td>
<td>1.8289 ; 2.2420</td>
</tr>
<tr>
<td>$\sigma = 8$</td>
<td>2.4677 ; 3.0314</td>
<td>2.5343 ; 3.0887</td>
<td>3.5348 ; 3.0996</td>
</tr>
</tbody>
</table>
It is evident from the above plots and tables that the developed model produces satisfactory results. For example, for the case ‘$\sigma = 2$, $\tau = 3$, $\lambda = 10$’, we have an error of only 1 vehicle per time step; however, this error increases as we increase the value of the parameters related error terms of the state space model. We can similarly demonstrate for other values of parameters. Hence, we demonstrated that the developed model works well for this set of conditions.

3.6.2 Effect of Light Traffic Conditions

In this subsection we will consider lighter traffic conditions where the number of input vehicles is less than the moderate scenario. We say this number can be anything between 3 to 5 vehicles per 20 seconds of time step. Figure 3.4 displays the actual and KF estimated vehicle counts respectively, where there were 3 vehicles per 20 seconds entering into the section. The count error is considered as 1 vehicle per time step and speed error as 2 km/hr.

![Figure 3.4](image_url)

**Figure 3.4** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed transformation method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma = 1$, $\tau = 2$ and $\lambda = 3$
The experiment was repeated 100 times. Table 3.4 displays the corresponding RMSE values for the developed method for the case where input vehicles are 3 per 20 seconds. Now, considering ‘\(\sigma = 2, \tau = 3, \lambda = 3\)’ from the Table 3.4, we can see the error of approximately 1 vehicle. This is more or less true in almost all other cases as seen from the above plots and table. Hence, the developed methodology does provide good results for light traffic scenario too where less vehicles enter the motorway section.

### Table 3.4 Average RMSEs for two roadway segments over 100 simulation runs using the developed method with \(\lambda=3\)

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>(\tau = 1)</th>
<th>(\tau = 2)</th>
<th>(\tau = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma = 1)</td>
<td>0.7732 ; 1.0101</td>
<td>2.6739 ; 3.3629</td>
<td>2.5701 ; 3.1999</td>
</tr>
<tr>
<td>(\sigma = 2)</td>
<td>1.1497 ; 1.4981</td>
<td>1.0701 ; 1.3357</td>
<td>1.0069 ; 1.2217</td>
</tr>
<tr>
<td>(\sigma = 3)</td>
<td>1.4915 ; 1.9521</td>
<td>1.3765 ; 1.7362</td>
<td>1.3245 ; 1.6160</td>
</tr>
<tr>
<td>(\sigma = 5)</td>
<td>2.0804 ; 2.7015</td>
<td>1.9414 ; 2.4607</td>
<td>1.8726 ; 2.2930</td>
</tr>
<tr>
<td>(\sigma = 8)</td>
<td>2.8170 ; 3.6798</td>
<td>2.6766 ; 3.3546</td>
<td>2.6074 ; 3.2090</td>
</tr>
</tbody>
</table>

3.6.3 The Impact Of the Choice For The Parameters (\(\tau\) and \(\sigma\))

It is of interest to investigate the impact of parameters on the accuracy of the estimation. Tables 3.2, 3.3 & 3.4 display the estimation errors in terms of RMSE when \(\tau\) and \(\sigma\) take values from 1 - 3 and 1 - 8 respectively. It is clearly evident that results in Tables 3.2, 3.3 & 3.4 are more or less accurate in almost all of the different situations as we move from \(\lambda = 3\) to 10. Over the range of different values, the similarity in estimation results suggests the accuracy of the developed methodology.
3.6.4 Comparison

In this subsection we will carry out the direct comparison between the developed methodology (based on the transformation approach) and the classical method, i.e. Gazis & Liu’s Model where vehicular density is estimated using the EKF.

We compared the classical method over the same values of parameters $(\tau, \sigma, \lambda)$ as we used for developed model. Figures 3.5 - 3.6 show the performance of the classical method over a range of traffic conditions.

**Figure 3.5** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis & Liu, (2003) method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=2$, $\tau=2$ and $\lambda=8$

**Figure 3.6** The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis & Liu, (2003) method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=1$ and $\lambda=10$
On comparing Figures 3.2, 3.3 & 3.4 with 3.5, 3.6 & 3.7, it is evident that the method developed by Gazis and Liu, (2003) shows significant variation in the estimation of densities which is due to the approximation done by the EKF during the estimating process.

Figure 3.7 The ‘true’ traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis & Liu, (2003) method for segment 1 (left) and segment 2 (right) in an experiment with $\sigma=1$, $\tau=2$ and $\lambda=3$

Thus, it is clear from the above comparative analysis that the developed method is more accurate and robust in nature as compared to the existing methodologies. Tables 3.5 – 3.7 represent the RMSE values for the Gazis and Liu, (2003) method for the situation where there were 8, 10 and 3 vehicles entering the section per 20 seconds of time step.

Table 3.5 Average RMSEs for two roadway segments over 100 simulation runs using the Gazis & Liu (2003) method with $\lambda=8$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td>0.9700 ; 1.0098</td>
<td>5.0170 ; 4.4732</td>
<td>11.0172 ; 10.3449</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>1.1493 ; 1.3842</td>
<td>3.6167 ; 3.2200</td>
<td>7.6187 ; 7.1864</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>1.4250 ; 1.7086</td>
<td>3.7018 ; 3.6776</td>
<td>6.5778 ; 6.9590</td>
</tr>
<tr>
<td>$\sigma=5$</td>
<td>1.9148 ; 2.4037</td>
<td>3.3630 ; 3.3420</td>
<td>6.0918 ; 6.1027</td>
</tr>
<tr>
<td>$\sigma=8$</td>
<td>2.6238 ; 3.3327</td>
<td>3.6667 ; 4.1538</td>
<td>6.0405 ; 6.3693</td>
</tr>
</tbody>
</table>
Table 3.6 Average RMSEs for two roadway segments over 100 simulation runs using the Gazis & Liu (2003) method with $\lambda=10$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td>0.9273 ; 0.9842</td>
<td>4.8887 ; 4.3649</td>
<td>11.1270 ; 10.0727</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>1.1069 ; 1.3109</td>
<td>3.4771 ; 3.5431</td>
<td>7.4076 ; 8.0536</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>1.3905 ; 1.6450</td>
<td>3.4038 ; 3.1777</td>
<td>6.5289 ; 7.0746</td>
</tr>
<tr>
<td>$\sigma=5$</td>
<td>1.8530 ; 2.3127</td>
<td>3.2032 ; 3.4549</td>
<td>6.2301 ; 6.7294</td>
</tr>
<tr>
<td>$\sigma=8$</td>
<td>2.5397 ; 3.2299</td>
<td>3.6254 ; 3.8918</td>
<td>5.9113 ; 6.6841</td>
</tr>
</tbody>
</table>

Table 3.7 Average RMSEs for two roadway segments over 100 simulation runs using the Gazis & Liu (2003) method with $\lambda=3$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td>1.2804 ; 1.2807</td>
<td>3.8987 ; 4.1185</td>
<td>6.1428 ; 6.5683</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>1.4714 ; 1.7192</td>
<td>3.9889 ; 4.1615</td>
<td>7.5427 ; 8.2706</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>1.7161 ; 2.1688</td>
<td>4.0599 ; 4.5865</td>
<td>7.0270 ; 7.2755</td>
</tr>
<tr>
<td>$\sigma=5$</td>
<td>2.2637 ; 3.0050</td>
<td>3.9614 ; 4.4617</td>
<td>6.2814 ; 6.4598</td>
</tr>
<tr>
<td>$\sigma=8$</td>
<td>3.0616 ; 4.1422</td>
<td>4.0412 ; 4.7097</td>
<td>5.8344 ; 6.7040</td>
</tr>
</tbody>
</table>

On comparing the values from the Tables (3.5 – 3.7) with the values in Tables 3.2 – 3.4, it is clear that the developed method has a better performance than that of the Gazis and Liu, (2003). For instance, with $\lambda = 3$, $\tau = 1$, $\sigma = 3$ (in Table 3.7) the estimation error in the Gazis and Liu method was about 6-7 vehicles and this increases as we move from ($\lambda = 3$) to ($\lambda = 10$) reaching11. However, the developed method is far more accurate with an error of just 1 vehicle in the former and 2-3 in the latter case. This clearly indicates the accuracy of the developed method along with the robustness to perform under different scenarios of road traffic conditions.
This is not surprising as it was pointed out earlier that the classical method is the EKF based recursive system. During our simulation experiments it was not unusual that this estimator completely broke down (i.e. greatly deviated from the trajectory of actual vehicular density) even if all the other conditions were favourable. Moreover, this classical model is applicable only under moderate traffic conditions in the section (meaning that the number of vehicles is moderate), however, the developed model always performs better irrespective of the road traffic conditions.

3.7 Summary

In this chapter we have presented a newly developed transformation approach and investigated the density estimation on a 1-lane 2-segment motorway. We have discussed the existing methodology and its shortcomings in detail and presented the developed transformation model along with simulation results. Simulation results have shown the accuracy of the developed methodology as compared to the existing method (Gazis & Liu, 2003). It has also been shown that the developed method is applicable over a wide range of traffic conditions.

Towards the end, whilst undertaking comparisons, we have shown that the reason behind the occasional failure of Gazis and Liu (2003) was the use of non-linear measurement equations and the linearisation during the estimation process using the EKF. The developed method is based on a transformation which leads to a linear state space model. When applying the KF, the approximation due to the linearization is avoided during estimation.

It is to be noted that we have not yet discussed drivers’ lane-changing behaviour in this research. However, in the coming chapters we will investigate the effect of drivers’ lane change behaviour on the density estimation.
CHAPTER IV

MARKOV CHAIN LANE CHANGE MODEL

In this chapter we will present a Markov chain lane change model. To start with, we will first discuss the problem of vehicular lane change on multi-lane motorways. We will explain how existing methods have estimated densities considering lane change behaviour. We will then discuss shortcomings of these existing methodologies and summarise Markov chain theory in general. Then, we introduce a newly developed Markov chain model for lane change and present an approach to estimate densities considering lane change behaviour. Finally, we will use EKF to estimate the vehicular densities on multi-segment multi-lane motorways.

4.1 Lane Change Problem

Lane change, as a real time inter-lane variable, is fundamental to traffic control and management. It has long been recognised that lane change manoeuvres can influence the underlying relationships of traffic flow. Moreover, these relationships may even be disrupted if lane change manoeuvres are not accounted in the respective traffic model. In the past, considerable effort has been made by some researchers to estimate and model density along with lane-change behaviour. However, most of these models are not representative or applicable to real traffic scenario due to their own various limitations which have been discussed in this section.

Traffic density was estimated considering lane change as infrequent by Gazis & Knapp (1971). They considered lane-changes in small amounts as a part of system noise, with known statistical properties. However, their method cannot withhold extensive lane changing and consequently adds
significant error to the system equations. To modify the above algorithm and to allow extensive lane-change movements, Chang and Gazis (1975) designed a model which uses detector data along with aerial data of vehicle counts. They presented linear and non-linear lane changing models and showed that consideration of lane-changing, using either of the methods, improved (measurably) the accuracy of the density estimation. They used discrete (general) KFT to estimate traffic densities. However, their model was applicable to a short segment length as long as ¾ mile. Their model was having limited applicability since they considered that the data (from loop detectors) is error free. Also another drawback for the usage of this model is the aerial data which is both costly and limited in availability.

Then, in the late 1990’s Sheu (1999) used lane traffic counts, detected from loop detectors, and proposed a stochastic system utilising six types of state variables. Using these defined variables and set of assumptions, a discrete-time, non-linear stochastic model was proposed. The model explicitly suggested the feasibility of estimating real-time section-wide inter-lane and intra-lane traffic variables such as lane-changing fractions and lane density. No doubt, the proposed methodology was promising, but must be regarded as preliminary. This is due to the limited field data. More data in terms of time-varying lane density and lane-changing fractions was required to adequately evaluate the proposed method. Further application of the techniques such as truncation and normalisation may not present the actual traffic problem and would only yield an approximate picture of the underlying scenario. Also, issues such as lane-blocking incidents and short term queue overflow occurrences have not been discussed.

To study queue occupancies, Coifman (2003), assumed that vehicular inflow and outflow should be balanced to get good results, i.e., the density estimation would be degraded if they are unbalanced and the estimates will drift apart. The algorithm was condition specific and it fails to work over
mixed traffic conditions on the segment: free flow at the upstream station and congested at the downstream station. Apparently, when the tail of the queue is within the link, drivers exhibit different lane-changing patterns. It was also found that as vehicle speed increased, the dual loop detector’s measurement resolution was degraded.

Linking speed and density together, Gazis & Liu (2003) used a well-known speed-density relationship (Drake et al. 1967) to estimate traffic densities ignoring lane-change behaviour. Along with no lane change, this method also had to undergo linearisation and approximation which led to the occasional failure of the methodology. Hence, this approach is applicable only if the lane change is not significant when estimating densities over multi-lane motorways.

In our research we will explicitly consider the drivers’ lane-change behaviour and it will be shown (later in this chapter) that inclusion of lane-change behaviour improves the results significantly.

Many of the lane-change models assumed accurate measurement of density; however one should accept the range of uncertainties due to the conventional operation of vehicle loop detectors. The small amount of available spatial data for validation limits the applicability of almost all the models. Therefore there is a need to deal with the problem of estimating densities considering accurate lane-change information which is derived from real time traffic information. In this chapter we present a novel approach to estimate densities using a Markov chain based driver’s lane-change behaviour where different states of the Markov chain are represented as different lanes of the motorway. Before explaining our developed model further, a brief discussion of the Markov chain theory is outlined as follows.
4.2 Markov Chain Theory

Andrew Markov, a Russian Mathematician, proposed a stochastic process model called the Markov process to describe a time varying random phenomenon for which the Markov property holds. To understand a Markov process $X_t$ we need to understand the Markov property (or memorylessness) which states that given the current state $X_n=x$, the conditional probability distribution of $X_{n+1}$ for the system at the next step $n+1$ depends only on the current state of the system, and not on the state of the system at previous steps, i.e.

$$P(X_{n+1}|X_1, X_2, \ldots, X_n) = P(X_{n+1}|X_n),$$

where $X_n$ denote the state vector at time $n$.

A Markov chain is a Markov process which has a discrete (finite or countable) state-space. In other words, a process which is randomly discrete (a discrete random system which is in a certain state at each time ‘step’ and the state changes randomly between steps) with the property that the next state depends only on the current state is termed as Markov chain. It is used as a mathematical tool for statistical modelling in modern applied mathematics, operational research and information sciences, etc.

We use the following example to explain the concept of a Markov chain. Suppose we have a frog jumping among several lily-pads, where the frog's memory is short enough that it is unable to remember what lily-pad it was last on, and so its next jump only depend upon where it is now.

Further, the steps can refer to time (such as in the frog and lily-pad example), or to physical distance or any other discrete measurement. In practice, these steps are just the integers or natural numbers, where the underlying random process maps these steps to states.

After summarising Markov theory, we are now in a position to introduce our Markov chain model for lane-change behaviour.
4.3 A Markov Chain Model For Lane Change

In this section, we present a novel approach of modelling the manoeuvre of a vehicle in a motorway segment as a Markov chain process where each lane is characterised as a state of the Markov chain.

To apply a stochastic process for estimating traffic density, a time period with a relatively uniform demand is considered in the analysis. The following tasks were undertaken: (i) define system states and state variables considering the probability of lane-change; (ii) estimate traffic density for each lane considering the lane-change effect using the KFT.

![Figure 4.1 A schematic plot showing an N-segment and M-lane roadway section](image)

Now consider a single motorway segment with $M$ lanes. We let denote the state vector at time $t$. Here describes how the states of a given motorway segment changes every time unit. includes entries, where is the number of vehicles in lane $i$ in the step $t$, i.e.

We assume that each vehicle in the motorway segment has a certain probability to stay in the current state (lane), or to change from one state (lane) to another. We also assume that the traffic flow is stable so that no subsequent lane-changes are possible.
The changes in the state of the system are termed as transitions and the corresponding probabilities are termed as transition probabilities. A Markov chain can be completely characterised by a set of all the states and transition probabilities. Now let \( p \) denote the transition probability that a vehicle moves from state (lane) \( j \) to state (lane) \( k \).

Figure 4.2 represents a basic structure of a Markov chain where states (in our case lanes) are presented by circles, and arrows represent various transitions with probabilities of changing from one state \( i \) to another state \( j \) or probabilities remaining in the same stage \( i \).

![Figure 4.2 Markov Chain state change diagram](image)

4.4 A General State Space Model

In this section we develop a general state space model for traffic flow in \( N \) motorway segments. Let \( \mathbf{x} \) denote the state vector, where \( x_j \) is the number of vehicles in lane \( j \) of segment \( i \) in time step \( t \). We first focus on the scenario of single motorway segment. With the lane-change behaviour being taken into consideration, the numbers of vehicles entering (leaving) a lane from (to) the neighbouring lane(s) must also be included into the state equation. For
ease of exposition, we consider a 3-lane problem first. For the middle lane \( j \) of segment \( i \), the state equation can be written as:

\[
\text{Likewise, for the outermost lane } j=1, \text{ the state equation is:}
\]

\[
\text{where } \text{ and } \text{ denote the counts of vehicle entering and leaving the upstream and downstream detectors in lane } j \text{ of segment } i \text{ in time step } t. \\
\text{ and } \text{ denote the corresponding counting errors. For the innermost lane, we can have a similar equation.}
\]

We define the system matrix for the \( n \)th motorway segment via the Markov transition probabilities:

\[
\text{where to simplify notation, the subscript for motorway segment } n \text{ is suppressed in the above definition. It is also assumed that each time step is short so that moving is feasible only between neighbouring lanes.}
\]

The above state equation can thus be written in matrix form as follows:

\[
\text{,} \\
\text{(4.2)}
\]
where, each entry of \( \alpha \) is the net gain and each entry of \( \beta \) is _____.

We use the same observation equation (3.3) in (chapter 3) based on the speed-density relationship, which can be stated as follows.

\[
\text{\begin{array}{ccc}
\end{array}}
\]

\( \text{\begin{array}{ccc}
\end{array}} \)

(4.3)

where each entry of \( \gamma \) denote the vector of the nonlinear functions and is _____.

\subsection{4.5 Density Estimation - Kalman Filter Formulation}

The observation equation (4.3) obtained above is non-linear. Since the KF can be applied to a linear system only, the above observation needs to be linearised. Specifically, let \( \hat{\mathbf{x}}_t \) denote the estimate of the state vector in each time step \( t \). Then the Jacobian evaluated at the current estimate is:

\[
\text{\begin{array}{ccc}
\end{array}}
\]

Through algebra it can be shown that the matrix \( H \) is diagonal with each diagonal entry equal to _______. The entries of \( H \) are actually taken from the speed-density relationship derived by Drake \textit{et al.} (1967) as discussed in Chapter 3 (see section 3.2). The observation equation can thus be approximated as:

\[
\text{\begin{array}{ccc}
\end{array}}
\]

(4.4)
The estimate of traffic densities can be updated on the basis of the state space model (4.2) and (4.4). During each time step, the non-linear system is linearised about the current estimate of the state vector (as outlined in equation 4.4). This estimation procedure is explained as follows.

Let us denote the predicted estimate of the state vector as \( \hat{x}_k \), the covariance matrix of the state vector as \( P_k \) and \( P_k \) for \( t=0 \). Using the equation (4.2) & (4.4) and applying the EKF methodology, the traffic density vector of a multi-lane multi-segment motorway can be estimated as follows:

1) To initialize the KF we set the state variables and covariance matrix as:

\[
\hat{x}_0, P_0, \text{ and } t=0.
\]

(4.5)

2) Predict the state vector and the covariance matrix

\[
\hat{x}_{k+1}, \quad P_{k+1}. \quad (4.6)
\]

\[
. \quad (4.7)
\]

where ‘ ‘ is defined in 4.2 and \( R \) is the covariance of the counting error for the quantity \( y \) and defined in Chapter 3 (see section 3.2). Here \( \Phi \) acts as the transition matrix containing transition probabilities of vehicles changing lanes from one lane to other.

3) Compute the Kalman gain Matrix

\[
K_k, \quad (4.8)
\]
where \( w \) can be defined as above and \( \Gamma \) is the covariance of \( \epsilon \) which is the error associated with the speed estimation and \( \sigma_{\varepsilon} \) is defined in Chapter 3 (see section 3.4).

4) Update the state vector and its covariance matrix

\[
\begin{align*}
\mathbf{x}_t & = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{w}_t, \\
\mathbf{P}_t & = \mathbf{F}_t \mathbf{P}_{t-1} \mathbf{F}_t^T + \mathbf{Q}_t.
\end{align*}
\]

where \( \mathbf{u}_t \) are the independently determined observed traffic speeds defined by the observation equation (4.4) and

\[
\begin{align*}
\mathbf{Q}_t & = \begin{bmatrix}
\sigma_{\varepsilon}^2 & 0 \\
0 & \sigma_{\varepsilon}^2
\end{bmatrix}, \\
\mathbf{B}_t & = \begin{bmatrix}
0 \\
0
\end{bmatrix}.
\end{align*}
\]

5) Let \( t = t + 1 \) and return to Step 2.

In this way, we estimate traffic densities recursively. After developing the model and KF equations, we are ready for the practical application of the developed model using simulation experiments which are presented in the next section.

4.6 Simulation Description

4.6.1 Traffic Data Simulation

We used a self-developed microscopic simulator to simulate traffic densities in a 3-lane motorway with a single segment. The length of the motorway segment was set equal to 400 meters. We considered a simulation scenario where the time period of interest was 2 hours. It was assumed that the upstream and downstream loop stations deployed at the two extreme ends of the segment measure traffic every 20 seconds. Hence,
the corresponding estimation time step was 20s. The system matrix was set to be:

$$
A
$$

where due to the assumption of stable traffic condition, we assumed that no lane-change is possible from the innermost lane to the outermost lane or vice versa. Moreover, to change subsequent lanes within 20 seconds time step is anyways difficult in real terms. The elements of the matrix ‘A’ represent the probability of the vehicles changing lanes from one to other. Clearly the middle lane is the most affected as compared to side way lanes, i.e. the high speed lane and heavy vehicle lane. The real time traffic density in each lane was simulated using the state equation (4.2). All the other experiment settings were kept to be the same as in the Chapter 3 (see Section 3.4).

Each experiment was repeated 100 times. The evaluation of each method was based on the RMSE between the ‘true’ and estimated vehicle counts. Also, we have similar research issues to tackle as in the previous chapter, so we will discuss them next with the help of plots and RMSE tables.

4.6.2 Addressed Research Issues

In this subsection we list various research issues which are of interest. These issues are carefully selected so that they can explain the real traffic scenarios when applied to a motorway.

(i) Effect of Moderate Number of Entering/Leaving Vehicles – In the first issue, we investigated the performance of the developed method under the moderate traffic condition. We consider that a moderate level
(8-10 vehicles per 20 seconds step) of traffic is flowing throughout the experiment.

(ii) **Effect of Light Traffic Conditions** – Then we discussed what happens if fewer vehicles enter the section, how it would affect the performance of the newly developed model in this chapter. Here we considered that every 20 sec of time step, 3-5 vehicles are entering the section.

(iii) **The Impact of the Choice of Parameters** – We have three main parameters namely the count error ($\sigma$), the speed error ($\tau$) and the number of vehicles ($\lambda$) entering per 20 sec of time. So, we discussed how the different values of these parameters affected the model estimation results.

(iv) **Comparison** – To show the accuracy, we compared the developed model with the existing methodologies.

(v) **Robustness** – Lastly, we discussed the issue of robustness of the developed model and show some of the numerical evidence.

### 4.7 Simulation Results

In this section we discussed the above mentioned research issues one by one considering the simulation settings discussed in Section 4.6.

#### 4.7.1 Effect of Moderate Number of Entering/Leaving Vehicles

As in the previous chapter, we first considered a scenario where a moderate number of vehicles enter the motorway section per time step of 20 seconds. We follow the same definition of moderate as in the previous chapter. Figure 4.3 displays the actual and KF estimated vehicle counts
respectively in one experiment run, where there were 10 vehicles per 20 seconds entering into the section.

To initiate the representation of the simulation results for the above model, we first consider the situation where during every 20 seconds there were 10 vehicles entering the motorway segment in each of the three lanes. When compared to 32 (maximum allowed), 10 is quite a generous number (moderate level) to test the model application. The count error was kept at 1 vehicle per time step and speed error as 1 km/hr. Figure 4.3 displays the ‘actual’ (broken red line) and the KF (blue line) estimated vehicle counts respectively in one run of the experiment. The three plots in Figure 4.3 shows the estimated densities for three lanes respectively where the middle-lane density estimation is represented by the middle plot and first and third lanes have been represented by the left and right plots, respectively.

![Figure 4.3 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single segment freeway with $\sigma=1$, $\tau=1$ and $\lambda=10$](image)

Looking at the above plots it is clear that the estimated traffic densities were close to the simulated traffic densities. Hence the developed model performed well if a sensible number of vehicles enter into the motorway segment.

In the experiments, different values of the parameters ($\sigma$, $\tau$, and $\lambda$) have been used. To assess the developed approach quantitatively, each
experiment was repeated 100 times. Values of RMSEs from simulation experiments are displayed in Table 4.1.

Table 4.1 suggests that the error is very low. For most of the different combinations of the parameters, it is about 1 vehicle.

Table 4.1 Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=10$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\tau=1$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>$\tau=2$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>$\tau=3$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83</td>
<td>0.77</td>
<td>0.88</td>
<td>1</td>
<td>0.76</td>
<td>0.72</td>
<td>0.78</td>
<td>1</td>
<td>1.66</td>
<td>1.40</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.79</td>
<td>0.69</td>
<td>0.82</td>
<td>2</td>
<td>0.75</td>
<td>0.69</td>
<td>0.81</td>
<td>2</td>
<td>1.86</td>
<td>1.74</td>
<td>1.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.79</td>
<td>0.69</td>
<td>0.85</td>
<td>3</td>
<td>0.85</td>
<td>0.85</td>
<td>0.87</td>
<td>3</td>
<td>2.60</td>
<td>3.28</td>
<td>2.90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.7.2 Effect of Light Traffic Conditions

In this case, we investigate the scenario in which there are only a few vehicles per time step entering the segment. In particular, we assumed that , i.e., only three vehicles entered the section with a count error of 1 vehicle and a speed error of 1 km/hr every 20 seconds. The simulation results are presented in Figure 4.4.

Figure 4.4 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single segment freeway with $\sigma=1$, $\tau=1$ and $\lambda=3$
To assess the developed approach quantitatively, the experiment was repeated 100 times. Table 4.2 displays the corresponding RMSE values for the developed method averaged over 100 runs of the experiment. It is clearly evident from the plots and table that the developed model performed well with an average error of less than 1 vehicle per 20 seconds.

**Table 4.2** Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=3$

<table>
<thead>
<tr>
<th></th>
<th>$\tau=1$</th>
<th></th>
<th>$\tau=2$</th>
<th></th>
<th>$\tau=3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Lane 3</td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Lane 3</td>
</tr>
<tr>
<td>1</td>
<td>0.85</td>
<td>0.75</td>
<td>0.89</td>
<td>1.00</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0.70</td>
<td>0.68</td>
<td>0.78</td>
<td>0.85</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.76</td>
<td>0.85</td>
<td>0.80</td>
<td>0.69</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Next, we assume five vehicles ($\lambda=5$) entering the motorway segment per 20 seconds in a three lane motorway. The following plots show this scenario for the developed model. We kept the rest of the settings as in previous case (where $\lambda=3$). It is evident from the plots that the estimated densities are very close to the ‘real’ values and all the three lanes indicate the similar results.

**Figure 4.5** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method for three lane single segment freeway with $\sigma=1$, $\tau=1$ and $\lambda=5$
Again, to assess the developed the approach quantitatively, the experiment was repeated 100 times. Table 4.3 shows the respective RMSE results for the developed methodology. It can be seen that the average error in this case is also less than 1 vehicle per time step.

<table>
<thead>
<tr>
<th>Table 4.3</th>
<th>Average RMSEs for three lane motorway section over 100 simulation runs using the developed method with $\lambda=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau=1$</td>
<td>$\sigma$ Lane 1 Lane 2 Lane 3 $\sigma$ Lane 1 Lane 2 Lane 3 $\sigma$ Lane 1 Lane 2 Lane 3</td>
</tr>
<tr>
<td>1</td>
<td>0.81 0.80 0.84</td>
</tr>
<tr>
<td>2</td>
<td>0.84 0.70 0.83</td>
</tr>
<tr>
<td>3</td>
<td>0.86 0.83 0.86</td>
</tr>
</tbody>
</table>

4.7.3 The Impact of the Choice of Parameters

It is of interest to investigate the impact of the parameters on the accuracy of estimation. Tables 4.1, 4.2 & 4.3 display the estimation errors in terms of RMSE with values of different parameters. It is clearly evident that results in Tables 4.1 - 4.3 are promising and do not change much if we keep the error under control. Not surprisingly, the estimation error was increased when the quality of observations deteriorated ($\tau=3$).

4.7.4 Comparison with Existing Research

Next, we present the comparative results between the developed method and the existing methodology (Gazis and Liu, 2003). We draw the comparison by the means of plots and tables with RMSE value. Theoretically, the main difference between the two methodologies is the inclusion of lane-change behaviour in the developed method. The Figures 4.6 & 4.7 show that the estimated densities differ from the ‘actual’ values when lane-changes were ignored as in the case of Gazis & Liu (2003).
On comparing the plots for the developed model for similar settings of parameters (see Figures 4.4 & 4.5), it is evident from the Figures 4.6 & 4.7 that the estimation process for the existing method (Gazis & Liu, 2003) occasionally lose the estimation path if we ignore lane-changes over multi-lane motorways. For instance, the performance of Gazis & Liu’s method was poor for the middle lane for the reason already explained in the previous section.

**Figure 4.6** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis and Liu, (2003) method for three lane single segment freeway with $\sigma=1$, $\tau=1$ and $\lambda=10$.

**Figure 4.7** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the Gazis and Liu, (2003) for three lane single segment freeway with $\sigma=1$, $\tau=1$ and $\lambda=5$.

The plots (Figure 4.4 & 4.5) have shown the importance of lane-change inclusion to the traffic flow model. Tables 4.4 & 4.5 display the RMSE values for the 100 experiments which show poor performance of the existing method under the same settings as used for the developed methodology (see Figure 4.3 & Table 4.1).
For instance, for $\sigma = 1$, $\tau = 2$, $\lambda = 10$, the developed model’s RMSE value is 0.6929 (Table 4.1), however for the existing method it is coming out to be 10.790 (Table 4.4). In other words, the developed method resulted in an error of only one vehicle, however for the existing methodology the corresponding error is 11 vehicles per time step of 20 seconds.

**Table 4.4** Average RMSEs for three lane motorway section over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=10$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.74</td>
<td>10.10</td>
<td>2.12</td>
<td>1</td>
<td>1.56</td>
<td>7.93</td>
<td>2.87</td>
<td>1</td>
<td>3.67</td>
<td>3.46</td>
<td>4.23</td>
</tr>
<tr>
<td>2</td>
<td>1.58</td>
<td>9.48</td>
<td>2.47</td>
<td>2</td>
<td>4.12</td>
<td>10.79</td>
<td>2.36</td>
<td>2</td>
<td>3.23</td>
<td>3.95</td>
<td>3.13</td>
</tr>
<tr>
<td>3</td>
<td>1.90</td>
<td>12.45</td>
<td>1.78</td>
<td>3</td>
<td>2.02</td>
<td>9.17</td>
<td>1.76</td>
<td>3</td>
<td>3.43</td>
<td>7.99</td>
<td>4.37</td>
</tr>
</tbody>
</table>

**Table 4.5** Average RMSEs for three lane motorway section over 100 simulation runs using the Gazis and Liu, (2003) method with $\lambda=5$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
<th>$\sigma$</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.33</td>
<td>4.97</td>
<td>1.56</td>
<td>1</td>
<td>4.03</td>
<td>6.15</td>
<td>2.76</td>
<td>1</td>
<td>1.58</td>
<td>7.01</td>
<td>3.03</td>
</tr>
<tr>
<td>2</td>
<td>1.79</td>
<td>7.98</td>
<td>2.39</td>
<td>2</td>
<td>3.99</td>
<td>8.12</td>
<td>1.90</td>
<td>2</td>
<td>4.07</td>
<td>6.32</td>
<td>3.37</td>
</tr>
<tr>
<td>3</td>
<td>1.42</td>
<td>11.88</td>
<td>1.88</td>
<td>3</td>
<td>2.64</td>
<td>5.96</td>
<td>2.68</td>
<td>3</td>
<td>1.55</td>
<td>7.43</td>
<td>2.09</td>
</tr>
</tbody>
</table>

Values from the table shows:

- an error as high as 11% can occur if lane-change behaviour is ignored.
- the middle lane is the most affected lane as expected since it has the tendency to allow lane-change from both the neighbouring lanes.

Thus it has been shown that the developed model does, not only perform well in general traffic conditions but in specific situations also. However, for
a model to be more appropriately applicable it is important that it should be robust which is discussed in the next section.

4.8 Robustness

This section deals with an important issue, robustness. In this section, we will present the motivations behind this key issue followed by explaining the simulation set-up. We will then present the results to show that the developed model is robust.

4.8.1 Motivations

The lane-change behaviour has been accounted by the transition matrix having transition probabilities for the vehicles changing from one lane to other. For this part of the research we used fixed values of the transition probabilities (lane-change probabilities). However, if the matrix we used is subject to errors, then it might not represent the actual traffic scenario. In other words, if the values in the matrix are subject to statistical errors, then we might get different results. For example, suppose entry of the transition matrix $A$ is 0.783. What if this entry was mistakenly set as 0.789 or even 0.791? We aim to check whether the developed model can handle these kinds of errors. If yes, then to what extent? To address this issue, we carried out further experiments changing some simulation settings which are explained in the next sub-section.

4.8.2 Simulation Setting Up

To investigate this situation, we simulated ‘true’ densities using transition matrix $A$ in the usual way. We then thought of introducing random noise to the transition probabilities in the transition matrix $A$. We introduced random error of up to 20% of lane change behaviour as follows:
We added noise (up to 20%) in the lane-change behaviour and the respective matrix thus obtained was . In other words, we can say that the lane-change transitions have some error (up to an acceptable level). After adding this noise, the next step is to discover how this change may impact the respective results.

We estimated densities (via EKF) using . We discussed that even if the ‘true’ transition probabilities are \( A \) but due to estimation errors, we obtained , the estimate of traffic densities are still very close to the ‘true’ values.

### 4.8.3 Results

RMSE values have been employed to discuss the robustness issue. We added noise (as stated above) and carried out the estimation procedure using non-linear state space model and EKF. As an example, for \( \lambda = 5 \), \( \tau = 1 \) and \( \sigma = 2 \) (here speed error was 1 km/hr, count error from ILDs was 2 and 5 vehicles are entering the chosen segment of motorway), we found the respective RMSEs for three-lanes in single-segment problem are as follows: 0.86, 0.76 and 0.91 Further, for the parameter setting as \( \lambda = 10 \), \( \tau = 1 \) and \( \sigma = 3 \), the RMSE values are 1.52, 1.49 and 1.53. On comparing these values with those in Tables 4.1 & 4.3, we found that the results after adding robustness noise are promising and variations (up to 4%) are acceptable if we compare the overall performance of the two approaches. Thus, we can say that the developed model is robust even under different traffic scenarios.

### 4.9 Summary

In this chapter, we have presented the second model developed in this research which introduces the Markov chain for modelling lane-change behaviour and for the vehicle densities. We first summarised the drawbacks and limitations of the existing methodologies and then presented the Markov
chain lane-change model followed by the state space equations. We then used the KFT and applied the EKF for estimating densities. In this model, we assumed that the traffic is stable inside the segment and thus lane-change probabilities are not changing dynamically but rather they are fixed.

It should be noted that fixed lane-change behaviour may not represent real traffic scenarios since the traffic is usually not stationary and drivers may behave differently in different time periods. To overcome this limitation, we will propose further extensions to this model where we actually estimate the lane-change probabilities using the discrete choice theory. This model is the subject of the next chapter.
CHAPTER V

DISCRETE CHOICE BASED LANE-CHANGE MODEL

In this chapter, we are presenting our third and final model developed in this research study. This model is based on discrete choice theory and builds up on the model presented in Chapter 4. In fact, this discrete choice based model is an explicit extension of the Markov chain based lane-change model. In this chapter, first of all, we will investigate the shortcomings of the model two (Markov chain based lane-change model) followed by a brief discussion about discrete choice theory. Then we will present the idea of discrete choice based lane-change behavior and discuss the transition matrix in this case. Finally, we will present the estimation scheme for the vehicle densities on multi-segment multi-lane motorways using two filters namely

- EKF;
- UKF.

Then we present the simulation studies for these two estimation approaches and close the chapter by presenting a short summary. We will show that the computational cost for updating the estimate of traffic densities is kept at a very low level so that online applications are feasible in practice. Consequently, the traffic densities can be monitored and the relevant information can be fed into the traffic management system of interest.

5.1 Shortcomings of the Markov chain model in Chapter 4

In Chapter 4, we used the Markov chain model for traffic modelling by considering the driver’s lane-change behavior in the existing model of Gazis & Liu (2003). However, we assumed that this lane change behavior is fixed
in nature. In other words, we assumed that the traffic condition is stable and that drivers would switch in-between the different lanes with fixed probabilities, which may not be true in a real traffic situation.

The contribution of this chapter lies in this gap itself, that instead of assuming the stable traffic condition or the driver’s fixed lane-change behavior, we are assuming traffic conditions may not be stable and rather dynamic. So, we are actually estimating driver’s choice behavior in terms of lane-change probabilities using a discrete choice based relationship combining various traffic variables. This model is based on previous work done by Chang & Kao (1991), who developed an interesting relationship relating various traffic variables such as speed, headway and density to estimate vehicular densities. In the next section, we will present a brief introduction on discrete choice models and logit models.

5.2 Discrete Choice Analysis

5.2.1 Introduction

Discrete choice models were introduced in 1920s by Thurstone (1927) and since then they have been studied, modified and extended. The idea of discrete choice models is to describe the decision maker’s choices among all possible available alternatives. The set of these alternatives is called the choice set and should exhibit the following three characteristics:

- Mutual exclusiveness
- Exhaustiveness
- Finiteness

Let us suppose that a decision maker, labelled \( n \), faces a choice among \( J \) alternatives. Let the utility that decision maker \( n \) obtains from alternative \( j \) is

\[
U_n(j)
\]

The decision maker would be choosing the alternative that provides the greatest utility. Let \( U_n \) be the representative utility
which depends on parameters that are unknown and therefore estimated statistically. More specifically, let be the utility representation where are either known or estimated, and is the vector consisting of the attributes of decision maker \( n \) and alternative \( j \) that affect the person’s utility and hence their choice.

Logit (discussed in the next section) is by far the most widely used discrete choice model (Train, 2003). As discussed by Train (2003), the underlying assumption that (which captures the factors that affect utility but are not included in ) are independent and identically distributed (iid) extreme-value variates for all \( i \) and \( n \). The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives. This assumption provides a very convenient form for the choice probability. It may be that unobserved factors related to one alternative might be similar to those related to another alternative.

5.2.2 Logit Model

Luce (1959) originally derived the logit formula from the assumptions about the characteristics of choice probabilities, namely the independence from irrelevant alternatives (IIA). It was shown by Marschak (1960) that these axioms implied that the model is consistent with utility maximisation. As mentioned in the previous section, the logit model is obtained by assuming that is iid extreme-value variates. The probability density function for each unobserved component of utility is

\[
\frac{1}{\pi} \exp \left( -\frac{1}{2} \frac{x^2}{\pi^2} \right)
\]

and the cumulative distribution is
The difference between two extreme-value variables is logistically distributed. i.e., if \( x \) and \( y \) are iid extreme value variates, then follows the logistic distribution

\[
\log \left( \frac{e^x}{e^x + e^y} \right) = \log \left( \frac{e^{x-y}}{1 + e^{x-y}} \right)
\]

This formula is sometimes used to describe a model with two alternatives and is also known as the binary logit model. We are going to use this approach for our next model (discussed in this chapter). However, it is to be noted here that using the logistic distribution for the error differences or the extreme-value distribution for errors is nearly the same as assuming that the errors are independently normal.

Using above definition, Macfadden (1974), suggested a closed form of logit choice probabilities as follows:

\[
\pi_j = \frac{e^{u_j}}{\sum_{i=1}^{J} e^{u_i}}
\]

Representing the above choice probabilities with utility specified with linear parameters: where is a vector of observed variables related to alternative \( j \), the above relation can be written as follows:

\[
\pi_j = \frac{e^{x_jb}}{\sum_{i=1}^{J} e^{x_i b}}
\]

Finally, gives the choice probabilities of the \( j \)th alternative. After this short introduction of the logit model we will now present our discrete choice based model for lane-change behaviour.

\subsection{5.3 Discrete Choice Based Lane-Change Behaviour}

In this subsection, we present the idea behind this model and propose an advanced version of the transition matrix using discrete choice theory. We fuse three commonly used modelling techniques in traffic and transport
research, i.e. discrete choice model, Markov chain model, and state space model to capture the dynamics of traffic flow.

The lane-change behavior is accounted for by the transition matrix of the state space model (which is matrix $A$ in our case). In the previous chapter, the manoeuvre of a vehicle on a motorway segment is modelled as a Markov chain process where each lane is characterised as a state and we kept the transition matrix of lane changing probabilities ($A$) fixed due to assumption of the stable traffic condition. However, in this chapter we are dropping the stability assumption. Instead we are assuming that the traffic condition may change with time and hence we are actually proposing a model of estimating the transition matrix rather than using a fixed transition matrix as in the case of the previous model. In other words, drivers’ lane-change decisions are based on variable traffic condition which depends on various traffic factors (variables) such as vehicle speed, vehicle headway and density.

Due to the dynamic nature of traffic flow, the traffic condition varies with different time periods. Hence, the transition matrix ($A$) should be estimated to account for the dynamically changing traffic on motorways. To work out transition matrix $A$, we consider a logit function based model developed by Chang and Kao (1991) in the form of linear combination relating various traffic variables. Chang and Kao presented a utility function relating average headway, average density and density ratio of vehicles on a motorway. Following Chang & Kao (1991), we will apply a logistic transformation to this utility function to present our lane-changing probabilities (transition probabilities).

Let us suppose that we have an $M$-segment $N$-lane system. As explained above, the lane change probabilities between two lanes are not assumed to be constant but rather they depend on the current traffic conditions. To capture drivers’ lane-change movements, we use a discrete choice model (Ben-Akiva and Lerman, 1985; Train, 2003). In other words, a drivers’
choice probabilities are modelled by several traffic attributes such as vehicle time headway, traffic density and speed (Chang and Kao, 1991) as follows:

\[ q(i) = \frac{d_i}{s_i h_i}, \quad (5.2) \]

where \( d_i \) denote the density, average speed, and average time headway in lane \( i \) respectively. The function \( q \) takes values on the interval \([0, 1]\) only.

We follow the discrete choice model (5.1) and assume that driver’s lane-change decision depends on the traffic variables. Specifically, the lane-change probabilities along with average headway of vehicles in the same lane (\( AHD \)), average density ratio (\( ADR \), vehicle/lane/mile) and average lane speed (\( AS \), mile/hour), are specified as:

\[ p(i) = \frac{d_i}{s_i h_i} AHD + \frac{d_i}{s_i h_i} ADR + \frac{d_i}{s_i h_i} AS, \quad (5.3) \]

where \( d_i \) are coefficients which can be chosen appropriately.

We mentioned earlier that no subsequent lane changes are allowed within a 20 second time period. This is due to the fact that the time period of 20 seconds under the current system is too small to make any subsequent lane change. However, under other systems where the time period is larger than 20 seconds (say 90 seconds or 150 seconds), it is possible to make another lane change.

Now suppose that we have three or more alternatives to make a choice from. These may include staying in the same lane or changing to the left side lane or to the right side lane. The above equation (5.3) includes the two alternatives and can be used for the case binary model. For multi-lane models the equation (5.1) can be referred and the respective utility function
can be derived. As mentioned earlier, in this research study, the utility function is a combination of three traffic variables i.e. average headway, average density and average speed.

Finally, we use these (see equation 5.3) as transition probabilities which represent the driver’s lane-change behavior. These constitute the matrix $A$ which is used further in the state space model, discussed in the next section.

5.4 State Space Model

The output of the above discrete choice model (5.2) in the form matrix $A$ is then fed into the Markov chain model (Chapter 4) to describe how the lane change behavior evolves over time as traffic condition varies. Finally, the Markov chain model (Chapter 4) for lane change behavior is combined with the following state space model to characterise the dynamics of traffic flow:

\begin{align}
\frac{d}{dt} \mathbf{x}(t) &= f(\mathbf{x}(t), \mathbf{y}(t), \mathbf{z}(t)) \\
\mathbf{x}(0) &= \mathbf{x}_0
\end{align}  

\begin{align}
\mathbf{y}(t) &= \mathbf{y}(t) \\
\mathbf{z}(t) &= \mathbf{z}(t)
\end{align}  

(5.4)  

(5.5)

where

$f(\mathbf{x}) = A(\mathbf{x})$ and

denotes the state vector whose entries are the numbers of vehicles in each lane of a motorway segment in time step $t$.

denotes the vector consisting of the average speed in each lane in time step $t$.

denotes the vector of net gains of vehicles (i.e. the number of vehicles entering each motorway segment minus the number of vehicles leaving the motorway segment) in time step $t$. 

In addition, and are noise vectors of the state equation and observation equation respectively.

We will now estimate the densities using the above definition of the state space model. We will carry out estimation using two different techniques called the EKF and the UKF.

5.5 Density Estimation via the Extended Kalman Filter

In this section, the model under consideration is non-linear in nature since both state and measurement equations are non-linear. Further, density estimation presented here is different from that presented in the section 4.5 where we considered a linear state equation and non-linear measurement equation. A step by step detailed algorithm has been presented considering a non-linear state space system with lane-changing behavior. The other difference in the two approaches is that instead of the fixed (static) transition matrix (A) in the previous approach, we used a time-varying transition matrix on the basis of (5.3) to describe the changing traffic conditions. Equations 5.6 – 5.11 explain the estimation of traffic densities using EKF.

Let us denote the predicted estimate of the state vector as , the covariance matrix of the state vector as and for .

1) To initialize the EKF we set the state variables and covariance matrix as:

, and .

(5.6)

2) Predict the state vector and the covariance matrix

, .

(5.7)

, .

(5.8)
where $F = \ldots$ and $A$ can be estimated using equation 5.3 and is the covariance of the counting error for the quantity and defined in Chapter 3 (see section 3.2). Here acts as the transition matrix containing time-varying transition probabilities of vehicles changing lanes from one lane to other.

3) Compute the Kalman gain Matrix

$$K = \ldots$$

(5.9)

where can be defined in similar way as in section 4.5 (Chapter 4) and is the covariance of which is the error associated with the speed estimation and is defined in Chapter 3 (see section 3.4)

4) Update the state vector and its covariance matrix

$$\Sigma = \ldots$$

(5.10)

where are the independently determined observed traffic speeds defined by the observation equation (5.5) and

$$\hat{x} = \ldots$$

(5.11)

5) Let $t=t+1$ and return to Step 2.

5.6 Density Estimation via the Unscented Kalman Filter

Traffic modelling in this research study (and in general too) is highly non-linear (Linesch & Perez, 2007). So we thought of applying a more
robust and accurate (Julier and Uhlmann, 2004) filter, UKF, which is a non-linear filter. In the context of the state estimation, Julier and Uhlmann demonstrated the superior performance of UKF for non-linear systems.

As discussed in section 2.9.2, UKF directly applies the UT to the recursive KF framework. To estimate vehicular densities using UKF, we let denote the estimate of the state vector in each time step \( t \). In the implementation of the UKF, the state random variance is redefined as

\[
\text{(5.12)}
\]

where is the original state, is the process noise and is the measurement noise. We use equation 2.16 (see Chapter 2) to select sigma points to apply on so as to calculate the corresponding sigma points is the original state dimension, is the process noise dimension and is the measurement noise dimension. In a similar way, the state covariance is established by the individual covariance of

\[
\text{(5.13)}
\]

where and are the process noise covariance and the measurement noise covariance.

Implementation of the UKF algorithm is shown as follows (Julier & Uhlmann, 1997):

- Initialisation

,
where

- For time steps

1) Calculation of sigma points

\[ \text{(5.14)} \]

2) Time update

\[ \text{(5.15)} \]
\[ \text{(5.16)} \]
\[ \text{(5.17)} \]

where is again the weight of mean and denotes the weight of covariance for the \( i \)th sigma point.

3) Measurement update.

\[ \text{(5.18)} \]
\[ \text{(5.19)} \]
This UKF algorithm provides the optimal estimates of vehicle densities for multi-segment multi-lane motorways.

5.6 Simulation Description

5.6.1 Traffic Data Simulation

We used a self-developed microscopic simulator to simulate traffic in a 1-segment 2-lane motorway. The length of the motorway segment was set equal to 400 meters. Chapter 5 differs from Chapter 4 in that we have included the discrete choice based transition matrix accounting for the drivers’ lane-change behaviour. The system matrix (transition matrix) was set according to the above discussed logit choice model using utility function (see equation 5.3). Drivers’ choice (Lane-change) probabilities are modelled using several traffic variables.

The utility function (5.3) along with the discrete choice model (5.1) provided transition matrix $A$. It is important to mention that the constants in the utility function were carefully chosen so as to match the units of respective traffic variables. All the other experiment settings are the same as in Chapter 4.
Figure 5.1 shows the variations in a drivers’ choice behaviour in the two lane system. The upper plot is for lane 1 and the lower plot for lane 2. It can be seen that the drivers exhibit a range of choice probabilities which varies from 0.2 to 0.8 in different time periods. This can be interpreted as occurring due to the dynamic nature of the traffic flow and the lane change behaviour changes with changing time periods. The model in this chapter is well capable of tracking these changes and estimating the densities accordingly, which is explained later.

Following the same scheme as in previous chapters, each experiment was repeated 100 times. The evaluation of each method was based on the RMSE between the ‘true’ and estimated vehicle densities. For the ease of comparison, we are considering a case of two lane single segment motorway scenario.

![Figure 5.1 Lane change probability representation using the developed method](image-url)
scenario.

5.6.2 Addressed Research Issues

We will present the simulation studies separately for the EKF and UKF based lane change model developed above and address the following research issues one by one in both cases.

(i) Effect of Moderate Number of Input Vehicles – In the first issue, we will investigate the performance of the developed method under the normal traffic condition. We will consider that a moderate number (8-10 vehicles per 20 seconds) of traffic are flowing throughout the experiment.

(ii) Effect of Light Traffic Conditions – To analyse the light traffic scenario, we will discuss how the developed get affected if less number of vehicles (3-5) enter the segment. Here we consider that every 20 seconds of time step, 3-5 vehicles are entering the segment.

2. Comparison between EKF and UKF based approaches - Here, we compare the differences between two approaches and discuss how these differences have led to the consequent results.

3. Comparison with Existing Methodologies - Lastly, we compare the two newly developed approaches with the existing methodologies.

5.7 Simulation Results

In this subsection we discuss the above mentioned research issues one by one considering the above simulation settings.
5.7.1 Simulation Analysis Using EKF Approach

(i) Effect of Moderate Number of Input Vehicles – Here 8 vehicles per time step (of 20 seconds) are entering the given segment of motorway. The vehicle count error ($\sigma$) from the loop detectors is about 1 vehicle per time step, and the speed error ($\tau$) of the speed detectors is about 1 km/hr.

Figures 5.2 & 5.3 represent the scenario mentioned above. Like the previous models, simulated traffic counts are represented by broken lines (red) and the corresponding estimated traffic counts are shown as solid lines (blue).

The RMSEs averages over the repeated 100 runs are displayed in Table 5.1. We investigated the performance of the developed approach based on EKF in various scenarios by setting the parameters equal to different values in the state space model: $\tau$ and $\sigma$ with number of entering vehicles being 8 in both the lanes respectively.

Figure 5.2 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (solid line) using the developed method with EKF for lane 1 (left) and lane 2 (right) with $\sigma$=1, $\tau$=1 and $\lambda$=8.
Figure 5.3 Lane change probability representation using the developed method with EKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=8$.

Figure 5.3 provides important information about the variation in a drivers’ choice behavior which keeps on varying with time and is dynamic in nature which is expected.

It is clear that the developed approach (EKF based) performed well: the estimated traffic density was fairly close to the ‘true’ traffic density. It can

<table>
<thead>
<tr>
<th>$\sigma=1$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1</td>
<td>1.24</td>
<td>3.17</td>
<td>3.71</td>
</tr>
<tr>
<td>Lane 2</td>
<td>1.56</td>
<td>3.19</td>
<td>3.82</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>$\tau=1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane 1</td>
<td>1.21</td>
<td>2.43</td>
<td>5.60</td>
</tr>
<tr>
<td>Lane 2</td>
<td>1.92</td>
<td>2.36</td>
<td>6.06</td>
</tr>
</tbody>
</table>

Table 5.1 Average RMSEs for a two-lane motorway over 100 simulation runs using the developed state equation with $\lambda=8$ with EKF
be seen from the Table 5.1 that overall, the estimated traffic densities are quite accurate. When the parameter $\sigma$ and $\tau$ is small, the average estimation error is about one vehicle per 20 seconds. When the two parameters become larger, the average estimation error increase but are still at a low (acceptable) level with a few exceptions.

(ii) Effect of Light Traffic Conditions – Here, we assume that there were 4 vehicles entering/leaving the chosen segment in every 20 second time step. Again, Figure 5.4 displays the ‘true’ traffic counts and the corresponding estimated values in one run of the experiment, where the left and right plots correspond to the two lanes respectively. Figure 5.5 displays the corresponding lane-change behavior probabilities for two lanes. It can be observed that probabilities vary between 0.2 – 1. Moreover, although lane-changes have taken place in a healthy manner, most of the time steps fluctuate around 0.5 showing that half of the vehicles took part in the lane-change process. We assume the magnitude of measurement and observation errors to be the same as in the previous case i.e. $\sigma = 1$, $\tau = 1$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.4.png}
\caption{Simulated traffic counts (broken line) and the corresponding estimated traffic counts (solid line) using the developed method with EKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=4$}
\end{figure}
Figure 5.5 Lane change probability representation using the developed method with EKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=4$.

Table 5.2 contains the RMSE values for this traffic scenario where a comparatively small number of vehicles are entering/leaving the motorway segment. It can be observed from Table 5.2 that an error ranging from 0 - 2 vehicles can occur with the different choice of parameters. On comparing with the Figure 5.3, we realised that vehicles in the light traffic situation change lanes more often. This may be due that fact that they more room to do so (assuming they are travelling safely).

Table 5.2 Average RMSEs for a two-lane motorway over 100 simulation runs using the developed state equation with $\lambda=4$ with EKF

<table>
<thead>
<tr>
<th></th>
<th>$\sigma=1$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1</td>
<td>0.58</td>
<td>1.33</td>
<td>2.36</td>
<td></td>
</tr>
<tr>
<td>Lane 2</td>
<td>0.63</td>
<td>1.48</td>
<td>2.17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\sigma=2$</th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1</td>
<td>0.82</td>
<td>1.71</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>Lane 2</td>
<td>0.75</td>
<td>1.26</td>
<td>1.72</td>
<td></td>
</tr>
</tbody>
</table>
5.7.2 Simulation Analysis Using UKF Approach

Using a similar choice of parameters and the exact same scheme of entering and leaving vehicles, we carried out simulation experiments for the UKF approach.

(i) Effect of Moderate Number of Input Vehicles – Figure 5.6 and 5.7 represent the scenario where a moderate number of vehicles enter or leave the respective segment. Again results can be interpreted in a similar way as in the previous section along with variable lane change behavior which truly represents the dynamic nature of traffic flow.

Figure 5.6 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (solid line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=8$.
Figure 5.7 Lane change probability representation using the developed method with UKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=8$.

We use RMSE values to investigate the impact of the parameters $\sigma$ and $\tau$ on the accuracy of estimation along with the $\lambda = 8$ (number of vehicles). Table 5.3 displays RMSE values averaged over the repeated 100 runs for different scenarios using the UKF based method when $\lambda = 8$.

<table>
<thead>
<tr>
<th></th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane 1</td>
<td>2.32</td>
<td>3.74</td>
<td>3.78</td>
</tr>
<tr>
<td>Lane 2</td>
<td>2.50</td>
<td>3.89</td>
<td>4.24</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane 1</td>
<td>1.80</td>
<td>3.14</td>
<td>5.39</td>
</tr>
<tr>
<td>Lane 2</td>
<td>2.00</td>
<td>2.78</td>
<td>5.42</td>
</tr>
</tbody>
</table>

(ii) Effect of Light Traffic Conditions - Figures 5.8 and 5.9 represent the scenario which exhibits light traffic inside the respective motorway segment. We are assuming that only 4 vehicles are entering every 20 seconds in the selected segment in each of the two lanes.

As we did in the case of EKF, we first set $\sigma = 1$, $\tau = 1$ and $\lambda = 4$ in the simulation experiment for UKF. Similar results can be observed in the following Figures (5.8 & 5.9) for the UKF simulation. Also it is to be noted that lane-change behavior is concentrated around the middle of the plot, showing that half of the vehicles took part in the lane-change process as in the case of EKF.
Chapter V – DISCRETE CHOICE BASED LANE CHANGE MODEL

**Figure 5.8** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (solid line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=1$, $\tau=1$ and $\lambda=4$.

**Figure 5.9** Lane change probability representation using the developed method with UKF for lane 1 (up) and lane 2 (down) with $\sigma=1$, $\tau=1$ and $\lambda=4$.

The RMSEs averages over the repeated 100 runs are displayed in Table 5.4 using the developed method with UKF when $\lambda = 4$.

**Table 5.4** Average RMSEs for a two-lane motorway over 100 simulation runs using the developed state equation with $\lambda=4$ with UKF

<table>
<thead>
<tr>
<th></th>
<th>$\tau=1$</th>
<th>$\tau=2$</th>
<th>$\tau=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td>Lane 1</td>
<td>1.45</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>Lane 2</td>
<td>1.28</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.94</td>
<td>1.43</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>Lane 1</td>
<td>2.01</td>
<td>1.41</td>
</tr>
</tbody>
</table>
The Table 5.4 shows that there can be an error of 1-3 vehicles depending upon the choice of parameters when light traffic is running through the chosen segment in a multi-lane motorway.

From the above results in the form of plots and tables explaining various research issues, it is clear that UKF is also an efficient estimator with a high level of accuracy and low levels of error.

After presenting simulation studies of the EKF and the UKF based approaches, we will compare their performances and analyse which one is better for this research study and why.

5.7.3 Comparison between EKF and UKF Based Approaches

The performance of any filter depends on various factors such as low noise & good initial estimates. With EKF the performance varies depending on how Jacobians (partial derivative matrices) are updated and as the number of Jacobians increases, the improvement in the performance becomes less dramatic. With UKF, Jacobians are not calculated but computational effort is typically high due to the need for multiple simulations at each time step of the underlying dynamic system (Simon, 2007). The matrix square root is the biggest computational difficulty in the case of UKF, although it can be reduced by skipping the time update equation for the sigma points and simply using the most recent sigma points in the succeeding equations.

In the literature, this account of comparative performance of EKF and UKF has been discussed by many researchers (for details, see Malan and Steyn, 2005; La Viola, 2003; Kumar et al., 2010; Simon, 2007).

From Tables 5.1 - 5.2 and 5.3 - 5.4, it is evident that the EKF based approach performs better in this research than the UKF. However, it should
be noted that UKF is not a complete failure. In the literature too, this ironic performance of these two filters have been observed. The effectiveness and advantages of using EKF over UKF have been studied in the literature (for details see Kumar et al., 2010; Simon, 2007; Malan and Steyn, 2005).

It has been already discussed in previous chapters that EKF uses the linearization approach to determine the current mean and variance of the states, however, UKF utilizes an UT which is a deterministic sampling approach to calculate the current mean and covariance of states. The performance of UKF can sometimes be poor because its predictive variances can be far too small if the sigma points are placed at inconvenient locations.

As a result of too small predictive variances, observations have too much weight in the measurement update and hence lead to the poor performance of UKF. “In the most extreme cases, the UKF can give a delta spike predictive distribution” (Turner and Rasmussen, 2011). This phenomenon is also known as sigma point collapse. The marginal likelihood is extremely unstable in the regions that experience sigma point collapse. In the situation of sigma point collapse, the predictive variances become far too small making the marginal likelihood much more susceptible to noise. The mean of the predictive distribution is used (Turner and Rasmussen, 2011) as a diagnostic for sigma point collapse.

Now we discuss the reasons for the observed differences between the two filtering approaches (EKF and UKF) for the above developed model in this chapter. Firstly, we know that the transportation system or more specifically traffic flow is non-linear in nature. This indicates that the EKF and UKF should outperform the KF and for a highly non-linear system we would expect the UKF to perform better than the EKF. However, for the traffic flow problem under the current setting the non-linearities are not severe enough to cause the UKF to perform better than the EKF.
Results may have been different if we had considered the Exit or On Ramps or lane-changing between all possible lanes, hence making the system more complicated and thus more non-linear in nature. This all indicates that the current system under consideration is non-linear but not so much that extra computational effort (e.g., with UKF or a computer intelligence based approach) is warranted. Once we get past the complexity of the EKF, we have reached a point where we achieve significantly accurate traffic density estimates.

To reduce the estimation errors which are due to non-linearities, it is natural to consider the use of the approach of higher-order Taylor expansion. However, work presented in this research indicates that the traffic flow non-linearities are small enough that the UKF does not provide more accurate performance than the EKF and it is doubtful that other higher order approaches will result in much improvement either.

Therefore considering cost as the major real world driver, EKF model seems a good trade-off solution for developing TMS where underlying non-linearities are mild in nature.

5.7.4 Comparison with Existing Method

Next, we compare the performance of the discrete choice based model developed in this chapter with that used in Gazis & Liu (2003). As mentioned earlier, the model considered in Gazis & Liu (2003) ignores lane-change behavior. It works well only if no vehicles make lane-changes. When there are a substantial number of vehicles that change their lanes, the state equation (Gazis & Liu, 2003) fails to reflect the true dynamics of traffic flow, and thus we expect a poorer performance of the model.

We are presenting the comparison of the UKF based approach as the EKF based approach with fixed transition matrix has already been compared in the last chapter (see Section 4.7.4) showing the superiority of the
developed model over the existing methodologies and estimating more accurate vehicle densities.

Figure 5.10 displays the ‘true’ traffic counts (broken line - red) and the corresponding estimated values (solid line - blue), where the left and right plots correspond to the innermost and outermost lane respectively. It can be seen that the quality of the estimates is poor when compared to Figure 5.11 which exhibits the results of the developed model under the same setting of parameters. In particular, the estimated density greatly departed from the ‘true’ values due to lane-changes being ignored in the traffic flow modelling.

![Figure 5.10](image)

**Figure 5.10** Simulated traffic counts (broken line) and the corresponding estimated traffic counts (solid line) using the existing method (Gazis & Liu, 2003) for lane 1 (left) and lane 2 (right) with $\sigma=2$, $\tau=2$ and $\lambda=8$

Table 5.5 & 5.6 show the RMSE values averaged over 100 repeated runs for different settings of the parameters. It can be seen from Table 5.5 & 5.6 that the estimation errors are larger than their counterparts in Tables 5.3 to 5.4. Clearly this is the consequence that the lane-change behavior was not
Figure 5.11 Simulated traffic counts (broken line) and the corresponding estimated traffic counts (real line) using the developed method with UKF for lane 1 (left) and lane 2 (right) with $\sigma=2$, $\tau=2$ and $\lambda=8$ taken into account in the state equation. It is clear from the Tables 5.5 & 5.6 that with the increase in the error parameters, results get deteriorated and the estimation error of up to 12 vehicles can occur if we ignore the lane-change behavior.

Thus, the existing methodology fails to provide accurate estimates of traffic densities because they do not include the lane-change behavior which represents more real life like situations of road traffic systems.

Table 5.5 Average RMSEs for a two-lane motorway over 100 simulation runs using the existing method (Gazis & Liu, 2003) state equation with $\lambda=4$

<table>
<thead>
<tr>
<th></th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 1</th>
<th>Lane 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=1$</td>
<td>3.32</td>
<td>4.37</td>
<td>3.56</td>
<td>6.74</td>
<td>4.64</td>
<td>5.86</td>
</tr>
<tr>
<td>$\tau=1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau=2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau=3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6: Average RMSEs for a two-lane motorway over 100 simulation runs using the existing method (Gazis & Liu, 2003) state equation with $\lambda=8$

<table>
<thead>
<tr>
<th></th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 1</th>
<th>Lane 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=2$</td>
<td>16.66</td>
<td>9.36</td>
<td>12.05</td>
<td>7.62</td>
<td>11.36</td>
<td>7.70</td>
</tr>
<tr>
<td>$\tau=1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau=2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau=3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter V – DISCRETE CHOICE BASED LANE CHANGE MODEL

<table>
<thead>
<tr>
<th></th>
<th>Lane 1</th>
<th>Lane 2</th>
<th>Lane 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ=1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.73</td>
<td>3.02</td>
<td>3.20</td>
</tr>
<tr>
<td></td>
<td>8.72</td>
<td>7.04</td>
<td>4.24</td>
</tr>
<tr>
<td>σ=2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.45</td>
<td>10.48</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>7.34</td>
<td>6.67</td>
<td>5.55</td>
</tr>
</tbody>
</table>

5.8 Summary

In this chapter, we have presented a novel approach to modelling lane change behavior in a dynamic traffic scenario using the logit model technique. This approach relates various traffic variables to give lane-change information based on real traffic scenarios. We have used two different approaches, the EKF and UKF techniques and undertaken simulation studies. We have shown that a discrete choice based lane change model works well in different traffic scenarios. We further compared the two developed approaches, EKF and UKF, with each other and with existing method (Gazis & Liu, 2003). The developed multi-segment multi-lane motorway based lane change estimation model is more accurate than the existing methods and provides equally accurate estimates with both EKF and UKF techniques.

So far we have developed three different models and presented respective simulation studies for each model. Our next chapter discusses the real data analysis considering various multi-lane scenarios.
CHAPTER VI

REAL TRAFFIC DATA ANALYSIS

This chapter deals with the final part of this research study which is the real traffic data analysis for the novel models developed in this research study. First of all, we will discuss the process of data collection. Then we will explain the data filtration (screening) process undertaken to remove the impure/incorrect data and to handle the missing data values from the respective data sets. We will then summarise the technique of system identification (a MATLAB toolbox) which is used to identify the state space system under study. We will finally carry out the real traffic data studies for 2, 3 and 4-lane cases and discuss the performance of the developed model in each of these cases individually for multi-segment multi-lane motorways. We conclude the chapter by summarising at the end.

6.1 Data Collection

As discussed in Section 1.3.2 of Chapter 1, the data was collected from a traffic database known as TDAD (Traffic Data Acquisition and Distribution), a Washington State Department of Transportation (http://www.ivhs.washington.edu/tdad/) funded project. ILDs have been used to collect volume and speed data from different sections (cabinets) on US based motorways namely SR-99, I-5 and I-90. The loops were placed approximately 500 meters apart in the segment. These segments have ILD loops and speed traps installed. In addition, they are free from exits and entries, thus have a relatively simple motorway network layout. The data in the form of volume and speed values was collected during two different time periods of a working day i.e., peak traffic hours (8 – 10 am) and off-peak
traffic hours (10 am – 12 noon) for 2 consecutive hours on the 28\textsuperscript{th} and 29\textsuperscript{th} May 2007. The data obtained on 28\textsuperscript{th} May was used for modelling using system identification technique and data obtained on 29\textsuperscript{th} May was used for testing the model (to estimate the density using the KF approach). The data obtained from different sections is for studying the applicability of the developed models under varied traffic conditions during different times of a working day.

Data from the source was downloaded in the form of ‘txt’ files and saved as ‘tab limited’ format to facilitate further data screening and conditioning which is explained in the next section.

6.2 Data Screening

After collecting the field data, extensive data screening was carried out. The ‘.txt’ file obtained from the data source contained all the speed or count values in one column $((r \times n) \times 1$ vector; where $r$ is the number of time periods and $n$ is the number of lanes) for all different lanes. So these values needed to be segregated and re-formatted before further use. So we separated these values lane-wise. After obtaining the speed and count values separately for all the lanes in separate columns, quality control was carried out to eliminate incorrect and impure data values to avoid any discrepancies at the later stages of analysis.

This data purification process also includes checking for any missing data values (speed and volume) individually and in combination at different locations. Data was available every 20 seconds from the detector system, and sometimes due to various reasons (e.g. no vehicle passing through) detectors skip to track the record of the vehicles, leading to missing data. To handle this situation we prepared a scheme which is as follows.

Considering the case of a two-lane motorway, data was extracted in the raw form (see Table 6.1) and rearranged (count and speed) values lane wise
After rearranging, we computed net counts by taking the difference of entering and leaving vehicle counts and if both speed values are available, we averaged two speed values one from each of the detectors and then eliminated the zero values.

**Table 6.1** Raw data values i.e. counts and speeds from two ILDs at the two ends of the chosen segment for the peak traffic time analysis.

<table>
<thead>
<tr>
<th></th>
<th>Counts</th>
<th></th>
<th>Speeds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loop1</td>
<td>Loop2</td>
<td>Loop1</td>
</tr>
<tr>
<td>Lane 1</td>
<td>2</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Lane 2</td>
<td>1</td>
<td>0</td>
<td>33.2</td>
</tr>
<tr>
<td>Lane 1</td>
<td>3</td>
<td>2</td>
<td>32.7</td>
</tr>
<tr>
<td>Lane 2</td>
<td>3</td>
<td>1</td>
<td>34.4</td>
</tr>
<tr>
<td>Lane 1</td>
<td>1</td>
<td>0</td>
<td>36.5</td>
</tr>
<tr>
<td>Lane 2</td>
<td>5</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Lane 1</td>
<td>0</td>
<td>2</td>
<td>28.8</td>
</tr>
<tr>
<td>Lane 2</td>
<td>3</td>
<td>3</td>
<td>34.4</td>
</tr>
<tr>
<td>Lane 1</td>
<td>4</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Lane 2</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Lane 1</td>
<td>0</td>
<td>0</td>
<td>26.9</td>
</tr>
<tr>
<td>Lane 2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lane 1</td>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Lane 2</td>
<td>0</td>
<td>2</td>
<td>25.8</td>
</tr>
<tr>
<td>Lane 1</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Lane 2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lane 1</td>
<td>6</td>
<td>0</td>
<td>26.6</td>
</tr>
<tr>
<td>Lane 2</td>
<td>4</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Lane 1</td>
<td>1</td>
<td>2</td>
<td>8.7</td>
</tr>
<tr>
<td>Lane 2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 6.2 Rearranged data values i.e. counts and speeds from two ILDs at the two ends of the chosen segment for the peak traffic time analysis.

<table>
<thead>
<tr>
<th>Counts</th>
<th></th>
<th></th>
<th></th>
<th>Speeds</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loop1</td>
<td>Loop2</td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Loop1</td>
<td>Loop2</td>
<td>Lane 1</td>
</tr>
<tr>
<td></td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Lane 1</td>
<td>Lane 2</td>
<td>Lane 1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>21</td>
<td>33.2</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>32.7</td>
<td>34.4</td>
<td>24.2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>36.5</td>
<td>28</td>
<td>3</td>
<td>22.2</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>28.8</td>
<td>34.4</td>
<td>4</td>
<td>9.7</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td>37.5</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>26.9</td>
<td>0</td>
<td>9.7</td>
<td>31.3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>25.8</td>
<td>7</td>
<td>11.8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>28.6</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>26.6</td>
<td>8</td>
<td>14.5</td>
<td>28.7</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8.7</td>
<td>0</td>
<td>9.5</td>
<td>29</td>
</tr>
</tbody>
</table>

The Second step after collecting data is to identify the system (in our case state space matrices). We used the system identification toolbox (MATLAB) to estimate the system or the state space model. Our system (model) closely resembles the linear grey-box model structure as per MATLABs’ specification, so we used the grey-box modelling approach (within the toolbox) to identify the system. After identifying the system (estimating the system matrices), we coded the model in MATLAB and finally estimated densities. Before presenting the actual real data analysis for different multi-lane motorway structures, it is important to explain the system identification toolbox, especially the technique with which we estimated our state space model (grey-box model). We will discuss this important issue in the next section.

6.3 System Identification

In control engineering, the statistical methods are used to build mathematical models of dynamical systems from the measured data. This is termed as system identification which also includes generating informative data
efficiently (by optimally designing the experiments) for fitting such models as well as model reduction.

MATLAB has a toolbox called ‘system identification’ which can be used to construct mathematical models of dynamic systems from measured input-output data. This data-driven methodology helps researchers to describe the systems that are difficult to model otherwise (for example, from first principles or specifications, such as chemical processes and engine dynamics).

This toolbox facilitates the estimation of parametric models, such as transfer functions or state space models. Parametric models are those which use a small number of parameters to capture system dynamics. The toolbox estimates model parameters and their uncertainties. Further, it allows researchers to estimate a user-defined (grey-box) model which can be stated as a set of differential or difference equations with some unknown parameters. This toolbox let researchers specify the model structure and estimate its parameters using non-linear optimisation techniques. For linear models, researchers can specify the structure of the state space model (system matrices) and impose constraints on identified parameters. However, for non-linear models, differential equations can be specified as MATLAB, C, or FORTRAN code.

Researchers can estimate linear discrete-time and continuous-time grey box models using single-output and multiple-output time domain data or output only time-series data. In order to estimate the grey-box model, it must be represented in the state space form. The state space model uses state variables $\mathbf{x}(t)$ to describe a system as a set of first-order differential equations.
This research deals with discrete-time models, so we will present the estimation procedure of a system for discrete-time grey-box models only. Consider the discrete-time state space model structure as follows:

After specifying the model in the state space form, the very first step in grey-box modelling is to write a function which does a job of returning state space matrices as a function of user-defined parameters and information about the model.

The following is the format to implement the linear grey-box model in a MATLAB file:

\[
[A, B, C, D, K, x0] = \text{myfunc}(par, T, aux),
\]

where the matrices \(A, B, C, D, K\), and \(x0\) represent discrete-time state space description of the system,

\(\text{myfunc}\) is the name of the file,

\(par\) contains the parameters as a column vector,

\(T\) is the sampling interval,

\(aux\) contains information of auxiliary variables (if any) in the system. We may wish to use auxiliary variables to vary system parameters at the input to the function, and avoid editing the file again and again. Next is to use \(\text{pem}\) to estimate the grey-box model.

‘\(\text{pem}\)’ stands for prediction-error minimisation method. This command estimate model parameters using iterative method which minimises the prediction errors to obtain maximum-likelihood values. We can use \(\text{pem}\) command for constructing and estimating linear state space models.

After creating the nominal model, we must specify which parameters to estimate and which to set specific values. To accomplish this task, we must
edit the structure of the following model properties: $A_s$, $B_s$, $C_s$, $D_s$, $K_s$ and $x_0$s which are structure matrices of the same size and represent properties of the nominal model created above.

To specify structure matrices we must:

- set a $\text{par}$ value to flag free parameters (to be estimated) at the corresponding locations in $A$, $B$, $C$, $D$, $K$ and $x_0$.
- specify the values of fixed parameters (either 0 or 1) at the corresponding locations in $A$, $B$, $C$, $D$, $K$ and $x_0$.

**Example:** We explain the above definition of the linear state space model with the help of the following example.

Consider the discrete-time state space model single-input and single-output system (SISO) as follows:

$$
A = \begin{pmatrix}
\end{pmatrix}, \quad B = \begin{pmatrix}
\end{pmatrix}, \quad C = [\text{par3} \quad \text{par4}] 
$$

and $w$ and $e$ are independent white-noise terms with covariance matrices $R_1$ and $R_2$, respectively. $R_1 = \text{E}\{ww'\}$ is a 2-by-2 matrix and $R_2 = \text{E}\{ee'\}$ is a scalar. Unknown parameter values to be estimated are represented by $\text{par1}$, $\text{par2}$, $\text{par3}$, and $\text{par4}$.

We assume that the variance of the measurement noise $R_2$ to be 1. $R_1$ (1, 1) is unknown and is treated as an additional parameter $\text{par5}$. The remaining elements of $R_1$ are known to be zero.
Next we express the above SISO system as an *idgrey* (grey-box) model using a function. Then this file, together with the *pem* command, would be used to estimate the model parameters based on initial parameter guess.

To estimate the system in the form of the grey-box model, we apply the following procedure:

1. Create the file *sysnoise* that computes the state space matrices as a function of unknown parameters and the auxiliary variable (*R2*).

   ```
   function [A, B, C, D, K, X0] = sysnoise (par, T, aux)
   R2 = aux (1); % known measurement noise variance
   A = [par(1) par(2) ; 1 0];
   B = [1;0];
   C = [par(3) ; par(4)];
   D = [0;0];
   R1 = [par(5) 0; 0 0];
   [est , K] = kalman (ss (A, eye(2), C), T), R1, R2); % Uses Control System toolbox product
   x0 = [0 , 0];
   ```

2. Specify the initial guess for the five unknown parameter values (par(1),...,par(5)) and the auxiliary value.

3. Construct *idgrey* model using the *sysnoise* file created above

   ```
   Minit = idgrey ('sysnoise', Pvec, 'd', auxval);
   ```

   here *d* stands for discrete-time system.

4. Estimate the model using *pem*

   ```
   Model = pem (data, Minit)
   ```

where *data* is the name of the iddata object containing input-output data which is defined as follows.

   ```
   data = iddata ([y1 y2], [u1 u2]);
   ```

where *u1* and *u2* are inputs and *y1* and *y2* are outputs of the system.
By using the iterative scheme, the system matrices can be estimated and hence the system can be identified. Note that in the real traffic data analysis, the model specification for the system identification would be different from the example shown above. For instance, during real data analysis we may consider scenario of four lanes and hence the matrix $A$ would be a 4x4 matrix instead of 2x2 as above.

After identifying the state space system, we used this system along with real traffic data to estimate vehicular densities. We used the Markov chain based lane change with the transformed observation equation and fixed lane change probabilities to carry out real data analysis. We presented this application (traffic data analysis) explicitly in the form of plots and presented the comparison between the developed method and the existing methodology. The analysis section is divided into two sub-sections and we have discussed the applications for peak-traffic time and non-peak traffic time.

6.4 Applications
In this section, we will present density estimation results considering the Markov chain based lane change behaviour using the transformed observation equation with linear state equation (model developed in chapter 4).

We divide this section into two subsections for presenting results for peak time traffic and non-peak time traffic separately. In each of these two sub-sections we will discuss the case of a 2, 3 and 4-lane motorway. The results are presented to explain different multi-lane scenarios having a different number of lanes under each case. We will also further compare our results with the existing model (Gazis & Liu, 2003).
6.4.1 Peak Traffic Time Analysis (8:00 am – 10:00 am)

6.4.1.1 Two-Lane Case

(i) Modelling - For 2-lane motorways, the modelling data we collected is a 2 hour count including speed data taken on 28th May 2007 between 8:00 am & 10:00 am (peak time traffic) from the sections on SR-99 motorway namely, S 102nd St and Cloverdale St (see Figure 6.1). After obtaining the data in the raw from as ‘.txt’ files, it was pre-screened and formatted as mentioned in Section 6.2

![Figure 6.1 A section of study from Motorway SR 99](image)

After obtaining the processed data, the state space system was identified using MATLAB as explained in Section 6.3. As a result we got the transition matrix $A$ which represents the lane change behaviour over multi-lane motorways. We used the modelling data to identify the system and then used the application data to actually estimate the densities via the KF approach. Modelling data provides the system specification (the matrices of the system) which was then used along with application data to estimate the densities using developed model. The obtained transition matrix is
Matrix $A$ can be interpreted in the following way. In the first lane, around 89% of the vehicles remain in the same lane in which they were initially travelling, whereas the rest of the vehicles change lane in the due course of their travel. Similarly, for the second lane, around 86.3% of the vehicles remain in the same lane but the rest (around 13.7%) of them change to the first lane respectively.

(ii) Application – For using our model with real traffic data, we obtained 2 hour data count including speed data taken on 29th May 2007 between 8:00 am & 10:00 am (peak time traffic) from the sections on SR-99 motorway namely, S 102nd St and Cloverdale St. The collected data was then used to estimate densities using the model identified earlier. Figures 6.1 & 6.2 display the density estimation during peak time traffic scenario for 2-lane motorways.

Figure 6.2 The traffic counts using the developed method for 2-lane motorway using peak time real traffic data.
In Figure 6.2 the plots represents the estimation results from the developed model and in Figure 6.3 the plot depicts the estimation results from Gazis & Liu (2003) model. The left plot in both the Figures represents the first lane and the right plot represents the second lane. A comparative study is presented in the form of plots. It is important to mention that the idea behind this real traffic data analysis is to show the application of the developed methodology using real traffic data. It can be seen from the above plots, the existing methodology (Gazis & Liu, 2003) failed to provide efficient estimates as compared to the developed approach in this research. The estimated densities via the existing methodology provide incorrect results since lane change behaviour is ignored due to linearization and approximation issues as explained earlier in the thesis.

To study the behaviour of the developed methodology over different multi-lane scenarios, it was important to carry out experiments with 3-lane and even 4-lane motorways as these numbers of lanes are common in practical settings. So, next we will show the 3-lane and 4-lane case.

6.4.1.2 Three-Lane Case
(i) Modelling - For 3-lane motorways, the modelling data was collected on 28\textsuperscript{th} May 2007 between 8:00 am & 10:00 am (peak time traffic) from the
sections on I-5 named as of S Oregon St and S Pearl St (see Figure 6.4).

![Figure 6.4 A section of study from Motorway I - 5](image)

After obtaining the pre-screened and the processed data, we got the transition matrix $A$ in a similar way as explained in the previous section which explicitly represents the lane change behaviour over multi-lane motorways. The obtained transition matrix can be stated as follows:

$$A = \begin{pmatrix} \ldots \end{pmatrix}.$$

Matrix $A$ can be interpreted as follows. In the first lane, around 86.7% of the vehicles remain in the same lane in which they were initially travelling, whereas the rest (around 13.3%) of the vehicles change to the second lane in due course of their travel. It is worth remembering here that we are assuming no lane change should be taking place to the next-to-next lane. Similarly, for the middle lane, around 67.5% of the vehicles remain in the middle lane, 11.5% of them change to the first lane and around 21% change to the third lane. It should be noted that in a three-lane case, the middle lane
is the one which has greater ability to allow users to change lane to or from either side which is not the case with first or third lane. Lastly, in the third lane 78% of the vehicles remain in the same lane, however around 22% changes to second lane. It is to note here that the accuracy of the matrix ‘A’ depends on the source and time of the day of the real data collected/obtained. After obtaining the transition matrix, the next step is to test the developed model.

(ii) Application– For application purposes, we collected a 2-hour count and speed data on 29th May 2007 between 8:00 am & 10:00 am (peak traffic time) from the cabinets on I-5 namely of S Oregon St and S Holgate St.

Using the above transition matrix and application data, Figures 6.3 & 6.4 represent the results of density estimation during peak time scenario for 3-lane motorways.

**Figure 6.3** The traffic counts using the developed method for 3-lane motorway using peak time real traffic data.

**Figure 6.4** The traffic counts using the Gazis & Liu (2003) method for 3-lane motorway using peak time real traffic data.
The plots in Figure 6.5 represent the estimation results of traffic density from the developed model and the plots in Figure 6.6 depict the estimation results from the Gazis & Liu (2003) model. Different lanes have been shown by using different plots; the first plot showing the first lane, the second plot showing the middle and the third plot representing the third lane respectively. It can be seen from the above plots that estimation results from the existing methodology are not in favour as compared to the developed methodology and fail to provide efficient density estimates.

After representing real traffic data analysis for 2, 3–lanes, we will discuss the 4-lane scenario in the next section.

6.4.1.3 Four-Lane Case

(i) Modelling - For 4-lane motorways, the modelling data was collected on 28th May 2007 between 8:00 am & 10:00 am (peak time traffic) from the cabinets I-90 including 246th Ave SE, EB and 256th Ave SE (see Figure 6.7).

![Figure 6.7 A section of study from Motorway I - 90](image-url)
After obtaining the polished data, the transition matrix $A$ can be stated as follows:

$$A = \begin{pmatrix}
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
\end{pmatrix}.
$$

In this case, matrix $A$ can be interpreted by saying that in the first lane, around 81% of the vehicles remain in the same lane in which they were initially travelling, whereas the rest around 19% of the vehicles, change to the second lane in due course of their travel. However, no changes take place to the 3rd and 4th lanes, so we have ‘zero’ values at respective positions of $A$. Similarly, for the second lane around 25.3% of the vehicles remains in the same (second) lane, 24.8% of them change to the first lane. Further, around 49.9% change to the third lane. It should be a point to note that in a four-lane case, the second and third lanes are the ones which have more ability to allow users to change lane to or from either side which is not the case with the first and fourth lane. In the third lane 24% of the vehicles remain in the same (third) lane, 56% change to second lane and around 18% change to the fourth lane respectively. Almost 75% of vehicles remain in the fourth lane itself however the other 25% changes to the third lane. After obtaining the transition matrix, the next step is the model application.

(ii) Application – For application purposes, we collected a 2-hour count including speed data taken on 29th May 2007 between 8:00 am & 10:00 am (peak time traffic) from the sections I-90 including 246th Ave SE, EB and 256th Ave SE.

Using the 4-dimensional transition matrix obtained via system identification technique and the real traffic data, the results of density
estimation during a peak time scenario for a 4-lane motorway is shown as follows.

**Figure 6.8** The traffic counts using the developed method for 4-lane motorway using peak time real traffic data.

**Figure 6.6** The traffic counts using the Gazis & Liu (2003) method for 4-lane motorway using peak time real traffic data.

In Figures 6.8 & 6.9, the upper plots represent the estimation results from the developed model and the lower plots depict the estimation results from Gazis & Liu (2003) model. Different lanes have been shown by using 4 different plots from left to right for 1st, 2nd, 3rd & 4th lane respectively.

It is evident that the estimation results using the existing methodology (due to the lack of lane-change information) completely lose the estimation track and provide unsatisfactory results (inaccurate) as compared to the developed methodology.

**6.4.1.4 Summary (for peak traffic time)**

From the above real traffic data analysis, it is clearly evident (from the Figures 6.1- 6.9) that the developed (Markov chain based lane change) model provides satisfactory results for real traffic data collected for 2, 3 and
4 – lane motorways. For all the three cases explained above, it worked well and provided accurate results. Moreover, the developed linear model excels over the existing model (Gazis & Liu, 2003). The plots in the Figures 6.8 & 6.9 clearly show that the developed model is applicable to the variable multi-lane scenarios during peak-traffic hours and hence a major contribution to the traffic theory for density estimation considering lane change behaviour.

6.4.2 Non-Peak Traffic Time Analysis (10:00 am – 12:00 noon)

6.4.2.1 Two-Lane Case

(i) Modelling – For 2-lane motorways, we start with the modelling data which consists of a 2-hour count including speed data collected taken on 28th May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the sections on SR-99 motorway namely, S 102nd St and Cloverdale St. After obtaining the data in the raw from of .txt files, it was pre-screened and formatted as mentioned in Section 6.2.

As in the Section 6.4.1, after obtaining the processed data, the state space system was identified using MATLAB (as explained in Section 6.3). The transition matrix $A$ thus obtained can be stated as follows:

$$
\begin{pmatrix}
0.942 & 0.058 \\
0.880 & 0.110
\end{pmatrix}
$$

Matrix $A$ can be interpreted as, around 94.2% of the vehicles remains in the same lane in which they were initially travelling (the first lane), whereas the rest of the vehicles (5.8%) change the lane in due course of their travel. Similarly, for the second lane, around 88% of the vehicles remain in the same lane but 11% of them change lane.
(ii) **Application** – We collected a 2-hour count including speed data on 29th May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the sections on SR-99 motorway namely, S 102nd St and Cloverdale St. The collected traffic count and speed data was then used to estimate traffic densities in the time period using the model identified earlier. Figures 6.10 & 6.11 represent the density estimation during a non-peak traffic time scenario for a 2-lane motorway.

**Figure 6.10** The traffic counts using the developed method for 2-lane motorway using non-peak time real traffic data.

**Figure 6.8** The traffic counts using the Gazis & Liu (2003) method for 2-lane motorway using non-peak time real traffic data.

The plots in Figure 6.10 represent the estimation results from the developed model and the plots in Figure 6.11 depict the estimation results
from the Gazis & Liu (2003) model. Different lanes have been shown by using two different plots i.e. the left plot for the first lane and the right plot for the second lane in both the Figures. Clearly, the developed method provides more accurate results when compared to existing methodologies which ignore lane change behaviour on motorways.

6.4.2.2 Three-Lane Case

(i) Modelling - For 3-lane motorways, the modelling data was collected on 28\textsuperscript{th} May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the sections on I-5 namely of S Oregon St and S Holgate St. After obtaining the pre-screened and the processed data, we got the transition matrix \( A \) which represents the lane change behaviour over multi-lane motorways. The obtained transition matrix can be stated as follows:

\[
\begin{pmatrix}
 & & \\
 & & \\
& & \\
\end{pmatrix}
\]

In this case, matrix \( A \) can be interpreted as before that in the first lane, around 86.3\% of the vehicles remains in the same lane in which they were initially travelling, whereas the rest, around 13.7\% of the vehicles, change to the second lane in due course of their travel. Similarly, for the middle lane around 62.2\% of the vehicles remains in the middle lane, 11.6\% of them change to the first lane and around 26.2\% changes to the third lane. Lastly, in the third lane 73\% of the vehicles remain in the same lane, however 27\% change to the second lane. After obtaining the transition matrix, the next step is the model application.
(ii) Application – To test the model for a 3-lane structure, we collected a 2-hour data on 29th May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the cabinets on I-5 namely of S Oregon St and S Holgate St.

Using the above transition matrix and application data, Figures 6.12 & 6.13 represent the results of density estimation during peak time scenarios for a 3-lane motorway.

Figure 6.12 The traffic counts using the developed method for 3-lane motorway using non-peak time real traffic data.

Figure 6.10 The traffic counts using the Gazis & Liu (2003) method for 3-lane motorway using non-peak time real traffic data.

The plots in Figure 6.12 represent the estimation results from the developed model and the plots in Figure 6.13 depict the estimation results from the Gazis & Liu (2003) model. Different lanes have been shown by using different plots. In other words, the left and right plots represent first & third lane and the middle plot represent the middle lane. Estimation results are not in favour of the existing methodology as compared to the developed methodology and occasionally failed to estimate correctly via the KF technique.

Next, we will discuss the 4-lane case as we did in the peak traffic time real data analysis.
6.4.2.3 Four-Lane Case

(i) Modelling – In this case, the modelling data was collected on 28\textsuperscript{th} May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the sections I-90 including 246th Ave SE, EB and 256th Ave SE. After obtaining the conditioned data, we got the transition matrix $A$ which explicitly represents the lane change behaviour over multi-lane motorways. The obtained transition matrix can be stated as follows:

$$
A = \begin{pmatrix}
\end{pmatrix}.
$$

In this case, matrix $A$ can be interpreted that around 72.4\% of the vehicles remain in the same lane in which they were initially travelling, whereas the rest around 27.6\% of the vehicles, change to the second lane in due course of their travel. Similarly, for the second lane, around 37.6\% of the vehicles remain in the same (second) lane, 33.7\% of them change to the first lane. Further, around 28.7\% change to the third lane.

In the third lane, 33\% of the vehicles remain in the same (third) lane, 33\% changes to second lane and around 33\% changes to fourth lane respectively. Coming to the case of fourth lane, almost 40\% remain in the fourth lane itself however the rest 59\% changes to third lane. After obtaining the transition matrix, the next step is the model application.

(ii) Application – We collected a 2-hour count including speed data taken on 29\textsuperscript{th} May 2007 between 10:00 am & 12:00 noon (non-peak time traffic) from the sections I-90 including 246th Ave SE, EB and 256th Ave SE.
Using the 4-dimensional transition matrix obtained via the system identification technique and the application data, the results of density estimation during peak time scenarios for a 4-lane motorway are shown using the following Figures.

**Figure 6.11** The traffic counts using the developed method for 4-lane motorway using non-peak time real traffic data.

**Figure 6.12** The traffic counts using the Gazis & Liu (2003) method for 4-lane motorway using non-peak time real traffic data.

In the Figures 6.14 & 6.15, the plots in Figure 6.14 represent the estimation results from the developed model and plots in Figure 6.15 depict the estimation results from the Gazis & Liu (2003) model. Again different lanes have been shown by using different plots. The plots suggest that estimation results’ using the existing methodology completely lose the estimation track and provide unsatisfactory results (inaccurate) as compared to the developed methodology. The developed methodology results are promising due to the inclusion of the lane change behaviour unlike the existing methodologies (Gazis & Liu, 2003).
6.4.2.4 Summary (for non-peak traffic time)

From the above real traffic data analysis, it is clearly evident (from the Figures 6.10 - 6.15) that the developed (Markov chain based lane change) model provides satisfactory results for real traffic data on 2, 3, 4 – lane motorways. For all the three cases explained above it worked well and provided better results. Moreover, the developed model excels over the existing model (Gazis & Liu, 2003) is applicable to different multi-lane scenarios during non-peak-traffic hours.

6.5 Numerical Comparison

In this sub-section we will present a numerical comparison regarding the performance of the developed method and the existing methods (with Gazis & Liu, 2003). It have been discussed earlier (see Chapter 1 for details) that traffic densities cannot be measured directly. Since the time step of 20 seconds is too short, so it would be too costly if we want to install ILDs at a distance of every 30-50 meters so as to use the retroactive forecast approach (for details, see Mason & Mimmack, 2002) to validate the results. Hence, in general it is quite difficult to get hold of actual densities. Thus, to handle this issue we have used the fault detection process approach using ‘innovation process’ technique to compare the performance of developed method with that of Gazis & Liu (2003).

The innovation process (or error signal) is defined as the difference between the actual system output and the expected output based on the model and the previous output data. The latter is generated directly by a statistical filter (in case the system is stochastic), i.e. subject to random inputs and variations. It is called the innovation process since it represents the new information brought by the latest observation. Under normal conditions, the error signal is "small" and corresponds to random fluctuations in the output since all the
systematic trends are eliminated by the model. However, under faulty conditions, the error signal is "large" and contains systematic trends because the model no longer represents the physical system adequately (Mehra & Peschon, 1971).

The statistics of the innovation process (error signal) are obtained from the filter which is used to predict the output of the system.

Kalman Filters can be used to recover the possible malfunctions in the estimation system (Sasaidak & Wang (1999), Zhang & Wei (2003)). For both linear and non-linear dynamic systems with Gaussian random inputs, a Kalman filter generates both the innovation sequence and its statistics (Kailath 1968, Okatan et al. 2007). Faults in multidimensional dynamic systems can be detected with the aid of an innovation sequence of Kalman filter. Generally, fault detection algorithms developed to check the statistical characteristics of the innovation sequence in real-time are based on the following fact. If a system of estimating operates normally, the normalized innovation sequence in the Kalman filter coordinated with a dynamics model represents the white Gauss noise with zero average value (and unitary covariance matrix). Any change in these statistical characteristics of the normalized innovation sequence suggests that the system is imperfect and this may caused due to a variety of problems (Okatan et al 2007). In our case, the non-accountability of probabilities of lane change in the existing methodologies is the reason of imperfectness (for Gazis & Liu, 2003)

Mathematically, innovation sequence ( and normalised innovation sequence ( ) can be defined as follows:

\[ 
\text{innovation sequence} = y_t - \hat{y}_t \\
\text{normalised innovation sequence} = \frac{y_t - \hat{y}_t}{\sigma_t} 
\]
where have been defined same as in Chapter 3.

The mean of normalised innovation sequence ( ) is estimated as

where \( N \) is the number of observations.

The respective values of mean of normalised innovation sequences have been presented in following tables 6.3 – 6.4 for 2-lane and 4-lane motorways respectively.

Table 6.3 Mean of normalised innovation sequence values considering 2-lane motorways for developed and Gazis & Liu, (2003) method

<table>
<thead>
<tr>
<th></th>
<th>Developed Method</th>
<th>Gazis &amp; Liu Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lane-1</td>
<td>Lane-1</td>
</tr>
<tr>
<td>Peak</td>
<td>0.6670</td>
<td>26.7818</td>
</tr>
<tr>
<td>Non-Peak</td>
<td>0.2929</td>
<td>26.6905</td>
</tr>
</tbody>
</table>

Table 6.4 Mean of normalised innovation sequence values considering 4-lane motorways for developed and Gazis & Liu, (2003) method

<table>
<thead>
<tr>
<th></th>
<th>Developed Method</th>
<th>Gazis &amp; Liu Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lane-1</td>
<td>Lane-1</td>
</tr>
<tr>
<td>Peak</td>
<td>1.8652</td>
<td>27.3079</td>
</tr>
<tr>
<td>Non-Peak</td>
<td>2.0105</td>
<td>26.5674</td>
</tr>
</tbody>
</table>

The values from the above tables are self-explanatory and show the difference between the estimated and actual speed values. For instance, on comparing peak-time mean of normalised innovation sequence values results under 2-lane system, it is clearly evident that the dynamic system under the developed method is more close to normal (zero mean) conditions (as explained above) as compared to the Gazis & Liu (2003) results. This further indicates that the lane change probabilities do account for
the estimation of vehicle densities. The lower mean value depicts the more normal-like and higher values as in the case of Gazis & Liu (2003) method is due to the ignorance of the lane change behaviour which eventual effect the performance of the Extended Kalman filter.

6.6 Summary

In this chapter, we have presented an important element of this research study. We have gathered data from different motorways of the USA and chose specific sections to get the data in the required format. After acquiring data from ILDs, we identified the state space system and applied the KFT to estimate the vehicular densities on multi-lane motorways. Results obtained are promising and satisfactorily explain the contribution of the developed model to the traffic studies. We have shown that the developed methodology in this research is applicable over motorways under suitable conditions for variable traffic times, i.e. peak time traffic and non-peak time traffic. It was shown that no matter what the lane structure is, the developed model always exceeded the existing methodologies since lane change had been ignored in them (existing methodologies).
CHAPTER VII

SUMMARY AND CONCLUSIONS

7.1 Summary

In this research study, we developed a statistical model for traffic flow using the state space modelling approach, and the Kalman filtering approach to estimate traffic flow on the multi-lane motorways. The recursive estimation of traffic flow (density) is investigated using a vehicular speed-density relationship. Using traffic information from ILDs (embedded in the pavements), the KF method is employed to improve the accuracy of vehicle count estimates. The proposed recursive estimation method uses a suitable transformation so that the state space system becomes linear to avoid the linearisation of observation equations used in the existing method (Singh & Li, 2010). Following the work of Gazis & Liu (2003), it is also shown that the estimated densities can be improved by considering the fact that the counting error for vehicles leaving a given section is the same as the error for the vehicles entering the very next section in tandem. The transformed equation minimises the chances of approximation of density estimates at the filtering step; hence it is not surprising that the developed method has a better performance than the existing methods.

Further, to present a more realistic picture of the multi-lane freeways’ traffic situation, we included the lane-change behavior of the drivers in the existing methodology and newly developed model. We used the Markov chain processes to describe lane-change behavior so that the state equation can better reflect the movements of real traffic flow between lanes. States of Markov chain represent different lanes of the freeways. We assume that each vehicle in the motorway segment has a certain probability to stay in the
current state (lane), or to change from one state (lane) to another.

So, the above Markov chain based lane-change model estimated the
densities considering that drivers are undergoing lane-changes during the
estimation time period (20 seconds) and that lane-changes are only allowed
between the next nearest/adjacent lane and are not allowed to other lanes.
This is due to the fact that the time period of 20 seconds under the current
system is too small to make any subsequent lane change. However, under
other systems where the time period is larger than 20 seconds (say 90
seconds or 150 seconds), it is possible to make another lane change. The
transition matrix of the state space model represents the lane-change
probabilities. The performance of this approach has been investigated in the
simulation studies. It is shown that, on comparing with the existing model
used in Gazis & Liu (2003), the improvements for the state space model on
the state equation have substantially increased the accuracy of the traffic
density estimation.

In the above model, we assumed that the lane-change probabilities are
fixed (not time varying). In other words, we can say that the traffic condition
is stable inside the particular segment. However this may not be true in
practice as there are many external factors which may affect this
assumption. To overcome this drawback, we further extended the above lane
change model where we dropped the assumption of stable traffic flow and so
these lane-change probabilities are not assumed to be fixed but rather they
depend on the current traffic condition. Instead we estimate the transition
matrix representing the lane-change manoeuvre. We modeled the drivers’
lane-change behavior using discrete choice methodology. The drivers’
choice probabilities are modeled using several traffic-condition related
attributes such as vehicle time headway, traffic density and speed. Hence the
matrix of transition (lane-change) probabilities forms our state matrix and
developed state space model is thus used to estimate the vehicular traffic densities on motorways.

This research study is based on the availability of traffic data measured by ILDs. In practice, loop detectors are widely deployed in strategic motorway networks so potentially the developed models can be widely applied as an effective approach to traffic surveillance. The estimated traffic density can facilitate effective traffic management of networks, and also provide inputs for the planning and controlling of both short term and long term urban transport.

The simulation and empirical studies suggest that the developed models perform satisfactorily in different scenarios. Moreover these developed models can be applied to estimate traffic density for multi-lane multi-segment motorways.

7.2 Conclusions

This thesis resulted in a number of conclusions which are listed as follows:

- The existing studies mainly focus on scenario where a vehicle’s lane-change movements are not common in a motorway and can thus be ignored. This attitude does not reflect the true movements of real traffic flow between various lanes. Studies carried out in this research take into account the lane-change behaviour and propose a Markov chain model for drivers’ lane-change movements and incorporate a Markov chain model to describe the lane-change behaviour assuming stable traffic conditions. We assumed the fixed lane change behaviour inside the selected segment of motorway for this research.

- Further, the assumption of stableness is dropped and the lane-change probabilities between two adjacent lanes are not assumed to be fixed anymore, but rather they depend on the current traffic condition such as
vehicle time headway, traffic density and speed. A discrete choice model is used to capture drivers’ lane choice behaviour. Hence, the developed model uses real time information rather than just fixed parameters as found in literature.

- The newly developed UKF approach is used to tackle the issue of high non-linearity of the state space model
- This study also incorporates a suitable transformation to deal with the non-linear observation equation. In this way biasness due to approximation can be avoided when using the KF to estimate the traffic density.

Moreover, the models developed in this research are applicable to a wider range of traffic infrastructure including multi-lane multi-section motorways. The numerical studies presented the accuracy and efficiency of the proposed models as compared to the existing methodologies and hence can be used as future density estimation models for multi-lane motorways. Results also proved the robustness of the developed models as they are applicable under variable road traffic conditions.

Great care must be taken while discussing the applicability of the developed models includes the parameters’ (such as free speeds and threshold density) values in the different lanes. These parameters are dependent on the individual motorways and represent local conditions. As mentioned in Chapter 5, we followed the scheme used by Gazis & Liu, (2003) and Wang and Kao (1990) regarding the choice of these parameters.

7.3 Future Work

The developed models are efficient in estimating densities from ILD data but there is a need to explore a couple of issues so that they can be applied to more comprehensive motorway structures which represent more sophisticated scenarios. We have several things to consider in the future studies which can be listed under the following headings.
• Possible extensions to the current developed model.
• The estimation/prediction of other traffic variables such as vehicle speed and travel time.
• The use of other filtering techniques such as non-linear filters, particle filters to compare the performances of the developed models.

We will briefly discuss the above mentioned points in the following subsections.

7.3.1 Possible Extensions Of The Current Developed Model

Firstly, the performance of the developed methodology is worth exploring under congestion-like scenarios where we have sections densely populated with vehicles. In this case, changing between lanes is quite difficult and hence the effect of lane-changing behaviour may be null on the estimated densities. Moreover, the relationship between speed-density would also get affected with the increase in the number of vehicles in the section for a particular lane. How this may be handled is a question for future research.

Secondly, scenarios need to be considered where one of the lanes within the motorway section is blocked due to some reason (accident, road works, vehicle breakdown, etc.). In this case we would not have values of counts for the few time steps which fall in the blocked part of the lane. Therefore, it is difficult to handle the situation when it comes to the estimation of densities for the given section of the freeway. So, there is a need that a proper algorithm would be made to include these rare but real-time situations which may affect the final results drastically.

Thirdly, there is scope in the research to consider entry and exit of vehicles via entry and exit roads. In the present model we have not included this possibility but it can be accommodated and researched further. In other words we can say that we need an extended ‘\( w \)’ in the state equation which
can accommodate the incoming and outgoing counts or the net gain via ramp-in and ramp-out.

Suppose, we have ‘ramp gain’ as the net gain of the ramp-in-out count, then the new term ‘w*’ in the state equation can be written as

\[ w^* = w + \text{ramp gain}, \]

and the resulting state equation becomes

In this way this model can be generalised to represent more comprehensive and complicated motorway structures.

### 7.3.2 The Estimation/Prediction Of Other Traffic Variables

Next is the estimation of other traffic variables using the estimated densities from the developed methodology in this research. Traffic density is implicitly related to travel times, incidents and vehicle speeds in the sense that the change in one would ultimately affect the others. In today’s fast paced hi-tech world, accurate estimation of travel times over motorways has become an important issue. How to utilise the existing surveillance infrastructure for better travel time estimation is an urgent issue for future research.

Traffic density estimates can be used to estimate respective travel times in motorway networks. Based on the concept of the section density Oh et al. (2002) presented an algorithm using a hydrodynamic kinematic traffic model. Traffic densities are also used as a state equation in a state space model to estimate travel times using adaptive KF by Chu et al. (2005) that incorporates two data sources, i.e. point detector data and area wide probe-data. Hence, there is scope to explore the developed models further so that they would be applicable for the estimation of travel times over the motorways.
Lane-change manoeuvres or a driver’s lane choice model can be taken further to predict incidents and make further assessments. The improvement of lane-change behaviour or explicit knowledge of this traffic variable would give more chances to understand real traffic scenario and hence proper steps or measures could be taken to reduce the number of incidents or to assess them if they had already happened.

7.3.3 Use Of Other Filtering Techniques

In this research, to estimate the state space model, we have used the KFT which poses the limitation to the underlying model to be of linear nature. However, the dynamic and non-linear nature of traffic related problems would suggest that we should further explore other filtering methodologies such as particle filter also known as Sequential Monte Carlo methods (SMC). Particle filters are the sequential analogue of Markov chain Monte Carlo (MCMC) batch methods. It is believed that well-designed particle filters are much faster than MCMC. Studies also reveal that they are often used as an alternative to the EKF or UKF with the advantage that, with sufficient samples, they approach the Bayesian optimal estimate, so they can be more accurate than either the EKF or UKF. Some studies have used particle filters for estimating traffic variables but nobody has used it (as per our knowledge) for the density estimation. To employ this filter it may be our next step towards more accurate and robust methodology for density estimation. However it is still a topic which needs more exploration and better understanding before its application can be made into current research themes.


Markov process (mathematics) - Britannica Online Encyclopedia. accessed on 13/07/2010.


