Evaluating Fluctuations in Urban Traffic Data and Modelling Their Impacts

Author:
Chakka, Mohana Naga Sai Chand

Publication Date:
2019

DOI:
https://doi.org/10.26190/unsworks/3659

Evaluating Fluctuations in Urban Traffic Data and Modelling Their Impacts

By

Sai Chand

M. Tech. (Transportation), B. E. (Civil)

A thesis presented in the fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Civil and Environmental Engineering

Faculty of Engineering

The University of New South Wales

March 2019

THESIS/DISSERTATION SHEET

THE UNIVERSITY OF NEW SOUTH WALES

Thesis/Dissertation Sheet

Surname or Family name: CHAKKA

First name: MOHANA NAGA SAI CHAND

Other name/s:

Abbreviation for degree as given in the University calendar:

School: Civil and Environmental

Faculty: Engineering

Title: Evaluating Fluctuations in Urban Traffic Data and Modelling Their Impacts

Abstract 350 words maximum: (PLEASE TYPE)

Predictability is important in decision making in many fields, including transport. The fluctuations of time-varying and short-term traffic processes pose critical challenges for real-time traffic predictions, which are the backbone of the intelligent transportation systems (ITS). The success of different prediction techniques depends on the structure of the phenomenon, in particular whether it is easily predictable or complex. Therefore, understanding the various aspects leading to high/low complexity in traffic time series data is critical in evaluating the performance of urban traffic systems.

This thesis presents the application of the Hurst exponent metric from Fractal theory in quantifying fluctuations in different micro and macroscopic parameters of urban traffic. Data from three distinct sources and geographical locations is used for the analysis.

i) Vehicle trajectory data collected every 0.5sec from an arterial road in India and a freeway in the USA are utilised to evaluate fluctuations in lateral movement and speeds of several vehicles. The effect of the average lateral position and vehicle type on fluctuations is also examined.

ii) Loop detector data collected every 30sec from several monitor sites on an urban motorway in Australia is used to quantify fluctuations in speed and flow. The effect of the time of the day, weekend, proximity to ramps, and the presence of a horizontal curve, on fluctuations is explored using the analysis of variances. Further, latent class and random parameters Tobit models are estimated to predict the effect of the Hurst exponent (long-range dependence) of speed on crash rates at motorway sites.

iii) Traffic count data collected every 5min at several signalised intersections in Sydney, Australia is studied to understand the patterns of predictability. The effect of the day of the week, public holidays, special events, weather, etc. on predictability is discussed using a random effects linear regression model.

This thesis is data driven; the empirical results suggest the need to evaluate fluctuations in traffic time series data before using different prediction and simulation techniques. Furthermore, this thesis could lead to many applications of fractal analysis on highways and urban traffic.

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ………………………………………………..............

Date ………………………………………………..............

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed ……………………………………………..........................

Date ……………………………………………..........................

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed ……………………………………………..........................

Date ……………………………………………..........................

INCLUSION OF PUBLICATIONS STATEMENT

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not.

☐ *This thesis contains no publications, either published or submitted for publication (if this box is checked, you may delete all the material on page 2)*

☑ *Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement (if this box is checked, you may delete all the material on page 2)*

☐ *This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below*

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Name	Signature	Date (dd/mm/yy)
SAI CHAND		

Postgraduate Coordinator's Declaration (to be filled in where publications are used in lieu of Chapters)

I declare that:

- the information below is accurate
- where listed publication(s) have been used in lieu of Chapter(s), their use complies with the Thesis Examination Procedure
- the minimum requirements for the format of the thesis have been met.

PGC's Name	PGC's Signature	Date (dd/mm/yy)

ABSTRACT

Predictability is important in decision making in many fields, including transport. The fluctuations of time-varying and short-term traffic processes pose critical challenges for real-time traffic predictions, which are the backbone of the intelligent transportation systems (ITS). The success of different prediction techniques depends on the structure of the phenomenon, in particular whether it is easily predictable or complex. Therefore, understanding the various aspects leading to high/low complexity in traffic time series data is critical in evaluating the performance of urban traffic systems.

This thesis presents the application of the Hurst exponent metric from Fractal theory in quantifying fluctuations in different micro and macroscopic parameters of urban traffic. Data from three distinct sources and geographical locations are used for the analysis.

i) Vehicle trajectory data collected every 0.5sec from an arterial road in India and a freeway in the USA are utilised to evaluate fluctuations in lateral movements and speeds of thousands of vehicles. The effect of the average lateral position and vehicle type on fluctuations is also examined.

ii) Loop detector data collected every 30sec from monitor sites on an urban motorway in Australia are used to quantify fluctuations in speed and flow. The effect of the time of the day, weekend, proximity to ramps, and the presence of a horizontal curve, on fluctuations is explored using the analysis of variances. Further, latent class and random parameters Tobit models are estimated to predict the effect of the Hurst exponent (long-range dependence) of speed on crash rates at motorway sites.

iii) Traffic count data collected every 5min at signalised intersections in Sydney, Australia are studied to understand the patterns of predictability. The effect of the day of the week, public holidays, special events, weather, etc. on predictability is discussed using a random effects linear regression model.

This thesis is data-driven; the empirical results suggest the need to evaluate fluctuations in traffic time series data before using different prediction and simulation techniques. Furthermore, this thesis could lead to many applications of fractal analysis on highways and urban traffic.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my supervisors, Associate Professor Vinayak Dixit and Professor Travis Waller. It was a wonderful experience working under their supervision. I will be grateful to them for giving me the freedom to explore a topic of my interest. Travis and Vinayak involved me in several research projects through which I refined my technical skills. I thank Travis for encouraging and funding me to attend several international and domestic conferences. Vinayak's energy and never-say-no attitude always kept me positive about research and other things in life. I always cherish the moments working under his supervision.

Thanks to Dr. Lauren Gardner, Dr. Taha Rashidi, Dr. David Rey, Dr. Emily Moylan, Dr. Hanna Grzybowska, Dr. Zhitao Xiong, and other staff at rCITI for their comments and suggestions during my progress reviews. Emily was especially conscientious in sharing safety data for this thesis. Special thanks to Maria Lee for her administrative and infrastructural help. Thanks to Gregory Aouad and Ankit Agrawal for their help in data processing.

I would like to thank Divya Nair, Ted Robson, Neeraj Saxena, and Shantanu Chakraborty for their love, support and of course so many lively conversations. I would like to acknowledge Neeraj for many valuable technical suggestions as well. Cheers to my mates: Abdul, Ahmed, Ali, Amarin, Chenyang, Fernando, Greg, Hassan *bhai*, Kasun, Kiran, Malith, Melissa, Milad, Nima, *Prince* Raed, Raghavi, Sisi, Tao, Xiang.

I would also like to thank Mrs. Padma Arasaratnam, my lessor for her grandmotherly affection and welcoming nature. I thank Divya, Prasanna and Aishu for their love.

I will always be grateful to my teachers at IIT Roorkee, Professor Satish Chandra, Professor Manoranjan Parida and Dr. Indrajit Ghosh for motivating me to pursue PhD in this beautiful field of transportation.

My heartfelt regard goes to my aunts, uncles, and cousins, for their love and encouragement. My deepest gratitude to my wife, Sonali for her unending care and understanding of my goals and aspirations which kept me positive throughout. I am lucky to have wonderful parents, Dhana Lakshmi and Vidya Sagar and a loving brother, Sai Charan, who are always there to support and stand by my side. This journey would not have been possible without the support of my family.

LIST OF RELEVANT PUBLICATIONS AND AWARDS

The following provides a list of the conference and journal publications that have contributed towards the development of this thesis.

Journals (published):

1. **Chand, S.,** Dixit, V.V., 2018. Application of Fractal theory for crash rate prediction: Insights from random parameters and latent class Tobit Models. *Accident Analysis and Prevention 112,* pp 30-38.
2. **Chand, S.,** Aouad, G., Dixit, V.V., 2017. Long-range dependence of traffic flow and speed of a motorway: Dynamics and correlation with historical incidents. *Transportation Research Record: Journal of the Transportation Research Board 2616*, pp 49-57.
3. **Chand, S.,** Dixit, V.V., Waller, S.T., 2016. Evaluation of fluctuating speed and lateral movement of vehicles: Comparison between mixed traffic and homogeneous traffic. *Transportation Research Record: Journal of the Transportation Research Board 2581*, pp 104–112.

Conferences:

1. **Chand, S.**, Theofilatos, A., Dixit, V.V., 2018. A novel application of Catastrophe and Fractal theories for crash and incident modelling. Presented at the 97th Annual Meeting of the Transportation Research Board (TRB), Washington DC, USA.
2. **Chand, S.**, Dixit, V.V., 2018. Latent class Tobit modelling of crash rates. Presented at the 97th Annual Meeting of the Transportation Research Board (TRB), Washington DC, USA.

3. **Chand, S.**, Aouad, G., Dixit, V.V., 2017. Long-range dependence of traffic flow and speed of a motorway: Dynamics and correlation with historical incidents. Presented at the 96th Annual Meeting of the Transportation Research Board (TRB), Washington DC, USA.
4. **Chand, S.**, Dixit, V.V., Waller, S.T., 2016. Evaluating speed and lateral movement fluctuating behavior of vehicles: Comparison between mixed traffic and homogeneous traffic. Presented at the 95th Annual Meeting of the Transportation Research Board (TRB), Washington DC, USA.

Journal papers under preparation:

1. **Chand, S.**, Theofilatos, A., Dixit, V.V., 2018. A novel application of Catastrophe and Fractal theories for crash and incident modelling. *Accident Analysis and Prevention (submitted)*.
2. **Chand, S.,** Dixit, V.V., Waller, S.T., 2018. Modelling predictability of intersection traffic counts. *Journal of Transportation Engineering-Part A (under preparation).*

Award:

Recipient of the David Willis Honourable Prize for the best poster at the Australasian Transport Research Forum (ATRF), Sydney, 2015 for his collaborative work on the development of Dynamic Traffic Assignment Model of Sydney.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADT --- average daily traffic

AADT --- average annual daily traffic

AIC --- Akaike information criteria

AID --- automatic incident detectors

ANOVA --- analysis of variance

ARIMA --- autoregressive moving average

ATMS --- advanced traffic management systems

ATIS --- advanced traveller information systems

AVL --- automatic vehicle location

BIC --- Bayesian information criteria

BOM --- Bureau of Meteorology

CA --- cellular automata

CBD -- central business district

DFA --- detrended fluctuation analysis

DSRC --- dedicated short-range communications

FD --- fractal dimension

FHWA --- Federal Highway Administration

GARCH --- generalised autoregressive conditional heteroskedasticity

GLS --- generalised least squares

HSD --- honest significant differences

ITS --- intelligent transportation systems

KNN --- k-nearest neighbours

LCT --- latent class Tobit

LCV --- light commercial vehicles

LRD --- long-range dependence

MLE --- maximum likelihood estimation

NB --- negative Binomial

OLS --- ordinary least squares

PDF --- probability density function

PI --- Predictability Index

NGSIM --- next generation simulation

RELR --- random effects linear regression

RPT --- random parameters Tobit

SARIMA --- seasonal autoregressive moving average

SCATS --- Sydney coordinated adaptive traffic system

SDIC --- sensitive dependence on initial conditions

TDM --- travel demand management

TMC --- transport management centre

TRAZER --- Traffic Analyzer and Enumerator

VMS --- variable message sign

VSLS --- variable speed limit signs

CHAPTER 1. INTRODUCTION

1.1 BACKGROUND

Traffic congestion can have significant negative impacts on the economy, environment, transport network reliability, and performance. Traffic congestion is estimated to cost around $16.5 billion per year in damages for Australian capital cities (BITRE 2015). The cost comes from vehicle delays (lost productivity), travel time variability, fuel consumption, and air pollution. Investigating ways to reduce the impact of congestion will be increasingly important, as projected population growth and increased motorisation will increase congestion and costs.

Transport authorities employ a variety of strategies focussing on managing both the demand and supply side of transport systems. Supply management strategies focus on building new roads or investing otherwise in quality infrastructure; such strategies tend to be expensive. On the other hand, travel demand management (TDM) strategies focus on reducing the number of low-occupancy trips, by encouraging off-peak travel, promoting active transport modes (biking and pedestrianisation), incentivising the use of public transport, creating public awareness campaigns, and by giving real-time information so that drivers can make better travel decisions.

In the recent years, data science has taken the world by storm because of its ability to extract insights from data in various forms through the unification of statistics, machine learning, visualisation, and computer science. Therefore, many government agencies are increasingly embracing ultramodern technologies and setting up high-quality infrastructure to gather numerous forms of data in the hopes of saving money, becoming cleaner, reducing congestion, and improving quality of life. Such data are becoming readily available to the public as the city authorities are more and more open to suggestions from researchers and wide sections of communities. The processing and analysis of huge volumes of data have become significantly easier than they used to be a decade or two earlier.

With regard to traffic management, several cities have been installing detection systems, such as loop detectors, sensors, and automatic vehicle location (AVL) systems, to gather traffic data such as speeds, volumes, vehicle type, etc. Such data will help authorities in understanding driver behaviour and in predicting future traffic conditions. In fact, predictions have become a critical component for many Intelligent Transportation System (ITS) applications such as Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS) (Shang et al. 2006).

In the past, researchers focussed mainly on planning models where the emphasis used to be on long-term predictions such as forecasting vehicle ownership, oil prices, etc. On the other hand, studies on non-recurrent congestion, short-term traffic flow and speed predictions were only a handful. The increased levels of vehicle ownership and the ensuing congestion made traffic authorities and researchers realise the significance of predicting short-term processes and proposing proper traffic management strategies. Therefore, several research studies have appeared in the recent past on the topic “short-term traffic” in the field of transportation (please see Figure 1-1). This surge could also be due to the significant increase in state-of-the-art data collection technologies, smart infrastructure and emergence of big data analysis techniques.

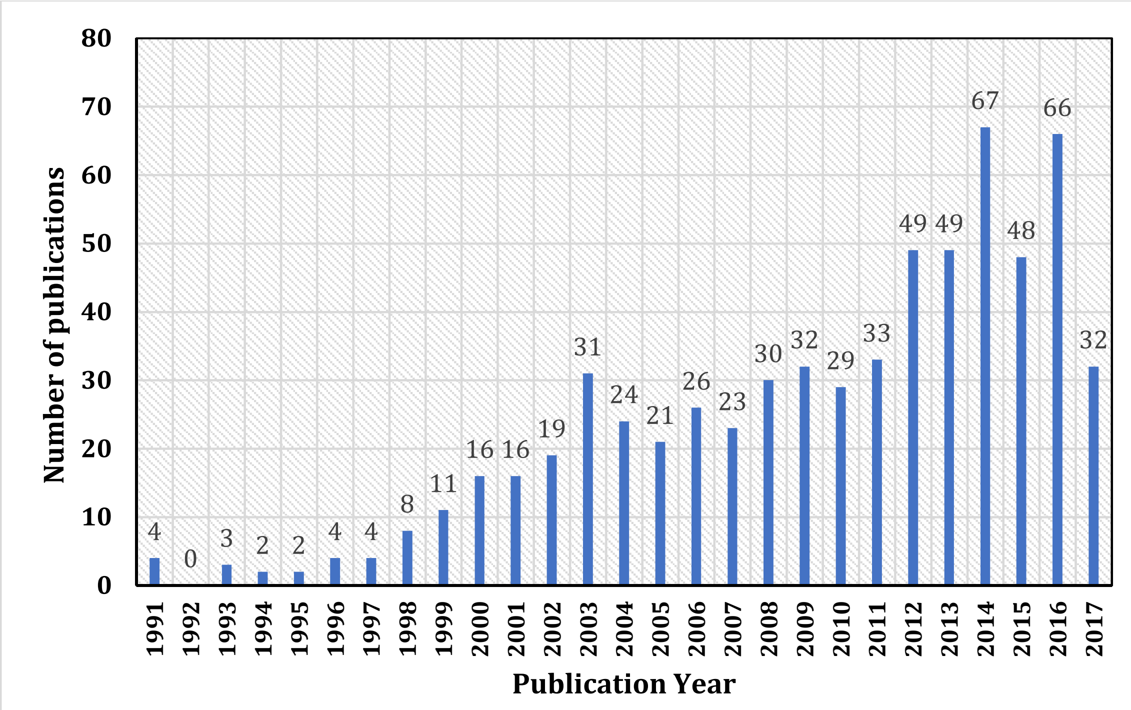


Figure 1-1 Timeline of the number of scientific publications on the topic "short-term traffic" {Source: Web of science}

Nowadays, short-term traffic predictions are of substantial commercial interest for some leading tech players in the market (Upadhyay 2016; Bell 2017). For example, traffic navigation applications such as Google maps, TomTom, and Here maps make short-term predictions of traffic conditions and provide route guidance to users. Poor prediction of such short-term processes could lead users to lose trust in the navigation applications. Therefore, these companies put much effort in short-term predictions, through a variety of strategies including machine learning and big data analytics. In a nutshell, short-term predictions are moving from "unimportant" to "important" (van Zuylen et al. 1999).

There are many techniques such as neural networks, autoregressive moving average (ARIMA), seasonal ARIMA (SARIMA), generalised autoregressive conditional heteroskedasticity (GARCH), k-nearest neighbours (k-NN), and support vector regression (SVR) to predict future values of an entity. The fit of these prediction models is often vastly reliant on the specific data available; prediction algorithms that work well in one context may not work well in other contexts (Tiesyte and Jensen 2009). Besides, the success and transferability of these mathematical models remain on the structure of the particular phenomenon, whether it is readily predictable or not (Turkay 2014). For instance, if the prediction obtained from a specific method is weak, but the time series contains an excellent predictive structure, one can practically conclude that the employed prediction technique is unsuitable to the task and that one should try a different technique (Garland et al. 2014). Therefore, it is paramount to evaluate whether the structure of evolution of an entity (for example, speed, travel time, traffic counts, etc.) is complex or easily predictable before resorting to advanced prediction techniques.

1.2 MOTIVATION

Traffic system is dynamic and is traditionally a complex system, because of the continuous interactions between different vehicles (Nair et al. 2001). These interactions, coupled with changes in road geometry, weather and road surface conditions, result in braking (fluctuations in speeds of individual vehicles) and lane changing (fluctuations in lateral movement) of vehicles, which eventually lead to fluctuations in macroscopic variables such as; speed, flow, and occupancy. These fluctuations will have implications for road safety, crashes, flow predictions, travel time and the reliability of these system performance indicators. For a transport system model to be efficient and reliable, it is essential to understand these fluctuations, so as to simulate and forecast traffic system performance. Further, analysis of fluctuations in traffic data can offer some interesting insights on congestion patterns, and their relationships with road geometry, time of the day, incidents, crashes, and weather.

If a dataset in hand has low predictive structure (in other words, random fluctuations), the ability to predict future values will be low as well. Therefore, predictability is considered necessary for decision making in many fields, including transport. The uncertainty (or ill-predictability) of time-varying and short-term processes such as non-recurrent congestion poses severe problems to traffic authorities. When observed at wider intervals (say 1 hour, 2 hours, etc.) and for a longer duration (1 month, 6 months, etc.), one may not observe significant disorder in traffic flow data. However, with short-term traffic flow data, rapid and intense fluctuations is a widespread problem. Smoothing of these fluctuations may lead to inappropriate predictions (Breslin and Belward 1999). To better understand the traffic dynamics, it is more practical to study the irregularities as they are, by assessing their magnitudes instead of smoothing.

Furthermore, distinguishing between several nonlinear behaviours (by assessing their complexity) is crucial as often the more random the process is, the more ‘‘nonlinear’’ the prediction methods should be (Li and Shang 2007). Stated differently, quantification of how much the process departs from a linear one is critical to decide whether linear methods with small corrections can be used, or what sort of more complex models are needed. At first look, some dynamical systems may appear to be regular and periodic, whereas others will appear strictly random; in both cases, closer examination topples these assumptions (Frazier and Kockelman 2004).

The power spectrum analysis (Fourier transformation) is a common practice to evaluate irregular time series data. When the power spectrum follows the power law; $P(f) \propto f^{-\alpha}$, the exponent α is considered the index for representing the irregularity of a time series. However, the main limitation of Fourier transform lies in the assumption that the fluctuations are stationary over a long interval where an ensemble average of the power spectra is taken to get a stable power law index. Nevertheless, the statistical characteristics of fluctuations often vary for a short time interval, and therefore it is not appropriate to take an average of the power spectra over a long interval (Higuchi 1988). Furthermore, ARIMA and GARCH models are some of the widely used prediction techniques because of their well-established theoretical foundation and ease of application. Nevertheless, they cannot capture the long memory properties and do not jointly treat the mean and variance (variability) of a time series (Karlaftis and Vlahogianni 2009). Treating both mean and variance of a time series is important as it ensures that the analysis takes care of the ordering of the time series. Therefore, there is a need to analyse the fluctuations using alternative techniques.

Fractal analysis of the Chaos theory[1] is one such technique to evaluate the complexity of time series. One of the main advantages of the fractal analysis is its ability to find patterns in large and complex datasets. The underlying concept of the Fractal theory is that many natural phenomena are better described using a non-integer dimension, called fractal dimension (FD). This dimension depends on the complexity of the shape of the object, i.e. a shape with a higher FD is more

[1] More about this in the 2nd Chapter of this thesis.

complicated or rough than the one with a lower FD and fills more space (Breslin and Belward 1999).

The Hurst exponent is the most widely used method to calculate the FD of a time series. The Hurst exponent evaluates the long-range dependence[2] (LRD) (also called long memory) of a time series, by estimating the relationship between the decreasing rates of autocorrelations and the increasing lag between pairs of values (Hurst 1951, 1956). It provides a measure of whether the data of a time series is a pure white noise random process or has underlying trends. The main advantage of using the Hurst exponent rather than the typical coefficient of variation for the measurement of fluctuations is that the former is a numerical representation of the randomness through the history of a dynamical process, while the latter is a statistic that is independent of the temporal evolution (Valle et al. 2013).

Current applications of Fractal and Chaos theories are restricted to inter-city freeways where the congestion problems are less severe than say arterial roads. The main contribution of this thesis is the application of the Hurst exponent metric of Fractal theory to evaluate fluctuations in urban traffic data that include roadways other than freeways. As mentioned throughout this thesis, a sound understanding of the complexity of datasets is needed before using different prediction techniques. The research provides novel insights on fluctuations in

[2] A phenomenon is usually considered to have LRD if the dependence (autocorrelation) decays more slowly than an exponential decay, i.e. a power-like decay {Short-range dependence (SRD)}.
If the traffic flow data is not completely random (which means the data of t0 is in correlation with t0 +t, or the time series data has t-long memory), then we can use the correlationship to infer the variation trend of data t0 +t using data t0. On the other hand, if the traffic flow data is completely random, the existing traffic flow time series data may not be of much use to predict the future traffic flow variation trends (because there is no correlation between the data of t0 and t0 +t) (Yuan and Lin 2017).

trajectory data, loop detector data, and intersection count data, which were collected at different time resolutions, i.e. 0.5-sec, 30-sec, and 5-min respectively.

1.3 CONTRIBUTION

The contributions of this thesis can be summarised as follows.

- Comparing fluctuations in lane-based and non-lane-based traffic conditions in the USA and India, respectively.
- Evaluating the effect of the vehicle type and the average lateral position on fluctuations in speed and lateral movement.
- Evaluating and visualising spatiotemporal patterns of fluctuations in loop detector data of an urban motorway.
- Analysing the effects of the time of the day, day of the week, proximity to ramps and geometry on fluctuations in macroscopic traffic variables, namely speed and flow.
- Using the long-range dependence property of speed in crash rate modelling.
- Proposing a latent class Tobit model to estimate historical crash rates and comparing the results with a random parameters Tobit model.
- Evaluating fluctuations in traffic counts at several signalised intersections for one year.
- Estimating a random-effects linear regression model to quantify the effects of the day of the week, public holidays, special event days, weather, and parking on the theoretical predictability of traffic counts.

1.4 THESIS ORGANISATION

This thesis is data-intensive and analyses fluctuations using different datasets. Chapter 2 presents the methodology, Chapters 3 to 6 are the core chapters of this thesis, and Chapters 1 and 7 introduce and conclude the thesis respectively. The following sections provide an overview of the chapters.

1.4.1 Chapter-2: Methodology

As noted earlier, evaluating fluctuations in urban traffic data is important for selecting appropriate short-term prediction techniques and devising effective traffic management strategies. This chapter introduces the concepts of Chaos and Fractal theories, the Hurst exponent metric, and various methods to estimate the Hurst exponent. Furthermore, this chapter reviews several studies applying the Chaos and Fractal theories within the transport and traffic domain.

1.4.2 Chapter-3 Fluctuations in Microscopic Traffic Variables: An Application of Vehicle Trajectory Data

Urban roads in most developing countries are characterised by mixed traffic conditions, resulting in the complex interactions among various kinds of vehicles. Drivers show poor lane discipline and thus the vehicles laterally traverse on whatever road space that is available. Further, driver and vehicle heterogeneity in mixed traffic lead to frequent braking and lane changes, thus affecting the quality of traffic flow. These lane change manoeuvres create voids in traffic streams and could reduce travel speed and roadway capacity (Coifman et al. 2005; Laval and Daganzo 2006). Furthermore, a lane change manoeuvre could perturb the spacing-speed relation of the vehicles immediately following the manoeuvre and hence could create oscillations (Mauch and Cassidy 2002; Wang and Coifman 2008).

Traditional car-following[3] and lane-changing[4] models are not applicable to the mixed traffic conditions unless modifications accounting heterogeneity are made.

A vehicle is more likely to keep a higher headway with a leading vehicle that changes its speed and lateral position more often. If more vehicles behave haphazardly, more significant gaps in the stream may occur which leads to a significant capacity drop. Therefore, the magnitude of fluctuations in speed and lateral movement should be evaluated before developing car-following and lane-changing models.

This chapter utilises vehicular trajectory data collected every 0.5-sec from two distinct traffic conditions, the heterogeneous or mixed traffic from India and the lane-based homogeneous traffic from the USA. The Hurst exponent concept is then applied to evaluate fluctuations in microscopic variables, specifically, speed and lateral movement of individual vehicles. Then the chapter shows the effect of vehicle type and average lateral position on fluctuations in mixed traffic.

1.4.3 Chapter-4 Fluctuations in Macroscopic Traffic Flow Variables: An Application of Loop Detector Data

Technological developments have resulted in the ubiquitous use of smart phones on a global scale, with over 5 billion people having access to a mobile phone (Kanhere 2013). Inbuilt Bluetooth and GPS devices have formed a new option for traffic data collection. User locations, travel patterns, route selections, and travel times and speeds can all be collected using the crowdsourced smartphone data. However, traffic studies based on crowdsourced data are less frequent because

[3] Car-following models are used to figure out how vehicles follow one another on a roadway
[4] Lane-changing models describe the lateral movement dynamics of vehicles.

such data are not widely available. On the other hand, loop detectors provide 95% of the macroscopic data used by traffic agencies and analysts worldwide (Bickel et al. 2007; Xiong et al. 2014; Ambühl et al. 2017). These detectors provide aggregated traffic measures, such as speed, flow, and occupancy every 20-30 seconds. These data are useful in near-real-time traffic control, management and building traffic flow models for planning. However, loop detector data are prone to fluctuations, which can arise from driver heterogeneity, road geometry, changes in traffic demand, crashes (or incidents in general), weather, road surface conditions, lane changing, and ramp inflows and outflows. Understanding the patterns in fluctuations will provide better insights on traffic flow models, and related considerations such as fuel consumption, emissions, driver comfort, etc.

This chapter analyses fluctuations in macroscopic traffic variables, namely speed and flow of an urban motorway in Sydney, Australia. Yet again, the Hurst exponent is utilised. This chapter discusses the spatial and temporal variation of the long-range dependence properties of flow and speed at several monitor sites. Furthermore, the chapter investigates the effects of several factors such as the number of lanes, time of the day, day of the week, proximity to ramps, the presence of horizontal curve, on flow and speed fluctuations.

1.4.4 Chapter-5 Application of the Fractal Theory in Crash Rate Modelling

Speed is one of the critical factors in crashes. Mean speed has been found to have mixed effects on aggregated crashes in the literature. While some studies found that increased speed reduces safety, other studies found the opposite. Further, there are a few studies of the view that speed variation is likely to be a more critical determinant of crashes than speed itself. However, the results are inconsistent which could be due to the lack of high-resolution data. This chapter introduces the Hurst exponent metric of speed as one of the explanatory variables in crash-rate modelling.

Further, this chapter estimates a random parameters model and a latent class Tobit regression model to examine the effect of the long-memory property of speed on historical crash rates, while also accounting for unobserved heterogeneity. Using a latent class modelling approach, the motorway sections are probabilistically classified into two segments, based on the presence of entry and exit ramps. Such segmentation will allow transportation agencies to implement appropriate safety/traffic countermeasures when addressing crash hotspots or inadequately managed sections of motorway.

1.4.5 Chapter-6 Fluctuations in Macroscopic Traffic Flow Variables: An Application of SCATS Intersection Count Data

Traffic counts at intersections are typically consistent and repetitive on the one hand, and yet can be variable and less predictable on the other hand in which on any given time, unusual circumstances such as crashes and adverse weather, can dramatically change the condition of road traffic. Understanding the various aspects leading to high/low complexity in the dataset is essential for better prediction results and the choice of prediction methods.

This chapter analyses fluctuations in traffic count data at signalised intersections. Data collected from 37 intersections in the downtown of Sydney, Australia for one year are used. There are several techniques to quantify the complexity (or predictability) of time series data. However, they do not offer intuitions on what makes a time series data difficult to predict. So, this chapter estimates a random-effects linear regression model to quantify the effect of factors such as the day of the week, special event days, public holidays, rainfall, temperature, bus stops, and parking lanes on the predictability of traffic counts.

1.4.6 Chapter-7 Conclusions

This chapter concludes the thesis with a synopsis of the significant findings and future research directions and applications of the research to advance the approaches presented in this thesis.

CHAPTER 2. METHODOLOGY

This chapter introduces the core methodology of this thesis. First, it provides an outline of the Chaos theory, including a review of studies applying the theory in transportation. Then the chapter provides an overview of the Fractal theory, with a particular focus on the Hurst exponent, the metric that is extensively used in this thesis to evaluate fluctuations. Then, this chapter presents various techniques to compute the Hurst exponent, followed by a review of studies applying the Fractal theory in transportation.

2.1 CHAOS THEORY

2.1.1 Background

Traffic flow models (and the field of mathematics, in general) have been restricted to the linear world for many decades. In the past, mathematicians and physicists have regarded dynamical systems as random and unpredictable (Donahue III 1997). They gave their primary attention to the systems that were believed to be linear, i.e. systems that were highly predictable. However, the world is full of unpredictability and surprises. In real-world systems, order or predictability always breaks down at one point, increasing entropy and chaotic behaviour (Thomas and Dia 2006). There are some areas which are hard to explain by the traditional linear methods. Weather patterns, ocean currents, and traffic flow are just a few. Scientists have long investigated ways to pursue non-linearity in systems to better model the real-world phenomenon. In this regard, Chaos theory is a promising branch of mathematics that studies the behaviour of non-linear dynamic systems which are extremely sensitive to initial conditions.

According to Williams (1997), chaos is a sustained, and disorderly-looking long-term evolution that satisfies specific unique mathematical criteria and that occurs in a deterministic nonlinear system. While most traditional methods deal with apparently predictable phenomena like gravity, electricity, or chemical reactions, Chaos theory deals with nonlinear phenomenon such as turbulence, weather, the stock market, and our brain states, that are extremely difficult to predict (Fractal Foundation 2017). Chaos theory states that within the apparent randomness of complex chaotic systems, there are underlying patterns, constant feedback loops[5], repetition, self-similarity[6], fractals, self-organisation[7], and reliance on programming at the initial point known as sensitive dependence on initial conditions (SDIC)[8]. Due to the SDIC property, the presence of chaos will hamper the success of long-term prediction.

Chaos analysis helps to find the underlying mechanism of the system, whether it is genuinely chaotic or random. Unlike random behaviour (say Brownian motion), chaotic behaviour is typically predictable over a short duration using some simple deterministic equations. The main advantage of the Chaos theory is that it reveals information about the system and relationships without having to find the laws or

[5] Outputs of a system are routed back as inputs as part of a chain of cause-and-effect that forms a circuit or loop

[6] Object that is exactly or approximately similar to a part of itself, such as coastlines (show the same statistical properties at many scales).

[7] A spontaneous process where some form of overall order arises from local interactions between parts of an initially disordered system. Examples are the organising of birds into an orderly flock, weather elements forming into a hurricane, etc.

[8] Zheng and van Zuylen (2010) present an interesting analogy for SDIC in the context of traffic. Two consecutive vehicles entering the network at the same time can have entirely different travel times when the first vehicle just passes an intersection at the end of the green phase and the following vehicle has to stop. This small difference may have an impact on delays at the later intersections, so that the first vehicle may have a significantly shorter travel time than the second one. This phenomenon, where minor differences in the initial state can have large differences in the final condition, is called "bifurcation" and systems with bifurcations are ill-predictable (Zheng and Van Zuylen 2010).

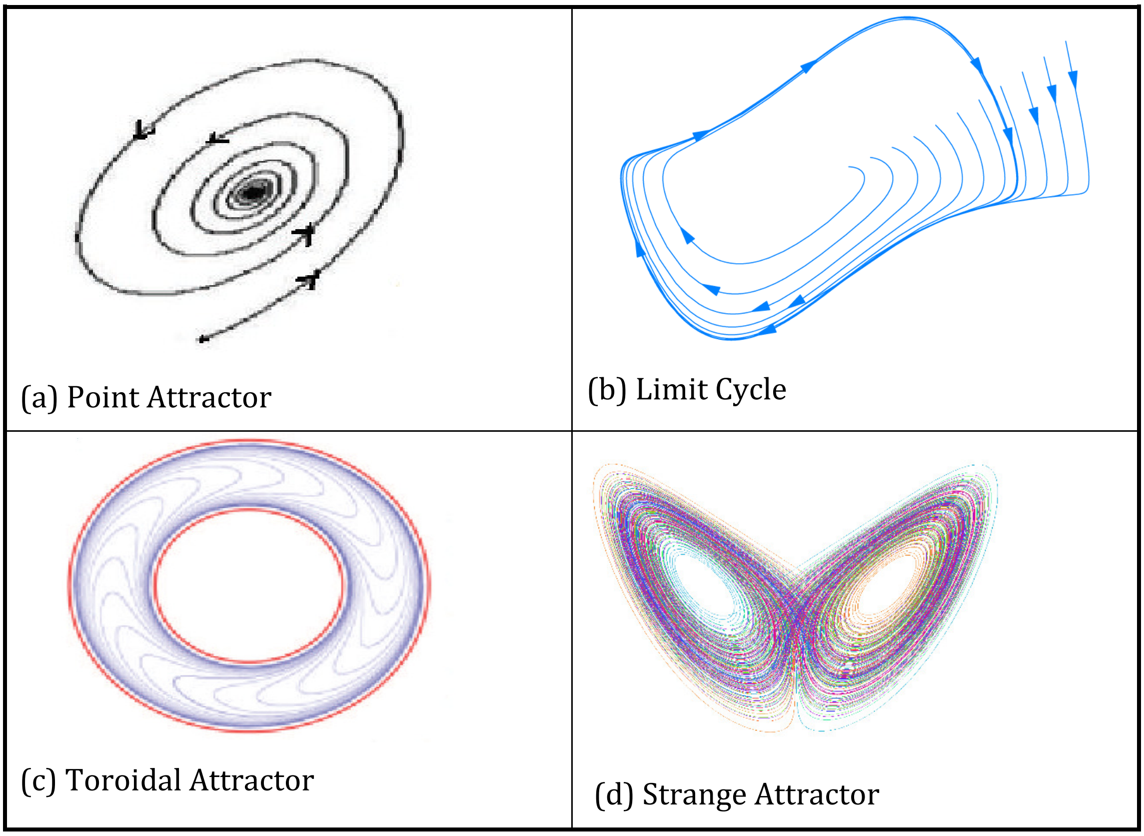
equations of the fundamental dynamics (Shang et al. 2005). Chaos theory has a wide range of applications in meteorology, stock markets, population dynamics, biology, physics, and engineering.

There are several methods, such as the power spectrum analysis, the largest Lyapunov exponent method, the Poincare's method, the Kolmogorov entropy method, and the surrogate data method to identify chaos in time series data. Nevertheless, there is no unique method to identify chaos phenomena firmly (Sivakumar 2000). Use of the Lyapunov exponent has been made in some of the studies, but other techniques are relatively unexplored in transportation.

2.1.2 Attractors

Possibly the most central concept of the Chaos theory is that of an attractor. An attractor is a point or curve to which the system tries to attain conformity. In other words, it is a set of stable conditions for a dynamical system, i.e. a compact and a comprehensive depiction of all possible steady states of a system. An attractor shows the long-term behaviour of a system (Williams 1997).

Attractors are broadly classified into two types: nonchaotic and chaotic. Nonchaotic attractors are further classified into three types: point, periodic, and toroidal attractors (in order of increasing complexity). Nonchaotic attractors consist of regular and predictable trajectories, i.e. predictions of long-term evolution can be entirely accurate, even far into the future. The point attractor (shown in Figure 2-1(a)) represents all systems that come to rest with the passage of time or that progress to a state where they no longer vary with time. Examples include a bouncing ball that comes to rest, a pendulum, and a marble rolling in a bowl and coming to rest at the bottom. The periodic attractor or "limit cycle" consists of two or more values that keep repeating in the same order (Figure 2-1(b)). An example is a mechanically boosted pendulum that repeatedly swings back and forth between the same two endpoints. Finally, the toroidal attractor (Figure 2-1 (c)) is a combined representation of two or more limit cycles, generally taking the shape of a 3-d doughnut in phase space (Williams 1997).



(a) Point Attractor

(b) Limit Cycle

(c) Toroidal Attractor

(d) Strange Attractor

Figure 2-1 Different types of attractors

On the other hand, chaotic or strange attractors or fractals as shown in Figure 2-1(d), arise only after the onset of chaos. They take on many interesting and complex shapes in phase space. There is no way to predict long-term evolution on those attractors with any reliability (Williams 1997). The fractal is not a point or a simple and smooth continuous curve but could be an infinite set of unconnected points (e.g. a Cantor dust), or a smooth curve with mathematical discontinuities, or a curve that is fully connected but discontinuous everywhere. Fractals are the objects that can be viewed with a similar appearance at different magnification scales. In case of nonchaotic attractors, trajectories for two different values of starting conditions get closer together or remain equidistant, for the same value of the control parameter. However, in case of a chaotic attractor, the trajectories diverge.

2.1.3 Dimensions

Dimensions are quantitative measures to evaluate or compare the geometric complexity of attractors (or objects or systems) of different size, shape, and structure. Computing the dimension can help differentiate between chaos and randomness. There are many types of dimensions in Chaos theory; however, there is no single type of dimension that is more useful than others. According to Williams (1997), the following are some of the commonly applied dimensions:

- Similarity dimension (or box-counting dimension)
- Capacity dimension
- Hausdorff dimension
- Information dimension
- Correlation dimension

The primary objective

of all these dimensions is to designate an object's geometry by a single number. Most types of dimensions are interrelated, and some even have the same numerical value for specific conditions. The "Fractal dimension" (FD) is a basket term (a more general) that can refer to any of a wide variety of exponent dimensions listed above. The first three dimensions in the above list measure only the attractor's geometry/complexity while the last two dimensions consider not only geometry but also informational features of the attractor (Williams 1997).

2.2 FRACTAL THEORY

The Fractal theory, as such was introduced by Benoit Mandelbrot to study the complex irregularities (Mandelbrot 1967, 1983, 2004). One of the main advantages of the fractal analysis is its ability to find patterns in large and complex datasets. A fractal can be defined as an irregular geometric object that has finite details at all scales.

Unlike Euclidean shapes, fractals have no typical sizes and are self-similar and independent of scaling. In other words, while the classical geometry deals with objects of integer dimensions (such as 1-d lines and curves, 2-d polygons, 3-d volumes and so on), Fractal theory deals with objects of non-integer dimensions. The underlying concept of the Fractal theory is that many natural phenomena are better described using a non-integer dimension. This dimension depends on the complexity of the shape of the object, i.e. a shape with a higher FD is more complicated or rough than the one with a lower dimension and fills more space (Breslin and Belward 1999).

A straight line has a dimension of precisely one. However, the line shown in Figure 2-2 (left side) is complex and tends to cover an area. The flatter the line fills the plane, the closer the FD approaches to two. Similarly, the surface on the right side is more complicated than a simple square or circle. The dimension approaches three if the surface is more complicated (or irregular). To summarise, the Fractal dimension describes how much space an entity fills and is a measure of the irregularities of that entity (Thomas and Dia 2006). Gneiting et al. (2012) give an elegant explanation of the Fractal dimension. According to them, the graph of a smooth, differentiable surface indexed in $\mathbb{R}^d$ has topological and Fractal dimension d. If the surface is nondifferentiable and rough, the FD takes values between the topological dimension, d, and $d+1$.

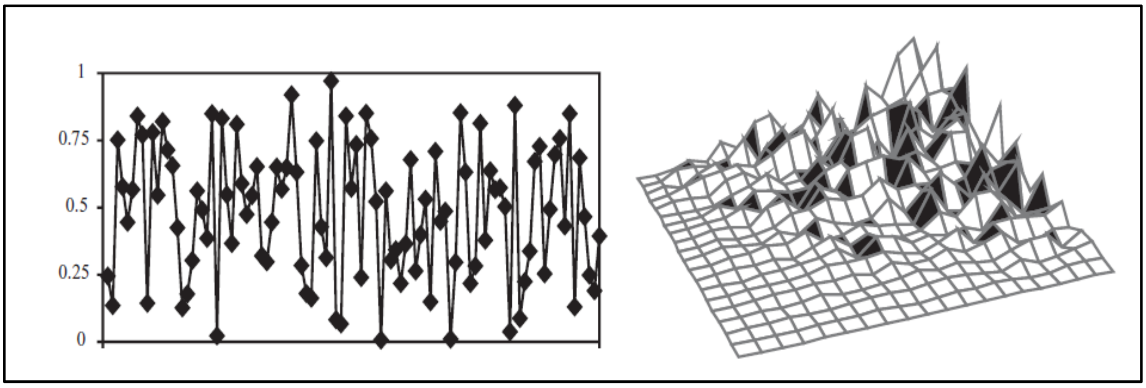


Figure 2-2 Fractal line and surface {Source: Thomas and Dia (2006)}

Williams (1997) pointed out a subtle difference between the Fractal theory and the Chaos theory. According to him, fractals deal with geometric patterns and quantitative ways of characterising those patterns. On the other hand, chaos analysis deals with time evolution of a system and its underlying or distinguishing characteristics. Fractals are a class of geometric forms; while chaos is a class of dynamical behaviour. Fractals and chaos are closely related, and fractals can help detect chaos (Williams 1997). The focus of this thesis is on quantifying fluctuations (i.e. the application of the Fractal theory) in traffic data rather than the evolution of dynamics of a system (i.e. the chaos analysis).

2.3 HURST EXPONENT

The Hurst exponent (denoted by H) is easily the most widely used method to calculate the Fractal dimension (D) of a time series, because of their direct relationship with each other, as shown in Equation 2.1 (Bo and Rashed 2004).

$$D=2-H \tag{2.1}$$

It should be noted that the Hurst exponent was proposed in the 1950s by a British hydrologist, Harold Edwin Hurst (Hurst 1951, 1956; Hurst et al. 1965), while Mandelbrot recognised its potential in the 1960s and introduced it later to the Fractal theory. Hurst proposed a technique called the rescaled range analysis (estimation procedure shown in the next section) to evaluate the flooding patterns of the Nile river. He studied 800 years of records of flooding and observed that a high flood year is likely to be followed by another high flood year, and a low flood year is likely to be followed by another low flood year. In other words, the annual levels of the Nile were found to be not independent, but they show a property called, the long memory property or the long-range dependence (LRD).

The Hurst exponent evaluates the LRD of a time series, by estimating the relationship between the decreasing rates of autocorrelations and the increasing lag between pairs of values (Hurst 1951, 1956). It provides a measure of whether the data of a time series is a pure white noise random process or has underlying trends. The Hurst exponent falls in a range of 0 to 1, with values higher than 0.5 (i.e. D < 1.5) indicating a persistent series, less than 0.5 (i.e. D > 1.5) showing an anti-persistent time series and close to 0 (i.e. $D \simeq 1.5$) indicating a random walk (Brownian motion).

The Hurst exponent was used by Rangarajan and Sant (1997) to develop a predictability index *(PI),* which has the same range of 0 to 1 as the Hurst exponent. They are related as shown in Equations 2.2 and 2.3. If *PI* is close to zero, then the corresponding process approximates the usual Brownian motion and is therefore unpredictable. If it is close to one, the process is highly predictable. This *PI* can be considered as theoretical predictability of a dataset, and it is different to the goodness of fit obtained by any prediction technique (which can be higher or lower than the *PI*).

$$PI = 2\,|D - 1.5| \tag{2.2}$$

$$PI = 2\,|H - 0.5| \tag{2.3}$$

For a time series,

i. A value of H in the range 0.5 to 1 is indicative of a time series having a long-term positive autocorrelation. In such cases, a high value in the series will likely be followed by another high value, i.e. the future trend is more likely to follow an established trend. For example, a very high H value (say H = 0.9) means a strong long-range dependence and a greater level of determinism (as shown in Figure 2-3 [a]), i.e. good predictability (PI = 0.8).
ii. H values close to 0.5 (Figure 2-3 [b]) indicates a completely uncorrelated series. It means that the values in the time series are random and potentially indicating Brownian motion. The *PI*, in this case, gets closer to 0 because it becomes extremely challenging to 'precisely' predict the stochastic variations.
iii. H values in the range of 0 to 0.5 suggest the long-term fluctuation between high and low values in adjacent pairs of observations in the time series. A low H value (say H = 0.1) indicates strong determinism in the time series. Although the

data are not random, they do fluctuate as depicted in Figure 2-3 (c). It is because a single high value will likely be succeeded by a low value or vice-versa. Due to strong determinism, the *PI* of the time series is high, even though *H* is low (*PI* = 0.8 for the series shown in Figure 2-3 [c]). Therefore, the *PI* for a time series is the same, if the value of *H* is either 0.9 or 0.1. Furthermore, the *PI* increases when *H* approaches either 1 or 0 and decreases when it approaches 0.5.

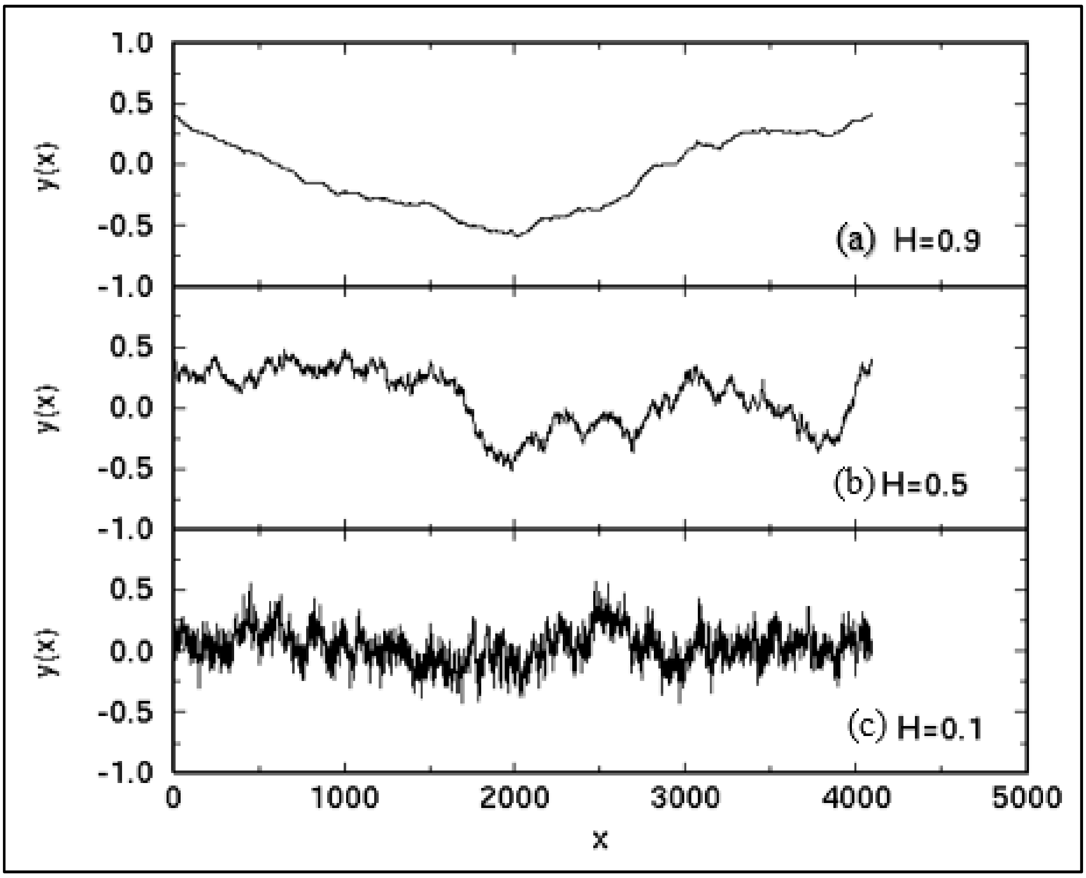


Figure 2-3 Demonstration of the Hurst Exponent. (Source :(The Cooperative Phenomena Group 1998))

To summarise, large values of *H* imply weak dynamics, while small values of *H* denote frequent changes, i.e. high dynamics. Additionally, high values of PI indicate less stochasticity (or strong determinism); whereas, low values of PI show a sign of high stochasticity (or weak determinism).

The main advantage of using the Hurst exponent rather than the usual coefficient of variation for the measurement of fluctuations in this thesis is that the former is a numerical representation of the randomness through the history of a dynamical process, while the latter is a statistic that is independent of the temporal evolution (Valle et al. 2013).

There are many studies on the application of the Hurst exponent technique to evaluate trends in financial markets (Bo and Rashed 2004; Eom et al. 2008; Matos et al. 2008). Most economic and financial time series data are *"trend- reinforcing"* or persistent with H higher than 0.5. This technique has been used extensively in medicine (DePetrillo et al. 1999; Havlin et al. 1999), ecology (Peng et al. 2012), seismology (Yulmetyev et al. 2001) and hydrology (Pelletier and Turcotte 1997). In the transport domain, there have been only a few applications of the Fractal theory and the Hurst exponent on real data.

There are several estimators of the Hurst exponent, such as the R/S method (Hurst 1951, 1956), the aggregate variance method, the absolute value method, Higuchi method (Higuchi 1988), residuals of regression method (Peng et al. 1994), periodogram method (Geweke and Porter-Hudak 1983), Whittle estimator (Whittle 1953), and Abry Veitch Method (Abry and Veitch 1998). However, the Hurst exponent method calculated using the classical rescaled range (or R/S) analysis (the one originally proposed by Hurst himself and later popularised by Mandelbrot) is widely followed by researchers (Breslin and Belward 1999; Bo and

Rashed 2004; Haleem et al. 2016; Yuan and Lin 2017). The following sections provide a summary of some of the Hurst estimation procedures.

2.3.1 Rescaled Range Analysis (R/S) Method

The process is shown below from Equations 2.4 to 2.9.

For a time series, $X = X_1, X_2, \ldots X_n$

(1) Calculate average,

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.4}$$

(2) Calculate mean adjusted series Y,

$$Y_t = X_t - m \qquad t = 1, 2, \ldots, n \tag{2.5}$$

(3) Calculate cumulative deviate series Z

$$Z_t = \sum_{i=1}^{t} Y_i \qquad t = 1, 2, \ldots, n \tag{2.6}$$

(4) Calculate range series R

$$R_t = \max(Z_t) - \min(Z_t) \qquad t = 1, 2, \ldots, n \tag{2.7}$$

(5) Calculate the standard deviation series S

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(X_i - m)^2} \qquad t = 1, 2, \ldots, n \tag{2.8}$$

(6) Calculate rescaled range series (R/S)

$$\left({}^{R}/_{S}\right)_t = {}^{R_t}/_{S_t} \qquad t = 1, 2, \ldots, n \tag{2.9}$$

It should be noted that $(R/S)_t$ is averaged over the regions $[X_1, X_t]$, $[X_{t+1}, X_{2t}]$ until $[X_{(m-1)t+1}, X_{mt}]$ where m = floor (n/t). The value of *t* is normally chosen so that it is divisible by n.

(7) Finally, Hurst exponent (H) is the slope of the regression line that estimates the relation between (R/S) versus t in log-log axes.

As may be seen above, the R/S process is non-parametric, and the main advantage of this method is that the robustness of the results will not be affected based on the distribution of data points of the time series (Yuan and Lin 2017).

2.3.2 Aggregated Variance Method

In this method, the original time series data are divided into subseries of size *m*[9], and average within each subseries is computed. The sample variance within each subseries is then calculated. The slope of the least square fit of the logarithm of the sample variances versus the logarithm of the size of the subseries is estimated as $2H - 2$, from which the Hurst exponent (H) can be calculated (Haleem et al. 2013).

---

[9] The choice of the ranges for "m" depend on the judgement of the analyst. In financial markets, analysts use ranges that match their investment time horizon. Few researchers created algorithms that look for sustained Hurst exponents by utilising thousands of different time horizons. The length of the sub-series influences the standard deviation and the mean. Therefore, the Hurst exponent may not be accurate if only calculated from a single data set. However, it is to be noted that the Hurst exponent estimated in the thesis is calculated from a multiple data sets (several vehicles in chapter 3, several time windows in chapter 4, and several days in chapter 6).

Literature says that the Hurst exponent calculated from a 'single' short data set (less than 100 data points) may deviate from its theoretical value by as much as 35%. However, when a spectrum of Hurst values is calculated from 'multiple' data sets, the average H value will be closer to the theoretical value. As the sample size increases, the deviation from theoretical value becomes smaller. For samples greater than 256 data points, the deviations become extremely low (Granero et al. 2008; Katsev and L'Heureux 2003).

2.3.3 Absolute Values of the Aggregated Series

This method (also called just as absolute values method) is similar to the aggregated variance method. Instead of computing the sample variance, the sum of the absolute values of the aggregated series should be computed. Then the slope of the least square fit of the logarithm of the calculated statistic versus the logarithm of m is estimated as H – 1 (Taqqu et al. 1995).

2.3.4 Higuchi Method

This method was proposed by Higuchi (1998), and it is similar to the absolute values of the aggregated series method, except that it implements a sliding window (partial sums) of the original time series and then finds the normalised length of the curve. The slope of the least square fit of the logarithm of the normalised curve length versus the logarithm of the window sizes equals D, the Fractal Dimension. H can then be computed using Equation 2.1.

2.3.5 Detrended Fluctuation Analysis (DFA)

This method is also popularly known as residuals of regression method, initially proposed by Peng et al. (1994) to study the long-range correlations of DNA sequences. The following steps detail the process of DFA (Bărbulescu et al. 2010).

(1) Starting with a time series, (X_t) with the length N, it is integrated, obtaining:

$$Y_k = \sum_{i=1}^{k}(X_k - \bar{X}) \qquad (2.10)$$

(2) Then the integrated series is split into d blocks of equal length m.

(3) In each block, fit X_t, using a polynomial function of order l which represents the trend of that block. The ordinate of the fit line in each box is denoted by $y_m(k)$.

(4) The integrated series is detrended by subtracting the local trend $y_m(k)$ in each block of length m.

(5) For a given block length, m, the root-mean-square fluctuation for the integrated and detrended series is computed:

$$F(m) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[Y_k - y_m(k)]^2} \tag{2.11}$$

(6) The preceding calculation is repeated for a broad range of scales (m) to provide a relationship between $F(m)$ and the box size m.

(7) A power-law relation between the average root-mean-square fluctuation function $F(m)$ and the box size m indicates the presence of scaling: $F(m) \sim m^{2H}$

2.3.6 Periodogram Regression

The periodogram is an estimator of the spectral density of time series data within finite variance domain (Haleem et al. 2013). In this method, one should estimate spectral density, $I(\lambda)$:

$$I(\lambda) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j^{\,eij\lambda}\right|^2 \tag{2.12}$$

where λ is frequency, N is the number of data points in the time series, and X_j is the time series data. A series with long-range dependence should have a periodogram which is proportional to $|\lambda|^{1-2H}$. Therefore, a regression of the log of the periodogram on the log of the frequency λ should give a coefficient of $1-2H$, through which an approximation to H can be obtained (Geweke and Porter-Hudak 1983; Taqqu et al. 1995).

2.4 STUDIES ON CHAOS THEORY IN TRANSPORTATION

In transportation systems, it was only in the late 20th century that researchers explored the application of Chaos theory. Disbro and Frame (1989) identified chaotic tendencies of the solution of the GHR generalised car-following equation[10]. Dendrinos (1994) analysed real traffic count data on a primary collector road in the U.S. for over two weeks. He observed that the traffic counts at finer time resolutions than 15-min were less periodic and more random. He used the power spectrum density method and could not find evidence of chaos in traffic count data at 15-min resolution, which he believed could be due to the lack of congested conditions on the road examined. Chaos in traffic queues based on a non-linear car-following model was then studied by Addison and Low (1998).

Zhang and Jarrett (1998) examined the dynamic gravity model for O-D pairs, and the confirmed the existence of chaos (albeit in higher dimensions) by the positive Lyapunov exponents and Fractal dimensions. This analysis showed that the equilibrium in gravity model might not be unique and stable. Further, the authors opined that it would be deceptive to predict an O–D flow pattern (at a supposed equilibrium) from a static gravity model when the underlying system is not at an equilibrium steady state.

[10] The Gazis-Herman-Rothery (GHR) family of models is the most popular car-following model (Gazis et al. 1961). In these models, the fundamental relationship between a leader and a follower vehicle is a stimulus-response type of function. The GHR model states that the follower's acceleration is proportional to the speed of the follower, the speed difference between follower and leader, and the space headway (Brackstone and McDonald 1999).

Using Lyapunov exponent, Nair et al. (2001) observed chaos in traffic data and suggested the use of phase space reconstruction[11] techniques to analyse and predict the traffic flow. They used the morning peak hour traffic data (6 AM-10 AM) collected for about six months in the year 1999, on an 11-kilometre stretch of a San Antonio highway. The raw data for speed, volume and occupancy were collected every 20 secs for each lane at more than 20 loop detectors. However, they aggregated raw data into 5-minute estimates of average speed, total volume and average occupancy after screening out erroneous data. This aggregation procedure would have smoothened out some irregularities in the data.

Safonov et al. (2002) showed that chaotic behaviour in traffic could arise due to the delays in driver reaction. Jamison and McCartney (2007, 2009) found that accounting overtaking manoeuvres in car-following models could produce chaos in traffic streams.

Lan et al. (2003) conducted different tests of chaos analysis on morning peak hour (6 AM-9 AM) traffic flow time series data at around 20 loop detector stations of the Miami Freeway in the USA. They found substantial evidence of chaos based on these tests. Further, using the concept of spatial and temporal nearest neighbours in the reconstructed phase space, they developed a confined space fuzzy neighbourhoods' difference (CSFND) model to predict the traffic flow variations.

Frazier and Kockelman (2004) collected freeway traffic data from Sacramento, the USA at 30-sec resolution. However, they aggregated the raw data to 5-min counts

[11] It is a complement to a time series plot and provides a different view of the evolution by condensing all the data into a manageable space on a graph. Structure one might not see on a time series plot often comes out in remarkable fashion on the phase space plot. For more details of phase space reconstruction, please refer to Williams (1997).

to smooth higher frequency fluctuations. They used several chaos analysis techniques and found traffic flow data to be chaotic. However, they were sceptical that the result could be due to high-frequency noise in the data rather than pure chaos.

Later, Lan et al. (2005) used a three-step parsimony procedure and observed chaos in traffic flow data collected at 1-min resolution during the morning peak hours (6 AM-9 AM) on the United States I-35 Freeway in Minneapolis, Minnesota. Using the same dataset, Lin and Lan (2005) examined nonlinear dynamics of traffic flow data at different timescales. They found that flow data measured in 1-min, exhibit chaotic phenomena, but if the same data were measured in 5-min and 10-min timescales, chaotic structures might disappear. In a separate study, Lan et al. (2007) proposed a novel filtering technique to analyse the traffic flow data of the Taiwan Freeway. They used the Fourier power spectrum, the Lyapunov exponent and the correlation dimension at different resolutions such as 20-sec, 1-min, 3-min and 9-min and observed different non-linear patterns and chaos phenomena.

Xu and Gao (2008) collected traffic data for a period of five weekdays during morning rush hours from a freeway in the Netherlands. During uncongested conditions, the structure of the traffic flow series inclined to be a point attractor or period-2 attractor. However, during congested conditions, they found evidence of chaos, and drivers were either accelerating or decelerating more often. They applied the maximum Lyapunov exponent and the Fractal dimension methods for the analysis.

Krese and Govekar (2013) collected traffic flow data in a ring road-based network (two detectors on highways and one on ring road) in the city of Ljubljana, Slovenia.

They implemented a 0-1 test for chaos {see Gottwald and Melbourne (2004) for details of the test} and found that traffic flow was chaotic at all the locations. However, using other standardised tests such as the Lyapunov exponent, they found that the ring road was less sensitive to initial conditions (small traffic perturbations) than the highways.

The significant implication of the presence of chaos in traffic systems is the inability to predict long-term behaviour. However, few studies employed methods such as zero-order local region (Fu and Wei 2013), neural network (Fu et al. 2005), support vector machine (SVM) theory (Wang and Shi 2013), the largest Lyapunov exponent method (Adewumi et al. 2016), and phase-space reconstruction (Hu et al. 2003) to predict the short-term behaviour[12] of chaotic systems. Compared with the long-term traffic flow predictions, short-term predictions focus on the real-time control. Therefore, chaos analysis can help in the real-time inducement and signal control.

2.5 APPLICATIONS OF FRACTAL THEORY IN TRANSPORTATION

Although fractal analysis of time series data has broad applications in many scientific fields, it has been hardly explored in detail in the transportation domain. Presently, research on the Fractal theory in transport systems is limited to observing fractal nature of traffic flow variables, rather than looking at potential real-world applications.

[12] Short-term predictions are not within the scope of this thesis and therefore, the interested readers are encouraged to refer to the cited references for more insights on prediction techniques.

2.5.1 Identification of Fractal Characteristics in Traffic Data

Rhodes (2000) analysed fluctuations in traffic speed data of the M25 motorway in the United Kingdom using the Hurst exponent. He collected data for one week during the morning peak hours and found anti-persistence in speed data. In other words, speed fluctuations were strongly constrained by the mean flow velocity.

The earlier studies mainly focused on observing self-similarity and LRD in traffic flow variables using simulated data (Török and Kertész 1996; Campari and Levi 2002; Nagatani 2005). These studies confirmed the non-linearity of the system by observing the fractal behaviour and further stated that it is also a signature of deterministic chaos. Campari and Levi (2002) simulated highway traffic using a simple cellular automata (CA) model and found self-similarity in car density and car flow. They used the box-counting method to calculate the Fractal dimension and noticed that the dimension increases moving from free flow conditions to the congested regime. Similarly, Nagatani (2005) performed a simulation experiment and concluded that a single vehicle passing through an infinite sequence of traffic lights exhibits self-similar behaviour.

Researchers in Germany analysed highway speed data collected from 60 detectors in free-flow conditions for approximately one year and observed substantial evidence of LRD (Belomestny and Siegel 2003). Shang et al. (2007) collected traffic data on a highway in China over a period of 40 months at a 20-sec interval. Further, they screened the raw data for errors and aggregated into 1-min average data. They used several methods such as the power spectrum, statistical moment scaling, and autocorrelation function and identified the fractal characteristics of speed data during free-flow conditions.

Thakur et al. (2013) analysed more than 25 million Google traffic images at more than 800 locations from six locations (Connecticut, London, Seattle, Sydney, Toronto and Washington DC) around the world. They extracted vehicular density time series using the background subtraction algorithm. Further, using several Hurst exponent estimation methods, they found that the density time series is self-similar and long-range dependent.

Lin et al. (2013) used techniques like approximate entropy[13] and the Hurst exponent to quantify the predictability of traffic volume time series of different highways in China and the USA. Further, they applied three prediction techniques (SARIMA, SVR, and k-NN) and correlated their performance with the results from the predictability quantification methods. They found that the SVR method was suitable for non-linear datasets and SARIMA and k-NN were suitable for linear datasets. The predictability quantification methods helped them in selecting the parameters of the different prediction methods. Thus, they showed the benefits of evaluating the predictability of datasets before resorting to different prediction techniques.

Krause et al. (2017) analysed traffic flow, density and speed data averaged over 1-min intervals over a period of one year from German motorways. Using the Hurst exponent, they observed the robust anti-persistent behaviour of traffic flow time series. They opined that anti-persistence in traffic flow is a crucial property for understanding traffic congestion patterns. Recently, Yuan and Lin (2017) studied

[13] Entropy, in some sense, is similar to the Hurst exponent. While Hurst exponent quantifies the complexity of the past information, Entropy measures the uncertainty in the next information produced by the dynamical process given complete knowledge of the past.

traffic data from China and found H to be higher than 0.5 for flow, speed, and occupancy time series, which means they all had long-term memory property.

2.5.2 Multi-fractal Analysis

Almost all the aforementioned studies observed the monofractal behaviour of time series data, i.e., a single exponent for the entire time series, irrespective of its length. Since a single exponent is not enough to explain the dynamics of a traffic system for an extended period, some researchers explored the use of multifractals (Vojak et al. 1994; Storkey 1996; Shang et al. 2006, 2008; Li and Shang 2007; Di et al. 2016). Multifractal analysis helps in calculating a continuous spectrum of exponents by decomposing a large series into several subsets.

Hu et al. (2010) analysed urban traffic volume data from China and freeway traffic flow data from the USA, collected over a period of four weeks. They opined that the multi-fractal behaviour of traffic flow could be exploited for efficient traffic control and modelling. Yue et al. (2010) proposed a new measure of variability, called the time-dependent Hurst exponent $H(t)$ to capture the degree of traffic flow complexity at each time t.

Zaksek and Schreckenberg (2016) analysed traffic flows and velocities obtained from loop detectors and analysed multi-fractal features of different vehicle types. Zhang et al. (2017) used the multi-fractal properties of traffic flow and proposed a hybrid short-term forecasting model for a freeway in the Twin Cities area (Minneapolis–Saint Paul), USA.

Recently, Feng et al. (2018) analysed freeway traffic data collected every 5-min at more than 1,000 loop detectors in China. They used the detrended fluctuation analysis (DFA) and multifractal detrended fluctuation analysis (MFDFA) techniques and observed long-range dependence in freeway traffic data.

2.5.3 Vehicle Arrival Patterns

Meng and Khoo (2009) analysed traffic data from several highways in Texas, USA and Malaysia. They employed different techniques such as periodogram, aggregate variance, R/S analysis to estimate the Hurst exponent. Finally, they determined that the highway vehicle arrival pattern during moderate to heavy traffic conditions do not follow the Poisson process, but it could be modelled as a self-similar process. Perati et al. (2012) evaluated vehicle arrival patterns at a toll plaza in India using the Hurst exponent method. They too observed the self-similar pattern of vehicle arrivals and found that the mean queue length and the busy period distribution increase as the Hurst exponent increases. They opined that the findings would be helpful to determine the optimal number of toll counters and to reduce congestion.

2.5.4 Safety Analysis

Thomas and Dia (2006) collected speed and occupancy data and analysed 100 incidents in the Australian state of Victoria. They found that the downstream Fractal dimension of occupancy was lower than the upstream fractal dimension over the period of an incident. The traffic data upstream of the incident showed complex behaviour as traffic slowly crawled over the loop detectors. On the other hand, the traffic behaviour was more regular downstream, as a relatively constant number of vehicles passed. The authors used these findings in developing an automated incident detection model.

Ye et al. (2014) used the Hurst exponent technique and developed an accident warning model. The authors used a simulation model and found that an accident is likely to occur when the average box dimension value of the heart rate, rotational speed and the acceleration is higher than 1.7.

Haleem et al. (2016) used fractal analysis and explored trends in the annual crash rate between the years 1990 and 2011 at ten random signalised intersections in Florida. They opined that if the Hurst exponent of crash rate is higher than 0.5 for an intersection that is already in the high crash list, then the intersection is expected to continue to have an increasing trend even in the future. Similarly, if the Hurst exponent is less than 0.5, the increasing trend at the intersection is expected to reverse in the future. Inadequate data size is a limitation of this study.

2.6 SUMMARY

From the above literature review, the following conclusions can be drawn:

1. Several studies have been focussing on detecting the fractal behaviour in traffic flow data, but only a handful of studies have explored the potential applications of the Fractal theory.
2. Current applications of the Fractal and Chaos theories focus mainly on inter-city freeways.
3. There is limited application of fractal analysis in mixed traffic of developing countries.
4. Most studies have focussed on the fractal characteristics of macroscopic variables, such as speed, flow and occupancy than the microscopic variables, such as lateral movement, and longitudinal speed.

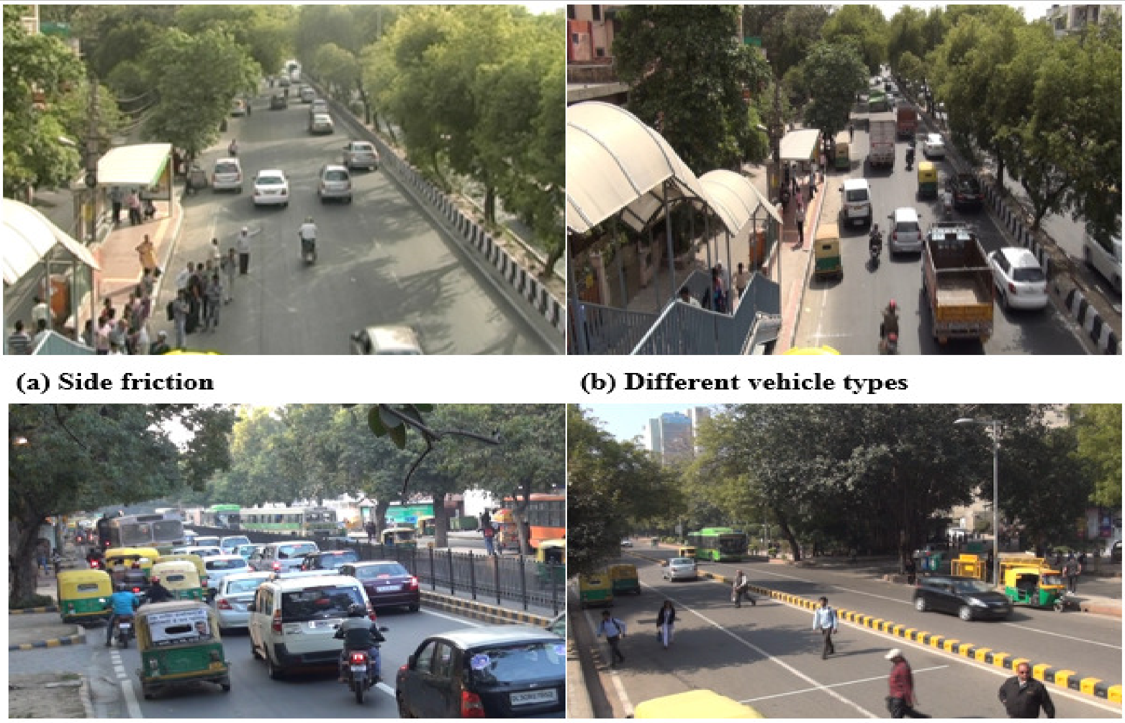
5. There are only a handful of studies on the application of the Fractal theory in safety.
6. There are no studies on the application of the Fractal theory to identify the possible reasons behind high/low fluctuations in traffic data.

CHAPTER 3. FLUCTUATIONS IN MICROSCOPIC TRAFFIC VARIABLES: AN APPLICATION OF VEHICLE TRAJECTORY DATA

3.1 INTRODUCTION

Traffic congestion is a growing concern in almost all the major cities across the world. Congestion is particularly alarming in developing countries where mixed traffic conditions exist, i.e. different types of vehicles with wide variations in physical and operational characteristics compete for available road space. This heterogeneity results in complex interactions between vehicles. The road width is shared by both motorised and non-motorised vehicles with speeds ranging from as high as 100 kmph to as low as 5 kmph. Lane discipline by drivers is very poor and thus drivers operate vehicles on whatever road space that is available without regard to road delineation (Arasan and Koshy 2005). Furthermore, traffic in such countries is characterised by loose traffic regulations, side friction (street vendors, people waiting for buses by standing on the road, etc.), honking, etc. Such fringe activities, as shown in Figure 3-1, result in the chaotic and congested traffic conditions on the urban road network.

On roads in most developing countries, vehicles keep lesser headways and lateral clearance distances with other vehicles, which might be considered unacceptable in many developed nations. While the lane changing of vehicles is discrete in homogeneous lane-based traffic conditions, the lateral movement is continuous in mixed traffic conditions. Drivers tend to change their lateral position and speed often to keep moving.



(a) Side friction

(b) Different vehicle types

(c) No lane discipline

(d) Undesignated pedestrian crossing (loose traffic regulation)

Figure 3-1 Typical mixed traffic conditions (Source: (Chand et al. 2016)).

Driver and vehicle heterogeneity in mixed traffic along with frequent braking of vehicles and lane changes affect the quality of traffic flow. Lane change manoeuvres create voids in traffic streams and could reduce travel speed and result in a capacity loss (Coifman et al. 2005; Laval and Daganzo 2006; Dhamaniya et al. 2017). The increased presence of lane changes in the surroundings could negatively influence workload and driving performance (Teh et al. 2014). If the lateral separation between a leading vehicle and the following vehicle decreases, the following distance increases (Gunay 2007). In other words, a vehicle is more likely to maintain a higher headway with a leading vehicle that changes its speed and lateral position more frequently. If more vehicles behave haphazardly, more significant gaps in the stream may occur which ultimately leads to a capacity drop.

Furthermore, a lane change manoeuvre could perturb the spacing-speed relation of the vehicles immediately following the manoeuvre and hence could create oscillations (Mauch and Cassidy 2002; Wang and Coifman 2008). Traditional car-following and lane-changing models are not applicable to these traffic conditions unless modifications accounting heterogeneity are made. Therefore, the fluctuations in speed and lateral movement should be evaluated before developing car-following and lane-changing models.

Researchers have tried to model mixed traffic conditions using several techniques such as agent-based modelling (Lee and Wong 2016), neural networks (Mathew and Ravishankar 2012), heterogeneous cell transmission model (CTM) (Kays et al. 2017), cellular automata (Mallikarjuna and Rao 2009; Chen and Wang 2016; Pandey et al. 2017), porous flow approach (Nair et al. 2011; Ambarwati et al. 2014;), strip-based approach (Mathew et al. 2013), social force and intelligent driver model (Babu et al. 2015), fuzzy logic (Oketch 2000), survival analysis (Guo et al. 2013; Zhao et al. 2015) and latent class approach (Choudhury and Islam 2016).

Although vehicles in mixed traffic appear to be travelling haphazardly, there seems to be a hidden underlying mechanism. Fractal theory is a promising research area to study the irregular movement of vehicles. Most of the studies applying the Fractal and Chaos theories in traffic modelling used traffic flow (volume counts and speed at aggregated level) data or simulated trajectory data. On the other hand, the application of the Fractal theory in real-life trajectory data is limited. Therefore, this chapter demonstrates the applicability of the Hurst exponent concept on vehicle trajectory data in mixed traffic. Furthermore, the results are compared with that of homogeneous data to examine the severity of chaos in mixed traffic.

The organisation of this chapter is as follows. First, the chapter presents a brief review of studies on the applications of vehicle trajectory data in mixed traffic. Then the chapter describes the sources for homogeneous and heterogeneous trajectory data, along with some descriptive statistics of the data. Next is an analysis of trajectory data, with a discussion on the effect of the vehicle type and the average lateral position on speed and lateral movement trends. Finally, the chapter summarises the work and concludes with potential future work in this regard.

3.2 APPLICATIONS OF VEHICLE TRAJECTORY DATA IN DEVELOPING COUNTRIES

Traffic flow is dynamic as it varies with time of the day and is traditionally a complex system because of the continuous interactions among different vehicles, the roadway, and the environment (Nair et al. 2001). The real-world trajectories reflect these interactions. Many researchers have analysed trajectory data using various mathematical techniques to examine the properties of shockwaves (Lu and Skabardonis 2007), fundamental diagrams (Chiabaut et al. 2009), emissions (Sun et al. 2015; Jiang et al. 2017), car-following and lane-changing dynamics (Kesting and Treiber 2008; Thiemann et al. 2008), road-user classification (Zaki and Sayed 2013), O-D estimation (Yang et al. 2017) and safety (Pande et al. 2017; Park et al. 2018). Most of the studies were based on the Next Generation Simulation (NGSIM) trajectory data which were collected on various freeways and arterials from the lane-based traffic conditions in the USA (FHWA 2007a; Kovvali et al. 2007).

Due to the unique and challenging traffic operations in developing countries, the typical trajectory extraction tools that were originally developed for lane-based traffic environment are not suitable. Therefore, some researchers have developed indigenous tools to transcribe trajectory information from video data (Mallikarjuna et al. 2009; Munigety et al. 2014b). For example, Mallikarjuna et al. (2009) developed TRaffic AnalyZer and EnumeratoR (TRAZER), an offline video image processing-based tool. The authors used this tool to capture lateral movements and track small sized vehicles under dense traffic conditions in New Delhi, India. However, Munigety et al. (2014b) noticed some drawbacks of TRAZER, such as low accuracy at the macroscopic level, occlusion, and its limitation to capture only up to 50m length of the road section. Therefore, they developed a traffic data extractor (TDE) on a Visual Basic (VB) platform. They used the concept of vanishing point-based camera calibration technique and claimed that the tool was able to capture trajectory data for up to 250m. Several studies used these tools to extract and explore the potential of trajectory data in calibrating or validating simulation models, analysis of motorcycle movements, and empirical analysis. The following section gives a synopsis of researches on the application of vehicle trajectory data on empirical analysis of microscopic characteristics in non-lane-based mixed traffic conditions. Appendix A presents a review of studies on other applications of trajectory data.

3.2.1 Empirical Analysis

In mixed traffic conditions, the real-life trajectory data were used in earlier studies to analyse lateral distribution (position on road width), lateral movement and lateral and longitudinal gaps (clear distances) for different categories of vehicles (Chunchu et al. 2010; Mahapatra and Maurya 2013). Chunchu et al. (2010)

collected traffic video data from New Delhi, India and extracted trajectory data using TRAZER (Mallikarjuna et al. 2009). They found that the traffic volume and composition significantly influence the lateral distribution of vehicles, while the proportion of the majority vehicle type in the traffic stream influences the lateral gap keeping behaviour. On the other hand, the longitudinal gap was found to be affected by speed, vehicle type, and the difference in lateral positions of vehicles.

Mahapatra and Maurya (2013) analysed the lateral movement of vehicles using the trajectory collected through GPS equipped vehicles. They observed that lateral acceleration is higher in case of auto-rickshaws[14] and lower in case of cars. Further, they also noticed an inverse relationship between lateral acceleration and longitudinal speed irrespective of the vehicle type. Recently, Pal and Chunchu (2017) extracted trajectory data on arterial roads from three major cities in India using TRAZER and analysed lateral gap maintenance behaviour of vehicles. They found that the lateral gap was significantly influenced by the size and speed of the subject vehicle, and speed of the adjacent vehicles.

Nguyen et al. (2014) developed an integrated model to evaluate the traffic conflicts of motorcycles in congested traffic conditions of Ho Chi Minch City, Vietnam. They found that motorcycles apply sudden braking in dense traffic and there is a high probability of crashes.

Munigety et al. (2014a) evaluated the effect of nearby traffic characteristics on lateral movement decisions of the vehicles based on trajectory data. They collected traffic video data on an urban arterial in Mumbai, India and extracted trajectory

[14] Auto-rickshaw or tuk-tuk is typically a three-wheeled motor scooter with a large cab mounted on it. It is a common form of urban transport, both as a vehicle for hire and for private/commercial use, in many developing countries around the world.

data using the semi-automated approach proposed by Munigety et al. (2014b). The dataset included the trajectories of about 3,200 vehicles. They found that the lateral movement duration was dependent on vehicle type, with two-wheelers exhibiting as low as 3 seconds and three-wheelers (auto-rickshaws) taking 12 seconds. This difference was attributed to the variations in the manoeuvrability of different vehicle types. Further, they found that cars and two-wheelers typically preferred faster path, however, heavy vehicles and three-wheelers preferred slower path.

Zhao et al. (2015) studied lateral interactions between motor vehicles and bicycles using trajectory data from Beijing, China. They used survival analysis for this study and observed that the motor vehicle drivers try to keep more lateral gap with non-motorised vehicles, while the non-motorised drivers accept even shorter lateral gaps. Recently, Li et al. (2016) also analysed interactions between motorised and non-motorised vehicles using the concept of the static potential energy. The model was aimed to predict precarious lateral positions for non-motorised vehicles in mixed traffic flow. They calibrated the model using real-life trajectory from an arterial in Beijing.

Asaithambi and Basheer (2017) proposed a multinomial logit model to predict the type of the following behaviour (typical car-following or staggered following or the in-between following) in mixed traffic. They applied the model on the Chennai trajectory dataset provided by Kanagaraj et al. (2015) and observed that a vehicle follows another if the size of the trail vehicle is smaller than the lead vehicle. However, the staggered following was more prevalent when the trail vehicle is larger in size than the lead vehicle indicating that the size of the vehicles has a critical role in the type of following behaviour in mixed traffic.

Recently, Budhkar and Maurya (2017) collected vehicle trajectories from six major cities in India and analysed the staggered vehicle following behaviour. They observed that when there was a second leading vehicle, there was a decrease in longitudinal gap compared to the situation with a single leader. Moreover, the longitudinal gap between leading and trailing vehicles increased as lateral gap between the two leading vehicles decreased.

Although these studies show the importance of substantial lateral movement in mixed traffic, the fluctuations were not thoroughly investigated. This thesis analyses the vehicle trajectory data obtained from Kanagaraj et al. (2015) and quantifies the fluctuations in lateral position and speeds of vehicles in mixed traffic. Also, this thesis analyses trajectory dataset in homogeneous conditions of the US {NGSIM (FHWA 2007a; Kovvali et al. 2007)} for comparison purposes.

3.3 DATA DESCRIPTION

The microscopic analysis of vehicle behaviour needs detailed and accurate vehicle trajectories. While NGSIM provides trajectories of thousands of vehicles on two freeways and two arterials in the USA, there was no publicly accessible dataset for mixed traffic conditions as recently as three years ago (till 2015). The limited availability of trajectory data in mixed traffic hindered the rigorousness in the estimation of microscopic models. Realising this problem, Kanagaraj et al. (2015) collected traffic videos of an urban arterial road in Chennai, India, transcribed trajectory information using the Trajectory Extractor tool developed by (Lee et al. 2008) and made the data publicly available.

3.3.1 Mixed Traffic Dataset

The dataset included vehicle category, length, width, instantaneous speed, acceleration and deceleration and lateral position of all the vehicles. The arterial section was a 6-lane divided road (three lanes in each direction) located on a bridge, as shown in Figure 3-2. The road geometry was uniform and away from intersections, bus stops, on-street parking and other side activities that might affect driver behaviour. Furthermore, the pedestrian walkway was segregated by a barrier, which limited the interaction between the vehicle traffic and pedestrians. A video camera mounted on the roof of a nearby building was used for collecting traffic video.

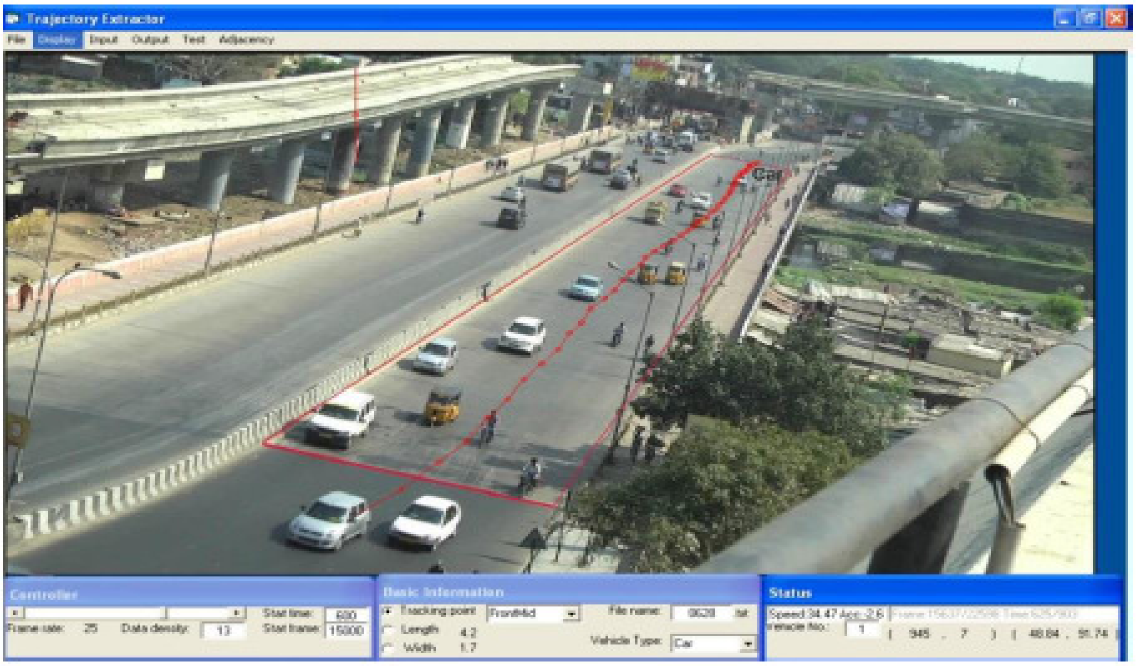


Figure 3-2 Urban arterial in Chennai for vehicle trajectory data estimation {Source: Kanagaraj et al. (2015)}

The Trajectory Extractor tool developed by Lee et al. (2008) was used to obtain the information on coordinates, dimensions and vehicle type. The dataset was extracted for 30 minutes between 2:45 PM and 3:15 PM on Thursday 13th February 2014. The authors selected that time-window as it apparently represented medium level traffic flows and showed both vehicle following and lateral movement behaviours. There were around 3,000 vehicles during the data collection period, and trajectories for all the vehicles were extracted at a 0.5-sec resolution. The average longitudinal space-mean speed was 21.17 kph during the observation period. The traffic included various categories of vehicles such as passenger cars (26.7%), motorcycles (56.4%), light commercial vehicles (LCV) (1.3%), heavy vehicles such as trucks and buses (3.5%) and auto-rickshaws (12.1%).

3.3.2 Homogeneous Traffic Dataset

Data for the homogeneous traffic was obtained from the NGSIM project (FHWA 2007a; Kovvali et al. 2007). Ideally, the mixed traffic characteristics on an arterial should be compared with that of another arterial with similar geometric and traffic conditions in homogeneous traffic. While the arterial under consideration in mixed traffic was devoid of intersection and other side activity affects, the arterial data (Lankershim - Boulevard) available from NGSIM has four intersections within a distance of 500m (FHWA 2007b). Therefore, in this study, the trajectory data from the U.S. Highway 101 (also called the Hollywood Freeway) in the Universal City neighbourhood of Los Angeles was used for comparison purposes (FHWA 2007c). The site was approximately 640m in length, with five lanes throughout the section. An auxiliary lane was also present through a part of the corridor. Figure 3-3 shows the aerial photograph and the schematic diagram of the study area.

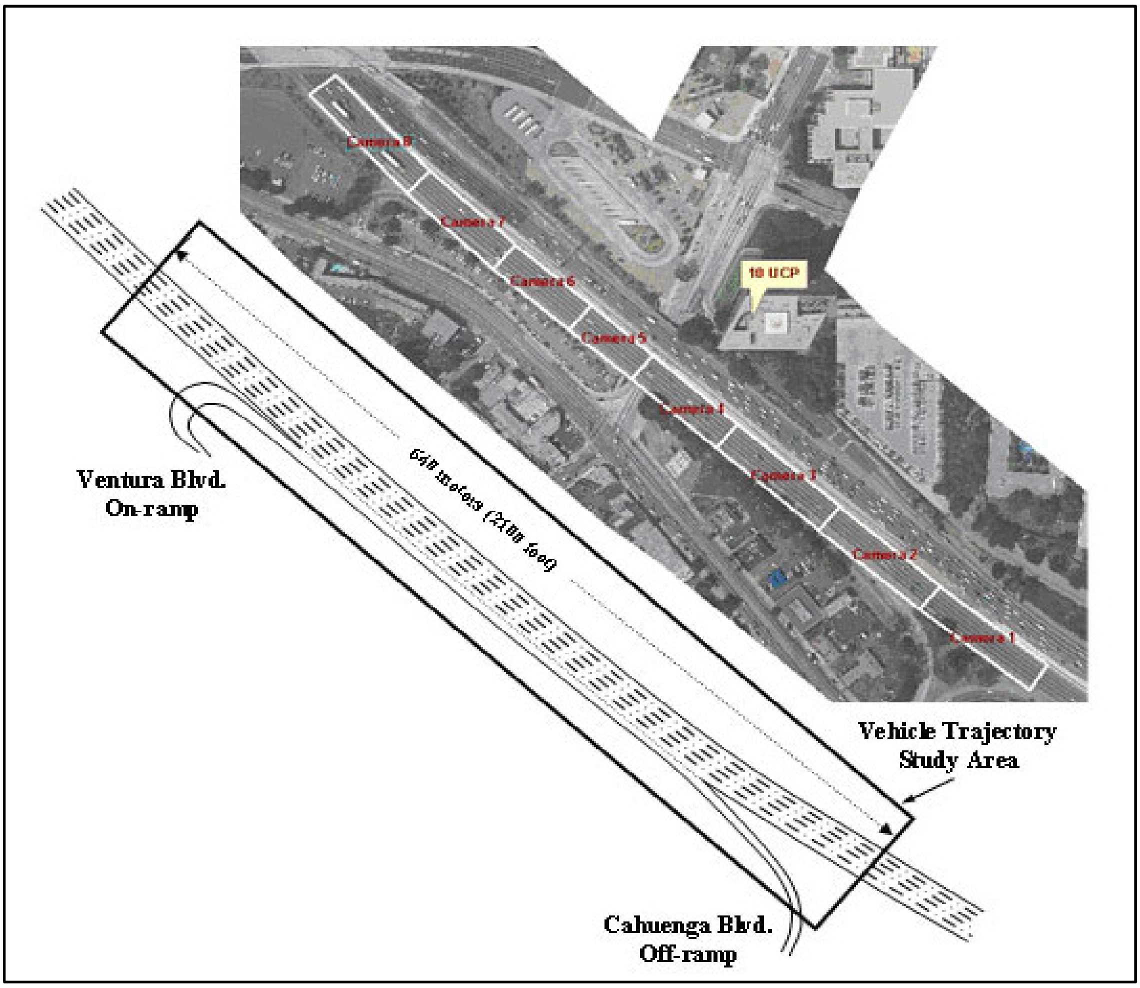


Figure 3-3 Aerial photograph showing the extent of the US 101 study area. The schematic drawing on the bottom shows the number of lanes and the location of the on-ramp and the off-ramp {Source: (FHWA 2007c)}

Eight synchronised digital video cameras were used to record vehicles passing through the study area. Then the trajectory information was extracted using NG-VIDEO, a customised software application developed for the NGSIM program. The data collected at 0.1-sec[15] resolution for a 15 min period between 7:50 AM and

[15] Although the homogeneous traffic dataset has been collected at 0.1 sec interval, it has been aggregated and only 0.5sec data was used for comparative purposes.

8:05 AM on Wednesday 15th June 2005 were analysed in this chapter. A total of 2,169 vehicles were observed during the 15 min period. The time mean speed and space mean speed were 44.90 kmph and 41.06 kmph respectively, which were significantly higher than that in the mixed traffic dataset.

The traffic consisted of cars (96.2%), motorcycles (1.4%) and trucks (2.4%). As may be seen in Figure 3-4, the distribution of modes is not as diverse as that in the mixed traffic dataset. While cars are predominant in the homogeneous dataset, motorcycles take the lion's share in the mixed traffic dataset.

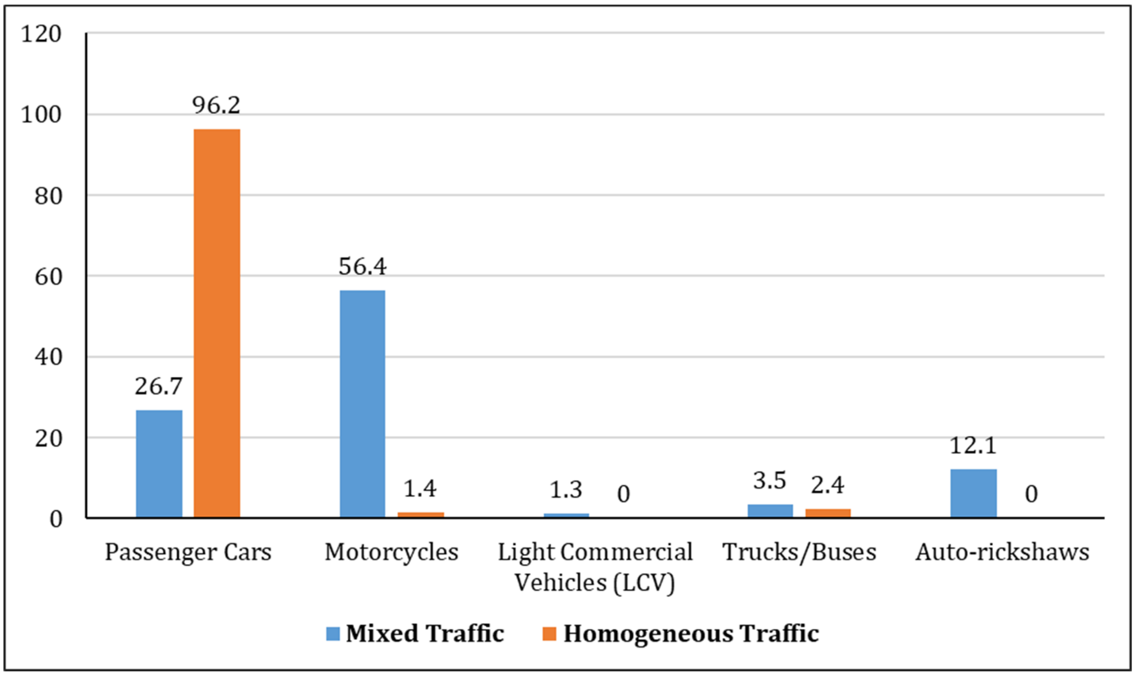


Figure 3-4 Traffic composition in mixed and homogeneous datasets

3.4 DATA ANALYSIS

Substantial lateral movement and lack of lane discipline characterise the mixed traffic. Figure 3-5 shows the variation in the lateral position of vehicles. The x-axis represents the lateral position (m) of the vehicles when they were first observed in the video. The y-axis represents their position (m) after 10 sec. There is a substantial scattering of the data points, particularly around the centreline of the road, indicating extensive and continuous lateral movement of vehicles.

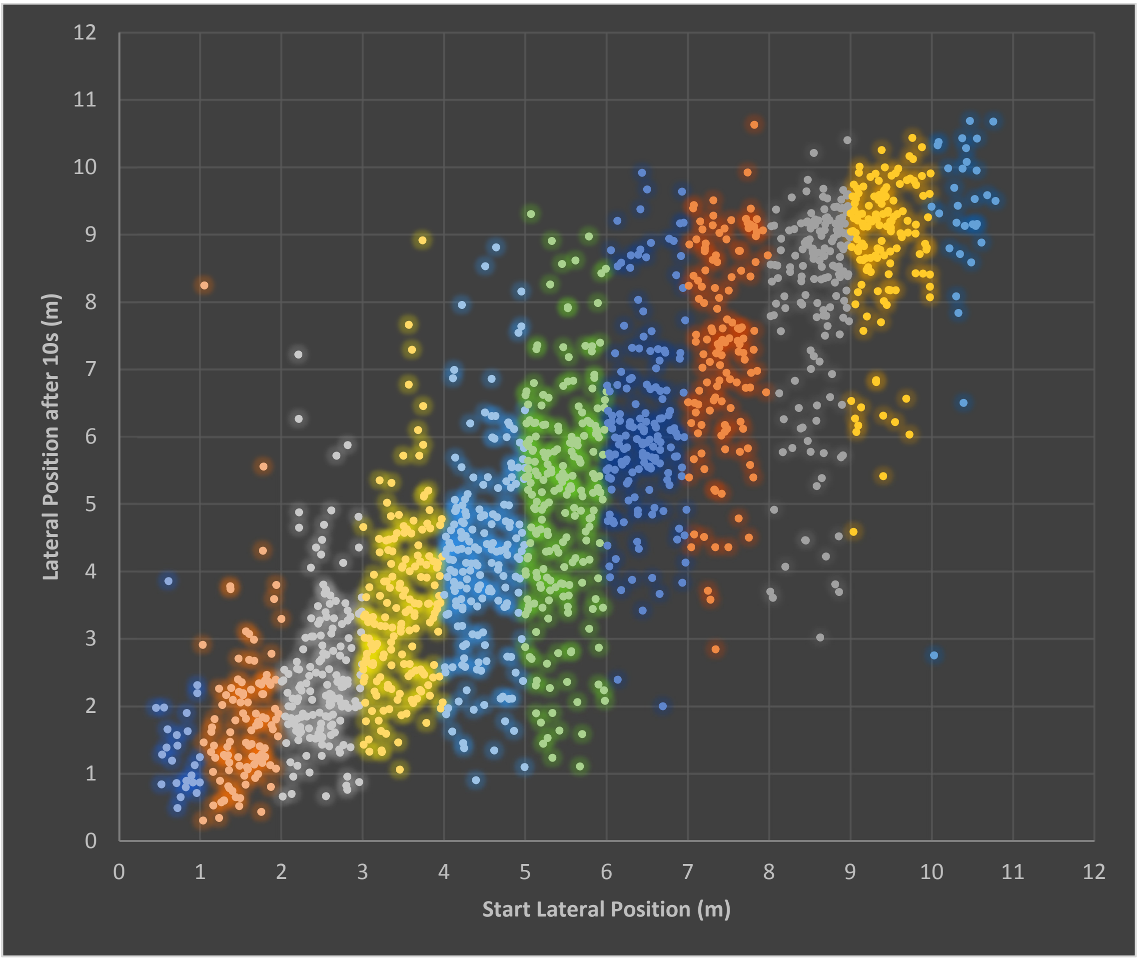


Figure 3-5 Variation in the lateral position of vehicles

Furthermore, we estimated a correlation matrix between the lateral positions[16] of the vehicles at a 3-sec[17] time interval (Table 3-1). The table shows that as time

[16] The lateral position was measured with respect to the edge of the road. Therefore, it is always positive. However, lateral movement (the difference between successive lateral positions) could be either +ve or -ve, depending on whether the vehicle is swerving towards left or right from its position in the previous timestamp.

elapses, the correlation with the starting lateral position, i.e. *Pos0* (the initial lateral position when a vehicle appears for the first time in the video) decreases. The correlation between *Pos3* (position of vehicles after 3 seconds of travel) and *Pos0* is around 0.98 indicating almost perfect correlation. However, the correlation decreases to 0.729 when the vehicles travel for 21 seconds. Although this value is still high, the significant decrease in correlation indicates that many drivers frequently change their lateral positions to keep moving in mixed traffic.

Table 3-1 Correlation matrix for lateral position of vehicles

Time (sec)	Pos0	Pos3	Pos6	Pos9	Pos12	Pos15	Pos18	Pos21
Pos0	1							
Pos3	0.962	1						
Pos6	0.913	0.987	1					
Pos9	0.869	0.968	0.986	1				
Pos12	0.833	0.946	0.967	0.986	1			
Pos15	0.786	0.929	0.949	0.969	0.987	1		
Pos18	0.763	0.913	0.934	0.954	0.973	0.991	1	
Pos21	0.729	0.896	0.921	0.941	0.961	0.979	0.989	1

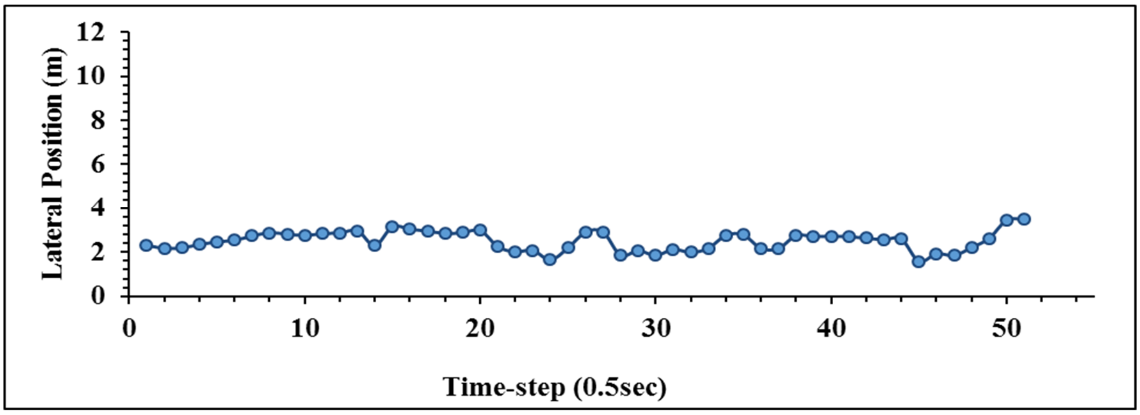
As noted earlier in Chapter 2.3, the Hurst exponent serves as an excellent statistical measure as it quantifies the fluctuations by estimating the relationship between the decreasing rates of autocorrelations and the increasing lag between pairs of values (Hurst 1956). The following sections explain the application of the Hurst exponent on both the trajectory datasets and evaluates the effect of vehicle type and average lateral position on fluctuations in speed and lateral movement of vehicles in mixed traffic.

[17] The 3-sec interval was only considered to illustrate the magnitude of the change in lateral positions.

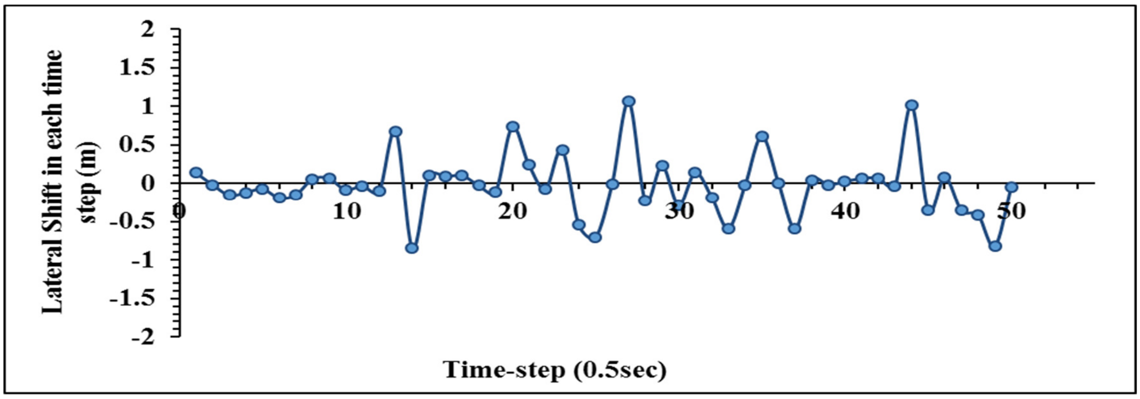
3.4.1 Hurst Exponent Demonstration

The rescaled R/S method was used to quantify the fluctuations in speed and lateral movement of vehicles. Figure 3-6 (a) shows the fluctuating lateral position of a motorcycle in different time steps. This fluctuation may not be a problem if the lateral movement, which is the difference between lateral positions in consecutive time steps, is approximately constant throughout. However, the lateral movement profile of the same vehicle presented in Figure 3-6 (b) shows that the movement is irregular and complex. The vehicle swerves from left to right and vice-versa in successive time steps and does not maintain the same lateral position for a long time. The calculated Hurst exponent for the lateral movement time series (H_{lat}) of this vehicle was 0.34, i.e. less than 0.5. It confirms the fluctuating behaviour of the time series. As the lateral clearance distance between vehicles in mixed traffic is typically low due to lack of lane discipline, even slight irregular movements may lead to chaotic traffic conditions. Further, using Equation 2.3, the theoretical predictability index (*PI*) of such time series can be calculated as 0.32, which is very low. Advanced techniques (e.g. non-linear and machine learning) would be needed to make predictions for this vehicle.

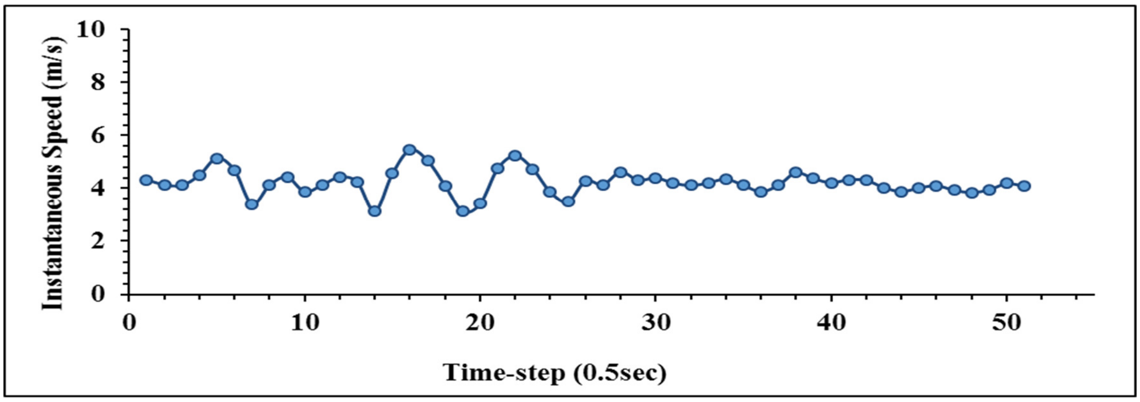
The longitudinal speed profile of this motorcycle shown in Figure 3-6 (c) is also irregular (standard deviation of 0.5 m/s) although the vehicle travels at a low speed of around 4.5 m/s. The calculated Hurst exponent for the speed time series (H_{speed}) was 0.38 (with PI = 0.24) which further confirms the haphazard movement of this vehicle. The lower H values indicate that the vehicle keeps moving in the traffic stream by substantially varying the speed and lateral movement in successive time steps. The plots of lateral position vs. speed and lateral movement vs. speed are included in Appendix B.



(a) Lateral position profile



(b) Lateral movement profile (H=0.34 and PI =0.32)



(c) Instantaneous longitudinal speed profile (H=0.38 and PI=0.24)

Figure 3-6 Speed and lateral movement profiles of a motorcycle in mixed traffic

In the same manner as discussed above, the Hurst exponent values for all vehicles were estimated in both mixed and homogeneous streams. Figure 3-7 shows the frequency distribution of the Hurst exponent values in mixed traffic. Most of the vehicles have H_{speed} values in the range of 0.6 to 0.9 and H_{lat} between 0.6 and 1. The H values above 0.5 indicate a good trend in financial markets (Bo and Rashed 2004).

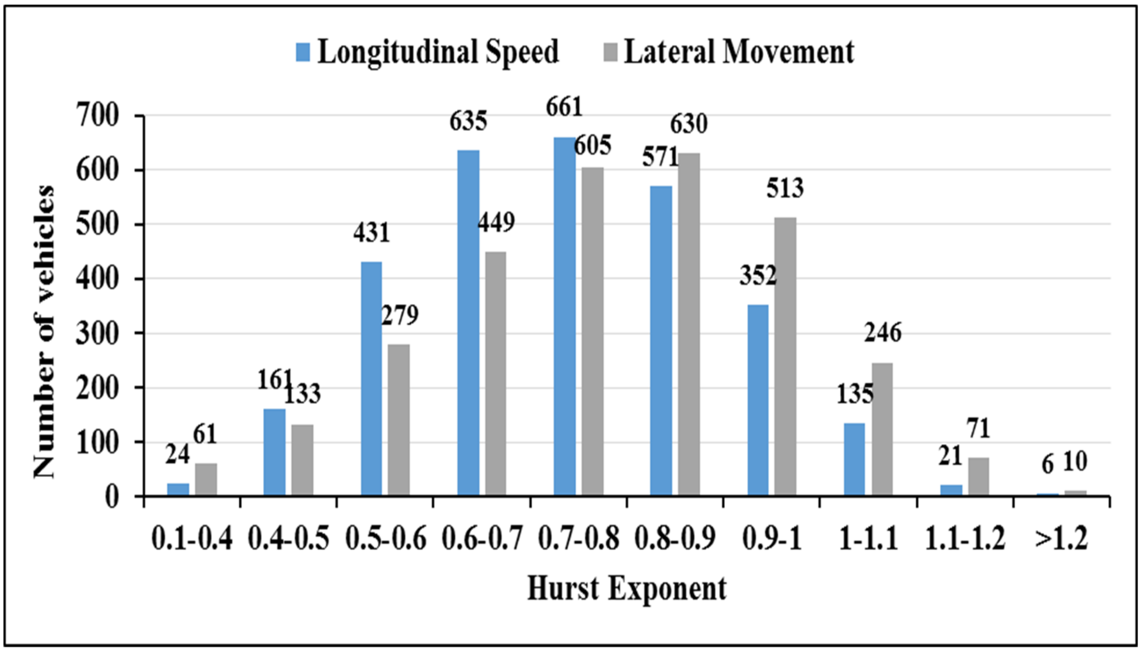


Figure 3-7 Frequency distribution of Hurst exponent in mixed traffic

In the current mixed traffic dataset, around 93 percent of the vehicles have both H_{speed} and H_{lat} values greater than 0.5. In fact, the average H_{lat} and H_{speed} values of all vehicles was 0.78 and 0.74 respectively, indicating a decent overall trend. Furthermore, using Equation 2.3, the average *PI* of lateral movement and speed was calculated as 0.56 and 0.48 respectively.

Only 7 percent of the total number of vehicles have *H* values less than 0.5, indicating they are all anti-persistent. It is worth mentioning here that a very small percentage of the vehicles have H greater than 1, which could potentially be due to the lower number of data points of the vehicles in the dataset (some vehicles were observed for shorter duration, say as low as 10-sec). Having said that, values of H greater than 1 can sometimes be obtained although there would not be any significance using the conventional fractal geometric approach (Racine 2011).

In the homogeneous dataset, 218 out of 2,169 vehicles (about 10 percent) used either auxiliary lanes or on-ramps for at least some time duration. As it was essential to change lanes for these vehicles, they are naturally expected to have substantial lateral movement. These vehicles were not considered in the analysis as the focus of the chapter is to evaluate the fluctuations at a mid-block section. Figure 3-8 shows the frequency distribution of H_{lat} and H_{speed} for the remaining 1,951 vehicles. Most vehicles have H_{lat} and H_{speed} in the range of 0.8 to 1, with average values being 0.93 and 0.91, signifying the trend-reinforcing property for most of the vehicles. Furthermore, the average PI of lateral movement and speed were calculated as 0.86 and 0.82 respectively, indicating a high predictive ability of microscopic characteristics in the homogeneous dataset. There were more trajectory points of every vehicle in the homogeneous dataset than that in the mixed traffic dataset, which could be due the differences in the extent of the study areas. The number of vehicles with $H > 1$ is, therefore, less in the homogeneous dataset when compared to mixed traffic.

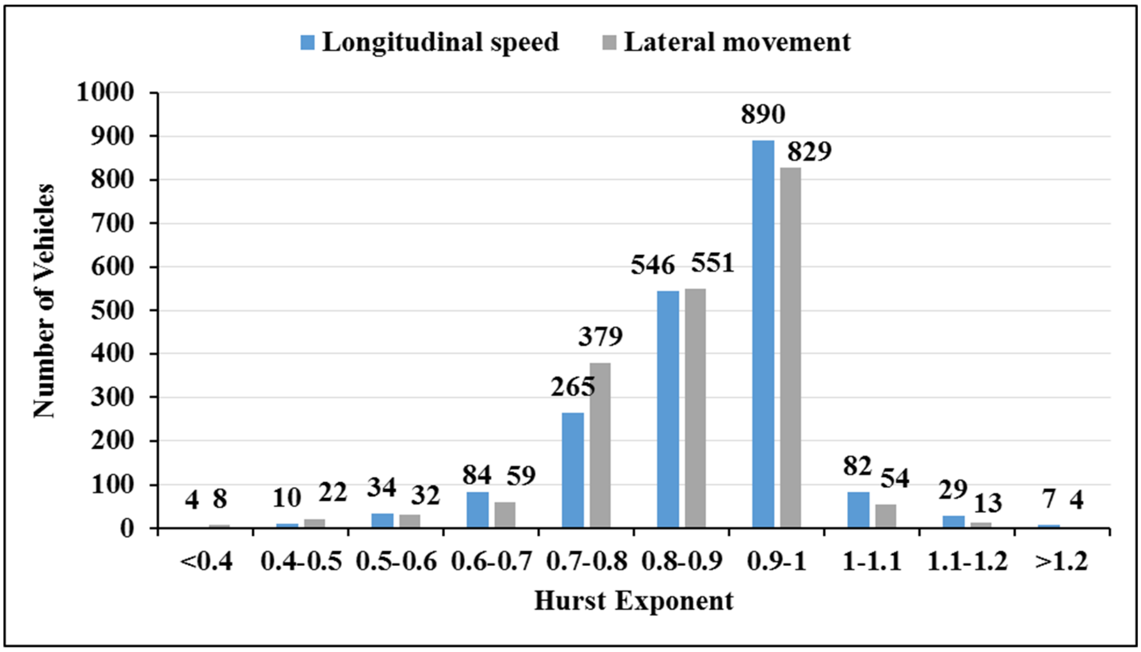


Figure 3-8 Frequency distribution of Hurst exponent in homogeneous traffic

3.4.2 Effect of Vehicle Type on the Hurst Exponent in Mixed Traffic

Several studies in the literature acknowledged the importance of heterogeneity in microscopic traffic models (Ossen and Hoogendoorn 2011; Aghabayk et al. 2014; Ravishankar and Mathew 2011). For example, the behaviour of a car following a heavy vehicle like bus/truck will be different from a bus/truck following a car under similar conditions. Larger vehicles will have better sight distance and field of vision and hence will behave differently. Therefore, these studies suggested that heterogeneity concerning vehicle type should be accounted to obtain better models.

As mixed traffic constitutes several types of vehicles with different physical and operating characteristics, this study investigated the effect of vehicle type on Hurst exponent and predictability. Table 3-2 shows the H_{speed} and H_{lat} values based on the vehicle type. The average H_{speed} is found to be higher for cars and lower for trucks.

On the other hand, H_{lat} of cars and motorcycles is higher when compared with other vehicle types. Nevertheless, all vehicle types have significantly higher H_{lat}. It should also be noted that the section of the road was just about 250 m and the vehicles were tracked only for around 20-25 sec. Longer road section (or more extended observation periods) may have given different results.

Table 3-2 Hurst Exponent and vehicle type

Vehicle type	Sample Size	H_{speed}					H_{lat}				
		Mean	Min	Max	Std. dev.	Mean PI_{speed}	Mean	Min	Max	Std. dev.	Mean PI_{lat}
Car	801	0.78	0.28	1.30	0.17	0.56	0.78	0.20	1.28	0.19	0.56
LCV	39	0.74	0.38	1.08	0.15	0.48	0.75	0.45	1.18	0.17	0.50
Motorcycle	1,692	0.73	0.32	1.27	0.15	0.46	0.76	0.19	1.22	0.18	0.52
Bus	60	0.68	0.40	1.04	0.17	0.36	0.75	0.38	1.17	0.17	0.50
Truck	45	0.62	0.48	0.99	0.16	0.24	0.72	0.48	0.95	0.16	0.44
Auto - Rickshaw	363	0.68	0.36	1.15	0.15	0.36	0.71	0.28	1.16	0.17	0.42

Cars have the strongest trend-reinforcing property in both speed and lateral movement. Their speeds and lateral movements are more predictable than those of the other vehicle types. This can be attributed to the superior operating capabilities of cars compared to other categories of vehicles. Similar stronger trends in speed and lateral movement can be seen for LCVs.

Motorcycles have a good long-memory property in speed and an even better LRD in lateral movement. It would be interesting but not surprising to note that motorcycles have a higher average lateral movement than other vehicles in mixed traffic (Kanagaraj et al. 2015). However, it seems that the lateral movement of motorcycles in different time steps is consistent, i.e. either increasing or decreasing. They are better predictable than other vehicle types (excluding car). It could be due to their smaller physical size which helps them to manoeuvre easily through the gaps in a traffic stream.

For the remaining vehicle categories, H_{speed} values are lower than 0.7, although H_{lat} values are higher than 0.7. Buses and trucks must uphold a strong trend in the lateral movement because of their bulky size. If they fluctuate their lateral position severely, it will lead to complete chaos in the traffic flow. Therefore, they fluctuate their speeds to not move laterally. This finding is consistent with earlier studies that reported that trucks are significantly more conservative than cars as they maintain higher headways (Punzo and Tripodi 2007; Ossen and Hoogendoorn 2011).

Finally, auto-rickshaws, because of the tendency of drivers to pick up or drop off the passengers and also due to inferior operating capabilities, do not maintain a stronger trend in either speed or lateral movement. Moreover, they change their existing path to avoid impeding the fast-approaching vehicles from the rear (Munigety et al. 2014a). In most cases, the sudden lateral movement of auto-rickshaws is followed by a substantial fluctuation of speed. The findings in this study are consistent with an earlier study that reported that the lateral acceleration is very high in case of auto-rickshaws and lower in case of cars (Mahapatra and Maurya 2013).

3.4.3 Effect of Average Lateral Position on the Hurst Exponent in Mixed Traffic

Breiman et al. (1977) observed that long-range correlation in speed data is lane-dependent, increasing as one goes from outer to inner lane. The lateral position distribution of vehicles is not uniform in mixed traffic. Kanagaraj et al. (2015)

observed that auto-rickshaws and motorcycles prefer to stay in the leftmost lane[18], cars are likely to travel in the rightmost lane and heavy vehicles in the middle lane. Figure 3-9 shows the influence of the overall average lateral position of vehicles on the Hurst exponent values. Lane-1 refers to the leftmost lane; lane-2 denotes the middle lane; and lane-3 refers to the rightmost lane. The average Hurst exponents for corresponding vehicles in these lanes was evaluated. A two-sample analysis of variance (ANOVA) test was conducted to check whether the *H* values are statistically different. The results from Table 3-3 show that the mean H_{lat} is different for Lanes 1 and 2, which could be due to the side friction for Lane 1 created by the edge of the road and the natural tendency of drivers to stay away from it.

On the other hand, the mean H_{speed} value in Lane 3 is significantly higher than those for the other lanes. This could be due to the higher number of cars cruising on this lane.

Table 3-3 ANOVA results for lane-wise fluctuations of lateral movement and speed

Sample 1	Sample 2	Lateral movement			Speed		
		F	Fcr (p<0.05)	p ($F<=F_{cr}$)	F	Fcr (p<0.05)	p ($F<=F_{cr}$)
Lane 1 (df = 927)	Lane 2 (df = 1273)	1.300	1.106	< 0.01*	1.086	1.106	0.089
Lane 2 (df = 1273)	Lane 3 (df = 794)	1.099	1.112	0.072	1.159	1.110	0.01*

* indicates that the samples are statistically different at p = 0.05

[18] Note that driving in India is on the left side of the road.

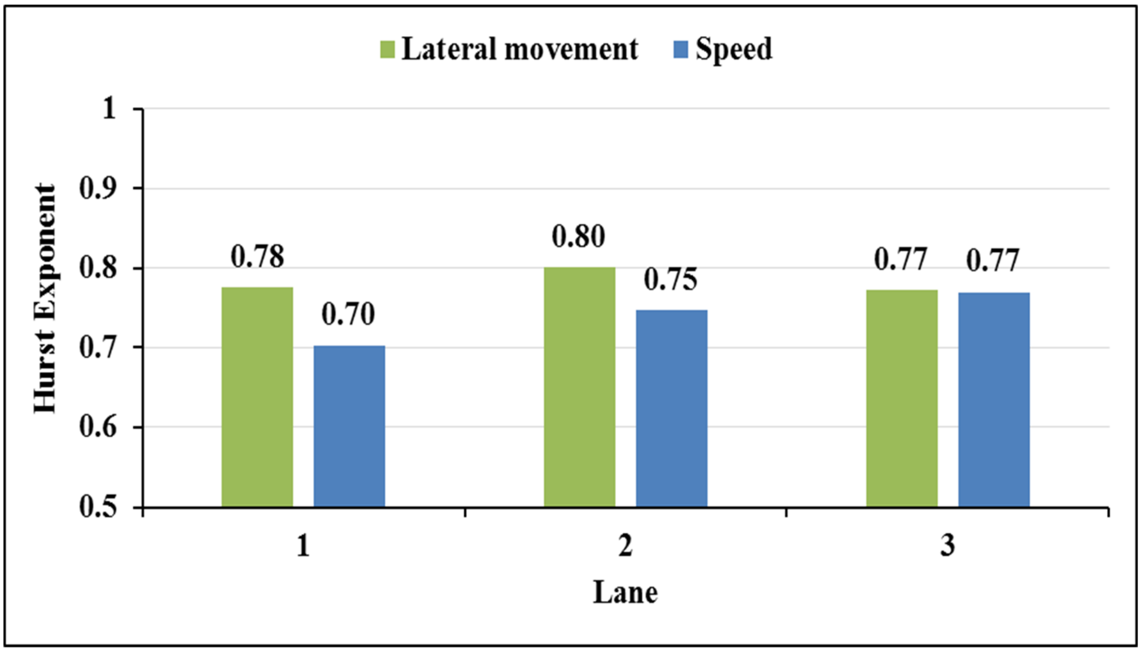


Figure 3-9 Effect of overall average lateral position on Hurst exponent (Lane number v/s Hurst exponent).

We divided the roadway into 1 m width strips to better capture the effect of lateral position on the Hurst values. Figure 3-10 shows that the H_{lat} is much lower at both extreme strips near the footpath and the median/divider, indicating that vehicles fluctuate their lateral position to avoid side friction. However, the mean H_{speed} increases gradually from left to the right side of the road. As stipulated earlier, slow-moving vehicles like auto-rickshaws typically move on the leftmost side of the road, while vehicles with superior operating capabilities like cars cruise on the rightmost side. This could be the reason for the contrasting H_{speed} values on the left and right sides of the road.

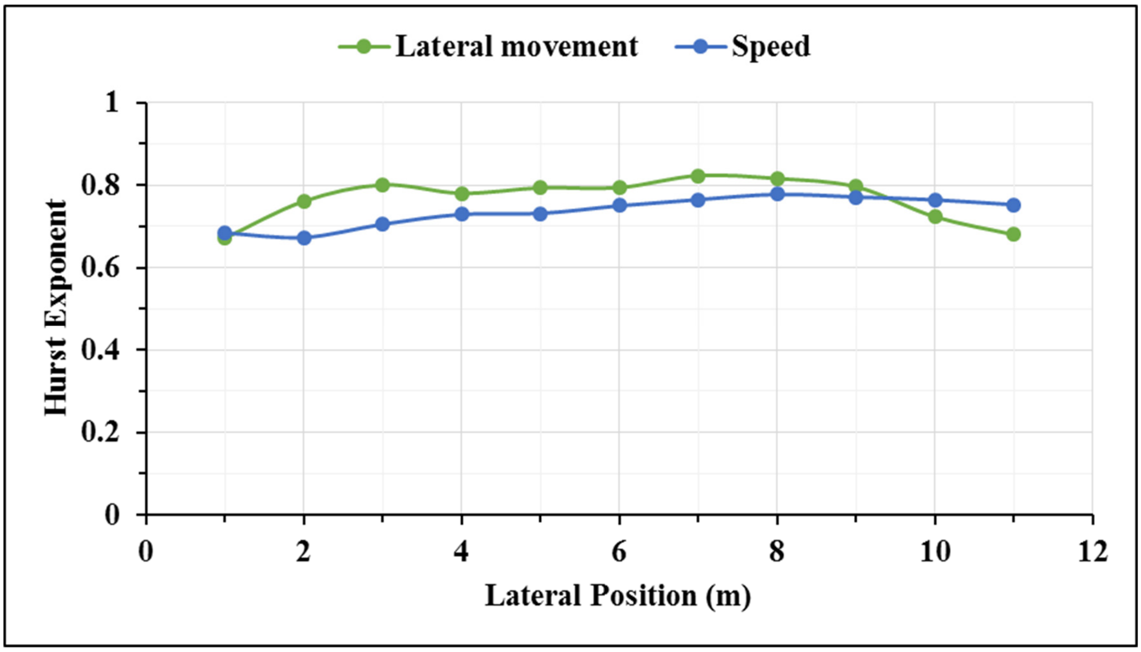


Figure 3-10 Effect of overall average lateral positions on Hurst exponent (Strip number v/s Hurst exponent).

3.5 SUMMARY

In mixed traffic conditions, there are only a few studies exploring vehicle trajectory data. This could be due to the enormous effort required to obtain trajectory data. However, with the advent of several video image processing software, the trajectory extraction process has become less cumbersome. Therefore, more studies that address this topic are expected to emerge.

Different issues in mixed traffic complicate the implementation of car-following or lane-changing theories. It is challenging to develop effective identification procedures to study the interactions and the resulting chaotic phenomena in traffic time series data, particularly in mixed traffic (Lan et al. 2005). The analysis of trajectory data on urban roads (non-freeways) forms a critical element in diversified research and policymaking decisions. Given this, the current chapter of the thesis analysed vehicle trajectory data in mixed traffic and then compared the fluctuations of speed and lateral movement with that of homogeneous traffic. We applied the Hurst exponent concept, a popular time series method in financial markets and a relatively unexplored method in the transportation domain.

The average Hurst exponent values of 0.78 (Predictability Index, PI=0.56) and 0.74 (PI=0.48) for lateral movement and speed showed good trend-reinforcing property in the mixed traffic. However, these values were still lower than that in the lane-based conditions of the USA, where the corresponding H values were 0.93 (PI=0.86) and 0.91 (PI=0.82). Thus, the theoretical predictabilities of lateral movement and speed in mixed traffic were lower than that of homogeneous traffic by 54% and 70% respectively.

As different types of vehicles characterise the mixed traffic, the fluctuations based on vehicle type were also evaluated. Cars, LCVs and motorcycles were found to have stronger trend-reinforcing property and higher predictability. On the other hand, we found heavy vehicles to have strong LRD in lateral movement but comparatively weaker LRD in speed. This finding indicated that the heavy vehicles fluctuate their speed (stop-and-go) to uphold strong trend in lateral movement. Finally, auto-rickshaws, had large fluctuations. This could be because of their inferior operating capabilities. Furthermore, they change their existing path to avoid impeding the fast-approaching vehicles from the rear. The average lateral position of the vehicles was also found to be affecting the fluctuations, with the vehicles that were closer to the road edge and the median having severe fluctuations compared to the ones that travel in the middle of the roadway.

Data size is a limitation of the Chaos and Fractal theories. There is no clear-cut guideline on the minimum data size required. Nevertheless, large datasets might be necessary to achieve more accurate results (Williams 1997; Sivakumar 2000). The problems of data size and data sampling frequency arise in almost every field of natural and physical phenomena, and vehicle trajectory data is not an exception to it. The trajectory time series analysed in this study had only a few observations for vehicles for analyses that resulted in H values greater than 1. Hurst exponent evaluation needs more observations as the length of the time series might affect the Hurst exponent value. For this reason, vehicles with a very low number of observations (less than 20) were not considered in this study. As the primary aim of the chapter is to demonstrate the application of the Hurst exponent concept to vehicle trajectory data, the variation of the number of observations for different vehicles was not considered.

The trajectory extraction process sought by Kanagaraj et al. (2015) was semi-automated, as a human operator was needed to click the mouse pointer to identify the edges of vehicles on the computer screen. This process is still tedious and involves enormous manual effort. Trajectory data are prone to errors, and so it should be inspected appropriately before using (Punzo et al. 2011; Coifman and Li 2017). Kanagaraj et al. (2015) used a locally weighted regression smoothening technique to overcome missing observations (occlusion) and reduce measurement errors. One of the limitations of the video-based trajectory data is the lack of observation for more extended periods. Therefore, alternative techniques for trajectory extraction such as GPS enabled vehicles, emerging technologies of sensor networks such as data from mobile phones, DSRC (Dedicated Short-Range Communications) should be analysed for more accurate and realistic analysis.

Furthermore, the current study should only be treated as a preliminary step towards the application of the Fractal theory on trajectory data. It is to be noted that the results are based on just the two datasets each of 15-min duration. So the results cannot be generalised, and data from several locations with distinct driving behaviours should be evaluated.

CHAPTER 4. FLUCTUATIONS IN MACROSCOPIC TRAFFIC FLOW VARIABLES: THE APPLICATION OF LOOP DETECTOR DATA

4.1 INTRODUCTION

Urban highways and freeways act as arteries and veins to carry people out of the heart of the city to the remaining parts. High performing arteries and veins are essential for an excellent circulatory system. Similarly, uncongested and reliable highways and freeways are vital for a transport system to be efficient. Therefore, urban freeways are often termed as the single greatest element to cure the problems of a city (DiMento 2009). Freeways in developed countries are often equipped with better infrastructure that can reduce the occurrences of congestion, thus easing the smoother movement of traffic. For example, automatic incident detectors (AID) can identify the flow breakdowns (congested conditions) based on near real-time traffic data from loop detectors. Then centralised transport management centres (TMC) can initiate the incident and demand management measures such as variable speed limits (VSL) and ramp metering, which can divert traffic to alternate routes. This proactive approach can yield significant economic, social, environmental, and safety benefits to the society.

Traffic system is dynamic as it is time-dependent and is traditionally a complex system, because of the continuous interactions between different vehicles (Nair et al. 2001). These interactions, in turn, result in the fluctuations of macroscopic variables such as; speed, flow, and occupancy. Small variations in traffic flow can make the difference between free flow and congestion (van Zuylen et al. 1999).

Furthermore, fluctuations in these macroscopic variables have implications on road safety, crashes, flow predictions, fuel consumption, emissions, travel time, and reliability. For a transport system model to be efficient and reliable, it is essential to understand these fluctuations, to simulate and forecast traffic behaviour. Further, analysis of fluctuations in traffic flow data can offer some interesting insights on congestion patterns, and their relationships with factors such as road geometry, time of day, incidents, crashes, and weather.

Investigating traffic patterns can significantly reduce congestion. Lin et al. (2014) analysed freeway traffic patterns using a data mining based method called the "frequent pattern tree". They found that season, the day of the week, and the hour of the day have impacts on macroscopic traffic characteristics namely speed, flow and density. Breiman et al. (1977) observed solid long-range correlation (correlation does not fall rapidly with the lag) in speed data of a freeway in the USA. They found that the correlation is lane-dependent, increasing as one goes from the outer to the inner lane.

Traffic flow and speeds show intense oscillations[19] depending on the time of day and demand. Several techniques were found in the literature to assess oscillations in traffic flow. Some of them are, spectral analysis (Lam and Rothery 1970), cumulative curves (Cassidy and Windover 1995), and wavelet transformation (Zheng et al. 2011).

Evaluating the temporal dependence of traffic flow and speed would be highly beneficial in understanding their fluctuations. As noted earlier in Chapter-2, the

[19] Traffic oscillations are stop-and-go (speed drops, rises, and drops again over time) or slow-and-go conditions in congested traffic. Moreover, this observed pattern should propagate upstream against the direction of traffic flow so as to be referred to as an oscillation (Zielke et al. 2008).

Fractal theory is instrumental in studying the irregularities in ime series data by assessing the long-range temporal dependencies (Mandelbrot 1967). The Hurst exponent serves as an excellent statistical measure to evaluate the long-range dependency (LRD) of a time series, by estimating the relationship between the decreasing rates of autocorrelations and the increasing lag between pairs of values (Hurst 1956).

This chapter presents the analysis of fluctuations in macroscopic traffic variables (speed and flow to be precise) using the Hurst exponent method. We utilised traffic data from the M4 motorway in Sydney, Australia for demonstration purposes. In this chapter, we discuss the spatial and temporal variation of LRD for flow (H_{flow}[20]) and speed (H_{mac_speed}[21])[22] at several traffic monitor sites. Furthermore, we explore the effects of road geometry on flow and speed fluctuations.

4.2 STUDY AREA AND DATA DESCRIPTION

4.2.1 Study Area

Sydney is the largest city in Australia with an estimated population of 5 million. Traffic congestion is a major problem in the city, although it noticed a substantial drop in the congestion rankings in the recent years (TomTom 2017). The M4 is a 47-kilometre long freeway, running from Concord in the inner-west of the city to

[20] Small magnitude H values in flow can be observed on downstream links at signalised intersections, particularly when the measurement interval is smaller than the cycle time of the signal.
[21] Strong long-range dependence in speed, i.e. large H_{mac_speed} values can be observed after the occurrence of an incident or removal of a bottleneck. Incident does not necessarily imply an accident; vehicle breakdown, fire hazard, police pull over, sudden braking and lane changing in dense traffic can result in high H_{mac_speed} values.
[22] The notation is different from the one used in Chapter-3. In the current chapter, H_{mac_speed} refers to the Hurst exponent of speed, which is a macroscopic variable. However, in Chapter-3, speed was a microscopic variable because it was for each individual vehicle.

the Blue Mountains in the far west (Figure 4-1). There are 16 interchanges along the M4, including a connection to the M7 Motorway at Eastern Creek. The M4 enables safe and efficient travel between key regions of western Sydney and services over 100,000 commuters each day. The M4 provides various forms of ITS, such as; variable message signs (VMS), variable speed limit signs (VSLS) and vehicle detection systems, also known as loop detectors. Despite the application of ITS on the M4, incidents of different magnitudes still occur along this motorway. These types of incidents include breakdowns, towing, police activities, hazards, and accidents.

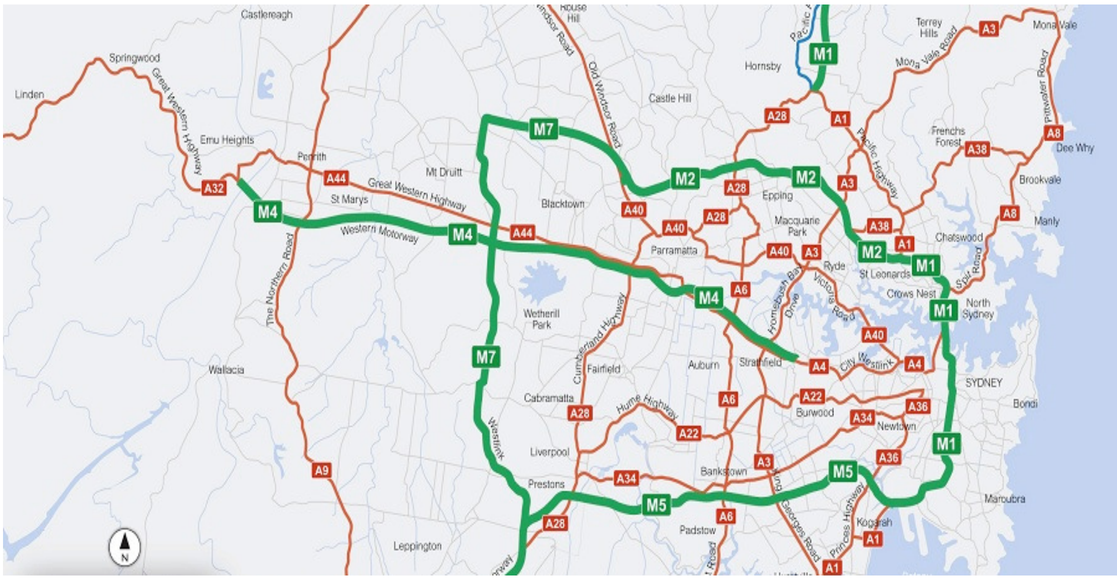


Figure 4-1 The M4 Motorway in Sydney {Source: RMS (2017) }

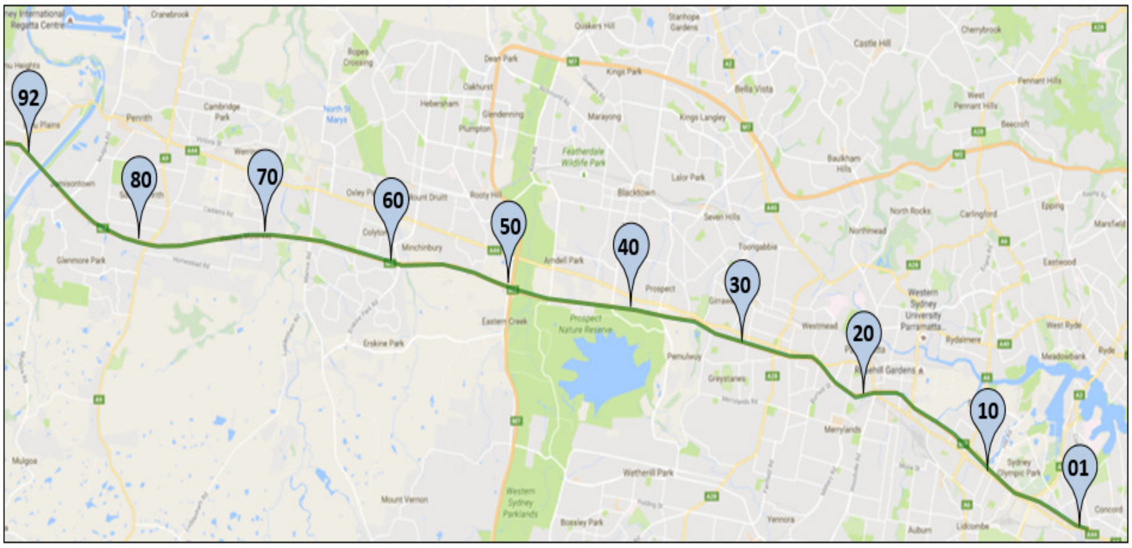


Figure 4-2 The M4 Motorway (with some of the monitor sites labelled)

4.2.2 Data Description

The data used in the current chapter were obtained from 92 monitor sites (Figure 4-2) in the westbound direction (away from the city) during February 2013. These monitor sites are spaced around 400m to 500m apart, with the site: 1 at the start of the Motorway and site: 92 at the end. Monitor sites 1 to 6 and 87 to 92 are on 2-lane road sections, sites 14 and 21 to 24 are on 4-lane sections, and the remaining sites are on 3-lane sections.

Each monitor site records traffic conditions, e.g., average speed, traffic flow rate, occupancy, every 30 seconds for each lane. Weighted average speed and the total flow across all lanes were calculated. To understand the temporal variations of the Hurst exponent for speed (H_{mac_speed}) and flow (H_{flow}), we divided the total observations of 1-month into smaller windows of 3-hours, non-overlapping and starting from 12 AM on the 1st of February 2013. The time windows 12 AM-3 AM,3 AM-6 AM and 9 PM-12 AM represent the off-peak periods; 6 AM-9 AM and 3 PM-6 PM signify morning and evening peak periods respectively; and 9 AM-12 PM, 12

PM-3 PM and 6 PM-9 PM indicate interpeak periods. Thus, we obtained 224 windows (= 28 days * 8 windows/day) of observations at each monitor site. In other words, we estimated 224 values of H_{mac_speed} and H_{flow} at each site.

Monitor sites 17, 18 and 19 were not functional throughout the data collection period. For a few sites in some instances, data were either not recorded or exhibited unrealistic values for some part of the day. Such records were not analysed in the study. The data cleaning exercise resulted in 1,348 erroneous time windows (6.5%) out of a total possible 20,608 windows (= 92 sites * 224 windows/site).

4.3 DATA ANALYSIS

We calculated the Hurst exponent for speed and flow time series for every 3-hour time window[23], at each monitor site for the entire one-month duration. We coded a MATLAB program to automate the process. Figure 4-3 and Figure 4-4 show the probability density function (PDF) plots of H_{flow} and H_{mac_speed}. The median value of H_{flow} is 0.78, while it is 0.66 for H_{mac_speed}. Using Equation 2.3, we can say that the flow PI is slightly higher than that of speed, which is expected because of the varying operating capabilities of vehicles and heterogeneous behaviour of drivers. Additionally, Figure 4-3 and Figure 4-4 show a wide variation in the values of Hurst for both flow and speed. Furthermore, the kernel density plots (the lines in red ink) estimated using an Epanechnikov kernel function (Silverman 1986) with

[23] The three-hour windows were primarily selected for mathematical convenience. Nevertheless, there is some practical significance for the selection of 3-hour time windows. For example, the time windows 12am-3am,3am-6am and 9pm-12am represent the off-peak periods; 6am-9am and 3pm-6pm signify the morning and evening peak periods respectively. Furthermore, 9am-12pm,12pm-3pm and 6pm-9pm indicate the interpeak periods.

20 bins show that the distributions are bimodal. This bimodality, which is strongly noticeable in Figure 4-4 could be due to the spatial and temporal variation of traffic dynamics, which is discussed in the subsequent sections.

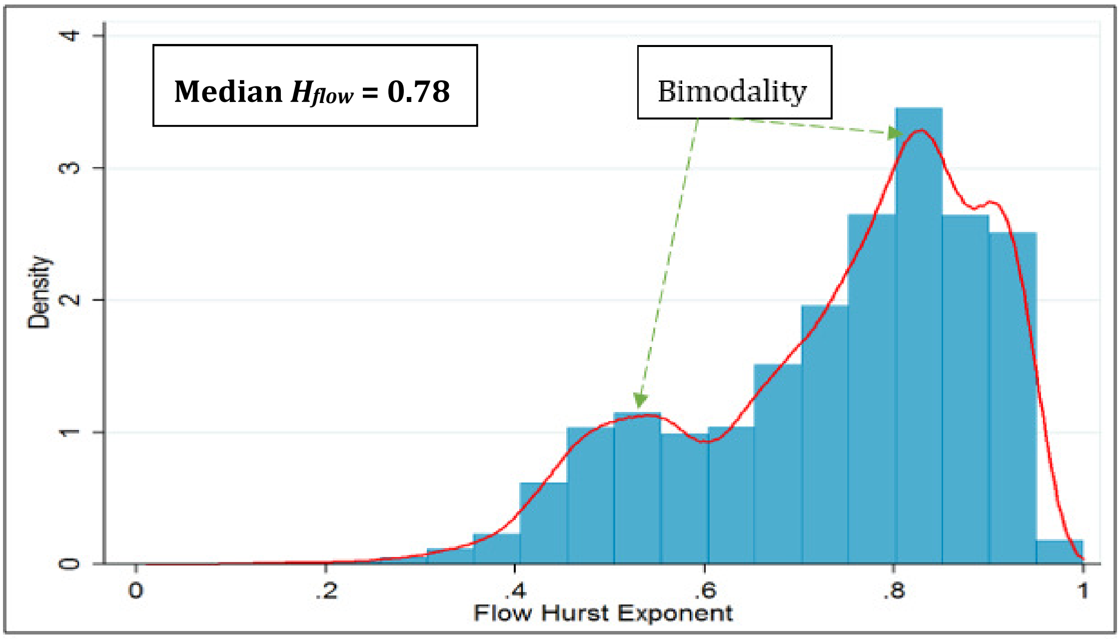


Figure 4-3 PDF and kernel density plots of H_{flow} of the M4 Motorway in the outbound direction from the city

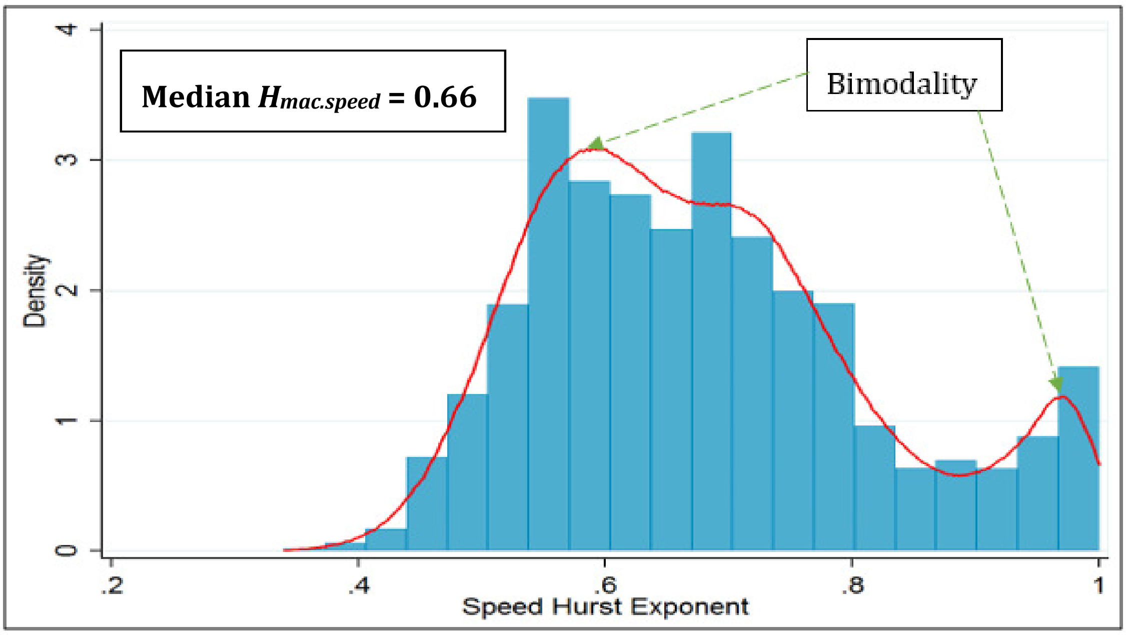


Figure 4-4 PDF and kernel density plots of H_{mac_speed} of the M4 Motorway in the outbound direction from the city

4.3.1 Weekday vs Weekend and Spatial Variation of Hurst Exponent

Heat maps depicted in Figure 4-5, and Figure 4-6 show the spatial and temporal variation of the Hurst exponent of flow and speed respectively. Each cell represents the corresponding H value spatially and temporally. Thick red cells specify time windows with high H values; solid blue cells indicate low H values and the white cells indicate the missing/erroneous values.

The heat map of H_{flow} (Figure 4-5) shows periodic light-blue bands across all the monitor sites, representing low H values every weekend (Saturday and Sunday). Single-factor ANOVA test showed that the two samples, i.e. weekdays and weekends are significantly different ($p < 0.01$) with a mean H_{flow} value of 0.76 on weekdays and 0.68 on weekends. Thus, weekdays had stronger trend-reinforcing property compared to the weekends, which further indicates that the weekday traffic flow (PI = 0.52), is better predictable than the weekend traffic (PI = 0.36). This could be due to the greater number of work-related trips on the weekdays, which are better predictable than the leisure trips that mostly occur on weekends. Furthermore, sites 1 to 6 displayed consistently lower H_{flow} values compared with other sites. Their low H values could be due to a consequence of their location, as they are located at the start of the Motorway and operate with two lanes.

On the other hand, the H_{mac_speed} heat map (Figure 4-6) presents extensive spatial and temporal variations. Monitor sites 1 to 60 have a more significant proportion of cells with high H_{mac_speed} values, than sites at the far west of the Motorway, i.e. 61 to 92. The mean H_{mac_speed} was calculated as 0.64 on weekends across several sites, while on weekdays, it was 0.69. Reinforcing this further, the ANOVA test shows that these two-sample means are statistically different ($p < 0.01$). The predictability index of speed on weekdays (PI = 0.38) is higher than weekends (PI = 0.28).

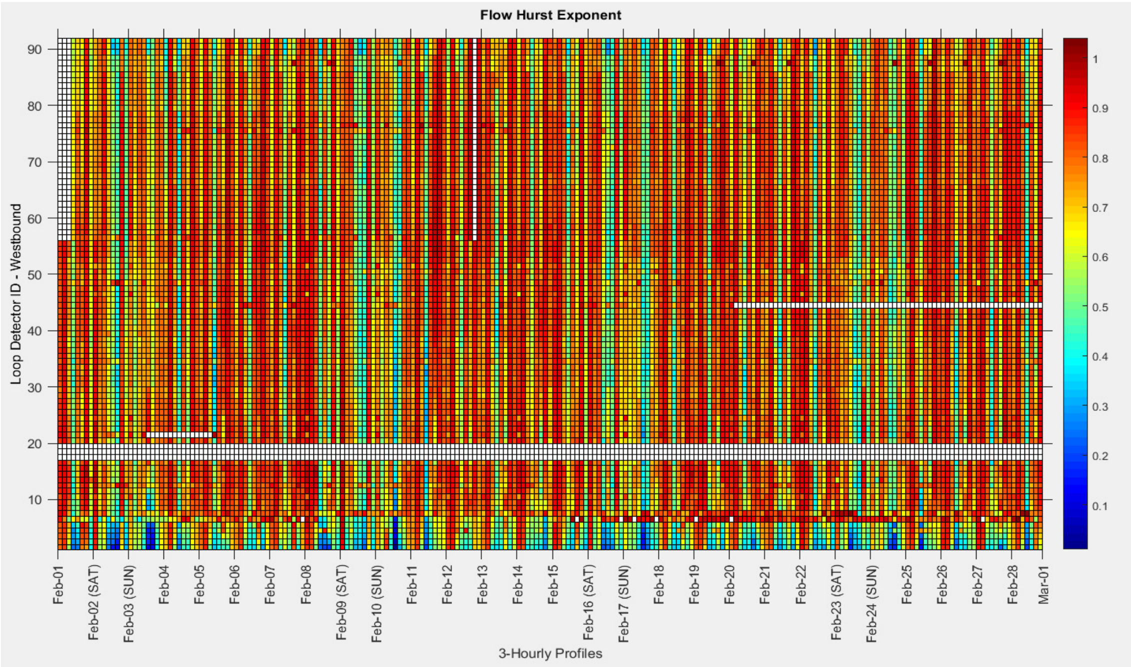


Figure 4-5 Heat map showing the spatial and temporal variation of H_{flow}

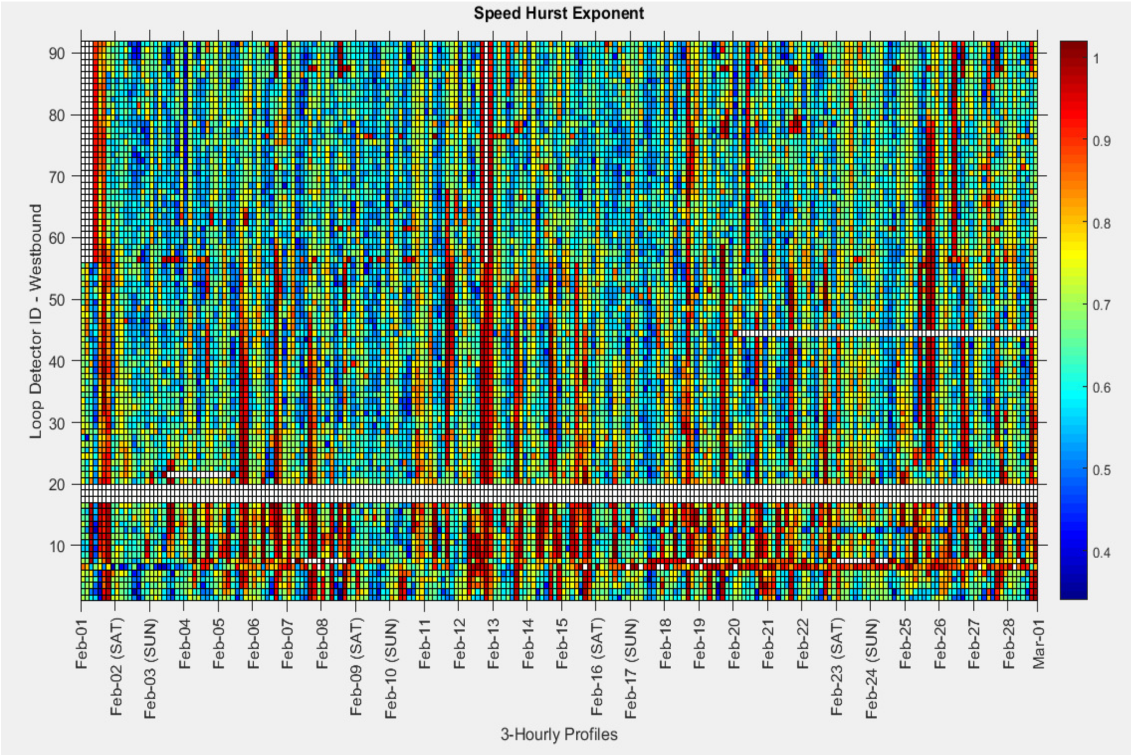


Figure 4-6 Heat map showing the spatial and temporal variation of H_{mac_speed}

4.3.2 Variation of the Hurst Exponent with Time of the Day

The variation of Hurst exponent with the time of the day provides useful insights for operational purposes. The predictability of flow and speeds can be ascertained such that appropriate information can be passed onto the road users. Figure 4-7 and Figure 4-8 are the box-plots of H_{flow} and H_{mac_speed} for all the detectors combined, showing the variation with the time of the day.

H_{flow} is high for all the time windows, except 9am-12pm and 3pm-6pm where the average values are 0.55 and 0.64, respectively. Consequently, the flow *PI* during these time periods is very low, when compared with other periods. Using single-factor ANOVA (F = 1854; F_{cr} = 2.01; p [F > F_{cr} = < 0.01]) and Tukey's posthoc honest significant difference (HSD)[24] tests (Tukey 1949), we observed that H_{flow} in each time-window is statistically different to every other time-window (l.o.s, α = 0.01).

H_{mac_speed}, on the other hand, is higher during the evening peak hours of 3pm-6pm. There is a high LRD during this time window, indicating that the speed series is persistent. A low value of speed is likely to be followed by another low value, which is a sign of traffic congestion. A high value followed by another high value is a potential scenario, but this is more likely to occur during the removal of a bottleneck or clearance of an accident site. High LRD results in a high *PI* of speed, compared with other time periods. The H_{mac_speed} for other time windows is generally low indicating more complexity, hence less predictability. Furthermore, single-factor ANOVA ($F = 462$; $F_{cr} = 2.01$; p [$F > F_{cr}$ = < 0.01]) and Tukey's HSD tests showed that H_{mac_speed} in each time-window is statistically different from that of every other time-window (l.o.s, α = 0.01).

[24] Tukey's HSD test is a single-step multiple comparison procedure that can be used on raw data or in conjunction with an ANOVA (post-hoc analysis) to find means that are significantly different from each other. It compares all possible pairs of means and is based on a studentised range distribution. Reject H_0: $\alpha_i = \alpha_j$ if

$$|t| = \frac{|y_i - y_j|}{\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}} > \frac{1}{\sqrt{2}} q_{\alpha,k,N-k}$$

where $q_{\alpha,k,N-k}$ is the upper 100*(1-α)th percentile of the studentised range distribution with parameter k and N – k degrees of freedom. k is the number of categories and N is the total number of observations.

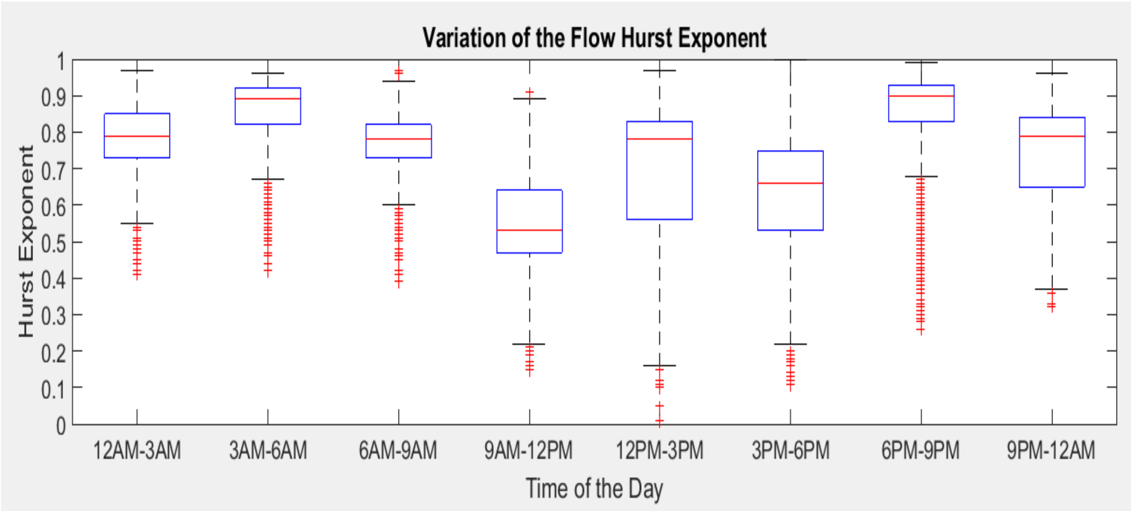


Figure 4-7 Variation of H_{flow} with the time of the day (box plot).

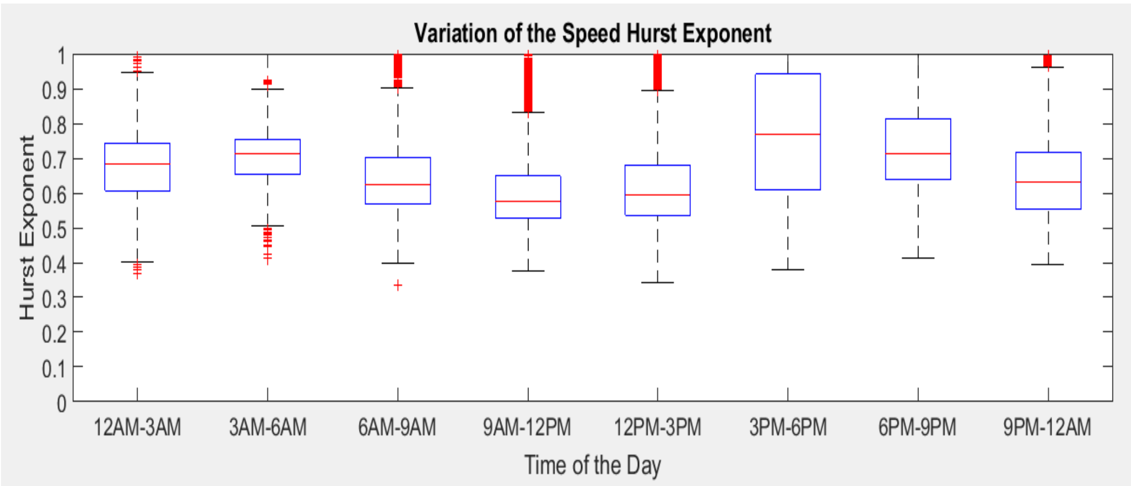


Figure 4-8 Variation of H_{mac_speed} with the time of the day (box plot).

4.3.3 The Effect of Number of Lanes on the Hurst Exponent

Since the M4 motorway has a different number of lanes at different sites, it is reasonable to analyse the effect of the number of lanes on the Hurst exponent. The mean H_{flow} for 2-lane sections is only 0.64, whereas it is 0.75 and 0.76 for 3-lane and 4-lane sections respectively. All the two-lane sections are located either near the entry or the exit of the motorway, where the presence of traffic signals result in intermittent traffic flow, leading to low H values. Further, using ANOVA and Tukey's HSD tests, we observed that H_{flow} for lane-2 sections is statistically different from other sections (Table 4-1). On the other hand, the H_{mac_speed} analysis shows the means of 2-lane, 3-lane, and 4-lane sections as 0.67, 0.67 and 0.71, respectively. In this case, the 4-lane sections are observed to have statistically different mean values than the other two types of road sections. This implies that the 4-lane sections have a high LRD than other sections on an average and better speed predictability.

Table 4-1 ANOVA and Tukey's posthoc HSD test results for the effect of the number of lanes on the Hurst exponent

Test	**Label**	$\boldsymbol{H_{flow}}$	$\boldsymbol{H_{mac\_speed}}$
ANOVA: Single factor	Null hypothesis H_0:	$\mu_{2\text{-lane}} = \mu_{3\text{-lane}} = \mu_{4\text{-lane}}$	$\mu_{2\text{-lane}} = \mu_{3\text{-lane}} = \mu_{4\text{-lane}}$
	Level of significance, α	0.01	0.01
	F	527	40
	F-critical, F_{cr}	4.61	4.61
	Probability, p ($F < F_{cr}$)	< 0.01	< 0.01
	Decision	Reject null hypothesis	Reject null hypothesis
Pair-wise comparisons via Tukey's HSD test	What pairs are significantly different at 0.01 level of significance?	2-lane and 3-lane 2-lane and 4-lane	2-lane and 4-lane 3-lane and 4-lane

4.3.4 The Effect of Entry and Exit Ramps

The M4 motorway has 13 entry and 14 exit ramps, which could affect the traffic characteristics on the mainline. Therefore, we evaluated the differences between the monitor sites that are closer to the ramps and those away from the ramps. We identified the monitor sites on the mainline of the motorway that are nearby the ramps. Out of the 92 monitor sites, 13 are near the entry ramps, 14 are near the exit ramps, and the remaining 65 are away from the ramps. Nevertheless, these 65 monitor sites could still be affected by the ramps, particularly during accidents and congested conditions. The mean H_{flow} for monitor sites that are away from the ramps is 0.76, whereas it is 0.74 and 0.73 for sites closer to entry and exit ramps respectively. ANOVA and Tukey's HSD tests (Table 4-2) showed that H_{flow} for sections away from ramp and closer to exit ramp are statistically different.

Table 4-2 ANOVA and Tukey's posthoc HSD test results for the effect of ramps on the Hurst exponent

Test	Label	H_{flow}	H_{mac_speed}
ANOVA: Single factor	Null hypothesis H_0:	$\mu_{no\text{-}ramp} = \mu_{entry\text{-}ramp} = \mu_{exit\text{-}ramp}$	$\mu_{no\text{-}ramp} = \mu_{entry\text{-}ramp} = \mu_{exit\text{-}ramp}$
	Level of significance, α	0.01	0.01
	F	6.56	35.95
	F-critical, F_{cr}	4.61	4.61
	Probability, p ($F < F_{cr}$)	< 0.01	< 0.01
	Decision	Reject null hypothesis	Reject null hypothesis
Pair-wise comparisons via Tukey's HSD test	What pairs are significantly different at 0.01 level of significance?	No ramp and exit ramp	No ramp and exit ramp No ramp and entry ramp

The H_{mac_speed} analysis shows the means of sites away from the ramp, closer to entry ramp, and closer to exit ramp sites as 0.67, 0.70 and 0.70, respectively. The sites away from the ramps have a statistically different and lower mean than the sites closer to the ramps. There will be increased turbulence at ramp locations, due to entering/exiting vehicles and frequent lane changes. These phenomena create traffic bottlenecks/congestion, and therefore high LRD of speed.

4.3.5 *The Effect of Horizontal Curve*

Some of the monitor sites (21 out of 89) are located on curved sections of the motorway. The mean H_{flow} at monitor sites on the straight and curved sections is 0.73 and 0.75 respectively. ANOVA test showed that the means are statistically different (F=22.76; F_{cr} = 6.64; p[F>F_{cr} = <.01]). Furthermore, the mean H_{mac_speed} on the straight and curved sections is 0.67 and 0.70 respectively. ANOVA test showed that even the mean H_{mac_speed} values are different for both the cases (F=39.42; F_{cr} = 6.64; p[F>Fcr = <.01]). Most drivers drive cautiously on curved sections of motorways. Still, few drivers attempt to navigate curves at high speeds, which could create turbulence in the downstream and frequent slowing down of vehicles, which could result in congestion and therefore, the strong trend-reinforcing property of speed.

4.3.6 *Summary Statistics*

Finally, Table 4-3 shows the summary statistics of all the considered variables. It shows the mean and standard deviation values of the Hurst exponent and Predictability Index for the variables[25]. Typically, flow is better predictable than speed which is intuitive because speed depends on more factors than traffic flow.

[25] The analysis of aggregated or average values of flow and speed and their relationship with accidents are well documented in earlier studies. The focus of the chapter (and the thesis in general) is to show the magnitude of fluctuations in traffic data and the reasons behind high/low fluctuations and their implications on predictability and crashes. The analysis of aggregated or average values of flow and speed and their relationships with fluctuations and crashes would deviate from the basic premise of the thesis and therefore the analysis is not within the scope of the study. Nevertheless, a table to summarise descriptive statistics for flow, flow per lane and speed is provided in Appendix C of the revised thesis. Graphs depicting the profile of flow and speed over a 24-hour period, and the PDF plots of H_{flow} and H_{mac_speed} are also shown in Appendix C.

Nevertheless, the mean *PI* is mostly less than 0.7, which indicates that simple linear prediction methods may not suffice, and sophisticated techniques should be sought.

Table 4-3 Summary statistics of all the considered variables

Variable	No. of valid time windows	Mean H_{flow}	Mean PI_{flow}	Std. Dev. H_{flow}	Mean H_{mac_speed}	Mean PI_{speed}	Std. Dev. H_{mac_speed}
Weekend v/s Weekday							
Weekend	5,579	0.68	0.36	0.16	0.64	0.28	0.11
Weekday	13,681	0.76	0.52	0.15	0.69	0.38	0.14
Time of the day							
12AM-3AM	2,399	0.79	0.58	0.09	0.68	0.36	0.10
3AM-6AM	2,399	0.85	0.70	0.10	0.70	0.40	0.08
6AM-9AM	2,398	0.77	0.54	0.07	0.65	0.30	0.13
9AM-12PM	2,435	0.55	0.10	0.12	0.62	0.24	0.13
12PM-3PM	2,435	0.70	0.40	0.17	0.62	0.24	0.12
3PM-6PM	2,398	0.65	0.30	0.15	0.77	0.54	0.17
6PM-9PM	2,398	0.87	0.74	0.11	0.73	0.46	0.13
9PM-12AM	2,398	0.75	0.50	0.12	0.64	0.28	0.11
Number of lanes							
2-lanes	2,488	0.64	0.28	0.18	0.67	0.34	0.14
3-lanes	15,669	0.75	0.50	0.14	0.67	0.34	0.13
4-lanes	1,103	0.76	0.52	0.16	0.71	0.42	0.14
Ramps							
Away from ramp	13,424	0.76	0.52	0.16	0.67	0.34	0.13
Closer to entry ramp	2,730	0.74	0.48	0.15	0.69	0.38	0.14
Closer to exit ramp	3,106	0.73	0.46	0.15	0.69	0.38	0.14
Horizontal curve							
On straight section	14,967	0.73	0.46	0.16	0.67	0.34	0.13
On curved section	4,293	0.75	0.50	0.15	0.70	0.40	0.14

4.4 SUMMARY

This chapter analysed traffic data obtained from several monitor sites on the M4 Motorway (outbound direction from the city) in Sydney. We used a 3-hour time window to evaluate the Hurst exponents of flow and speed across all the monitor sites over one month. We plotted heat maps to visualise the spatial and temporal variations of the Hurst exponent. With the empirical analysis, we demonstrated that the Hurst exponent is predominantly time-dependent. Furthermore, we found that the predictability of flow and speed is significantly lower on weekends than the weekdays. This could be due to lower (and intermittent) traffic volumes/flows on weekends. Also, a greater number of work-related trips occur on the weekdays, which are better predictable than the leisure trips that mostly occur on weekends.

Using ANOVA and Tukey's posthoc HSD tests, the monitor sites located on 2-lane sections were found to have a low average H_{flow} and a low "flow" predictability compared to other sections. On the other hand, 4-lane sections exhibited a significantly high "speed" predictability. This analysis of the effect of the number of lanes on the Hurst exponent would provide useful insights into traffic flow and speed prediction algorithms. Finally, we evaluated the effect of the proximity to an entry or exit ramp and the presence of a horizontal curve on fluctuations in speed and flow.

Nevertheless, several other factors could also affect fluctuations in macroscopic traffic variables such as weather, the degree of curvature, the proportion of heavy vehicles, driver heterogeneity, the age of vehicles, the location of study area, etc. These aspects should be considered in future studies. Furthermore, only one month of traffic data is considered for the analysis. Seasonal variations could have an influence on fluctuations, which could be evaluated by collecting and analysing additional data.

The main limitation of this chapter is the usage of a fixed time window of 3-hours. This assumption helped in calculating and visualising long-range-dependence at an aggregated level, yet it may not efficiently depict traffic operations. Considering a *"moving time window"* (by incrementally shifting the three-hour interval) will help in overcoming this limitation. While fixed time windows are helpful in perceiving traffic trends at an aggregated level, moving windows will be useful for operational purposes. For example, by spotting the differences between Hurst exponents of time series from 12 AM to 3 AM and 12:05 AM to 3:05 AM, one can infer the dynamics of the 5-minute period between 3 AM and 3:05 AM. The consideration of moving time window would also enable one to study the propagation of long-term dependence or shockwaves at a macroscopic level.

An exciting extension of this study is to use cross-correlation technique to evaluate how the fluctuations travel upstream. Another interesting study could be lane-wise fluctuation analysis to find whether fluctuations are correlated (also called synchronised) and observe how fluctuations propagate laterally.

CHAPTER 5. APPLICATION OF THE FRACTAL THEORY IN CRASH RATE MODELLING

5.1 INTRODUCTION

Reducing congestion and crashes are the primary goals of any transport agency. These two issues are interdependent as significant improvements to one could result in substantial impacts on the other (Chang and Xiang 2003). Furthermore, relieving congestion could also reduce travel delays, such that air quality, fuel economy, and economic productivity can be improved. The association between congestion and road safety is an issue debated by transport planners and safety experts. Some studies argue that the increased level of traffic congestion reduces traffic speeds and therefore leads to less severe crashes. While that may be true, there will also be an increase in the likelihood of exposure and the number of potential conflicts in the congested conditions. This may lead to more crashes, although of a less severe nature (Quddus et al. 2009).

Existing studies explore the applications of different statistical models in estimating crash frequencies, crash rates, and injury severities at specific locations. For example, count data models such as Poisson and Negative Binomial models were extensively used to estimate crash frequencies (Shankar et al. 1995; Parida et al. 2006; Bhat et al. 2014; Dong et al. 2017); Tobit model was used to study the accident rates (Anastasopoulos et al. 2008); the Mixed Logit (Moore et al. 2011), the Ordered Logit and the Ordered Probit models (O'Donnell and Connor 1996; Kockelman and Kweon 2002) were used to predict injury severity. These studies considered road geometric variables (ramps, shoulder width, and gradient), pavement characteristics (roughness, rutting, and friction), environmental factors (rainfall, crosswind speed, and snow) and traffic variables (ADT or AADT, mean speed, posted speed limit and the heavy vehicle proportionality).

Many studies accounted for traffic congestion by considering proxy variables. For example, the volume-capacity ratio (V/C) (Shefer 1994), employment density (Noland and Quddus 2005), AADT (Kononov et al. 2008) were some of the proxies considered. However, modelling the aggregated crashes at a road segment level with such proxies for congestion may obscure the actual relationships (Quddus et al. 2009). In this regard, some researchers have used disaggregated crash records and a measure of traffic congestion during the crash period (Zheng et al. 2010; Ahmed et al. 2012; Yeo et al. 2013). Nevertheless, studies of this category need extensive traffic and crash data, which can be difficult to obtain and computationally cumbersome.

Speed is one of the critical factors in crashes. Mean speed has been found to have mixed effects on aggregated crashes in the literature (Wang et al. 2013). While some studies found that increased speed reduces safety (Taylor et al. 2002; Elvik et al. 2004; Nilsson 2004), other studies found the opposite (Garber and Gadiraju 1989). Further, there are some studies of the view that speed variation is likely to be a more critical determinant of crashes than speed itself (Garber and Gadiraju 1989). Again, the results are inconsistent which could be due to the lack of high-resolution data (Theofilatos and Yannis 2014).

As mentioned earlier in Chapter 1, the Hurst exponent, i.e. the long-range dependence metric is a numerical representation of the randomness through the history of a dynamical process, while the coefficient of variation is independent of the temporal evolution. Therefore, we use the Hurst Exponent of speed (based on one-month traffic data) as one of the explanatory variables in modelling four-year historical crash rates. The hypothesis is that typical congestion levels at a site could be assessed for a month and inferences can be made about historical crash rates. In this way, this chapter presents a reconciliation effort between the aggregated and disaggregated crash modelling approaches. We estimated random parameters and latent class Tobit models to account for unobserved heterogeneity. The following sections describe the procedure in detail.

5.2 METHODOLOGY

5.2.1 Tobit Model

There has been an enormous emphasis in past research on factors that determine the frequency of crashes (Anastasopoulos et al. 2008). Using exposure-based crash rates (continuous variable) instead of traditional crash frequencies (count variable) as the dependent variable has significant appeal because crash rates are typically used in crash reporting, i.e., the number of crashes per 100 million vehicles. The use of crashes per vehicles or crashes per vehicle kilometres has an intuitive appeal in highway safety—providing a standardised measure of relative safety on roadway segments, which is more easily interpreted than the number of crashes per time period (Anastasopoulos et al. 2008).

Crash rates that are assessed over a finite period along highway sections can result in numerous segments having either zero or minimal crashes reported during the analysis period. This results in a clump of zeros (20 out of 172 within this study) in the dependent variable, which generates biased and inconsistent parameter estimates of modelled crash rates, by standard ordinary least squares

(Anastasopoulos et al. 2008; Washington et al. 2010). Tobit model is typically applied in these circumstances, where crash rate is considered as a censored dependent variable (Tobin 1958). A few extensions or variations of the Tobit model have been used in the literature. For example, multivariate Tobit (Anastasopoulos et al. 2012b), random parameters Tobit (Anastasopoulos et al. 2012a; Anastasopoulos 2016; Park and Lee 2017), Bayesian spatial random parameters Tobit (Zeng et al. 2017) and random effects Tobit models (Chen et al. 2014) were used to analyse both crash rates and injury severities.

In a Tobit model, there is a latent variable Y_i^* which is observable only above a given threshold (zero in this study). The latent variable Y_i^* can be formulated as:

$$Y_i^* = \beta X_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \tag{5.1}$$

The model assumes that the latent variable, Y_i^* is linearly dependent on the independent vector of variables X_i via the vector of parameters β. Also, there is a normally distributed error term ε_i, with zero mean and constant variance σ^2 across observations, i. Finally, the observable variable Y_i is defined as:

$$Y_i = \begin{cases} Y_i^*, & if\ Y_i^* > 0 \\ 0, & if\ Y_i^* \leq 0 \end{cases} \tag{5.2}$$

The likelihood function for Tobit model over N number of observations is as follows:

$$L = \prod_{i=1}^{N} \left[1 - \Phi(\frac{\beta X_i}{\sigma})\right]^{1-I(Y_i)} \left[\emptyset(\frac{Y_i - \beta X_i}{\sigma})\right]^{I(Y_i)} \tag{5.3}$$

where

$\Phi(.)$ = the cumulative density function,

$\emptyset(.)$ = the probability density function,

σ = the standard deviation of the error term, and

$I(.)$ = the indicator function which takes value zero if Y_i = 0 and 1 otherwise.

The expected value of the dependent observed variable Y_i is:

$$E(Y_i|X_i) = \beta X_i \Phi(z) + \sigma \emptyset(z) \tag{5.4}$$

where

$z = \beta X/\sigma$,

Φ(z) = the probability of Y_i being greater than zero,

$\emptyset$(z) = the standard normal probability density function, and

σ = the standard deviation of the error term.

5.2.2 Unobserved Heterogeneity

Crashes are complicated and less predictable events, making it next to impossible to collate all the data that could determine the likelihood of a roadway crash. The absence of relevant explanatory variables can have serious model specification problems for traditional statistical analyses, which can potentially lead to biased and inconsistent parameter estimates, erroneous inferences and inaccurate accident predictions (Mannering et al. 2016). This particular problem is referred to as 'unobserved heterogeneity'.

Researchers typically have adopted statistical approaches such as random parameters models (Milton et al. 2008; Anastasopoulos and Mannering 2009; Anastasopoulos et al. 2012a; Anastasopoulos 2016; Venkataraman et al. 2013; Pande et al. 2017), latent class (finite mixture) models (Park and Lord 2009; Eluru et al. 2012; Zou et al. 2013; Shaheed and Gkritza 2014; Yasmin et al. 2014), a combination of latent class and random parameters models (Xiong and Mannering 2013; Buddhavarapu et al. 2016), Markov-switching models (Malyshkina et al. 2009; Malyshkina and Mannering 2009, 2010), and Markov-switching models with random parameters/latent classes (Xiong et al. 2014) to address the problem. The last three approaches are complex processes and computationally cumbersome, as noted by (Mannering et al. 2016). This chapter estimates the random parameters and latent class Tobit models to account for unobserved heterogeneity in crash rate prediction. A brief overview of both the approaches is presented in the following sections.

5.2.3 Random Parameters Model

In random parameter (RP) models, estimated parameters can vary across observations according to an analyst-specified distribution, typically normal, log-normal, triangular ora uniform, to account for the heterogeneity.

The estimable parameter can be viewed as:

$$\beta_i = \beta + \emptyset_i,\ i = 1,2,\dots N \qquad (5.5)$$

where $\emptyset_i$ is a randomly distributed term with mean μ and variance σ^2. The log-likelihood function (LL) can be written as follows:

$$LL = \sum_{i=1}^{N} ln \int_{\phi} g(\phi) P({Y_i}^{*}|\phi) d\phi \qquad (5.6)$$

where g (.) is the probability density function of the randomly distributed term, ϕ. The maximum likelihood estimation (MLE) procedure of random parameter models is computationally cumbersome (Anastasopoulos et al. 2012a). Therefore, a simulation-based ML method is typically employed using Halton draws, which has been shown to provide an efficient distribution of draws for numerical integration (Halton 1960; Train 2000; Bhat 2001).

The RP formulation presented above assumes the same random parameter mean for all the individuals (highway segments). This approach was widely used in many earlier studies (Anastasopoulos and Mannering 2009; Venkataraman et al. 2013; Pande et al. 2017). There are a few extensions, such as RP models with correlated parameters and RP models with means (and variances) as functions of explanatory variable that were lately applied in accident modelling (Kim et al. 2013; Venkataraman et al. 2014; Mannering et al. 2016; Behnood and Mannering 2017; Seraneeprakarn et al. 2017). These approaches offer a more general formulation of unobserved heterogeneity. However, these approaches are not within the scope of the present study.

5.2.4 Latent Class Model

A latent class (LC) model is similar to an RP Model, in the sense that it also accounts for unobserved heterogeneity. LC models have been applied in past studies on crash models that focussed on injury severity (Yasmin et al. 2014; Sasidharan et al. 2015). In LC models, unobserved heterogeneity is accounted by classifying observations into a set of Q different latent classes, with similar variable effects within each group. LC models consist of two components: a class membership model and a class-specific regression model (Vij et al. 2013). In the case of crash analysis, the class membership model formulates the probability that a highway section belongs to a specific latent segment, as some function of the section's characteristics. Conditional on the class that a section belongs to, the class-specific Tobit regression model evaluates the effect of different explanatory variables on the crash rate. The main advantage of an LC model is that it does not impose a distributional assumption on the parameter, thus making it semi-parametric (Greene and Hensher 2003; Mannering et al. 2016). The estimated parameters are assumed to have identical effects across elements belonging to that segment, but different effects between classes. The class membership probability for a location *i* to belong to class *q* follows a logit model:

$$P_{iq} = \frac{\exp(\theta_q z_i)}{\sum_{q=1}^{Q} \exp(\theta_q z_i)}, \quad q = 1,2, \dots Q \tag{5.7}$$

where

θ_q = a corresponding vector of estimable parameters for the latent segment *q*, and

z_i = a set of attributes (covariates) that are associated with class membership.

The unconditional probability for a location *i* is the expectation (over all classes) of the class-specific contributions:

$$P_i = \sum_{q=1}^{Q} P_{iq} P_{i|q} \tag{5.8}$$

where $P_{i/q}$ is the expected number of crashes at location *i*, conditional on latent class *q*. Finally, the log likelihood is the summation of the logs of all the probabilities:

$$LL = \sum_{i=1}^{N} \ln P_i \tag{5.9}$$

The class membership model is rich regarding interpretation and provides potential applications for policymakers (Vij et al. 2013) in implementing suitable crash countermeasures for a targeted segment. Without covariates in a class membership model, one can only infer that there are few distinct classes in the observed data and different effects of variables on each of the classes. However, suggesting crash countermeasures for these classes may be challenging because the class that an observation belongs to will not be known to the analyst.

Both RP and LC models have their drawbacks. RP models need distributional assumptions and are not flexible enough to identify groups of highway segments with shared unobserved heterogeneity. On the other hand, LC models are not flexible to account for unobserved heterogeneity within the identified latent classes (Mannering et al. 2016). Therefore, some researchers proposed latent class models with random parameters within classes (Xiong and Mannering 2013; Buddhavarapu et al. 2016). Nevertheless, such models are computationally expensive and also can severely complicate the estimation routine (Mannering et al. 2016). This hybrid approach is not within the scope of this study.

Furthermore, it should be noted that the results of both the LC and RP models are highly data-specific and therefore, one particular method may not always be necessarily superior to the other (Mannering et al. 2016). This is because unobserved heterogeneity can follow different distribution in different datasets and sometimes LC models may appear to fit data better, and sometimes the RP models may seem to fit better.

5.3 STUDY AREA AND DATA DESCRIPTION

The data used in this chapter were obtained from 184 monitor sites in the eastbound and westbound directions of the M4 motorway during February 2013. The westbound is away from the Sydney CBD, and the eastbound is towards the CBD. Weighted average speed and the total flow across all lanes were calculated. 1-month traffic observations were then discretised into 3-hour, non-overlapping windows, starting from 12 AM on 1st of February 2013. There are 224 windows

(28 days * 8 windows/day) of observations at each monitor site. The data cleaning exercise resulted in the omission of 12 monitor sites from the analysis. Figure 5-1 shows the frequency distribution plots and the summary statistics of the Hurst exponent of speed (H_{mac_speed}) in the westbound and eastbound directions. Although H can theoretically take values from 0 to 1, the smallest value observed was 0.31.

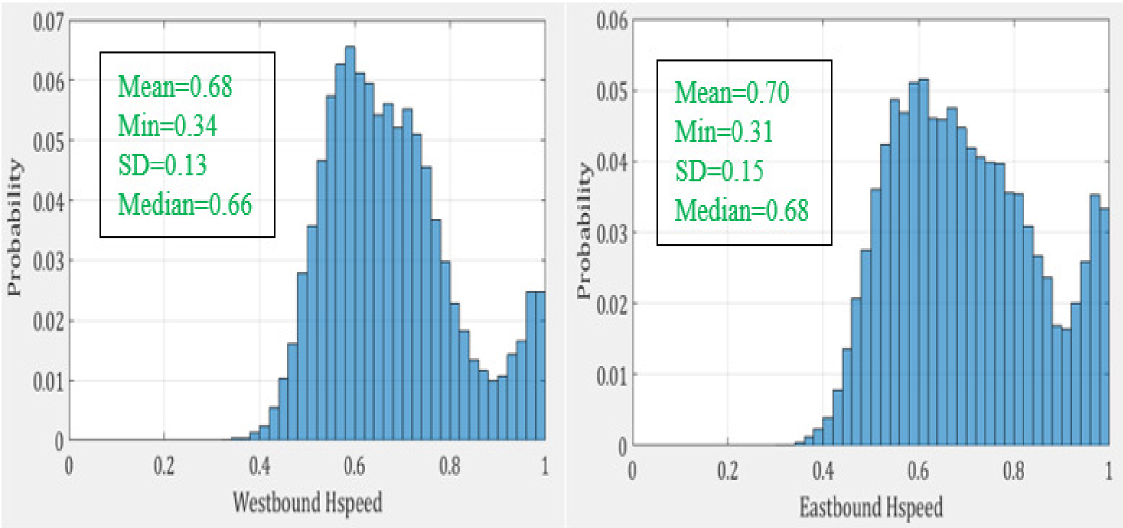


Figure 5-1 Frequency distributions of the westbound and eastbound H_{mac_speed}

Long-term crash counts/rates are a commonly accepted measure to classify safe and unsafe road sections (Hauer 1996). For this study, four-year historical crash data between 2012 and 2015 were obtained. The dataset included information on time, location, duration and crash type. Any crash occurring on the carriageway was attributed to the two nearest monitor sites (one upstream and one downstream). For example, if a crash happened at a 1/4th distance from site A and 3/4th distance from B, then the 3/4th proportion of the crash was attributed to site A and 1/4th to site B. The total number of crashes at each site was then calculated, by summing all these proportions and rounding to the nearest integer. Table 5-1 provides descriptive statistics of all variables considered for the model estimation. Entry and exit ramps were assigned to the nearest monitor sites. As seen in Table 5-1, 17% of the monitor sites are located near an entry ramp, and 16% near an exit ramp.

Table 5-1 Descriptive statistics of all the considered variables

Variable	Variable Type	% of observations (categorical/ binary variables)	For continuous variables			
			Mean	SD	Min	Max
Crash rate (number of crashes per 100 million vehicles)	Continuous	---	3.04	3.56	0.00	18.92
Direction of travel	Binary		---	---	---	---
0 --- Outbound (Westbound)		51.75				
1 --- Citybound (Eastbound)		48.25				
Horizontal curve	Binary		---	---	---	---
0 -- No curve		76.75				
1 -- Presence of curve		23.25				
Ramp	Categorical		---	---	---	---
0 -- No ramp		67.44				
1 -- Entry ramp		16.86				
2 -- Exit ramp		15.70				
ADT per lane (in 10,000)	Continuous	---	1.69	0.48	0.65	2.80
% heavy vehicles	Continuous	---	4.97	1.38	0.86	8.11
H0.85[26] (in percentage of 10)	Continuous	---	1.50	1.07	0.28	5.41

[26] Description of H0.85 is in section 5.4.1.

5.4 DATA ANALYSIS

5.4.1 Usability of the Hurst Exponent for Crash Modelling

As noted in the data section, we estimated H_{mac_speed} values for every 3-hour window at each monitor site. Then we obtained the percentage frequency distributions obtained across all 3-hour periods at each location. Next, a set of 13 binary variables based on threshold value "*H*" were defined as follows:

High_Hurst_*H*=1; If the Hurst exponent is greater than "*H*"

High_Hurst_*H*=0; Otherwise.

The threshold value *H* varied from 0.35 to 0.95 in 0.05 increments. This changing value of *H* produced 13 different binary variables defined as "High_Hurst_0.35" through to "High_Hurst_0.95". This classification is similar to the "jerk" (rate of change of acceleration) classification approach used in Pande et al. (2017).

Figure 5-2 shows the Pearson's correlation coefficient between crash rates and percentage of observations with High_Hurst_*H*=1, based on the 13 defined thresholds. The correlation coefficient rises, as the threshold value *H* increases from 0.35 to 0.95. The rise is noteworthy for the three highest threshold values, i.e. 0.85, 0.90 and 0.95. Thus, the sites with a higher proportion of high-value *H* (a potential indicator of congestion as noted in Chapter-4) are correlated strongly with long-term crashes. Therefore, the crash rate modelling in this study is based on the threshold, *H0.85*, i.e. the percentage of observations with *H* > 0.85.

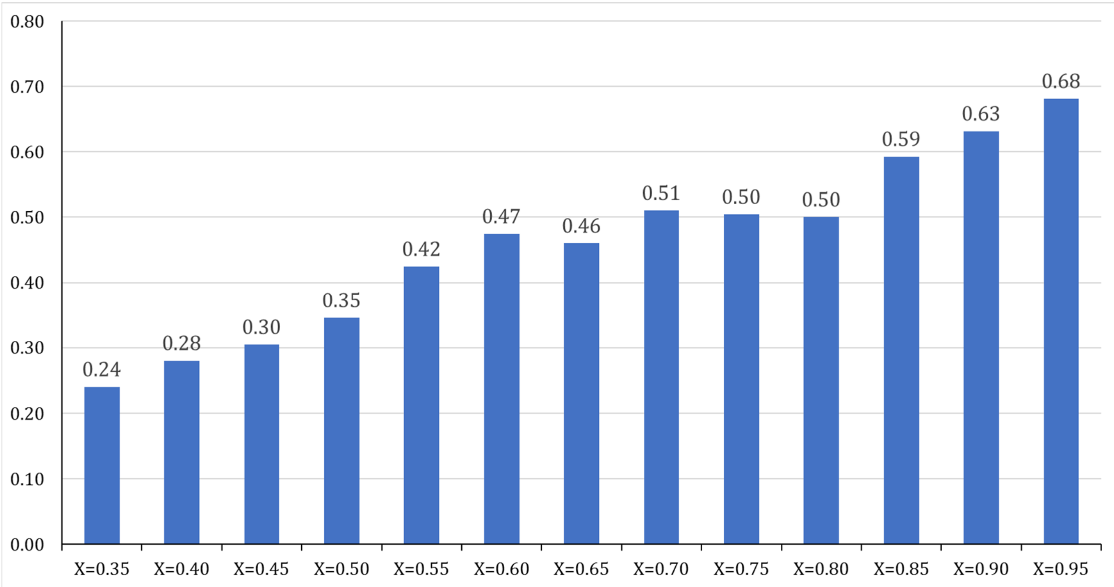


Figure 5-2 Correlation coefficient between crash rates and percentage of observations with High_Hurst_*H*=1 (based on 13 threshold H values)

Next, we estimated a fixed parameters Tobit model with fundamental variables (ADT per lane, curve, ramp, bound, the proportion of heavy vehicles). Then we included the *H0.85* (the percentage of observations with H > 0.85) variable in the set of explanatory variables and re-estimated the model. The latter model specification had a significantly better fit in terms of the Akaike Information Criteria (AIC) (717 compared to 740 in the first model), which signifies the importance of capturing traffic congestion in crash models. However, we estimate crash rates using advanced statistical approaches to account for unobserved heterogeneity and improved model fit. The following section describes the modelling approach.

5.4.2 Crash Rate Modelling

The next step in the analysis is to predict long-term crash rates using random parameters, and latent class approaches at the 172 monitor sites under consideration. We utilised a simulation-based ML with 500 Halton draws to estimate the random parameters Tobit (RPT) model. In deciding the best model specification for the sample, we estimated several models by changing the mixing distributions (normal, log-normal, uniform, and triangular). The variables that are significant at p = 0.10 were considered. From all the RPT model specifications explored, the best model in terms of goodness-of-fit outputted the following: *H0.85* variable with a normally distributed random parameter; entry ramp and exit ramp with a normal distribution; remaining variables with fixed parameters. Table 5-2 shows the key results of the model estimation.

For the latent class model, we estimated different model specifications by: i) changing the number of latent classes and ii) including different explanatory variables (covariates) in the class membership model. Covariates are helpful to assess the probability that an observation belongs to an exclusive segment. Specific variables are only included in the class membership model if they contribute significantly to the segmentation of the motorway sections, i.e. only if the estimated covariate parameters are significant at p = 0.10. Similarly, within each class-specific Tobit model, the variables that are significant at p = 0.10 for all classes, were considered. Out of all the evaluated LC specifications, the one with two latent segments and the presence of entry and exit ramps as covariates in the class membership model was found to give the best fit. Table 5-2 presents the estimation results along with the goodness of fit measures for the final selected RPT and LCT models. The class membership part of the LCT model is also presented in Table 5-2.

Table 5-2 Estimation results of RPT and LCT models

	Random Parameters Tobit (RPT) Model	Latent Class Tobit (LCT) Model	
		Class-1	Class-2
Constant	1.44 (2.31)	14.41 (4.25)	-0.49 (0.39)
H0.85 (in percentage of 10)	1.89 (4.45)	4.26 (5.90)	0.84 (3.56)
Std. dev. of H0.85	0.61 (3.31)	---	---
ADT per lane (in 10,000)	-1.57 (-3.59)	-6.20 (-4.23)	0.81 (0.16)
Direction of travel	-1.07 (-3.86)	-2.86 (-2.98)	-0.42 (-1.26)
Proportion of heavy vehicles	---	-0.87 (-1.92)	0.09 (0.63)
Presence of horizontal curve	0.56 (1.96)	---	---
Presence of entry ramp	3.21 (5.28)	Included in the class membership model	
Std. dev. of entry ramp	2.66 (4.34)	---	---
Presence of exit ramp	4.59 (4.48)	Included in the class membership model	
Std. dev. of exit ramp	3.40 (3.93)	---	---
Number of observations	172	172	
LL	-329	-313	
AIC	680	656	
BIC	714	703	
		Class Membership component	
		Class-1	**Class-2**
Constant	NA	-2.60 (-3.54)	*Fixed*
Presence of entry ramp	NA	4.08 (4.09)	*Fixed*
Presence of exit ramp	NA	4.85 (3.59)	*Fixed*

t-statistic in parenthesis

As noted earlier, the results are data-specific as heterogeneity patterns are likely to vary from one dataset to another. Therefore, the inferences made in this study are data-specific as well. In the analysed dataset, based on AIC and BIC, the LCT model offered a better fit than the RPT model. Upon inspecting the estimation results in Table 5-2, one may observe that the signs of the significant parameters are consistent across the RPT and LCT model frameworks, despite having different magnitudes (due to the underlying parametric distribution assumptions). The results presented in Table 5-2 for the class membership model in conjunction with the class-specific Tobit model, offer valuable insights into how classes differ, with respect to location types and the varying effect of explanatory variables on each class. The following section compares the model estimation results of both frameworks.

5.4.2.1 H0.85 of Speed:

The RPT estimation results in Table 5-2 show that *H0.85* variable has a normally distributed random parameter with a mean of 1.89 and standard deviation 0.61. Given this distribution, one may calculate that 99.9% of the locations show an increasing rate of crashes with an increase in congestion. Therefore, according to the RPT model, increased levels of congestion cause higher crash rates at almost all the locations in the analysed dataset. Even in the case of LCT model, *H0.85* has a similar effect on the crash rate in both the classes. However, the effect of congestion in class-1 is more profound than class-2, showing the congestion severity in this class. The results from the RPT and LCT models concerning the effects of congestion are consistent, demonstrating the significance of capturing congestion in crash models.

5.4.2.2 ADT per lane:

The relationship between ADT (or ADT per lane) and crash rate in the past studies is slightly ambiguous. Some studies reported a negative relationship between crash rate and ADT (Anastasopoulos et al. 2012a) and others reported it as having an inverse U-shaped relation (Hall and Pendleton 1989; Garber and Subramanyan 2001; Qi et al. 2007). The inverse U-shape suggests that at low values of ADT, the crash rate increases and then stabilises/peaks, and finally decreases as ADT increases further. These findings should not be confused with the relation between the "number" of crashes and ADT per lane, which is typically positive and sometimes nonlinear (Abdel-Aty and Radwan 2000). In the current study, ADT per lane has shown a negative relationship with the crash rate in class-1 of the LCT model. However, in class-2, this variable is not significant, albeit with a positive sign. It is plausible that the locations in class-1 have a high ADT per lane and are stacked towards the right side of the inverse U-curve and therefore, crash rates decrease with increasing ADT per lane. Locations associated with class-2 probably have lower ADT per lane than that of class-1 and clustered near the peak (slightly towards left) of the inverse U-curve, and so, are insignificant. On the other hand, it is likely that the RPT model considers most of the locations to have ADT per lane on the right side of the curve. It is emphasised once again that all these findings are strictly data-specific.

5.4.2.3 Direction of travel:

The citybound locations were observed to show a negative relation with crash rates in both the RPT and class-1 of LCT frameworks. Similar to ADT per lane, the "direction of travel" variable is found to be insignificant in class-2. The citybound locations capture drivers predominantly driving toward their workplaces. To contextualise, concentration levels are expected to be higher at these times. Conversely, outbound locations will predominantly capture drivers travelling away from their workplaces and present fatiguing or lowered driver concentration levels. Thus, resulting in higher crash rates in the outbound direction.

5.4.2.4 Proportion of heavy vehicles:

The proportion of heavy vehicles in the traffic stream is found to be insignificant in the RPT model. On the other hand, the variable has an adverse effect in class-1 and positive (but not significant) effect in class-2 of the LCT model. It is likely that the negative and positive effects are balanced out, resulting in the non-significance of the variable in the RPT model. It was discussed in section 5.4.2.2 that the class-2 locations are likely to have lower ADT per lane than class-1. Therefore, it is also likely that there are more heavy vehicles in class-1, which could heighten the concentration of light vehicle drivers (Anastasopoulos et al. 2008, 2012a). However, another study reported weak evidence of a negative relationship between crash rate and the proportion of heavy vehicles (Aljanahi et al. 1999). The significance of the variable in class-1 and non-significance in class-2 indicates that the effect of the proportion of heavy vehicles on crash rate can be highly data and site-specific, and heterogeneity considerations are to be made while modelling.

5.4.2.5 Presence of a horizontal curve:

The presence of horizontal curves may have heterogeneous effects across highway locations due to unobserved time-varying traffic and weather conditions as well as heterogeneous reactions of drivers to these curves (Mannering et al. 2016). Occasionally, drivers will attempt to navigate curves at high speeds, resulting in the tires skidding. This increases the likelihood of a crash occurence. However, there is an alternate hypothesis that sharp curves make motorists to drive cautiously. The RPT model predicts a positive relation between crash rate and the presence of a horizontal curve. Conversely, the LCT model identifies the variable as not significant in the current dataset. This discrepancy could be attributed to the different distribution assumptions and heterogeneity considerations of both the approaches. Further, the characteristics of a curve, such as the radius, length, etc. should be captured to precisely estimate the effect of horizontal curves on crash rate.

5.4.2.6 Presence of entry and exit ramps:

The RPT estimation results show that the presence of entry and exit ramps have normally distributed parameters. Given the distributions presented in Table 5-2, it can be inferred that most of the locations (88.7%) show a positive relationship with crash rate when there is an entry ramp. Similarly, the presence of exit ramps also shows a positive correlation with crash rates at around 91% of the locations. The comparison between the RPT and the LCT models concerning the ramp variable is not simple because these variables are included in the class membership component of the LCT model. While the ramp variables have a direct effect on the crash rate in the RPT model, their effect is indirect in the LCT specification. There will be increased turbulence at ramp locations due to entering/exiting vehicles and frequent lane changes. These phenomena create traffic bottlenecks/congestion and subsequently, higher crash rates than locations without ramps. The class membership component presented in Table 5-2, shows that the locations with either entry or exit ramps are more likely to belong to class-1 than class-2 (because of their positive signs).

5.4.3 A Note on the Latent Segments

As stipulated earlier, the class membership model formulates the probability that a highway section belongs to a particular latent class. In other words, for each location in the dataset, we calculated the probabilities to categorise which class a location belongs to, i.e., class-1 and class-2. Then, we summated the class membership probabilities for each location to generate the final proportions of each class. Around 1/3rd of the 172 locations are estimated to be in class-1 and the remaining 2/3rd in class-2. Following this, we assigned each location to a single class, based on the highest probability (Magidson and Vermunt 2004; Nagin 2005; de Oña et al. 2013). The average observed crash rates for class-1 and class-2 are calculated as 6.33 and 1.49 crashes per 100 million vehicles, respectively. Class-1 exhibits a higher average crash rate than class-2. Apart from congestion, there are significant contributions to crash rate from other explanatory variables in class-1. Conversely, the congestion indicator variable is significant in class-2, despite having a lower magnitude than class-1. This analysis signifies the importance of including congestion metrics in crash prediction models. Furthermore, the findings indicate that the segments belonging to class-1 should be a higher priority (potential hotspots) for transport authorities than those in class-2. More significant consideration should be given to ramp locations, due to their increased likelihood of contributing to a high priority segment. Congestion could be reduced at these bottlenecks by implementing traffic management countermeasures such as ramp metering algorithms, variable speed limits, and variable message signs, to adequately inform the driver before accessing the ramp.

5.5 SUMMARY

This chapter showed the application of the Hurst exponent from Fractal theory for the prediction of crash rates. We estimated the random parameters and latent class Tobit models to account for unobserved heterogeneity. The best RP model fitted for the dataset had the Hurst variable, the presence of entry and exit ramps with normally distributed parameters. On the other hand, the final LC model (the best-fit LCT among several explored specifications) suggested the presence of two classes, with the entry and exit ramps as covariates in the class membership model. Segmenting motorway locations into different classes is helpful for transport authorities to implement crash countermeasures.

From the explored literature, this is one of the first applications of LCT analysis for "crash rate" modelling. Other studies have used multivariate, random parameters and Bayesian spatial random parameters Tobit models to predict crash rates. While RP models account for unobserved heterogeneity and offer a better fit than a fixed parameters Tobit model, they are time-consuming as the analyst is required to make parametric assumption relating to the distribution of heterogeneity across observations. When considering LC models, choosing the number of classes and covariates in a class membership component is also cumbersome, but computationally minimal for each specification.

To summarise, this chapter presented two significant insights into: i) the application of the Fractal theory and ii) the estimation of Random parameters and Latent class Tobit models for crash rate modelling. Congestion indication, in terms of the Hurst exponent, was found to be significant in both the high and low priority segments, suggesting the importance of capturing congestion information in crash prediction models. The results of this analysis should be treated as exploratory and data-specific because of the different heterogeneity considerations of the RP and LC approaches. Subsequent investigation at a larger scale, and other explanatory variables, such as curvature, pavement quality, weather, traits of motorists, etc. could be explored to reinforce the findings. Despite this, the proposed methodology can easily be applied to newly constructed roads and in developing countries, with insufficient crash data and no standardised police reports. Congestion patterns in terms of the Hurst exponent can be assessed for a limited duration (say one month), for the potential identification of crash hotspots.

The consideration of H0.85 variable is only an engineering call and any of the other dynamic Hurst variables could be considered. In fact, H0.95 might offer even better model fit. The occurrences of LRD of speed are not as rare as accidents and that the Hurst exponent can be calculated easily from loop detector data. From this demonstration, results from this research may lead to the ability to proactively flag locations with higher crash potential.

CHAPTER 6. FLUCTUATIONS IN MACROSCOPIC TRAFFIC FLOW VARIABLES: THE APPLICATION OF SCATS INTERSECTION COUNT DATA

6.1 INTRODUCTION

Signalised intersections play a crucial role in improving the performance of urban road networks. Statistical data indicate that two-thirds of urban vehicle miles travelled in the U.S. are on signal-controlled roadways (FHWA 2012). Ensuring the efficient operation of traffic signals can minimise vehicle delays, maximise vehicle throughput, reduce fuel consumption, and improve air quality to serve the ever-rising travel demand. To achieve these goals, understanding the temporally varying patterns of demand and intersection vehicle throughput is essential. Intersection traffic counts have been traditionally used to estimate trip matrices (Maher 1983; Nihan and Davis 1989), optimising signal time (Li and Prevedouros 2004; Zhang and Wang 2011), modelling air quality (Gokhale and Raokhande 2008), capacity analysis (Chen et al. 2009), and modelling safety (Miaou and Lord 2003; Huang et al. 2017).

Most urban intersections show a predictable and smooth profile of traffic counts when measured at wider time-intervals (say, 5-min, 10-min, etc.). These count profiles typically consist of morning and evening peaks, mid-night off-peak and afternoon inter-peak as shown in Figure 6-1. These typical profiles show high prognostic structure, i.e. future values can be predicted from past values without much effort. However, the profiles can sometimes be irregular as shown in Figure 6-2, which could be due to abnormal weather conditions, intersection-specific geometry, random driver behaviour, crashes, or road construction activity in the vicinity. Commuters might experience different delays and congestion levels at the intersections than a typical day because of such conditions.

The regular occurrence of such abnormal days of traffic would result in commuters to often deviate from their usual preferred (or perceived optimum) routes, and the historical information will be insufficient in picking their best route. This reduces the reliability of travel time, and only real-time information could provide significant benefits in such cases (Rakha and Van Aerde 1995). Since reliability is a crucial factor for the performance assessment of highway segments and systems, transport agencies should be able to find those irregular profiles in advance. This identification further enables them in devising efficient traffic management strategies and proper information could be passed on to road users.

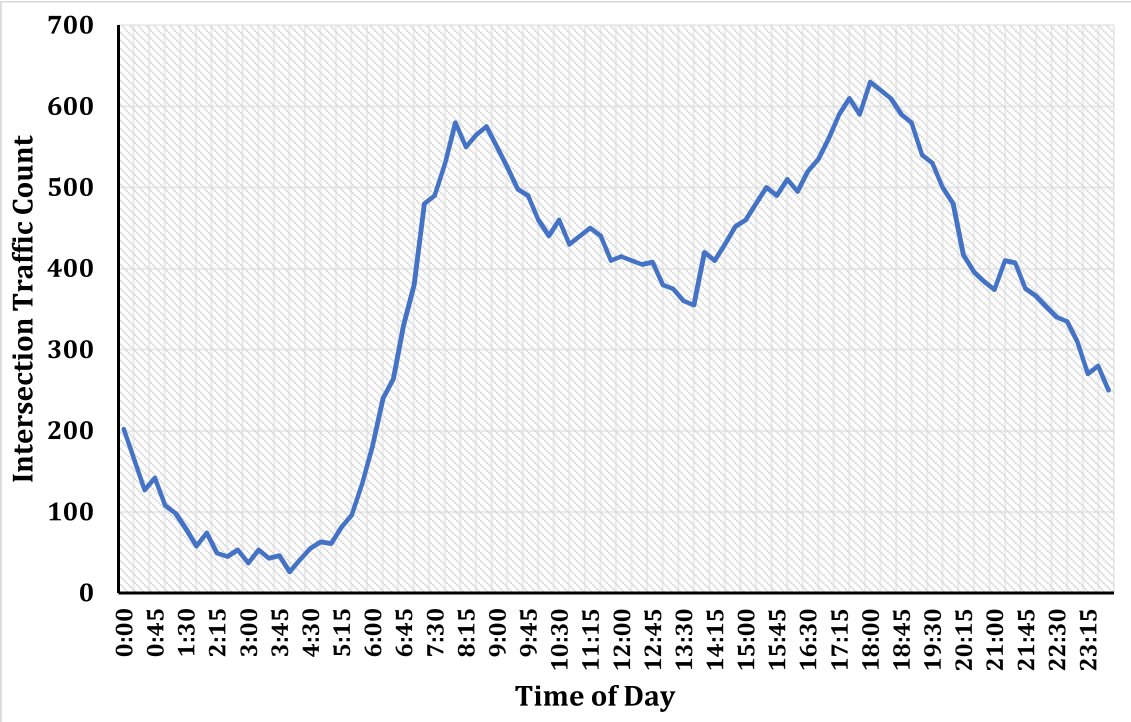


Figure 6-1 Typical intersection count profile (smooth changes)

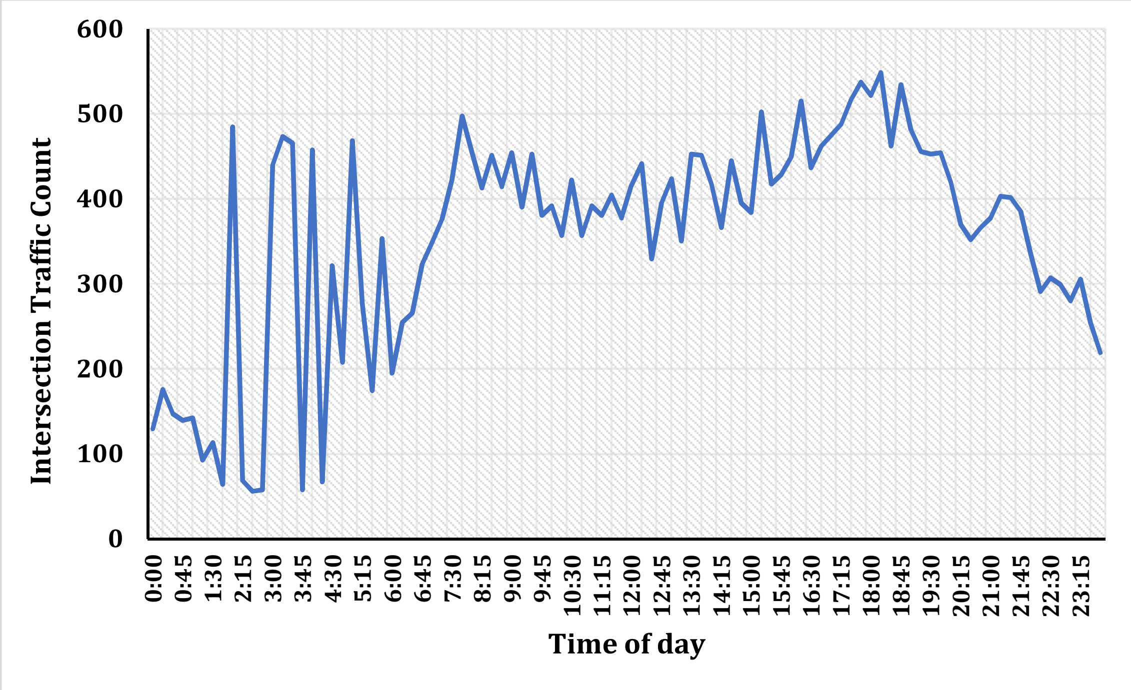


Figure 6-2 Atypical intersection count profile (severe fluctuations)

In this chapter, we use the Hurst Exponent method from the Fractal theory to evaluate complexity/predictability of traffic count data at several signalised intersections in the central business district (CBD) of Sydney, Australia. Nevertheless, merely evaluating the level of complexity of data does not offer intuitions on the underlying factors. In this regard, we estimate a random effects linear regression (RELR) model to evaluate the contributions of several factors on predictability.

First, the chapter gives an overview of studies on predictability in transportation. Then it shows the mathematical framework of the RELR model, followed by a description of the study area, the data collection effort and basic descriptive statistics of the data. This is followed by a discussion of the model results. Finally, the chapter ends with a summary of findings.

6.2 LITERATURE REVIEW

Predictability is considered important in decision making in many fields, including transport. In particular, the uncertainty (or ill-predictability) of time-varying and short-term processes such as non-recurrent congestion poses severe problems to traffic authorities. Van Zuylen et al. (1999) characterised the sources of ill-predictability in traffic phenomena into two categories; one due to uncertainty and incompleteness of data and models, and the other due to the complexity of the processes itself.

There are numerous methodologies (for example, neural networks, SARIMA, ARIMA, etc.) to forecast time series data, yet the success of these methods remains on the structure of the particular phenomenon, whether it is easily predictable or not (Turkay 2014). For instance, if the prediction obtained from a specific method is weak, but the time series contains an excellent predictive structure, one can practically conclude that the employed prediction technique is unsuitable to the task and that one should try a different technique (Garland et al. 2014). Therefore, it is paramount to find whether the structure of time series data is complex or easily predictable before resorting to advanced prediction techniques (Yue et al. 2007). Nevertheless, there are only a handful of studies in the literature that evaluate the predictability of traffic entities.

Song et al. (2010) used the entropy method to quantify predictability in human mobility patterns. They analysed mobile phone call data (frequency and sequence of location visits) of 50,000 individuals and found a 93% potential theoretical predictability in user mobility. Similarly, Lu et al. (2012) evaluated mobile phone records of 1.9 million users in Haiti after an earthquake and quantified mobility patterns using the entropy method. They found that the upper-bound predictability is still high at 85%. Later, Lu et al. (2013) found similar high predictability of 88% using data from mobile phone users in Ivory Coast of West Africa. Further, they used Markov chain (MC) models and found that the theoretical limit of predictability can be approached. All these studies proved that regardless of the locations travelled by users, there is high predictability in their mobility patterns.

However, mobile phone data have been typically of sparse spatiotemporal resolution. Addressing this issue, Lin et al. (2012) used high-resolution GPS data, and still found 90% predictability in mobility sequences at an hourly sampling rate. Li et al. (2014) used taxi GPS data from Shanghai and Beijing in China and found using the entropy method that the theoretical predictability of the location of the taxis ranges from 78% to 99%. Similarly, Wang et al. (2015) used GPS data from 12,000 taxis in Beijing and found that the predictability of their movements is more than 80%. Moreover, they found that the daily traffic patterns on weekdays are of very similar predictability, despite the differences in commuter demand and are only slightly less predictable than the weekends. Recently, Xu et al. (2017) also analysed taxi data from Shanghai and found very high predictability (> 90%) of daily "travel time" time series along an expressway section at a 5-min resolution.

Lin et al. (2013) used techniques such as approximate entropy and the Hurst exponent to quantify the predictability of traffic volume time series of different highways in China and the USA. Further, they applied three prediction techniques (SARIMA, SVR, and k-NN) and correlated their performance with the results from the predictability quantification methods. They found that the SVR method was suitable for non-linear datasets and SARIMA and k-NN were suitable for linear datasets. The predictability quantification methods helped them in selecting the parameters of the different prediction methods. Thus, they showed the benefits of evaluating the predictability of datasets before resorting to different prediction techniques.

The findings from these studies are essential for the design and improvement of prediction algorithms. Almost all the studies found that predictability is very high, in the range of 80%-95% in human and vehicle mobility. However, the majority of the studies reviewed above used the entropy method to quantify predictability, without much attention to the other methods. Moreover, they focussed mainly on mobile phone call data, which is of sparse resolution and prone to errors.

6.3 METHODOLOGY

Panel datasets are widely used to study the effect of spatiotemporal variations of the explanatory variables on dependent variables. In such datasets, the unobserved effects associated with a specific region will remain the same over time, thus resulting in the dependent variable to be correlated over time. Similarly, there can be correlation over space, because regions that are nearby may share unobserved effects. These correlations violate the assumptions of ordinary least squares (OLS) regression and misestimate the errors on the model coefficients. The random-effects (RE) and random parameters (RP) models are typically considered to account for these correlations (Coruh et al. 2015; Xu and Huang 2015; Truong et al. 2016; Chand et al. 2018). In the case of the RE model, the common unobserved effects are assumed to be distributed across the spatial and temporal units according to some distribution and shared unobserved effects are assumed to be uncorrelated with explanatory variables (Lord and Mannering 2010). Therefore, the intercept term is represented by a distribution in RE models. In the case of RP models, each estimable parameter (including intercept) of the model can vary across observations in the dataset. In this regard, RP model can be considered as a more-flexible extension of RE model.

As noted earlier in Chapter 5, the RP models account for unobserved heterogeneity and offer a better fit than fixed-parameters models, yet, they are time-consuming and complicated to estimate, due to the simulation-based likelihood estimation. Furthermore, the analyst is required to select the random parameters and their appropriate distribution. The RP approach may not necessarily improve the model, and for studies with several data points (13,468 in this study) and explanatory variables, using an RP approach can be computationally intensive due to simulation-based Halton sequences; subject to errors in specification because the modeller needs to select the variables with distributed parameters; and non-parsimoniousness because of the many parameters to be estimated (Chen and Tarko 2014; Pande et al. 2017; Shugan 2006; Washington et al. 2010). Therefore, in the current study, we estimate a random-effects linear regression model (RELR) to model the theoretical predictability of traffic counts. The RELR model has the following form:

$$Y_{it} = \alpha + \beta^{|} X_{it} + \varepsilon_{it} + u_i \qquad (6.1)$$

where

Y_{it} = the dependent variable where i = entity and t = time,

α = the intercept term,

X_{it} = value of the independent variable for group i at time t,

$\beta^{|}$ = coefficient of independent variables,

ε_{it} = within entity error term, and

u_i = between entity error term.

Furthermore, the RE model assumes that the error term of the entity is not correlated with the predictors, so that the time-invariant variables can also be treated as independent variables in the regression model (Greene 2016).

$$\mathrm{Cov}(u_i, X_{it}) = 0 \text{ for all } t, \tag{6.2}$$

$$\mathrm{Exp}[u_i | X_{it}] = 0, \tag{6.3}$$

$$\mathrm{Var}[u_i | X_{it}] = \sigma_u^2, \text{ and} \tag{6.4}$$

$$\mathrm{Cov}[\varepsilon_{it} u_i | X_{it}] = 0. \tag{6.5}$$

The random effects model is a generalised regression model. It is homoscedastic, as all disturbances have variance, which is:

$$Var[\varepsilon_{it} + u_i] = \sigma^2 = {\sigma_u}^2 + {\sigma_\varepsilon}^2 \tag{6.6}$$

But, for a given i, the disturbances in different periods are correlated because of their common part, u_i, which is:

$$Corr[\varepsilon_{it} + u_i, \varepsilon_{is} + u_i] = \frac{{\sigma_u}^2}{\sigma^2} \tag{6.7}$$

Although there is an ML-based estimation method for RELR model, the estimates based on generalised least squares (GLS) is efficient (Greene 2016). This study uses the GLS estimator.

6.4 STUDY AREA AND DATA COLLECTION

6.4.1 Study Area

The central business district (CBD) area of Sydney, Australia is the focus area for this study. Sydney is the largest city in Australia and the eighth largest in the southern hemisphere with an estimated population of 5 million. The city is famous for its sporting events, new year fireworks, three-week lighting festival in May and

June, and several other special events that run throughout the year. Further, the busy, modern and vibrant lifestyle attracts millions of tourists from across the world. The CBD of Sydney employs 13% of the Sydney region's workforce and generates almost a quarter of the Sydney economy (Grattan Institute 2014). Public transport is the main mode for commuters during weekdays. As shown in Figure 6-3, access to jobs by public transport is very good for the CBD when compared with other areas of Sydney. Traffic congestion is a major problem in the city, being ranked 29th among the cities evaluated by TomTom (TomTom 2017). The morning and evening peak hours are especially notorious for congestion, increasing travel times by about 67% compared to the free-flow conditions.

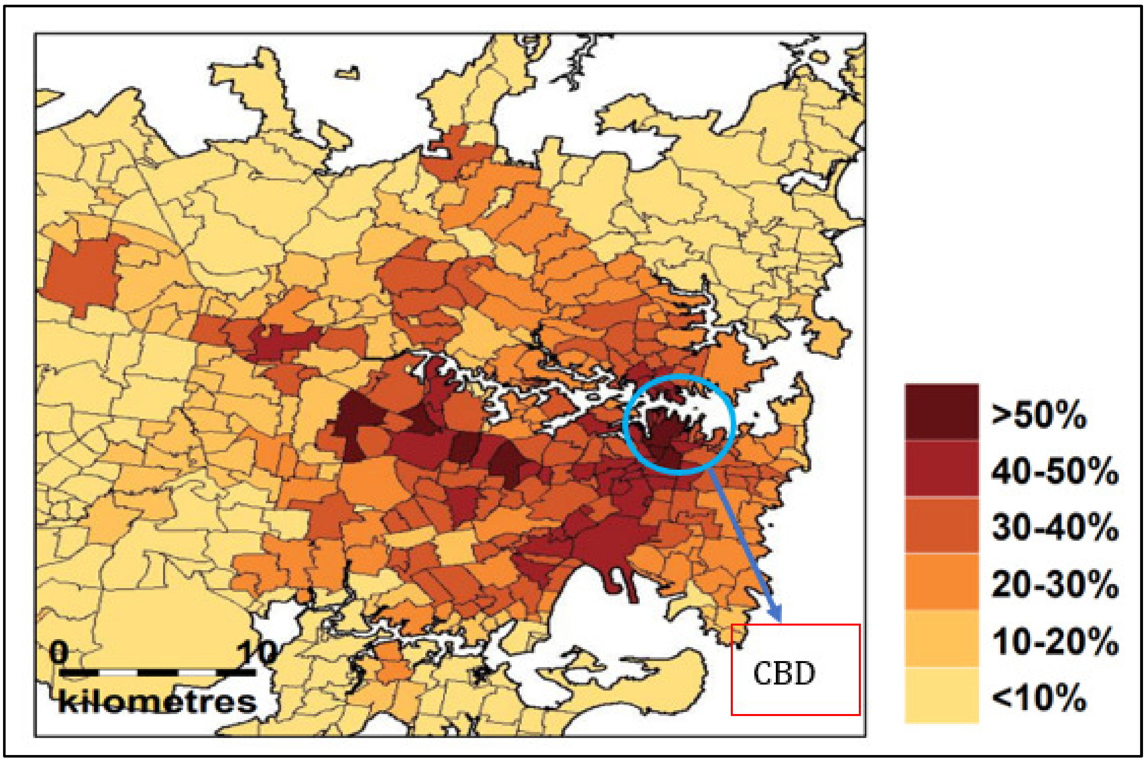


Figure 6-3 Percentage of jobs that can be reached in 60 minutes by public transport in Sydney as of 2011 {Source: (Grattan Institute 2014)}

6.4.2 Data

The data for this study were obtained from Sydney Coordinated Adaptive Traffic System (SCATS) intersection counts in the Sydney CBD. SCATS is a fully adaptive urban traffic control system that optimises traffic flow at intersections. It gathers traffic count data in real-time at each intersection and then makes incremental adjustments to traffic signal timings based on the variations of traffic counts at the intersections (Lowrie 1982; Hunter et al. 2012; Zhang et al. 2013). Therefore, SCATS is an "on-line" algorithm, in which the designed control strategy "matches" the current traffic conditions to the "best" pre-calculated off-line timing plan (Yu and Recker 2006). Its self-calibrating software minimises manual intervention, which can result in substantial operational cost savings. Loop detectors are located at intersection approaches on all the lanes at the stop line, which record the counts of vehicles for every pre-specified time interval. Figure 6-4 shows the schematic of a SCATS signal depicting detector locations and phases.

Figure 6-4 Layout of a SCATS traffic signal showing detector locations and phases (Courtesy: Roads and Maritime Services, Sydney)

We collected traffic count data at a 5-min resolution at around 180 signalised intersections in and around the Sydney CBD, starting from 15th October 2014 to 13th October 2015, translating to 364 days of observations. However, after careful examination for missing and erroneous data, we selected 37 intersections for the analysis. Figure 6-5 shows the layout of these 37 intersections along with their average daily intersection approach volumes. Although the SCATS interface outputs count data at each detector, we considered the aggregated[27] data for the analysis. The geometric features of the intersections vary with respect to the number of approaches, number of bus stops in the vicinity, number of lanes and number of parking lanes. We estimated the Hurst exponent and the Predictability Index (using Equation 2.3) for every day at each intersection. In total, there are 13,468 (37 signals * 364 days) observations of *H* and *PI*.

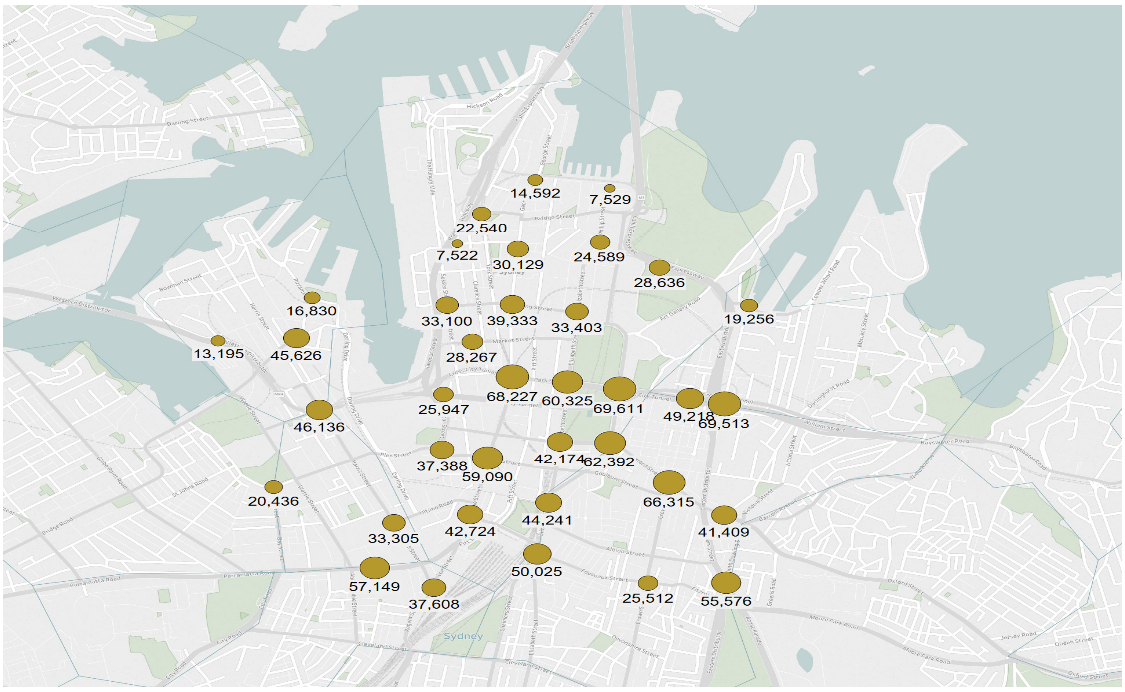


Figure 6-5 Layout of the analysed intersections in the Sydney CBD

[27] There are so many detectors at each intersection that record traffic counts. It is a cumbersome process to do the analysis in a disaggregate manner. Therefore, it was decided to aggregate total traffic count at intersection.

Figure 6-6 shows the PDF and kernel density plots of *PI*. A sizeable portion of the *PI* observations is close to 1, indicating a highly predictable trend at most intersections on most days. However, a non-trivial proportion of observations exist with slightly lower predictability (PI<0.8). Furthermore, the kernel density estimation shows three modes of the *PI*. This chapter investigates the influence of various factors behind such low/high predictability.

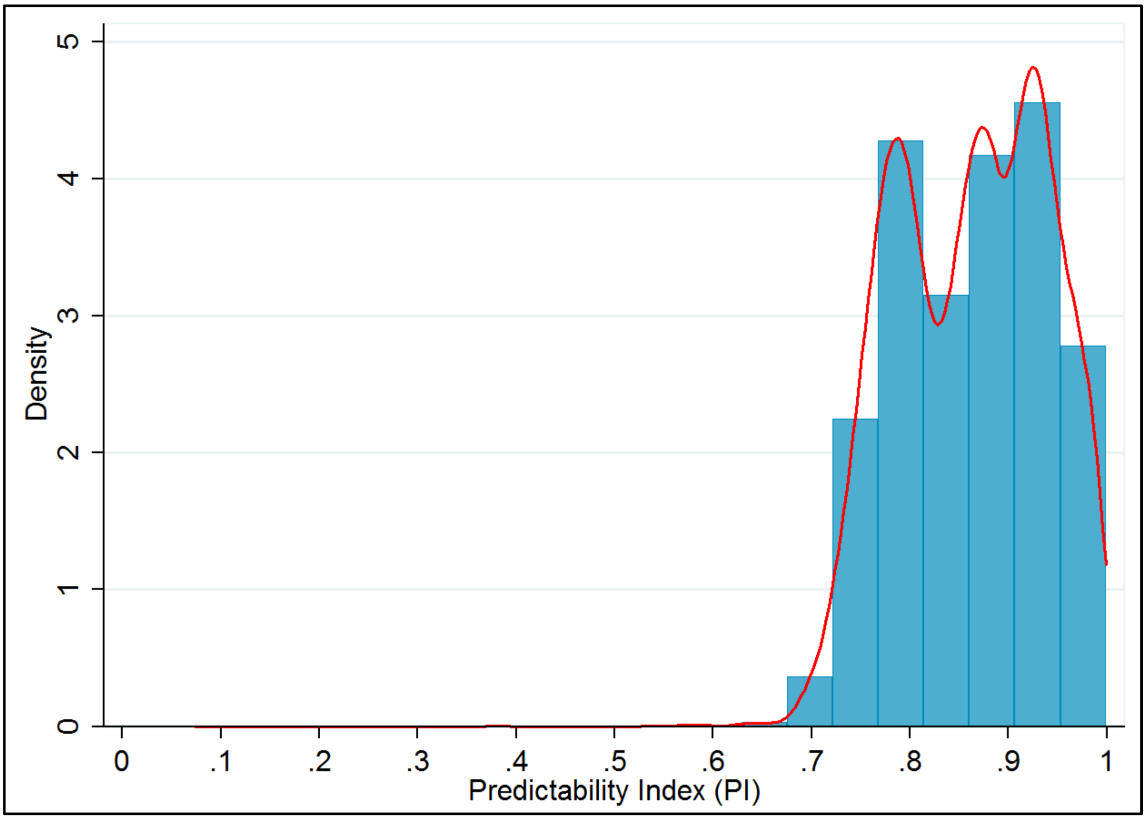


Figure 6-6 PDF and kernel density plots of Predictability Index *(PI)*

6.5 DATA ANALYSIS

6.5.1 Preliminary Analysis

It is surmised that the day of the week will have a profound impact on predictability. Figure 6-7 shows the variation of average *PI* of all the intersections with the day of the week. The weekends (Saturdays and Sundays) tend to have higher predictability, compared to the normal weekdays. Moreover, the variation of *PI* is low on weekends. It is also seen that the predictability is low on Mondays and gradually increases as the week goes on.

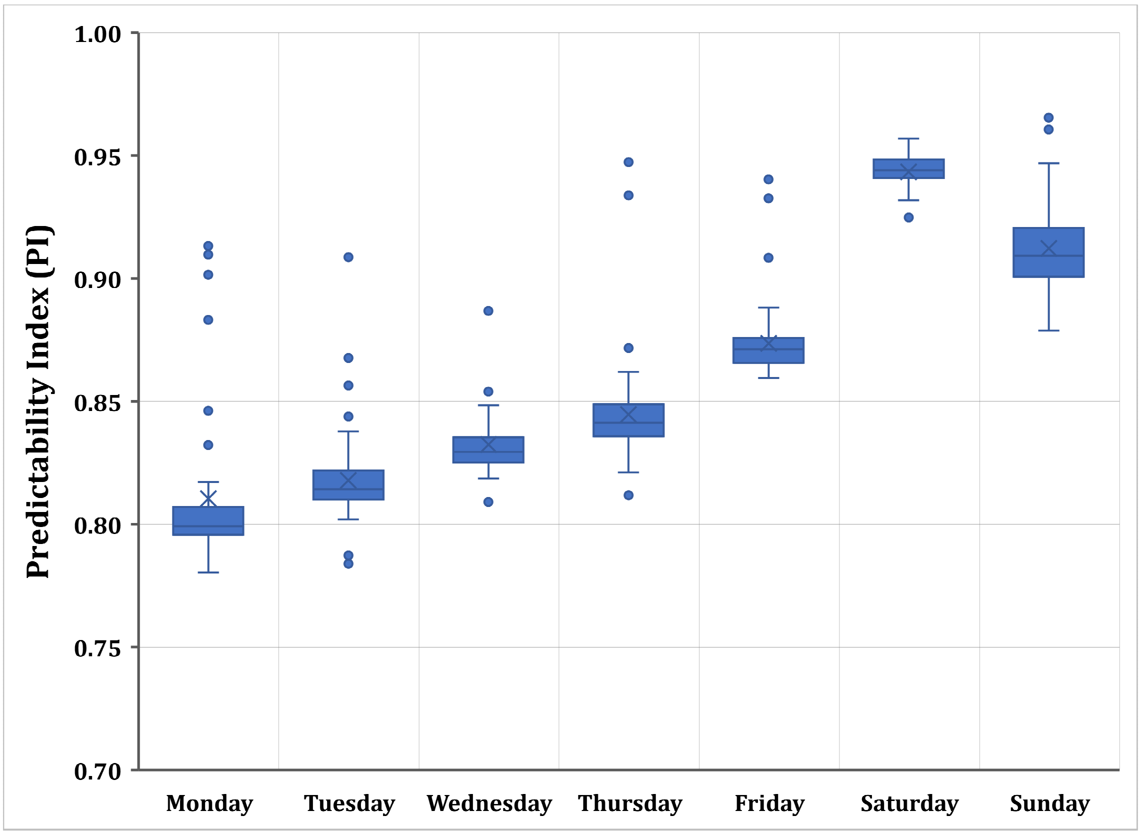


Figure 6-7 Variation of average *PI* (of all intersections) with the day of the week

We performed a one-way ANOVA test for the average *PI* of the day of the week. This test showed that the differences are statistically significant [F(6, 13461) = 1246.53, p-value < .001]. Further, Tukey's posthoc HSD test showed that the mean *PI* of each day of the week is statically different from every other day, implying a strong influence of the day of the week on intersection count predictability.

However, there are several special event days[28], public holidays, days with heavy rain, during the analysis period which should be accounted while modelling predictability. Additionally, the role of intersection geometry and the presence of bus stops and parking cannot be ignored. The following section presents the results of the RELR model which considers the above specified factors.

6.5.2 Statistical Model

Table 6-1 shows the descriptive statistics of all the potential variables considered in the model estimation. We arranged the dataset in a panel data format, consisting of daily estimated theoretical predictability indices of traffic counts at 37 signalised intersections. Thus, the dataset is a strongly balanced panel with 13,468 observations including 37 panels (all of them have the same number of observations, i.e. 364). We collated data of the weather[29], public holidays, special events, and other relevant variables from various sources for the model estimation. Then, we classified the variables into two types, namely, space-varying and time-varying, that vary across the intersections and days respectively.

[28] There were several special events during the data collection period that could have significantly affected the traffic movement in the CBD. Cricket matches (ICC world cup and BBL), Football league games (NRL and AFL), NYE fireworks, marathon runs, Vivid lights festival, Mardi Gras parade are some of these events. As can be seen in Table 6-1, there are 33 of them during the study period. Because of the varying nature of individual events, it is difficult to classify which event as special. In this study, all the events that had attendance of at least 10,000 people are considered as "special events".

[29] The daily rainfall data obtained from the Bureau of Meteorology (BOM), Australia is used. There are two rainfall stations in the study area and they make observations of daily rainfall at 9 AM and record the total for the previous 24 hours. Rainfall includes all forms of precipitation that reach the ground, such as rain, drizzle, hail and snow (BOM 2018). Since the fluctuation analysis in this study is based on daily count data, i.e. from 12 AM to 12 AM, a weighted average of rainfall based on the number of hours is considered. On the other hand, the temperature data was obtained from BOM in the form of average temperature of the day, i.e. from 12 AM to 12 AM. So, temperature data were directly used for the analysis without any modification.

Table 6-1 Descriptive statistics of all variables

Type of Day	**Count**	**Type of Variable**	**% of obs. (categorical variables)**	***for continuous variables***			
				Mean	Std. Dev.	Max	Min
Day of week							
Monday	52	TV	14.29				
Tuesday	52	TV	14.29				
Wednesday	52	TV	14.29				
Thursday	52	TV	14.29				
Friday	52	TV	14.29				
Saturday	52	TV	14.29				
Sunday	52	TV	14.29				
Type of day							
Public holiday	11	TV	3.02				
Special event day	33	TV	9.07				
Season							
Spring	90	TV	24.73				
Autumn	92	TV	25.27				
Winter	92	TV	25.27				
Summer	90	TV	24.73				
Rainfall (mm)	364	TV	---	3.65	10.27	116.02	0
Temperature (^{0}C)	364	TV	---	19.03	4.44	29.2	9.6
AADT	37	SV	---	36,360	15,772	64,189	7,237
Lanes	37	SV	---	9.54	3.33	16	3
<=8 lanes	7	SV	18.92				
8-12 lanes	23	SV	62.16				
> 12 lanes	7	SV	18.92				
Approaches	37	SV	---	3.43	0.59	4	1
1, 2	2	SV	5.41				
3	17	SV	45.95				
4	18	SV	48.64				
Parking Lanes	37	SV	---	1.35	1.07	4	0
No parking	10	SV	27.03				
1-2 lanes	22	SV	59.46				
3-4 lanes	5	SV	13.51				
Bus Stops	37	SV	---	1.27	1.22	5	0
No bus stop	12	SV	32.43				
1-2 bus stops	20	SV	54.05				
3,4 and 5 bus stops	5	SV	13.51				
Crashes	37	SV	---	18.16	14.70	72	1

*TV --- time-varying

*SV --- space-varying

In determining the best model specification for the sample, we estimated several RELR models by changing the variables. The variables that are significant at p = 0.10 were considered. After each model specification, we performed multicollinearity diagnostics by estimating the variance-covariance matrix and correlation matrix of the estimated coefficients. Table 6-2 shows the best model in terms of goodness-of-fit among all the RELR model specifications explored. The Table shows the coefficients of the variables, along with robust standard errors, *z* and *p* values. Furthermore, we conducted the Breusch-Pagan Lagrange multiplier test (Breusch and Pagan 1980). Based on the test, we rejected the null hypothesis that variance across entities is zero, indicating that there are significant differences across panels units. In other words, a random-effects model is better than a simple pooled OLS regression. The final model presented in Table 6-2 has only the relevant variables. The model has an R-square value of 0.61.

There are two categorical variables in the model, namely the day of the week and the parking indicator, which have "*Monday*" and "*No parking*" as base categories respectively. The coefficients of other categories are estimated with respect to these categories. The model has a significant constant of 0.768, suggesting that the mean *PI* on Mondays at intersections with parking restrictions is 0.768 when all the other variables have zero values.

The day of the week has a strong influence on the *PI*. The predictability of all other days is significantly higher than that of Mondays. Notably, the weekends are highly predictable, with mean predictability increasing by 10-14 %, while the other variables are held constant. Moreover, it is likely that congestion occurs on the weekdays and so, it becomes easy to reroute on urban roads rather than say motorways, resulting in more fluctuations and therefore, low predictability on weekdays.

Table 6-2 Results of the Final RELR model

Variable	Coefficient	Robust Std. Error	Z	p>|z|
Day of the week				
Monday	Base			
Tuesday	0.0135	0.0027	4.98	<0.01
Wednesday	0.0276	0.0037	7.41	<0.01
Thursday	0.0367	0.0046	7.90	<0.01
Friday	0.0642	0.0064	10.02	<0.01
Saturday	0.1336	0.0086	15.52	<0.01
Sunday	0.1035	0.0063	16.55	<0.01
Type of Day				
Special event day	0.0119	0.0018	6.47	<0.01
Public holiday	0.0705	0.0057	12.42	<0.01
Weather				
Rainfall (in 10 mm)	-0.0049	0.0007	-6.43	<0.01
Temperature (in 10^0C)	0.0093	0.0021	4.49	<0.01
Parking				
No parking	Base			
1-2 lanes	Insignificant	---	---	---
3-4 lanes	0.0680	0.0209	3.25	<0.01
Constant	0.7675	0.0128	60.03	
Sigma_u	0.0404			
Sigma_e	0.0422			
Rho	0.4781			

Previous studies also found that the type of day can have an enormous impact on highway traffic pattern. For example, Rakha and Van Aerde (1995) found that traffic flows on core weekdays will be different from Mondays, Fridays, Saturdays and Sundays. Similarly, Weijermars and Berkum (2005) classified working days into 1) Mondays, 2) core weekdays, 3) Fridays, and 4) days within vacation period. In these studies, the core weekdays were found to have similar traffic flow patterns. Even in the current study, the predictability varies within the range of 4% from Monday to Thursdays but is significantly higher on other days of the week.

Days with special events are found to have better predictability. One can theorise that the demand rises fairly linearly before the start of the event, and then drops similarly after the event, resulting in high *PI*.

Public holidays increase *PI* of traffic counts by 0.07 when all other variables are held constant. The traffic profile on public holidays tends to be similar to that of Sundays (Keay and Simmonds 2005), which could be a reason for similar predictability levels.

Rainfall is found to have a negative impact on *PI*. While a light shower may not significantly reduce the predictability index, rainfall of more than 20 mm per day could reduce it by 0.01 when other factors are held constant. Travel behaviour could significantly change when there is a massive shower. Heavy rains could persuade the otherwise public transit users to seek car or even cancel their trips. Past research suggested the importance of accounting for rainfall in short-term traffic predictions (Tsirigotis et al. 2012) as rainfall was found to reduce travel demand and average speeds, particularly on weekends (Chung et al. 2005). Recreational trips were found to have greater sensitivity to weather changes (Sabir et al. 2010). Therefore, even the theoretical predictability index of traffic count reduces due to heavy rainfall.

Temperature has a small yet positive and significant effect on the *PI*. A 10^0C increase in temperature could increase *PI* by 0.01 if all other variables were held constant.

Table 6-2 shows that the intersection geometry also has some role in predictability. The intersections with at least three parking lanes (i.e. on at least three approaches) are found to have higher *PI* than the ones with either no parking or parking on up to two approaches. On-street parking is limited in the CBD, and therefore, the intersections near roads with more parking spaces get a continuous stream of vehicles throughout the day. Therefore, such intersections tend to have less fluctuations and so, higher *PI* of traffic counts.

6.6 SUMMARY

Traffic counts at intersections are consistent and repetitive on the one hand, and yet can be variable and less predictable on the other hand in which at any given time, unusual circumstances such as crashes and adverse weather, erratic driver behaviour, etc. can dramatically change the condition of road traffic. These anomalies can create congestion and uncertainty in the transport networks and therefore, from an operational standpoint, it is crucial to detect as early as possible potential irregular traffic patterns.

This chapter used the Hurst exponent method to quantify fluctuations in traffic count data at several intersections in the Sydney CBD. *PI* is typically upwards of 0.80 (i.e. 80%) for most of the days at most signals. This finding is in line with the earlier studies that found that human mobility is highly predictable, with the upper range of 80% to 95%.

Although the techniques such as the Entropy, Fractal dimension, and Hurst exponent are useful in quantifying the predictability of time series, they do not offer intuitions on what makes a time series hard to predict. Understanding the various aspects leading to high/low complexity in the dataset is critical for better prediction results and the choice of prediction methods. In this regard, this chapter estimated a random effects linear regression model to identify the variables that significantly influence the predictability of traffic counts. The statistical analysis revealed that the count predictability is strongly associated with the day of the week, with lowest on Mondays and highest on Saturdays. Public holidays and special event days are found to have a positive impact on the *PI*. Rainfall has an adverse effect, but the temperature has a small yet significant positive effect. Finally, the intersections with parking on more than two approaches are more predictable than the ones with no parking or parking on up to two approaches.

The predictability index calculated in this study is only a theoretical one, and independent of the prediction method, which can be slightly higher or lower. However, it can aid the development of prediction methods in choosing appropriate parameters. It is inappropriate to claim superiority of a prediction technique over other techniques without evaluating the complexity of the time series data. The proposed technique would be genuinely superior if the *PI* of the time series data is low, but still, the predictions are right. On the other hand, decent predictions for a dataset with already high *PI* by "sophisticated" methods may not be of much help.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

The success of ITS deployment is hugely dependent on accuracy and reliability of the real-time information. For this, accurate short-term predictions of traffic variables such as volumes, speeds and occupancies are essential. However, a traffic system can exist in different states; deterministic, random or chaotic, inherently making the system complicated. Due to traffic management infrastructure and a drivers' instinct to avoid a collision, real traffic flow will not stay chaotic for long and usually maintains its order (Fu et al. 2005). However, traffic variables show strong oscillations depending on the time of the day, demand, weather, driver behaviour, and interactions among different vehicles.

There are several prediction techniques to forecast time series data. However, most models are not transferable from one dataset to another, because the properties of the data vary considerably across settings (Tiesyte and Jensen 2009). The presence of anomalies in traffic flow data can make predictions easily go wrong. Besides, the performance of traffic forecasting models is not only influenced by the effectiveness of the model itself, but also by reliability and predictability/complexity of temporal variations of traffic flow characteristics. Therefore, it is essential to understand the structure of the time series data, whether it is complex or not. Fluctuation analysis provides valuable guidelines for the further design and choice of prediction techniques. Also, from a forecasting point of view, evaluating the stochastic fluctuations in the traffic patterns will be of great help in improving the accuracy, reliability, and robustness of forecasting models.

In this context, this thesis utilised the Hurst exponent from the Fractal theory to quantify fluctuations in three different datasets, namely trajectory, loop detector, and intersection counts. This study could lead to many applications of fractal analysis for traffic operations on highways and urban roads. The primary objective of the thesis is evaluating fluctuations in urban traffic data using the Hurst exponent metric of the Fractal theory. Comparing the Hurst exponent with other methods is not within in the scope of this study. The following sections give an overview of the significant findings of the thesis and suggestions for some future research extensions.

7.1 SIGNIFICANT FINDINGS

The following are some significant findings of this thesis:

Chapter-3

1. Speed and lateral movement in homogeneous traffic are 70% and 54% respectively predictable than that in mixed traffic.
2. Out of all vehicles in mixed traffic, cars are highly predictable, and auto-rickshaws are less predictable.
3. The average lateral position impacts the fluctuations in lateral movement and speed. There will be severe fluctuations when vehicles are closer to the median and the outer lane, than the centreline of the road.

Chapter-4

1. Weekdays, peak hours, proximity to ramps, the presence of horizontal curves, etc. can implicate strong LRD of speed, which could be a potential indicator for congestion.

2. Traffic flow is always more predictable than speed. Nevertheless, Predictability Index in the case of loop detector is always less than 0.70, suggesting the need for advanced prediction techniques.

Chapter-5

1. The high magnitude of the Hurst exponent of speed has a positive effect on the crash rate.
2. Locations closer to entry and exit ramps should be of high priority for crash countermeasures.
3. Citybound locations showed a negative relation with the crash rate.

Chapter-6

1. Theoretical predictability of 5-min traffic counts at signalised intersections is upwards of 0.80 (i.e. 80%) for most of the days.
2. Intersection traffic count predictability is strongly associated with the day of the week.
3. Public holidays, special event days, and weekends are better predictable than typical weekdays.
4. Rainfall decreases *PI*, and intersections with more parking spaces have higher *PI*.

7.2 FUTURE RESEARCH RECOMMENDATIONS

Lack of superior quality data is a concern for the analysis of fluctuations. Errors in measurements, faulty instruments, sampling at wrong frequency will not efficiently capture system dynamics. Therefore, care should be taken in collecting data, and appropriate sanity checks should be performed.

Existing studies utilise various statistical or empirical methods, focusing on finding the fractal behaviour or long-range dependence of different traffic flow variables. However, it is equally important to observe the real-world applications of fractal analysis. There is potential for applying fractal and chaotic analysis in transportation problems.

Particularly, the proposed methodology in Chapter 5 of this thesis can easily be applied to newly constructed roads and in developing countries, with insufficient crash data and no standardised police reports. Congestion patterns in terms of the Hurst exponent can be assessed for a limited duration (say one month), for the potential identification of crash hotspots.

Fluctuations in lateral position and speed could change based on traffic density and the type of lead vehicle. This aspect needs to be explored. Furthermore, the effect of turbulent and smooth leader vehicles on the behaviour of the following vehicles and the throughput of the roadway will be an exciting extension of this analysis. The type of following behaviour (typical car-following, staggered following, etc.) could influence the Hurst values and need to be considered. Further, it would be interesting to evaluate fluctuations in speed and lateral movement using different trajectory datasets and develop an empirical relationship with the road capacity.

A study evaluating instabilities in driver behaviour (through the measurement of heart rate, head and eye movement, speed and acceleration) before accident and during normal driving situations would be extremely helpful in developing accident warning models. Naturalistic driving data could be used for such studies.

Evaluating fluctuations in speed, flow and occupancy and relating it with capacity could give some interesting insights into the macroscopic traffic behaviour.

Furthermore, fractal and chaos analysis could be applied to crowd flow modelling, particularly during emergency situations. Person-specific trajectories could be explored to see whether mobility is predictable or not in those situations.

Cross-correlation technique could be used to evaluate how the fluctuations in speed and flow travel upstream from one detector to another. Another exciting study could be lane-wise fluctuation analysis, identifying whether fluctuations are correlated (also called synchronised) and observing how fluctuations propagate laterally.

The application of fractal analysis can provide a comparative measure when the geometric configuration of the motorway has been upgraded or changed. Explicitly speaking, locations correlated to historically high frequency of large Hurst values of speed, can be utilised as a surrogate study, to ascertain the level of improvement or decline at the same location, following the upgrade.

Predictability of traffic counts should be evaluated to identify intersections that are more reliable than others. Thus, predictability could be used as one of the performance measures of intersections such as vehicle-throughput, delays, queues, and fuel emissions.

Intersection count predictability could be affected by the geographical location of the intersection. For example, intersections in a suburb may have an intermittent throughput of vehicles than say CBD, where a continuous stream of vehicles may exist. The driving habits of the populace, number of parking manoeuvres, the percentage of heavy vehicles, pedestrian movement, and side friction could influence the predictability. These aspects should be considered in the future studies. Furthermore, application of random parameters models will provide a better understanding of the heterogeneity impacts of different variables.

BIBLIOGRAPHY

Abdel-Aty, M. A., and Radwan, A. E. (2000). "Modeling Traffic Accident Occurrence and Involvement." *Accident Analysis & Prevention*, 32(5), 633–642.

Abry, P., and Veitch, D. (1998). "Wavelet Analysis of Long-Range-Dependent Traffic." *IEEE Transactions on Information Theory*, 44(1), 2–15.

Addison, P. S., and Low, D. J. (1998). "A Novel Nonlinear Car-Following Model." *Chaos (Woodbury, N.Y.)*, 8(4), 791–799.

Adewumi, A., Kagamba, J., and Alochukwu, A. (2016). "Application of Chaos Theory in the Prediction of Motorised Traffic Flows on Urban Networks." *Mathematical Problems in Engineering*, 2016.

Aghabayk, K., Sarvi, M., and Young, W. (2014). "Attribute Selection for Modelling Driver's Car-Following Behaviour in Heterogeneous Congested Traffic Conditions." *Transportmetrica A: Transport Science*, 10(5), 457–468.

Ahmed, M., Abdel-Aty, M., and Yu, R. (2012). "Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather, and Traffic Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2280, 51–59.

Aljanahi, A. A. M., Rhodes, A. H., and Metcalfe, A. V. (1999). "Speed, Speed Limits and Road Traffic Accidents Under Free Flow Conditions." *Accident Analysis & Prevention*, 31(1), 161–168.

Ambarwati, L., Pel, A. J., Verhaeghe, R., and van Arem, B. (2014). "Empirical Analysis of Heterogeneous Traffic Flow and Calibration of Porous Flow Model." *Transportation Research Part C: Emerging Technologies*, 48, 418–436.

Ambühl, L., Loder, A., Menendez, M. & Axhausen, K. W. Empirical Macroscopic Fundamental Diagrams: New Insights from Loop Detector and Floating Car Data. 96th Annual Meeting of the Transportation Research Board, Washington, DC, 2017.

Anastasopoulos, P. C. (2016). "Random Parameters Multivariate Tobit and Zero-Inflated Count Data Models: Addressing Unobserved and Zero-State Heterogeneity in Accident Injury-Severity Rate and Frequency Analysis." *Analytic Methods in Accident Research*, 11, 17–32.

Anastasopoulos, P. C., and Mannering, F. L. (2009). "A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models." *Accident Analysis & Prevention*, 41(1), 153–159.

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., and Haddock, J. E. (2012a). "A Study of Factors Affecting Highway Accident Rates Using the Random-Parameters Tobit Model." *Accident Analysis & Prevention*, 45, 628–633.

Anastasopoulos, P. C., Shankar, V. N., Haddock, J. E., and Mannering, F. L. (2012b). "A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates." *Accident Analysis & Prevention*, 45, 110–119.

Anastasopoulos, P. C., Tarko, A. P., and Mannering, F. L. (2008). "Tobit Analysis of Vehicle Accident Rates on Interstate Highways." *Accident Analysis & Prevention*, 40(2), 768–775.

Arasan, V. T., and Koshy, R. Z. (2005). "Methodology for Modeling Highly Heterogeneous Traffic Flow." *Journal of Transportation Engineering*, 131(7), 544–551.

Asaithambi, G., and Basheer, S. (2017). "Analysis and Modeling of Vehicle Following Behavior in Mixed Traffic Conditions." *Transportation Research Procedia*, World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016, 25, 5094–5103.

Babu, F. A. M., Vortisch, P., and Mathew, T. V. (2015). "Modelling of Motorcycle Movements in Mixed Traffic Conditions." *Transportation Research Board 94th Annual Meeting*, Washington DC, USA.

Bărbulescu, A., Serban, C., and Maftei, C. (2010). "Evaluation of Hurst Exponent for Precipitation Time Series." *Proceedings of the 14th WSEAS International Conference on Computers: Part of the 14th WSEAS CSCC Multiconference - Volume II*, ICCOMP'10, World Scientific and Engineering Academy and Society (WSEAS), Corfu Island, Greece, 590–595.

Behnood, A., and Mannering, F. (2017). "The Effect of Passengers on Driver-Injury Severities in Single-Vehicle Crashes: A Random Parameters Heterogeneity-in-Means Approach." *Analytic Methods in Accident Research*, 14, 41–53.

Bell, K. (2017). "Google Maps' New Feature Will Help You Avoid Traffic Jams." Mashable Australia. https://mashable.com/2017/07/17/google-maps-app-traffic-predictions/#1BMj0aeUFPq3. (Mar. 10, 2018).

Belomestny, D., and Siegel, H. (2003). *Stochastic and Self-Similar Nature of Highway Traffic Data*. DFG-Schwerpunktprogramm, 1114, pp.1-13.

Bhat, C. R. (2001). "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." *Transportation Research Part B: Methodological*, 35(7), 677–693.

Bhat, C. R., Born, K., Sidharthan, R., and Bhat, P. C. (2014). "A Count Data Model with Endogenous Covariates: Formulation and Application to Roadway Crash Frequency at Intersections." *Analytic Methods in Accident Research*, 1, 53–71.

Bickel, P. J., Chen, C., Kwon, J., Rice, J., Zwet, E. van, and Varaiya, P. (2007). "Measuring Traffic." *Statistical Science*, 22(4), 581–597.

BITRE. (2015). *Traffic and Congestion Cost Trends for Australian Capital Cities*. Bureau of Infrastructure, Transport and Regional Economics, Department of Infrastructure and Regional Development, Government of Australia., Canberra, Australia.

Bo, Q., and Rashed, K. (2004). "Hurst Exponent and Financial Market Predictability." *IASTED Conference on Financial Engineering and Applications (FEA 2004)*, Cambridge, MA, USA, 203–209.

BOM. (2018). "Climate Data Online, Bureau of Meteorology, Australian Government." <http://www.bom.gov.au/climate/data/> (Feb. 22, 2018).

Brackstone, M., and McDonald, M. (1999). "Car-Following: A Historical Review." *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4), 181–196.

Breiman, L., Lawrence, R., Goodwin, D., and Bailey, B. (1977). "The Statistical Properties of Freeway Traffic." *Transportation Research*, 11(4), 221–228.

Breslin, M. C., and Belward, J. A. (1999). "Fractal Dimensions for Rainfall Time Series." *Mathematics and Computers in Simulation*, 48(4–6), 437–446.

Breusch, T. S., and Pagan, A. R. (1980). "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *The Review of Economic Studies*, 47(1), 239–253.

Buddhavarapu, P., Scott, J. G., and Prozzi, J. A. (2016). "Modeling Unobserved Heterogeneity Using Finite Mixture Random Parameters for Spatially Correlated Discrete Count Data." *Transportation Research Part B: Methodological*, 91, 492–510.

Budhkar, A. K., and Maurya, A. K. (2017). "Multiple-Leader Vehicle-Following Behavior in Heterogeneous Weak Lane Discipline Traffic." *Transportation in Developing Economies*, 3(2), 20.

Campari, E. G., and Levi, G. (2002). “Self-Similarity in Highway Traffic.” *The European Physical Journal B - Condensed Matter and Complex Systems*, 25(2), 245–251.

Cassidy, M. J., and Windover, J. R. (1995). “Methodology for Assessing Dynamics of Freeway Traffic Flow.” *Transportation Research Record: Journal of the Transportation Research Board*, (1484), 73–79.

Chand, S., Dixit, V.V., Waller, S.T., 2016. Evaluation of Fluctuating Speed and Lateral Movement of Vehicles: Comparison between Mixed Traffic and Homogeneous Traffic. *Transportation Research Record: Journal of the Transportation Research Board 2581*, pp 104–112.

Chand, S., Moylan, E., Waller, S. T., and Dixit, V. V. (2018). “Factors Affecting Vehicle Breakdown Frequency: A Macro-Level Model of Breakdowns in New South Wales, Australia.” *Transportation Research Board 97th Annual MeetingTransportation Research Board*, Washington DC, USA.

Chang, G.-L., and Xiang, H. (2003). *The Relationship Between Congestion Levels and Accidents*. Research Report, University of Maryland, College Park.

Chen, E., and Tarko, A. P. (2014). “Modeling Safety of Highway Work Zones with Random Parameters and Random Effects Models.” *Analytic Methods in Accident Research*, 1, 86–95.

Chen, F., Ma, X., and Chen, S. (2014). “Refined-Scale Panel Data Crash Rate Analysis Using Random-Effects Tobit Model.” *Accident Analysis & Prevention*, 73, 323–332.

Chen, Q., and Wang, Y. (2016). “A Cellular Automata (CA) Model for Motorized Vehicle Flows Influenced by Bicycles Along the Roadside.” *Journal of Advanced Transportation*, 50(6), 949–966.

Chen, X., Shao, C., Li, D., and Dong, C. (2009). “Capacity Reliability of Signalized Intersections with Mixed Traffic Conditions.” *Tsinghua Science and Technology*, 14(3), 333–340.

Chiabaut, N., Buisson, C., and Leclercq, L. (2009). “Fundamental Diagram Estimation through Passing Rate Measurements in Congestion.” *IEEE Transactions on Intelligent Transportation Systems*, 10(2), 355–359.

Choudhury, C. F., and Islam, M. M. (2016). “Modelling Acceleration Decisions in Traffic Streams with Weak Lane Discipline: A Latent Leader Approach.” *Transportation Research Part C: Emerging Technologies*, 67, 214–226.

Chunchu, M., Kalaga, R. R., and Seethepalli, N. V. S. K. (2010). "Analysis of Microscopic Data under Heterogeneous Traffic Conditions." *Transport*, 25(3), 262–268.

Chung, E., Ohtani, O., Warita, H., Kuwahara, M., and Morita, H. (2005). "Effect of Rain on Travel Demand and Traffic Accidents." *Proceedings of the IEEE Intelligent Transportation Systems Conference*, Vienna, Austria, 1080–1083.

Coifman, B., Krishnamurthy, S., and Wang, X. (2005). "Lane-Change Maneuvers Consuming Freeway Capacity." *Traffic and Granular Flow '03*, S. P. Hoogendoorn, S. Luding, P. H. L. Bovy, M. Schreckenberg, and D. E. Wolf, eds., Springer Berlin Heidelberg, 3–14.

Coifman, B., and Li, L. (2017). "A Critical Evaluation of the Next Generation Simulation (NGSIM) Vehicle Trajectory Dataset." *Transportation Research Part B: Methodological*, 105, 362–377.

Coruh, E., Bilgic, A., and Tortum, A. (2015). "Accident Analysis with Aggregated Data: The Random Parameters Negative Binomial Panel Count Data Model." *Analytic Methods in Accident Research*, 7, 37–49.

Dendrinos, D. S. (1994). "Traffic-Flow Dynamics: A Search for Chaos." *Chaos Solitons and Fractals*, 4(4), 605–617.

DePetrillo, P. B., Speers, d'Armond, and Ruttimann, U. E. (1999). "Determining the Hurst Exponent of Fractal Time Series and Its Application to Electrocardiographic Analysis." *Computers in Biology and Medicine*, 29(6), 393–406.

Dhamaniya, A., Chand, S., and Chandra, S. (2017). "Lane Change as a Measure of Capacity Reduction Near Curbside Bus Stops Under Mixed Traffic Conditions." *96th Annual Meeting of the Transportation Research Board*, Washington DC, USA.

Di, B., Shi, K., Zhang, K., Svirchev, L., and Hu, X. (2016). "Long-Term Correlations and Multifractality of Traffic Flow Measured by GIS for Congested and Free-Flow Roads." *Fractals*, 24(01), 1650012.

DiMento, J. F. (2009). *Stent (or Dagger?) in the Heart of Town: Urban Freeways in Syracuse, 1944-1967*. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY.

Disbro, J. E., and Frame, M. (1989). *Traffic Flow Theory and Chaotic Behavior*. New York (State). Dept. of Transportation, New York, NY, USA.

Donahue III, M. J. (1997). “An Introduction to Chaos Theory and Fractal Geometry.” <http://plouffe.fr/simon/math/An%20Introduction%20to%20Mathematical%20Chaos%20Theory%20and%20Fractal%20Geometry(1).pdf>.

Dong, C., Dong, Q., Huang, B., Hu, W., and Nambisan, S. S. (2017). “Estimating Factors Contributing to Frequency and Severity of Large Truck–Involved Crashes.” *Journal of Transportation Engineering, Part A: Systems*, 143(8), 04017032.

Eluru, N., Bagheri, M., Miranda-Moreno, L. F., and Fu, L. (2012). “A Latent Class Modeling Approach for Identifying Vehicle Driver Injury Severity Factors at Highway-Railway Crossings.” *Accident Analysis & Prevention*, 47, 119–127.

Elvik, R., Christensen, P., and Amundsen, A. (2004). *Speed and Road Accidents: An Evaluation of the Power Model*. Institute of Transport Economics TOI, Oslo, Norway, 2004.

Eom, C., Choi, S., Oh, G., and Jung, W.-S. (2008). “Hurst Exponent and Prediction Based on Weak-Form Efficient Market Hypothesis of Stock Markets.” *Physica A: Statistical Mechanics and its Applications*, 387(18), 4630–4636.

Feng, S., Wang, X., Sun, H., Zhang, Y., and Li, L. (2018). “A Better Understanding of Long-Range Temporal Dependence of Traffic Flow Time Series.” *Physica A: Statistical Mechanics and its Applications*, 492, 639–650.

FHWA. (2007a). *NGSIM-Next Generation Simulation*. Federal Highway Administration (FHWA), U.S. Department of Transportation.

FHWA. (2007b). “Lankershim Boulevard Dataset, FHWA-HRT-07-029.” <https://www.fhwa.dot.gov/publications/research/operations/07029/index.cfm> (Feb. 16, 2018).

FHWA. (2007c). “US Highway 101 Dataset, FHWA-HRT-07-030.” <https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm> (Feb. 16, 2018).

FHWA. (2012). *Highway Statistics 2010. US Department of Transportation, Federal Highway Administration (FHWA)*. Washington, DC. USA.

Fractal Foundation. (2017). “What is Chaos Theory?” <https://fractalfoundation.org/resources/what-is-chaos-theory/> (Feb. 5, 2018).

Frazier, C., and Kockelman, K. (2004). “Chaos Theory and Transportation Systems: Instructive Example.” *Transportation Research Record*, 1897(1), 9–17.

Fu, H., Xu, J., and Xu, L. (2005). “Traffic Chaos and its Prediction Based on a Nonlinear Car-Following Model.” *Journal of Control Theory and Applications*, 3(3), 302–307.

Fu, Y., and Wei, W. (2013). “Research on the Short-Term Traffic Flow Forecasting Method Based on Chaos Theory.” *Fourth International Conference on Transportation Engineering*, American Society of Civil Engineers, Chengdu, China, 2637–2642.

Garber, N. J., and Gadiraju, R. (1989). “Factors Affecting Speed Variance and Its Influence on Accidents.” *Transportation Research Record: Journal of the Transportation Research Board*, 1213, 64–71.

Garber, N., and Subramanyan, S. (2001). “Incorporating Crash Risk in Selecting Congestion-Mitigation Strategies: Hampton Roads Area (Virginia) Case Study.” *Transportation Research Record: Journal of the Transportation Research Board*, 1746, 1–5.

Garland, J., James, R., and Bradley, E. (2014). “Model-Free Quantification of Time-Series Predictability.” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 90(5–1), 052910.

Gazis, D. C., Herman, R., and Rothery, R. W. (1961). “Nonlinear Follow-the-Leader Models of Traffic Flow.” *Operations Research*, 9(4), 545–567.

Geweke, J., and Porter-Hudak, S. (1983). “The Estimation and Application of Long Memory Time Series Models.” *Journal of Time Series Analysis*, 4(4), 221–238.

Gipps, P. G. (1981). “A Behavioural Car-Following Model for Computer Simulation.” *Transportation Research Part B: Methodological*, 15(2), 105–111.

Gneiting, T., Ševčíková, H., and Percival, D. B. (2012). “Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data.” *Statistical Science*, 27(2), 247–277.

Gokhale, S., and Raokhande, N. (2008). “Performance Evaluation of Air Quality Models for Predicting Pm10 and Pm2.5 Concentrations at Urban Traffic Intersection During Winter Period.” *Science of The Total Environment*, 394(1), 9–24.

Gottwald, G. A., and Melbourne, I. (2004). “A New Test for Chaos in Deterministic Systems.” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2042), 603–611.

Granero, M. A. S., Segovia, J. E. T., and Pérez, J. G. (2008). “Some Comments on Hurst Exponent and the Long Memory Processes on Capital Markets.” Physica A: Statistical Mechanics and its Applications, 387(22), 5543–5551.

Grattan Institute. (2014). *Mapping Australia's Economy: Cities as Engines of Prosperity*. Grattan Institute, Melbourne, Australia.

Greene, W. H. (2016). *LIMDEP Version 11.0*. Econometric Software Inc., Plainview, NY, USA.

Greene, W. H., and Hensher, D. A. (2003). "A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit." *Transportation Research Part B: Methodological*, 37(8), 681–698.

Guo, H., Wang, W., Guo, W., and Zhao, F. (2013). "Modeling Lane-Keeping Behavior of Bicyclists Using Survival Analysis Approach." *Discrete Dynamics in Nature and Society*, 2013.

Haleem, K., Alluri, P., and Gan, A. (2016). "Investigating Fractal Characteristics in Crash Trends for Potential Traffic Safety Prediction." *Transportation Research Board 95th Annual Meeting*, Washington DC, USA.

Haleem, K., Alluri, P., Gan, A., Li, H., and Li, T. (2013). *Synthesis of the Advance in and Application of Fractal Characteristics of Traffic Flow*. Final Report, State of Florida Department of Transportation, Tallahassee, FL, USA.

Hall, J. W., and Pendleton, O. J. (1989). *Relationship Between V/C Ratios and Accident Rates*. New Mexico State Highway and Transportation Department, New Mexico, USA.

Halton, J. H. (1960). "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals." *Numerische Mathematik*, 2(1), 84–90.

Hauer, E. (1996). "Identification of Sites with Promise." *Transportation Research Record: Journal of the Transportation Research Board*, 1542, 54–60.

Havlin, S., Amaral, L. A. N., Ashkenazy, Y., Goldberger, A. L., Ivanov, P. C., Peng, C.-K., and Stanley, H. E. (1999). "Application of Statistical Physics to Heartbeat Diagnosis." *Physica A: Statistical Mechanics and its Applications*, 274(1), 99–110.

Higuchi, T. (1988). "Approach to an Irregular Time Series on the Basis of the Fractal Theory." *Physica D: Nonlinear Phenomena*, 31(2), 277–283.

Hu, J., Wang, Y., Zhang, Z., and Li, D. (2010). "Analysis on Traffic Flow Data and Extraction of Nonlinear Characteristic Quantities." *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Madeira Island, Portugal, 712–717.

Hu, J., Zong, C., Song, J., Zhang, Z., and Ren, J. (2003). "An Applicable Short-Term Traffic Flow Forecasting Method based on Chaotic Theory." *IEEE Intelligent Transportation Systems Conference*, IEEE, Shanghai, China, 608–613.

Huang, H., Zhou, H., Wang, J., Chang, F., and Ma, M. (2017). "A Multivariate Spatial Model of Crash Frequency by Transportation Modes for Urban Intersections." *Analytic Methods in Accident Research*, 14, 10–21.

Hunter, M. P., Wu, S. K., Kim, H. K., and Suh, W. (2012). "A Probe-Vehicle-Based Evaluation of Adaptive Traffic Signal Control." IEEE Transactions on Intelligent Transportation Systems, 13(2), 704–713.

Hurst, H. E. (1951). "Long-Term Storage Capacity of Reservoirs." *Transactions of the American Society of Civil Engineers*, 116(1), 770–799.

Hurst, H. E. (1956). "Methods of Using Long-Term Storage in Reservoirs." *Proceedings of the Institution of Civil Engineers*, 5(5), 519–543.

Hurst, H. E., Black, R. P., and Simaika, Y. M. (1965). *Long-Term Storage: An Experimental Study*. London : Constable.

Imran, M. A. (2009). "Neuro - Fuzzy Model for Car Following Behaviour in Heterogeneous Road Traffic Condition." Master's Thesis, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

Jamison, S., and McCartney, M. (2007). "A Vehicle Overtaking Model of Traffic Dynamics." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(3), 033116.

Jamison, S., and McCartney, M. (2009). "Discrete vs. Continuous Time Implementation of a Ring Car Following Model in Which Overtaking Is Allowed." *Nonlinear Analysis: Real World Applications*, 10(1), 437–448.

Jiang, Z., Chen, X. (Michael), and Ouyang, Y. (2017). "Traffic State and Emission Estimation for Urban Expressways Based on Heterogeneous Data." *Transportation Research Part D: Transport and Environment*, 53, 440–453.

Kanagaraj, V., Asaithambi, G., Toledo, T., and Lee, T.-C. (2015). "Trajectory Data and Flow Characteristics of Mixed Traffic." *Transportation Research Record: Journal of the Transportation Research Board*, 2491, 1–11.

Kanhere, S. S. Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces. International Conference on Distributed Computing and Internet Technology, 2013. Springer, 19-26.

Karlaftis, M. G., and Vlahogianni, E. I. (2009). "Memory Properties and Fractional Integration in Transportation Time-Series." *Transportation Research Part C: Emerging Technologies*, 17(4), 444–453.

Katsev, S., and L'Heureux, I. (2003). "Are Hurst Exponents Estimated from Short or Irregular Time Series Meaningful?" Computers & Geosciences, 29(9), 1085–1089.

Kays, H. M. I., Shimu, T. H., Hadiuzzaman, M., Muniruzzaman, S. M., and Rahman, M. M. (2017). "H-CTM for Simulating Non-Lane-Based Heterogeneous Traffic." *Transportation Letters*, 0(0), 1–9.

Keay, K., and Simmonds, I. (2005). "The Association of Rainfall and Other Weather Variables with Road Traffic Volume in Melbourne, Australia." *Accident Analysis & Prevention*, 37(1), 109–124.

Kesting, A., and Treiber, M. (2008). "Calibrating Car-Following Models by Using Trajectory Data: Methodological Study." *Transportation Research Record: Journal of the Transportation Research Board*, 2088, 148–156.

Kim, J.-K., Ulfarsson, G. F., Kim, S., and Shankar, V. N. (2013). "Driver-Injury Severity in Single-Vehicle Crashes in California: A Mixed Logit Analysis of Heterogeneity Due to Age and Gender." *Accident Analysis & Prevention*, 50, 1073–1081.

Kockelman, K. M., and Kweon, Y.-J. (2002). "Driver Injury Severity: An Application of Ordered Probit Models." *Accident Analysis & Prevention*, 34(3), 313–321.

Kononov, J., Bailey, B., and Allery, B. (2008). "Relationships Between Safety and Both Congestion and Number of Lanes on Urban Freeways." *Transportation Research Record: Journal of the Transportation Research Board*, 2083, 26–39.

Kovvali, V. G., Alexiadis, V., and Zhang P. E., L. (2007). "Video-Based Vehicle Trajectory Data Collection." *Transportation Research Board 86th Annual Meeting*, Washington DC, USA.

Krause, S. M., Habel, L., Guhr, T., and Schreckenberg, M. (2017). "The Importance of Antipersistence for Traffic Jams." *EPL (Europhysics Letters)*, 118(3), 38005.

Krese, B., and Govekar, E. (2013). "Analysis of Traffic Dynamics on a Ring Road-Based Transportation Network by Means of 0–1 Test for Chaos and Lyapunov Spectrum." *Transportation Research Part C: Emerging Technologies*, 36, 27–34.

Lam, T., and Rothery, R. (1970). "The Spectral Analysis of Speed Fluctuations on a Freeway." *Transportation Science*, 4(3), 293–310.

Lan, L. W., Lin, F.-Y., and Kuo, A. Y. (2003). "Testing and Prediction of Traffic Flow Dynamics with Chaos." *Journal of the Eastern Asia Society for Transportation Studies*, 5, 1975–1990.

Lan, L. W., Lin, F.-Y., and Kuo, A. Y. (2005). "Identification for Chaotic Phenomena in Short-Term Traffic Flows: A Parsimony Procedure with Surrogate Data." *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1518–1533.

Lan, L. W., Sheu, J.-B., and Huang, Y.-S. (2007). "The Characteristics of Temporal Traffic Flow Dynamics." *Journal of the Eastern Asia Society for Transportation Studies*, 7, 2246–2260.

Laval, J. A., and Daganzo, C. F. (2006). "Lane-Changing in Traffic Streams." *Transportation Research Part B: Methodological*, 40(3), 251–264.

Lee, T.-C., Polak, J. W., and Bell, M. G. H. (2008). "Trajectory Extractor User Manual Version 1.0." Centre for Transport Studies, Imperial College London, United Kingdom.

Lee, T.-C., and Wong, K. I. (2016). "An Agent-Based Model for Queue Formation of Powered Two-Wheelers in Heterogeneous Traffic." *Physica A: Statistical Mechanics and its Applications*, 461, 199–216.

Li, H., and Prevedouros, P. D. (2004). "Traffic Adaptive Control for Oversaturated Isolated Intersections: Model Development and Simulation Testing." *Journal of Transportation Engineering*, 130(5), 594–601.

Li, H., Xie, K., and Pan, Q. (2016). "A Comparison Study of Motorized Vehicles' Trajectory in Mixed-Traffic Flow with and without Non-Motorized Vehicles." *16th COTA International Conference of Transportation Professionals*, Proceedings, American Society of Civil Engineers (ASCE), Shanghai, China.

Li, X., and Shang, P. (2007). "Multifractal Classification of Road Traffic Flows." *Chaos, Solitons & Fractals*, 31(5), 1089–1094.

Li, Y., Jin, D., Hui, P., Wang, Z., and Chen, S. (2014). "Limits of Predictability for Large-Scale Urban Vehicular Mobility." *IEEE Transactions on Intelligent Transportation Systems*, 15(6), 2671–2682.

Lin, F., and Lan, L. W. (2005). "Traffic Flow Analysis with Diffeerent Time Scales." *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1624–1636.

Lin, L., Wang, Q., and Sadek, A. (2013). "Short-Term Forecasting of Traffic Volume: Evaluating Models Based on Multiple Data Sets and Data Diagnosis

Measures." *Transportation Research Record: Journal of the Transportation Research Board*, 2392, 40–47.

Lin, L., Yuan, K., and Ren, S. (2014). "Analysis of Urban Freeway Traffic Flow Characteristics Based on Frequent Pattern Tree." *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 1719–1725.

Lin, M., Hsu, W.-J., and Lee, Z. Q. (2012). "Predictability of Individuals' Mobility with High-resolution Positioning Data." *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, ACM, Pittsburgh, Pennsylvania, USA, 381–390.

Lord, D., and Mannering, F. (2010). "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A: Policy and Practice*, 44(5), 291–305.

Lowrie, P. R. (1982). "The Sydney Coordinated Adaptive Traffic System - Principles, Methodology, Algorithms." *International Conference on Road Traffic Signalling*, London, United Kingdom.

Lu, X., Bengtsson, L., and Holme, P. (2012). "Predictability of Population Displacement After the 2010 Haiti Earthquake." *Proceedings of the National Academy of Sciences*, 109(29), 11576–11581.

Lu, X., Wetter, E., Bharti, N., Tatem, A. J., and Bengtsson, L. (2013). "Approaching the Limit of Predictability in Human Mobility." *Scientific Reports*, 3, 2923.

Lu, X.-Y., and Skabardonis, A. (2007). "Freeway Traffic Shockwave Analysis: Exploring NGSIM Trajectory Data." *Transportation Research Board 86th Annual Meeting*, Washington D.C., USA.

Magidson, J., and Vermunt, J. (2004). "Latent Class Models." *Handbook of Quantitative Methodology for the Social Sciences*, SAGE, Newbury Park, CA, USA, 175–198.

Mahapatra, G., and Maurya, A. K. (2013). "Study of Vehicles Lateral Movement in Non-lane Discipline Traffic Stream on a Straight Road." *Procedia - Social and Behavioral Sciences*, 2nd Conference of Transportation Research Group of India (2nd CTRG), 104, 352–359.

Maher, M. J. (1983). "Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach." *Transportation Research Part B: Methodological*, 17(6), 435–447.

Mallikarjuna, C., Phanindra, A., and Rao, K. R. (2009). “Traffic Data Collection under Mixed Traffic Conditions Using Video Image Processing.” *Journal of Transportation Engineering*, 135(4), 174–182.

Mallikarjuna, C., and Rao, K. R. (2009). “Cellular Automata Model for Heterogeneous Traffic.” *Journal of Advanced Transportation*, 43(3), 321–345.

Malyshkina, N. V., and Mannering, F. L. (2009). “Markov Switching Multinomial Logit Model: An Application to Accident-Injury Severities.” *Accident Analysis & Prevention*, 41(4), 829–838.

Malyshkina, N. V., and Mannering, F. L. (2010). “Zero-State Markov Switching Count-Data Models: An Empirical Assessment.” *Accident Analysis & Prevention*, 42(1), 122–130.

Malyshkina, N. V., Mannering, F. L., and Tarko, A. P. (2009). “Markov Switching Negative Binomial Models: An Application to Vehicle Accident Frequencies.” *Accident Analysis & Prevention*, 41(2), 217–226.

Mandelbrot, B. (1967). “How long is the coast of Britain? Statistical self-similarity and fractional dimension.” *Science (New York, N.Y.)*, 156(3775), 636–638.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. WH Freeman New York.

Mandelbrot, B. B. (2004). “Fractals.” *Encyclopedia of Statistical Sciences*, John Wiley & Sons, Inc.

Mannering, F. L., Shankar, V., and Bhat, C. R. (2016). “Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data.” *Analytic Methods in Accident Research*, 11, 1–16.

Mathew, T. V., Munigety, C. R., and Bajpai, A. (2013). “Strip-Based Approach for the Simulation of Mixed Traffic Conditions.” *Journal of Computing in Civil Engineering*, 29(5), pp. 04014069-1-9.

Mathew, T. V., and Ravishankar, K. V. R. (2012). “Neural Network Based Vehicle-Following Model for Mixed Traffic Conditions.” *Eurpean Transport \ Trasporti Europei*, (52).

Matos, J. A. O., Gama, S. M. A., Ruskin, H. J., Sharkasi, A. A., and Crane, M. (2008). “Time and Scale Hurst Exponent Analysis for Financial Markets.” *Physica A: Statistical Mechanics and its Applications*, 387(15), 3910–3915.

Mauch, M., and Cassidy, M. J. (2002). “Freeway Traffic Oscillations: Observations And Predictions.” *Transportation and Traffic Theory in the 21st Century:*

Proceedings of the 15th International Symposium on Transportation and Traffic Theory, Emerald Group Publishing Limited, Adelaide, Australia, 653–673.

Meng, Q., and Khoo, H. L. (2009). “Self-Similar Characteristics of Vehicle Arrival Pattern on Highways.” *Journal of Transportation Engineering*, 135(11), 864–872.

Miaou, S.-P., and Lord, D. (2003). “Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods.” *Transportation Research Record: Journal of the Transportation Research Board*, 1840, 31–40.

Milton, J. C., Shankar, V. N., and Mannering, F. L. (2008). “Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis.” *Accident Analysis & Prevention*, 40(1), 260–266.

Minh, C. C., and Sano, K. (2007). “Acceleration and Deceleration Models of Motorcycle at Signalized Intersections.” *Journal of the Eastern Asia Society for Transportation Studies*, 7, 2396–2411.

Minh, C. C., Sano, K., and Matsumoto, S. (2012). “Maneuvers of Motorcycles in Queues at Signalized Intersections.” *Journal of Advanced Transportation*, 46(1), 39–53.

Moore, D. N., Schneider, W. H., Savolainen, P. T., and Farzaneh, M. (2011). “Mixed Logit Analysis of Bicyclist Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-Intersection Locations.” *Accident Analysis & Prevention*, 43(3), 621–630.

Munigety, C. R., Mantri, S., Mathew, T. V., and Rao, K. V. K. (2014a). “Analysis and Modelling of Tactical Decisions of Vehicular Lateral Movement in Mixed Traffic Environment.” *Transportation Research Board 93rd Annual Meeting*, Washington DC, USA.

Munigety, C. R., Vicraman, V., and Mathew, T. V. (2014b). “Semiautomated Tool for Extraction of Microlevel Traffic Data from Videographic Survey.” *Transportation Research Record: Journal of the Transportation Research Board*, 2443(1), 88–95.

Nagatani, T. (2005). “Self-Similar Behavior of a Single Vehicle Through Periodic Traffic Lights.” *Physica A: Statistical Mechanics and its Applications*, 347, 673–682.

Nagin, D. S. (2005). *Group-Based Modeling of Development*. Harvard University Press, MA, USA.

Nair, A. S., Liu, J.-C., Rilett, L., and Gupta, S. (2001). "Non-Linear Analysis of Traffic Flow." *2001 IEEE Intelligent Transportation Systems, 2001. Proceedings*, IEEE, Oakland, (CA), USA, 681–685.

Nair, R., Mahmassani, H. S., and Miller-Hooks, E. (2011). "A Porous Flow Approach to Modeling Heterogeneous Traffic in Disordered Systems." *Procedia - Social and Behavioral Sciences*, Papers selected for the 19th International Symposium on Transportation and Traffic Theory, 17, 611–627.

Nguyen, L., Hanaoka, S., and Kawasaki, T. (2012). "Describing Non-Lane-Based Motorcycle Movements in Motorcycle-Only Traffic Flow." *Transportation Research Record: Journal of the Transportation Research Board*, 2281, 76–82.

Nguyen, L. X., Hanaoka, S., and Kawasaki, T. (2014). "Traffic Conflict Assessment for Non-Lane-Based Movements of Motorcycles Under Congested Conditions." *IATSS Research*, 37(2), 137–147.

Nihan, N. L., and Davis, G. A. (1989). "Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D Matrices from Traffic Counts." *Transportation Science*, 23(2), 77–90.

Nilsson, G. (2004). "Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety." *Bulletin-Lunds Tekniska Högskola, Inst för Teknik och Samhälle, Lunds Universitet*, 221.

Noland, R. B., and Quddus, M. A. (2005). "Congestion and Safety: A Spatial Analysis of London." *Transportation Research Part A: Policy and Practice*, 39(7), 737–754.

O'Donnell, C. J., and Connor, D. H. (1996). "Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice." *Accident Analysis & Prevention*, 28(6), 739–753.

Oketch, T. (2000). "New Modeling Approach for Mixed-Traffic Streams with Nonmotorized Vehicles." *Transportation Research Record: Journal of the Transportation Research Board*, 1705, 61–69.

de Oña, J., López, G., Mujalli, R., and Calvo, F. J. (2013). "Analysis of Traffic Accidents on Rural Highways Using Latent Class Clustering and Bayesian Networks." *Accident Analysis & Prevention*, 51, 1–10.

Ossen, S., and Hoogendoorn, S. P. (2011). "Heterogeneity in Car-Following Behavior: Theory and Empirics." *Transportation Research Part C: Emerging Technologies*, Emerging theories in traffic and transportation and methods for transportation planning and operations, 19(2), 182–195.

Pal, D., and Chunchu, M. (2017). “Modeling of Lateral Gap Maintaining Behavior of Vehicles in Heterogeneous Traffic Stream.” *Transportation Letters*, 0(0).

Pande, A., Chand, S., Saxena, N., Dixit, V., Loy, J., Wolshon, B., and Kent, J. D. (2017). “A Preliminary Investigation of the Relationships Between Historical Crash and Naturalistic Driving.” *Accident Analysis & Prevention*, 101, 107–116.

Pandey, G., Rao, K. R., and Mohan, D. (2017). “Modelling Vehicular Interactions for Heterogeneous Traffic Flow Using Cellular Automata with Position Preference.” *Journal of Modern Transportation*, 25(3), 163–177.

Parida, M., Jain, S. S., and Landge, V. S. (2006). “Stochastic Modelling for Traffic Crashes on Non Urban Highways in India.” *22nd ARRB Conference*, Canberra, Australia.

Park, B.-J., and Lord, D. (2009). “Application of Finite Mixture Models for Vehicle Crash Data Analysis.” *Accident Analysis & Prevention*, 41(4), 683–691.

Park, H., Oh, C., Moon, J., and Kim, S. (2018). “Development of a Lane Change Risk Index Using Vehicle Trajectory Data.” *Accident Analysis & Prevention*, 110, 1–8.

Park, M., and Lee, D. (2017). “Analysis of severe injury accident rates on interstate highways using a random parameter tobit model.” *Mathematical Problems in Engineering*, 2017, e7273630.

Pelletier, J. D., and Turcotte, D. L. (1997). “Long-Range Persistence in Climatological and Hydrological Time Series: Analysis, Modeling and Application to Drought Hazard Assessment.” *Journal of Hydrology*, 203(1), 198–208.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). “Mosaic Organization of DNA Nucleotides.” *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 49(2), 1685–1689.

Peng, J., Liu, Z., Liu, Y., Wu, J., and Han, Y. (2012). “Trend Analysis of Vegetation Dynamics in Qinghai–Tibet Plateau Using Hurst Exponent.” *Ecological Indicators*, 14(1), 28–39.

Perati, M. R., Raghavendra, K., Koppula, H. K. R., Doodipala, M. R., and Dasari, R. (2012). “Self-Similar Behavior of Highway Road Traffic and Performance Analysis at Toll Plazas.” *Journal of Transportation Engineering*, 138(10), 1233–1238.

Punzo, V., Borzacchiello, M. T., and Ciuffo, B. (2011). “On the Assessment of Vehicle Trajectory Data Accuracy and Application to the Next Generation

Simulation (NGSIM) Program Data." *Transportation Research Part C: Emerging Technologies*, 19(6), 1243–1262.

Punzo, V., and Tripodi, A. (2007). "Steady-State Solutions and Multiclass Calibration of Gipps Microscopic Traffic Flow Model." *Transportation Research Record: Journal of the Transportation Research Board*, 1999, 104–114.

Qi, Y. (Grace), Smith, B. L., and Guo, J. (2007). "Freeway Accident Likelihood Prediction Using a Panel Data Analysis Approach." *Journal of Transportation Engineering*, 133(3), 149–156.

Quddus, M. A., Wang, C., and Ison, S. G. (2009). "Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models." *Journal of Transportation Engineering*, 136(5), 424–435.

Racine, R. (2011). "Estimating the Hurst Exponent." Bachelor Thesis, ETH Zurich, Switzerland.

Rakha, H., and Van Aerde, M. (1995). "Statistical Analysis of Day-to-Day Variations in Real-Time Traffic Flow Data." *Transportation Research Record*, 1510, 26–34.

Rangarajan, G., and Sant, D. A. (1997). "A Climate Predictability Index and its Applications." *Geophysical Research Letters*, 24(10), 1239–1242.

Ravishankar, K. V. R., and Mathew, T. V. (2011). "Vehicle-Type Dependent Car-Following Model for Heterogeneous Traffic Conditions." *Journal of Transportation Engineering*, 137(11), 775–781.

Rhodes, C. J. (2000). "Fluctuation Analysis of M25 Loop Data." *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 214(6), 653–658.

RMS. (2017). "Documents and maps, Roads and Maritime Services (RMS), Government of New South Wales (NSW)." *Roads and Maritime Services, New South Wales*, <http://www.rms.nsw.gov.au/documents/roads/using-roads/alpha-numeric/sydney-map.jpg> (Jul. 20, 2017).

Sabir, M., Ommeren, J., Koetse, M. J., and Rietveld, P. (2010). *Impact of Weather on Daily Travel Demand*. Working paper, Department of Spatial Economics, VU University, Amsterdam, The Netherlands.

Safonov, L. A., Tomer, E., Strygin, V. V., Ashkenazy, Y., and Havlin, S. (2002). "Multifractal Chaotic Attractors in a System of Delay-Differential Equations Modeling Road Traffic." *Chaos (Woodbury, N.Y.)*, 12(4), 1006–1014.

Sasidharan, L., Wu, K.-F., and Menendez, M. (2015). “Exploring the Application of Latent Class Cluster Analysis for Investigating Pedestrian Crash Injury Severities in Switzerland.” *Accident Analysis & Prevention*, 85, 219–228.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., and Milton, J. (2017). “Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances.” *Analytic Methods in Accident Research*, 15, 41–55.

Shaheed, M. S., and Gkritza, K. (2014). “A latent class analysis of single-vehicle motorcycle crash severity outcomes.” *Analytic Methods in Accident Research*, 2, 30–38.

Shang, P., Li, X., and Kamae, S. (2005). “Chaotic Analysis of Traffic Time Series.” *Chaos, Solitons & Fractals*, 25(1), 121–128.

Shang, P., Lu, Y., and Kama, S. (2006). “The Application of Hölder Exponent to Traffic Congestion Warning.” *Physica A: Statistical Mechanics and its Applications*, 370(2), 769–776.

Shang, P., Lu, Y., and Kamae, S. (2008). “Detecting Long-range Correlations of Traffic Time Series with Multifractal Detrended Fluctuation Analysis.” *Chaos, Solitons & Fractals*, 36(1), 82–90.

Shang, P., Wan, M., and Kama, S. (2007). “Fractal Nature of Highway Traffic Data.” *Computers & Mathematics with Applications*, 54(1), 107–116.

Shankar, V., Mannering, F., and Barfield, W. (1995). “Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies.” *Accident Analysis & Prevention*, 27(3), 371–389.

Shefer, D. (1994). “Congestion, Air Pollution, and Road Fatalities in Urban Areas.” *Accident Analysis & Prevention*, 26(4), 501–509.

Shugan, S. M. (2006). “Editorial: Errors in the Variables, Unobserved Heterogeneity, and Other Ways of Hiding Statistical Error.” *Marketing Science*, 25(3), 203–216.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, Boca Raton.

Sivakumar, B. (2000). “Chaos Theory in Hydrology: Important Issues and Interpretations.” *Journal of Hydrology*, 227(1–4), 1–20.

Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. (2010). “Limits of Predictability in Human Mobility.” *Science*, 327(5968), 1018–1021.

Storkey, A. (1996). "The Fractal and Multifractal Nature of Traffic." *Universities' Transport Study Group (UTSG) Conference*, Huddersfield, West Yorkshire, England.

Sun, Z., Hao, P., Ban, X. (Jeff), and Yang, D. (2015). "Trajectory-Based Vehicle Energy/Emissions Estimation for Signalized Arterials using Mobile Sensing Data." *Transportation Research Part D: Transport and Environment*, 34, 27–40.

Taqqu, M. S., Teverovsky, V., and Willinger, W. (1995). "Estimators for Long-Range Dependence: An Empirical Study." *Fractals*, 3, 785–798.

Taylor, M. C., Baruya, A., and Kennedy, J. V. (2002). *The Relationship Between Speed and Accidents of Rural Single-Carriageway Roads*. Transport Research Laboratory, Crowthorne, Berkshire, UK.

Teh, E., Jamson, S., Carsten, O., and Jamson, H. (2014). "Temporal Fluctuations in Driving Demand: The Effect of Traffic Complexity on Subjective Measures of Workload and Driving Performance." *Transportation Research Part F: Traffic Psychology and Behaviour*, 22, 207–217.

Thakur, G. S., Hui, P., and Helmy, A. (2013). "On the Existence of Self-Similarity in Large-Scale Vehicular Networks." *9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, Sardinia, Italy, 1756–1761.

The Cooperative Phenomena Group. (1998). *Coopphenbooklet: Projects and Methods: Self-Affine Fractals*. http://laplace.ucv.cl/Patterns/coopphen_book97/boxes/selfAffineFractals.html. Last accessed on 25-10-2015.

Theofilatos, A., and Yannis, G. (2014). "A Review of the Effect of Traffic and Weather Characteristics on Road Safety." *Accident Analysis & Prevention*, 72, 244–256.

Thiemann, C., Treiber, M., and Kesting, A. (2008). "Estimating Acceleration and Lane-Changing Dynamics Based on NGSIM Trajectory Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2088, 90–101.

Thomas, K., and Dia, H. (2006). "Comparative Evaluation of Freeway Incident Detection Models Using Field Data." *IEEE Proceedings - Intelligent Transport Systems*, 153(3), 230–241.

Tiesyte, D., and Jensen, C. S. (2009). "Assessing the Predictability of Scheduled-Vehicle Travel Times." *Proceedings of the 17th ACM SIGSPATIAL*

International Conference on Advances in Geographic Information Systems, GIS '09, ACM, Seattle, WA, USA, 416–419.

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables." *Econometrica*, 26(1), 24–36.

TomTom. (2017). "Full ranking | TomTom Traffic Index." https://www.tomtom.com/en_gb/trafficindex/list?citySize=LARGE&continent=ALL&country=ALL. Last accessed on 17-07-2017.

Török, J., and Kertész, J. (1996). "The Green Wave Model of Two-dimensional Traffic: Transitions in the Flow Properties and in the Geometry of the Traffic Jam." *Physica A: Statistical Mechanics and its Applications*, 231(4), 515–533.

Train, K. (2000). "Halton Sequences for Mixed Logit." *Department of Economics, University of California, Berkeley*.

Truong, L. T., Kieu, L.-M., and Vu, T. A. (2016). "Spatiotemporal and Random Parameter Panel Data Models of Traffic Crash Fatalities in Vietnam." *Accident Analysis & Prevention*, 94, 153–161.

Tsirigotis, L., Vlahogianni, E. I., and Karlaftis, M. G. (2012). "Does Information on Weather Affect the Performance of Short-Term Traffic Forecasting Models?" *International Journal of Intelligent Transportation Systems Research*, 10(1), 1–10.

Tukey, J. W. (1949). "Comparing Individual Means in the Analysis of Variance." *Biometrics*, 5(2), 99–114.

Turkay, C. (2014). "Visualizing Time Series Predictability." *IEEE VIS 2014 Workshop on Visualization for Predictive Analytics*, Paris, France.

Upadhyay, A. (2016). "Know live Congestion or Future Traffic on Google Map: Desktop and Mobile" IGIS Map. https://www.igismap.com/know-live-congestion-or-future-traffic-on-google-map-desktop-and-mobile/ (Mar. 10, 2018)

Valle, M. A. V., García, G. M., Cohen, I. S., Klaudia Oleschko, L., Ruiz Corral, J. A., and Korvin, G. (2013). "Spatial Variability of the Hurst Exponent for the Daily Scale Rainfall Series in the State of Zacatecas, Mexico." *Journal of Applied Meteorology and Climatology*, 52(12), 2771–2780.

Venkataraman, N., Shankar, V., Ulfarsson, G. F., and Deptuch, D. (2014). "A Heterogeneity-in-Means Count Model for Evaluating the Effects of Interchange Type on Heterogeneous Influences of Interstate Geometrics on Crash Frequencies." *Analytic Methods in Accident Research*, 2, 12–20.

Venkataraman, N., Ulfarsson, G. F., and Shankar, V. N. (2013). "Random Parameter Models of Interstate Crash Frequencies by Severity, Number of Vehicles Involved, Collision and Location Type." *Accident Analysis & Prevention*, 59, 309–318.

Vij, A., Carrel, A., and Walker, J. L. (2013). "Incorporating the Influence of Latent Modal Preferences on Travel Mode Choice Behavior." *Transportation Research Part A: Policy and Practice*, 54, 164–178.

Vojak, R., Levy Vehel, J., and Danech-Pajouh, M. (1994). "Multifractal Description of Road Traffic Structure." *IFAC Proceedings Volumes*, 27(12), 877–882.

Wang, C., and Coifman, B. (2008). "The Effect of Lane-Change Maneuvers on a Simplified Car-Following Theory." *IEEE Transactions on Intelligent Transportation Systems*, 9(3), 523–535.

Wang, C., Quddus, M. A., and Ison, S. G. (2013). "The Effect of Traffic and Road Characteristics on Road Safety: A Review and Future Research Direction." *Safety Science*, 57, 264–275.

Wang, J., Mao, Y., Li, J., Xiong, Z., and Wang, W.-X. (2015). "Predictability of Road Traffic and Congestion in Urban Areas." *PLOS ONE*, 10(4), e0121825.

Wang, J., and Shi, Q. (2013). "Short-Term Traffic Speed Forecasting Hybrid Model Based on Chaos–Wavelet Analysis-Support Vector Machine Theory." *Transportation Research Part C: Emerging Technologies*, Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII), 27, 219–232.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2010). *Statistical and Econometric Methods for Transportation Data Analysis*. CRC press.

Weijermars, W., and Berkum, E. van. (2005). "Analyzing Highway Flow Patterns Using Cluster Analysis." *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, 308–313.

Whittle, P. (1953). "Estimation and Information in Stationary Time Series." *Arkiv för Matematik*, 2(5), 423–434.

Williams, G. P. (1997). *Chaos Theory Tamed*. Joseph Henry Press, Washington, D.C.

Xiong, Y., and Mannering, F. L. (2013). "The Heterogeneous Effects of Guardian Supervision on Adolescent Driver-Injury Severities: A Finite-Mixture Random-Parameters Approach." *Transportation Research Part B: Methodological*, 49, 39–54.

Xiong, Z., Rey, D., Mao, T., Liu, H., Dixit, V. V. & Waller, S. T. A Three-Stage Framework for Motorway Travel Time Prediction. Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, 2014. IEEE, 816-821.

Xiong, Y., Tobias, J. L., and Mannering, F. L. (2014). "The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity." *Transportation Research Part B: Methodological*, 67, 109–128.

Xu, M., and Gao, Z. (2008). "Nonlinear Analysis of Road Traffic Flows in Discrete Dynamical System." *Journal of Computational and Nonlinear Dynamics*, 3(2), 021206-021206-6.

Xu, P., and Huang, H. (2015). "Modeling crash spatial heterogeneity: Random parameter versus geographically weighting." *Accident Analysis & Prevention*, 75, 16–25.

Xu, T., Xu, X., Hu, Y., and Li, X. (2017). "An Entropy-Based Approach for Evaluating Travel Time Predictability Based on Vehicle Trajectory Data." *Entropy*, 19(4), 165.

Yang, X., Lu, Y., and Hao, W. (2017). "Origin-Destination Estimation Using Probe Vehicle Trajectory and Link Counts." *Journal of Advanced Transportation*, Research article, <https://www.hindawi.com/journals/jat/2017/4341532/> (Feb. 13, 2018).

Yasmin, S., Eluru, N., Bhat, C. R., and Tay, R. (2014). "A Latent Segmentation Based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity." *Analytic Methods in Accident Research*, 1, 23–38.

Ye, C., Li, H., and Xu, G. (2014). "An Early Warning Model of Traffic Accidents Based on Fractal Theory." IEEE, Qingdao, China, 2280–2285.

Yeo, H., Jang, K., Skabardonis, A., and Kang, S. (2013). "Impact of Traffic States on Freeway Crash Involvement Rates." *Accident Analysis & Prevention*, 50, 713–723.

Yu, X.-H., and Recker, W. W. (2006). "Stochastic Adaptive Control Model for Traffic Signal Systems." *Transportation Research Part C: Emerging Technologies*, 14(4), 263–282.

Yuan, P., and Lin, X. (2017). "How Long Will the Traffic Flow Time Series Keep Efficacious to Forecast the Future?" *Physica A: Statistical Mechanics and its Applications*, 467, 419–431.

Yue, J., Shang, P., and Dong, K. (2010). "Time-Dependent Hurst Exponent in Traffic Time Series." *2010 IEEE International Conference on Information Theory and Information Security (ICITIS)*, Beijing, China, 744–746.

Yue, Y., Yeh, A. G. O., and Zhuang, Y. (2007). "Prediction Time Horizon and Effectiveness of Real-time Data on Short-term Traffic Predictability." *2007 IEEE Intelligent Transportation Systems Conference*, Seattle, WA, USA, 962–967.

Yulmetyev, R., Gafarov, F., Hänggi, P., Nigmatullin, R., and Kayumov, S. (2001). "Possibility Between Earthquake and Explosion Seismogram Differentiation by Discrete Stochastic Non-Markov Processes and Local Hurst Exponent Analysis." *Physical Review E*, 64(6), 066132.

Zaki, M. H., and Sayed, T. (2013). "A Framework for Automated Road-Users Classification Using Movement Trajectories." *Transportation Research Part C: Emerging Technologies*, 33, 50–73.

Zaksek, T., and Schreckenberg, M. (2016). "Fractal Analysis of Empirical and Simulated Traffic Time Series." *Traffic and Granular Flow'15*, Springer, 435–442.

Zeng, Q., Wen, H., Huang, H., and Abdel-Aty, M. (2017). "A Bayesian Spatial Random Parameters Tobit Model for Analyzing Crash Rates on Roadway Segments." *Accident Analysis & Prevention*, 100, 37–43.

Zhang, G., and Wang, Y. (2011). "Optimizing Minimum and Maximum Green Time Settings for Traffic Actuated Control at Isolated Intersections." *IEEE Transactions on Intelligent Transportation Systems*, 12(1), 164–173.

Zhang, H., Wang, X., Cao, J., Tang, M., and Guo, Y. (2017). "A Hybrid Short-Term Traffic Flow Forecasting Model Based on Time Series Multifractal Characteristics." *Applied Intelligence*, 1–12.

Zhang, L., Garoni, T. M., and de Gier, J. (2013). "A Comparative Study of Macroscopic Fundamental Diagrams of Arterial Road Networks Governed by Adaptive Traffic Signal Systems." *Transportation Research Part B: Methodological*, 49, 1–23.

Zhang, X., and Jarrett, D. F. (1998). "Chaos in a Dynamic Model of Traffic Flows in an Origin-Destination Network." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 8(2), 503–513.

Zhao, F., Guo, H., Wang, W., and Jiang, X. (2015). "Modeling Lateral Interferences Between Motor Vehicles and Non-Motor Vehicles: A Survival Analysis Based

Approach." *Transportation Research Part F: Traffic Psychology and Behaviour*.

Zheng, F., and Van Zuylen, H. (2010). "Uncertainty and Predictability of Urban Link Travel Time." *Transportation Research Record: Journal of the Transportation Research Board*, 2192, 136–146.

Zheng, Z., Ahn, S., Chen, D., and Laval, J. (2011). "Applications of Wavelet Transform for Analysis of Freeway Traffic: Bottlenecks, Transient Traffic, and Traffic Oscillations." *Transportation Research Part B: Methodological*, 45(2), 372–384.

Zheng, Z., Ahn, S., and Monsere, C. M. (2010). "Impact of Traffic Oscillations on Freeway Crash Occurrences." *Accident; Analysis and Prevention*, 42(2), 626–636.

Zielke, B., Bertini, R., and Treiber, M. (2008). "Empirical Measurement of Freeway Oscillation Characteristics: An International Comparison." *Transportation Research Record: Journal of the Transportation Research Board*, 2088, 57–67.

Zou, Y., Zhang, Y., and Lord, D. (2013). "Application of Finite Mixture of Negative Binomial Regression Models with Varying Weight Parameters for Vehicle Crash Data Analysis." *Accident Analysis & Prevention*, 50, 1042–1051.

van Zuylen, H., van Geenhuizen, M., and Nijkamp, P. (1999). "(Un)predictability in Traffic and Transport Decision Making." Transportation Research Record: Journal of the Transportation Research Board, 1685, 21–29.

Appendix A REVIEW OF STUDIES ON THE APPLICATION OF VEHICLE TRAJECTORY DATA IN MIXED TRAFFIC

Chapter-3 presented a review of studies on the application of trajectory data in the empirical analysis of microscopic characteristics. Now, Appendix A presents a review of studies focussing on validation or calibration of car-following and lane-changing models and analysis of motorcycle movement.

A.1 Validating or Calibrating Car-following and Lane-changing Models

In developing countries (and with mixed traffic), a considerable number of studies used real-life trajectory data for either calibrating or validating microsimulation models. Imran (2009) extracted trajectory information using traffic video data from Dhaka, Bangladesh and used it for calibration of a hybrid neuro-fuzzy car-following model for mixed traffic conditions.

Ravishankar and Mathew (2011) analysed data from global positioning system (GPS) equipped vehicles in Mumbai and studied the effect of vehicle type on car-following behaviour. They developed a simulation model by accommodating vehicle-specific parameters in the widely used Gipps's[30] car-following model and found that the model was able to efficiently predict the following behaviour. In a later study, the authors realised the need to represent the complicated relationship between various vehicle-types in a better way, and therefore, applied neural network technique (Mathew and Ravishankar 2012). They applied this technique to the same dataset and found it to be better performing than the Gipp's model.

[30] Gipp's car-following model assumes that the following vehicle will maintain a safe distance with the vehicle in front and will select its speed such that the vehicle can stop safely to avoid a rear-end collision (Gipps 1981).

Mathew et al. (2013) proposed a space discretisation-based simulation framework, where they divided the road width into several small strips. The movement of vehicles was then assumed to be governed by these strips instead of the lanes. The proposed framework accounted for both the multiple leaders and vehicle-type dependent following behaviour, which are common in mixed traffic conditions. By defining smaller strip widths, they were also able to model the continuous lateral movement. They implemented the simulation model in SUMO, an open-source simulator for both midblock and intersections. Further, they validated the model using real data collected on an urban arterial in Mumbai, India and found that their model replicated the mixed traffic conditions in a better way than the lane-based approaches.

Recently, Choudhury and Imran (2016) proposed a latent class approach to model acceleration decisions in mixed traffic. Since multiple leaders could potentially influence the subject driver in mixed traffic, and only the applied accelerations could be observed in the trajectory data, the governing leader was considered unobserved or latent. The model had two components: a random utility based "dynamic" class membership model (latent leader component) and a class-specific acceleration model (acceleration component). They used trajectory data collected from Dhaka, Bangladesh to calibrate the model.

Pandey et al. (2017) developed a CA model to evaluate interactions in heterogeneous traffic conditions. They used trajectory data from Ludhiana city in India for calibration purposes. Using the model, they observed more interactions involving motorcycles compared to other vehicle types.

A.2 Analysis of Motorcycle Movement

Few studies analysed vehicle trajectory data to evaluate motorcycle driving behaviour. Most of these studies were conducted in the South East Asian countries such as Vietnam and Indonesia, where the motorcycle is the predominant mode of transport. Minh and Sano (2007) collected video data and extracted vehicle trajectories at signalised intersections from two cities in Vietnam (Ho Chi Minh City and Hanoi). Then they used the dataset for validation of acceleration and deceleration models that they developed for evaluating motorcycle behaviour. In a later study, the authors further developed a motorcycle manoeuvrability model and applied on the same trajectory datasets (Minh et al. 2012).

Nguyen et al. (2012) extracted vehicle trajectories from video data on two arterial sections in Ho Chi Minh City, Vietnam and used the dataset to validate a model developed to evaluate acceleration and deceleration patterns of a subject motorcycle concerning the speed changes of lead vehicles. In a follow-up study, the authors developed an integrated model to evaluate the traffic conflicts of motorcycles in congested traffic conditions. They found that motorcycles apply sudden braking in dense traffic and there is a high probability of crashes (Nguyen et al. 2014).

Ambarwati et al. (2014) studied the complex interactions in mixed traffic using a porous flow approach, which was initially proposed by (Nair et al. 2011). They calibrated the model using vehicle trajectory data on two arterials in Surabaya City, Indonesia, with a strong focus on the movement of motorcycles. Similarly, Babu et al. (2015) focussed on the movement of motorcycles in mixed traffic and collected traffic video data in Mumbai and extracted trajectory information using the tool developed by (Munigety et al. 2014b). They suggested the use of the social force model and the intelligent driver model for modelling the movement of motorcycles in mixed traffic conditions.

Appendix B SUPPLEMENTARY DATA (FOR CHAPTER-3)

The following are the plots between lateral position vs. instantaneous speed and lateral movement vs. instantaneous speed respectively.

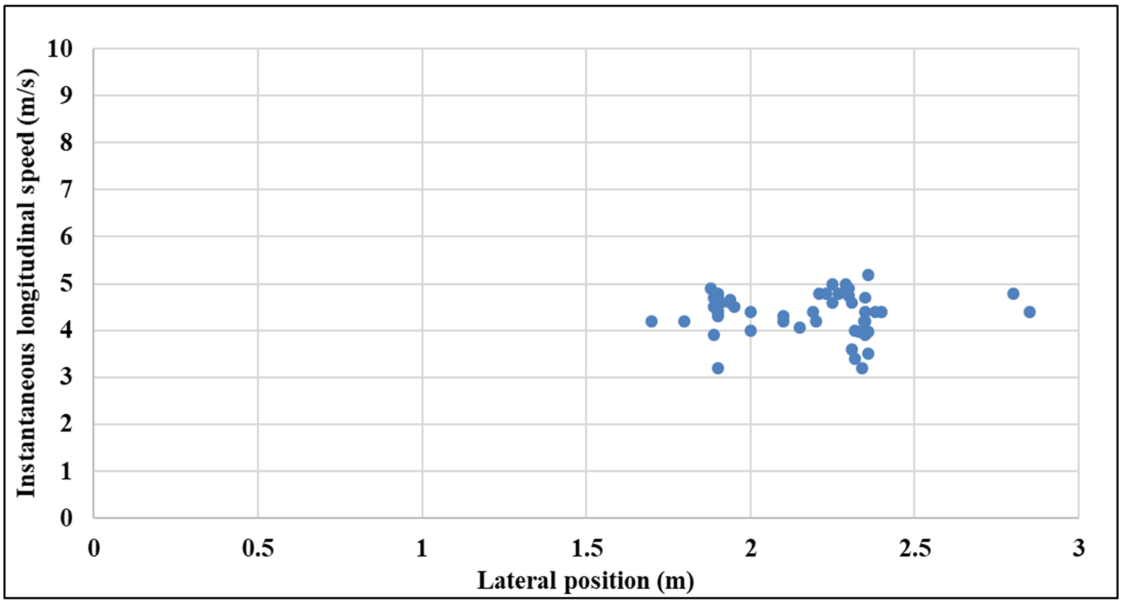


Figure B- 1 Lateral position vs. instantaneous speed

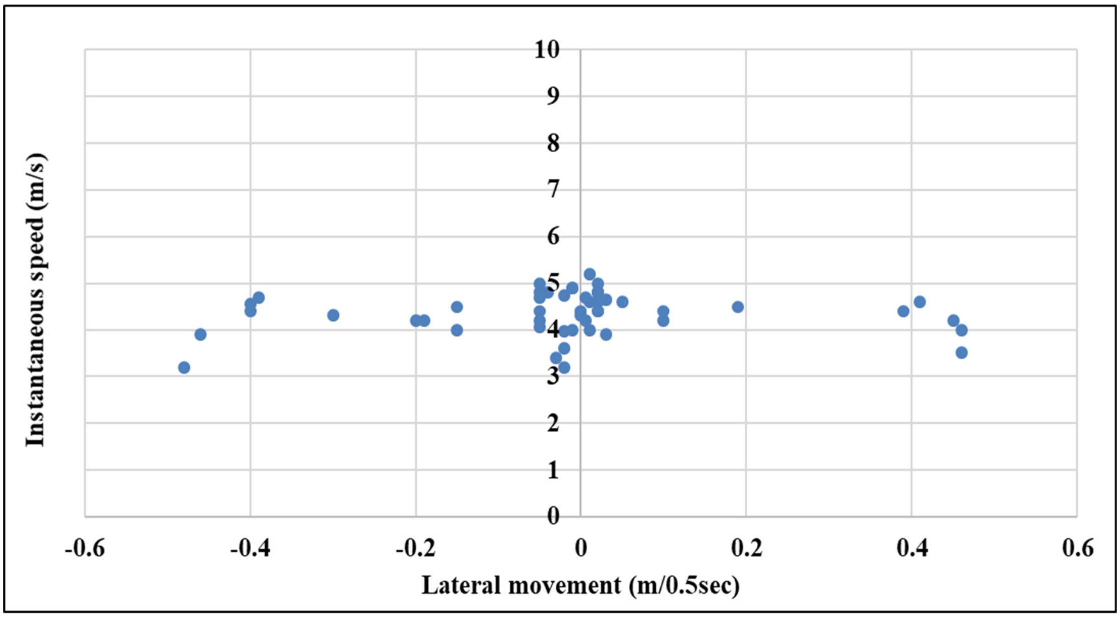


Figure B- 2 Lateral movement vs. instantaneous speed

Appendix C SUPPLEMENTARY DATA (FOR CHAPTERS 4 AND 5)

Chapter-4 presented the analysis of the effect of variables such as the time of the day, the number of lanes, proximity to a ramp, and the presence of horizontal curve on speed and flow fluctuations. Appendix C presents the distributions of H_{flow} and H_{mac_speed} with regard to the above mentioned variables.

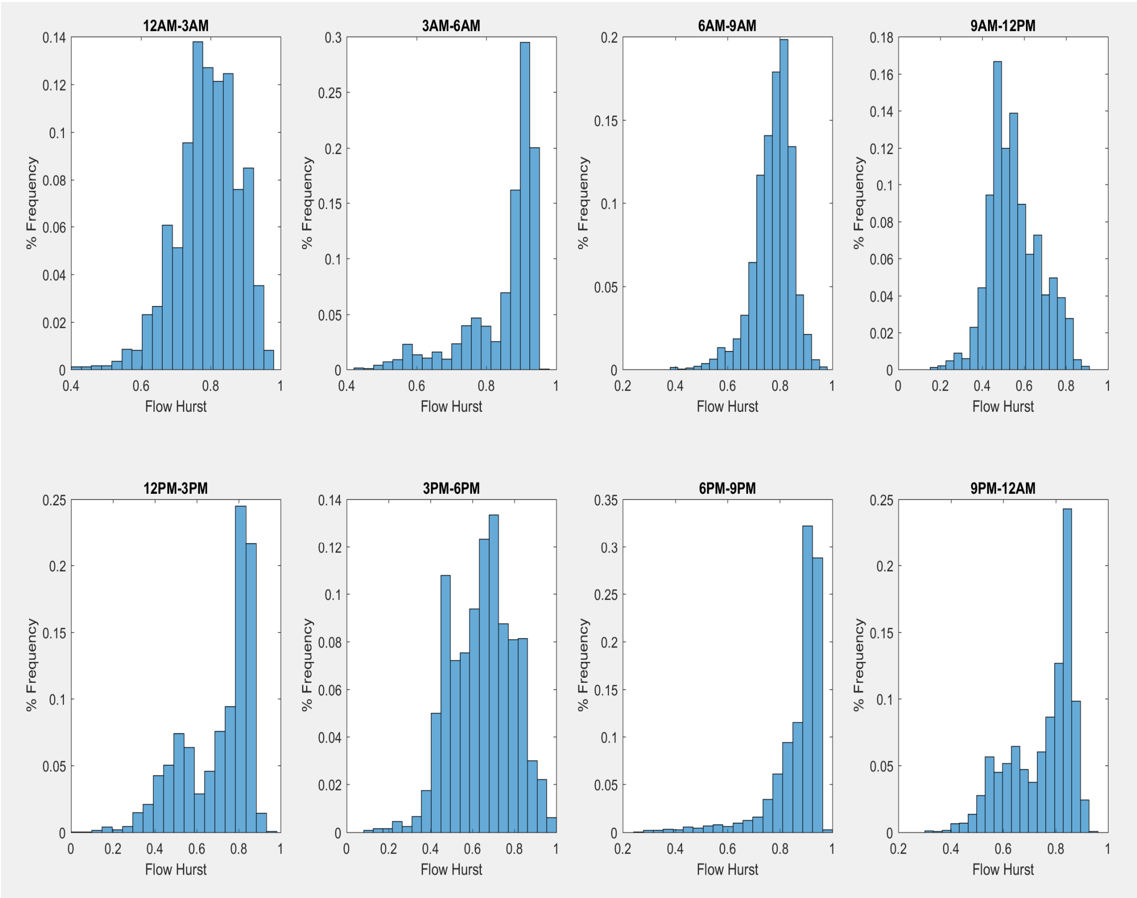


Figure C- 1 Distribution of H_{flow} across different time windows

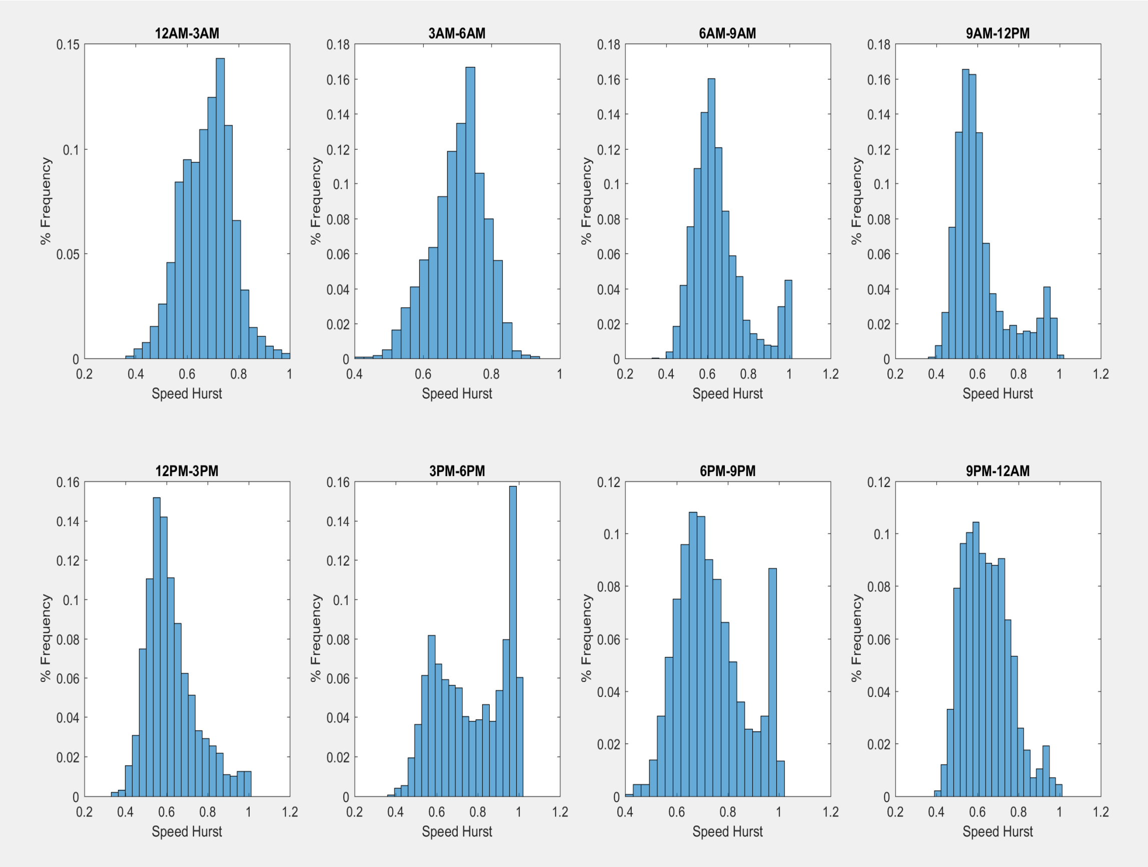


Figure C- 2 Distribution of H_{mac_speed} across different time windows

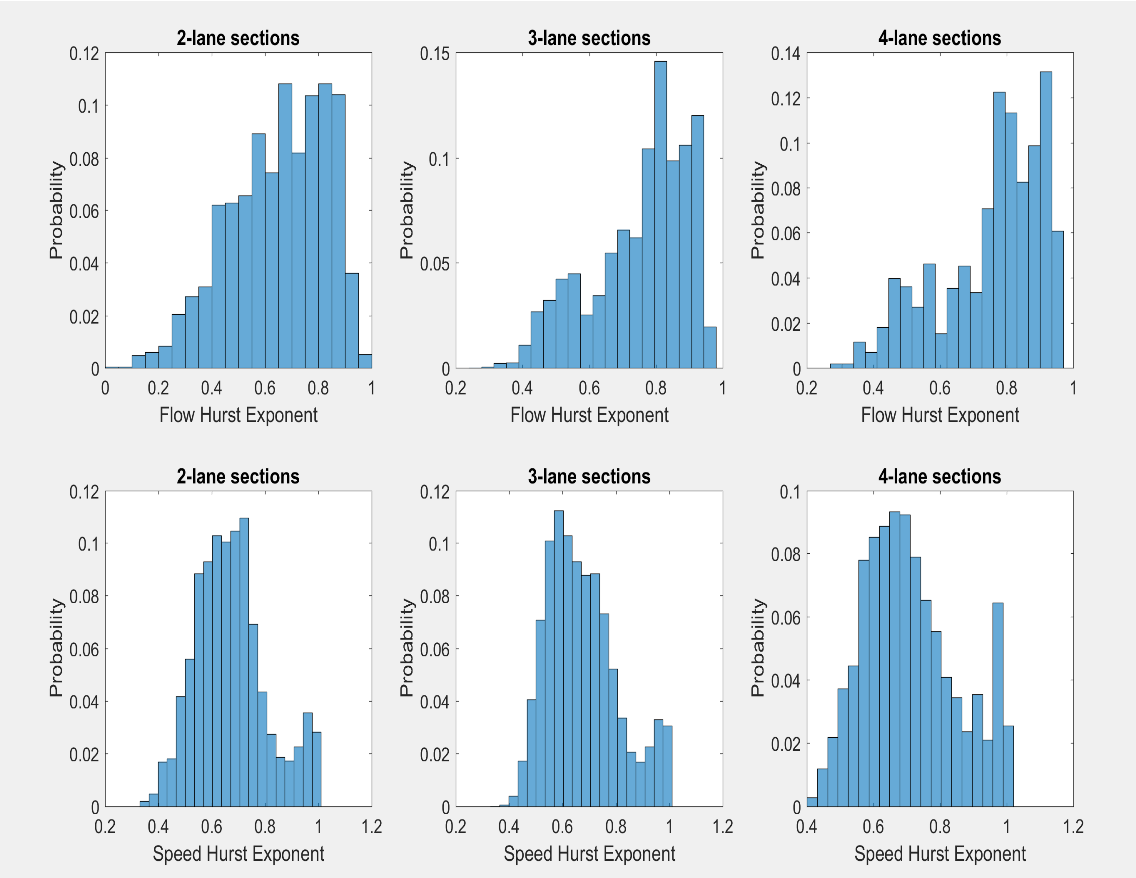


Figure C- 3 Variation of H_{flow} and H_{mac_speed} with the number of lanes

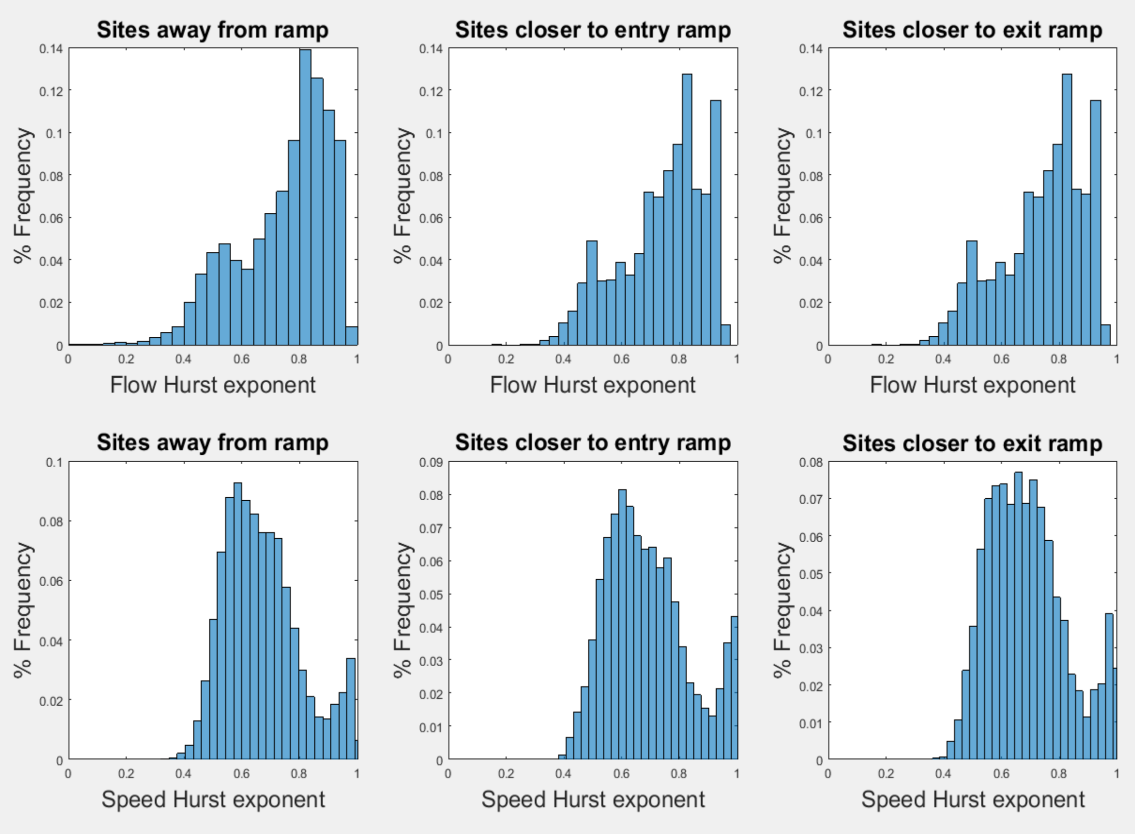


Figure C- 4 Variation of H_{flow} and H_{mac_speed} with the presence of a ramp

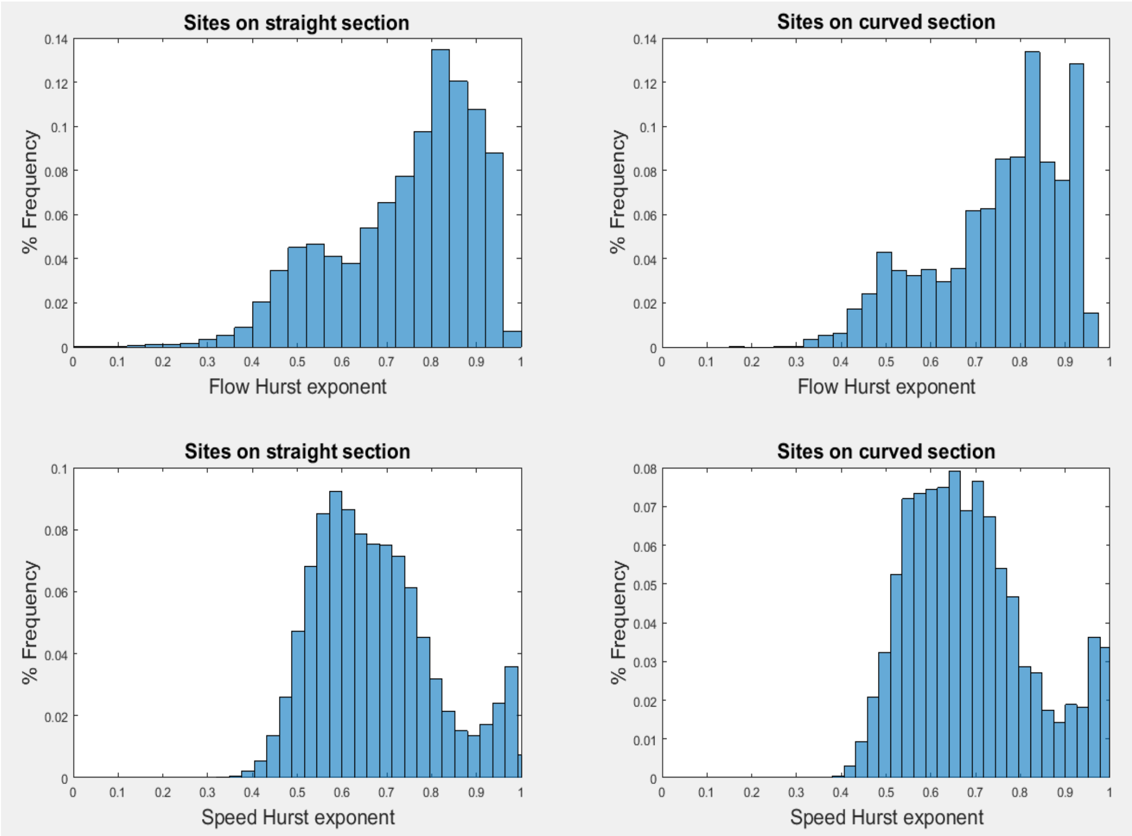


Figure C- 5 Variation of H_{flow} and H_{mac_speed} due to a horizontal curve

Table C-1 Summary statistics of traditional aggregated traffic variables

Variable	No. of time windows	Mean speed	Std. Dev. Speed	Mean Flow	Std. Dev. Flow	Mean Flow per lane	Std. Dev. Flow per lane
Weekend v/s Weekday							
Weekday	13,681	94	11	2226	779	756	230
Weekend	5,579	97	9	1708	615	584	194
Time of the day							
12AM-3AM	2,399	96	8	506	275	173	94
3AM-6AM	2,399	98	7	592	281	201	89
6AM-9AM	2,398	95	11	2449	895	834	282
9AM-12PM	2,435	96	9	2419	799	825	238
12PM-3PM	2,435	96	10	2963	1024	1007	299
3PM-6PM	2,398	87	18	3889	1121	1319	301
6PM-9PM	2,398	95	12	2415	808	821	235
9PM-12AM	2,398	96	9	1392	657	475	218
Number of lanes							
2-lanes	2,488	90	6	1421	762	711	381
3-lanes	15,669	96	6	2085	1224	695	408
4-lanes	1,103	85	8	3366	1735	842	434
Ramps							
Away from ramp	13,424	96	5	2134	1244	725	423
Closer to entry ramp	2,730	92	7	1945	1081	681	378
Closer to exit ramp	3,106	94	6	1887	1052	640	357
Horizontal curve							
On straight section	14,967	96	6	1997	1170	691	404
On curved section	4,293	92	7	2296	1258	752	412

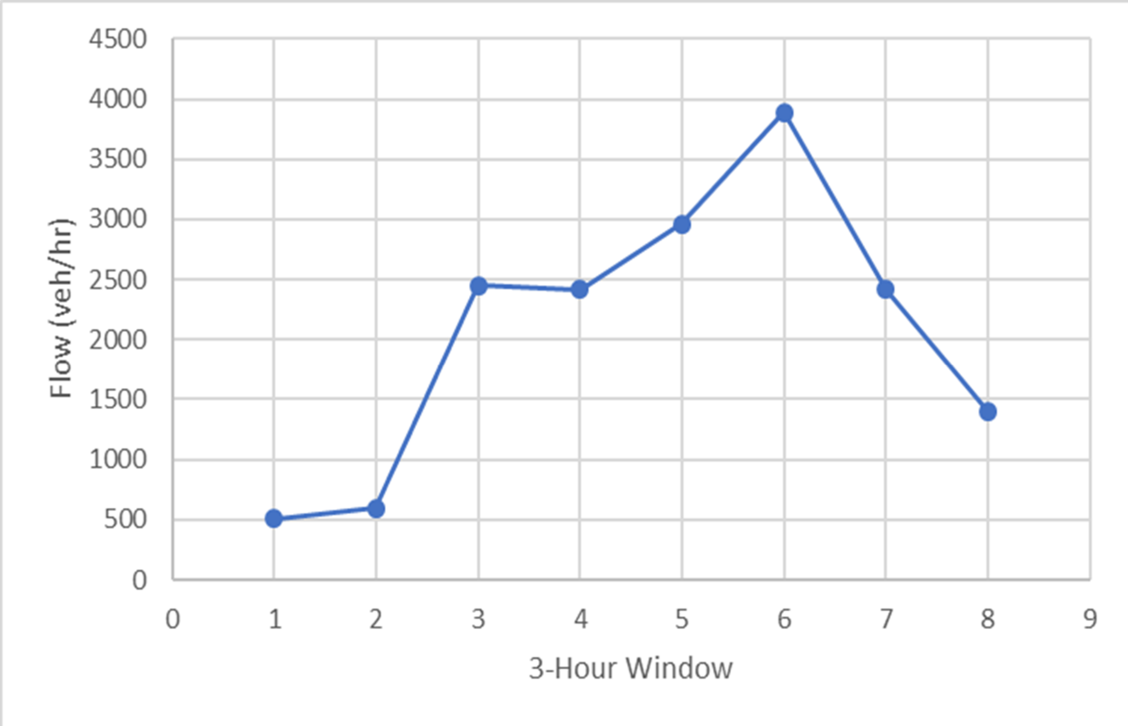


Figure C- 6 Variation of flow with the time of the day

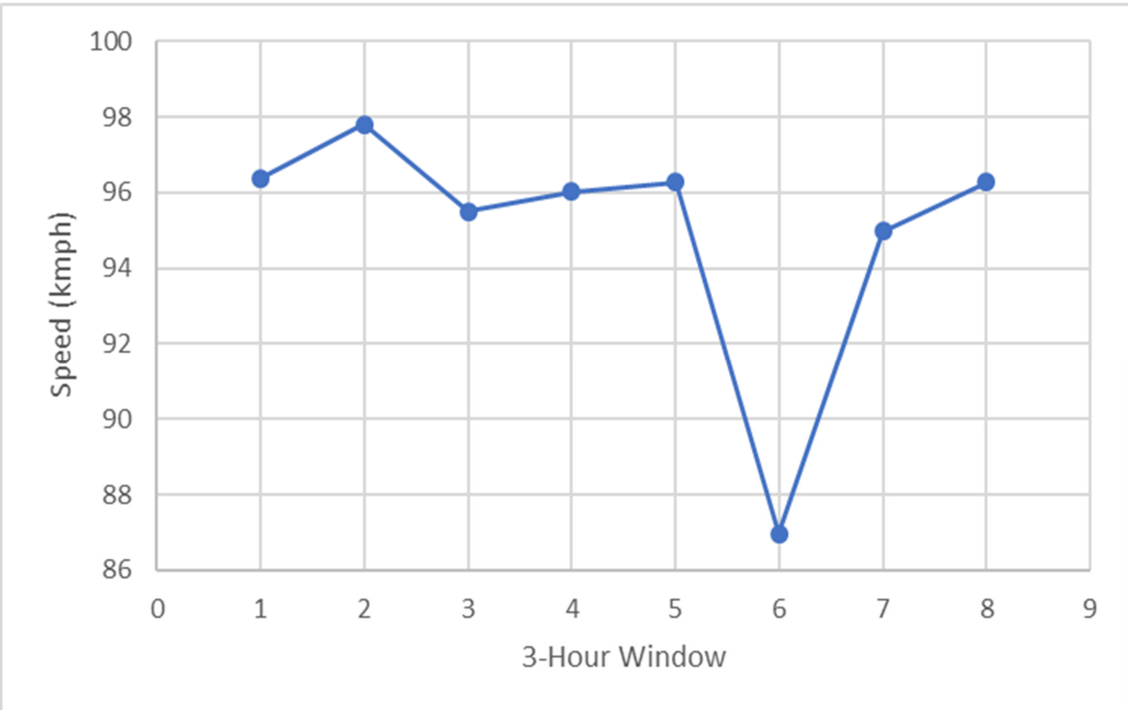


Figure C- 7 Variation of average speed with the time of the day