Spatio-Temporal Clustering for Non-Recurrent Traffic Congestion Detection on Urban Road Networks

Berk Anbaroğlu

Civil, Environmental and Geomatic Engineering

University College London

Supervisors:

Professor Tao Cheng
Professor Benjamin Heydecker

Thesis submitted to University College London for the degree of Doctorate of Philosophy

September 2013

I, Berk Anbaroğlu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature ……………………..

Date ..……………………

ABSTRACT

Non-Recurrent Congestion events (NRCs) frustrate commuters, companies and traffic operators because they cause unexpected delays. Most existing studies consider NRCs to be an outcome of incidents on motorways. The differences between motorways and urban road networks, and the fact that incidents are not the only cause of NRCs, limit the usefulness of existing automatic incident detection methods for identifying NRCs on an urban road network.

This thesis contributes to the literature by developing an NRC detection methodology to support the accurate detection of NRCs on large urban road networks. To achieve this, substantially high Link Journey Time estimates (LJTs) on adjacent links that occur at the same time are clustered. Substantially high LJTs are defined in two different ways: (i) those LJTs that are greater than a threshold, (ii) those LJTs that belong to a statistically significant Space-Time Region (STR). These two different ways of defining the term ‘substantially high LJT’ lead to different NRC detection methods. To evaluate these methods, two novel criteria are proposed. The first criterion, high-confidence episodes, assesses to what extent substantially high LJTs that last for a minimum duration are detected. The second criterion, the Localisation Index, assesses to what extent detected NRCs could be related to incidents.

The proposed NRC detection methodology is tested for London’s urban road network, which consists of 424 links. Different levels of travel demand are analysed in order to establish a complete understanding of the developed methodology. Optimum parameter settings of the two proposed NRC detection methods are determined by sensitivity analysis. Related to the first method, LJTs that are at least 40% higher than their expected values are found to maintain the best balance between the proposed evaluation criteria for detecting NRCs. Related to the second method, it is found that constructing STRs by considering temporal adjacencies rather than spatial adjacencies improves the performance of the method. These findings are applied in real life situations to demonstrate the advantages and limitations of the proposed NRC detection methods. Traffic operation centres could readily start using the proposed NRC detection methodology. In this way, traffic operators could be able to quantify the impact of incidents and develop effective NRC reduction strategies.

To *Pınar* & *Osman Anbaroğlu*

and

Yurdakul Oral Ayden

ACKNOWLEDGMENTS

There is a long list of individuals that made the completion of this thesis possible and very rewarding for me. Firstly, I am very fortunate to have Prof. Tao Cheng and Prof. Benjamin Heydecker as my supervisors. I am indebted for the patience Tao showed throughout my thesis and during countless hours of meetings. From Tao, I learned the extent to which I should be critical of my own research and the importance of explaining my research clearly. From Ben, I learned how to approach a research problem in a systematic way. I could not have achieved a thorough understanding of some of the concepts and methods that I used in this thesis without his guidance. It is an honour for me to complete this thesis under their supervision. I am also thankful to Dr Andy Chow who was like my informal supervisor. Finally, I am thankful to my examiners, Prof. Chris Brunsdon and Prof. Michael Batty, whose comments improved the quality of the thesis as well as expanded my horizon.

I am also grateful to have the members of 'Team Tao' as my colleagues. I am thankful to the post-doctoral researchers, Artemis Skarlatidou and Suzy Moat, who read and commented on the draft of my thesis, and provided invaluable information by sharing their expertise with academia. Jiaqiu Wang helped me with various methodological issues. Lastly, I had the pleasure of collaborating with Ioannis Tsapakis, for which I am especially grateful. I also thank my fellow PhD students, Garavig Taraksaranond, James Haworth, Adel Bolbol Fernandez, and Ed Manley, for their help with various technical issues and comments on my research during/after the pleasant coffee meetings.

I had the opportunity to work with Andy Emmonds, Jonathan Turner and Alex Santacreu from Road Network Performance of Transport for London. They kindly provided the traffic data used within this thesis and shared their expertise. I especially thank Jonathan, who helped me understand the importance of realising the issues that arise in real-life traffic data.

I am also thankful to have to my supervisor in Baskent University, Dr. Emre Sümer, who provided me with the opportunity to perform research in spatial information science. Members of the Geomatics Engineering Department of Hacettepe University, Prof Mustafa Türker and Ali Osman Demirer, supported me in the early stages of my career. Without their encouragement I could not have obtained the scholarship from the Council of Higher Education, which allowed me to study at UCL.

This thesis is mostly a product of interdisciplinary research, and I received help from various academics by email correspondence. Specifically, I offer gratitude to Prof. Martin Kulldorff and Prof. Daniel B. Neill for helping me to clarify some of the concepts in spatial and spatio-temporal scan statistics. Thanks to Dr. Simone Severini and Dr. David Gleich for helping me to understand the computational complexity of some of the graph theoretical concepts that I explored. Thanks to Dr. Varun Chandola for explaining me the difficulties regarding the evaluation of an event detection method in the absence of ground truth data. Thanks to Prof. Douglas A. Reynolds for helping me to clarify some of the results regarding the evaluation part of my thesis. Thanks to the online community within the forums, 'Stack Overflow', 'Cross Validated', 'Matlab' and 'Esri', where I received good quality answers to my technical questions in a short amount of time. Last, but not least, thanks to my friends Valentino Gianuzzi, Santosh Bhattarai and Chris Atkins who proofread parts of my thesis and provided valuable comments regarding the proper usage of English.

Above all, I am very grateful and lucky to have my parents. My mother, Pınar Anbaroğlu, has been so selfless in supporting me at all stages of my life. She listened to my thesis related rehearsals over and over again and helped me greatly in clarifying some parts of my thesis. I must read more poetry to describe her meaning and importance in my life! My father, Osman Anbaroğlu, has shown the importance of finding the right balance in whatever I do in life. He has always been a source of inspiration by showing me the importance of time management. Without my parents support I could not have started or completed this thesis. I am dedicating this thesis to my parents for their endless love and support; and to my late grandfather, Yurdakul Oral Ayden, whom I wish could have been with us longer.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AID	Automatic Incident Detection
ANPR	Automatic Number Plate Recognition
CCTV	Closed Circuit Television
GAM	Geographical Analysis Machine
ITS	Intelligent Transportation Systems
LJT	Link Journey Time
LTIS	London Traffic Information System
MADM	Multi-Attribute Decision Making
NRC	Non-Recurrent Congestion
RC	Recurrent Congestion
STR	Space - Time Region
STN	Spatio-Temporal Neighbourhood
STSS	Space-Time Scan Statistics
TfL	Transport for London
WPM	Weighted Product Model

1. INTRODUCTION

Traffic congestion detection has been extensively researched for decades due to the growing necessity to manage road networks more effectively. However, most of the research on congestion detection focuses on motorways, which are not subject to interruptions due to traffic lights or pedestrian crossings. Uninterrupted traffic flow on motorways allowed scientists to develop physical models to explain the formation and development of traffic congestion as a ‘cluster of densely moving vehicles’ (Kerner and Konhäuser, 1994; Treiber et al., 2000). However, investigation of the characteristics of traffic congestion on urban road networks remained a challenge due to difficulty in modelling irregular interruptions such as traffic lights. Nevertheless, an understanding of traffic congestion on urban road networks is required, as congestion has a substantial impact on society and nature (Beevers and Carslaw, 2005; Goodwin, 2004).

With the advancement and common usage of traffic sensors, vast amounts of traffic data are collected by traffic operation centres. This provides the opportunity to mine the data and obtain useful information. However, the quality of the obtained information usually depends on the quality of the collected traffic data. Moreover, traffic operation centres often rely on *processed* traffic data, such as Link Journey Time estimates (LJTs)[1]. Individual vehicle travel times obtained from vehicle identification sensors have to be processed to estimate an LJT. The quality of the LJTs depends on many factors (e.g. traffic flow, technology of the sensor or weather) which are difficult to model. Therefore, researchers often rely on simulated traffic data to conduct their investigations (Hawas, 2007). However, most of the simulation frameworks cannot handle all the subtleties of real-life situations. Ignoring the issues that arise in a real-life context may lead to theoretically powerful solutions, but those solutions would not necessarily be used in a traffic operation centre (Williams and Guin, 2007). Therefore, it is important to consider the issues that arise in a real-life context, if the developed solution is to be used in a traffic operation centre.

Traffic congestion is classified into two types: Recurrent Congestion (RC) and Non-Recurrent Congestion (NRC) (Ozbay and Kachroo, 1999, p.1; Dowling et al., 2004;

---

[1] A link is defined as a subset of a road network on which traffic data are collected and it may contain any number of traffic junctions with or without traffic signals. Estimation of an LJT is discussed in detail in section 2.2.

OECD, 2007, p.14; Varaiya, 2007). RC is the congestion observed at morning or afternoon peak periods, and exhibits a daily pattern. Location and duration of an RC event is usually known by regular commuters and traffic operators. It is mainly caused by excess travel demand, inadequate traffic capacity or poor signal control (Han and May, 1989). On the other hand, NRC can occur at any time of day, and its location and duration usually depends on the local conditions of the road network, as well as travel demand and traffic capacity. An NRC event is mainly caused by unexpected events like traffic accidents or vehicle breakdowns. However, it can also occur due to planned engineering works, special events (e.g. football matches or concerts) or inclement weather (FHWA, 2012a; Kwon et al., 2006).

1.1. Motivation

The major source of travel time variability is considered to be NRC events (Noland and Polak, 2002). Occurrence of an NRC event frustrates commuters and traffic operators. FHWA (2012b) explains this by stating "*Most travellers are less tolerant of unexpected delays because they cause travellers to be late for work or important meetings, miss appointments, or incur extra childcare fees. Shippers that face unexpected delay may lose money and disrupt just-in-time delivery and manufacturing processes*". Therefore, research is required to gain a better understanding of the causes of NRC events and how they are related to incidents. Such an understanding will provide valuable information for traffic operation centres on (Hallenbeck et al., 2003b):

- Development of NRC mitigation strategies based on their cause;
- Understanding how much NRC is caused by particular type of incidents; and the development of incident response strategies accordingly.
- Effective management of planned events like engineering works or social/cultural/sports fixtures.

Most of the existing research focused on determining the total amount of annual congestion that is due to recurrent or non-recurrent congestion (Dowling et al., 2004; Skabardonis et al., 2003). Therefore, existing research are solely developed to measure NRC, which limits their usefulness to detect NRC events due to the following two reasons:

1. Substantial amount of data including travel demand, traffic capacity, incidents and weather is required to measure NRC. Usually the required data cannot be feasibly collected over a large road network. Consequently, NRC is measured on a small subset of a road network.

However, analysis of large road networks is important, as a growing number of traffic operation centres manage such networks, especially in an urban road network context. Therefore, an NRC detection method should be able to operate on a large urban road network to detect NRC events[2] on an analysed time interval.

2. Existing methodologies provide a generic estimation of the total amount of NRC, which does not allow a traffic operator to study the impact of each NRC event separately.

However, it is also important to identify and quantify each NRC event separately so a traffic operator can hypothesise the cause of an observed NRC. If, for instance, an incident is found to cause an NRC, then the impact of that incident could be characterised with that NRC.

These two uses are necessary in a traffic-operation centre, and this need is illustrated by a screenshot taken from the London Congestion Analysis Project in Figure 1-1.

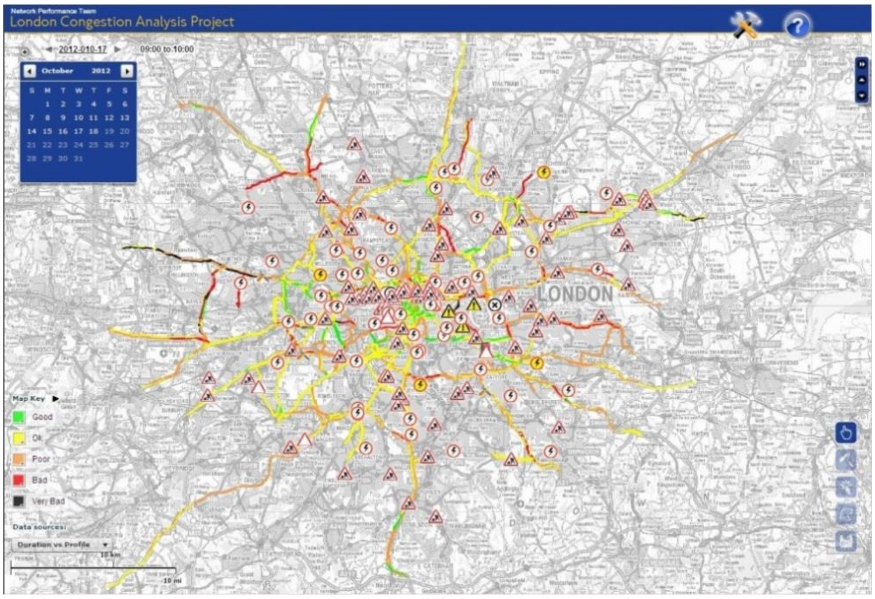


Figure 1-1 A screenshot that illustrates congestion levels and incidents

[2] Hereafter, the term 'NRC' is used interchangeably to refer to the name of the phenomenon and the realisation of it (i.e. an NRC event). The term 'NRCs' refers to NRC events.

The links are coloured based on the relationship between the estimated LJTs and their expected values during the analysed time interval stated on the top left of the screenshot. A link is represented by any one of five colours each indicating a different congestion level, which are shown on the bottom left of the screenshot. Incidents that took place during the analysed time interval are denoted with a symbol that resembles the incident type.

The two uses of an NRC detection method can be observed by inspecting the screenshot illustrated in Figure 1-1. The road network is very large, which necessitates effective methods to automatically detect NRCs. Second, it is important to understand why the detected NRCs happened, which could be achieved by relating the detected NRCs with the reported incidents.

An important challenge should further be addressed by an NRC detection methodology: evaluation of the performance of an NRC detection method. Absence of ground truth data (i.e. lack of information regarding whether a traffic observation belongs to an NRC) makes the performance evaluation of an NRC detection method a difficult task. This challenge could be addressed by using an incident data set, as traffic incidents[3] are regarded as the main cause of an NRC (Cheu and Ritchie, 1995; Thomas and van Berkum, 2009; Yuan and Cheu, 2003). However, it has also been observed that many large delays occur that cannot be explained via the incident data set (Hallenbeck et al., 2003b). As a result, more research is required to understand the reasons that prevent the use of incident data sets to evaluate the detected NRCs.

Last, but not least, urban road networks are heterogeneous in two aspects. Firstly, link lengths vary from hundreds of metres to tens of kilometres. Handling day-to-day variations on short links, and at the same time capturing the variations on longer links which may be due to an NRC, is a challenging task. Secondly, the quality of the traffic data estimates varies spatially depending on, amongst others, the sensors being used and the positioning of the sensors. Reliance on few vehicles data to estimate traffic data may result in sampling errors. Furthermore, the quality of traffic data depends also on the traffic situation. For example, the dense movement of vehicles during congestion may reduce the quality of estimated traffic data. These two reasons lead to a heterogeneous

[3] An 'incident' is a transient event that temporarily decreases the capacity of a link in which the incident occurred. Examples of incidents include traffic accidents, engineering works or special events.

road network, which makes it difficult to detect NRCs, but results in a very interesting research topic.

1.2. Research Aims and Objectives

The aim of this thesis is '*to support the accurate detection of NRCs on a large urban road network"*. In order to achieve this aim, the objectives are stated as follows:

- Develop an NRC detection methodology for a large urban road network.
- Quantify the impact of an NRC event.
- Understand the difficulties in evaluating the effectiveness of an NRC detection method by using an incident data set.
- Determine criteria to evaluate different NRC detection methods that consider high LJT as an indicator of NRC.

1.3. Research Framework

The research framework is structured around the following two main questions regarding an urban road network consisting of hundreds of links:

- How could NRCs be detected accurately on such a network?
- How could the performance of an NRC detection method be evaluated?

Regarding the first question, it has been observed that spatio-temporal clustering has been previously used within the context of traffic congestion detection. Theoretical research on this topic explains how a small perturbation in a dense road, which is characterised by a cluster of vehicles, can cause traffic congestion. This type of congestion is known as 'phantom jams' or traffic congestion without a specific cause, and is common on motorways (Helbing, 2001; Kerner and Konhäuser, 1994; Sugiyama et al., 2008). Empirical research also demonstrates the effectiveness of spatio-temporal clustering to detect traffic congestion (Anbaroglu and Cheng, 2011; Li et al., 2007; Ying et al., 2009). However, these empirical approaches have two important limitations. Firstly, they are not able to differentiate between RC and NRC. Secondly, their application to an urban road network consisting of hundreds of links is computationally infeasible.

Spatio-temporal clustering has been used in different contexts with two different meanings: grouping of observations based on a similarity measure and detection of statistically significant regions. The former strategy relies on the definition of a similarity measure (usually expressed in terms of a distance function) that groups similar observations in clusters (Dubes and Jain, 1976; Han and Kamber, 2006, p.383). The latter strategy conducts a hypothesis testing procedure and determines whether observed data could have arisen by chance alone (Besag and Newell, 1991). This thesis proposes a method to detect NRCs for each of these strategies. The first method clusters spatio-temporally overlapping episodes, where an episode is a maximal interval on a link during which all the estimated LJTs are excessive. The second method develops a modified version of expectation-based Space-Time Scan Statistics (STSS), the state-of-the-art method for detecting statistically significant clusters. Different NRC detection methods would yield to different NRCs, and it is necessary to evaluate the effectiveness of these different methods.

Evaluating the performance of an NRC detection method is a difficult task, because LJTs do not contain information on whether or not they are estimated in an NRC situation. In other words, it is not known which LJT belongs to an NRC. It is therefore necessary to investigate the possible ways of evaluating the performance of an NRC detection method, so that different NRC detection methods can be compared to determine the best performing method.

1.4. Organisation of the Thesis

The thesis is organised as illustrated in Figure 1-2 and consists of six chapters and three appendices.

Chapter 2 is the literature review on NRC detection and space-time clustering. Section 2.1 is the background to NRC detection. It describes the differences between motorways and urban road networks, traffic data types and traffic congestion in the macroscopic fundamental diagram. Section 2.2 focuses on LJT, which is the main traffic data type used throughout this thesis. It briefly describes the estimation process of an LJT, and distribution of LJTs. Section 2.3 describes the existing methodology on NRC detection. As the literature on NRC detection is very limited, it explores whether the knowledge obtained from a relevant research area, incident detection, might be useful for NRC detection. The literature relevant to the methodology of this thesis, spatio-temporal clustering, is introduced and discussed in detail in section 2.4. Finally, section 2.5 investigates the different ways to evaluate the performance of an NRC detection method.

Chapter 3 proposes two novel NRC detection methods. The first method builds upon similarity based clustering, and detects NRC by clustering excessive LJTs. This method is discussed in section 3.2. The second method detects NRCs based on STSS. A novel expectation-based STSS that is suitable for NRC detection is developed in section 3.3. The detected NRCs are represented in their evolution in order to quantify the impact of an NRC. Subsection 3.2.3 describes the two characteristics of an NRC, lifetime and severity, and how they can be estimated.

Chapter 4 investigates the possible ways to evaluate the performance of NRC detection methods. Section 4.1 introduces the two data sets that are used throughout this thesis, which are LJT and incident data sets. Section 4.2 identifies five reasons why the incident data set is not suitable to evaluate the performance of NRC detection methods that are based on LJT data. Finally, section 4.3 proposes two performance evaluation criteria to compare different NRC detection methods.

Chapter 5 analyses the performance of the proposed NRC detection methods on London's urban road network, which consists of 424 links. In section 5.1 the distribution of LJTs is determined based on raw and cleaned LJTs. In section 5.2 the effect of weekdays on estimated LJTs, the effective analysis period and the adjacency of a special kind of topological relationship between links, namely anti-parallel links are determined. In section 5.3 the two NRC detection methods proposed in sections 3.2 and 3.3 are compared. The experiments are conducted on three types of dates from 2010: bank holidays, normal days and tube strikes, in order to demonstrate the effectiveness of both NRC detection methods in situations of low, normal and high travel demand days respectively. Section 5.4 demonstrates the effectiveness of the proposed NRC detection methods on different situations. Firstly, link-based analyses demonstrate the advantages and limitations of NRC detection methods under different scenarios. Secondly, the impact of an incident is determined by different NRC detection methods.

Chapter 6 concludes the thesis by providing an overview of the study. The main contributions of the thesis are provided in section 6.2, where the contributions are linked with the objectives stated in this chapter. A critical assessment of the methodology is provided in section 6.3. This assessment includes both the NRC detection methods and the criteria that are proposed to evaluate the detected NRCs. Finally, future research directions are provided in section 6.4.

There are four appendices:

- Appendix A discusses the wide array of research using the concepts of spatial and spatio-temporal neighbourhoods. This research area is closely with defining the adjacency matrix of a road network, which is an important input to the proposed NRC detection methodology.
- Appendix B demonstrates the importance of the quality of the estimated LJTs. It consists of two empirical observations regarding this issue; firstly, ignoring the sampling errors may lead to misleading conclusions. Secondly, the quality of the estimated LJTs depends on, amongst other reasons, congestion levels.
- Appendix C details the additional results that are not included within the thesis in order to simplify the main context and to improve the presentation of the reported results.
- Appendix D lists the outcomes of this thesis. It lists the published work, honours/awards and professional activities.

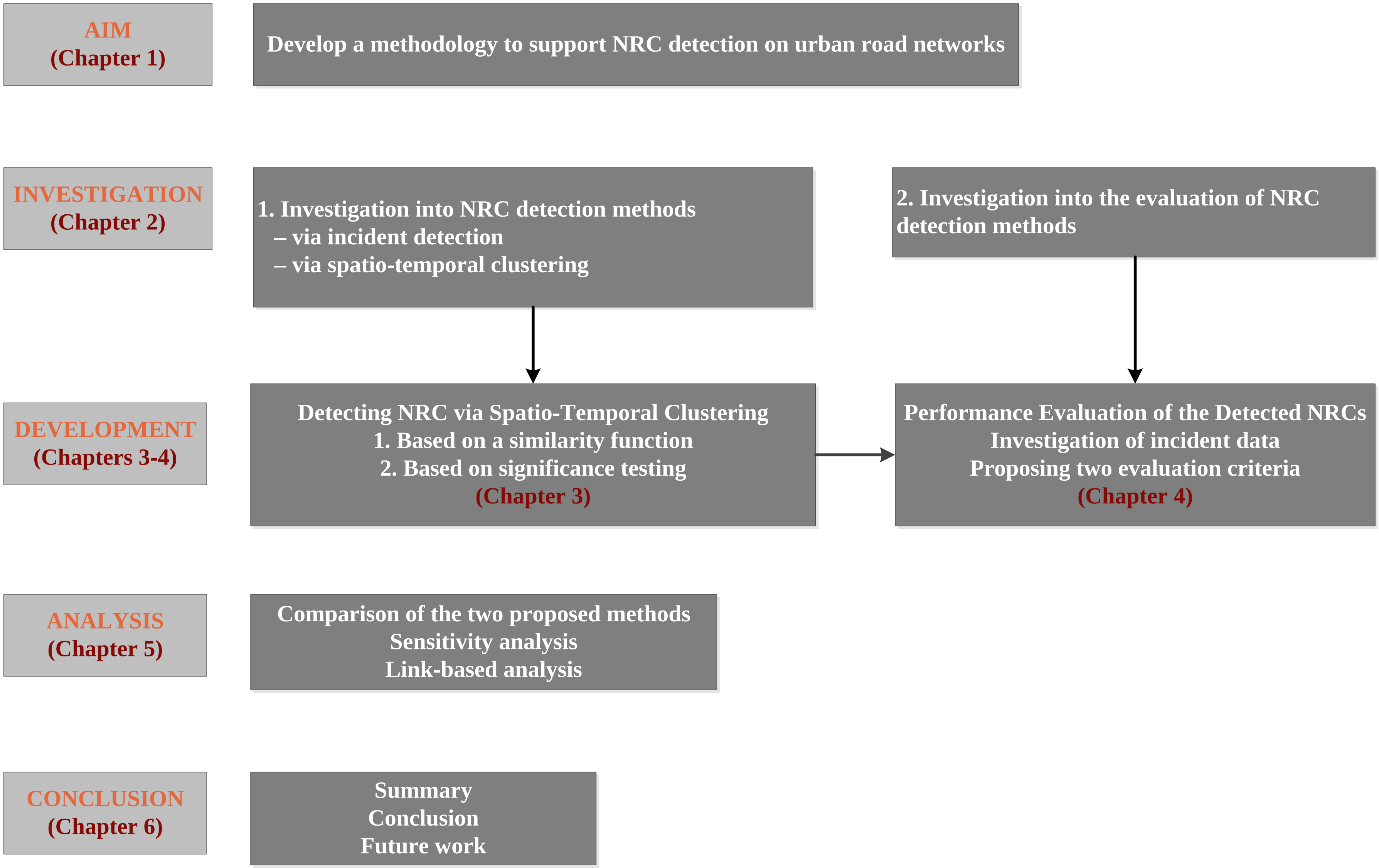


Figure 1-2 Organisation of the thesis

2. NRC DETECTION AND SPATIO-TEMPORAL CLUSTERING

Road network congestion has been observed since 45 BC. One of the first preventive measures against congestion was taken by Julius Caesar, who only allowed selected citizens' carriages to enter Rome between 06:00 and 16:00 (Black, 2010, p.14). Even though the means of land transportation has changed from horses to vehicles, research on traffic congestion is still an active area. This is because; traffic congestion has substantially impacts on society (Eddington, 2006; Goodwin, 2004) and the environment (Beevers and Carslaw, 2005). These negative impacts have become more noticeable, as a finding from a recent survey suggests that congestion has *become worse* according to the perception of the majority of traffic operators (Bertini, 2006). Therefore, management and reduction of traffic congestion is a focus of most developed countries' transportation related polices (DfT, 2012; USDOT, 2006).

There are three main strategies to reduce traffic congestion:

- Increasing the capacity of a road network is the traditional approach to reducing traffic congestion. The solutions included in this strategy are building new roads, widening existing roads or using reversible lanes. However, implementation of such solutions is usually cost prohibitive and requires elaborate planning, especially in urban road networks, which makes this solution unsustainable (National Research Council, 1994).
- Managing travel demand by introducing policies like congestion charges, increasing travel fares or rewarding commuters who prefer to travel at designated times (Ben-Elia and Ettema, 2009). Although some of these policies, such as congestion charges, decrease congestion levels with immediate effect, their long-term effects are unquantified.
- Managing the road network more effectively via Intelligent Transportation Systems (ITS) (Chu et al., 2004). The usage of ITS, from both commuters and traffic operation centres' points of view, is becoming more frequent due to the collection of vast amounts of traffic related data (Zhang et al., 2011). Commuters benefit from ITS in two phases: pre-trip and in-trip. In the pre-trip phase, commuters plan their journey according to the information gathered using online applications, and in the in-trip phase commuters can change their plans based on the most current traffic

information (Sansò and Milot, 1999). Traffic operation centres benefit from ITS via applications including dynamic traffic light control or automatic incident detection. Such systems can be useful in reducing both RC and NRC. Since most of the incidents are associated with an NRC, this thesis focuses on the third strategy, especially considering the desire to reduce congestion by providing a better understanding of NRC for the benefit of traffic operation centres.

This chapter reviews the literature on NRC detection and spatio-temporal clustering and is divided into two parts. The first part illustrates the state of the art in NRC detection by providing background information in section 2.1, introducing the Link Journey Time (LJT) traffic data type in section 2.2 and describing the existing methods used to detect NRC in section 2.3. The second part illustrates commonly used methods in spatial and spatio-temporal clustering, as this is the methodology proposed to detect NRCs. The two main spatio-temporal clustering strategies, similarity based clustering and significance testing based clustering, are discussed in sections 2.4.1 and 2.4.2 respectively. Thereon, section 2.5 discusses the evaluation of the detected NRCs. Finally, section 2.6 concludes this chapter by summarising the literature review.

2.1. Background to Urban Road Traffic Congestion

2.1.1. Types of Road Network

Road networks can be grouped according to their purpose, capacity or maximum legally allowed speed. There are different classifications of road types; for example, Pline (1992, p.314) categorises roads into 11 types. However, the level of detail reached by such a scheme is mostly useful to road constructors. On the other hand, OECD (2007, p.37) uses a four-level hierarchy, which is usually sufficient to provide the details that a commuter and traffic operator would need. In this four-level classification, road categories are referred to as motorways (also known as ‘freeways’ or ‘autobahn’), arterial (also known as ‘urban’), collectors and local networks. Such a classification is illustrated in Figure 2-1.

Figure 2-1 Classification of roads according to their purpose

Source: OECD, 2007 with permission

The highest speeds and traffic flows are achieved on motorways as they are designed to connect cities. Within a city, arterials are the main roads which connect boroughs, and collector roads are used to connect the neighbourhoods within a borough; finally, local roads are used to find a specific address.

In the literature, however, road networks are usually divided into two main types: motorways and urban (or arterial), as traffic operation centres collect traffic data mainly on these types of roads. The main difference between these two road types is their dynamical characteristics: vehicle movement on a motorway is uninterrupted, as there are no traffic lights or pedestrian crossings, which makes motorways suitable to apply fluid dynamic models to explain traffic congestion formation (Helbing, 2001). On the other hand, the movement on an urban road network is interrupted due to the presence of traffic signals, side street traffic, on-street parking or pedestrian crossings.

Although traffic junctions can also be modelled, the developed models usually assume the conservation of traffic flow. Due to side-street parking or vehicles entering/exiting car parks, traffic flow is less conserved in urban road networks than motorways. Thus, the models developed for junctions cannot be applied to a large network consisting of hundreds of links (Helbing et al., 2007). Therefore, spatial correlation amongst the traffic data collected on motorways on adjacent sensors is higher than urban road networks (Hawas, 2007; Yang et al., 2009).

Another factor that may be attributed to the spatial correlation amongst the traffic data on adjacent sensors is the positioning of traffic sensors. On motorways, traffic sensors are usually located frequently and homogeneously, every 500 metres to one kilometre (Dowling et al., 2004), whereas the location of traffic sensors on an urban road network is much more heterogeneous. Hence, understanding the occurrence of traffic congestion on an urban road networks is more difficult compared to motorways. OECD (2007, p.87) stresses that '*congestion on uninterrupted flow facilities such as motorways does not occur in the same manner nor for the same proximate causes as congestion arising on interrupted flow facilities such as those found in dense urban cores*'. Furthermore, as vehicles move in a denser fashion in urban road networks, their detection and differentiation from other vehicles becomes more complex due to occlusion. In this way, the quality of traffic data varies substantially on an urban road network.

Another difference between the two types of road network is that a motorway has a linear topology, whereas urban road networks are in the form of a graph. Motorways are designed to connect cities which are sparse in space, whereas urban road networks connect districts of a city. Therefore, there are few plausible routes on a motorway to connect an origin and a destination, whereas in an urban road network many more alternative routes can be selected to connect two locations. This becomes an important issue when determining the downstream – upstream link relationship[4] in an urban road network, which is trivial in motorways due to the links' linear structure. In addition, longer stretches of roads in motorways allow a traffic operator to surveil much longer distances than in an urban road network. This distinction is illustrated in Figure 2-2, where a screenshot of a CCTV operating on a motorway and urban road network is illustrated.

(a)

(b)

Figure 2-2 Surveillance of a motorway segment (a) and an urban road segment (b)

[4] Suppose that traffic flows from link A to link B. Then, link A is referred to as being upstream of link B; and link B is downstream of link A.

Due to these differences, analysis of urban road networks is often regarded as more complex than motorways (Ozbay and Kachroo, 1999, p.79; Lee et al., 2011). It has been stated that a different theoretical consideration from that used for the well-developed motorway networks is required to understand traffic flows in an urban road network context by considering all these aforementioned differences (Kerner, 2004, p.582). The differences between a motorway and an urban road network are summarised in Table 2-1.

Table 2-1 Differences between urban road networks and motorways

		Urban Road Network	**Motorway**
Traffic flow		Interrupted. Not conserved.	Uninterrupted. Conserved.
Data Resolution	**Spatial**	Heterogeneous. Link lengths vary substantially.	Homogeneous. Similar link lengths.
	Temporal	In minutes.	In seconds.
Geometry		Graph	Linear

2.1.2. Traffic Data Types

In order to automatically detect congestion, developed methodologies should rely on traffic data collected in a real-life context. Traffic data are collected through two main types of sensors. The first is spatially fixed sensors which are embedded on a road network and these are the main types of sensors used in traffic operation centres. This is because spatially fixed sensors can be located on a road network based on the operational purposes of a traffic operation centre. The commonly used spatially fixed sensors include Automatic Number Plate Recognition (ANPR) cameras, Closed Circuit Television (CCTV) cameras and inductive loops. The second is spatially mobile sensors which are attached to commuters or vehicles (also called as ‘probe vehicles’).

Spatial coverage is another important distinction between traffic data types, which can be divided into two groups: measurements over a short section and measurements along a length of a road (Hall, 2001). Measurements over a short section collect traffic data on a short section of a road. Traffic data obtained from inductive loops are examples of such traffic measurements, as inductive loops operate on a small section of road. On the other hand, measurements along a length of road collect traffic data on roads that are usually longer than 500 metres, which is the main reason this type of measurement is suitable for a network-wide analysis. Link Journey Time (LJT) is a prominent example of measurements along a length of road and is therefore suitable for a network-wide analysis. How an LJT observation is estimated is discussed in section 2.2.

Remotely sensed visual data are collected via surveillance cameras. The main traffic sensors used for surveillance are CCTV cameras that are located based on the operational purposes of traffic operation centres; usually on a high pole (e.g. traffic lights). Visual data are used for automated methods to detect congestion, and also by traffic operators for surveillance. Automated methods use computer vision algorithms which can be used for various purposes; a comprehensive review is presented in Buch et al. (2011). Traffic operators rely on visual data as it is one of the fastest and reliable ways to surveil a road network. Remotely sensed visual data are commonly used to detect incidents and traffic violations, and for surveillance.

The main drawback of remotely sensed visual data, however, is its limited geographical coverage, which further depends on the viewing angle of the cameras. Unless surveillance takes place on a motorway network, where roads can be linear for several kilometres, CCTV cameras often have a limited area of surveillance in urban road networks. Furthermore, weather may also substantially affect the quality of remotely sensed visual data (Shehata et al., 2008; Weil et al., 1998).

Traffic flow measures the number of vehicles that pass a given location at a fixed temporal interval. It is usually measured in vehicles per hour per lane. The main traffic sensor used to measure traffic flow is inductive loops, which are buried under a road surface. The principle of inductive loops relies on the law of induction, where a vehicle’s metal surface passing through the inductive loop alters the electrical flow within the loop detector. Therefore, an inductive loop has two states: occupied and not occupied. When a vehicle is present on top of the inductive loop, the loop is occupied due to the magnetic interaction between the loop and vehicle’s metal surface.

The importance of traffic flow arises from the fact that it is used to determine a road’s *capacity*, which is defined as the maximum flow rate. On the other hand, the main issue regarding the usage of loop detectors is the lane/road closure that is required to install an inductive loop.

The ***speed*** of a vehicle is defined as the distance travelled in a fixed time interval, and it is usually measured in miles per hour. A vehicle's speed can be measured using several sensors, including inductive loops, radar, remotely sensed visual data or via a location aware device that is attached to the commuter or the vehicle. Speed can also be measured as an aggregate measure of vehicles. There are two types of such aggregate speed measures: time-mean speed and space-mean speed. Time-mean speed measures the average instantaneous speed of the vehicles passing through the speed detector (e.g. inductive loop). Thus, time-mean speed does not take into account the variations in journey time of different vehicles. On the other hand, space-mean speed takes space as the reference, and measures speed based on the average journey time of the vehicles (i.e. firstly averaging the journey times, and then converting this value to speed).

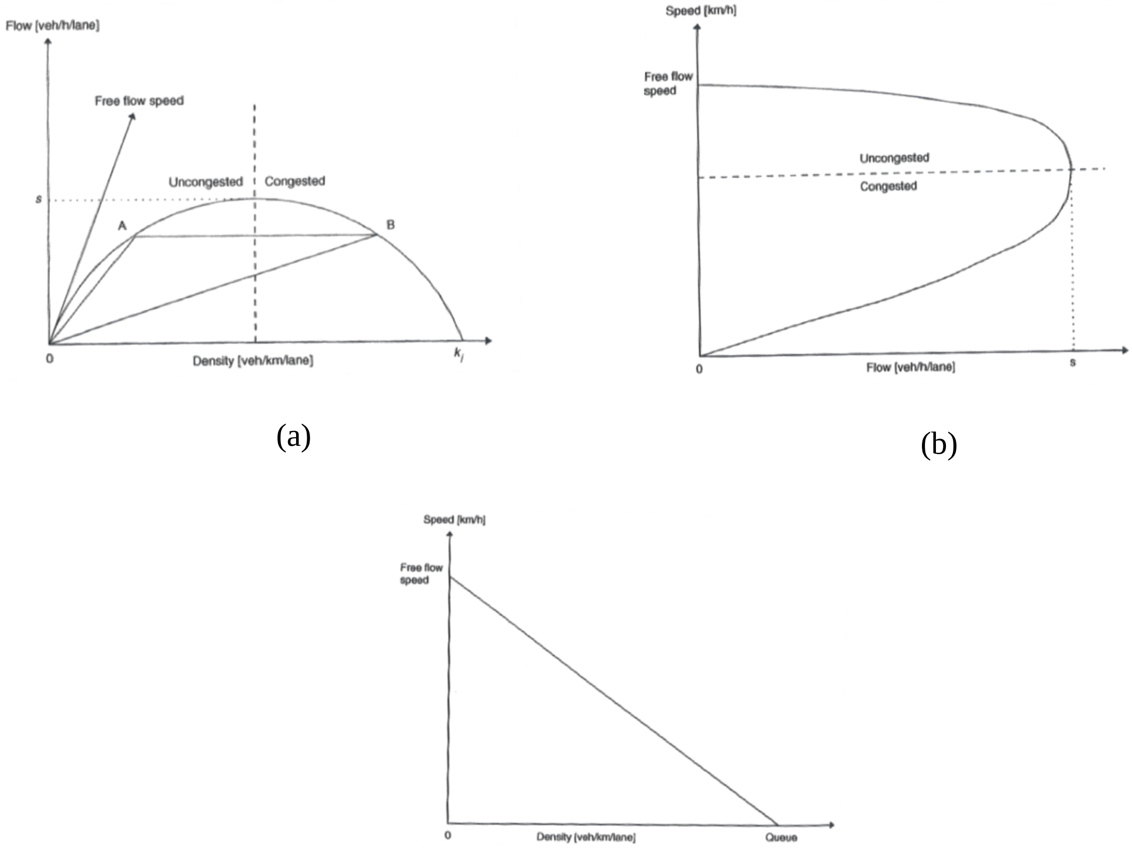
Density is used to measure the number of vehicles at an instant of time on a unit space. It is usually measured in vehicles per kilometre per traffic lane.

Occupancy measures the amount of time for which an inductive loop is occupied by a vehicle. Occupancy is a commonly used traffic data type in NRC detection, as the occurrence of NRC can be characterised by low speeds; hence, higher occupancies. Thus, most incident detection algorithms, as will be discussed in section 2.3.1, rely on occupancy data. However, occupancy data provides information on a section of a road as it is measured by inductive loops. This limits its usage for a large network consisting of hundreds of links.

These traffic data types are commonly used by traffic operation centres. However, various types of traffic data could be collected through traffic sensors. A comprehensive review of traffic data types is provided by Klein (2001).

2.1.3. Traffic Congestion in Macroscopic Fundamental Diagram

There is a fundamental relationship between flow, density and speed: Flow (veh/hr) = density (veh/km) $\times$ space-mean speed (km/hr). The relationship between traffic flow and density is often referred to as the Macroscopic Fundamental Diagram (MFD), and is illustrated in Figure 2-3.



(a) (b) (c)

Figure 2-3 The macroscopic relationship between density and flow (a), flow and speed (b) and density and speed (c)

Source: Bell and Iida (p.69)

The first observation that can be drawn from Figure 2-3 is that there are two traffic states: uncongested and congested. The second observation is that the traffic flow is zero in two situations: either when there is no vehicle on the road (i.e. density is zero), or the road is full of vehicles and all the vehicles are still (i.e. density is at maximum, and this is referred to as queue density, k_j), so no new cars can enter the road. In addition to the maximum flow (i.e. capacity, *s*), there are two density values for each flow value, one in the uncongested regime and the other in the congested regime (Bell and Iida, 1997).

Space-mean speed can be calculated from MFD, which is the slope of line connecting the origin with the point of interest. For example, the speed at point A is calculated as the slope of the line connecting the points zero and A. An important conceptual speed value is "*free-flow speed*" and there are differing definitions of what is meant by this. In theory, as shown in Figure 2-3(a), it is the highest achievable speed, but in practice it is regarded as the average speed between 02:00 and 05:00 AM, the period when travel demand is at its lowest. It should also be noted that a comparison of speed observations with free-flow speed might be misleading, since it is extremely rare for a vehicle to travel at free-flow speed.

The relationship between flow and speed is shown in Figure 2-3(b). This suggests that there are two possible flow values (except at the boundary of the transition from uncongested to congested regimes), one for uncongested and the other for congested regimes for one flow value. The logic behind this is that the flow is usually low in a congested situation and a free-flowing situation, since not many vehicles can be identified in a congested state, and not many vehicles are present in a free-flow state.

The association between density and speed is illustrated in Figure 2-3(c). Speed and density are inversely proportional. That is, as density increases, speed decreases. Maximum density is reached when traffic is standing still (i.e. speed is equal to zero). The intuitive reasoning behind this is, the more vehicles on a road network, the lower the possibility they will move at higher speeds due to increased interaction between drivers.

Although these traffic theoretical approaches are well accepted, analysed data usually do not fit into these smooth curves, and highly scattered MFD plots are observed, especially in congested regime (Sugiyama et al., 2008). However, there is empirical evidence that when the whole network is considered, these theoretical connections start to emerge again (Geroliminis and Daganzo, 2008).

Although theoretically well developed and accepted, MFD cannot be used to differentiate between RC and NRC. This is because MFD focuses on the generic characteristics of traffic congestion, which are common in both RC and NRC (e.g. reduced speeds and increased density). Due to this commonality, MFD is not suitable to differentiate between the types of congestion.

2.2. Link Journey Time

Link journey time (LJT) is an approximation of the journey time through a link at an established time interval. It is suitable for network-wide analysis, as it is a prominent traffic data type collected on a length of a road (Hall, 2001). LJT is treated as an important indicator of the performance of a link due to the following three reasons.

Firstly, it is the primary indicator of traffic flow. Secondly, almost all other measures of link performance are correlated with travel time. Thirdly, it is easy to measure (Sheffi, 1985, p.16). In addition, delay, which is calculated by using LJT, is reported to be most important indicator of congestion by both commuters and traffic operators (OECD, 2007, p.43). Furthermore, travel time is a major resource that commuters allocate for commuting and most people plan their journeys based on its estimated duration. Last, but not least, LJT is cumulative across adjacent links to estimate the LJT of an aggregated link. For example, LJT from location *a* to *b* and *b* to *c* can be summed to obtain the LJT from location *a* to *c* (Wardrop, 1952, p.331; D'Angelo et al., 1999).

2.2.1. Estimation of an LJT

Estimation of an LJT requires the calculation of a vehicle's travel time. A vehicle's travel time can be calculated by matching two sensors' observations. These two sensors are located at the beginning and ending of a link and can identify a vehicle when the vehicle passes underneath them. There are two main sensors that can be used to identify a vehicle: an Automatic Number Plate Recognition (ANPR) camera and an Automatic Vehicle Identifier (AVI). An ANPR camera captures the licence plate of a vehicle, whereas an AVI sensor identifies a vehicle by a radio-frequency identifier. ANPR cameras are commonly used in city centres for law enforcement (e.g. congestion charging), whereas AVI sensors usually operate on motorways to collect tolls. Having calculated a number of vehicles' journey times, LJT can be estimated based on those vehicles journey times.

Suppose a link is defined as the shortest route between two ANPR cameras, as shown in Figure 2-4. There are other roads (dashed lines) that intersect with the link (thick line), with arrows indicating traffic directions. When a vehicle passes underneath an ANPR camera, an observation is recorded with the licence plate of the vehicle and the time of the capture. If two ANPR cameras record the same licence plate, a *match* occurs, and the *travel time* of the vehicle can be calculated from the capture time at both cameras by simply differencing the capture times.

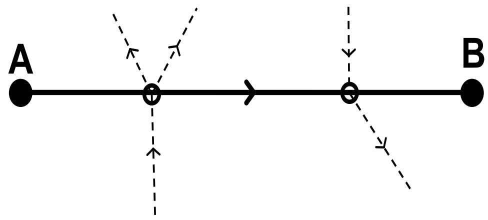


Figure 2-4 A link defined by two ANPR cameras, A and B

Suppose that during the data collection period, nine vehicles are identified by the two ANPR cameras. The recorded observations are illustrated in Table 2-2.

Table 2-2 Observation recorded at ANPR cameras A and B, as shown in Figure 2-4

Vehicle ID	Capture Time at A	Capture Time at B	Travel Time (min)
1	09:00	09:07	7
2	–	09:08	–
3	09:02	09:07	5
4	09:03	09:48	45
5	09:04	09:10	6
6	09:04	–	–
7	09:04	09:12	8
8	09:06	09:16	10
9	09:06	09:08	2

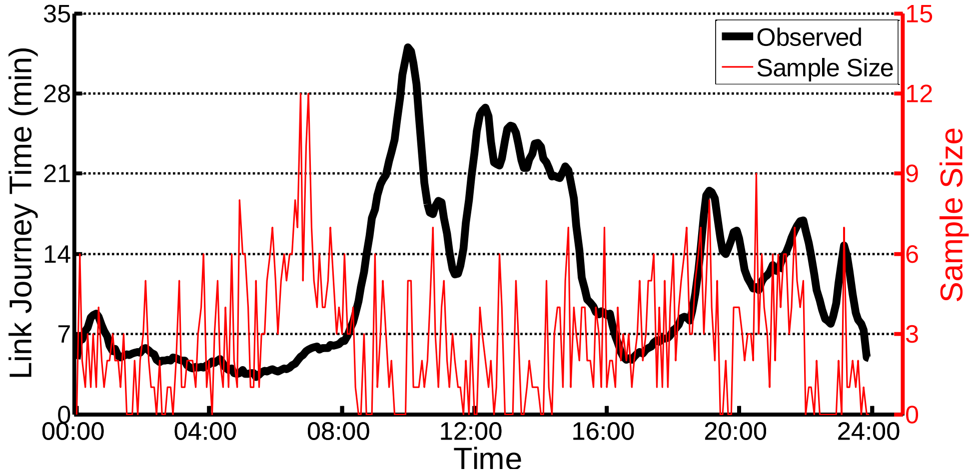
All of the observations are physically possible, including the two outliers. The first outlier is vehicle 4, as its journey took 45 minutes, whereas similar journeys conducted by vehicles 5 and 3 took six and five minutes respectively. This observation could be made in two possible situations. The first is that the vehicle might have taken a different route than that defined by the link (since there are two junctions within the link), and hence had a longer journey. The second is that the vehicle might have stopped after passing camera A, and then continued its journey through camera B, again resulting in a longer travel time. The second outlier is vehicle 9, which completed its journey in a much shorter time than other vehicles. This could be an official vehicle (police or ambulance) with alarms on or simply a fast moving vehicle; however, in both cases it does not represent the majority of vehicles. Therefore, these sorts of observations must be discounted before an LJT is estimated (Robinson and Polak, 2006).

Furthermore, the travel times of two vehicles (vehicle 2 and 6) could not be determined. Such an outcome may occur in several situations. The first is when, even though those vehicles travelled the link continuously, their licence plates might not have been captured by either of the ANPR cameras due to occlusion between the vehicles or simply because of the failure of the software on the ANPR camera. The second reason is that vehicles might have joined or exited the link at a junction in between two ANPR cameras. In both situations, it is clear that these vehicles contributed to the traffic, but their journey times cannot be used to estimate LJT. This is a limitation of using LJT, and the problem becomes worse as the link lengths increase.

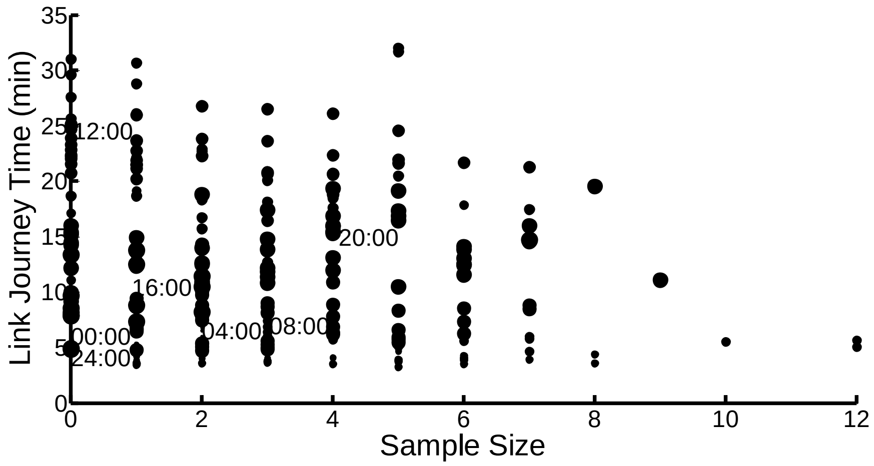
Another factor to be considered is the temporal granularity of an LJT observation (i.e. data collection interval); the higher the temporal granularity, the shorter the sampling interval. Suppose that temporal granularity is chosen as five minutes and LJT at 09:00 is to be estimated for the observations given in Table 2-2. All the acceptable observations, which were captured at camera A between 09:00 and 09:05, are used for this estimation. Thus, by using journey times of vehicles 1, 3, 5 and 7, the LJT is estimated as (7+5+6+8) / 4 = 6.5 minutes. As four vehicles' journey times are used, the ***sample size*** for the estimated LJT is four.

If the temporal granularity is increased to one minute, then it would be possible to estimate the LJT at 09:03. However, this calculation would rely on a single observation, which is an outlier (i.e. vehicle 4). This is because increasing temporal granularity decreases the number of observations that can be used to estimate an LJT. Furthermore, it is also possible that no vehicles might have been captured during a given temporal interval (e.g. between 09:05 and 09:06), and thus a missing LJT observation occurs. In order to avoid missing LJT observations, a process called *patching* is applied, which is used to estimate an LJT based on different strategies based on the amount of missing data (Haworth and Cheng, 2012).

Once the temporal granularity is determined and the LJTs are estimated, two time series can be used to describe a link for a given day of analysis. These two time series are the estimated LJTs and the sample size that is used to estimate each LJT, as each of them has a value for a corresponding time interval. These two time-series can be illustrated by two ways. First, a time-series plot is shown in Figure 2-5(a), where the estimated LJTs and the sample size to estimate each LJT is plotted with respect to time of day. Second, a scatter plot is shown in Figure 2-5(b), where the points' size increase by time of day (i.e. smallest point corresponds to 00:00 and largest point correspond to 24:00) and the correlation between the estimated LJTs and sample size, if any, could be better observed.



(a)



(b)

Figure 2-5 Data of a link at a given day in time-series format (a) and as a scatter-plot (b)

There are several limitations of using LJT estimates that are collected by ANPR cameras. Firstly, an ANPR camera cannot capture all the vehicles that pass beneath it. The sampling rate is reported to be between 50-90% (see references within Robinson and Polak (2006)). The sampling rate depends on factors including the topology of the road network, the viewing angle of the ANPR cameras, weather conditions and congestion levels (Chang et al., 2004). Thus, researchers focused on the accurate estimation of LJT from vehicle data with a low sample size (Dion and Rakha, 2006; Turner et al., 1998). In order to minimise sampling errors, a thorough data cleaning procedure is required to estimate LJTs (Robinson and Polak, 2006). Lastly, most of the analyses using LJT are retrospective in nature because of the necessary data cleaning procedure; also, an LJT can only be estimated *after* vehicles complete their trip on a link.

2.2.2. Distribution of LJTs

There is a mutual relationship between traffic congestion and travel time reliability. This relationship can be described in general terms as ‘congestion decreases travel time reliability’ and ‘the higher the travel time reliability, the lower the variation of travel times from the expected travel times’. It should also be noted that ‘expected travel times’ capture anticipated congestion during peak periods, including the bottlenecks observed during the AM/PM peak periods (referred to as recurrent congestion, RC). Thus, it is the NRC that degrades travel time reliability rather than the RC. In order to derive a better understanding of NRC and to conduct statistical analysis upon it, it is also necessary to develop a better understanding of travel time reliability, which is a broad research area (Bell and Cassir, 2001; Noland and Polak, 2002; Polus, 1979; Pu, 2011).

Modelling the statistical distribution of LJTs is an important component of research on travel time reliability (Hollander and Liu, 2008; Susilawati et al., 2011), and has spanned decades (Wardrop, 1952). However, there is still no agreement on the reported results, which is mainly due to the highly subjective nature of the research. Firstly, ‘distribution of travel time’ might assume different meanings in different studies. For example, Arezoumandi (2011) and Susilawati et al. (2011) attempt to find a distribution to characterise the travel times on a link, regardless of the temporal variations within a day, whereas Polus (1979), on the other hand, considers two time periods (i.e. to and from work trips). Secondly, different road characteristics may result in different outcomes. For example, link lengths have shown a distinctive effect (e.g. making the distribution bimodal) on the distribution of travel times (Susilawati et al., 2011). Furthermore, it is also seen that most of the existing literature does not mention the data cleaning procedure, which might have an important impact on the distribution of LJTs.

On the other hand, some researchers have developed an agreement on some aspects of LJT distribution. The most important of these is about the shape of the LJT distribution. The distribution should *not* be symmetric, as an LJT cannot take a negative value. In addition, the distribution must be positively skewed. This is because most trips are completed within a certain time interval, which is characterised by expected travel time. Thus, the majority of the travel times are located within a narrow travel time interval situated on the left side of the distribution. These two important characteristics of the distribution of journey times were noted decades ago (Wardrop, 1952) and are supported by current literature (Susilawati et al., 2011). These common properties are captured by *lognormal* distribution (Arezoumandi, 2011; Dandy and McBean, 1984; Hollander and Liu, 2008).

2.3. Existing Methodologies for NRC Detection

The importance of NRC detection is recognised by most traffic operation centres, yet few studies focus solely on this topic. It is observed that NRC detection and incident detection are two closely linked research areas. This is because an incident is usually assumed to cause NRC. Thus, it is reasonable to think that detection of an incident necessitates detection of an NRC. Therefore this section explores Automatic Incident Detection (AID) methods and investigates to what extent AID can be used for NRC detection. Please note that a comparison of AID methods requires the definition of evaluation measures such as Detection Rate (DR) and False Alarm Rate (FAR), which are discussed in section 2.5.

2.3.1. Incident Detection-based Methods

Automatic Incident Detection (AID) presents an interesting research problem that has attracted traffic scientists for decades. With the advancement of sensor technology, AID methods analyse the collected traffic data with the aim of detecting incidents immediately, thus lessening traffic operators' load. Since the advent of AID, a traffic operator only needs to validate the detected incidents, which enables him/her to spend less time surveilling. Research efforts have resulted in a substantial number of AID methods and several surveys have been written to summarise these techniques (Martin et al., 2001; Parkany and Xie, 2005; Weil et al., 1998).

In order to understand what AID methods seek to achieve, it is important to clarify what is meant by an 'incident'. Some definitions are as follows:

- "*An anomaly that disrupts the smooth flow of traffic*." Examples include debris on the road, vehicle accidents, roadworks, events (e.g. sports matches, concerts or festivals) or severe weather (Klein, 2001, p.134).
- "*Any non-recurring event that causes a reduction of roadway capacity or an abnormal increase in demand*." Such events include traffic crashes, disabled

vehicles, spilled cargo, highway maintenance and reconstruction projects, and special non-emergency events" (FHWA, 2010).

- "[An] *event which reduces the capacity of network links to carry traffic*" (Ash, 1997).
- "*Any non-recurrent event that has a perceptible negative impact on traffic operations*" (Williams and Guin, 2007). Examples of incidents include traffic accidents, debris on roadway and stalled vehicles.
- "[An] *unexpected and non-recurrent event that significantly reduces link capacity and reduces travel speeds, and has a significantly longer term impact at specific locations than on freeways*" (Luk et al., 2001). This explanation focuses on the differentiation of incidents based on the network type on which they occurred, and events which cause minor disruptions on an urban road network are not considered to be incidents.

Thereon, incident detection is defined as "*the process of identifying the spatial and temporal coordinates of an incident*". It consists of two steps; firstly, the existence of congestion is determined, and secondly, the available data are further analysed to determine whether the cause of congestion is an incident (Ozbay and Kachroo, 1999, p.61; Weil et al., 1998).

Various methodologies have been applied to detect incidents, and the boundaries between these different methodologies are not crisp. It is observed that AID methods can be classified into four main types: *pattern based algorithms*, *traffic theoretical approaches*, *statistical methods* and *machine learning* based methods. Although these headings cover the majority of AID methods, advances in *signal processing* also show some promising results on incident detection.

Pattern based algorithms consist of two main steps. In the first step, the 'expected' patterns (or profiles) of the examined traffic data type, which is usually occupancy, are determined. This step is achieved either by using historic data or by incorporating expert knowledge. Then, the second step constructs a decision tree based on the traffic data type which is being used to detect incidents. One of the first AID algorithms uses this logic is referred to as the 'California Algorithm', which soon became the benchmark to compare different AID algorithms (Payne and Tignor, 1978). Incidents are detected through the comparison of the occupancy measurements at two adjacent sensors. Occupancy measurements at upstream of an incident will increase (as the duration taken for vehicles to pass the incident's location will increase), while occupancy measurements at downstream will decrease.

There are three main limitations of the California algorithm. Firstly, it uses one observation per-time. This increases the false alarm rate (FAR), because it cannot alleviate the dynamic fluctuations in traffic data. Secondly, it requires a laborious calibration procedure, which involves determining the threshold values used in the construction of the decision tree. Thirdly, it uses raw traffic data, which are often highly stochastic. This might prevent optimum calibration and result in a high FAR. In order to alleviate these limitations, the Minnesota algorithm was proposed which operates on smoothed traffic data reduced the FAR (Stephanedes and Chassiakos, 1993). However, the smoothing process requires data that have not been collected at the time of analysis, which increases the response times.

Traffic theoretical approaches require two traffic data types to detect incidents. An incident is identified by observing an abrupt change in one of the traffic data types, whereas the observations of the other traffic data type remain steady. The requirement of an abrupt change in one of the traffic data types results in the association of this method with 'catastrophe theory'. For example, an incident can be characterised as speed observations dropping dramatically, while occupancy observations are steady. The most prominent algorithm that follows this methodology is the McMaster algorithm, which uses speed-occupancy and flow-occupancy connections based on a single sensor to detect incidents (Persaud and Hall, 1989). The McMaster algorithm can successfully differentiate between RC and NRC, which is indeed an important requirement of any AID method. However, it requires two traffic data types (i.e. speed and occupancy), which might not be readily available.

Statistical methods aim to generate a statistical model of the traffic data types that are being used. The first group uses a probabilistic approach to detecting incidents, which can be classified as Bayesian based models. The Bayesian approach uses probabilistic modelling by considering prior knowledge and posterior probabilities on the occurrence of incidents. Hence, Bayesian methods inherently consider the probabilistic nature of traffic phenomena. Designing a Bayesian network is not a straightforward process, as it requires a thorough analysis of the causal relationships within the road network (in other word, which link affects other links). For each causal relationship, a Bayesian network must be constructed. An important advantage of using Bayesian methods is their ability to deal with missing or noisy traffic data (Sun et al., 2006). On the other hand, Bayesian methods are usually applied at the individual link level, as the task becomes intractable if spatio-temporal autocorrelation amongst the adjacent links is to be modelled (Fei et al., 2011). Therefore, an integrated spatio-temporal Bayesian approach to detecting incidents on motorways and urban road networks is an open research area.

Another type of statistical method is the use of time series modelling via autoregressive moving average (ARIMA). It is shown that use of previous traffic occupancy observations to predict current occupancy can outperform benchmark algorithms like the California algorithm (Ahmed and Cook, 1980). The quality of the prediction in ARIMA models depends on the number of autoregressive terms, which varies with the traffic situation (e.g. the number of previous occupancy measures to be used might be different for congested and free-flowing situations).

Machine learning based incident detection relies on the fact that commonly used machine learning methods such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) can reveal hidden patterns in traffic data, thus indicating an incident. Two steps are required to develop artificially intelligent models: training and testing. In the training step, observations and the output corresponding to those observations are used to construct a model. The second step is the testing part, where the effectiveness of the model obtained from training is tested using observations that are not used in training.

ANNs are simple representations of the human brain, where the input is processed through an interconnected neural network. A neural network comprises of neurons, which are used to process information, and weights, which transfer information between neurons. Training is achieved by updating the weight values of the connections between the neurons. There are various different weight configurations which may all end up giving the same training performance, and most of these different configurations correspond to a local optimum. The main limitation of an ANN arises at this point, since training is completed when a local optimum is achieved, without providing information on whether or not it is a global optimum. Thus, often, lots of training runs are frequently required to establish a better idea of the performance of the ANN, a process called cross-validation. SVM, on the other hand, depends on the observations' distribution on the space, and finds a hyperplane to separate the observations. 'Support vectors' are the observations that the hyperplane passes through, and only these observations are required to construct an SVM model. Yuan and Cheu (2003) demonstrate the effectiveness of using SVM on incident detection while providing a concise overview of the similarities and differences between ANN and SVM. They experimentally show that SVM outperforms ANN in terms of DR and FAR.

The main limitation of ANN and SVM is the lack of knowledge on setting the given parameters of ANNs (e.g. activation function, learning rate, number of neurons in the hidden layer) and SVMs (i.e. the kernel and its parameters), which often require sensitivity analysis to determine these parameters (Wu et al., 2004). In addition, it is often difficult to derive physical insight from the parameters of ANN or SVM, as the way in which they function resemble a black-box (Sjöberg et al., 1995).

Signal processing methods convert time series signals into frequency domain signals via transformations like Fourier or Wavelet transforms. These transformations are applied in numerous application domains, and their effectiveness to detect incidents has also been tested. The main motivation in these transformations is to investigate the signal in the frequency domain, which might allow the researcher to detect patterns that do not appear in time domain. For example, Jeong et al. (2011) use wavelet transforms on the time series consisting of occupancy and speed measurements at upstream and downstream sensors. They also provide a detailed literature review on wavelet based methods, which are proposed to detect incidents. The main advantage of their method is to use adaptive thresholds that adjust themselves according to the traffic situation when detecting incidents. However, their experiments mostly used simulated data and the effectiveness of testing their method in a real-life context remains as a future research objective.

It has also been illustrated that combined usage of the aforementioned methods results in improved performance. For example, optimising the parameters of an ANN via genetic algorithms has resulted in higher performance than that exhibited by ANN alone (Roy and Abdulhai, 2003). Although their method was suggested to be transferable, only two links were analysed and a very high detection rate raises concerns with regard to overfitting of the developed models to the two links that were analysed. Karim and Adeli (2002) combine ANN with wavelet analysis using occupancy and speed measurements, which decreased the FAR and increased the DR. An important aspect of their research is the proposal of a new performance evaluation metric, which they refer to as ‘algorithm portability’. This new measure favours methods that require the least amount of calibration, so a method that works well for a small network can be scaled to a larger network without calibrating its parameters. However, their research focuses on motorways, and its usefulness when applied to an urban road network is an open research question.

2.3.2. Limitations of the Existing Incident Detection Methodologies

There are three main limitations regarding the current methodologies developed to detect incidents that could be applied to NRC detection in a large urban road network:

- Most AID methods require observations collected on two or more traffic data types, and occupancy is usually the common traffic data type that is required for almost any incident detection algorithm. This is mostly due to the common usage of loop detectors. However, as occupancy is measured only for a short section of a road (Hall, 2001), its application to a large urban road network is limited; as it is cost prohibitive to install loop detectors across a space of this size.
- Most of the AID methods have been designed for motorways, in which incident detection and NRC detection are treated analogously. Although attempts have been made to detect incidents on an urban road network, they are mostly incapable of differentiating RC from NRC (Cullip and Hall, 1997). The differences between incident detection on motorways and urban road networks have been recognised previously (Luk et al., 2001) and are summarised in section 2.1.1. These differences complicate the transferability of an AID method from motorways to an urban road network.
- Although most of the AID algorithms are reported to have very high DR and low FAR, only 12.5% of surveyed traffic operation centres have been continuously using an AID method (Williams and Guin, 2007). Considering the decades of research on AID, this finding is startling as it highlights the immense difference between theory and practice.

As a summary, most of the AID methods require the detection of NRC due to the common belief that an NRC is usually caused by an incident. However, the aforementioned three reasons prevent knowledge developed in AID from being transferred to be used for NRC detection on a large urban road network. Therefore, new techniques should be investigated for NRC detection on a large urban road network.

2.4. Detecting Traffic Congestion by Spatio-Temporal Clustering

Traffic congestion has been characterised as a clustering phenomenon, from both theoretical and empirical aspects. Theoretical aspects focus on how a cluster of vehicles moving at low speeds at critical traffic densities could cause congestion, which is referred to as a ‘phantom jam’ (Helbing, 2001; Kerner and Konhäuser, 1994; Sugiyama et al., 2008). These theoretical findings are developed mainly on motorways where traffic flow is uninterrupted for long distances. Investigation of such traffic congestions in an urban road network still remains as a research challenge, due to regular and irregular disturbances (e.g. traffic lights, pedestrian crossings) in traffic flow, as well as the higher rate of interaction between drivers. Therefore, empirical research has investigated how spatio-temporal clustering can be used to detect congestion patterns in an urban road network. Examples of such research include clustering links having increased LJTs (Anbaroglu and Cheng, 2011) or increased traffic density (Li et al., 2007). These pieces of empirical evidence mostly focus on congestion detection and their application to NRC detection requires further research. Nevertheless, these research findings illustrate the effectiveness of spatio-temporal clustering on congestion detection.

Traffic congestion, as most other spatial phenomenon, is dynamic. Therefore, a coherent combination of both spatial and temporal aspects is required when clustering spatio-temporal data (Deng et al., 2011; Shekhar et al., 2011). The wide usage of the term ‘clustering’, however, results in different meanings attached to the term *cluster* and there are two main strategies for clustering. The first strategy uses a similarity function to group similar observations into clusters (Dubes and Jain, 1976; Han and Kamber, 2006, p.383). The aim of a similarity based clustering algorithm is to maximise the similarity of observations within a cluster (i.e. minimise the distance between its observations) and maximise the dissimilarity between different clusters (i.e. maximise the distance between clusters). Some of the representative algorithms in this strategy include *k*-means (Macqueen, 1967) and DBSCAN (Ester et al., 1996). This strategy is discussed in detail in section 2.4.1.

The second strategy applies a statistical significance testing procedure to detect statistically significant clusters (Besag and Newell, 1991). Methods developed within this strategy use the statistical properties of the data and aim to detect ‘anomalous’ clusters. The detected clusters often correspond to unusual increases in the observed phenomenon, like a region having unusually high rates of disease or crime. The representative methods of this strategy include spatial scan statistics (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995) and space-time scan statistics (Kulldorff et al., 2005; Neill, 2008). The methods developed within this strategy are discussed in detail in Section 2.4.2.

An important point should be highlighted regarding the linkage between these two strategies. It has been suggested that hypothesis testing should be conducted to validate the detected clusters from the first strategy (Dubes, 1993; Halkidi et al., 2001). This, indeed, is very similar to the idea behind the second strategy, as it involves a significance testing process to determine whether the clusters are observed by chance alone or not. However, it has been observed that such hypothesis testing has not been applied in the case of most clustering algorithms based on the first strategy. Therefore, it is reasonable to distinguish the two strategies of clustering. Note that previous research distinguishes these two strategies by referring to them as 'clustering' and 'cluster detection' respectively (Neill, 2006, p.13; Wheeler, 2007). These two terms are very similar to one another, so in order to avoid ambiguity this thesis follows the aforementioned terminology.

2.4.1. Spatio-Temporal Clustering based on a Similarity Measure

The aim of clustering based on a similarity measure is to group the observations in clusters so that the *similarity* between observations of a cluster is maximised. Similarity between observations is usually defined by a distance measure, which is determined according to the objectives of the research (Dubes and Jain, 1976; Han and Kamber, 2006, p.383). Clustering via a similarity measure has been commonly used as a useful exploratory analysis tool. This is because vast amounts of data are being collected without a label being associated with them, and it is often necessary to group the data into meaningful portions for better modelling of data (Berkhin, 2006; Jain et al., 1999; Xu and Wunsch, 2005). Various methods can be used to conduct similarity based clustering and some commonly used examples are discussed as follows.

Partitioning based methods group observations into *k*-clusters. Most of the partitioning based methods iteratively update the centres of the clusters by moving observations to the closest clusters defined by a distance function and updating the cluster centre until convergence (i.e. clusters do not change with further iterations). A cluster's centre can be represented by the mean of the observations belonging to that cluster (Macqueen,

1967) or the medoid so that each cluster centre corresponds to an observation (Theodoridis and Koutroumbas, 2006).

Partitioning based methods became popular because of their simplicity. However, the output is dependent on the *k* value and the initial assignments of clusters' centres. Therefore, it is often necessary to vary the number of clusters, *k*, and the initial assignments (Adnan et al., 2010). Furthermore, arbitrarily shaped clusters may occur in spatial phenomenon and they cannot be detected by partitioning based methods (Ester et al., 1996; Han and Kamber, 2006, p.403).

The partitioning idea could also be applied when the data is transformed into a different space. The most commonly used method under this domain is 'spectral clustering', where data are represented as their eigenvectors. The main motivation of this approach is to investigate whether there is better clustering in the transformed data space. In this way, concave shaped clusters can be detected which cannot be detected via *k*-means (Ng et al., 2001). However, spatial information is usually lost through the transformation, which limits its application on spatial and spatio-temporal clustering (H. Q. Liu et al., 2010). The loss of spatial information can be observed in Türkeş and Tatlı (2011), where the authors apply spectral clustering and spatially non-continuous clusters are detected.

Mixture models assume that each cluster has a statistical representation and they are referred to as model-based clustering or statistical clustering. An important advantage of these methods is that they might uncover the underlying distribution of the data. In addition, rather than making the decision to assign an observation to a cluster solely based on the distance between the observation and the cluster centre (which is the case in partitioning based methods), mixture models also consider the variation within the clusters. Another important advantage of clustering using mixture models is that they consider that each observation has a 'likelihood' value which quantifies the decision to assign an observation to a cluster. In this way, mixture models provide a statistical framework to perform fuzzy clustering.

The Expectation Maximisation (EM) algorithm (Dempster et al., 1977) is commonly used for clustering by mixture models. It can be thought of as an extension to the *k*-means algorithm, since it can be used to model a larger class of data sets (Han and Kamber, 2006, p.429). The EM algorithm consists of two steps and alternates between these two steps. The first step, Expectation, calculates the expected cluster memberships of each observation. The second step, Maximisation, uses the previously calculated values to maximise the likelihood values of the parameters of the distribution that are used to define a cluster. These two steps are iteratively continued until convergence. In spatial clustering, the spatial adjacencies should be considered and it is desirable to detect the clusters which are compact in both spatial and temporal domains. One way of dealing with this issue is to assume the spatial domain is an extension of the thematic domain (i.e. non spatio-temporal attributes of a phenomenon). However, this might result in clusters that are not spatially continuous. In order to detect both spatially and thematically compact clusters, Ambroise and Govaert (1998) extend the EM algorithm to what they refer to as Neighbourhood EM, which adds a spatial penalty term that forces the method to detect spatially continuous clusters. However, the compactness of the clusters depends on the penalty term and setting this parameter was left as a future research.

The main limitation of mixture models is the necessity to determine the number of clusters a priori. In addition, mixture models cannot detect arbitrarily shaped clusters, since the statistical models assume that the data set is generated via these models. Furthermore, an EM-based algorithm requires many runs, since the obtained results only guarantee local optimum.

Hierarchical clustering methods create a hierarchical composition of clusters so that each level of the hierarchy corresponds to a clustering output. Therefore, this approach generates all the possibilities between representing each observation as a cluster and representing all the observations as one cluster. Unlike partitioning based methods, each cluster is represented by all its observations. When the similarity between two clusters is to be calculated, the closest pair of observations belonging to different clusters can be chosen (Han and Kamber, 2006, p.410).

Hierarchical clustering has also been used in traffic science. Wei and Peng (2009) apply hierarchical spatio-temporal clustering on a motorway network to determine similarly behaving links. Their method requires three parameters. The first is a spatial constraint, which determines whether two traffic sensors are spatially close. The second is a similarity threshold, which determines whether traffic observations of two sensors are similar. The third is a temporal window size, which determines the number of consecutive traffic observations that are considered when detecting clusters. However, these values are arbitrarily chosen and the resulting clusters are not associated with any physical phenomenon. Another interesting example on hierarchical clustering is illustrated in Yiu and Mamoulis (2004), where authors emphasise using network distance to define similarities between observations (represented by points) rather than Euclidean distance. Chen et al. (2008) extend the framework proposed by Yiu and Mamoulis (2004) by clustering moving objects (e.g. probe vehicles) on a road network. Although they briefly mention that the detected clusters correspond to congested areas, the discussion is limited mostly to the run-time aspect of their algorithm.

The main limitation of hierarchical clustering algorithms is their inability to rewind a merge (or split) decision and the next steps are based on newly generated clusters. Although this provides a simple way to detect clusters, it does not tolerate errors that could have been made when merging or splitting clusters (Han and Kamber, 2006, p.408).

Density based clustering methods are used to discover arbitrarily shaped clusters by creating a neighbourhood around the data object. The most prominent density based algorithms are DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999). These methods can handle the density differences among the data; however, they require predefined thresholds which are difficult to set, as they require visual validation of the results. Furthermore, since most density-based algorithms require a fixed distance threshold to generate the clusters, handling datasets which contain clusters of different densities might pose serious concerns. This is an important issue in spatial information science; because of the common occurrence of density differences within a study area has differences in density. For instance, in traffic science, a city centre will contain more loop detectors or ANPR cameras than the suburbs.

Most of the spatial and spatio-temporal clustering algorithms based on point data are extensions of density based methods. This is mainly because of the importance of the concept of density in spatial information sciences. Regions of high density (e.g. of patients, crime activities, forest fires) have long been of interest to researchers. This has led to the development of several algorithms, as presented by Birant and Kut (2007) and Deng et al. (2011). These researchers consider the temporal domain to be another thematic domain attribute, since the detected clusters are fixed over the examined period. However, it is also important to understand how clusters change their shape over time, which may allow a researcher to achieve a better modelling of a spatial phenomenon.

Limitations of Spatio-Temporal Clustering Based on a Similarity Measure

There are four main limitations of spatio-temporal clustering methods that rely on a similarity measure.

- Several thresholds are usually required to define which observations are close to one another in spatial domain, as well as thematic domain. There is usually a lack of knowledge on setting these parameters, and researchers often rely on empirical observations to determine these thresholds (Lin et al., 2005). Therefore, it is usually necessary to incorporate domain knowledge in a clustering algorithm (Jain et al., 1999; Wagstaff et al., 2001).
- The number of clusters is usually not known in advance, and the methods requiring this input have more limited use in a real-life context. It should also be highlighted that, even if the correct number of clusters is known, this is not sufficient to evaluate the performance of a clustering result, as many different clustering schemes of the same data set could yield a given number of clusters (Vendramin et al., 2010).
- It is usually difficult to compare different clustering algorithms due to the lack of ground truth data. This is because; the information suggesting the 'true' labels of all observations is usually not known in advance (Dubes and Jain, 1976). Validating the detected clusters often requires statistical analysis (Halkidi et al., 2001). However, most of the similarity based clustering methods do not consider such a validation process.
- Context plays an important role in the detection of meaningful clusters and should be considered when calculating the distances between observations. Ignoring a spatial or spatio-temporal context may yield results that are not meaningful (e.g. locating an automated teller machine on a river (Tung et al., 2001)).

2.4.2. Spatio-Temporal Clustering based on Significance Testing

Detection of statistically significant regions of anomalously high (or low) concentrations of the observed phenomenon (e.g. number of patients or crime incidents) has been of interest to researchers since the mapping of cholera patients by John Snow in 1854 (Snow, 1854). There are three main ways to detect statistically significant regions (Besag and Newell, 1991):

- Global tests aim to identify whether the whole study area is anomalous or not. A common global test is Moran's I (Moran, 1948). However, global tests make assumptions that may not reflect reality, such as data being sampled from Normal distributions or each rearrangement of data between different locations being equally likely (Goovaerts and Jacquez, 2005; Rogerson and Yamada, 2009, p.57). In addition, global tests have a low detection rate if there is a localised cluster in the study area. This is because the localised cluster may not be big enough to be detected as a global cluster.

- Focused tests aim to identify whether a pre-determined region is significant or not; however, such tests often suffer from pre-selection bias, as the pre-determined region is chosen for a reason (Besag and Newell, 1991; Yamada et al., 2009). These tests are not commonly used, because a researcher usually does not know the location to search a priori to the analysis.

- Tests for cluster detection aim to identify localised clusters without determining a focus, as was the case in focused tests (Choynowski, 1959). One of the first methods used to conduct tests for cluster detection is the Geographical Analysis Machine (GAM), which aims to identify anomalous regions regardless their size or location. GAM scans the entire study area via overlapping circles whose location and size vary and determine the significant regions. In this way, any localised cluster can be detected, regardless of its size and location (Openshaw et al., 1988, 1987). GAM, although a useful exploratory analysis method, does not handle an issue called *multiple-hypothesis testing*; hence, it reports many significant spatial regions. This problem arises due to conducting many tests that are spatially correlated. Lowering the significance level to determine whether a region is significant or not, by a process called the Bonferroni adjustment (i.e. using an extremely low significance level), is not the optimal choice. This is because the Bonferroni adjustment is not useful when tests are spatially correlated, and it also decreases the effectiveness of detecting real clusters (Anselin, 1995; Brunsdon and Charlton, 2011; Caldas de Castro and Singer, 2006).

Spatial scan statistics was proposed in order to overcome the multiple hypothesis testing problem, while aiming to detect any localised clusters regardless their size and location (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995). The test statistic is the region that maximises the likelihood ratio value, which is calculated by dividing the likelihood of the alternative hypothesis (i.e. investigated region is a cluster) with the null hypothesis (i.e. there are no clusters in the study region). Spatial scan statistics were originally developed for point data, where each point is referred to as a '*case*'. A Bernoulli model is used in which there are two possible outcomes: a case or at-risk. In an epidemiological context, cases represent patients infected with a specific type of disease and the locations of these patients are known. Population-at-risk constitutes the people who might possibly become cases, but do not fulfil this criterion at the time of data collection. Spatial scan statistics formulate the null hypothesis, H_0, with the assumption that the probability of being a case is equal for all the individuals at risk (i.e. q). On the other hand, the alternative hypothesis, $H_1(r)$, suggests that the region being tested (i.e. r) is anomalous and the probability of being a case within this region (i.e. p) is higher than the rest of the study area (i.e. $p > q$). The regions are generated by overlapping geometric figures (e.g. circles or ellipses) whose location and size vary, and the likelihood of each region is expressed as follows (Kulldorff, 2006; Kulldorff and Nagarwalla, 1995).

$$f(r, p, q) = p^{c_r}(1-p)^{n_r - c_r} q^{C - c_r}(1-q)^{(N-n_r)-(C-c_r)} \qquad (1)$$

Where f is used to calculate the likelihood of region r, with a probability of observing a case within/outside region r are denoted as p/q respectively. The c_r and n_r are the number of cases and population-at-risk within region r respectively. C and N denote the total number of cases and total population at-risk within the study area respectively. As each region r has a different population, the likelihood ratio has to be used to compare different regions. The likelihood ratio is calculated as shown in equation 2 (Kulldorff and Nagarwalla, 1995):

$$Ƒ(r) = \frac{f(\mathrm{H}_1(r)|\mathrm{Data})}{f(\mathrm{H}_0|\mathrm{Data})} = \frac{\Pr\left(\mathrm{Data}\middle|\mathrm{H}_1(r)\right)}{\Pr(\mathrm{Data}|\mathrm{H}_0)} \qquad (2)$$

The likelihood function for the null hypothesis (i.e. denominator of equation 2) is calculated by plugging the most likely estimate of p (i.e. C/N) into equation 1, which becomes; $\frac{C^C(N-C)^{(N-C)}}{N^N}$. Similarly, the likelihood function for the alternative hypothesis (i.e. numerator of equation 2) is calculated by plugging the most likely estimate of p (i.e. c_r / n_r) and of q (i.e. $C - c_r$ /$N - n_r$) into equation 1, and the likelihood function is obtained as illustrated in equation 3:

$$\left(\frac{c_r}{n_r}\right)^{c_r}\left(\frac{n_r-c_r}{n_r}\right)^{n_r-c_r}\left(\frac{C\ -\ c_r}{N-n_r}\right)^{C-\,c_r}\left(\frac{(N-\ n_r)-(C-\ c_r)}{N-\ n_r}\right)^{(N-\ n_r)-(C-\,c_r)} \quad (3)$$

The likelihood ratio function is calculated for each region of the study area, $r \in R$, and the region that has the highest likelihood ratio value (denoted as $Ƒ(r^*)$ is chosen as the most likely cluster that might reject the null hypothesis (Kulldorff and Nagarwalla, 1995).

If spatial data are not available in point form, then the discretised geographic context (e.g. zones, regions or countries) can also be used with a different formalisation of the null hypothesis. When the spatial data are represented in discrete units, then Poisson based scan statistics can be used (Kulldorff, 1997).

Having determined the likelihood ratio values of each region and found the region maximising the likelihood ratio, r^*, it is time to assess its significance. The likelihood ratio values depend on the inhomogeneous population distribution; hence, they do not have an analytical form to calculate the probabilities (i.e. *p*-values). In order to calculate the probabilities, Monte Carlo simulations are used. This procedure is summarised as follows;

1. Create S simulations (replications) based on the null hypothesis (i.e. there is no significant cluster).
2. Find the region with the highest likelihood ratio test statistic score of each simulation (denoted as s_i^* for simulation *i*). It should be noted that these regions occur due to chance, since the simulations are created based on the null hypothesis.
3. Count the number of simulations whose $Ƒ(s_i^*)$ is higher than the region with the highest likelihood ratio in the observed data (denoted as s_{beat}).
4. Calculate the *p*-value of the region under investigation by $(s_{beat} + 1)/(S + 1)$. It is a common practice to choose S as values ending with nine as 19, 99 or 999 to obtain round *p*-values, as the original data set is also considered in calculating the *p*-value. For example, if S is chosen as 99 then the lowest possible *p*-value will be 0.01.

There would be many statistically significant clusters that overlap with the most significant cluster, as adding or removing an area from the region with the highest likelihood ratio value does not change the likelihood ratio value substantially. Therefore, spatial scan statistics provide an ‘approximate’ location of the real cluster (Kulldorff and Nagarwalla, 1995). The remaining statistically significant clusters may spatially overlap with the most significant cluster (Kulldorff, 2010, p.20) or they may occur at a distinct location (Zhang et al., 2010), all of which are referred to as 'secondary clusters'. Figure 2-6 shows a screenshot on the possible options to report the secondary clusters taken from the commonly used software on spatial and spatio-temporal cluster detection, SaTScan™ (http://www.satscan.org/).

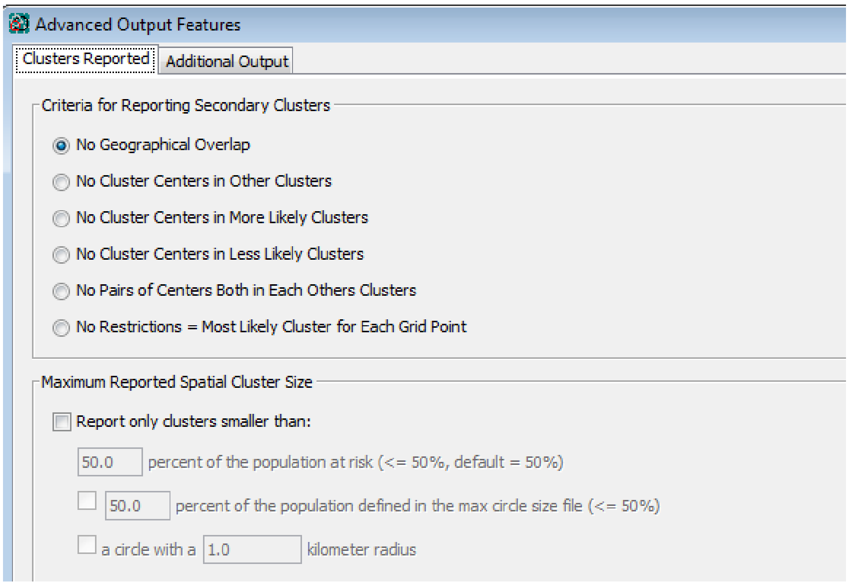


Figure 2-6 Different options to report secondary clusters on SaTScan™

Several spatial scan statistics methods have been proposed. One of the most prominent is expectation based scan statistics, which are useful when population at risk data are not available (Neill, 2008). Expectation based scan statistics compare the number of cases in regions with their expected values. Expected values are calculated based on the time series analysis conducted on historical data. The null hypothesis is formulated, as there is no clustering in a given region *r*. On the other hand, the alternative hypothesis assumes that region *r* is a cluster, which has a higher observed value than is expected. The count data can be generated by a Poisson process (Neill, 2009) or a Gaussian process (Neill, 2006, p.38).

Space– Time Scan Statistics (STSS) is an extension of spatial scan statistics so the temporal aspect of a spatial phenomenon is also included in the analysis. The existing research focuses on the detection of emerging clusters. For example, in disease outbreak detection, epidemiologists are interested in detecting an emerging disease outbreak as early as possible, so that preventive measures can be taken on time to prevent/minimise fatalities and economic loss (Kulldorff, 2001). The advantage of STSS, compared to earlier space-time cluster detection methods, for example Knox-text (Knox, 1964), is that STSS does not require any thresholds that define whether two observations are *close* in space and time, as it employs overlapping space-time regions (STRs) in a fixed geometrical shape like cylinders to scan the entire study area and determine regions that display an unusually high number of observations (e.g. patients).

Another way of detecting statistically significant spatio-temporal clusters is to apply spatial scan statistics at each temporal interval. Then, statistically significant spatial clusters that are found at each temporal interval can be combined to detect spatio-temporal significant clusters. However, this solution is not feasible for several reasons. If spatial scan statistics are applied continuously over time and significant spatial clusters are reported at these different times, then the spatial scan statistics might struggle to detect clusters that gradually develop. In addition, such an approach introduces multiple-hypothesis testing again, as the entire region is tested for all the temporal period. Therefore, STSS considers space and time in conjunction to achieve a better performance when seeking to detect statistically significant clusters.

Another type of STSS is space-time permutation scan statistics, which does not require population data (Kulldorff et al., 2005). This method generates permutations by shuffling the observations' occurrence times so that the total number of observations and their location stay constant. In this way, the method adjusts for any purely temporal variations, as well as any purely spatial variations, since the total number of cases is the same in the real and permutated data sets. However, permutation based STSS is derived by assuming the cases are distributed according to Poisson distribution and further research is required to derive models for phenomena that could be modelled with another distribution.

Currently, STSS is regarded as a state-of-the-art method to detect statistically significant clusters (Patil and Taillie, 2004). STSS is commonly used in epidemiology (most of the previously cited literature), crime science (Maciejewski et al., 2010; Nakaya and Yano, 2010) and environmental research (Tuia et al., 2008). An extensive list of researchers that use population based spatial scan statistics and STSS is provided in SaTScan (2010). However, investigation of spatial scan statistics or STSS in transportation science is a very recent research interest. Huang et al. (2009) use spatial scan statistics to detect clusters of active transportation (i.e. walking, cycling). The geographical coordinates of people's walking/cycling habits are mapped, and regions of high and low concentrations of citizens who undertake active transportation are detected by spatial scan statistics. There is no research evidence that develops an STSS based method for NRC detection.

Limitations of Spatio-Temporal Clustering based on Significance Testing

There are two important limitations of significance testing based cluster detection using STSS, which are summarised as follows:

- A spatial cluster is usually characterised by a geometric figure like a circle (Kulldorff, 1997) or an ellipse (Kulldorff et al., 2006). A significant spatial region, however, may have an irregular shape (e.g. a disease which spreads via a river will not result in a circular cluster). In such cases, the detection of irregularly shaped clusters may increase the effectiveness of detecting statistically significant clusters (Assunção et al., 2006; Duczmal et al., 2007; Tango and Takahashi, 2005). However, there is lack of research on determining how the cluster's shape has changed over time (Patil and Taillie, 2004). This motivated researchers to model the growth and displacement of significant clusters. Iyengar (2004) modelled the growth and displacement of a spatio-temporal cluster, however, the shape of the cluster was only modelled for a square pyramid shape, which limits its usefulness for more complex scenarios (e.g. an NRC growing in two or more different directions).
- Scanning a large spatial area containing hundreds of regions becomes computationally infeasible (Neill and Moore, 2004) as it requires analysis of many overlapping spatial regions. It should be noted that the run time depends on the chosen maximum spatial (and spatio-temporal) window size, which determines the number of searched regions. In spatial scan statistics a spatial window can contain at most 50% of the population (Kulldorff et al., 2009). However, most of the existing literature that uses spatial scan statistics does not mention the maximum spatial and temporal window size, and this choice has been shown to exert important effects on the results (Pfeiffer et al., 2008, p.54). Presumably, this is because most research relies on the default settings provided in the SatScan™.

Last, but not least, most of the spatial scan statistics and STSS methods have been developed using discrete distributions to model the null hypothesis like Poisson or Binomial. Although these issues have been recognised previously (Patil and Taillie, 2004), not much research attention has been given to develop STSS approaches to model a lognormal distribution, which is commonly used in transportation science to model journey times (Arezoumandi, 2011; Dandy and McBean, 1984; Hollander and Liu, 2008).

2.5. Evaluating the Performance of an NRC Detection Method

This section examines the strategies to evaluate the performance of an NRC detection method. In this way, it would be possible to compare the performance of different NRC detection methods. The previous sections introduced how NRC detection tasks can be classified as a topic under 'spatio-temporal clustering'. Therefore, this section investigates the evaluation criteria that are commonly used in clustering methods. Initially, cluster evaluation criteria (also referred to as cluster validation indices) are classified into three groups: internal, external and relative (Brun et al., 2007; Halkidi et al., 2001), but most of the current literature actually uses only the internal and external categories, as the distinction between internal and relative criteria is not well-defined (Kremer et al., 2011; Y. Liu et al., 2010).

Internal criteria measure the goodness-of-fit of the data to the resulting clusters. Therefore, the data that are used for clustering are also used for evaluating the detected clusters (Milligan, 1981). These criteria rely on estimating the two characteristics of the detected clusters: how well are the clusters separated from each other and how distant are the observations from one another within each cluster. These two characteristics are commonly referred to as separation and cohesion (or compactness) respectively. The preferable clustering method would result in well-separated clusters and the observations within each cluster would be similar. Tens of internal cluster evaluation criteria have been proposed to formalise this objective, and researchers feel the necessity to compare the effectiveness of these different criteria (Arbelaitz et al., 2013; Y. Liu et al., 2010).

Although internal criteria are commonly used to evaluate clustering methods that do not contain temporal domain, their application in the spatio-temporal domain poses a difficulty: defining a similarity function between the observations of a data set. In spatial clustering, common distance functions like Euclidean distance can be used to define the similarity between two observations. However, incorporating time into this similarity function usually requires a weighting between the distance in spatial and temporal domains (Kisilevich et al., 2010). For example, previous research on finding similar traffic speeds on a road network required three parameters: spatial constraints, similarity constraints and temporal window size. Spatial constraints determine whether two traffic sensors are spatially close, similarity constraints determine whether the estimated speed values of two sensors are similar and temporal window size determines the number of consecutive estimates that are considered when detecting clusters (Wei and Peng, 2009). However, there is a lack of research on determining the values of these parameters.

External criteria are used to evaluate a clustering method by comparing the detected clusters with the pre-specified clusters. The pre-specified clusters are regarded as 'true' (real) clusters and the data set that contains such pre-specified clusters is referred to as '*ground truth data*'. External criteria assess the overlap between the detected clusters and 'true' clusters by using measures such as the Jaccard coefficient (Brun et al., 2007) or precision and recall (Kremer et al., 2011). The aim of an NRC detection method is to report NRCs so that the reported NRCs match the ground truth data.

Ground truth data can be obtained in the following three ways:

- Manual labelling: a domain expert can manually label the traffic data and identify the links and times that experienced NRC. Previously, Guralnik and Srivastava (1999) relied on manual labelling to generate ground truth data for the purpose of detecting the times when traffic patterns change based on traffic flow data. However, they noted that experts may have disagreements regarding how they label the data. Sen et al. (2012) also rely on expert knowledge to generate ground truth data manually for the task of categorising traffic patterns into two, namely free-flow and congested. However, their research was limited to a single road. This is because manual labelling of traffic data is usually a labour intensive task, which prevents its usage on a large network.
- Use of simulation frameworks. Simulations use traffic models to generate traffic data. However, simulations often make simplistic, but unrealistic assumptions like limiting the number of incidents that can occur on a link (Emmerink et al., 1995). In commercial simulation packages, which are usually more acceptable in the scientific domain, these assumptions are even unknown to users. This, in turn, might lead to two simulation packages reporting different outcomes for the same purpose (e.g. estimating travel times in Bloomberg and Dale (2000)). Furthermore, there is a lack of research on modelling a large urban road network by a simulation framework that is capable of generating NRCs.

- Collecting it as a different data source. Research on AID is a good example of this, because researchers use an incident data set as ground truth data to evaluate the performance of their AID methods (Martin et al., 2001). Although incident data sets are commonly used to evaluate the performance of AID methods, they have also been reported to have inconsistencies, especially regarding the beginning and end times of incidents (Šingliar and Hauskrecht, 2009). This suggests that even ground-truth data may not be completely reliable.

The 'true' NRCs defined in ground truth data and the NRCs 'detected' by an NRC detection method can be represented in terms of the Venn diagram shown in Figure 2-7.

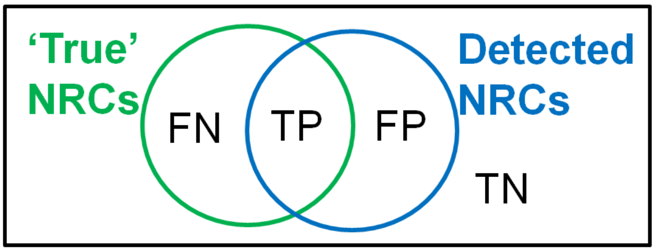


Figure 2-7 Comparison of 'True' NRCs with the 'Detected' NRCs

- FN denotes the total number of LJTs that are reported to belong to a 'true' NRC, but do not belong to a 'detected' NRC,
- TP denotes the total number of LJTs that are reported to belong to both to 'true' and 'detected' NRCs,
- FP denotes the total number of LJTs that belong to a 'detected' NRC, but not to a 'true' NRC,
- TN denotes the total number of LJTs that are not reported to belong to an NRC in both 'true' and 'detected' NRCs.

The comparison of the 'true' and 'detected' NRCs allows the calculation of the following commonly used criteria:

- Detection Rate (DR) is the ratio of correctly detected LJTs that belong to a 'true' NRC over the total number of LJTs that belong to 'true' NRCs. It is calculated by $\frac{\mathrm{TP}}{\mathrm{TP+FN}}$.
- False Alarm Rate (FAR) is the ratio of the wrongly classified LJTs that belong to an NRC over the total LJTs that were detected as belonging to an NRC. It is calculated

by $\frac{\mathrm{FP}}{\mathrm{TP+FP}}$. It is worth highlighting that FAR can take different meanings in different contexts and some researchers calculate FAR as $\frac{\mathrm{FP}}{\mathrm{TP+FP+FN}}$ (Parkany and Xie, 2005, p.53).

- False Negative Rate (FNR) is the ratio of total LJTs that belong to a 'true' NRC, but not a 'detected' NRC over the total number of LJTs belonging to a 'true' NRC. It is calculated by $\frac{\mathrm{FN}}{\mathrm{TP+FN}}$.

Given these definitions, an NRC detection algorithm aims to maximise DR while minimising FAR. This is a challenging task due to the relationship between the two performance criteria: most of the time, as DR increases, FAR also increases. For example, in the extreme case of reporting all LJTs that belong to an NRC, the method will result in a DR of one, as all NRCs would be detected, but also the maximum possible value of FAR. On the other hand, if an NRC method is very conservative and does not report any LJT as belonging to an NRC, then the DR will be zero, and so will the FAR (Martin et al., 2001).

The main limitation of external criteria is that in most real-life situations it is very difficult to obtain ground truth data. For example, Bar-Gera (2007) could not identify the method that measures travel time most accurately due to the lack of ground truth data on travel times. Even if there is ground truth data, it is often imperfect due to data collection process, which often involves subjective judgment of traffic operators. In order to avoid this, Šingliar and Hauskrecht (2009) suggest relabeling incident data sets so that the recorded incident times are changed without manual intervention. However, such a procedure assumes that all the incidents have already been recorded in the incident data set, which might not be the case. Furthermore, criteria like DR and FAR treat every NRC as being equally important. For example, failure to detect a mild NRC receives the same contribution as failing to detect a major NRC towards the calculation of FAR. Therefore, it is common for researchers not to rely on external criteria entirely and propose their own criteria. For example, Petty et al. (2000) propose a cost-benefit analysis of AID methods instead of comparing AID methods based on external criteria like DR and FAR.

Relative criteria are defined based on the consistency of the clustering methods. They require a comparison of the clusters obtained by the same method by varying the parameters of the method. The distinction between relative criteria and internal criteria is not well defined, and some researchers do not even consider relative criteria to be a different clustering evaluation strategy (Kremer et al., 2011; Y. Liu et al., 2010). It is also observed that some researchers use the term 'relative criteria' for measures like Dunn's index or Silhouette index (Vendramin et al., 2010), which are usually classified under internal criteria (Arbelaitz et al., 2013; Kremer et al., 2011).

Last, but not least, it is also common for traffic scientists to propose new performance evaluation criteria. This is especially true in AID, where an important measure is introduced: Mean Time to Detect (MTD) is used to assess the duration it takes for an AID method to detect incidents. Furthermore, Karim and Adeli (2002) propose a new performance measure called 'portability' to evaluate an AID method. This measure is motivated by the fact that most AID methods require parameter calibration, which makes them practically unfeasible with reference to a large motorway management system. Hence, 'portability' favours the method requiring the least recalibration when an AID method is applied in different contexts. Browne et al. (2005) propose two new performance measures to evaluate the accuracy of a method when correctly identifying the end of an incident. These measures are called 'nuisance rate' and 'false normal rate'. The nuisance rate measures how often an AID method detects the same incident more than once. The false normal rate measures how often an AID method prematurely detects normal circumstances following the initial detection of an incident. Although these examples assume the availability of ground truth data, they are useful in demonstrating that the proposal of new criteria to evaluate an event detection method is also common in the literature.

2.6. Summary

This chapter has provided the necessary background and literature to develop the methodology of the thesis. It is structured in two main parts: the first part focused on NRC detection. The lack of research on NRC detection led to the investigation of a closely related area: incident detection. The second part explained spatio-temporal clustering and provided evidence as to why it can be useful for NRC detection. The key findings from both of these parts are summarised as follows:

In NRC detection it is observed that;

- Two important distinctions between AID and NRC detection have been recognised. Firstly, planned events like concerts or football matches are usually *not* considered to be incidents and are hence not focused upon by AID methods. However, they may also cause an NRC (Dowling et al., 2004; Kwon et al., 2006). Thus, an NRC detection method also aims to detect NRCs caused by such planned events. Secondly, an AID method seeks to immediately detect incidents in order to assist traffic operators surveilling a road network; hence, designed to be used in *real-time*. On the other hand, an NRC detection method is designed to operate *off-line,* in which the purpose is to detect the NRCs that occurred in an investigated time period. In this way, the cause(s) of the detected NRCs could be investigated, and the impact of incidents could be quantified.
- Most of the existing research focuses on motorways, where incidents are considered to be the main cause of an NRC. The differences between an urban road network and motorway, and also the aforementioned two distinctions, prevent knowledge developed within AID methods from being applied to NRC detection.
- Although most of the published AID methods are reported to have a very high DR and very low FAR (for example, the results presented in Jeong et al. (2011) and Roy and Abdulhai (2003)), their application in the real life context is limited (Williams and Guin, 2007). This is probably due to overfitting the developed methods to certain circumstances, which is usually defined by a simulation framework, without considering the issues that arise in a real-life context.
- Traffic data types that are collected from inductive loops (e.g. occupancy or traffic flow) prevent a network wide analysis. This is because inductive loops provide traffic data on a short section of a road and it is cost prohibitive to install an inductive loop to cover a large urban road network. Network-wide analysis requires traffic data types to be collected along a length of a road (Hall, 2001). Link Journey Time (LJT) is found to be such a traffic data type, with some other additional advantages. It is regarded as the main indicator of congestion (OECD, 2007, p.43), and it can be aggregated in space (Wardrop, 1952; D'Angelo et al., 1999).

In spatio-temporal clustering, it is observed that:

- The occurrence of traffic congestion on motorway networks and urban road networks have been investigated by using spatio-temporal clustering. However, most of the existing researches are not able to differentiate RC and NRC. In addition, their application to a large urban road network is unfeasible. Thus, successful usage of spatio-temporal clustering for NRC detection on a large urban road network requires further research.
- There are two main strategies of spatio-temporal clustering, one is based on a similarity function and the other based on significance testing. Similarity based clustering relies on a similarity function which is defined between observations, whereas significance testing based clustering uses hypothesis testing to determine statistically significant clusters. Space-time scan statistics (STSS) is a state-of-art method to detect statistically significant clusters that allow a researcher to adjust for multiple hypothesis testing. However, STSS do not inform a researcher of the development of a significant cluster in space and time, as the focus is usually on the most significant cluster. Furthermore, STSS detects only fixed geometrically shaped clusters, whereas recent research has illustrated the importance of detecting arbitrarily shaped clusters (Duczmal et al., 2007; Tango and Takahashi, 2005). Last, but not least, as the study area becomes larger, calculation of the likelihood ratio scores of many overlapping space-time regions becomes computationally intractable (Neill and Moore, 2004).
- There are two main strategies to evaluate the performance of a clustering method. The first strategy is based on calculating commonly used performance criteria like DR and FAR. This strategy relies on data that is assumed to be error-free and represent reality, which is referred to as ‘ground truth data’. The main advantage is these criteria are commonly used by most researchers. However, ground truth data may not always be available. The second strategy, therefore, evaluates the quality of the detected clusters. Specifically, the evaluation focuses upon how well the observations within a cluster are connected and how well the clusters are separated.

3. SPATIO–TEMPORAL CLUSTERING TO DETECT NRC

This chapter proposes two novel methods for NRC detection from LJT data on an urban road network. Both of these methods is a realisation of a ‘spatio-temporal clustering’ strategy. As discussed in section 2.4, spatio-temporal clustering could be achieved via two main strategies. The first strategy is clustering by defining a similarity function between observations. The second strategy is clustering by significance testing. Although spatio-temporal clustering has previously been investigated for the detection of congestion (Anbaroglu and Cheng, 2011; Kerner and Konhäuser, 1994; Li et al., 2007), none of the previous research has addressed NRC detection in an urban road network consisting of hundreds of links. This chapter, therefore, proposes two novel methods for each of the spatio-temporal clustering strategies for NRC detection.

The aim of the proposed methodology is to identify NRCs accurately (i.e. recognition of all NRCs and no incorrect interpretations of recurrent events in traffic as NRCs) that occurred on a given date of analysis. The identified NRCs should allow a researcher to observe the development of an NRC in space and time. In this way, the dynamic nature of an NRC can be captured, and each NRC can be quantified based on its impact. The developed methodology also addresses how to evaluate the performance of an NRC detection method. Only after an evaluation procedure can the relative advantages and limitations of NRC detection methods be understood.

The methodological framework of this thesis is illustrated in Figure 3-1. Section 3.1 explains the necessary inputs to the developed methodology on NRC detection. Section 3.2 proposes a similarity based spatio-temporal clustering method for the detection of NRCs. Section 3.3 proposes an STSS based spatio-temporal clustering method for the detection of NRCs. The NRCs detected by these two different methods are denoted as $\mathbf{Z}_{\mathrm{CE}}$ and $\mathbf{Z}_{\mathrm{STSS}}$ respectively. Finally, section 3.4 summarises the chapter with a qualitative comparison of the two proposed NRC detection methods. Evaluation of the performance of NRC detection methods is discussed in Chapter 4.

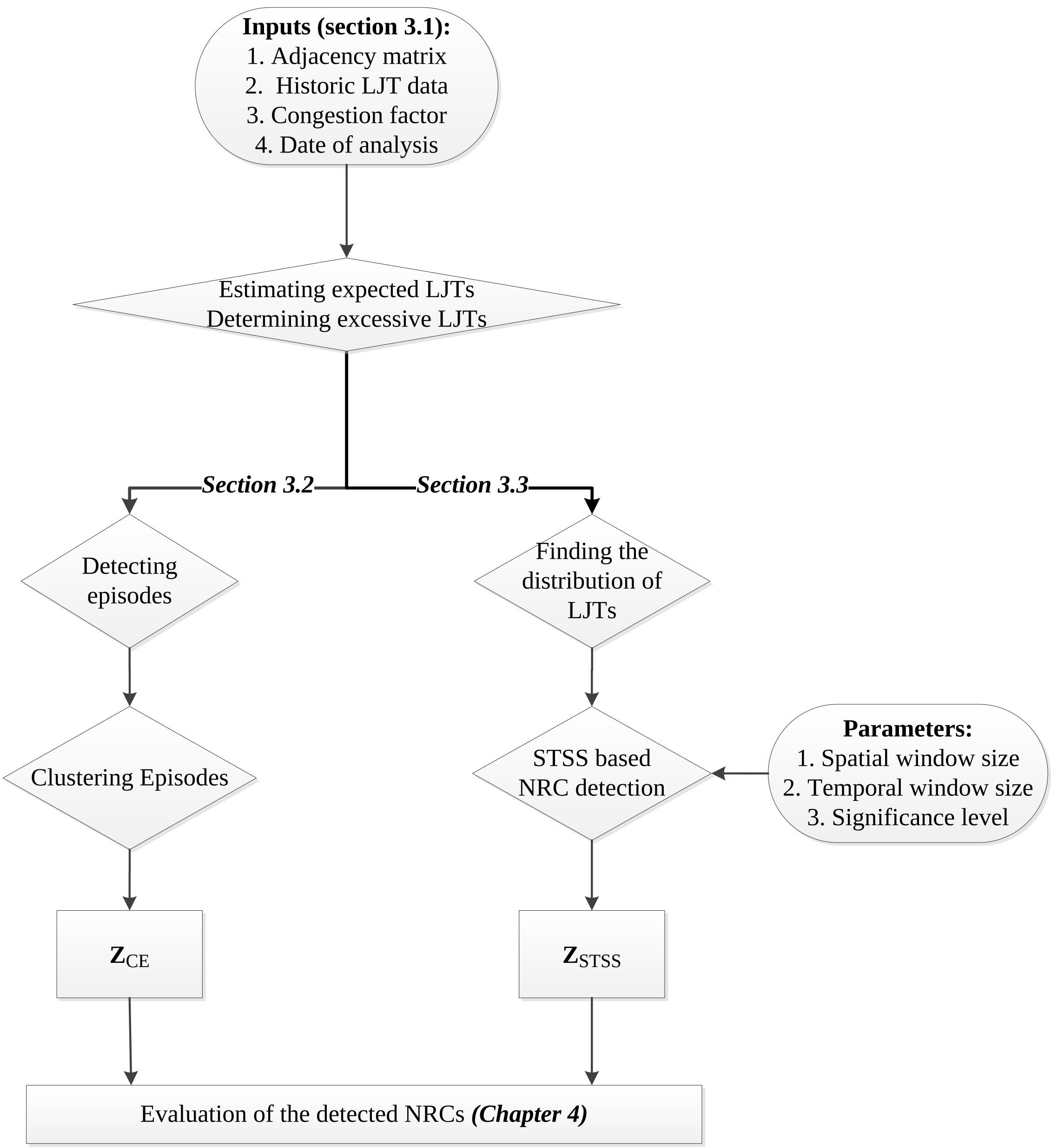


Figure 3-1 Spatio-temporal clustering to detect NRCs based on LJT data

3.1. Required Inputs for NRC Detection in Urban Road Networks

There are four main inputs to the developed methodology for NRC detection in urban road networks. In order to provide a complete understanding of these four inputs, the main data source of this thesis should be mentioned, which is Link Journey Time (LJT). The wide usage of ANPR cameras for traffic enforcement purposes allows traffic operation centres to estimate aggregate LJT from the vehicle journey times that are obtained from ANPR cameras. Although the procedure to estimate LJTs is not within the scope of this thesis, a brief introduction regarding the estimation process has been provided in section 2.2. Because LJT is a prominent traffic data type in network-wide analysis, this thesis uses LJT as the main traffic data type for NRC detection, where NRCs are considered to be a cluster of high LJTs. Thereon, a road network can be characterised by a directed graph that consists of a set of nodes (i.e. **N**) corresponding to ANPR cameras, and a set of arcs (i.e. **A**) corresponding to links. Link a is defined with two nodes, n_a(from) and n_a(to), where traffic on link a is flowing from n_a(from) to n_a(to).

The four main inputs, which are used in both of the proposed NRC detection methods, are described as follows:

Adjacency matrix **(M)** is the mathematical representation of a geographical road network. It is a |**A**| × |**A**| matrix, where |**A**| corresponds to the number of links. The value M(a, b) denotes the *adjacency relationship* between the two links, a and b. This thesis defines two links, a and b, to be *adjacent* if the ending node of link a is the beginning node of link b. The *adjacency relationship* is formally defined as; $\mathrm{M}(a, b) = 1$ iff $n_a(\text{to}) = n_b(\text{from}), \ \forall\, a, b \in \mathbf{A}$. In addition, a link is considered to be adjacent with itself. Formally, $\mathrm{M}(a, a) = 1, \forall\, a \in \mathbf{A}$.

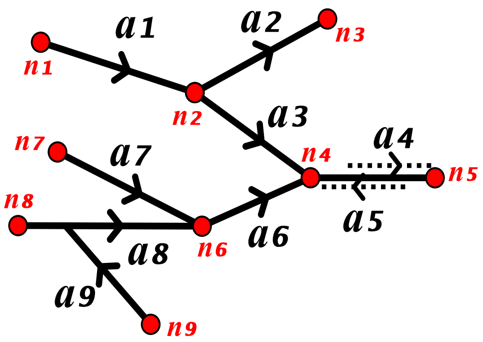
The main reason to define the adjacency relation as presented is related to the cause of an NRC, which is usually considered to be an incident. Because an incident decreases the capacity of the link where it occurred, the reduction in the capacity of that link could cause a vehicular queue to grow towards its upstream link (Ozbay and Kachroo, 1999, p.8; Weil et al., 1998). Therefore, the impact of an NRC is observed first on the link where the incident occurred, and then a vehicular queue grows towards the upstream of the link. The stated adjacency relationship aims to capture how an NRC grows spatially; however, it makes some simplifying assumptions regarding the following issues that arise in an urban road network:

- *Overlapping links* occur when two links share a common part of the road network. They are observed due the lack of ANPR cameras and due to the spatial distribution of ANPR cameras. The possible situations where two links overlap are illustrated in Figure 6-1. Even though an incident happening at a location where two links overlap might increase LJTs of both links causing an NRC, two overlapping links are not considered to be adjacent if they do not possess the adjacency relationship.
- *Anti-parallel links* occur as a special case of overlapping links, in which one link's beginning/ending is another link's ending/beginning. Formally, link a and link b are

anti-parallel links if $n_a(\text{to}) = n_b(\text{from})$ and $n_a(\text{from}) = n_b(\text{to})$. Most of the time, anti-parallel links collect traffic data on the same road, but for opposite directions of traffic flow. Even though anti-parallel links possess the adjacency relationship, this thesis do not consider them to be adjacent based on the empirical evidence provided in section 5.2.3, which suggests that the majority of incidents that occur on either of the anti-parallel links did not affect the other link.

- Link lengths vary substantially due to the heterogeneous distribution of ANPR cameras (i.e. more ANPR cameras in a city centre result in shorter links in the city centre). Therefore, an incident occurring on a shorter link is more likely to affect its adjacent links than an incident occurring on a longer link. However, this thesis does not consider link lengths when defining the adjacency relationship.

In order to demonstrate the adjacency relationship between links, a road network illustrated in Figure 3-2(a), where links are denoted with black letters and nodes (i.e. the beginning and end of a link) are illustrated with red letters. The traffic flow direction is indicated with an arrow, and the absence of arrow indicates that traffic is flowing on both directions. The adjacency relationships of this network are illustrated in Figure 3-2(b), where 'From' and 'To' columns indicate the beginning and end of a link respectively. Finally, 'Adjacent of' column indicates which links are adjacent of a given link, and it is synonymous to the adjacency matrix (e.g. M(a1, a2) = 1).



(a)

Links	**From**	**To**	**Adjacent of**
a_1	n_1	n_2	$a_{1,}$ $a_{2,}$ a_3
a_2	n_2	n_3	a_2
a_3	n_2	n_4	$a_{3,}$ a_4
a_4	n_4	n_5	a_4
a_5	n_5	n_4	a_5
a_6	n_6	n_4	$a_{4,}$ a_6
a_7	n_7	n_6	$a_{6,}$ a_7
a_8	n_8	n_6	$a_{6,}$ a_8
a_9	n_9	n_8	$a_{8,}$ a_9

(b)

Figure 3-2 Sample road network (a) and its adjacency relationships (b)

Links a4 and a5 are anti-parallel links as the beginning/ending of link a4 is the ending/beginning of link a5. Links a8 and a9 are overlapping, as some part of the road network is shared by both of the links. Because, link a8 begins from the node where link a9 ends, link a9 is considered to be adjacent of link a8.

A hierarchical structure can be built upon the adjacency relationships. First-order adjacent links are directly connected with each other, whereas there is another link to connect two second-order adjacent links and so on and so forth. For example, links $a1$ and $a3$ are first-order adjacent links, whereas links $a1$ and $a4$ are second-order adjacent links that are connected through link $a3$.

Historic LJT data are past LJT estimates that can be used for statistical analysis. The proposed NRC detection methods use historic LJT data in two aspects, firstly to estimate expected LJTs which are considered to be LJTs in normal traffic conditions; and secondly to determine the statistical distribution of LJTs. As LJTs are estimated by observations collected from ANPR cameras which work in an automated fashion, there is no extra operational cost to collect historical LJT data.

***Congestion factor* (*c*)** is a real valued number multiplied with the expected LJTs to determine the threshold to identify whether an LJT is excessive. It is denoted as c, where $c \geq 1$. 'Congestion factor' is a commonly used concept in traffic operation centres to classify LJTs into distinct congestion levels like moderate, serious or severe congestion (OECD, 2007, p.57; TfL, 2010, p.95; Transport Canada, 2007, p.9). Determination of a congestion factor usually depends on expert knowledge.

Date of analysis determines the day on which NRCs will be detected. Any date where LJT data are collected can be used for this input. The aim is to report the NRCs that occurred within the analysed date to traffic operators.

These four main inputs are the necessary inputs to the two proposed NRC detection methods. Using these inputs, two important measures can be determined:

- ***Expected LJT*** is the mean travel time to occur on a given link a at time interval t under normal traffic conditions. It is denoted as $\overline{y}_a(t)$ and measured in minutes. Expected LJT capture the recurrent nature of traffic. Historic LJTs are used to determine the expected LJTs, which allow a researcher to focus on NRC (Hallenbeck et al., 2003a; Varaiya, 2007). However, the estimation of expected LJTs causes two main issues. Firstly, the definition of the term *expected* might take different forms in different contexts. For example, a substantial increase in LJTs due to a planned event (e.g. engineering work or a concert) is *expected* by traffic operators, but it is *unexpected* by commuters who are unaware of that event. Secondly, there are numerous time series based methods to estimate expected LJTs, ranging from simple averaging to a more complicated weighted averaging to kernel regression. In order to alleviate the difficulty arising from these issues, domain knowledge is usually required to estimate the expected LJTs.

- ***Excess LJT*** is the difference between the observed and expected LJT if the observed LJT is higher than a threshold which is determined by multiplying the congestion factor with the expected LJT. Thus, if the estimated LJT is higher than the threshold, then it is excessive (substantially high). Formally, given a congestion factor *c*, the value of the excessive LJT at link *a* time interval *t* is defined as follows:

$$\delta_a(t) = \begin{cases} y_a(t) - \overline{y}_a(t), & \text{if } y_a(t) > \left(c \times \overline{y}_a(t)\right), c \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The special case of excessive LJTs occurs when c=1, in which case this thesis uses the term '***delay***' instead of excessive LJT. Excessive LJTs are the main component of an NRC, because the estimated LJT is substantially higher than its expected value. The simplifying assumption in determining excessive LJTs is that the congestion factor is the same value for all the links and times. This assumption, however, is in-line with the operational uses in a traffic operation centre.

3.2. Clustering Episodes for NRC Detection

This section explains how similarity based spatio-temporal clustering can be used to detect NRCs based on LJT data. A similarity measure is usually applied *between* the observations, and similar observations are clustered. For example, Anbaroglu and Cheng (2011) investigate the beginning and end of congestion events by determining similarities based on the sharp changes within consecutive LJTs on different links. A sharp change is determined empirically and an adjustment of more than 20% between consecutive LJT observations is regarded as a sharp change. Trajectory data have also been used to find congested routes on an urban road network based on the similarities between the trajectories of vehicles. If a number of trajectories are close in space and time, then a congestion event is detected (Li et al., 2007; Ying et al., 2009). The main limitation of these approaches, however, is that they cannot differentiate whether detected patterns belong to an RC or an NRC.

This section, therefore, takes a different approach to defining similarities on an urban road network. The proposed method defines a similarity between the estimated LJT and its expected value. In this way, it is aimed to focus on NRCs rather than congestion events. The proposed method consists of two steps, detecting episodes and clustering episodes, which are discussed in subsections 3.2.1 and 3.2.2 respectively. In subsection 3.2.3, the process to determine the impact of an NRC is introduced. Finally, subsection 3.2.4 provides an example to illustrate how clustering episodes method can be used to detect NRCs.

3.2.1. Detecting Episodes

An episode is defined as a maximal interval on a link during which all the LJTs are excessive. Formally, an episode is defined on link *a* as the time interval between *m* and *n*, as follows:

$$e\{a\} = \left\{ ([m\,n]) \left| \begin{array}{ll} \delta_a(t) > 0 \ \forall t \ \in [m\,n], n \ \geq \ m & \text{and} \\ \delta_a(m-1) = \delta_a(n+1) = 0 & \text{and} \\ a \in \mathbf{A}; \ \ t \in \{1, 2, \dots, T\} & \end{array} \right. \right\} \quad (5)$$

The set of links is denoted with **A**, and the total number of LJTs within the analysis period is denoted with *T*. The first requirement states that during an episode all the LJTs are excessive. The second requirement states that an episode is a maximal interval of consecutive excessive LJTs. Thus, the former/latter LJT from the beginning/end of the episode is not excessive. The number of episodes on a link varies from zero where no LJT is excessive to $\lceil T/2 \rceil$ where every other LJT is excessive.

Each episode can be quantified with two measures: *duration* and *severity*.

- The duration of an episode is the time lapse between the beginning and end times of the episode. It is calculated by $\mathrm{d}(e\{a\}) = (n - m + 1) \times \omega$, where $m, n \in e\{a\}$ and ω is the interval to estimate two consecutive LJTs in minutes. Thus, the duration of an episode is measured in minutes. The reason for adding one is that the beginning and end times are included within the definition of an episode *inclusively*.
- The severity of an episode is the total excess LJT that the episode exhibits. It is measured in minutes. The severity of an episode is calculated by $\mathrm{s}(e\{a\}) = \sum_{t=m}^{n} \delta_a(t)$, where $m, n \in e\{a\}$.

In order to illustrate the importance of the congestion factor, suppose that a traffic operation centre evaluates three different congestion factors. These congestion factors are multiplied with the expected LJTs to calculate the thresholds. These thresholds are used to identify whether an LJT is excessive or not. The LJT thresholds defined by the high, medium and low congestion factors are denoted as High c, Medium c and Low c respectively in Figure 3-3. Having defined these thresholds, after applying the definition of an episode, it is possible to identify all the episodes that occurred on a link. If the threshold is defined by the medium congestion factor, then four episodes are identified (denoted with *e1*, *e2*, *e3* and *e4*), as shown in Figure 3-3.

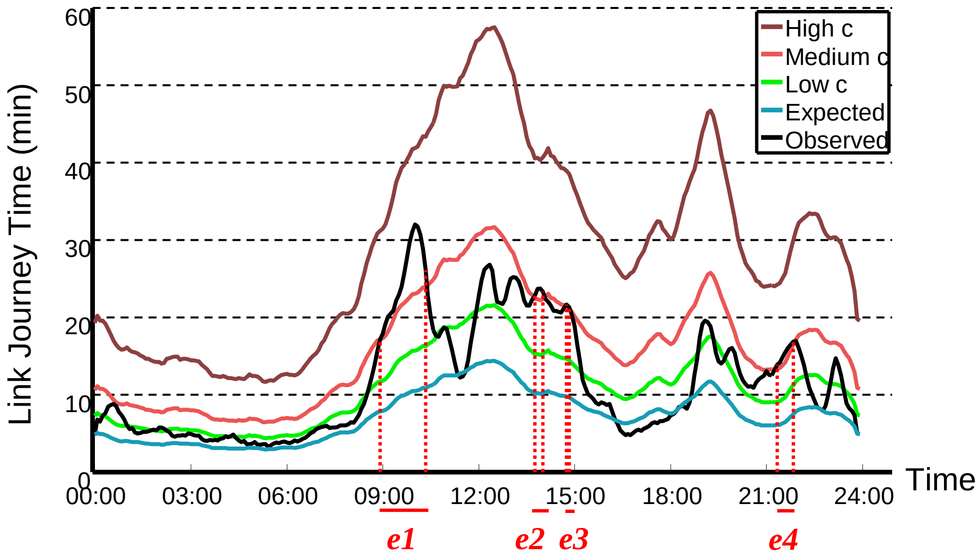


Figure 3-3 Detection of episodes using the medium congestion factor

Two observations can be made regarding the detected episodes:

- The congestion factor is the main criterion to determine an episode. If the congestion factor is increased to the high congestion factor, then no episodes are detected. If the congestion factor is reduced to the low congestion factor, then the second and third episodes would merge. The duration and severity of the resulting episode will be longer.
- Episodes can be prioritised based on their duration or severity or both. For example, given these four episodes, because the first episode had the longest duration and highest severity, its investigation can be prioritised amongst other episodes.

The example illustrated in Figure 3-3 is based on a single link. However, an urban road network usually comprises many links. Therefore, an NRC detection method should be able to detect NRCs that may occur on more than one link. Achieving this objective is the focus of the next subsection.

3.2.2. Clustering Episodes to Detect NRCs

Episodes can be used to detect NRCs that span any number of adjacent links. This is because all the LJTs that are included in an episode are much higher than their expected values, indicating that they are likely to belong to an NRC. The extent to which an LJT is higher depends on the congestion factor. Therefore, after determining a congestion factor, an episode can be considered to be *similar* to an NRC pattern.

This section describes the procedure to detect NRCs by clustering *spatio-temporally overlapping* episodes. Suppose that the two episodes are detected on links a and b, which are denoted with indices i and j as $e_i\{a\}$ and $e_j\{b\}$ respectively. These two episodes are spatio-temporally overlapping if they satisfy the following two conditions:

- $\mathrm{M}(a, b) = 1,\ \forall\, a, b\ \in \mathbf{A}$.
- $\exists t_i, t_j \middle| t_i = t_j$, where $t_i \in\ e_i\{a\}$ and $t_j \in\ e_j\{b\}$ and $1 \leq\ t_i, t_j\ \leq T$.

The first condition states that the episodes should occur at adjacent links. The second condition states that there is at least one time interval that is common in both of the episodes. These two conditions are required to cluster two episodes. By clustering spatio-temporally overlapping episodes, NRCs that span multiple links across a time period could be detected. The simplifying assumption, however, is that if two episodes are spatio-temporally overlapping, then they belong to the same NRC, even though their underlying causes are different. To clarify, suppose that two accidents happened at the same time on adjacent links and caused an episode on both of the links. These two episodes would be considered to belong to a single NRC, even though their causes are different. Therefore, an NRC could be a consequence of several distinct incidents.

The definition of an episode is applied to the entire road network across the study period to determine all the episodes, which is denoted with the set **E**. These episodes are converted into NRCs with the pseudocode illustrated in Figure 3-4[5], where the set of NRCs is denoted with **Z**.

---

[5] The visual appearance of the pseudocode is achieved by using the 'pseudocode' Latex package (Kreher and Stinson, 1999).

Algorithm 1: CLUSTER EPISODES(**E**)

```
Initialize:
Z ← ∅

for all i ∈ E                                                    (1)
  do {
    if addedPreviously(i)
      then continue − (1)
    else {
      Create a new NRC k by adding the episode i                 (2)
      k ← i
      Z ← Z ∪ k

      found ← 1
      while (found)                                              (3)
        do {
          found ← 0
          for all j ∈ E                                          (4)
            do {
              if addedPreviously(j)
                then continue −(4)
              if (j ⋈ k)
                then {
                  Add episode j to NRC k
                  k ← k ∪ j
                  found ← 1
```

Figure 3-4 Pseudocode of the algorithm to cluster episodes to detect NRCs

The set of NRCs (i.e. **Z**) is initialised to contain no episode. The algorithm starts to search all the episodes with step (1). If the analysed episode *i* has already been added to an NRC, then the algorithm proceeds with the next episode. If the analysed episode has not been added to an NRC, then a new NRC *k* is created with the episode *i* being its first component with step (2). The 'while' loop starting with step (3) ensures that the newly created NRC *k* contains all the episodes that are spatio-temporally overlapping. In order to achieve this, all the remaining episodes are searched. If an episode *j* is found to spatio-temporally overlap with any episode contained within the NRC *k* (assessed with symbol $\bowtie$), then the episode *j* is added to NRC *k*. This operation may extend the spatio-temporal extent of NRC *k*. Thus, it is necessary to search for all the episodes one more time starting with step (3) in order to ensure that the NRC *k* reaches its maximum spatio-temporal extent before creating a new NRC.

3.2.3. Determining the Evolution of NRCs

The NRC detection algorithm described in Figure 3-4 determines all NRCs such that no two episodes that belong to different NRCs are spatio-temporally overlapping with one another. However, the detected NRCs do not inform a researcher of the temporal *evolution* of an NRC. The temporal evolution of an NRC allows a researcher to observe which links were congested and when in an NRC. Determination of the evolution of an NRC allows a researcher to quantify an NRC. The process to determine the evolution of the detected NRCs is described in Figure 3-5.

Process 1: $\textsc{evolution}(\mathbf{Z})$

Initialize:

$\mathbf{L} = \{L_1, L_2, \ldots, L_{|\mathbf{Z}|}\} \leftarrow \emptyset$

for all $k \in \mathbf{Z}$ (1)

$$\textbf{do}\ \begin{cases} \textbf{for all}\ e\{a\} \in k & (2) \\ \textbf{do}\ \begin{cases} \textbf{for all}\ t \in e\{a\} & (3) \\ \textbf{do}\ \{L_k(t) \leftarrow L_k(t) \cup a \end{cases} \end{cases}$$

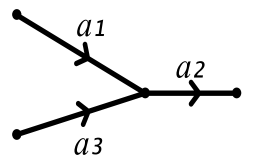
Figure 3-5 Determination of the evolution of NRCs

The set L_k represents the evolution of k^{th} NRC. L_k is a time-series of links that belong to the k^{th} NRC at each time interval of the NRC. Representation of each NRC in its evolution allows a researcher to quantify the impact of an NRC with the two characteristics that are analogous to quantify an episode: lifetime and severity.

- ***Lifetime*** is the temporal extent of an NRC. It represents the time period that an NRC contains at least one link. The lifetime of NRC k is expressed in terms of the beginning and end time of an NRC. The duration of an NRC can be determined by converting the lifetime to a continuous time period.
- ***The severity*** of an NRC is the total impact of the NRC in terms of excessive LJTs. All the excessive LJTs are added throughout the lifetime of an NRC to calculate its severity. The severity of NRC k is calculated with the following formulation: $\sum_{a,t \in L_k(t)} \delta_a(t)$. Calculation of the severity of each NRC would lead to quantifying the impact of NRCs. In this way, NRCs that had high severities could be prioritised for further investigation into possible causes.

3.2.4. An Example of Clustering Episodes to Detect NRCs

This section intends to clarify the concepts described in the previous sections with an example. Suppose a three-link network and its adjacency matrix as shown in Figure 3-6, where the adjacency matrix is determined based on the discussion in section 3.1.



	a_1	a_2	a_3
a_1	1	1	0
a_2	0	1	0
a_3	0	1	1

(a) (b)

Figure 3-6 A simple road network (a) and its adjacency matrix (b)

Suppose that eight LJTs are estimated for these three links, and those which are excessive are illustrated with a '+' sign in Figure 3-7. After determining the excessive LJTs, three, two and two episodes are detected for links a_1, a_2 and a_3 respectively. These episodes are illustrated with orange rectangles in Figure 3-7(a). Three NRCs are found by clustering these episodes based on the algorithm described in Figure 3-4. The boundaries of these NRCs are highlighted with red in Figure 3-7(b).

Links \ Times	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
a_1	+	+	+	–	+	–	–	+
a_2	–	–	+	+	+	–	+	–
a_3	+	+	+	+	+	–	+	–

\+ Excessive LJT
– Not excessive LJT
Boundary of Episode

(a)

Links \ Times	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
a_1	+	+	+	–	+	–	–	+
a_2	–	–	+	+	+	–	+	–
a_3	+	+	+	+	+	–	+	–

k_1 k_2 k_3

\+ Excessive LJT
– Not excessive LJT
Boundary of NRC

(b)

Figure 3-7 Grouping episodes to detect NRCs

Several observations can be made regarding this outcome:

- The shape and size of NRCs may vary. The shape and size of an NRC cluster depends on the spatio-temporal distribution of the excessive LJTs. The first NRC contains 12 excessive LJTs, whereas the third NRC contains only one excessive LJT. As this thesis does not make any assumption regarding the nature of the phenomenon that resulted in an NRC, it is reasonable to observe variations in the shape and size of the detected NRCs.

- The second and third NRCs are not merged into one NRC, as none of the episodes within these NRCs spatio-temporally overlap. Although links a_1 and a_2 are adjacent, the time intervals where the episodes occurred do not align; hence, they are not spatio-temporally overlapping.
- The first NRC can only be detected through the usage of the 'while' loop in Figure 3-4, as it has a concave shape. Therefore, the 'while' loop is necessary to detect NRCs in the pseudocode of grouping episodes.

The evolution of an NRC cannot be traced directly from the output of the grouping episodes, as the algorithm iteratively clusters episodes that are spatio-temporally overlapping. This is because the algorithm does not cluster episodes based solely on their beginning times. Therefore, the evolution of clusters is determined using Process 1 (illustrated in Figure 3-5), and the results are illustrated in Table 3-1.

Table 3-1 Evolution of NRC clusters detected in Figure 3-7

NRC ID	**Evolution**	
	Times (t)	**Links**
k_1	1	a_1, a_3
	2	a_1, a_3
	3	a_1, a_2, a_3
	4	a_2, a_3
	5	a_1, a_2, a_3
k_2	7	a_2, a_3
k_3	8	a_1

By representing each NRC in its evolution, two characteristics can be identified to quantify an NRC. These two quantifiable measures are lifetime and severity. The lifetime of the first NRC is between times one and five, the second and third NRCs are times seven and eight respectively. The severity of the NRCs is calculated by adding the excessive LJTs within an NRC. For example, the severity of the second NRC is found by $\delta_{a_2}(7) + \delta_{a_3}(7)$.

3.3. Space-Time Scan Statistics to Detect NRCs

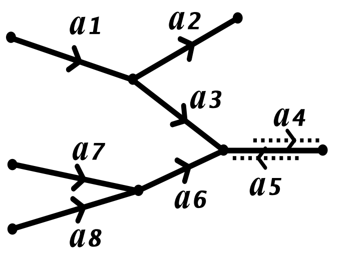
This section describes how Space-Time Scan Statistics (STSS) can be used to detect NRCs based on LJT data. An expectation based STSS is proposed for the purpose of NRC detection. The proposed STSS based method comprises of four steps, which are described in detail in the following subsections.

3.3.1. Generating Space-Time Regions

A space-time region (STR) is the aggregation of spatial regions in time, where links correspond to the spatial regions. In order to detect any NRC regardless of the number of links that it contains or its duration, it is necessary to scan an entire study area with overlapping STRs whose size and location varies. Most of the STSS based approaches use overlapping cylinders to scan the entire region. However, cylinders are not suitable to scan a road network as they cannot use the traffic flow direction information. Therefore, it is more reasonable to scan the road network by the network structure itself, which is captured in the adjacency matrix.

There are two important values that should be determined to generate all STRs. These are maximum spatial (i.e. ρ) and temporal window (i.e. τ) sizes. It was suggested that ρ should not exceed more than half of the study area (Kulldorff et al., 1998). However, scanning a large spatial area containing hundreds of regions becomes computationally unfeasible (Neill and Moore, 2004). This observation is also true for a large urban road network consisting of hundreds of links, where each link is represented by a time series consisting of hundreds of LJTs. Therefore, this thesis proposes to reduce the number of STRs that can be generated on a road network by considering the main difference between a road network and a point/areal data set: directional relationship. Firstly, traffic is flowing from one direction to the other. Secondly, an NRC is observed initially on the link itself and then it propagates to its first-order adjacencies. Considering these two points, spatial regions are created by only considering the link itself and its first-order adjacencies. The number of first-order adjacencies that are used to generate the spatial regions depends on the maximum spatial window size.

In order to illustrate how spatial regions are generated, suppose the road network given in Figure 3-8(a), whose adjacency matrix is illustrated in Figure 3-8(b) based on the discussion in section 3.1. Based on this adjacency matrix, and a given maximum spatial window size value (ρ), the spatial regions are illustrated as shown in Figure 3-8(c).



(a)

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
a_1	1	1	1	0	0	0	0	0
a_2	0	1	0	0	0	0	0	0
a_3	0	0	1	1	0	0	0	0
a_4	0	0	0	1	0	0	0	0
a_5	0	0	0	0	1	0	0	0
a_6	0	0	0	1	0	1	0	0
a_7	0	0	0	0	0	1	1	0
a_8	0	0	0	0	0	1	0	1

(b)

$\boldsymbol{\rho}$	**Spatial Regions**
$\rho = 1$	a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8
$\rho = 2$	$\{\rho = 1\} \cup \{a_1$-a_2, a_1-a_3, a_3-a_4, a_4-a_6, a_6-a_7, a_6-$a_8\}$
$\rho = 3$	$\{\rho = 2\} \cup \{a_3$-a_4-a_6, a_6-a_7-$a_8\}$
$\rho \geq 4$	$\{\rho = 3\}$

(c)

Figure 3-8 A road network (a), its adjacency matrix (b) and generated spatial regions (c)

Several observations can be made regarding Figure 3-8. Firstly, all spatial regions are generated based on the link itself and its first-order adjacencies, as each link starts having an NRC initially with itself and its first-order adjacencies. Secondly, the spatial regions are the same for ρ = 3 and ρ ≥ 4, where ρ corresponds to the maximum spatial window size. This is because there are no links with more than two first-order adjacencies. It should be highlighted that an NRC can be observed on more possibilities than the spatial regions illustrated in Figure 3-8(c), including $\{a_1 - a_3 - a_4\}$ or $\{a_3 - a_4 - a_6 - a_7 - a_8\}$. These possibilities are not represented by a spatial region, because the number of such possibilities increases dramatically if STRs are not generated by only considering the first-order adjacencies of a link, which would prevent the proposed NRC detection method from being applied to a large urban road network. Note that, even though these possibilities are not represented separately as a spatial region, they can be identified as belonging to an NRC with the modification that this thesis introduces to STSS which is discussed in section 3.3.4.

Generating STRs from this point is straightforward, as they are the three dimensional representation of spatial regions where the height of a spatial region varies between one and τ observations. STRs are determined, after applying this procedure to all spatial regions and for each observation during the analysis period.

3.3.2. Determining the Likelihood Ratio Function

In order to derive the likelihood ratio function, it is first necessary to have a model to describe LJTs. The evidence collected from the literature review (Arezoumandi, 2011; Dandy and McBean, 1984; Hollander and Liu, 2008) and the empirical analysis conducted in section 5.1 indicates that LJTs can be modelled by lognormal distribution. Thus, this section describes how an expectation based STSS can be derived by assuming that LJTs are distributed based on lognormal distribution.

The likelihood ratio function is derived from $\frac{f(\mathrm{H}_1|\mathrm{Data})}{f(\mathrm{H}_0|\mathrm{Data})}$, where $f(\mathrm{H}_1|\,\mathrm{Data})$ is the likelihood of observing the alternative hypothesis, and $f(\mathrm{H}_0|\,\mathrm{Data})$ is the likelihood of observing the null hypothesis.

The null (H_0) and the alternative hypothesis (H_1) are defined as follows:

- H_0 assumes that there is no NRC at an *STR*, and each LJT that is included within that *STR* is sampled from a lognormal distribution. Specifically, for a given *STR*, $\mathrm{H}_0(STR)\colon y_a(t) \sim \mathrm{LogNormal}(\mu_{a,t}, \sigma_{a,t})$ for all links a and times $t \in STR$. $\mu_{a,t}$ is referred to as the location parameter and is calculated by taking the mean of the natural logarithm of the historic LJTs. $\sigma_{a,t}$ is referred to as the scale parameter, and is the standard deviation of the natural logarithm of the historic LJT observations.
- H_1, on the other hand, assumes that a given *STR* belongs to an NRC. This can be modelled as an increase in the mean of the LJT. Formally, $\mathrm{H}_1(STR)\colon \mathrm{y}_a(t) \sim \mathrm{LogNormal}((\mu_{a,t} + q), \sigma_{a,t})$, for all links a and times $t \in STR$ for some positive constant $q > 0$, $q \in \mathbb{R}$. The reason for proposing an *additive* lognormal model is to build a linkage between the congestion factor parameter (c) in similarity based NRC detection and q. Specifically, the threshold to determine whether or not an LJT is excessive is calculated by $c \times \overline{y}_a(t)$ in similarity based NRC detection. The natural logarithm of this threshold is $\ln\big(c \times \overline{y}_a(t)\big)$, which is equal to $\ln(c) + \mu_{a,t}$. Therefore, q corresponds to $\ln(c)$.

In order to use such a likelihood ratio function, a simplifying assumption has to be made: each LJT contained within an STR is drawn independently from the associated lognormal distribution. Although LJTs possess temporal auto-correlation (Cheng et al., 2012), this assumption has to be made in order to calculate the likelihood ratio function efficiently. The reason for performing Monte Carlo simulations (discussed in section 3.3.3) is to reduce the impact of this simplifying assumption.

The likelihood ratio is calculated as follows:

$$
\text{Ƒ}(STR) = \frac{f(\mathrm{H}_1(STR)|\ \text{Data}\ )}{f(\mathrm{H}_0(STR)|\ \text{Data}\ )} = \frac{\Pr\left(\text{Data}\ |\ \mathrm{H}_1(STR)\right)}{\Pr\left(\text{Data}\ |\ \mathrm{H}_0(STR)\right)}
$$

$$
= \frac{max_{q>0} \prod_{a,t\ \in STR} \Pr\left(y_a(t) \sim \text{LogNormal}\left(\mu_{a,t} + q, \sigma_{a,t}\right)\right)}{\prod_{a,t\ \in STR} \Pr\left(y_a(t) \sim \text{LogNormal}\left(\mu_{a,t}, \sigma_{a,t}\right)\right)}
$$

$$
= \frac{max_{q>0} \prod_{a,t\ \in STR} \frac{1}{y_a(t)\sigma_{a,t}\sqrt{2\pi}} e^{-\frac{\left(\ln(y_a(t)) - (\mu_{a,t}+q)\right)^2}{2\sigma_{a,t}{}^2}}}{\prod_{a,t\ \in STR} \frac{1}{y_a(t)\sigma_{a,t}\sqrt{2\pi}} e^{-\frac{\left(\ln(y_a(t)) - \ \mu_{a,t}\right)^2}{2\sigma_{a,t}{}^2}}}
$$

$$
= \frac{max_{q>0} \prod_{a,t\ \in\ STR} e^{-\frac{\left(\ln(y_a(t)) - (\mu_{a,t}+q)\right)^2}{2\sigma_{a,t}{}^2}}}{\prod_{a,t\ \in\ STR} e^{-\frac{\left(\ln(y_a(t)) - \ \mu_{a,t}\right)^2}{2\sigma_{a,t}{}^2}}}
$$

$$
= \ max_{q>0} \prod_{a,t\ \in\ STR} \exp\left(\frac{\left(\ln\left(y_a(t)\right) - \ \mu_{l,t}\right)^2 - \left(\ln\left(y_a(t)\right) - \left(\mu_{a,t} + q\right)\right)^2}{2\sigma_{a,t}{}^2}\right)
$$

$$
= max_{q>0} \exp\left(\sum_{a,t\ \ \in\ STR} \frac{2q \times \ln\left(y_a(t)\right) - \ \mu_{a,t} - q^2)}{2\sigma_{a,t}{}^2}\right)
$$

$$
= max_{q>0} \exp\left(q \sum_{a,t\ \in\ STR} \frac{\ln\left(y_a(t)\right) - \ \mu_{a,t}}{\sigma_{a,t}{}^2} - \frac{q^2}{2} \sum_{a,t\ \in STR} \frac{1}{\sigma_{a,t}{}^2}\right)
$$

$$
= max_{q>0} \exp\left(qA\ - \frac{q^2}{2} B\right)
$$

where $= \sum_{a,t\ \ \in\ STR} \frac{ln\left(y_a(t)\right) - \mu_{a,t}}{\sigma_{a,t}{}^2}$, and $B = \sum_{a,t\ \ \in\ STR} \frac{1}{\sigma_{a,t}{}^2}$. From this equation, the maximum likelihood estimate of q can be found as $\frac{A}{B}$. If this is substituted into the previous function, the function illustrated in equation 6 is obtained:

$$Ƒ(STR) = \begin{cases} \exp\left(\frac{A^2}{2B}\right), & \text{if } \sum_{a,t \in STR}\left(\ln\left(y_a(t)\right) - \mu_{a,t}\right) > 0 \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

The likelihood ratio values are calculated for each STR. Determining the significance of an STR is discussed in the next subsection.

3.3.3. Determining Significant STRs

This step determines which STRs are significant. In order to decide whether an STR is significant or not, the likelihood ratio values' distribution under the null hypothesis should firstly be determined. This is achieved by conducting Monte Carlo simulations, as an analytical solution to compute the test statistic derived in equation 6 is not available.

Monte Carlo simulations generate LJTs based on the null hypothesis that there is no NRC. Then, STRs are used to scan the simulated data sets and their likelihood ratio values are calculated based on equation 6. At this point, it is important to keep in mind that the maximum spatial and temporal window sizes which are used to create STRs to scan the entire road network should be the same for the simulations. Since the simulated data are generated based on the null hypothesis, the highest likelihood ratio scores of each simulation (denoted as Ƒ*) indicate the scores which are expected to be observed when there is no NRC.

After conducting the Monte Carlo simulations and detecting the highest scoring regions of each simulation under the null hypothesis, it is then necessary to calculate the *p*-value of each STR in the observed data. Only those STRs whose individual LJTs are excessive are evaluated. In this way, detecting a significant STR that includes several links but only one having an NRC would be prevented. Focusing only on STRs whose individual LJTs are excessive also reduces the run-time of the STSS, as there would be fewer STRs in total to evaluate. After guaranteeing all the individual LJTs of an STR are excessive, the *p*-value of an STR is calculated, which indicates the likelihood of observing that STR if data are to be generated under the null hypothesis. Thus, the lower the *p*-value, the higher the likelihood that the LJTs within that STR are *not* generated according to the null hypothesis. The LJTs that belong to a statistically significant STR are considered to belong to an NRC.

The number of simulations which had a higher score than the STR are counted (denoted as R_{high}), and the *p*-value of an STR is calculated as; $\frac{R_{high}+1}{R+1}$, where R is the total number of simulations (or replications). In order to clarify this, consider the example illustrated in Figure 3-9, where the *p*-value of the investigated STR (highlighted in red) in the original data is to be determined. The likelihood ratio value of the STR is compared with the highest scores obtained in each simulation denoted with Ƒ*. Highlighted STRs should be considered as a three-dimensional entity, which can govern at most τ (i.e. maximum temporal window size) consecutive LJTs.

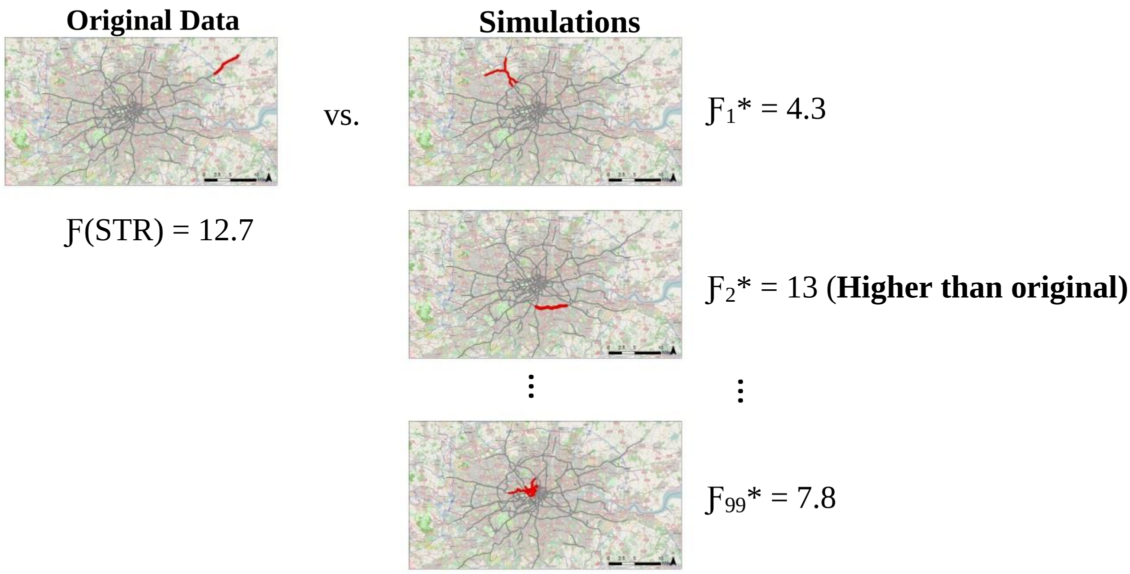


Figure 3-9 Determining the p-value of an STR

The significant STRs are detected as STRs whose *p*-values are less than a significance level (denoted as α), which is usually chosen as 0.05. Suppose that seven simulations out of 99 had a higher likelihood ratio than the one being investigated. Then, the *p*-value of the STR is calculated as $\frac{7+1}{99+1} = 0.08$. If the significance level is set as 0.05, the investigated STR would not be classified as a significant STR, as its *p*-value is higher than the significance level. As a result, a significant STR has the following four attributes: congested links, the times when the links are congested, a likelihood ratio value, and a *p*-value.

3.3.4. Clustering Significant STRs

There would be a number of significant STRs with the same *p*-values as the most significant STR, since adding or removing an LJT observation from an STR does not have a noticeable effect on the likelihood ratio value. This observation was reported when spatial scan statistics were proposed, where it was concluded that spatial scan statistics can just provide an '*approximation*' of the location and shape of the '*real*' cluster (Kulldorff and Nagarwalla, 1995). The other significant clusters are referred to as *secondary clusters*. Therefore, in most spatial scan statistics approaches, secondary clusters that spatially overlap with the most likely cluster are *not* reported (SaTScan, 2010). However, reporting only the most significant cluster leads to the provision of only a static snapshot of the significant cluster without providing an insight into how the cluster developed in space and time. Therefore, this step proposes to cluster the significant STRs found in the previous step, so that the evolution of an NRC can be observed.

The proposed way to determine the evolution of an NRC is to cluster all statistically significant STRs that spatio-temporally overlap. The usage of 'spatio-temporal overlap' is the same as in the previous section. Specifically, two STRs ($STR1$ and $STR2$) are spatio-temporally overlapping if they satisfy the following two conditions:

- $\exists\ a, b|\ \mathrm{M}(a, b) = 1$, where $a \in STR1$, $b \in STR2$ and $a, b \in \mathbf{A}$.
- $\exists t_i, t_j \middle| t_i = t_j$, where $t_i \in STR1$ and $t_j \in STR2$ and $1 \leq\ t_i, t_j\ \leq T$.

Two statistically significant STRs are spatio-temporally overlapping if they occur at adjacent links (i.e. the first condition), and there is a time interval which is common in both of the statistically significant STRs (i.e. the second condition). In this way, all the information present in secondary clusters is used to detect NRCs. Since all the STRs that are included within an NRC are known and an STR can be defined in terms of episodes, the evolution of an NRC can also be determined by using the process described in Figure 3-5.

3.4. Summary

This chapter proposed two novel methods to detect NRCs that could be applicable on a large urban road network. Each method is a realisation of a spatio-temporal clustering strategy. The primary data source indicating an NRC is high LJT estimations, and both of these methods require four inputs. The first is an adjacency matrix M, which is the mathematical representation of a geographical urban road network. The second is historical LJTs, namely the past LJTs that are used to estimate the expected LJTs and determine the statistical distribution to model an LJT. The third, a congestion factor, is used to identify whether an LJT is excessive or not. Lastly, NRCs are detected based on an input date.

The first NRC detection method centres on similarity-based clustering, where excess LJT is used to assess whether or not a given LJT belong to an NRC. An episode is defined as a maximal interval on a link during which all LJTs are excessive. The episodes that spatio-temporally overlap are clustered to detect NRCs. The advantage of this method is its simplicity, as it requires only a single parameter (i.e. congestion factor *c*). However, there is a lack of information on setting the correct congestion factor for the purpose of NRC detection. Using a high congestion factor might result in missing some NRCs or underestimating the impact of NRCs. Using a low congestion factor, on the other hand, might assign day-to-day traffic variations to an NRC. This will result in NRCs that are larger (e.g. more severe) than real NRCs.

The second NRC detection method centres on significance-testing based clustering. An expectation based STSS is proposed to detect NRCs. The method uses overlapping STRs to scan the entire road network and clusters spatio-temporally overlapping significant STRs to detect NRCs. The proposed method has two important advantages:

- The detected clusters are represented dynamically, so that the growth of a significant cluster could be observed. In this way, the impact of a cluster could be estimated more accurately. In order to achieve this, the proposed method does not solely rely on the *most* significant STR. Instead it clusters all the spatio-temporally overlapping significant STRs to detect NRCs.
- The proposed method generates STRs by considering the first-order adjacencies of a link, and does not evaluate an STR unless all the LJTs included within that STR are excessive. This adjustment decreases the number of STRs that should be evaluated, hence, increases the computational feasibility of the proposed method.

Both of the proposed methods allow NRCs to be represented in their evolution. Thus, it is possible to identify which links had NRC and when, so each detected NRC can be quantified by two measures: lifetime and severity. Lifetime denotes the temporal extent of an NRC, and severity denotes the total excess LJT that the NRC exhibits. If traffic flow data are available, then flow-weighted total excess journey time could also be considered to estimate the severity of an NRC. In this way, the amount of vehicular traffic would influence the way in which the severity of an NRC is estimated.

4. EVALUATING NRC DETECTION METHODS

Two methods have been proposed in Chapter 3 to detect Non-Recurrent Congestion events (NRCs) based on Link Journey Time (LJT) data. These methods do not make any assumption regarding the severity of an NRC event and use excess LJTs as the main indicator of an NRC. Naturally, different NRC detection methods or different parameter settings of the same method will detect different NRCs. Therefore, the aim of this chapter is to seek for an answer to the question '*how can the performance of these different NRC detection methods be evaluated?*'

This chapter is organised based on the discussion in section 2.5, which identified two strategies to evaluate the performance of a clustering method: external and internal.

- External strategy is used to calculate the commonly used performance criteria like detection rate (DR) and false-alarm rate (FAR). This strategy requires ground truth data, which includes the information whether or not a high LJT does really belong to an NRC. There are three main options to collect ground truth data of this kind: asking traffic operators to label the data, using traffic simulations, or collecting data that can be used for ground truth. The former two options are not suitable for this thesis' purpose; since it is cost prohibitive for a traffic operator to label all the LJTs on a large urban road network, and there is lack of research on simulating NRCs on a large urban road network. Therefore, this chapter investigates whether an incident data set can be used as ground truth data to evaluate the performance of an NRC detection method in section 4.2. It should be emphasised that the incident data set investigated in this chapter includes any event that may affect road network performance, including planned engineering works, cultural/sports events, even congestion, and so includes more types of events than those used in AID literature.
- Internal strategy derives performance criteria without relying on an external data source, but only the clustered data. A common way of adapting this strategy is to propose new criteria. Section 4.3 proposes two novel criteria to evaluate NRC detection methods that rely on LJT data. The first is called 'high-confidence episodes' and it assesses to what extent episodes that last at least a minimum duration were detected by different NRC detection methods. The second is called 'Localisation Index' and it assesses to what degree the detected NRCs comprise connected components throughout their lifetime.

4.1. London’s Urban Road Network

London’s urban road network can be studied using three main data sources. Each of these sources captures a specific component. The first is the Integrated Transport Network (ITN). This data source provides the highest spatial resolution by detailing every road segment and junction (Ordnance Survey, 2013). The second is CCTV cameras. Visual traffic data are collected on a daily basis by around 1,300 CCTV cameras. Such visual data are mainly used for surveillance and to detect incidents by visual inspection carried out by traffic operators. The third is automatic number plate recognition (ANPR) cameras. These cameras are mainly used for traffic enforcement, which covers schemes like the congestion charge or low emission zone. The data obtained from ANPR cameras are also used to estimate an LJT; hence, they are used for road network performance monitoring (TfL, 2012b). Estimated LJTs are the main data source of this thesis, which are kindly provided by Transport for London (TfL).

4.1.1. Automatic Number Plate Recognition Camera Network

A link is usually defined as the route between two ANPR cameras with the highest traffic flow. Therefore, as previously mentioned, it is possible for two links to overlap spatially. Some of the links created by TfL overlap in this manner and such a network representation is not ideal. This is because it creates redundancies, which unnecessarily increases the run time of the analysis. Therefore, TfL classifies the links into two as *core* or not. Core links minimise the spatial overlap between the links. Furthermore, core links are generally treated as having higher sample sizes to estimate LJTs, which increases the quality of the LJT estimates. Therefore, when reporting the performance of road networks, TfL relies on data collected from core links. In total there are 1,261 ANPR links and 424 of these are core links. In order to illustrate the spatial overlap between these two networks, it is useful to illustrate them on a map, as shown in Figure 4-1. As can be seen from the figure, most of central London is covered with core links and most of the non-core links are actually outside central London. As TfL uses core links for their road network performance monitoring, this thesis also uses the core ANPR network. Hereafter ‘link’ indicates a ‘core link’.

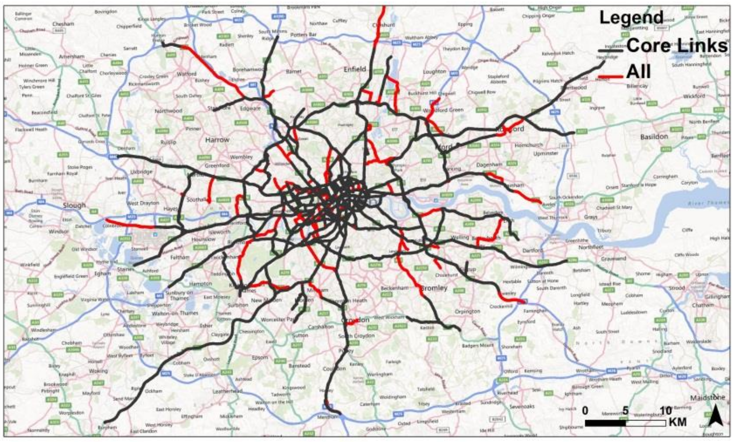


Figure 4-1 London's ANPR network

ANPR links display different lengths and characteristics. Link lengths vary from 175 to 13,487 metres, and the total length of the network is around 1,250 kilometres. Links also differ with respect to the traffic flow direction, the number of lanes they contain and the number of junctions through which they pass. Even though the core network is supposed to represent the road links which do not spatially overlap, some ANPR core links, especially in Central London, do spatially overlap.

Estimation of Expected LJTs

Expected LJTs represent the most likely estimate of journey time at a given link and time. Use of expected LJTs is intended to capture the recurrent nature of traffic, as the expected LJT in peak hours is higher than the expected LJT in off-peak hours. Estimates of expected LJTs depend on time-series analysis performed on historic data. There are different possible methods that can be used for this purpose, ranging from simple averaging of historic LJTs to a more complicated weighted averaging.

This thesis uses two different expected LJTs, one for similarity based NRC detection, and the other for significance testing based NRC detection. The first uses the expected LJTs adopted at TfL, where there are three main periods: weekdays, Saturdays and Sundays (TfL, 2010, p.100). For weekdays, expected LJTs are estimated by taking the mean of all the LJTs that happened on weekdays of March 2009 and no data cleaning procedure is applied. For weekends, however, approximately two months (March and April 2009) of data are used to estimate LJTs.

Once the expected LJTs have been estimated, links are categorised into distinct classes. This classification is based on the ratio calculated by dividing the cumulative of estimated LJTs to the cumulative of expected LJTs during the analysis period determined by a traffic operator (e.g. between 07:00 and 19:00). The congestion level on the link; hence its colour as displayed to the traffic operator is determined based on the ratio interval that the calculated ratio falls into. There are four congestion levels and they are illustrated in Table 4-1. Note that, these congestion factors are currently in use at TfL.

Table 4-1 Congestion factors used at TfL

Ratio Interval	Congestion Level	Colour
1.0 – 1.2	Minimal	Yellow
1.2 – 1.5	Moderate	Orange
1.5 – 2.0	Serious	Red
2.0 or more	Severe	Black

This classification allows a traffic operator to ‘see’ the spatial distribution of congestion on an analysed period. An example of such an illustration is illustrated in Figure 1-1.

4.1.2. Incident Data Set

London is one of the major cities of the world, and its traffic network is monitored by various sensors on a daily basis. An important data source to analyse London’s road network is the incident data set, which is referred to as the ‘London Traffic Information System’ (shortly LTIS). The incidents are recorded from various sources, including traffic operators surveilling the road network via CCTV cameras (approximately 1,300 cameras), police reporting a crime incident, bus drivers notifying the incidents and advance notice from road work contractors or event organisers. In other words, incidents that are recorded in LTIS are considered to affect the road network performance.

All of the recorded incidents are spatio-temporal in nature, as each incident is localised on the road network (i.e. spatial domain), spans a time interval (i.e. temporal domain), and each incident has characteristics like its type (i.e. thematic domain).

The spatial domain of LTIS consists of four attributes which are, ‘*Easting*’, ‘*Northing*’, ‘*Street*’, and ‘*Borough*’. Easting and Northing attributes, together, are used to define the location of the incident on the geographical Cartesian coordinates. Street specifies the name of the street, and the Borough attribute specifies the borough (one of 32 boroughs in London) in which the incident occurred, which is determined based on the values of Easting and Northing. As can be seen, these three attributes locate an incident in London at different spatial resolutions. Easting/Northing attributes locate the incidents as a point in London, which usually corresponds to the location of the CCTV that the traffic operator used to record the incident or the location of the faulty traffic signal. On the other hand, incidents are described as a line and an area by the Street and Borough attributes respectively.

The spatial location of an incident, however, is uncertain and cannot be easily captured by the aforementioned spatial attributes due to the following reasons. 'Easting/Northing' attributes locate an incident as a 'point' in space, which oversimplifies the location of the incident, since the incidents that occur on a road network are too big to be represented as a point in geographical Cartesian coordinates. On the other hand, the Street attribute usually provides the street name of the incident rather than pinpointing the exact location. Although some of the incidents might result in a road closure, in which case the Street attribute gives the most accurate information, this does not occur in the majority of incidents. Lastly, Borough provides the most generic locational information about the incident, since only few major events like State Opening of the Parliament could affect a region as large as a borough. For the convenience of the analysis, this thesis assumes that the locations of incidents are determined by Easting/Northing attributes, which provide the most precise information on the location of the incidents.

The temporal domain consists of two attributes which are '*Start Time*' and '*End Time*'. These attributes are used to specify the start and end times of the incident, respectively. The start and end times should be thought of as the times when the traffic operator became aware of the incident and recorded it, instead of the start and end times of the incident in reality. This is because an incident might have started and ended earlier than records show in the LTIS data set. The duration of an incident is measured in seconds. Hence, the start and end time of an incident consists of day-month-year, hour:minute:seconds. The difference between the '*End Time*' and '*Start Time*' is the duration of an incident. Duration of incidents varies from seconds to years, which illustrate the amount of the temporal variation.

The thematic domain characterises each incident by using three attributes, which are: '*Event Category*', '*Description*' and '*Comments*'. 'Event Category' specifies the type of incident, and there are 49 incident categories. These 49 event categories are divided according to whether the incident is planned or not. Fifteen of these categories describe incidents that are planned a priori and hence, do not require immediate attention. The remaining 34 categories describe incidents that are not planned, thus, requiring immediate attention. For example, unplanned incidents like ‘Accident’ or ‘Broken Down Vehicle’ require immediate attention, whereas planned events like ‘Sporting Event’ or ‘Bridge Lift’ do not require immediate attention. Thus, any event that might affect travel times are recorded in the incident data set. The traffic direction which is affected by the incident is indicated by the Description attribute (e.g. eastbound direction or all directions). Lastly, the Comments attribute provides a literal explanation of the incident logged by the traffic operator. Some generic characteristics of the incidents recorded in the LTIS data set for 2010 are summarised in Table 4-2.

Table 4-2 Generic characteristics of LTIS for 2010

Characteristic	**Value**
Total Number of Incidents	24,278
Duration of the Shortest Incident	3 seconds
Duration of the Longest Incident	1,854 days
Most Common Event Category / Number of Times it was Recorded	Accident / 5,048
Least Common Event Category / Number of Times it was Recorded	Planned Signalling Work / 3
Borough – Greatest Number of Incidents	Westminster (2,358 incidents)
Borough – Lowest Number of Incidents	Harrow (128 incidents)
Total Number of Planned / Unplanned Incidents	11,719 / 12,559

An important outcome of this result is that the incident data set is heterogeneous within the three domains of a spatio-temporal phenomenon. Heterogeneity is observed in the spatial domain, as some regions experienced a much higher number of incidents than others. The Borough of Westminster experienced the highest number of incidents, because of two main reasons. Firstly, it is surveilled more and therefore the likelihood of an incident being detected is higher. Secondly, traffic activity in Westminster is higher than most other regions due to its touristic, political and economic importance, which increases the likelihood of an incident occurring. On the other hand, the suburbs of London do not attract as much traffic activity; hence, they are not surveilled as much as the central boroughs. Heterogeneity is also observed in the temporal domain, where the duration of incidents varies from three seconds to almost six years. At both ends of these extremes, the uncertainty involved in the validity of the recorded incident increases, since an incident which lasted a few seconds is probably an erroneous recording or the end time of a planned engineering work that last for years (e.g. the CrossRail project) may not be determined accurately. Lastly, heterogeneity exists in the thematic domain, as some types of incidents are much more common than others. For example, accidents were recorded 5,048 times, whereas planned signalling works were recorded only three times. On the other hand, homogeneity is observed in the incident data set in which the number of planned incidents is close to the number of unplanned incidents.

Distribution of Incident Types

Some incident types are recorded more frequently than others, and some incident types are dependent on the season (e.g. a ‘slippery road surface’ occurs only in winter). In order to illustrate the variation amongst the distribution of incident types, the most common 10 incidents that began in 2010 are analysed. These most common 10 incidents constitute 75.7% of all recorded incidents (18,385 out of the 24,278 incidents), and their relative percentage of occurrence is illustrated in Figure 4-2.

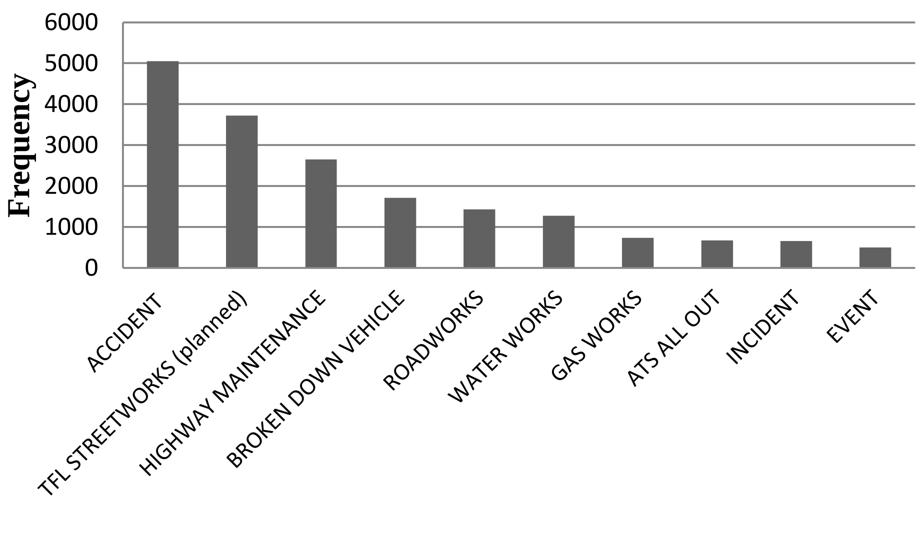


Figure 4-2 Top 10 most common incidents of LTIS (2010)

Two points should be mentioned regarding this outcome. The first is that the event categories overlap in terms of their effect on the road network. For instance, an ‘Accident’ and an ‘Incident’ may result in a road blockage or closure. Hence, the impacts of that Accident and Incident on the road network are not easily distinguishable from each other. The second point is that there is a balance between the incidents which require immediate attention, or the incidents which are planned a priori. Incidents which are planned are TfL street works, highway maintenance, road works, water works, gas works, and events. The remaining event categories (i.e. accident, broken down vehicle, ATS all out) are assigned to incidents which require immediate attention.

Classifying incidents as planned and unplanned is important, since there is more time to take the necessary precautions to decrease the impact of planned incidents on the road network. These precautions include using variable message signs and warning the public via passenger information systems. However, such planning time is not available in unplanned incidents. This observation is also justified by the recent findings of TfL, which state that unplanned incidents are responsible for approximately 74% of serious and severe delays (TfL, 2010, p.99).

The Linkage between the Number of Incidents and the Total Delay

This section demonstrates the relationship between the number of incidents and total delay. The number of incidents which begin each month and the total delay are calculated for the whole ANPR network. In order to compare total delay (measured in minutes) and number of incidents (measured as a count), which are in different domains, both of the observations are ranked. Figure 4-3 illustrates the outcome of this analysis, where the x-axis illustrates the ranked total number of incidents and the y-axis illustrates the ranked total delay. The month with the highest number of incidents and the greatest delay is ranked first. Total delay is calculated taking into account all the available data for consistency of analysis.

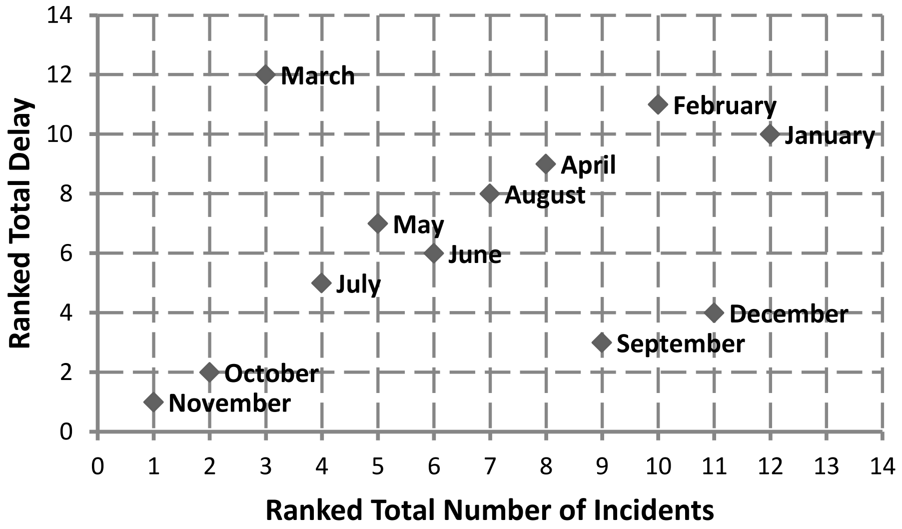


Figure 4-3 Comparison of the ranked incidents and ranked total delay

The highest number of incidents and total delay is observed in November, with 2,426 incidents and approximately 45,897 hours of delay. On the other hand, the lowest number of incidents occurred in January totalling 1,573. However, the lowest amount of delay occurred in March, due to maintenance work which affected the entire ANPR network between 1st and 28th March resulting in 3,261 hours of delay.

The results suggest that the number of incidents is not directly proportional to the total delay. For example, more incidents were recorded in August than in December; however, the total delay in December is higher than August. Similarly, more incidents were recorded in April, but September witnessed more delay. Thus, it is not suitable to compare these two different data sets based on the number of recorded incidents, as it is not an indicator for determining total delay.

Analysis of 'Congestion' Incidents

It is important to mention that congestion is a type of incident in LTIS, and it is defined in the metadata as '*This category should only be used if after investigation no cause for the congestion can be found. Please see the general remark requirements for guidance in this incidence.*' Therefore, if a congestion incident is recorded on the ANPR network, then there should also be a substantial increase in the LJTs indicating an NRC without a specific cause that is visible to traffic operator. That's why congestion incidents act as a valuable data source, as the researcher expects to observe a substantial increase in the LJTs had the congestion incident taken place on the ANPR network.

Congestion is not one of the top 10 most common event categories, and only 385 'congestion' incidents were recorded in 2010, which amounts to 1.5% of all incidents. However, not all of these congestion incidents are realistic. For instance, the longest occasion congestion incidents lasted for 22 days. After removing this incident, the histogram of the remaining congestion incidents is illustrated in Figure 4-4.

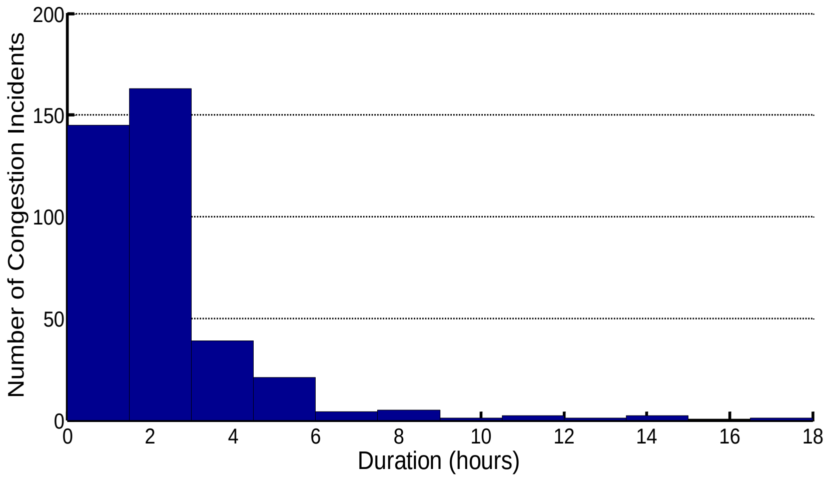


Figure 4-4 Distribution of the duration of congestion incidents

The analysis of the duration of congestion incidents reveals that the majority of them lasted less than three hours. The duration of each congestion incident completely depends on the perception of the traffic operator, due to the subjectivity involved to answer the questions, 'when did a congestion incident begin/end?'. For example, there are seven instances of congestion incidents which lasted less than five minutes, with a minimum of approximately one minute. Logging congestion incidents of such short duration raises further questions about defining a congestion incident.

4.2. Reasons for the Mismatch between LJT and Incident Data Sets

Incident and LJT data sets differ, but are still related to each other. If there is an NRC, the researcher would expect to observe this in both of the data sets via an incident in LTIS and a substantial increase in the LJTs. However, these data sets are generated using different mechanisms and priorities, which might prevent such a match. This fact is recognised in the literature (Jeong et al., 2011), but an empirical analysis of the reasons for this mismatch has not yet been conducted.

This section investigates the reasons which prevent the matching of the LJT and incident data sets. The examined reasons include the location of the sensors, unrecorded incidents, non-disruptive incidents, inactive cameras, and the identification of the start and end times of an incident. These findings suggest that in order to use an incident data set to validate the detected NRCs, these five factors must be satisfied.

4.2.1. Location of the Incidents

In order to observe the impact of an incident on an ANPR link, it is a necessity that the incident occurred on the ANPR network. Otherwise, the impact of the incident on the nearest ANPR link could only be partial, depending on the proximity of the incident to the link, as well as the nature of the incident (i.e. location, duration and number of lanes blocked) and the link (e.g. length). Modelling such a partial impact is difficult and is not considered within this thesis.

This section investigates the amount of spatial overlap between the locations of the incidents and the ANPR network. As discussed previously, incidents are represented by points (characterised by the Northing/Easting attribute pair), and the ANPR network is represented as lines. There are several reasons for the incidents not to overlap with the ANPR network. The first is that the incidents might be recorded elsewhere, since the ANPR network is limited to the major arterials within an urban road network. Although there is an overlap between the spatial coverage of CCTV cameras and the ANPR network, it is not perfect. Thus, it is possible for an incident to be recorded on a road segment which is not covered by an ANPR link. The second is that, when recording the location of the incident, there might be a data input error which prevents the incident from overlapping with the ANPR network. This situation can occur even though the location was actually covered by the ANPR network (which is verified by inspecting the location attribute of the incident). An example of this situation is illustrated in Figure 4-5, where the incidents (illustrated with an exclamation mark) did actually occur on the ANPR link. However, due to data input errors, when conducting spatial analysis on a geographical information system, those incidents would not be counted as incidents which overlap with the link.

Figure 4-5 Incidents not spatially overlapping with the ANPR link

Therefore, the incidents should be viewed as having a buffer around them which characterises this error margin. The size of the buffer, however, is not known and depends on the way in which the traffic operator recorded the incident. The size varies from 0 metres (perfect overlap) to 10 metres (i.e. the design standard for 2-3 lane carriageways in the UK (DfT, 1999). The percentages of incidents that overlap with the ANPR network are illustrated in Figure 4-6. It should also be noted that as the size of the buffer increases, the probability of an incident being incorrectly identified as occurring on an ANPR link also increases. This is because most of the CCTV cameras are located at junctions, and not all the roads entering/exiting a junction are part of the ANPR network.

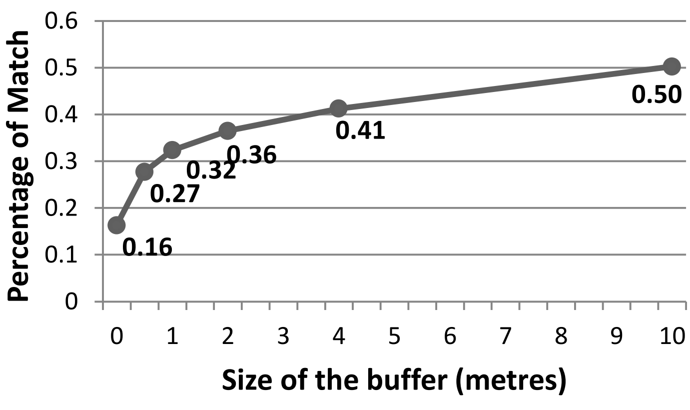


Figure 4-6 The amount of spatial overlap between the incidents and ANPR network

If the spatial analysis had been conducted solely based on the raw incident data set and the ANPR network, the percentage of incidents which overlap with an ANPR link would be 16% (i.e. 3,958 incidents out of 24,278), and once the size of the buffer is increased to 10 metres, the spatial overlap percentage increases to 50%.

From this analysis, it can be observed that at least half of the incidents which took place in 2010 did not occur on roads which are covered by the ANPR network. Those incidents that did not occur on the ANPR network may only partially impact LJTs. However, quantifying such an indirect impact requires further research, which depends on the nature of the incident and the closest ANPR link which may be affected by the incident.

4.2.2. Unrecorded Incidents

A link might have experienced substantial increase in LJTs, but it is possible that such an increase in LJTs cannot be explained via the incident data set. There are two main reasons why this may happen. Firstly, surveilling the entire link via CCTV cameras is cost-prohibitive. As most of the CCTV cameras are located around Central London, the suburbs are not as surveilled. Secondly, as previously discussed, incidents are reported based on the subjective interpretation of the traffic operators. This subjective interpretation may lead to ambiguities, where one traffic operator would consider a given traffic situation to belong an incident, whereas another would not.

In order to illustrate this, the following two cases are presented as examples:

Case 1: Link 573 on 20 July 2010, Tuesday, shows a substantial increase in LJTs around 08:00. This link is located at the border of Greater London and intersects with the M25. The length of the link is approximately 2.2 kilometres and the traffic flows from north to south-west. The estimated LJTs, expected LJTs, and sample size to estimate each LJT are illustrated in Figure 4-7.

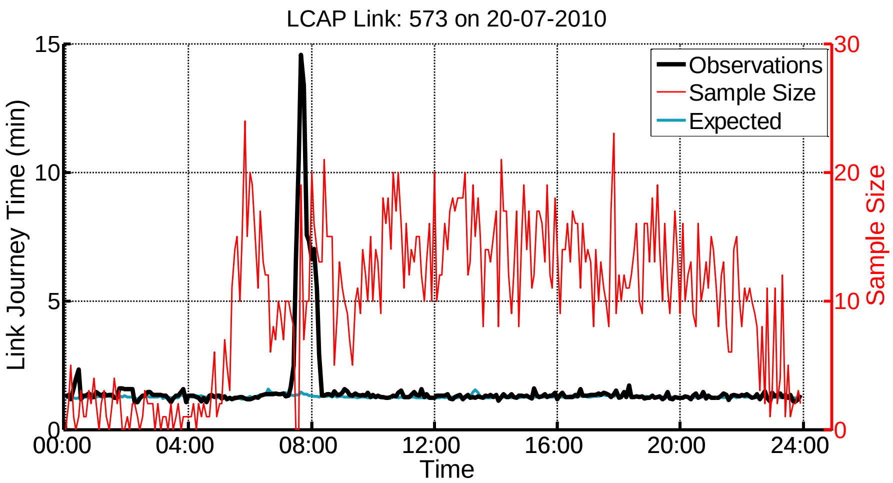


Figure 4-7 The LJTs and corresponding sample sizes of link 573 on 20 July 2010

It can be seen that there is a substantial increase in LJTs between 07:30 and 08:15 compared to the expected LJTs. In addition, the link is good-quality since the number of captured vehicles (i.e. the sample size) is usually greater than five per LJT observation, which assures the researcher that such an increase did actually occur, and was not a mere sampling error. However, this increase in LJTs could not be traced from the incident data set, since no incident was recorded on 20 July 2010 within 500 metres of the link. In order to be sure that there was no long-term incident which might affect the LJTs between 07:30 and 08:15, the LJTs observed on the same link on 19th and 21stJuly are also analysed. This analysis verifies that the increase in LJTs is specific to 20 July 2010, since the link does *not* show a substantial increase in LJTs on the previous or subsequent days. It should also be noted that all of the analysed days are weekdays, thus the expected LJTs remain the same for the analysis period, 19th – 21stJuly. The substantial increase in LJTs proves that there was an incident on 20 July 2010 which was not recorded on the LTIS data set. This is probably because of the lower level of surveillance in this area, as the link is located on the periphery of London, so it is more likely that an incident will be missed.

Case 2: This case examines link 550, which is centrally located, beginning at Hyde Park Corner and continuing to Brompton Road, on which the famous Harrods department store is located. The link is analysed on 22 August 2010, Sunday. The length of this link is 1,232 metres and the traffic flow direction is from north to south-west.

The link shows a substantial increase in LJTs between 14:20 and 16:35, as seen in Figure 4-8. The qualities of the estimated LJTs can be considered to be reliable, since the sample size rarely fell below five vehicles per LJT estimate. However, no incident which could possibly affect the LJTs was recorded on 22 August 2010 within 500 metres of this link. Therefore, the increase in LJTs cannot be explained using the incident data set. The total delay on 22 August on link 550 between 14:20 and 16:35 is calculated as 92.71 minutes, and this figure increases to 149.91 minutes if the whole day is considered. In other words, the substantial increase in LJTs between 14:20 and 16:35 represents approximately 62% of the delay.

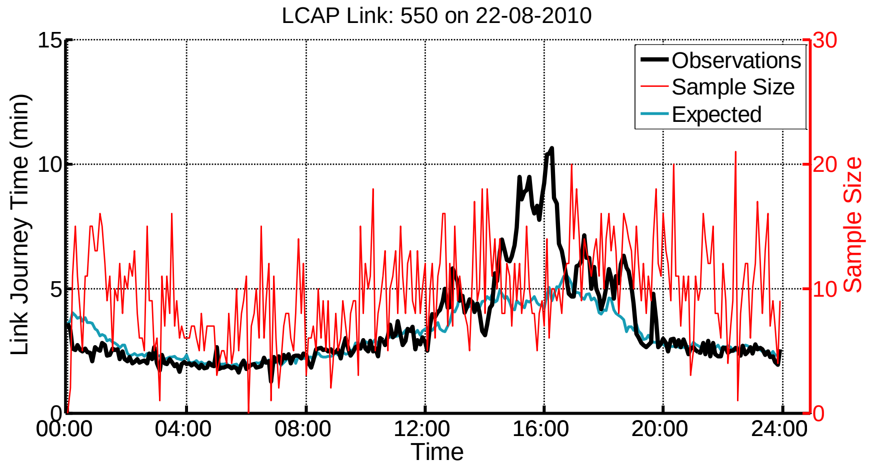


Figure 4-8 LJTs and their sample sizes for link 550 on 22 August 2010

In order to understand whether this increment is due to a long-term incident, the same link is investigated on 21st and 23rd of August 2010, and no substantial increase in LJT observations is observed.

In order to demonstrate that there was an 'unrecorded incident', the same link is analysed for 7 April 2010 when a congestion incident was reported. The reason why a congestion event is analysed is that they are rarely recorded, and the traffic operator should observe a substantial increase in traffic density in order to report an incident of this kind. The congestion incident started at 16:08 and ended at 16:22. The analysed day is a Wednesday, thus a different expected LJT is applicable which is shown in Figure 4-9.

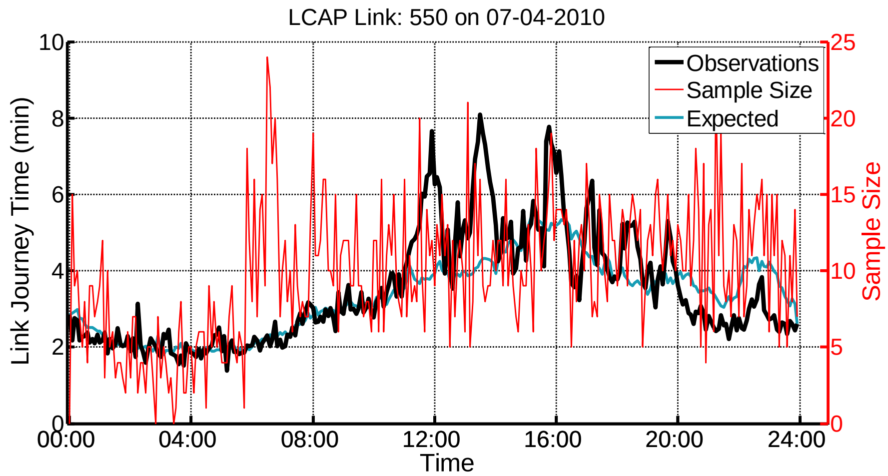


Figure 4-9 LJTs and their sample sizes for link 550 on 07 April 2010

It is observed that there is a substantial increase in LJTs between 11:10 and 14:00, and the excess travel time is calculated as 61.36 minutes; this increases to 115.38 minutes if the whole day is considered. This means that approximately 53% of the delay could be considered to be caused by the incident. This increases the likelihood that at least a congestion incident should have been recorded on 22 August.

Similar cases where there is an NRC when the LJTs are analysed which is not supported by the incident data set can be enumerated. However, the main conclusion that should be drawn is that LJT and incident data sets may not necessarily match due to the subjective component involved within reporting incidents.

4.2.3. Non-disruptive Incidents

There are several reasons why incidents' impact on LJTs cannot be observed. The first is that the incident might have taken place somewhere outside the ANPR network, as previously discussed in section 4.2.1. The second reason is that, even if the incident occurred on the ANPR network, its impact on LJTs may not be visible due to the length of the ANPR link. The rule that applies in this context is the longer the link, the higher the likelihood that the impact of the incident is diluted. The third is that the incident could be mild so that its impact on the road network is limited. For example, such a mild incident might occur when traffic flow is low, or the incident is cleared quickly. Hence, the impact of mild incidents on LJTs may not be easily noticeable enough to be detected as an NRC.

The impact of congestion incidents is discussed in section 4.1.2 and it is recognised that they usually result in a substantial increase in LJTs. A congestion incident occurred on 18 November 2010 between 18:44 and 20:25 on a link with a length of approximately 11.5 kilometres. In order to investigate the impact of this incident, the estimated LJTs are shown in Figure 4-10.

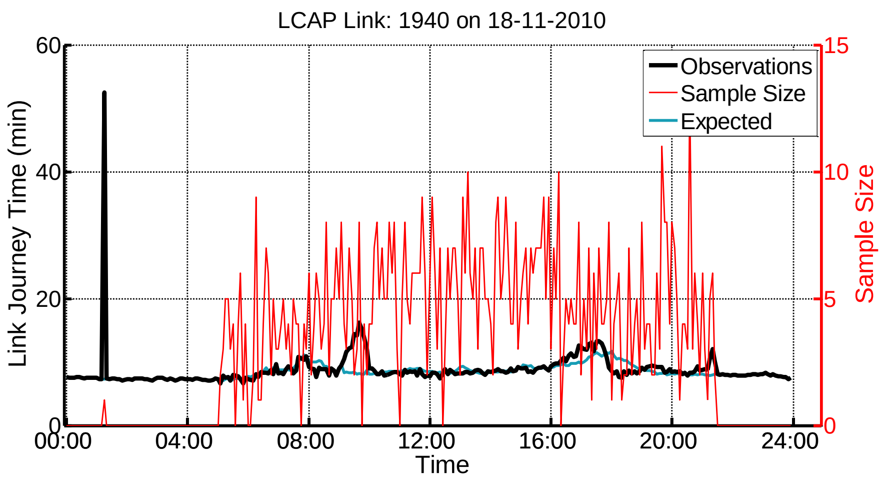


Figure 4-10 LJTs and their sample sizes for link 1940 on 18 November 2010

From the estimated LJTs, it can be observed that only a minor increase occurred on the link during the congestion's temporal extent (between 18:44 and 20:25). It can also be observed that the link can be considered to be good quality, since the sample sizes of LJTs are usually greater than five vehicles per five minutes (especially between 07:00 and 19:00). Note that an unusually high LJT is estimated at 01:15, mostly due to the only vehicle that was captured taking a detour. Only one vehicle was captured successfully and this vehicle displays a very long journey time, resulting in the estimated LJT being very high. Although this LJT estimate is most likely to be erroneous, the other LJTs that rely on a single vehicle during the day show a reasonable value, suggesting the difficulty to model the quality of link solely based on sample size.

A congestion incident is usually recorded when a traffic operator observes an unusual increase in traffic density, where no specific cause is observed. Therefore, substantial increase in LJTs is anticipated on the link from which it is reported. Nevertheless, even a congestion incident may not show a noticeable increase in the estimated LJTs, as it may be reported from a long link, which may dilute its effect.

As a summary, incidents which may initially appear to cause a substantial increase in estimated LJTs might not do so due to several possible reasons, including the impact of incidents being diluted across a long link, or they may be mild incidents[6].

4.2.4. Inactive Cameras

ANPR cameras might be unavailable for a period of time, and these periods are referred to as '*inactive periods*'. An ANPR camera becomes inactive when it is broken or undergoing maintenance. If an ANPR camera becomes inactive, the links which use that camera will not show any estimated LJT. This leads to missing LJT estimates, and such missing data are patched to their expected LJTs. An incident's impact on an ANPR link cannot be investigated if the link is inactive during the incident's duration. A realisation of this situation is illustrated in Figure 4-11, where an incident was reported on the link between 06:09 and 07:47. As there were no LJT estimates during this period, the impact of the incident cannot be investigated.

[6] If incidents that last for fewer than five minutes are considered to be mild, then 7% of the reported incidents are mild.

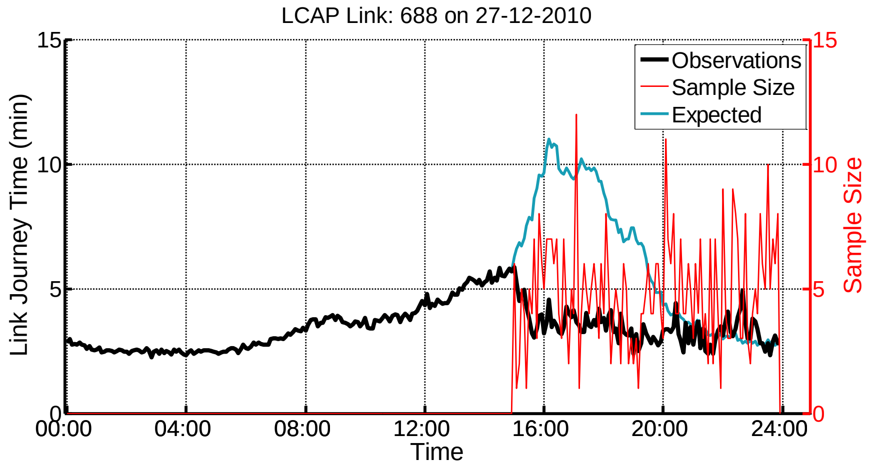


Figure 4-11 LJTs and their sample sizes for link 668 on 27 December 2010

In order to investigate to what extent ANPR cameras are inactive, the following analysis is applied. Except for days when the whole ANPR network was inactive in 2010 (which accounts for 32 days in total), the remaining 333 days are analysed between 07:00 and 19:00. The duration of the inactive period is varied from 30 minutes to 12 hours, and the number of links which were inactive is counted for each of these periods. Then, the average number of links which are continuously inactive during these intervals is calculated and the results are plotted in Figure 4-12.

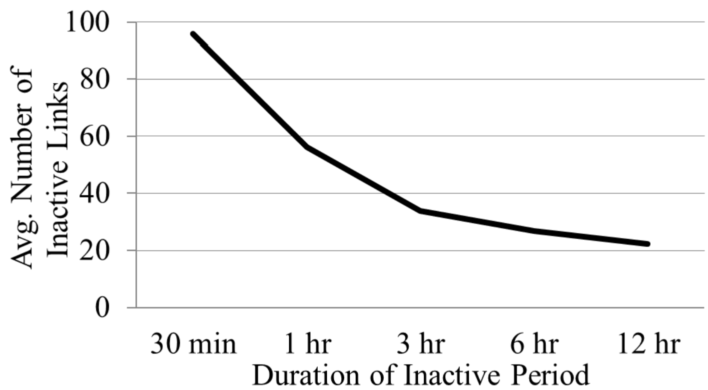


Figure 4-12 Average number of inactive ANPR links given a duration of inactive period

It can be seen that in an average day in 2010, approximately 20 links were inactive between 07:00 and 19:00, and this number increases to approximately 95 links when the inactivity period is reduced to 30 minutes. It should also be noted that the number of links which were inactive varies from day-to-day. For example, no links were inactive between 07:00 and 19:00 on 28 April 2010, whereas 162 links were inactive for the same period on 6 February 2010. This result illustrates that the ANPR road network is highly dynamic in terms of the variation in inactive links on a day-to-day basis. This characteristic of the LJT data makes it very difficult to remove links that are of low quality.

4.2.5. When does the Incident Begin/End?

There are many uncertainties involved in the generation of the incident data set. One of the most important is the start and end times of the incidents. There are two main reasons that explain the uncertainty involved in data recorded under these attributes. Firstly, an incident is recorded when the traffic operator initially becomes aware of it. This distinction becomes even more obscure when the incident depends on the subjective interpretation of the traffic operator. For example, when recording a congestion incident, '*When did the congestion start/end?*' or for a slippery road surface incident, '*When did the road become slippery?*' are the sorts of questions whereby answers depend on the perceptions of the traffic operator. The second is that there might be an error in the data input process. As illustrated previously, the shortest incident of 2010 lasted for three seconds (i.e. roadwork), and the longest congestion incident lasted for 22 days. These durations are not realistic. Furthermore, incidents are recorded in real-time as traffic operators inspect the CCTV cameras. In other words, when the traffic operator observes a traffic incident, s/he can only record the start time of the incident. Recording the end time of the incident (unless it is a planned incident) theoretically requires continuous surveillance. However, to what extent this is achievable in practice is an open question.

Uncertainties also exist in LJT data when detecting NRCs. For instance, if similarity-based clustering is used, different congestion factors result in different beginning and end times of NRCs. The rule of thumb, however, is the higher the congestion factor, the shorter the lifetime of an NRC, as there would be fewer excessive LJTs. On the other hand, if STSS is used, the spatial and temporal window size parameters are important in determining the beginning and end times of an NRC. As a result, matching the two data sets, where both contain uncertainties regarding the beginning and end of an incident/NRC, is a difficult objective.

4.3. Proposed Evaluation Criteria for NRC Detection Methods

This section proposes new criteria to evaluate the performance of an NRC detection method. This is achieved by considering the uses of an NRC detection method. Note that it is common to instigate the use of a clustering method when designing the evaluation criteria (Luxburg et al., 2012).

Two criteria are proposed that consider the practical uses of an NRC detection method. The first detects NRCs that result in substantial impact. The second relates the detected NRCs with incidents. These two criteria are called 'High-Confidence Episodes' and 'Localisation Index' respectively.

4.3.1. High-Confidence Episodes

This evaluation criterion assesses to what extent an NRC detection method detects 'High-Confidence' episodes (e^*). A 'high-confidence' episode is an episode on link *a* that lasts for a minimum duration during which all LJTs are excessive. A 'high-confidence' episode is formally defined on link *a* between times *m* and *n* as follows;

$$e^*\{a\} = \left\{ [m\,n] \middle| \begin{array}{ll} \delta_a(t) > 0\ \forall t\ \in [m\,n];\ m,n\ \in \{1,2,\dots,T\} & \text{and} \\ \delta_a(m-1) = \delta_a(n+1) = 0 & \text{and} \\ n - m + 1\ \ \geq t^{\min} & \end{array} \right\}$$

,where e^* was observed on link a, between time intervals indexed by *m* and n. The first two lines of the definition of an e^*satisfies that an e^* is an episode. Hence, it is a maximal interval during which LJTs are excessive. The third line is the most important since it differentiates an episode with an e^*. Specifically, an episode should occur for a minimum duration, which is denoted by $t^{\min}$. The continuous time of the minimum duration is $t^{\min} \times w$ minutes, where w is the LJT estimation interval.

It can be noted that the way in which a high-confidence episode is defined is very similar to detecting NRCs by Clustering Episodes. The only difference between detecting NRCs by e^* and Clustering Episodes is the minimum duration requirement within an e^*. The main reason for developing this evaluation criterion is to satisfy the practical requirement, as very high LJTs that last a minimum duration demand attention. Because an e^* commands this attention, an NRC detection method should be able to detect the e^*s that occurred on an analysed day.

In order to evaluate the performance of a method of NRC detection using the definition of a 'high-confidence' episode, a researcher must create a 2×2 table, which is used to illustrate the number of LJTs that:

- Belong to both the e^*s and the detected NRCs (True Positive, denoted by TP);

- Do not belong to the e^*s, but belong to the detected NRCs (False Positive, denoted by FP);
- Belong to the e^*s, but not the detected NRCs (False Negative, denoted by FN);
- Neither belong to the e^*s nor the detected NRCs (True Negative, denoted by TN).

These results can be visually represented, as shown in Figure 4-13.

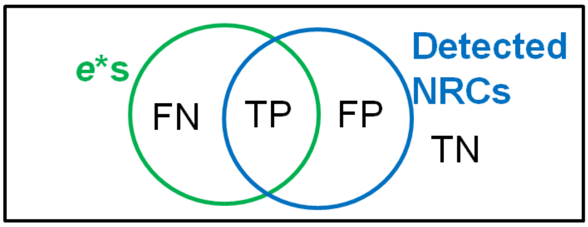


Figure 4-13 Comparison of the detected NRCs with 'high-confidence' episodes

Two important performance evaluation measures can be calculated by using the values recorded in Figure 4-13:

- False Alarm Rate (FAR), which is the proportion of all LJTs that are detected to belong to an NRC which do not belong to an e^*.

  Formally, $\mathrm{FAR} = \mathrm{FP}/(\mathrm{TP} + \mathrm{FP})$

- False Negative Rate (FNR), which is the proportion of all LJTs that belong to e^* but was missed by the NRC detection method to all LJTs that belong to e^*.

  Formally, $\mathrm{FNR} = \mathrm{FN}/(\mathrm{TP} + \mathrm{FN})$

These two measures are required to assess to what extent an NRC detection method identified and missed e^*s that are considered to be episodes that should ideally be detected. Of these two measures, FNR is critical as it determines the proportion of e^*s missed. It is possible for an NRC to last for fewer than the minimum duration defined in the definition of an e^*; hence, an ideal NRC detection model should have some FAR, but the ideal FAR is unknown due to the lack of ground truth data.

Determining the parameters of an e^*

Two parameters are required to define an e^*. The first is congestion factor c, which determines whether an LJT is excessive or not, and the second is the minimum duration

($t^{\min}$), which ensures that the episode lasted for a substantial amount of time. These two parameters are determined based on the following empirical analysis.

Determining the congestion factor

Link Journey Times (LJTs) within a 'high-confidence' episode are substantially high enough to attract a traffic operator's attention. This part aims to provide empirical evidence to quantify what is meant by '*substantially*'. In this way, the congestion factor that is used to define a 'high-confidence' episode can be estimated.

This thesis considers an LJT to be excessive enough to belong to an e^* if it is greater than or equal to its 95$^{\text{th}}$ percentile value. This is because 95$^{\text{th}}$ percentile values are commonly used in traffic science to indicate unusually high traffic data (Pu, 2011). Once the 95$^{\text{th}}$ percentile LJTs are determined, then the congestion factor that would classify it as excessive can be determined. The average of such congestion factors throughout the entire study area is used as the congestion factor to define an e^*. This process is summarised as follows:

1. Calculate the 95$^{\text{th}}$ percentile of LJTs using a historic data set. The 95$^{\text{th}}$ percentile LJT for link a at time interval t is denoted as $y_a^{95}(t)$, which is calculated by linear interpolation as follows[7]. Collect historical $y_a(t)$ and sort them in ascending order, and denote this data vector as $\mathbf{H}$, where $\mathbf{H}(1)$ and $\mathbf{H}(h)$ denotes the lowest and highest $y_a(t)$ in the historical data respectively. Find the rank of the 95$^{\text{th}}$ percentile in $\mathbf{H}$ by, $r = \lfloor (0.95 \times h) + 0.5 \rfloor$. Hereafter, $y_a^{95}(t)$ is calculated as the linearly interpolated value between $\mathbf{H}(r)$ and $\mathbf{H}(r+1)$, where the interpolation depends on the decimal of $(0.95 \times h) + 0.5$.
2. Calculate the congestion factor for link a at time interval t that would classify the 95th percentile LJT as excessive. Formally, $c_a^{95}(t) = y_a^{95}(t)/\overline{y}_a(t)$.
3. Repeat steps 1 and 2 for each LJT, $a \in \mathbf{A}$ and $t \in [1, 2, \ldots, T]$, where $\mathbf{A}$ denotes the set of links and T denotes the total number of LJTs within the analysis interval.
4. The congestion factor that is used to define an e^* is identified as the median of the $c_a^{95}(t)$ values.

[7] There are different ways to calculate 95$^{\text{th}}$ percentile of a data set. This thesis uses the one that is described in Matlab (2013).

These four steps are applied between 07:00 and 19:00 in London's urban road network consisting of 424 links. Each LJT is estimated every 5 minutes, so the analysis interval contains T = 145 LJTs (12 hours × 12 LJTs/hour + 1). The $c_a^{95}(t)$values are estimated using all the weekdays between 1st April and 1st August 2010, which constitutes 87 LJT observations. The box-plot of the estimated $c_a^{95}(t)$ values with respect to time is illustrated as shown in Figure 4-14 (those links with a $c_a^{95}(t) \geq 4$ are not illustrated to improve legibility. The original results are illustrated in Appendix C in Figure C- 1):

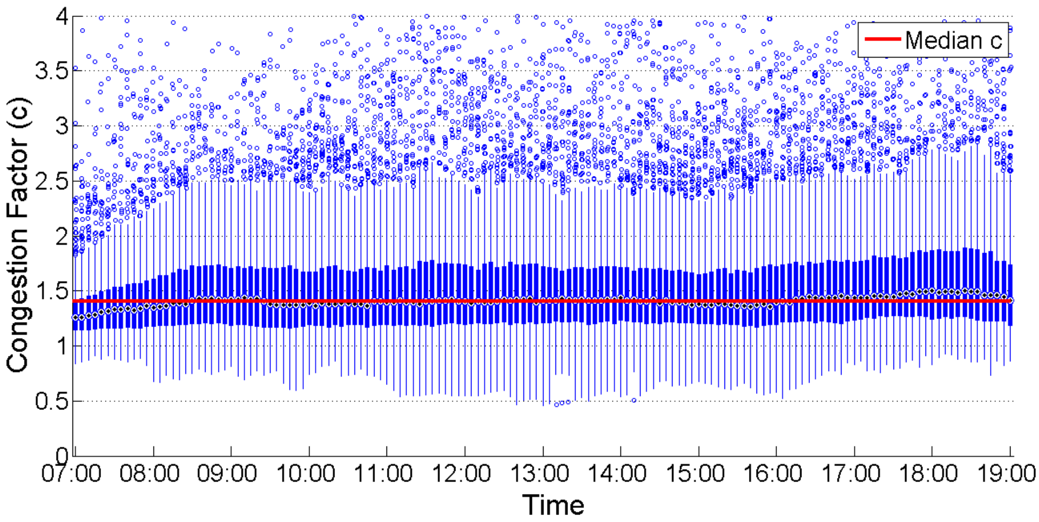


Figure 4-14 Determining the congestion factor of a high-confidence episode

These results illustrate that the $c_a^{95}(t)$ values varies spatially and temporally. Spatial variation is, however, more substantial than the temporal variation. This is an expected outcome, as some links may experience long-term engineering work during the data collection interval, which would result in high 95th percentile LJTs; hence, higher $c_a^{95}(t)$ values. This thesis makes the simplistic assumption to use a single congestion factor to define a 'high-confidence' episode. This value is estimated by taking the median of the $c_a^{95}(t)$ values that are calculated for all links at all times between 07:00 and 19:00. Median is used instead of mean, as median is a more robust estimator of central tendency than mean. The median of the $c_a^{95}(t)$ values for all the links across all the times between 07:00 and 19:00 is approximately 1.4. This suggests that whenever an LJT is 40% higher than its expected value, then it should be considered as excessive.

The variation of congestion factor values that would classify 95th percentile LJTs as excessive is very high. On average 6% of the road network (i.e. 24 links) are found to be outliers necessitating a very high congestion factor to classify 95th percentile LJTs as excessive. On such links major engineering works could have taken place during the data collection period, or they may be low quality due to the technical limitations of ANPR cameras that are used to define that link.

Determining the Minimum Duration

The second parameter in the definition of a ‘high-confidence’ episode is the minimum duration ($t^{\min}$). This parameter ensures a ‘high-confidence’ episode contains a number of consecutive LJTs that are excessive, implying persistency within the episode. This part aims to determine the $t^{\min}$ by using the information obtained from all the episodes during the analysis interval. In order to achieve this, the following question is asked: ‘what is the maximum value of the duration of an episode in which the severity of all episodes with durations greater than or equal to this value accounts for the majority of total delay?’ Therefore, all the episodes that last at least $t^{\min}$ time intervals would correspond to the majority of total delay.

This part builds upon the previous part, using a congestion factor of 1.4, and applies the following steps to determine $t^{\min}$:

1. Find all the episodes within the analysis interval of T = [07:00 19:00].
2. Calculate the duration and severity of all episodes. The duration is measured in terms of the number of consecutive LJTs that the episode contains. Severity is the total excess LJT of the episode and it is measured in minutes.
3. Sort all the episodes based on their duration in descending order, so that $\mathrm{d}(e_i) \geq \mathrm{d}(e_{i+1})$, where $\mathrm{d}(e_i)$ is the duration of the i^{th} episode.

   Find the duration of the episode when the cumulative severities of the episodes explain the majority of the total delay. Formally, it is necessary to find $\mathrm{d}(e_x)$, where $x = \min_x(\sum_{i=1}^{x} \mathrm{s}(e_i)/\Psi) \geq 0.5$, and Ψ is the total delay observed during the analysis period and $\mathrm{s}(e_i)$ is the total severity of the i^{th} episode.

4. The $\mathrm{d}(\mathrm{e}_x)$ is the minimum sufficient duration that is required to account for the majority of the total delay during the analysis period.
5. Apply the steps 1-4 for all the days between 1 April and 1 August 2010 and record all the $\mathrm{d}(e_x)$ values for different days. The histogram of these $\mathrm{d}(e_x)$ values for the analysis period is illustrated in Figure 4-15.

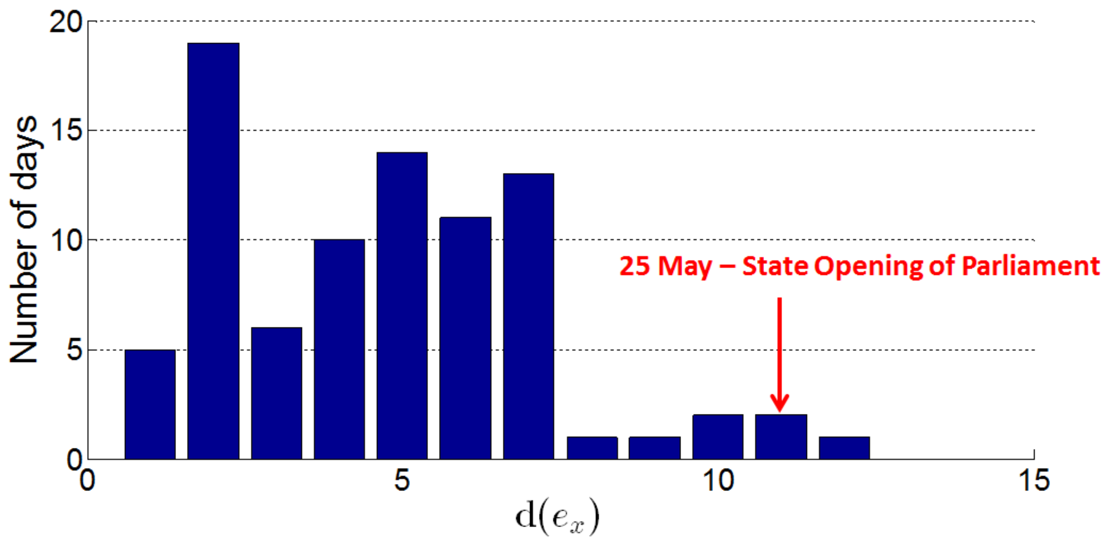


Figure 4-15 Histogram of the days where episodes lasting $\mathrm{d}(e_x)$ number of LJTs are sufficient to explain the majority of total delay

The most noticeable outcome in Figure 4-15 is the substantial decrease in the number of days with a $\mathrm{d}(e_x)$ value greater than 7. These dates are highly congested; for example, the 25th May 2010, State Opening of Parliament, is one such date. However, because congested days do occur, all of the 87 weekdays between 1 April and 1 August are considered to determine the minimum duration. It is found that, on average, episodes that last at least five consecutive time intervals account for the majority of the total delay. Since each LJT is estimated every five minutes, the minimum duration to define an e^* ($t^{\min}$) is determined to be 5×5 = 25 minutes.

These findings suggest that a 'high-confidence' episode (e^*) occurs when a link's estimated LJTs are 40% (i.e. c=1.4) higher than their expected values for a minimum of 25 minutes.

4.3.2. Localisation Index

An important application of a Non-Recurrent Congestion (NRC) detection method is the ability to relate the detected NRCs with an incident. In order to relate an NRC with an incident, the detected NRC should consist of links that are adjacent throughout its lifetime. The 'Localisation Index' is an evaluation criterion of an NRC detection method that quantifies the extent to which the detected NRCs consist of link groups that are adjacent throughout their lifetime. Its calculation depends on the number of connected components that an NRC exhibits throughout its duration, which can be described with an example.

Suppose links $a1$, $a2$, $a3$ and $a4$ are linearly aligned where the traffic is flowing from $a1$ towards $a4$, as shown in Figure 4-16. An incident occurs at link $a4$ and its effect propagates gradually towards its upstream links during the analysis period between times 1 and 5. The excessive LJTs are illustrated with a '+' sign in Figure 4-16, and the detected NRC is highlighted in red.

Traffic flow direction

Links \ Times	1	2	3	4	5
$a1$	–	–	–	+	+
$a2$	–	–	+	+	–
$a3$	–	+	+	–	–
$a4$	+	+	+	–	–

\+ Excessive LJT
– Not excessive LJT
Boundary of NRC

Figure 4-16 Propagation of the effect of an incident towards upstream links

The number of connected components is found by investigating the adjacency relationship between the links that were included in the NRC. For the NRC illustrated in Figure 4-16, there is only one connected component between times 1 and 5. However, the links that are included within the connected component change over time. For example, at time 1 the connected component consists of only link $a4$, whereas at time 3 the connected component contains links $a2$, $a3$ and $a4$. These three links constitute a single connected component, because $a2$ is adjacent to $a3$ and $a3$ is adjacent to $a4$. In other words, all the links that were in the NRC at time 3 are connected. On the other hand, for the NRC illustrated in Figure 4-17, it is more difficult to relate the detected NRC with an incident because the number of connected components is two, two and one for times 1, 2 and 3 respectively. The number of connected components is two for times 1 and 2 because the two links included within the NRC at times 1 and 2 (i.e. $a1$ and $a3$) are not adjacent.

Traffic flow direction

Links \ Times	1	2	3
$a1$	+	+	+
$a2$	–	–	+
$a3$	+	+	+

\+ Excessive LJT
– Not excessive LJT
Boundary of NRC

Figure 4-17 An NRC whose spatial localisation is more difficult than that illustrated in Figure 4-16

Calculation of the 'Localisation Index'

Calculation of the 'Localisation Index' consists of two steps:

1. For each detected NRC event $k \in \{1, 2, \ldots, |\mathbf{Z}|\}$, where $|\mathbf{Z}|$ denotes the number of detected NRCs, calculate the mean of the number of connected components that the k^{th} NRC exhibits throughout its lifetime. For the NRCs illustrated in Figure 4-16 and Figure 4-17, this number would be 1.0 (5/5) and 1.7 (5/3) respectively.
2. The maximum mean of the number of connected components for all the detected NRCs throughout their lifetime is the localisation index of an NRC detection method.

As a summary, the higher the Localisation Index is, the harder it is to relate an NRC with an incident. As the average number of connected components increases, it becomes more likely that each connected component could be related to a different incident, because different connected components are not spatially connected. On the other hand, the minimum value of the Localisation Index is one, which is the best value of Localisation Index as the detected NRCs consist of only a single connected component throughout their lifetime.

Last, but not least, the reasoning behind the proposal of 'Localisation Index' is supported within the literature on congestion formation on networks. It has been reported that traffic congestion events due to an incident (e.g. obstruction) in an idealised road network (i.e. a grid network where the speeds of all vehicles are assumed to be the same) mostly consist of a single connected component. However, it is also possible that such a congestion event may split into two connected components, depending on travel demand (Wright and Roberg-Orenstein, 1999). Once the congestion is split into two connected components, relating these two connected components to the initial cause becomes more difficult (i.e. a higher Localisation Index). Recent research on understanding the formation of congestion due to an incident has also focused on grid networks, and research on irregular road networks is left to future studies (Long et al., 2012).

4.4. Summary

The performance of every NRC detection method should be subject to a thorough evaluation. It is shown in section 2.5 that two main strategies can be used to evaluate the performance of an NRC detection method. The first compares the detected NRCs with an external data set, which is often referred to as 'ground truth data'. The 'ground truth data' is the data that defines the 'truth' and each NRC detection method aims to find solutions that closely match the 'ground truth data'. However, such 'ground truth data' are usually not available in practice, since it is cost-prohibitive for traffic operators to label LJT data and there is a lack of research on simulating NRCs on a large urban road network. As an NRC is usually assumed to be the outcome of an incident, section 4.2 investigated whether an incident data set can be used as ground truth data to evaluate the performance of an NRC detection method. This investigation determined five reasons that prevent incident data sets from being used for this purpose.

- Locations of ANPR and CCTV cameras do not necessarily overlap. It is observed that at least half of the recorded incidents occur in locations where there is no ANPR link. This is because these two data sources are collected from different sensors. Focusing only on those incidents that occur on ANPR networks is also not an optimal decision, because remaining incidents such as diverged traffic may affect the links of the ANPR network.
- There may be unrecorded incidents in which a substantial increase in LJTs cannot be explained by the incident data set. This might happen, for example, if the incident was not observed or reported by a traffic operator. Ignoring this observation would increase FAR, as there would not be a corresponding incident for each detected NRC.
- Non-disruptive incidents may be recorded, which occur when an incident that happened on a link does not cause a substantial increase in LJTs. An important reason for this is the length of the links; the longer the link, the less the impact of an incident on the link. In addition, some incidents last for a shorter time than the LJT estimation interval. It is observed that approximately 7% of the incidents last for fewer than five minutes, and observing the effect of these incidents on the estimated LJTs is very difficult. Ignoring this observation would decrease the Detection Rate (DR), since not all incidents cause substantial increases in LJTs that result in an NRC.
- Inactive ANPR cameras are a serious and commonly observed issue. If the link where an incident has occurred has missing data due to an inactive ANPR camera, then the impact of the incident cannot be investigated. It is observed that

approximately 95 links are inactive for at least 30 minutes between 07:00 and 19:00. Ignoring this would decrease DR, as those incidents that occur on a link where there is no LJT estimate cannot be detected.

- Determination of the beginning and end of an incident involves several uncertainties. These uncertainties not only arise due to the subjective interpretation of traffic operators reporting an incident, but they also depend on the location and time of the incident. Therefore, it is difficult to correctly identify which LJTs correspond to an NRC.

These five reasons prevent the available incident data set from being used as ground truth data for the detected NRCs. As a result, it is found that ground truth data are not available to evaluate the performance of NRC detection methods.

In order to compare two NRC detection methods using LJT data, section 4.3 proposed two novel evaluation criteria. Both of these criteria consider the uses of an NRC detection method:

1. High-Confidence Episodes (e^*): Substantially high LJTs lasting for a minimum duration would attract a traffic operator's attention. Such episodes are called 'High-Confidence Episodes'. An NRC detection method should ideally include all such episodes within its detected NRCs. However, detection of an NRC that is not an e^* is inconclusive. This is because there would be NRCs that last for less than the minimum duration value defined in e^*. It is found that whenever the estimated LJTs are at least 40% higher than the expected LJTs for at least 25 minutes, a high-confidence episode occurs.
2. Localisation Index: The detected NRCs should ideally be related to real life events such as incidents. In this way, the impact of incidents could be estimated and the reason for the occurrence of NRCs could be better understood. The number of connected components is used as an indicator to determine to what extent this could be achieved. The higher the average of the number of connected components of a detected NRC, the greater the difficulty in relating that NRC with an incident.

The main issue of comparing detected NRCs with e^*s is that all e^*s are treated the same. In other words, two e^*s that last for the same duration make the same contribution towards the calculation of FAR and FNR, regardless of their severity.

Furthermore, as the design of the criterion considers practical usage, it does not take into account the variability of LJTs in a link.

On the other hand, 'Localisation Index' does not consider the temporal order of the changes in the number of components of an NRC. For example, for the road network configuration given in Figure 4-18(a) where links $a1$ and $a3$ are upstream of link $a2$, both of the detected NRCs as illustrated in Figure 4-18(b) and Figure 4-18(c) would result in the same 'Localisation Index' (i.e. $(2 + 2 + 1)/3 \approx 1.7$), even though it is more likely that Figure 4-18(c) would be observed due to an incident.

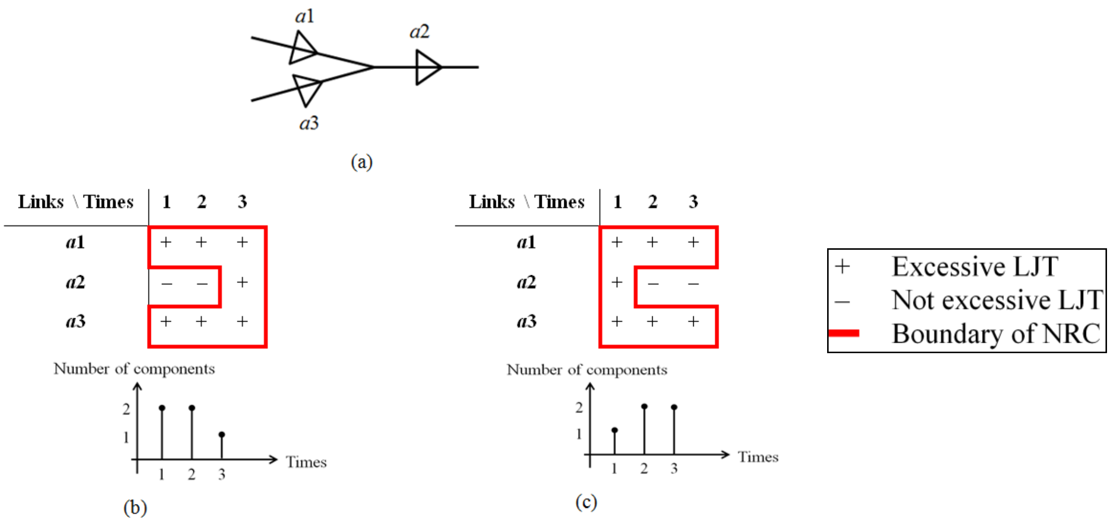


Figure 4-18 Limitation of 'Localisation Index'

The framework to evaluate NRC detection methods is summarised in Figure 4-19. It consists of two main strategies: external criteria and internal criteria. External criteria require ground-truth data, and there are three main ways to obtain it: manual labelling, using traffic simulations and investigating the effectiveness of a separate data source (e.g. incident data set). Because none of these three ways are possible, this thesis proposed two measures that are considered within the internal criteria as they do not require any external data (i.e. High-Confidence Episodes and Localisation Index).

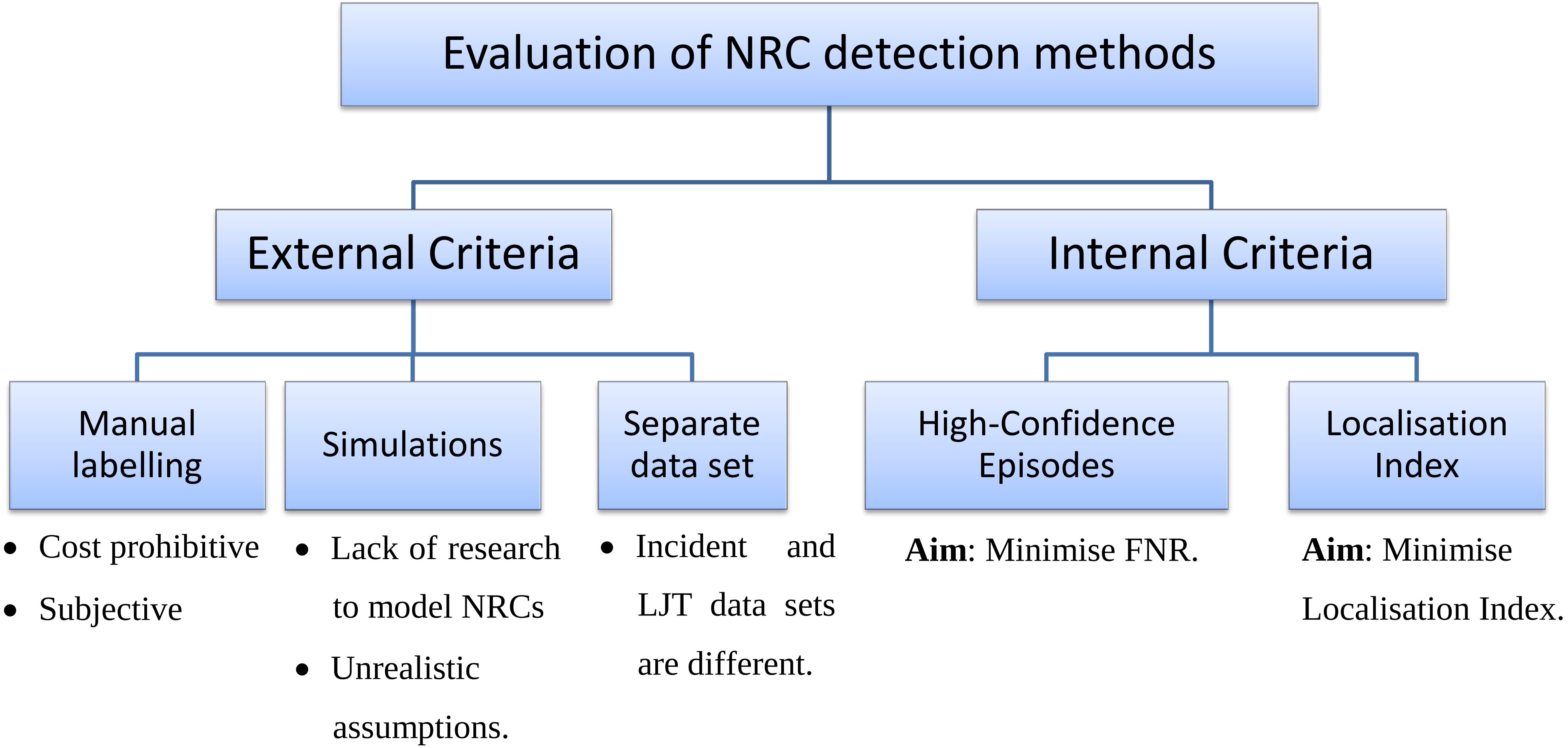


Figure 4-19 The framework to evaluate different NRC detection methods

5. CASE STUDY: DETECTING NRCs IN LONDON'S URBAN ROAD NETWORK

London, the capital city of the United Kingdom, has a long history of urban development which led commuters to experience traffic congestion as early as the 1700s (British Library, 2013). The infrastructure of London consists of an underground (commonly known as 'tube') network covering 402 kilometres (TfL, 2013) and a road network of 13,800 kilometres (TfL, 2012a). However, it still struggles to cope with increasing levels of traffic congestion, which account for almost 20% of the total UK traffic congestion (TfL, 2010, p.86). Approximately 24.4 million trips have been made to/from or within London within a day, and 58% of these require road network transportation (TfL, 2010, p.38). These facts illustrate the importance of the road network in London, and indeed make it very interesting to investigate NRCs in an urban road network.

This chapter applies the proposed methodology to detect NRCs in London's urban road network. The distribution of LJTs on both raw and cleaned LJT data are determined in section 5.1. Decisions regarding the experimental design to analyse London's urban road network are formalised in section 5.2. The performance of the two proposed NRC detection methods are compared in section 5.3. The effectiveness of NRC detection models are explored through some exemplar links in section 5.4. Finally, this chapter is summarised in section 5.5.

5.1. Determining the Distribution of LJTs

Detecting the distribution of LJTs is important for several reasons. Firstly, there is a strong relationship between the distribution of LJTs and travel time reliability. Travel time reliability is also related to NRC, because the frequent presence of NRCs leads to lower travel time reliability. Secondly, in order to derive the likelihood ratio function in in expectation based STSS, a statistical distribution is required to model LJTs (discussed in section 3.3.2). Due to these important aspects, this section aims to determine the distribution of LJTs.

The analysis period is the weekdays between 1 April and 1 August 2010. This period consists of 87 weekdays, which exceeds the number of reported LJTs required to derive reliable statistical estimates in order to model an LJT, which is reported to be approximately 40 (Hollander and Liu, 2008). All of the collected data are analysed between 07:00 and 19:00 as this interval covers the AM/Inter/PM peak periods (TfL, 2010, p.268). Within this interval there are 145 LJT estimations, as an LJT is estimated every 5 minutes. Therefore, 424×145 = 61,480 (i.e. the number of links × the number of time intervals between 07:00 and 19:00) LJTs are tested.

The null hypothesis formulates that an LJT follows one of the four hypothesised distributions: lognormal, gamma, normal (i.e. Gaussian) and exponential. The reason why normal and exponential distributions are also hypothesised is that they will be useful to compare the results, since both of the distributions are theoretically not justified to be eligible for modelling LJTs. This is because normal distribution is a symmetric distribution, whereas an LJT cannot have negative values. On the other hand, exponential distribution does not have a peak which is observed in an LJT distribution, as most trips are completed within a narrow LJT interval. Furthermore, because exponential distribution is a special case of gamma distribution; it is expected that gamma distribution would be more flexible to model LJTs compared to exponential distribution.

Four of the hypothesised distributions are fit using maximum likelihood estimation, and then the Kolmogorov-Smirnov (KS) test is applied to establish whether the difference between the fitted distribution and hypothesised distribution is significant or not at 5% significance level. The KS test is used to ascertain whether each LJT follows a given distribution due to its common usage in the literature to detect the distribution of travel times (Arezoumandi, 2011; Hollander and Liu, 2008; Susilawati et al., 2011). If the *p*-value of the KS-test statistic is lower than 5%, then the null hypothesis is rejected.

Two different analyses are conducted to assess the distribution of LJTs. The first uses all the available data, and the second uses the cleaned data (i.e. after the outliers are removed from the original data). These analyses are discussed in section 5.1.1 and 5.1.2 respectively.

5.1.1. Assessing the Distribution of LJTs on Raw Data

All the available observations for a given LJT (i.e. 87 observations corresponding to the data collection period) are fit to the hypothesized distributions using maximum likelihood estimation. Therefore, this analysis incorporates all the data including the NRCs that happened during the data collection period. The number of rejected null hypothesis (i.e. that the LJT cannot be modelled using the hypothesized distribution), and the percentage that indicates the hypothesised distribution can be used to model all the LJTs (Accuracy) is illustrated in Table 5-1.

Table 5-1 Number of null hypotheses that are rejected using the given distributions

Hypothesized Distribution	# rejected H_0	Accuracy
Lognormal	**22,909**	**63%**
Gamma	27,527	55%
Normal	36,888	40%
Exponential	61,480	0%

These results illustrate that the determined LJT distributions are in line with what has been reported in the literature. It is seen that lognormal distribution describes the LJTs best, as it rejected the fewest hypotheses. Lognormal distribution can be used to describe 63% of the LJTs (i.e. 1 – 22,909/61,480), whereas gamma, normal and exponential distributions describe LJTs by 55%, 40% and 0% respectively. Although the majority of LJTs can be modelled by lognormal distribution, it should also be noted that there is still a huge ratio of LJTs that cannot be modelled using a lognormal distribution. These LJTs that cannot be modelled via a lognormal distribution account for 37% of all possible LJTs, and the causes require further investigation. The effectiveness of the Kolmogorov-Smirnov test can also be observed, as it successfully rejected all the null hypotheses that describes LJTs with an exponential distribution.

5.1.2. Assessing the Distribution of LJTs on Cleaned Data

In the aforementioned analysis, a data cleaning procedure had not been applied, which may explain the substantial number of LJTs not being modelled by lognormal distribution. LJTs are often subject to sampling errors due to low sample sizes, as discussed previously. Thus, it may be that these LJTs that are estimated based on a low sample size prevent a higher modelling accuracy. It is therefore important to investigate how the results change after a data cleaning procedure. The purpose of the data cleaning procedure is to remove any substantially high LJTs that possibly result from a low sample size or a previous NRC event.

Data cleaning can be achieved largely via an outlier removal procedure. It should be noted that most of the existing outlier detection methods (e.g. z-score, Grubbs test) work well under the assumption that data are sampled from normal distribution. However, since LJTs cannot be modelled via a Normal distribution, a test that does not depend on the distribution of the observations should be used. In order to achieve this purpose, an interquartile range method is used to detect the outliers, which is described in Table 5-2 (Tukey, 1977).

Table 5-2 Interquartile range method used to detect outliers in LJT data

Detecting Outliers via the Interquartile Range Method

1. $Y_a(t)$ = Sort the observations in ascending order.
2. Find the 25th, 50th and 75th percentiles (i.e. Q1, Q2 and Q3 respectively).
   2.1. Q1 = median of the lower half of $Y_a(t)$.
   2.2. Q2 = median of $Y_a(t)$.
   2.3. Q3 = median of upper half of $Y_a(t)$.
3. Find the Interquartile range (IQR) = Q3 – Q1.
4. Detect the outliers.
   4.1.Observations which are smaller than (Q1 – 1.5×IQR).
   4.2. Observations which are larger than (Q3 + 1.5×IQR).
5. Remove the outliers.
6. Repeat steps 1-5 for all $a \in \mathbf{A}$, $t \in T$.

, where $Y_a(t)$ corresponds to the 87 LJT observations collected historically for link a at time interval t. The first step sorts these observations in ascending order. In the second step the quartiles are identified, where Q2 corresponds to the median of $Y_a(t)$ (50th percentile), Q1 corresponds to the median of the lower half (25th percentile) and Q3 corresponds to the median of the upper half (75th percentile). The third step calculates the interquartile range (IQR), which is the difference between Q3 and Q1. Finally, outliers are identified in the fourth step, which are identified as those observations that have a lower value than Q1 – 1.5×IQR or a higher value than Q3 + 1.5×IQR. The fifth step removes the outliers from $Y_a(t)$. The procedure described in the previous section is applied to determine the statistical distribution that can model each $Y_a(t)$, and the results are illustrated in Table 5-3.

Table 5-3 Number of null hypothesis rejected for modelling LJTs on cleaned data

Hypothesized Distribution	# rejected H_0	Accuracy
Lognormal	**7300**	**88%**
Gamma	8099	87%
Normal	11879	80%
Exponential	61480	0%

Removing outliers resulted as an improvement to model the LJTs, as the number of rejected null hypothesis decreased dramatically. Lognormal distribution can be used to model 88% of the LJTs and, again, is chosen as the most plausible distribution to model the LJTs amongst others. On the other hand, Gamma, Normal and Exponential distributions can be used to model 87%, 80% and 0% of all LJTs respectively. The similarity between results from lognormal and Gamma distributions is worth noting and requires further investigation. In addition, normal distribution has also been found to approximate 80% of the LJTs. However, as previously discussed, Normal distribution is theoretically not valid to model LJTs, as it is a symmetrical distribution that allows sampling of negative LJTs, which is not possible in reality. Nevertheless, as lognormal distribution results in the highest accuracy, it is assumed that LJTs can be modelled by this distribution. The parameters of each lognormal distribution (i.e. location and scale) that fit the data are used to estimate the expected LJTs that are used in STSS.

5.2. Experimental Design

This section describes the procedure to determine decisions regarding the design of the experiments that take place throughout this chapter. It comprises three decisions: the effect of weekdays on estimated LJTs, the effective analysis period and the adjacency of anti-parallel links. In addition, three types of days are determined, so that the developed NRC detection methodology can be tested on different travel demand levels.

5.2.1. Determining the Effect of Weekdays on LJTs

This section investigates whether there is a significant effect caused by working days of the week (i.e. Monday to Friday) on LJT estimates. The investigation carried out in this section seeks to find out whether or not two weekdays could be treated in the same manner for statistical analysis with regard to estimated LJTs. Weekends are excluded from this analysis because most home to work trips that contribute a substantial part of weekday traffic do not occur at the weekend (TfL, 2010, p.103).

One-way analysis of variance (ANOVA) is a hypothesis testing method that is suitable to investigate whether or not an estimated LJT on link a time interval t is the same on all the weekdays. The null hypothesis is that there is no weekday effect on the estimated LJTs; hence, the expected LJT is the same on all weekdays. On the other hand, the alternative hypothesis is that there is a significant difference between weekdays. Formally,

H_0: $\overline{y}_a^{\text{Mon}}(t) = \overline{y}_a^{\text{Tue}}(t) = \cdots = \overline{y}_a^{\text{Fri}}(t)$, where $\overline{y}_a^{d}(t)$ is the expected LJT on link a at time interval t on a weekday d.

H_1: At least one weekday is significantly different from the other weekdays regarding the expected LJT on link a time interval t.

In order to conduct ANOVA, historical data should be collected and estimated LJTs should be grouped based on the day of the week. Having grouped the historic LJT data based on the day, three types of variation across the historic LJT data can be identified:

- Variation between LJTs within the same weekday.
- Variation due to experimental error.
- Variation due to the differences between weekdays.

The first variation states that LJT estimates may vary between observations of the same weekday. This would suggest, for example, that an LJT estimate would be different from one Monday to another. The second variation is due to imperfections in the data collection process, including technical faults within the ANPR cameras that result in erroneous or missing LJT data. The third variation arises due to the differences between weekdays, and it is the main type of variation that ANOVA is designed to identify.

In order to determine the effect of weekdays on the estimated LJTs, two types of one-way ANOVA are worth investigating. The first is called 'independent measures ANOVA' and the second is 'repeated measures ANOVA'. The first type considers the first two variations together; hence, the variation between LJT estimates is not removed on different observations of the same weekday. The second is called 'repeated measures

ANOVA' and it neglects the first variation; hence, it considers all weekdays to be the same within themselves. Because an LJT estimate is based on the journey times of the captured vehicles, it is unlikely that the same vehicles will be captured to estimate an LJT during different observations of the same weekday. Therefore, this thesis relies on the independent ANOVA method.

In order to conduct one-way independent measures ANOVA, the following three assumptions must be validated. Firstly, ANOVA assumes that the observations follow a Normal distribution. The empirical analysis that was discussed in section 5.1 shows that LJTs are distributed according to a lognormal distribution. Therefore, instead of using the raw LJT estimates, their logarithms should be used within ANOVA, since the logarithm of the LJTs would follow Normal distribution. Secondly, it assumes that the variance of the observations (i.e. the logarithm of LJTs) is equal for all weekdays. Several tests can be used to investigate this assumption, including Bartlett's, Levene's and the Box-Andersen test. Because Levene's test is regarded as robust (Lim and Loh, 1996), it is selected to determine whether the logarithm of LJTs have equal variances between different weekdays. Based on the experimental setup that will be described shortly within this subsection, approximately 94% of the logarithm of LJTs has equal variances between different weekdays. Thirdly, the LJT estimates on different weekdays are assumed to be independent. This is because prevailing traffic conditions are different from day-to-day, so this thesis assumes that the estimated LJTs on different weekdays are independent from one another. These findings suggest that one-way independent measures ANOVA can be used to determine whether there is a significant effect of a weekday on the expected LJTs.

One-way independent measures ANOVA analysis consists of the following steps:

1. Collect *n* LJT estimates on link *a* at time interval *t* grouped according to weekday. The collected data can be illustrated as; $\begin{bmatrix} x_1^{\text{Mon}} & \cdots & x_1^{\text{Fri}} \\ \vdots & \ddots & \vdots \\ x_n^{\text{Mon}} & \cdots & x_n^{\text{Fri}} \end{bmatrix}$, where x_i^d denotes the logarithm of the LJT estimate on link *a* time interval *t* on a day indexed by *d*, where $d \in \{\text{Mon}, \ldots, \text{Fri}\}$. The subscript *i* denotes the observation ID of the LJT estimate, where $i \in \{1, 2, \ldots, n\}$. Note that the logarithm of the LJT estimates is used in order to comply with the assumptions of ANOVA.

2. Calculate the mean for each weekday and denote it with $\overline{x}^d = \sum_{i,d} x_i^d / n$. The grand mean of the data set is calculated by taking the means of weekdays and denoted as $\overline{\overline{x}} = \sum_d \overline{x}^d / D$, where D represents the total number of weekdays and is equal to five.
3. Calculate the Mean of Squares Between (MSB) the weekdays by $\mathrm{SSB}/(D-1)$. The numerator is the Sum of Squares Between (SSB) the weekdays, and is calculated by $\mathrm{SSB} = n\left(\sum_d \left(\overline{x}^d - \overline{\overline{x}}\right)^2\right)$. The denominator is the degrees of freedom to calculate SSB. The mean of squares between weekdays is used to determine the variation between different weekdays.
4. Calculate the Mean of Squares Within (MSW) the weekdays by $\mathrm{SSW}/D(n-1)$. The numerator is the Sum of Squares Within (SSW) for all weekdays and is calculated by $\sum_{i,d}\left(x_i^d - \overline{x}^d\right)^2$. The denominator is the degrees of freedom to calculate SSW. The MSW arises due to the variations between LJT estimates on each weekday and also the variation contributed due to experimental error. Therefore, MSW includes the first two variations, which were listed previously.
5. Calculate the empirical F-ratio by $\mathrm{MSB}/\mathrm{MSW}$ and denote it with $\mathrm{F}_{\mathrm{Empirical}}$. The theoretical F-ratio ($\mathrm{F}_{\mathrm{Theoretical}}$) is determined by using the degrees of freedom $(D-1)$ and $D(n-1)$ respectively in the F-distribution.
6. If $\mathrm{F}_{\mathrm{Empirical}} > \mathrm{F}_{\mathrm{Theoretical}}$, then reject the null hypothesis.
7. The steps between 1 and 6 are used to determine whether the means of a single LJT estimate can be assumed to be the same on different weekdays. In order to generalise the results, the steps between 1 and 6 are applied to all the links, $a \in \mathbf{A}$, across the entire time period $t \in \{1,2, \dots, T\}$. This yields testing $|\mathbf{A}| \times T$ hypothesis.

The analysis is conducted on all the available LJT data during 2010. Apart from three weeks in March for which there is no LJT data, the remaining 49 weeks are used within the analysis. Hence, the number of observations for each LJT is 49 (i.e. n = 49). No outlier removal process is performed in order to minimise the assumptions regarding LJT data. The period between 07:00 and 19:00 is analysed in order to cover the morning/inter/evening peak periods (TfL, 2010, p.268). The analysis period is inclusive, so the continuous time for the study period is 12 hours and five minutes. Each LJT is estimated every five minutes; hence, in total there are 145 LJTs within the analysis period (i.e. T = 12 LJTs/hr × 12 hr + 1 = 145 LJTs). Because there are 424 links, a total of 424×145 = 61,480 LJTs are tested with the aforementioned procedure.

Of these tests, 17,675 have been rejected which means that 28.8% of expected LJTs are *not* equal on different weekdays. Because the majority of expected LJTs are not significantly different on different weekdays, this thesis assumes that there is *no* statistically significant effect of day of weekday on the estimated LJTs. This finding adds further support to the existing literature (Dowling et al., 2004; Noland and Polak, 2002; Yeon et al., 2009) and state-of-practice (TfL, 2010, p.100).

5.2.2. Determining the Effective Analysis Period

Traffic data are estimated based on a sample of observations collected over time. Often, there is a high level of variation within the estimated traffic data, which might be due to missing data from certain data collection intervals. The traditional way of decreasing the variability of traffic data is to 'smooth' it (Stephanedes and Chassiakos, 1993). Various methods can be used to achieve this and vary from using moving or exponential averaging to more complex signal processing methods, including wavelet transform. However, there is usually a lack of information on choosing an appropriate smoothing function and its parameters. Therefore, researchers usually decide these parameters based on empirical observation (Luk et al., 2001). As the smoothing process alters traffic data by reducing the information content within it (e.g. the sample size to estimate each LJT will not be meaningful anymore) and also smooths the NRC patterns that are of interest to this thesis, smoothing is found to be incapable of handling the high variability of the estimated LJTs.

The quality of LJTs on a link does not remain the same all day, as the estimation of each LJT depends on sample size. Variations in travel demand, as well as local traffic conditions and weather, affect the sample size which is defined as the number of vehicles whose journey times are used to estimate an LJT. The lower the sample size used to estimate an LJT, the higher the likelihood that the LJT estimate is susceptible to sampling error. A consequence of sampling error is the observation of unrealistically high LJTs (for example see Figure 4-10). Failure to consider this fact might result in misleading conclusions. Such a situation is discussed in Appendix B. Therefore, it is reasonable to focus on a more limited analysis period which has a higher sample size and hence, a higher data quality. This period is referred to as the *effective analysis period*, and this section illustrates how it can be determined.

The *effective analysis period* is defined as the maximum time interval within a weekday that maximises the quality of the estimated LJTs. The following analysis is conducted to determine the effective analysis period. LJTs are collected for all the weekdays between 1 April and 1 August 2010, which amounts to 87 observations per LJT estimate. Then, LJTs which are estimated at five minute intervals are aggregated into hourly data. Hence, for each hour there are 87 × 12 = 1,044 samples for a link. The *median* sample size of these 1,044 samples is allocated to the categories illustrated in Table 5-4 to classify the quality of a link during the analysed hour.

Table 5-4 Categorisation of median sample size

Median Sample Size	**Label of the category**
0	Patch
1	Low
Between 2 and 5	Average
6 or more	High

If the median sample size is zero, then the link at that hour is considered to consist of only patched LJTs. If the median sample size is one, then the link at that hour is likely to show a sampling error, because a single vehicle's journey time is mostly used to estimate an LJT. If the median sample size is between two and five, then the link is considered to display an average quality LJT, and finally a link with a median of six or more sample size is considered to be a good quality link. This four-level classification of LJTs estimates is determined based on discussions with domain experts. The numbers of links falling into each category are plotted in Figure 5-1.

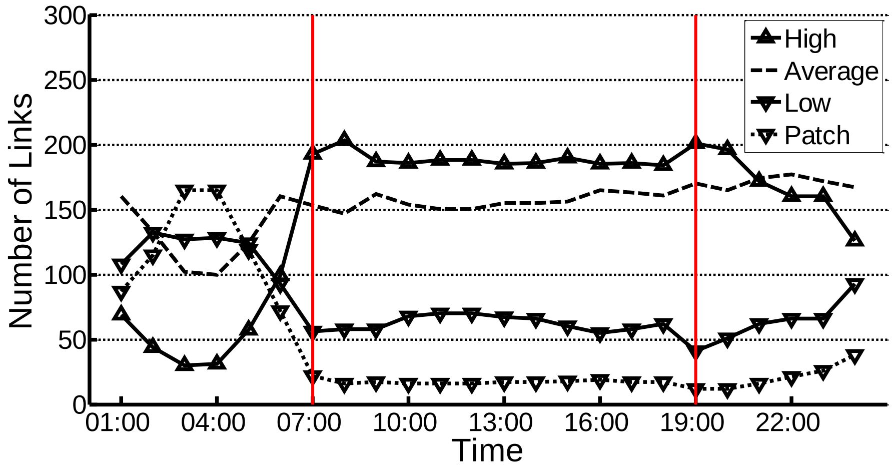


Figure 5-1 Variation in the number of links with different qualities

It is clearly visible that between 07:00 and 19:00, the number of good quality links is at its highest with approximately 200 links. It is observed that before 07:00 the number of highest quality links is at its lowest. In addition, after 19:00 some of the high quality links drop to average quality or are patched, as the number of good quality links decreases with a corresponding increase in the number of patched or biased links. In addition, a further 150 links achieve average quality in the period between 07:00 and 19:00. Because there is the potential for more than 75% of links to be reliable between 07:00 and 19:00, the effective analysis period is chosen as such. This empirical finding is also in line with operational purposes at TfL as it covers the AM/Inter and PM peak periods (TfL, 2010, p.268). There are a total of 145 LJTs within the effective analysis period (12 hours × 12 LJTs/hour + 1, where one is added as the effective analysis period is inclusive of 07:00 and 19:00).

5.2.3. Determining the Adjacency of Anti-Parallel Links

Two links are anti-parallel, a term which is introduced in section 3.1, if one link begins/ends where the other ends/begins. Even though anti-parallel links are not upstream of each other (i.e. most commuters do not make U-turns), it is still important to determine whether anti-parallel links should be considered adjacent or not. This is because an incident occurring on one of the anti-parallel links could affect the other, as the common junction may become congested. Most of the existing research overlooks the issue of anti-parallel links, which is quite a common phenomenon in an urban road network. In London's ANPR network, there are 102 links that count as anti-parallel, corresponding to almost a quarter of the entire network. Therefore, this section investigates whether or not anti-parallel links should be considered to be adjacent, based on the empirical evidence provided by the incident data set. If incidents that occur on anti-parallel links affect both links, then they should be considered to be adjacent. This is because; an NRC would likely to be observed on both of the anti-parallel links.

Incidents that occurred on anti-parallel links in 2010 are found by assuming a spatial buffer of 10 metres around each incident in order to have a better match between the incident data set and ANPR network. The total number of incidents that occurred on either of the anti-parallel links is 2,207. Of these incidents, 42.5% affected both directions of traffic flow (i.e. 938 incidents are reported to affect both traffic flow directions). As the majority of the incidents did not affect both directions of traffic flow, this thesis considers that anti-parallel links are not adjacent.

5.2.4. Types of Dates for Detecting NRCs

Travel demand may change considerably on an urban road network, depending on the date of analysis. This makes the analysis of road networks a difficult task. This thesis applies the developed NRC detection and evaluation methods to three types of dates. Firstly, what is referred to as 'normal' dates are analysed, and on such dates there is no reason for traffic operators to expect very high or low travel demand. Secondly, bank holidays are analysed and these dates are assumed to correspond with lowest travel demand (TfL, 2010, p.88). Thirdly, dates of tube strikes are analysed and on such dates most of the commuters who normally use the tube network use the road network instead, which increases the travel demand considerably (Tsapakis et al., 2013). Analysis of these three types of dates is intended to grant a thorough understanding of the developed NRC detection methodology with regard to its application on dates with different travel demands.

The analysed dates, classified according to their anticipated travel demand, for 2010 are illustrated in Table 5-5.

Table 5-5 Analysed dates of 2010

Bank Holidays	1 January , 2 April, 5 April, 3 May, 31 May, 30 August, 27 December, 28 December.
Normal Days	Weekdays of October except for a tube strike day (i.e. 4 October).
Tube Strikes	7 September, 4 October, 3 November, 29 November.

5.3. Detection of NRCs on Different Travel Demand Levels

Two NRC detection methods have been proposed in Chapter 3. These methods are referred to as Clustering Episodes (discussed in section 3.2), and clustering significant space-time regions, determined by expectation based STSS (discussed in section 3.3). These two methods are evaluated based on different parameter settings in the next two sections. The three types of travel demand levels illustrated in Table 5-5 are analysed during the effective analysis period (i.e. 07:00 – 19:00) to detect NRCs.

5.3.1. Detection of NRCs by Clustering Episodes

The main parameter for using the Clustering Episodes (denoted as ‘CE’) method to detect NRCs is the congestion factor. Five different congestion factors are investigated and the congestion factors for these five different NRC detection models[8] are set to 1.2, 1.4, 1.6, 1.8 and 2.0 in accordance with the values which are currently used at TfL (see Table 4-1).

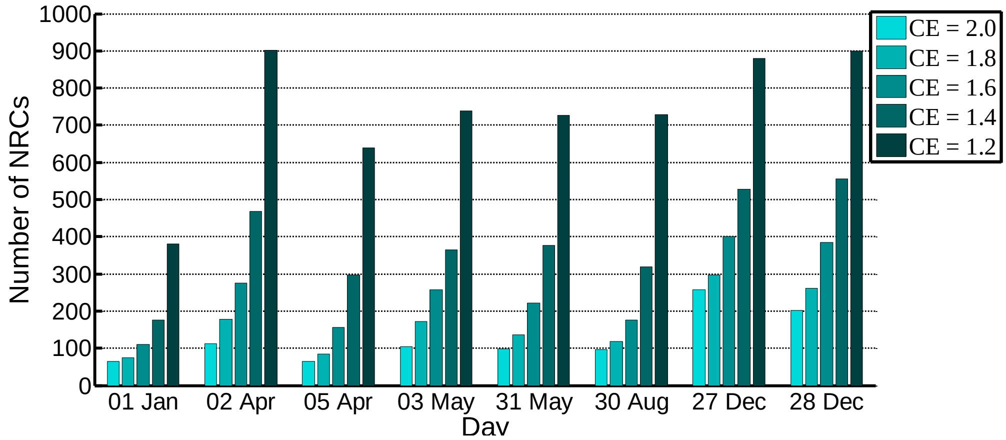
NRCs detected from these five NRC detection models are compared in the following aspects: firstly, the numbers of detected NRCs are compared. Because the correct number of NRCs is not known due to the lack of ground truth data, the number of NRCs is not a performance criterion. Nevertheless, considering that a traffic operator has a fixed allocated time to analyse the detected NRCs, the higher the number of NRCs, the less time to available to investigate each NRC. On the other hand, the lower the number of NRCs is the higher the likelihood that some NRCs are missed. Therefore, it is interesting to analyse the most severe NRC in detail, as it is probably the first NRC that would be investigated. Therefore, the second aspect focuses on the two characteristics of the most severe NRC, which are its severity and lifetime. In all of these different analyses, the number associated with each ‘CE’ denotes the congestion factor used to determine the episodes.

Number of Detected NRCs

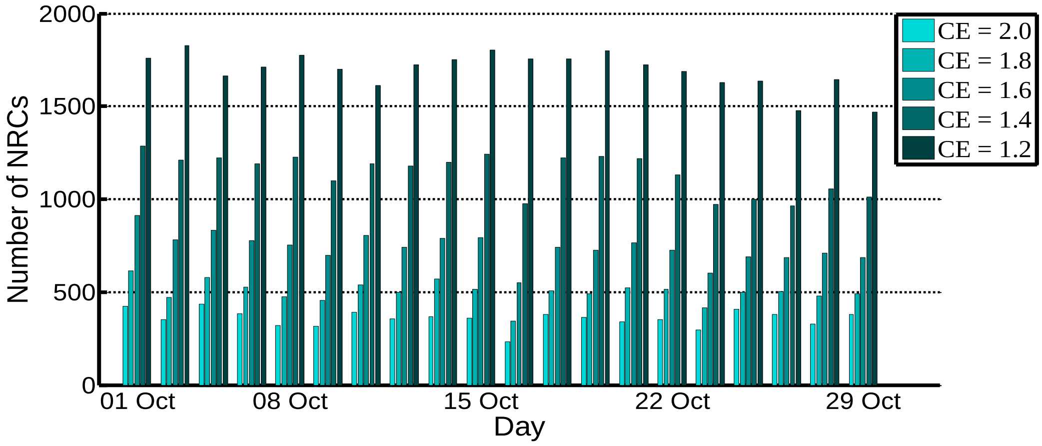
The numbers of NRCs detected by the five different CE models are illustrated for the bank holidays, normal dates and tube strikes in Figure 5-2 (a-c) respectively.

---

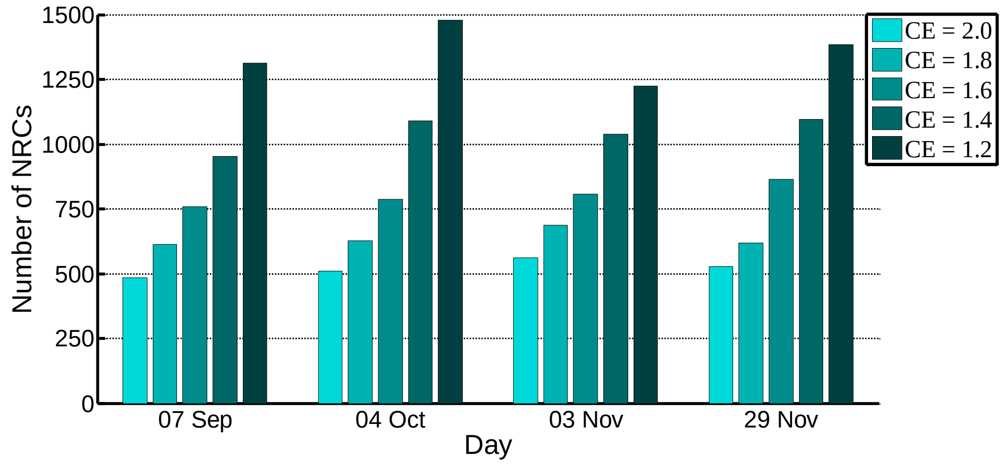
[8] An NRC detection model is a realisation of an NRC detection method by setting the required parameters of the method.



(a)



(b)



(c)

Figure 5-2 Number of detected NRCs on bank holidays (a) Normal days (b) and tube strikes (c)

Two notable outcomes can be noted from the observed results on the number of detected NRCs:

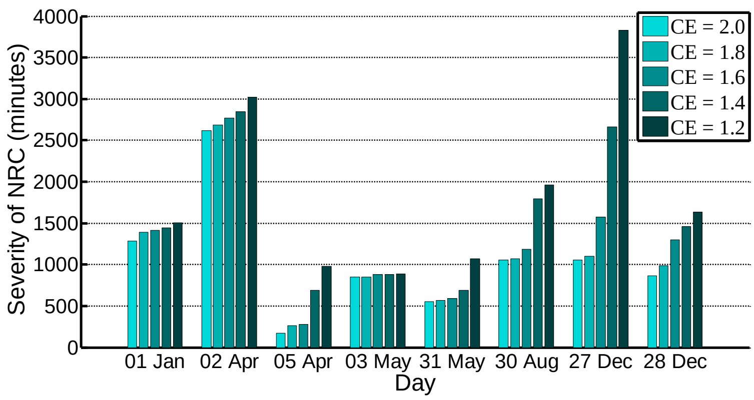
- Decreasing the congestion factor resulted in more NRCs. Initially, this may appear to be trivial as more excessive LJTs would be generated by decreasing the congestion factor; hence, more NRCs. However, it is also possible to merge distinct

NRCs by lowering the congestion factor, which may decrease the number of detected NRCs. This is especially true on highly congested dates like tube strikes. For example, when normal dates and tube strikes are compared by the most liberal way of detecting NRCs (i.e. CE=1.2), it appears that tube strike days experienced fewer NRCs than normal dates. This is because, as an NRC detection model becomes more liberal on highly congested days, the model merges more distinct NRCs than producing newly created NRCs.

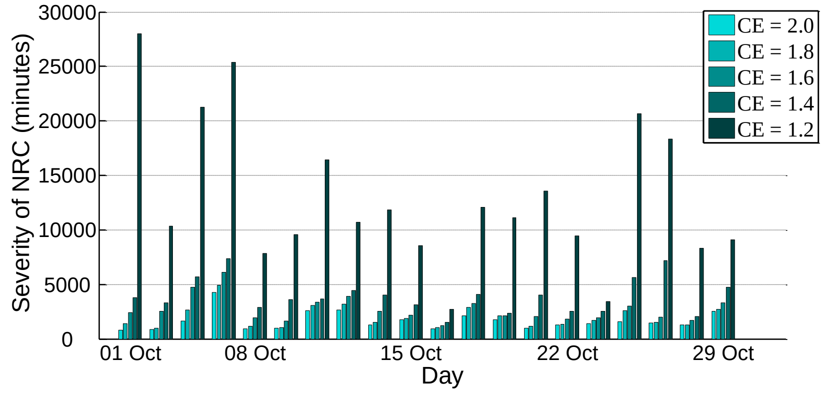
- As the congestion factor increases, the CE method becomes more conservative when detecting NRCs, because an LJT becomes less likely to be excessive. Once the most conservative model is analysed (i.e. CE=2.0), a clear trend between the different types of day appears; as travel demand increases, the number of detected NRCs also increases. Specifically, the number of NRCs is lowest on bank holidays and highest on tube strikes, while normal dates fall in between bank holidays and tube strikes when CE=2.0. This finding suggests that the merging of distinct NRCs is less likely to occur when the NRC detection model becomes conservative.

Characteristics of the Most Severe NRC

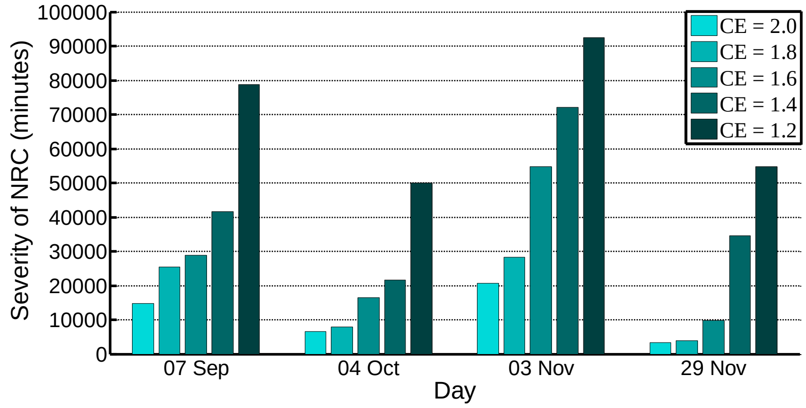
The severity and duration of each NRC can be calculated by representing each NRC in its evolution. In this way, NRCs could be prioritised based on their severity, duration or both. In this part, the focus is placed on the most severe NRC, as it caused the highest impact on the road network in terms of excessive LJTs. Analysis of the most severe NRC is interesting in two aspects. Firstly, road network performance monitoring can be performed effectively by investigating the reason(s) behind the most severe NRC. Secondly, a more comprehensive understanding of an NRC detection model can be achieved. The severity of the most severe NRC as detected by each model on bank holidays, normal dates and tube strikes are illustrated in Figure 5-3(a-c) respectively.



(a)



(b)



(c)

Figure 5-3 Severity of the most severe NRC on bank holidays (a), normal dates (b) and tube-strikes (c)

The following observations can be made regarding the severity of the most severe NRC on the three types of dates in 2010.

- There is a considerable difference between the severity of the most severe NRCs on bank holidays and tube strikes. The most severe NRC on tube strike days is approximately 10 to 35 times higher than bank holidays, depending on the congestion factor. This highlights the extent to which the most severe NRCs can vary. These substantial differences demonstrate the difficulty of relying on a single NRC detection model which is suitable for all different travel demand levels.

- Decreasing the congestion factor increases the severity of the most severe NRC. This outcome is expected, as decreasing the congestion factor results in more excessive LJTs. However, on some bank holidays this observation is not that obvious. For example, on 3 May decreasing the congestion factor barely increased the severity of the most severe NRC. This suggests that even the most severe NRC could be quite stable, regardless of the congestion factor. However, this situation is observed to be very rare.

The second characteristic of an NRC, its lifetime, is illustrated for the most severe NRC detected by each CE model on bank holidays and tube strikes in Figure 5-4(a) and Figure 5-4(b) respectively. Note that the lifetime for normal dates is shown in Appendix C (i.e. Figure C- 2), so that a better legibility of lifetimes on bank-holidays and tube-strikes can be achieved here. Nevertheless, the following findings are also valid for normal dates.

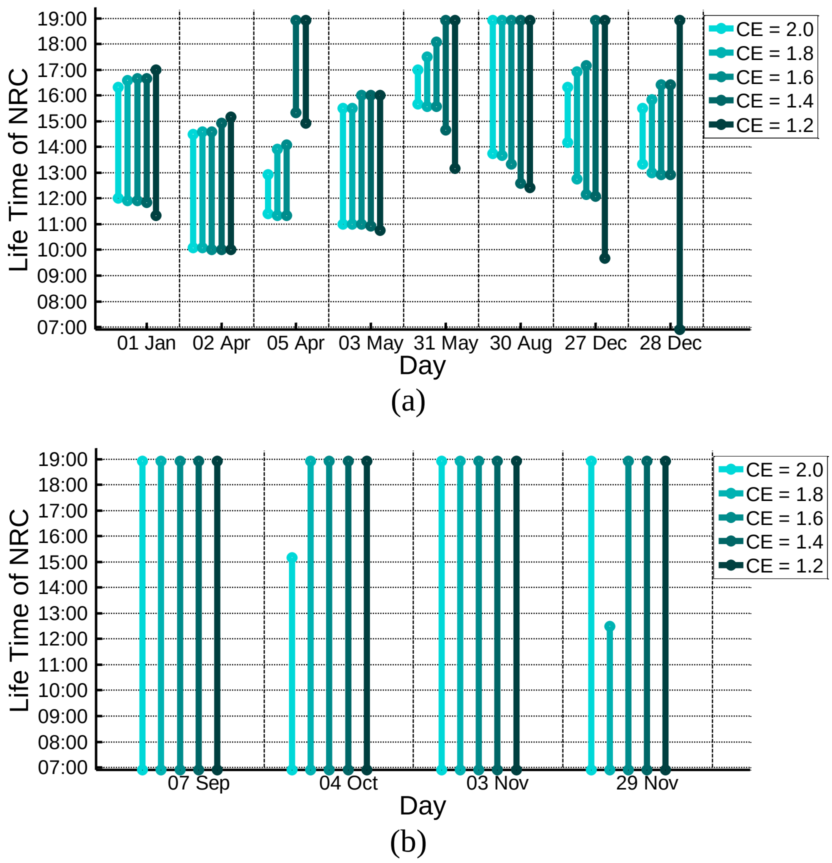


Figure 5-4 Lifetime of the most severe NRC on bank holidays (a) and tube strikes (b)

Decreasing the congestion factor in CE may lead to merging distinct NRCs. For example, on 5 April (a bank holiday), the lifetime of the most severe NRC when CE=1.6 does not coincide with CE=1.4. If the most severe NRC occurred on the same links, the lifetimes of these different models must have coincided. This is also validated by inspecting the severity of the most severe NRC in Figure 5-3(a). It can be observed that the severity of the most severe NRC increases substantially when the congestion factor decreases from 1.6 to 1.4 on 5 April. Empirical investigation validates that indeed two distinct NRCs are merged when the congestion factor is decreased from 1.6 to 1.4. A similar situation occurs on the tube strike day 29 November, as the lifetime of CE=1.8 is shorter than the lifetime of CE=2.0. The most severe NRC on tube strike days spanned the entire analysis period most of the time indicating the extent to which the road network is congested.

Evaluating the Performance of NRC Detection Methods

This section evaluates the detected NRCs based on the two criteria proposed in section 4.3. Specifically, 'high-confidence episodes' and 'Localisation Index' are evaluated for the aforementioned NRC detection models based on CE. In addition to the two proposed criteria, run-times[9] of the models are also compared in order to illustrate the required time to detect NRCs on the analysed dates. If two NRC detection models perform in similar ways, then it would be reasonable to prefer the faster one.

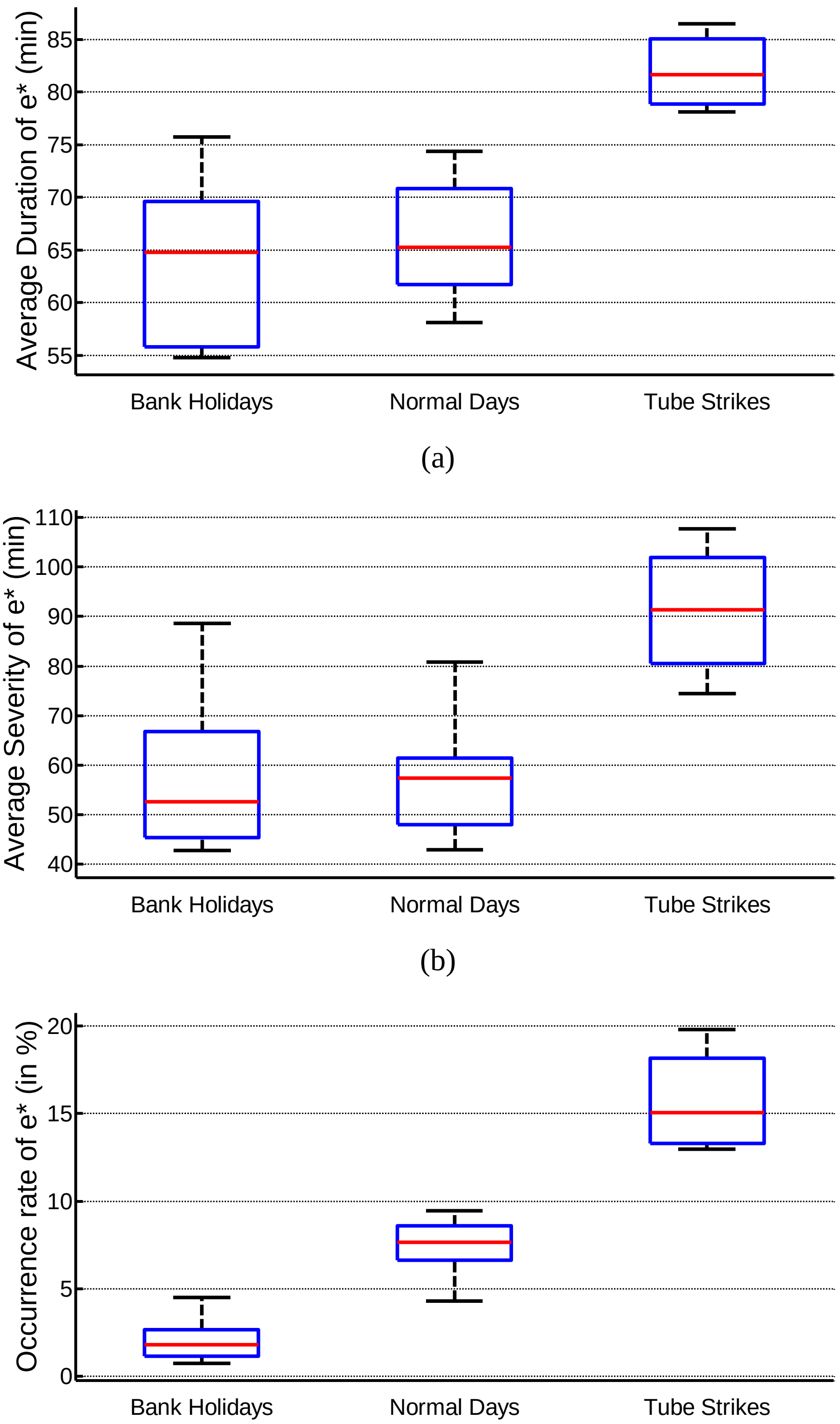
High-Confidence Episodes (e^*)

The empirical evidence provided in section 4.3.1 suggests that a high-confidence episode (e^*) occurs on a link when the estimated LJTs are at least 40% higher than their expected values for at least a minimum of 25 minutes. Previous analysis has shown that e^*s cause the majority of total delay; hence, their duration and severity is considered to be sufficient to attract a traffic operator's attention. The definition of an e^* is used to detect all the e^*s. Having determined all e^*s, this performance criterion assesses to what extent an NRC detection model detected e^*s. The analysis is conducted on the three types of dates: bank holidays, normal days and tube strikes. It is assumed that the more an NRC detection model detects e^*s (i.e. low FNR), the better it becomes.

In order to gain a better understanding of the comparison between the detected NRCs and e^*s, it is necessary to provide the characteristics of e^*s that have occurred on the analysed dates. For all the dates being analysed, the average duration and severity of e^* is illustrated in Figure 5-5(a-b) respectively for all the three types of dates. In addition, the percentage of $e^*$to the total number of estimated LJTs is shown in Figure 5-5(c). Note that there are a total of 61,480 LJT estimates (424 links × 145 time intervals).

---

[9] The analyses throughout this chapter are conducted on a workstation running on Windows 7 with a 12 GB RAM and 3.2 GHz processor. The code is written in Matlab R2012a.



(a)

(b)

(c)

Figure 5-5 Average duration (a) and severity (b) of high-confidence episodes and the percentage of LJTs that belong to high-confidence episodes (c)

The analysis of the characteristics of e^*s reveals some interesting outcomes:

- The average duration of high-confidence episodes is very close on bank holidays and normal days and is approximately 65 minutes. This suggests that there is no substantial difference between bank holidays and normal days regarding the duration of an e^*. On tube strike days, however, the average duration of e^*s increase to 82 minutes.

- The average severity of e^*s increases as travel demand increases. It is observed that the average severity of e^*s is lowest on bank holidays and cause approximately 53 minutes of excessive LJT. This increases to approximately 57 minutes on normal days and to 91 minutes on tube strike days. These results suggest that the severity of e^*s on bank holidays and normal days is not substantially different, whereas there is a noticeable difference between these two types of days and tube strikes.
- The amount of e^* to the total number of LJTs also increases as travel demand increases. On bank holidays, approximately 2% of LJTs are included within an e^*, whereas this ratio increases to 7.5% on normal days and to 15% on tube strikes. These results demonstrate the difficulty of detecting NRCs, because if all LJTs are considered *not* to belong to an NRC, then the accuracy of correct identification of traffic situation would be at least 85%.

The average False Negative Rate (FNR) is plotted against the average False Alarm Rate (FAR) in order to illustrate the relationship between the e^*s that are detected and missed by different NRC detection models. The FAR and FNR values for each NRC detection model are calculated for all the bank holidays, normal days and tube strikes. The calculated FAR and FNR values are then averaged and the results are illustrated in Figure 5-6. The congestion factor that is used to detect NRCs by CE is denoted with the number appearing next to the label 'CE='.

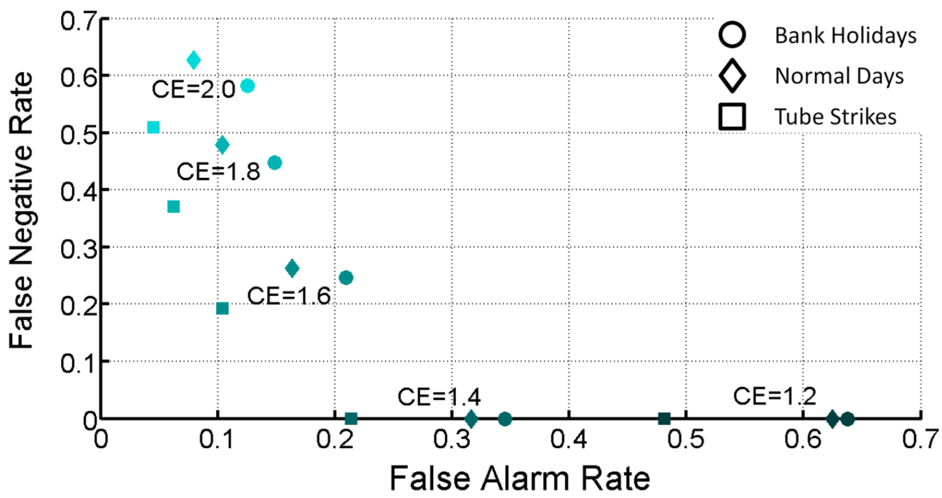


Figure 5-6 The performance of CE models when detecting high-confidence episodes on bank holidays, normal days and tube strikes

The best models with regard to detecting e^*s are the CE models that use a congestion factor of less than or equal to that used by e^*, which is 1.4. As can be seen in Figure 5-6, the CE=1.4 and CE=1.2 models detected *all* the e^*s, leading to an FNR of zero. This is because the underlying principle used to detect NRCs by CE and e^*s relies only on excessive LJTs. Unlike the e^*, however, CE models do not make an assumption regarding the duration of an NRC. Therefore, CE models detect NRCs that are not e^*s; hence, those models have an FAR value that is greater than zero. It is important to stress one more time that it is reasonable for an NRC detection model ***not*** to make an assumption regarding the minimum duration of an NRC, as there could be, for example, accidents which are cleared quickly, yet result in an NRC. Therefore, an ideal NRC detection model should have *some* FAR when compared with e^*s. However, the optimum value of FAR is unknown due to the lack of ground truth data. On the other hand, increasing the congestion factor for values greater than 1.4 makes the CE models more conservative; hence, more e^*s are missed. For example, when the congestion factor is set to 2.0, around 62% of e^*s are missed on normal days.

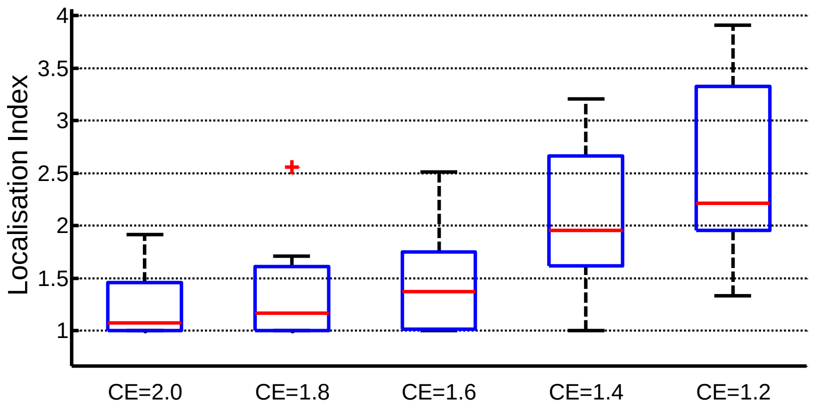
The distinction between tube strikes and the remaining two types of day (i.e. bank holidays and normal days) is noticeable. Specifically, tube strike days resulted in fewer FNR and FAR than the remaining two types. This is because the duration and severity of e^*s on tube strike days is substantially greater than the other types. This increases the performance of the CE model, because NRC is experienced on most parts of the road network for longer durations, which leads to a better match between the detected NRCs and e^*s.

On the other hand, the distinction between normal days and bank holidays is difficult to comprehend. Although normal days are more congested than bank holidays (i.e. there are more e^*s on normal days), more e^*s are missed (i.e. normal days have a higher FNR than bank holidays). The percentage of missed e^*s increases as the congestion factor increases. For example, when the congestion factor is 1.6, there is only a 1% difference in FNR values and this percentage increases to 5% when the congestion factor is increased to two. Although these differences between normal days and bank holidays are very small, the most likely reason behind these results is the more irregular disruptions occurring on normal days (e.g. accidents, engineering works etc.) degrade the performance of NRC detection models.

Localisation Index

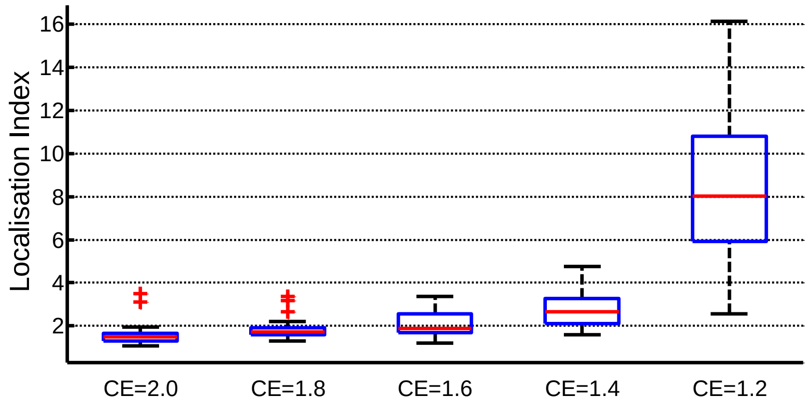
An important use of an NRC detection method is to be able to relate the detected NRCs with incidents, so that the impact of incidents can be estimated. In addition, the reason of the occurrence of the detected NRC would be better understood. The Localisation Index quantifies to what extent an NRC could be related to an incident. It is calculated based on the average of the number of connected components that an NRC exhibits throughout its lifetime (for further details, refer to section 4.3.2). After calculating the average of the number of connected components for each NRC, the NRC with the highest average number of connected components is the Localisation Index of the corresponding NRC detection model.

Localisation Index values are calculated for each bank holiday, normal day and tube strike day based on the five NRC detection models by using Clustering Episodes. The Localisation Index results are summarised for bank holidays, normal days and tube strikes in Figure 5-7(a-c)[10] respectively.

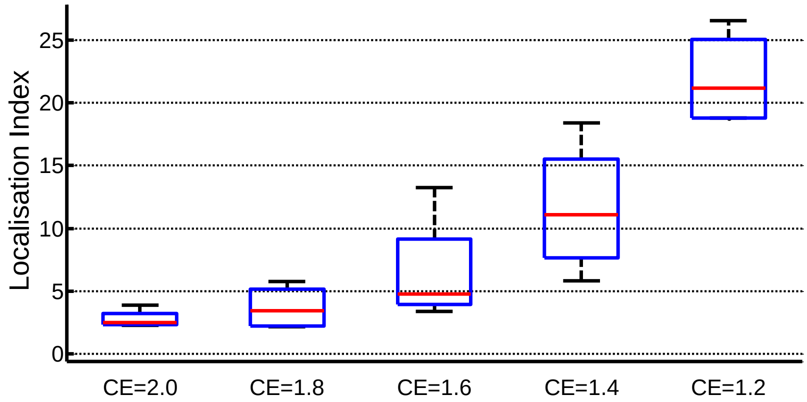


(a)

[10] When plotting box-plots, outliers are illustrated with a '+' sign and are determined based on the description given in 5.1.2 as described in Tukey (1977).



(b)



(c)

Figure 5-7 Localisation Index for bank holidays (a), normal days (b) and tube strikes (c)

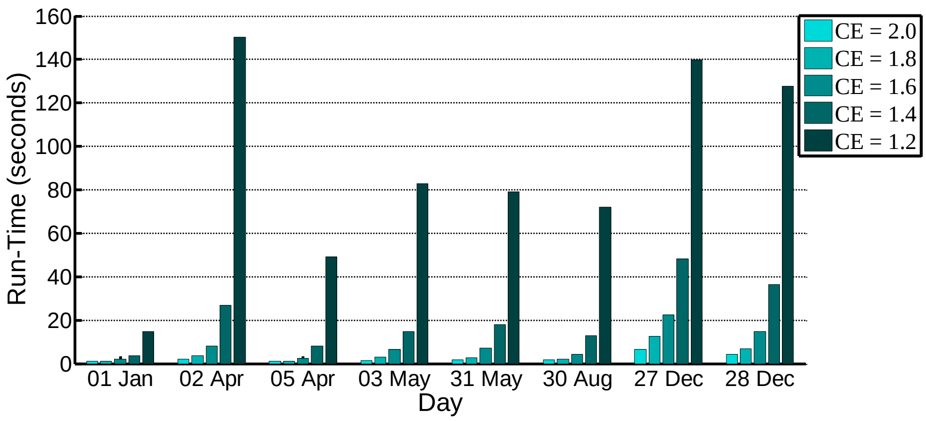
The results generally validate the theoretical expectation: liberal ways of detecting NRCs result in a worse performance in the 'Localisation Index'. This can be observed when the median of the 'Localisation Index' values are investigated. Localisation Index values gradually increase as the congestion factor decreases. This is because the likelihood of day-to-day traffic variations being considered to belong to an NRC increases as the NRC detection model becomes more liberal. Deeming day-to-day variations in LJTs to belong to an NRC may result in the merging of two or more distinct NRCs into a single NRC. Therefore, such liberal ways of detecting NRCs are not preferred in terms of relating detected NRCs with incidents.

It should also be highlighted that the 'Localisation Index' may also decrease if NRCs are detected in a more liberal way. For example, the maximum value of the 'Localisation Index' on a bank holiday when CE=1.8 is 2.55, which is actually detected as an outlier amongst the other bank holidays. On the other hand, the maximum 'Localisation Index' when CE=1.6 is 2.51. This highlights the possibility that more excessive LJTs may also contribute to a lower number of connected components throughout an NRC's lifetime. This is possible, as decreasing the congestion factor leads to more LJTs becoming excessive, and these excessive LJTs may be between the connected components of an NRC. Hence, different connected components will be merged by these additional excessive LJTs, which will then decrease the number of connected components[11]. However, the results show that this is a rare situation and most of the time liberal ways of detecting NRCs lead to a higher Localisation Index.

Another noticeable result regards the travel demand level of the analysed day. Specifically, as congestion levels increase from bank holidays to normal days to tube strike days, the Localisation Index values become higher, indicating the increased difficulty in localising the detected NRCs. For example, it is very difficult to localise the source of NRCs occurring on tube strike days. This is because 'tube strike' events lead to the entire road network being identified as the source of an NRC due to the substantial increase in travel demand across the road network. This means NRCs contain links that are not connected during their lifetime, making the interpretation of detected NRCs more difficult.

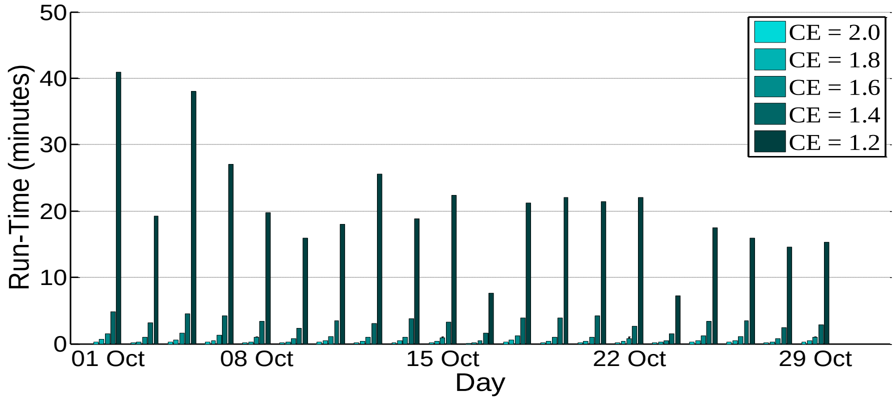
Run-Time

The last part of evaluating the performance of NRC detection models is their run-time. The run-times of the five CE models are illustrated for bank holidays, normal days and tube strikes in Figure 5-8(a-c) respectively.

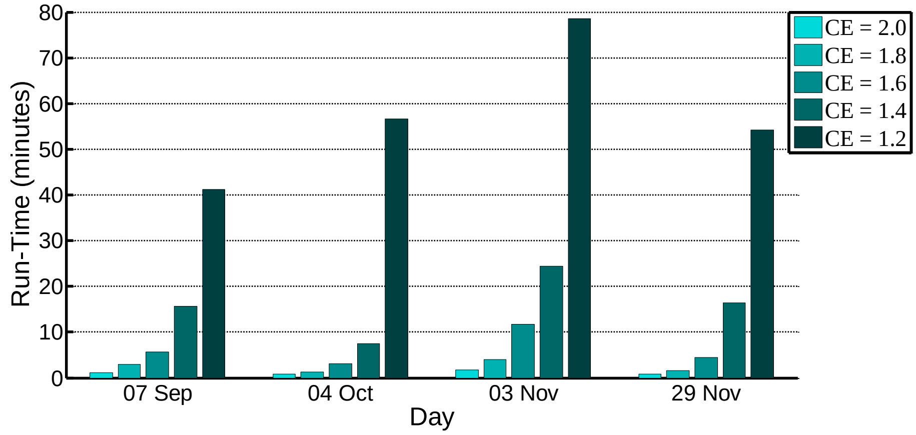


(a)

[11] Consider the example in Figure 4-17. If the LJT at link *a*2 at time 2 becomes excessive when the congestion factor is decreased, then the average number of connected components of the NRC will decrease.



(b)



(c)

Figure 5-8 Run-times of NRC detection models for bank holidays (a), normal days, and tube strikes (c)

The results indicate that on bank holidays, with the exception of CE=1.2, all models detected NRCs in less than one minute. On average, the run-time is 30 to 45 higher on tube strike days than bank holidays. Regardless of the type of analysed date, CE=1.2 resulted in an exceptionally high run-time. This is because it is the most liberal model used to detect NRCs and therefore is most likely to consider day-to-day traffic variations to belong to an NRC. On the other hand, increasing the congestion factor leads to more conservative ways of detecting NRCs, which means there are fewer episodes to cluster; hence, NRC detection is faster.

All the results obtained by detecting NRCs by Clustering Episodes (CE) are summarised for the most severe NRC and for overall results in Table 5-6 and Table 5-7 respectively. In these tables, bank holidays, normal dates and tube strikes are denoted with letters B, N and T respectively.

Table 5-6 Characteristics of the most severe NRC detected by each CE model

CE Model	**Duration (hours)**			**Severity (minutes)**		
	B	**N**	**T**	**B**	**N**	**T**
CE = 2.0	3.20	4.81	11.06	1,057	1,659	11,367
CE = 1.8	3.80	6.09	10.40	1,115	2,010	16,421
CE = 1.6	4.21	7.82	12.00	1,248	2,685	27,469
CE = 1.4	4.92	8.87	12.00	1,558	3,932	42,468
CE = 1.2	6.70	11.69	12.00	1,861	12,936	69,006

The difference exerted by different travel demand levels on the characteristics of the most severe NRC is startling! The most severe NRC is 10 – 35 times more severe on tube strike days compared to bank holidays. This substantial difference suggests there is a difficulty in relying on a single NRC detection model that is suitable for different travel demand levels.

These results illustrated in Table 5-7 illustrate the advantages and limitations of conservative and liberal models of detecting NRCs. The most liberal CE model (CE=1.2) and also CE=1.4 models detected every e^*; hence considered to be the best regarding the detection of e^*s. However, liberal NRC detection models also lead to higher Localisation Index values, as they are more susceptible to deeming day-to-day traffic variations as belonging to an NRC. On the other hand, the most conservative model (CE=2.0) is much more effective in terms of relating the detected NRCs with incidents; hence, it had the lowest Localisation Index. However, because such conservative NRC detection models are reluctant to detect an NRC, they also miss detecting e^*s, which are considered to be important to a traffic operator.

Table 5-7 Overall results of the comparison between different models on bank holidays (B), normal days (N) and tube strike days (T)

CE Model	**# NRCs**			**Run-time**			**e^***						**Localisation Index**		
	B	**N**	**T**	**B (sec)**	**N (sec)**	**T (min)**	**B**		**N**		**T**		**B**	**N**	**T**
							FAR	**FNR**	**FAR**	**FNR**	**FAR**	**FNR**			
CE = 2.0	125	358	521	2	12	1	0.13	0.58	0.08	0.63	0.05	0.51	**1.25**	**1.65**	**2.78**
CE = 1.8	165	501	637	4	24	2	0.15	0.45	0.10	0.48	0.06	0.37	1.39	1.89	3.70
CE = 1.6	248	738	805	8	60	6	0.21	0.25	0.16	0.26	0.10	0.19	1.47	2.08	6.56
CE = 1.4	386	1142	1045	21	198	16	0.34	**0.00**	0.32	**0.00**	0.21	**0.00**	2.09	2.84	11.59
CE = 1.2	737	1695	1350	89	1231	58	0.64	0.00	0.62	0.00	0.48	0.00	2.53	8.36	21.91

5.3.2. Detection of NRCs by Space–Time Scan Statistics

This section examines NRC detection by the expectation based space-time scan statistics (STSS) method developed in section 3.3. There are four parameters of the STSS method: congestion factor, maximum spatial (ρ) and temporal window size (τ) and the significance level. The developed STSS models use the minimum congestion factor used in CE (i.e. $c = 1.2$) and a significance level of 0.05, which is commonly used in statistical analysis. By fixing these two parameters of each STSS model, this thesis aims to understand the effect of maximum spatial and temporal window size values on the results.

After setting the congestion factor and significance level to detect a significant STR, ρ is varied between one and five, as STSS could be applied by solely considering the links themselves and by considering all their first-order adjacent links. Because no link has more than four first-order adjacent links, the maximum spatial window size is set to five. The maximum temporal window size is varied between one and six, as an NRC event is assumed to develop in 30 minutes at most, as the temporal difference between two consecutive LJTs is five minutes. This decision is empirically supported, as a previous study reports that 30 minutes is sufficiently long for a congestion event to develop (Jain et al., 2012). The total number of Space-Time Regions (STRs) that must be evaluated by each STSS model is shown in Table 5-8.

Table 5-8 Number of STRs in London's ANPR network on ρ and τ values

	ρ	**1**	**2**	**3**	**4**	**5**
τ	**# regions**	424	906	1141	1225	1241
1	145	61,480	131,370	165,445	177,625	179,945
2	289	122,536	261,834	329,749	354,025	358,649
3	432	183,168	391,392	492,912	529,200	536,112
4	574	243,376	520,044	654,934	703,150	712,334
5	715	303,160	647,790	815,815	875,875	887,315
6	855	362,520	774,630	975,555	1,047,375	1,061,055

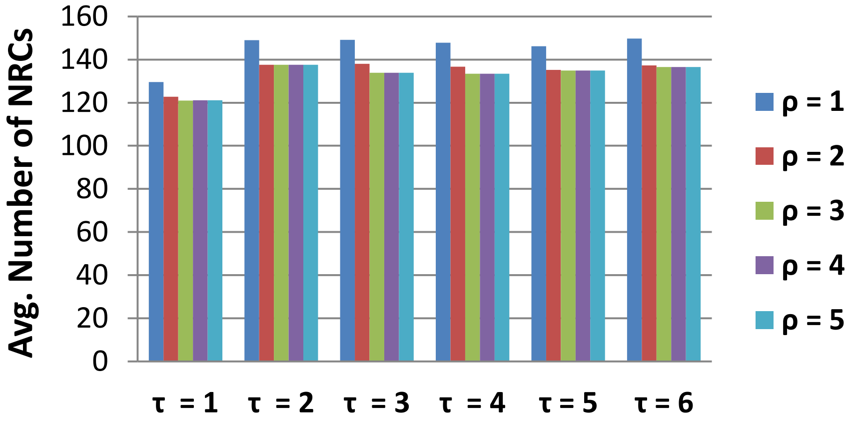
For example, when the maximum temporal window size is set to three, there are 432 overlapping temporal windows; 145 of them consist of a single LJT, because the analysis period is between 07:00 and 19:00 and each LJT is estimated every five minutes. Consequently, 144 and 143 temporal windows consist of two and three consecutive LJTs respectively. When the maximum spatial window size is set to two, there are 906 spatial regions. There are 424 links in the analysed network, and 482 spatial regions are derived that consist of two adjacent links. Therefore, with these choices, the total number of STRs that are to be evaluated is 432×906=391,392. Note that the difference between the total number of spatial regions decreases as the maximum spatial window size increases, as few links have that many adjacent links. Last, but not least, the most conservative STSS model occurs when both the spatial and temporal window size is one, because no temporal or spatial information is used to determine significant STR.

It should be highlighted once more that maximum likelihood ratio values are obtained for all different maximum spatial and temporal window size combinations. A total of 99 simulations are conducted prior to the analysis and maximum likelihood ratio values are recorded for each of the maximum spatial and temporal window size configurations. Increasing the maximum spatial and temporal window sizes also increases the maximum likelihood ratio values, as there are more STRs to evaluate. However, it is observed that the top four high scores (in order to detect significant STRs at the significance level of 0.05) converge to their maximum when ρ is three and τ is five. Therefore, an increase in the $\rho \geq 3$ or $\tau \geq 5$ or both could only result in an STSS model detecting more significant STRs; hence, making it more liberal.

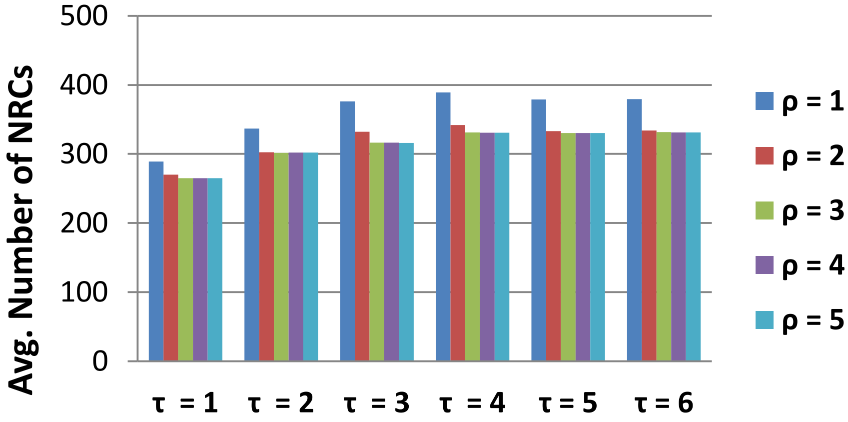
This section follows the same analyses as the previous section. Specifically, the investigation of different STSS models is based on several aspects. Firstly, the numbers of detected NRCs are compared. Secondly, the most severe NRC is investigated. Finally, performances of these different models are compared using the high-confidence episodes and Localisation Index criterion. In order to improve the legibility of the results, this section aggregates the days that are analysed for each travel demand level.

Number of detected NRCs

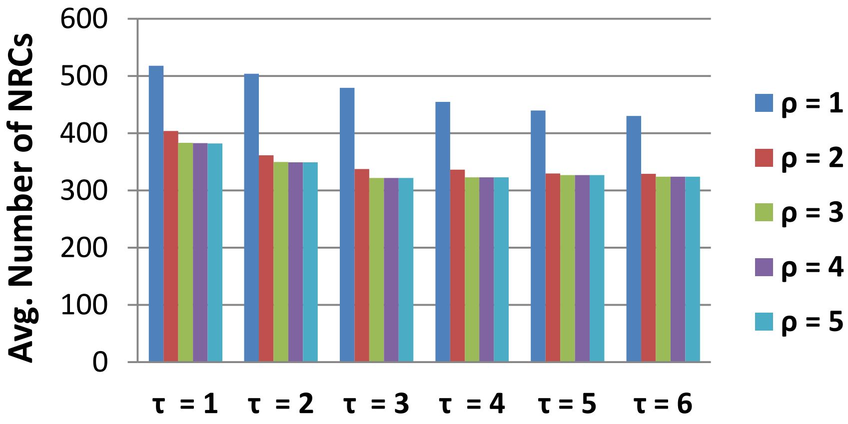
The number of NRCs detected by varying the maximum spatial (ρ) and temporal window sizes (τ) on bank holidays, normal days and tube strike days are illustrated in Figure 5-9(a-c) respectively.



(a)



(b)



(c)

Figure 5-9 Average number of NRCs by different STSS models on bank holidays (a), normal days (b) and tube strike days (c)

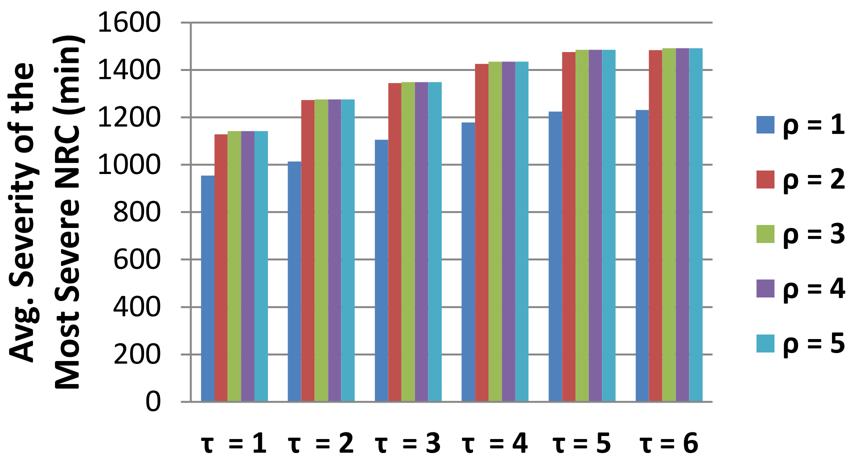
The following observations can be made regarding the number of detected NRCs by different STSS models:

- There is a noticeable decrement in the number of detected NRCs when ρ is increased from one to two, regardless of the value of τ. This observation is common to all the travel demand levels. By increasing ρ, an STSS model becomes more liberal when detecting NRCs, as it considers adjacent links together, rather than relying on individual links to detect significant STRs. This leads to merging of NRCs that could not previously have been merged. Note that this reduction could also be possible due to some NRCs becoming insignificant, because maximum likelihood ratio values increase as ρ is increased from one to two. However, this option is found be less likely during the empirical investigation.

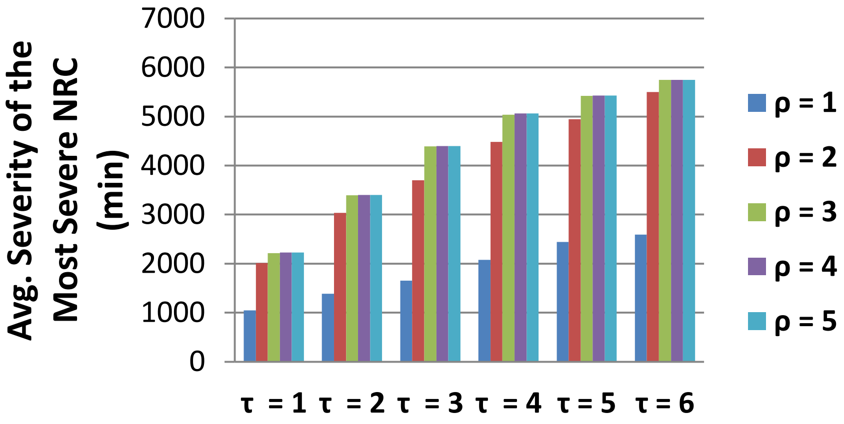
- Although increasing τ leads to more liberal STSS models, this is reflected differently by different types of day. For example, on bank holidays and normal days the number of NRCs escalates by increasing τ from one to two, whereas on tube strike days the number of NRCs decreases. Although in both cases STSS becomes more liberal when detecting NRCs, this is reflected by the detection of additional NRCs on bank holidays and normal days and the merger of previously distinct NRCs on tube strike days.
- The number of detected NRCs converge when $\rho \geq 3$ and $\tau \geq 5$, which is consistent with convergence values of maximum likelihood ratio values.

Characteristics of the Most Severe NRC

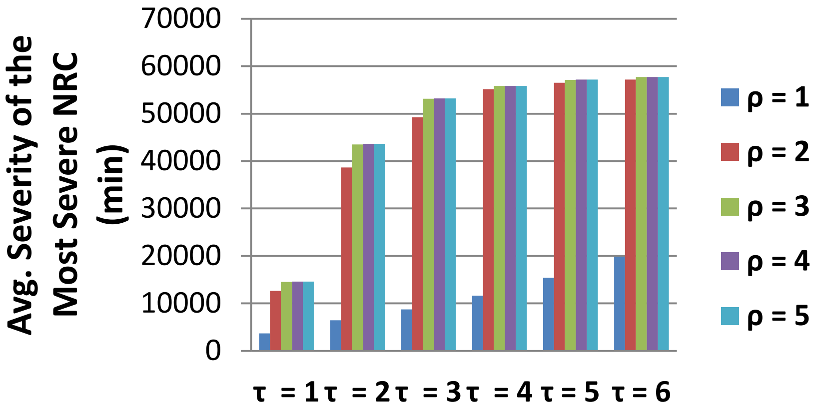
Two characteristics of the most severe NRC are analysed for different STSS models, namely its severity and duration. The severity of the most severe NRC are illustrated for bank holidays, normal days and tube strike days in Figure 5-10(a-c) respectively.



(a)



(b)



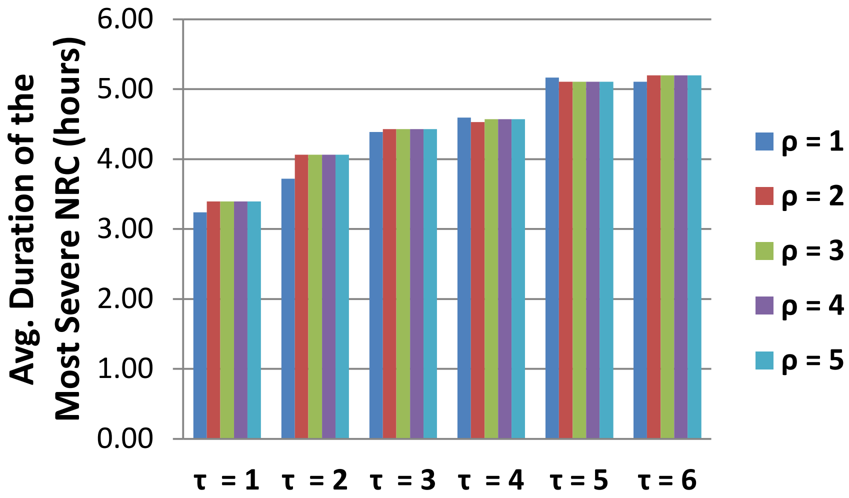
(c)

Figure 5-10 Severity of the most severe NRC detected by different STSS models on bank holidays (a), normal days (b) and tube strike days (c)

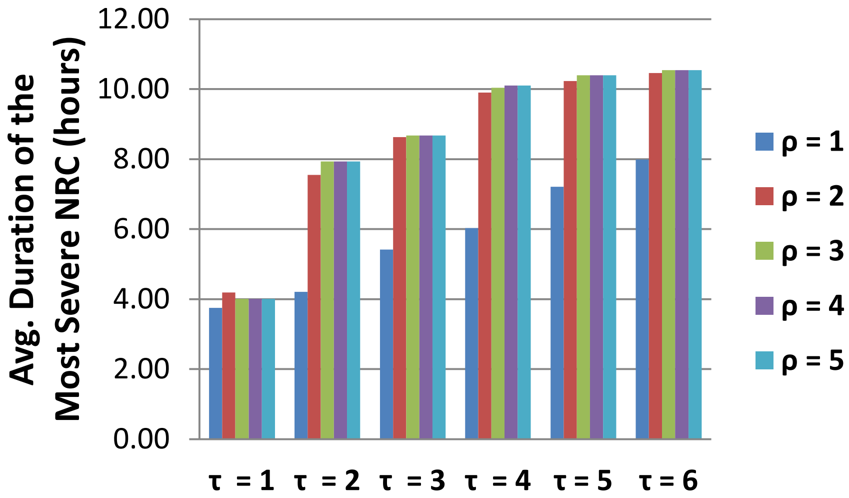
The analysis of the severity of the most severe NRC on different types of day validates that increasing τ or ρ leads to a more liberal STSS model, because the most severe NRC becomes even more severe when these parameters are increased. This situation is especially clear at low values of τ and ρ on tube strike days because consideration of additional information leads to a substantial change, since adjacent links or times would be more likely be congested. Furthermore, the results indicate that the severity of the most severe NRC converges to its maximum for $\tau \geq 4$ and $\rho \geq 3$ for both bank holidays and tube strike days, which is in-line with the convergence observed in maximum likelihood ratio values.

Another interesting result can be observed when ρ is increased from one to two, because the most severe NRC becomes much more severe during tube strikes, whereas this increase does not occur on bank holidays. This is related to the nature of the days, as more links would exhibit excessive LJTs simultaneously on tube strike days, whereas on bank holidays the spatial distribution of NRCs would be more localised.

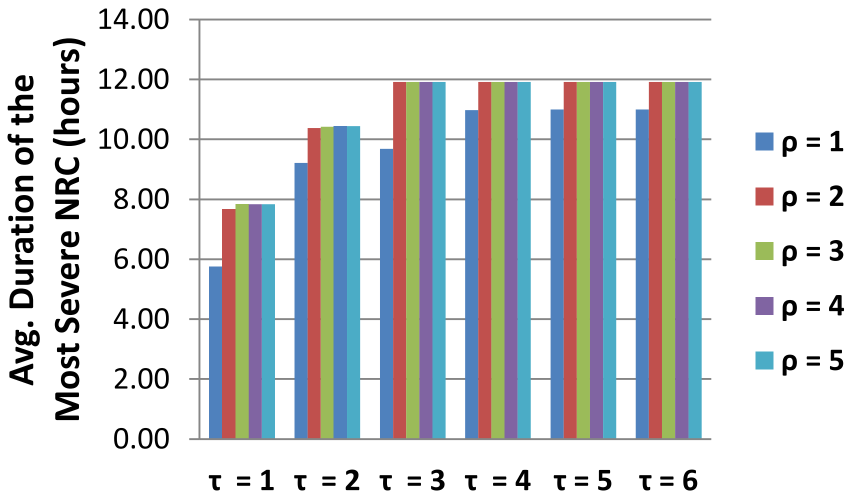
The average duration of the most severe NRC on bank holidays, normal days and during tube strikes is illustrated in Figure 5-11(a-c) respectively.



(a)



(b)



(c)

Figure 5-11 Average duration of the most severe NRC for bank holidays (a), normal days (b) and tube strike days (c)

The duration of the most severe NRC is the continuous time with which the lifetime of the most severe NRC corresponds to. These results validate the previous findings on the natural values regarding spatial and temporal window sizes. Increasing $\rho \geq 3$ does not change the average duration of the most severe NRC. In addition, the duration of the most severe NRC converges to its maximum value for $\tau \geq 5$.

On tube strike days the convergence of the duration of the most severe NRC is more noticeable. As can be seen, the duration of the most severe NRC reaches almost 12 hours, which is the total analysis period, for $\tau \geq 3$ and $\rho \geq 2$, which are smaller than normal days or bank holidays. This outcome validates intuitive expectations, because less information is sufficient to detect NRCs during tube strikes, as most of the road network is congested throughout the day.

The aforementioned characteristics provide an overview of the results when the τ and ρ are changed. It is observed that STSS models become more liberal as ρ and τ are increased. It is also important to understand how these changes affect the main evaluation criteria: high-confidence episode detection and Localisation Index.

Evaluation of the Performance of STSS Models on NRC Detection

This part evaluates the performance of the aforementioned STSS models with regard to NRC detection. The previous analyses on the characteristics of the STSS models demonstrated that they converge for $\rho \geq 3$. Therefore, the analyses conducted in this section do not consider STSS models that have a maximum spatial window size greater than three. The maximum temporal window size (τ) is varied between one and six.

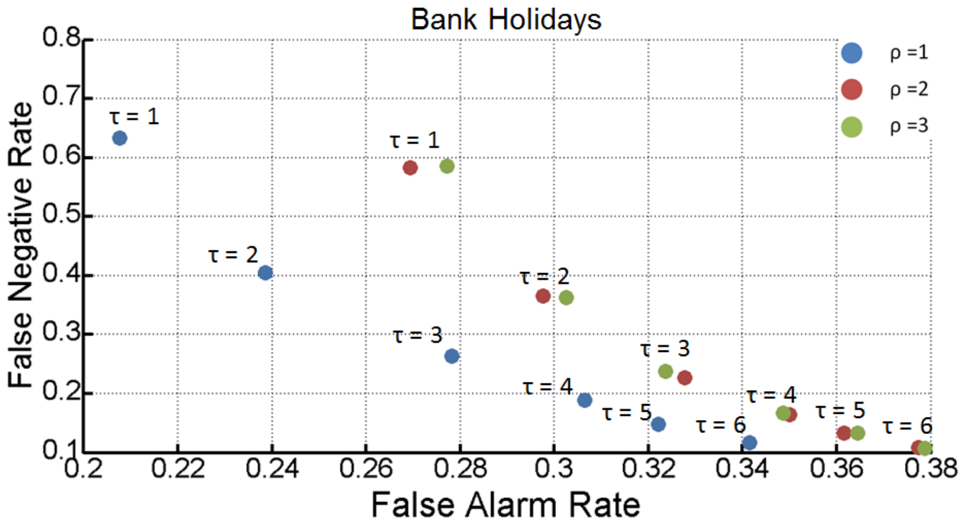
High-Confidence Episodes (e^*)

This part illustrates the results obtained from analysis of the performance of STSS models with regard to detecting e^*s[1]. The characteristics of the e^*s detected in this section are similar to the characteristics of e^*s discussed in the previous section. Specifically, the duration and severity of bank holidays and normal days are very close, whereas tube strike days feature a substantially higher duration and severity. In

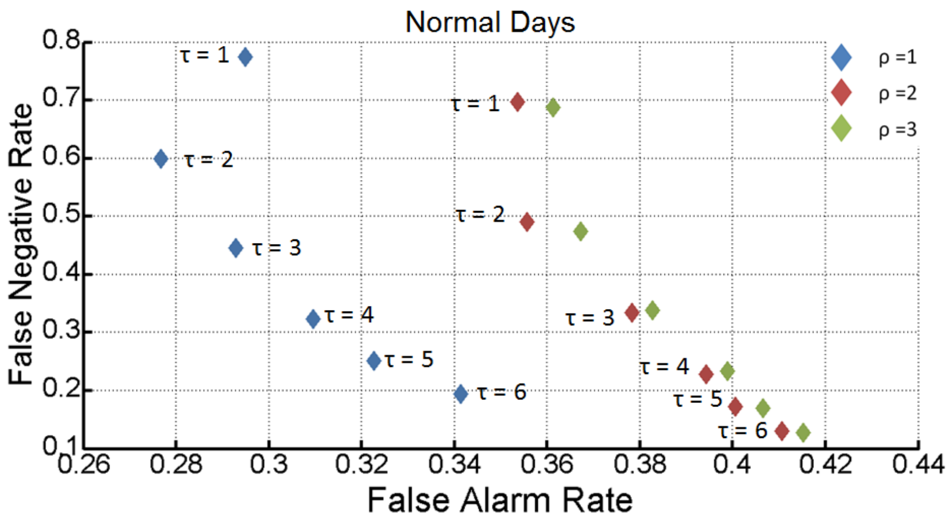
[1] The e^*s in this part are different than those detected in the previous section, since the expected LJTs are different in this section. Expected values are calculated based on the parameters determined to model each LJT by a lognormal distribution in section 5.1.

addition, as travel demand increases from bank holidays to normal days to tube strike days, so does the occurrence rate of e^*s. The characteristics of the e^*s highlighted in this part are illustrated in Appendix C in Figure C- 3.

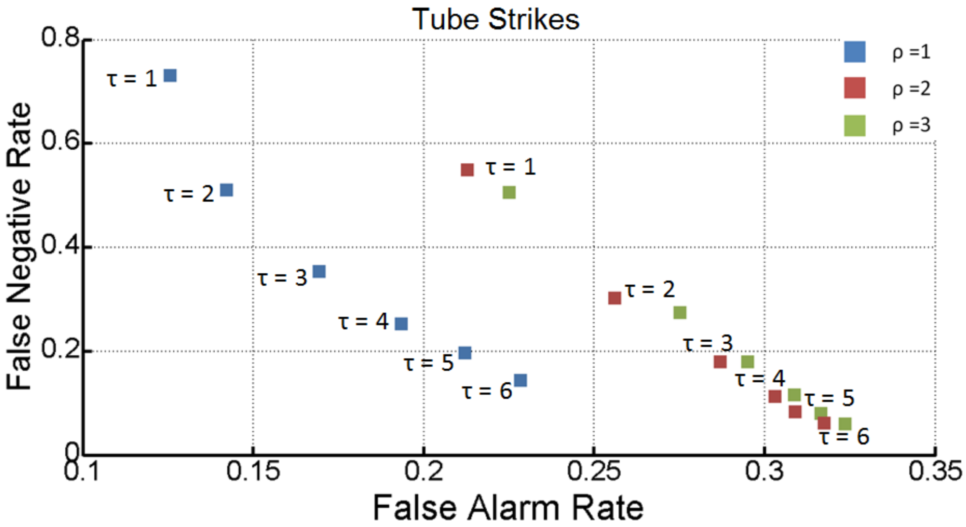
The False Negative Rate (FNR) and False Alarm Rate (FAR) results are obtained by comparing the detected NRCs with different STSS models. The analysis is conducted for bank holidays, normal days and during tube strike days and are illustrated in Figure 5-12(a-c) respectively.



(a)



(b)



(c)

Figure 5-12 Performance of STSS models regarding the detection of high-confidence episodes on bank holidays (a), normal days (b) tube strike days (c)

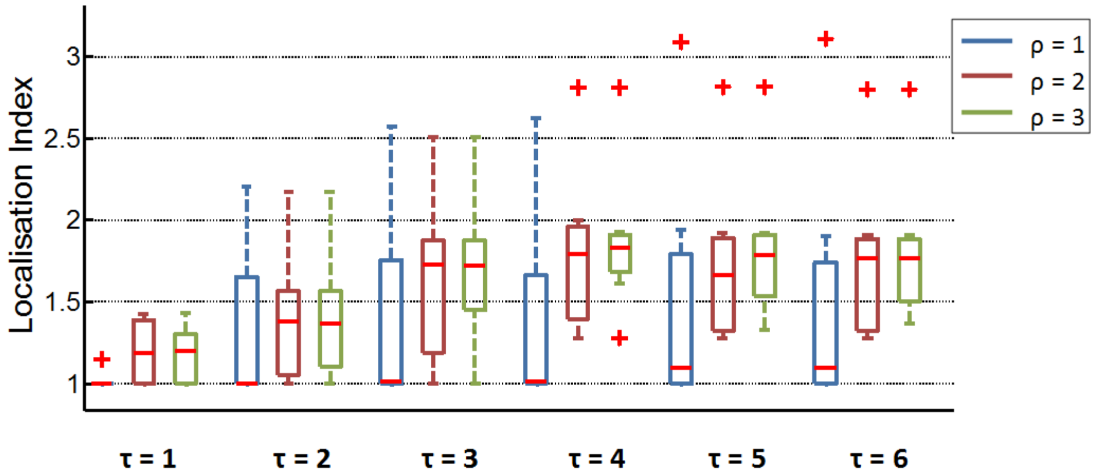
The results suggest that the temporal window size has an important effect on the detection of e^*s. When the temporal window size is increased, STSS models detect e^*s much more effectively. It is observed that an increase of approximately 60% in the detection of e^*s could be achieved by increasing τ from one to six. The corresponding increase in FAR suggests that more LJTs are also considered to belong to an NRC that do not belong to an e^*.

Another noticeable result is that performance increases (i.e. FNR decreases) are higher at lower values of τ. For example, increasing τ from one to two leads to a decrement in FNR of approximately 20%, whereas this percentage decreases to approximately 5% when τ is increased from five to six. Similarly, increasing ρ from one to two results in a greater decrease in FNR compared to increasing ρ from two to three. However, increasing ρ does not always necessitate better detection of e^*s. For example, on bank holidays it is observed that increasing spatial window size from two to three when τ is three actually decreased the performance. This is because increasing ρ leads to higher maximum likelihood ratio values. This might lead to the disintegration of some NRCs or to missing them entirely. On low travel demand days like bank holidays, this situation is possible. However, most of the time an increase in ρ leads to a more liberal way of detecting NRCs, which leads to a lower rate of missed e^*s (i.e. lower FNR).

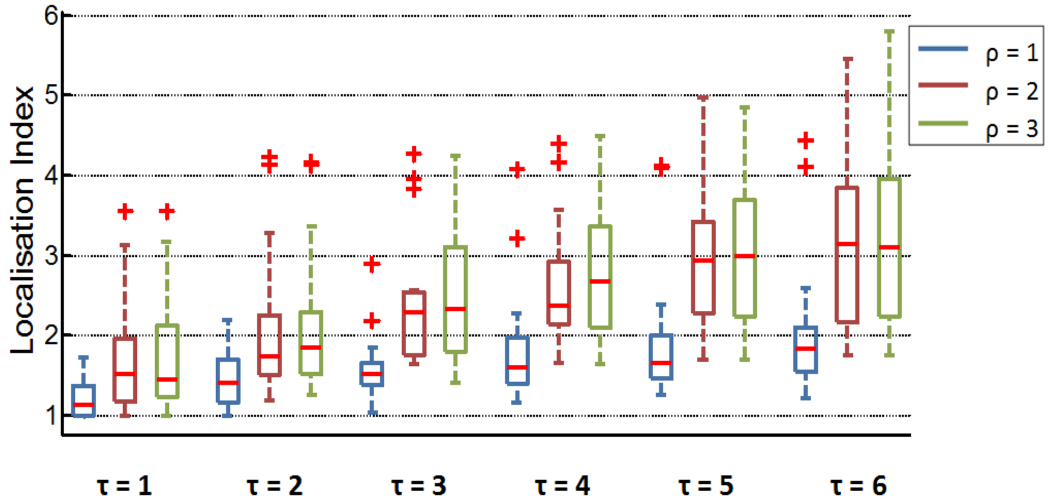
Once these results are compared with the results reported in Figure 5-6, it becomes clear that none of the STSS models reach the perfect situation of detecting all e^*s (i.e. an FNR of zero), which was possible for CE models using a congestion factor of less than or equal to 1.4. This finding illustrates that even considering six consecutive LJTs together coming from three adjacent links may lead to some e^*s being missed, which can be attributed to the significance testing procedure within STSS. Link-based analysis conducted in subsection 5.4.1 provides examples to support why STSS models miss to detect some e^*s.

Localisation Index

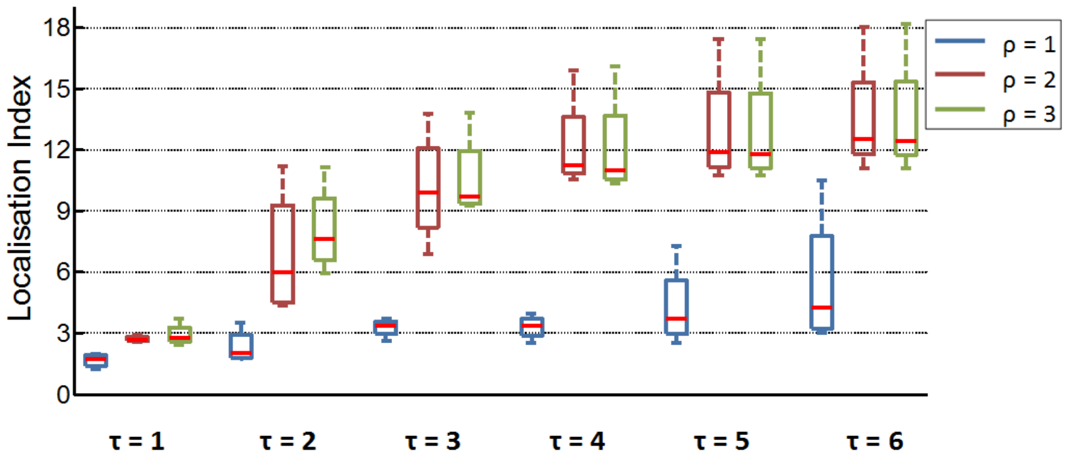
This part illustrates the results obtained from the analysis of the Localisation Index in different STSS models. As was the case in the previous part, the maximum spatial window size is varied between one and three, and the maximum temporal window size is varied between one and six. The Localisation Index values are summarised for bank holidays, normal days and tube strike days in Figure 5-13(a-c) respectively.



(a)



(b)



(c)

Figure 5-13 Localisation Index for bank holidays (a), normal days (b) and tube strike days (c) for different STSS models

The main outcome that is common to all the three types of day is that increasing the maximum spatial and temporal window size leads to the detection of NRCs that are harder to relate with incidents. This validates the previous findings, as an STSS model becomes more liberal as τ and/or ρ increases, because more STRs are detected as being significant. However, this also increases the likelihood that day-to-day traffic variations will be identified as belonging to an NRC. This explains why the Localisation Index values increase when the maximum spatial and temporal window size increases.

The Localisation Index increases noticeably when ρ rises from one to two, which is noticeable on all three types of day. This finding suggests that it is better to liberalise an STSS model by increasing τ than ρ. Otherwise, if ρ is increased, it becomes harder to localise detected NRCs, as the increase in the average number of connected components is greater, suggesting that an STSS model is more likely to include day-to-day variation in traffic as belonging to an NRC.

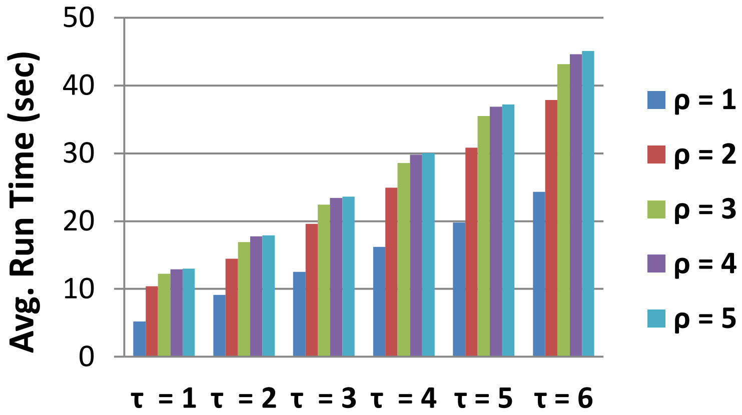
Two observations add further support to the findings observed when CE models are analysed:

- Liberalising an STSS model does not always lead an increase in the Localisation Index. For example, on a bank holiday increasing ρ from one to two actually resulted in a lower Localisation Index for $\tau \geq 5$, which is regarded as an 'outlier' and has a much higher Localisation Index value than other bank holidays. This phenomenon suggests that by liberalising an NRC detection method it is also possible to decrease the Localisation Index, as the additional excessive LJTs might decrease the number of components that an NRC exhibits. However, as was the case in CE, this situation is uncommon, and usually the Localisation Index increases as the NRC detection model becomes more liberal.

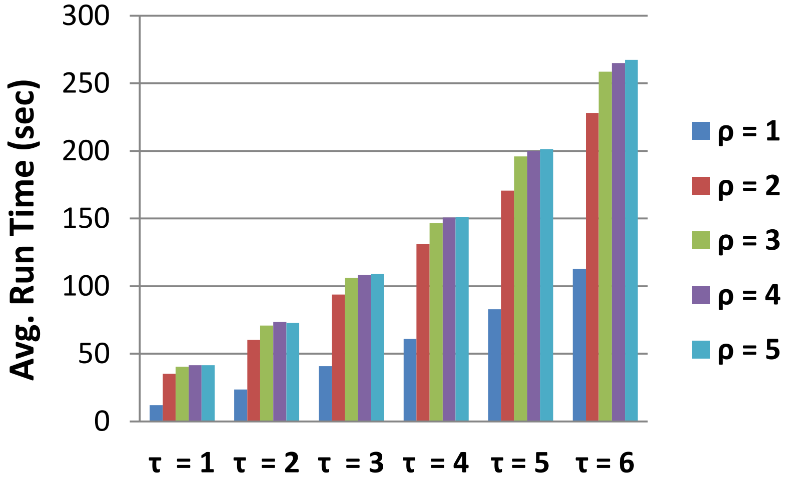
- There is a relationship between travel demand level and the Localisation Index. It can be seen that, as travel demand increases from bank holidays to normal days to tube strike days, Localisation Index values also increase. This is reasonable, because as travel demand increases, more and more links would become congested and these links could more easily be included in an NRC.

Run-Time

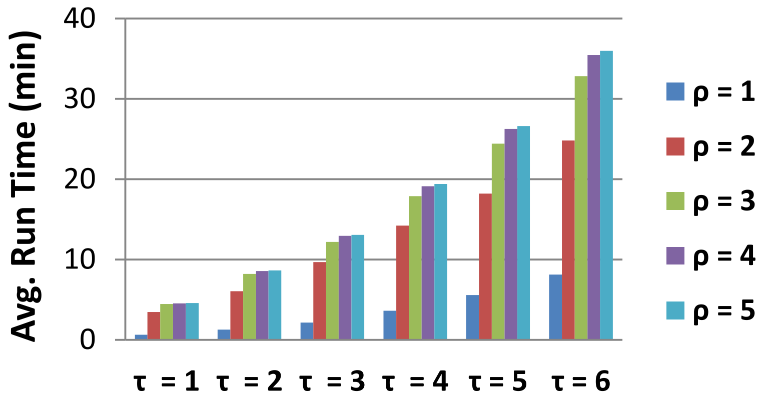
The last performance evaluation criterion is the run-time. The run-times of different STSS models are illustrated for bank holidays, normal days and tube strike days in Figure 5-14(a-c) respectively.



(a)



(b)



(c)

Figure 5-14 Average run-time for different STSS models on bank holidays (a), normal days (b) and tube strike days (c)

The results validate the previous findings. Specifically, increasing τ or ρ or both increases the run-time, as there would be more STRs to evaluate. It is worth highlighting the substantial difference between the average run-time for bank holidays and tube strike days. Analysis on tube strike days last approximately 26 times longer when τ is two, and this number increases to approximately 45 times when τ is increased to six compared to bank holidays. The main reason for this is that there are much more significant STRs on tube strike days due to the increased levels of traffic, and this becomes even worse for higher values of τ, as increasing the maximum temporal window size leads to a more liberal way of detecting NRCs. Last, but not least, on normal days the run-times vary between 12 and 267 seconds. This suggests that the proposed STSS models could readily be adapted in a traffic operation centre.

To sum up, increasing the maximum spatial and temporal window size leads to more liberal ways of detecting NRCs, even though the maximum likelihood ratio scores also increase. An increase in the maximum likelihood ratio scores could lead to the following two outcomes:

- An NRC becomes less severe: as maximum likelihood ratio values rise due to increasing the spatial and temporal window size, previously significant STRs may become insignificant. If the newly added STRs do not correspond with the increase that is observed in maximum likelihood ratio values, then some significant STRs may not be significant any more. This may reduce the severity of an NRC and could even result in an NRC splitting into two distinct NRCs. In the most extreme case, an NRC may not be significant anymore. This outcome is mostly observed on bank holidays when travel demand is very low.

- An NRC becomes more severe: LJTs that are at the periphery of an NRC (towards the beginning or end of an NRC) that were not significant may become significant once they are considered together with spatially or temporally adjacent LJTs. This is only made possible by using larger STRs. This outcome is found to be much more common in the analysis.

All the results obtained by detecting NRCs by STSS are summarised for the most severe NRC and overall in Table 5-9 and Table 5-10 respectively. In these results, bank holidays, normal dates and tube strike days are denoted with B, N and T respectively.

Table 5-9 Characteristics of the most severe NRC detected by each STSS model

STSS Model		**Duration (hours)**			**Severity (minutes)**		
τ	**ρ**	**B**	**N**	**T**	**B**	**N**	**T**
1	**1**	**3.24**	**3.75**	**5.75**	**954**	**1,049**	**3,698**
	2	**3.40**	**4.19**	**7.67**	**1,128**	**2,010**	**12,630**
	3	**3.40**	**4.00**	**7.83**	**1,141**	**2,214**	**14,532**
2	**1**	**3.72**	**4.20**	**9.21**	**1,013**	**1,386**	**6,470**
	2	**4.06**	**7.55**	**10.38**	**1,273**	**3,033**	**38,632**
	3	**4.06**	**7.93**	**10.42**	**1,276**	**3,391**	**43,498**
3	**1**	**4.39**	**5.41**	**9.69**	**1,105**	**1,650**	**8,759**
	2	**4.43**	**8.63**	**11.92**	**1,344**	**3,701**	**49,206**
	3	**4.43**	**8.67**	**11.92**	**1,349**	**4,393**	**53,130**
4	**1**	**4.59**	**6.03**	**10.98**	**1,178**	**2,074**	**11,603**
	2	**4.53**	**9.90**	**11.92**	**1,425**	**4,481**	**55,189**
	3	**4.57**	**10.04**	**11.92**	**1,435**	**5,038**	**55,804**
5	**1**	**5.17**	**7.21**	**11.00**	**1,224**	**2,444**	**15,407**
	2	**5.10**	**10.23**	**11.92**	**1,475**	**4,945**	**56,490**
	3	**5.10**	**10.40**	**11.92**	**1,484**	**5,423**	**57,138**
6	**1**	**5.10**	**7.99**	**11.00**	**1,231**	**2,593**	**19,848**
	2	**5.20**	**10.46**	**11.92**	**1,483**	**5,502**	**57,209**
	3	**5.20**	**10.55**	**11.92**	**1,491**	**5,746**	**57,715**

Table 5-10 Overall results of the comparison of different STSS models on bank holidays (B), normal days (N) and tube strike days (T)

STSS Model		**# NRCs**			**Run-time**			***e****						**Localisation Index**		
								B		**N**		**T**				
τ	**ρ**	**B**	**N**	**T**	**B (sec)**	**N (sec)**	**T (min)**	**FAR**	**FNR**	**FAR**	**FNR**	**FAR**	**FNR**	**B**	**N**	**T**
1	**1**	**130**	**289**	**518**	**5**	**12**	**1**	**0.21**	**0.63**	**0.29**	**0.77**	**0.13**	**0.73**	**1.02**	**1.22**	**1.67**
	2	**123**	**270**	**404**	**10**	**35**	**3**	**0.27**	**0.58**	**0.35**	**0.70**	**0.21**	**0.55**	**1.20**	**1.70**	**2.74**
	3	**121**	**265**	**383**	**12**	**40**	**4**	**0.28**	**0.59**	**0.36**	**0.69**	**0.23**	**0.51**	**1.18**	**1.74**	**2.94**
2	**1**	**149**	**337**	**504**	**9**	**24**	**1**	**0.24**	**0.40**	**0.28**	**0.60**	**0.14**	**0.51**	**1.31**	**1.43**	**2.36**
	2	**138**	**302**	**361**	**14**	**60**	**6**	**0.30**	**0.37**	**0.36**	**0.49**	**0.26**	**0.30**	**1.40**	**2.07**	**6.89**
	3	**138**	**302**	**350**	**17**	**71**	**8**	**0.30**	**0.36**	**0.37**	**0.47**	**0.28**	**0.27**	**1.41**	**2.11**	**8.09**
3	**1**	**149**	**376**	**479**	**13**	**41**	**2**	**0.28**	**0.26**	**0.29**	**0.45**	**0.17**	**0.35**	**1.39**	**1.59**	**3.28**
	2	**138**	**332**	**337**	**20**	**94**	**10**	**0.33**	**0.23**	**0.38**	**0.33**	**0.29**	**0.18**	**1.64**	**2.40**	**10.12**
	3	**134**	**316**	**322**	**22**	**106**	**12**	**0.32**	**0.24**	**0.38**	**0.34**	**0.30**	**0.18**	**1.70**	**2.53**	**10.64**
4	**1**	**148**	**389**	**455**	**16**	**61**	**4**	**0.31**	**0.19**	**0.31**	**0.32**	**0.19**	**0.25**	**1.37**	**1.83**	**3.30**
	2	**137**	**342**	**336**	**25**	**131**	**14**	**0.35**	**0.16**	**0.39**	**0.23**	**0.30**	**0.11**	**1.80**	**2.61**	**12.22**
	3	**133**	**331**	**323**	**29**	**146**	**18**	**0.35**	**0.17**	**0.40**	**0.23**	**0.31**	**0.12**	**1.87**	**2.74**	**12.11**
5	**1**	**146**	**379**	**440**	**20**	**83**	**6**	**0.32**	**0.15**	**0.32**	**0.25**	**0.21**	**0.20**	**1.48**	**1.91**	**4.30**
	2	**135**	**333**	**330**	**31**	**171**	**18**	**0.36**	**0.13**	**0.40**	**0.17**	**0.31**	**0.08**	**1.73**	**2.93**	**12.98**
	3	**135**	**331**	**327**	**36**	**196**	**24**	**0.36**	**0.13**	**0.41**	**0.17**	**0.32**	**0.08**	**1.83**	**3.02**	**12.94**
6	**1**	**150**	**379**	**430**	**24**	**113**	**8**	**0.34**	**0.12**	**0.34**	**0.19**	**0.23**	**0.14**	**1.47**	**2.02**	**5.51**
	2	**137**	**334**	**329**	**38**	**228**	**25**	**0.38**	**0.11**	**0.41**	**0.13**	**0.32**	**0.06**	**1.76**	**3.10**	**13.53**
	3	**137**	**332**	**324**	**43**	**258**	**33**	**0.38**	**0.11**	**0.42**	**0.13**	**0.32**	**0.06**	**1.81**	**3.19**	**13.53**

5.3.3. Overall Comparison of NRC Detection Methods

The previous two sections investigated the performance of the two proposed NRC detection methods (i.e. CE and STSS). The effect of the congestion factor on CE models and the maximum spatial and temporal window size parameters on STSS models are investigated. The formal investigation involved five CE and 18 STSS models. This section aims to compare all the investigated NRC detection models and identify the most suitable.

Before discussing the developed framework in order to compare different NRC detection models, it is necessary to develop a brief understanding of Multi-Attribute Decision Making (MADM, also referred to as multi-criteria decision making), which provides the necessary theoretical background to compare different models (also referred to as 'alternatives') based on two or more, often conflicting, attributes [1]. The aim of such a comparison is to determine the best-performing model. MADM consists of many methods, including, but not limited to, Weighted Sum Model (WSM), Weighted Product Model (WPM) and Analytic Hierarchy Process (AHP) (Triantaphyllou and Mann, 1989). The main issue, however, is that it is very difficult to choose the most suitable method from amongst these different MADM methods. This is because such a decision may yield rank reversals, which means that the introduction of a new model that is inferior to the existing models may alter the decision regarding the best-performing model (Leskinen and Kangas, 2005; Zanakis et al., 1998). Therefore, such a decision regarding the best MADM method is considered to be a paradox (Triantaphyllou and Mann, 1989).

It has been reported that WPM is not susceptible to rank-reversals due to its structure (Triantaphyllou and Mann, 1989), which relies on a pair-wise comparison of different NRC detection models. For a given attribute, the ratio of the values belonging to the two compared NRC detection models is calculated. The weight for each attribute is considered as an exponent to the corresponding ratio. Once this ratio is calculated for each attribute, these ratios are multiplied. In order to determine the best NRC detection model, Final Score is calculated as follows:

---

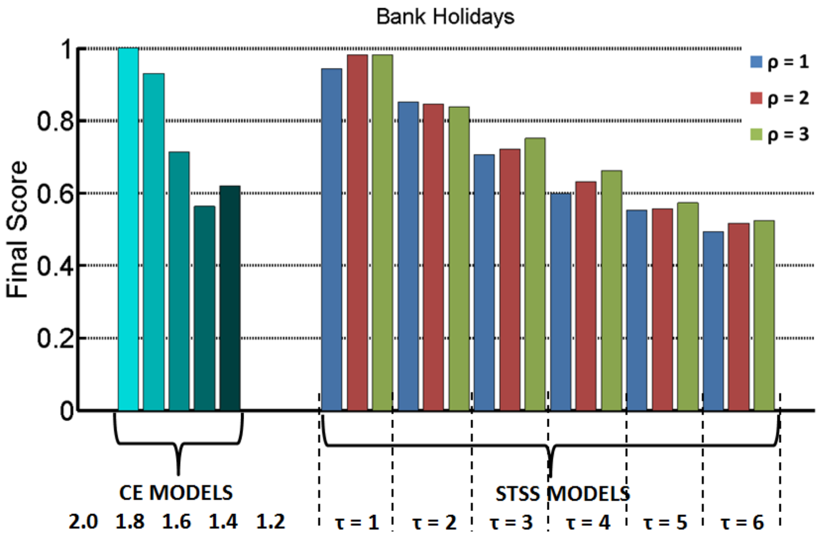
[1] In the current method of evaluating NRC detection methods there are two attributes: the False-Negative Rate obtained by comparing the detected NRCs with e^{*}s and the Localisation Index. Note that these two attributes are conflicting, where the aim is to minimise both.

$$\text{Final Score}(M_K, M_L) = \prod_{j \in \{e^*, \text{Localisation Index}\}} \left(M_{Kj}/M_{Lj}\right)^{w_j}, \text{where } \sum_j w_j = 1$$

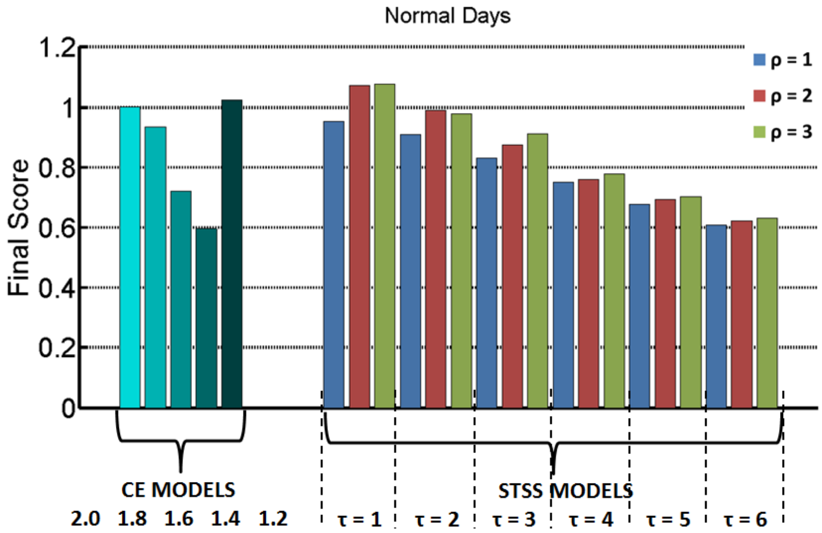
where M_K and M_L are the two NRC detection models and j denotes an attribute which can be either the False-Negative Rate in e^* or the Localisation Index. M_{Kj} and M_{Lj} denote the values of the j^{th} attribute of K^{th} and L^{th} NRC detection models respectively, and w_j is the weight of attribute j. Having determined all Final Score values, the best model is the one that has the ***smallest*** Final Score, because the smaller the values of both of the attributes, e^* and the Localisation Index, the better the NRC detection model. Once all NRC detection models are compared with one another, they can be ranked based on their Final Score values.

The WPM method is a simple, yet useful way of comparing different NRC detection models. An important advantage of WPM is that it provides a dimensionless analysis, as it relies on ratios. In this way, attributes with different units could be combined in one measure. On the other hand, as it relies on the calculation of a ratio, it assumes that no attribute value could be zero. This assumption is not satisfied within this thesis' context, because the FNR values of the two Clustering Episode models (i.e. CE ≤ 1.4) have an FNR of zero. Studies indicate that there is no published research evidence to handle this issue (Mela et al., 2012). Because WPM does ***not*** suffer from rank reversal, this thesis relies on WPM to compare different NRC detection models with the following adjustment: the zero FNR values for CE ≤ 1.4 models are replaced with the second smallest FNR value.

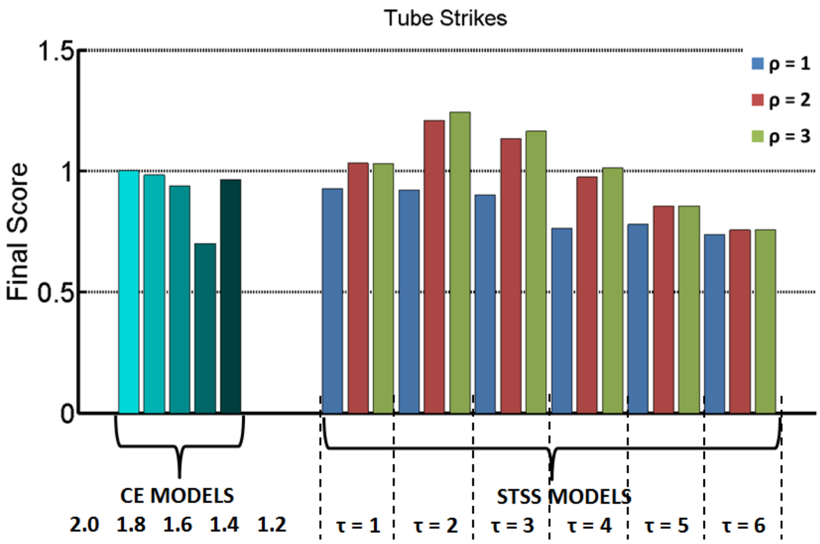
Both of the attributes are assumed to have equal importance (i.e. $w_{e^*} = w_{\text{Localisatation Index}} = 0.5$). The Final Scores are calculated for each of the NRC detection models on bank holidays, normal days and during tube strike days, which are illustrated in Figure 5-15(a-c) respectively.



(a)



(b)



(c)

Figure 5-15 Final Comparison of NRC detection models for bank holidays (a), normal days (b) and tube strike days (c)

The following conclusions can be drawn by analysing the final scores as illustrated in Figure 5-15:

- CE=1.4 models result in the best performance amongst all the CE models and most of the STSS models. This is probably due to the fact that e^* and CE=1.4 use the same congestion factor of 1.4 to detect episodes. It is important to highlight one more time that e^* is proposed through considering one of the uses of NRC detection models, which is to detect substantially high LJTs that last for a minimum duration. In traffic operation centres, 'substantially-high' LJTs are characterised by a multiplicative increase of expected LJTs using a congestion factor. Because CE and e^* use this idea to determine 'substantially-high' LJTs, it is likely that CE=1.4 performs better than most of the NRC detection models.
- It is possible to liberalise an STSS model by increasing either the maximum temporal window size (τ) or maximum spatial window size (ρ). Most of the time liberalising an STSS method improves overall performance. It appears that it is preferable to liberalise an STSS model by increasing τ rather than increasing ρ. This is because increasing ρ leads to considering more day-to-day variations to belong to an NRC; hence, increasing the Localisation Index considerably. As a result, by increasing ρ, overall performance is degraded most of the time.
- On normal days, it appears that increasing τ beyond the priori determined maximum temporal window size of six when ρ=1 could perform better than CE=1.4. This is because, by increasing τ an STSS model performs better with regard to detecting e^*s, and eventually a corresponding increase in Localisation Index values does not degrade the overall performance of STSS as much. It is, however, important to realise that increasing τ may not always be the best solution, and the reason for this is discussed in the next section by focusing on some exemplar links.

5.4. Investigating the Effectiveness of NRC Detection Methods

This section investigates the effectiveness of NRC detection methods based on some real life scenarios from two aspects. Firstly, detected NRCs are visualised on some individual links to demonstrate the relative advantages and limitations of the NRC detection methods, as well as e^*. Secondly, the impact of an incident which affected several links is determined by using different NRC detection models.

5.4.1. NRC Detection on Exemplar Links

The aim of this section is to illustrate the advantages and limitations of different NRC detection methods on some exemplar links. The estimated LJTs that are found to belong to an NRC are highlighted. Three NRC detection models are used for this purpose. The first is an STSS with a maximum temporal window size of four and a maximum spatial window size of one, as it performs better than most of the STSS models. The second is CE=1.4, as it could be considered to be the best NRC detection model. Finally, e^* is used as it is an evaluation criterion. The three NRC detection models used within this study and their parameters are illustrated in Table 5-11:

Table 5-11 Analysed methods and their parameters

	Space-Time Scan Statistics (STSS)	**Clustering Episodes (CE)**	**High-Confidence Episode (e^*)**
Congestion factor	1.2	1.4	1.4
Minimum duration	–	–	25 minutes
Spatial window size	1	–	–
Temporal window size	4	–	–
Significance level	0.05	–	–

The outcomes of these three NRC detection models are investigated according to some exemplar links in order to illustrate how the methods behave with regard to issues that arise in a real life context. Four important issues have been addressed, which are: missing data, short NRCs, unreliable LJTs, and the handling of day-to-day traffic variations on short links.

An exemplar link is found to illustrate how the three NRC detection methods react to these issues. The date of analysis is 23 June 2010 (Wednesday) between the times 07:00 and 19:00. Note that, in all of the exemplar links, the threshold is illustrated by multiplying the expected LJTs with the congestion factor of 1.4, as it is the congestion factor used within the two NRC detection models (i.e. CE and e^*) of the three analysed within this section.

Missing LJT Data

It is possible for Automatic Number Plate Recognition (ANPR) cameras to miss capturing a vehicle during a data collection interval. An important reason for this is congestion. The dense movement of vehicles during congested times prevents ANPR cameras from capturing the licence plates of the vehicles. In addition, various types of incidents occur in urban environments that result in road closures. These incidents vary from joyful carnivals to saddening vehicular accidents, where the duration of road closure depends on the nature of the incident.

An incident occurred on link 2111 that began at 18:15 and ended at 19:45. The incident was a ‘demonstration event’ that took place next to the British Parliament. The expected LJTs, threshold (expected LJTs multiplied by 1.4) and estimated LJTs are illustrated in the Figure 5-16. In addition, the sample size used to estimate an LJT (i.e. the number of vehicles that are captured at both ends of the link to estimate an LJT) is also illustrated. On the top of the illustration, the outcome of the three aforementioned methods are shown, where each LJT is plotted in green if it does not belong to an NRC or red if it belongs to an NRC. The first row shows the results obtained by applying STSS, the second row shows the results obtained by applying CE and the third row shows the results obtained by applying e^*, whose characteristics are described in Table 5-11.

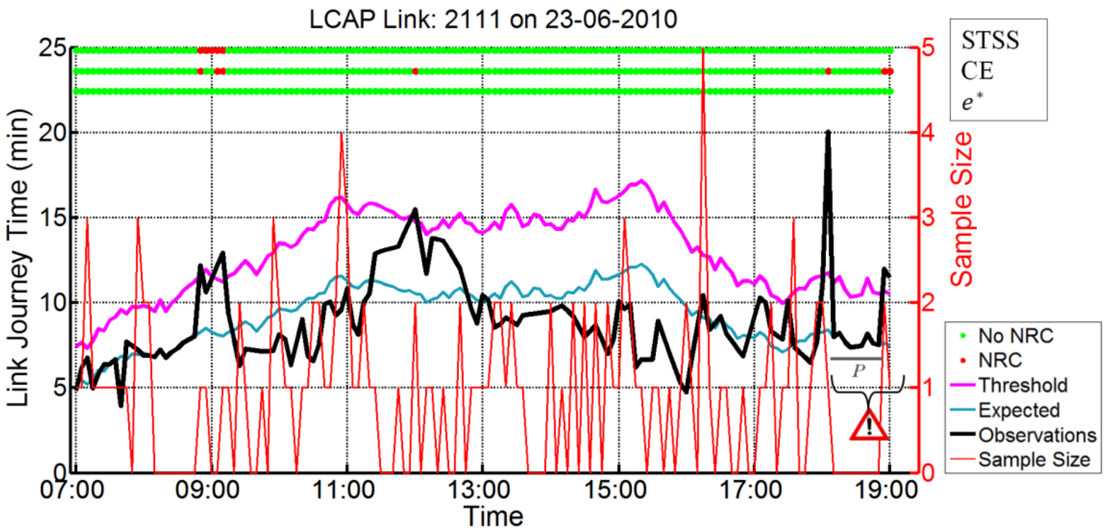


Figure 5-16 Missing LJTs due to an incident

The only NRC detection model that signalled an LJT during the incident was CE, which detected one episode at the beginning of the incident and one at the end. The former episode lasted for five minutes (i.e. a single LJT), and the latter lasted for ten minutes (i.e. two consecutive LJTs). These episodes were captured because the estimated LJTs are 40% higher than the expected values. However, neither STSS nor e^* was effective in detecting an episode that could be associated with this incident. STSS could not detect these episodes because the space-time regions that are used to scan the road network did not find any statistically significant space-time region that belongs to this link during the incident. These episodes could not be detected by e^*, because they did not last for longer than 25 minutes.

The advantage of STSS over CE can also be observed. The first and only episode detected by STSS lasted for 20 minutes (i.e. four consecutive LJTs). However, the same event was represented as two different episodes by CE, due to the fluctuations of the estimated LJTs around the threshold. In this context, STSS captured the first episode to its full extent more effectively than CE.

Last, but not least, during the period illustrated with 'P,' the LJTs were patched to their expected values. This period occurred during the incident and could have been caused by two possible reasons. Firstly, the traffic on the link might have been diverted as the incident was a demonstration event. Secondly, ANPR may not have captured any vehicle's license plate due to the dense movement of vehicles. These two possible reasons caused 45 minutes of missing data to start immediately after the beginning of the incident. Once the vehicles were captured, the estimated LJTs were higher than the threshold, indicating the presence of congestion caused by the incident.

Short NRCs

Tower Bridge was lifted three times between 10:15 and 23:00 on 23 June 2010. The estimated LJTs on the link that passes through Tower Bridge (link 2515) is illustrated in Figure 5-17. This event caused three visually apparent increases in LJTs, all of which were successfully detected only by the CE model. Therefore, the episodes that CE detected can be associated with the increases in LJTs due to the lifting of the bridge. STSS is effective at detecting the second visually apparent increase in LJTs. The duration of this episode is five minutes longer than CE and e^*, because STSS considers adjacent times to detect statistically significant STRs and also uses a lower congestion factor than CE. On the other hand, e^* did not detect an episode between 09:00 and 11:00, because none of the episodes that CE detected lasted for at least 25 minutes. The third visually apparent increase in LJTs were only detected by the CE model, because the episode was not found to be statistically significant by STSS, nor did it last for at least 25 minutes so it was not detected by e^*.

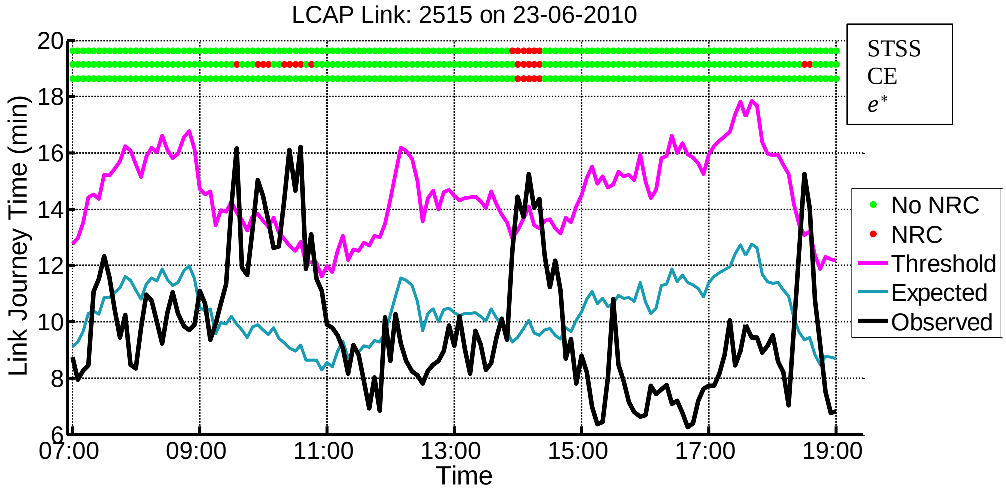


Figure 5-17 Three NRCs due to the lifting of Tower Bridge

Unreliable LJT Estimates

Some links LJTs are estimated based on a low sample size due to several possible reasons including the technological limitations of the ANPR cameras or their temporary failure. Such LJT estimations are prone to sampling error. A good example is link 2394, which has a length of 1.3 kilometres and is located near A2 New Cross Road in South London. The results regarding the analysis of this link are illustrated in Figure 5-18. In addition, the sample size used to estimate each LJT is also illustrated.

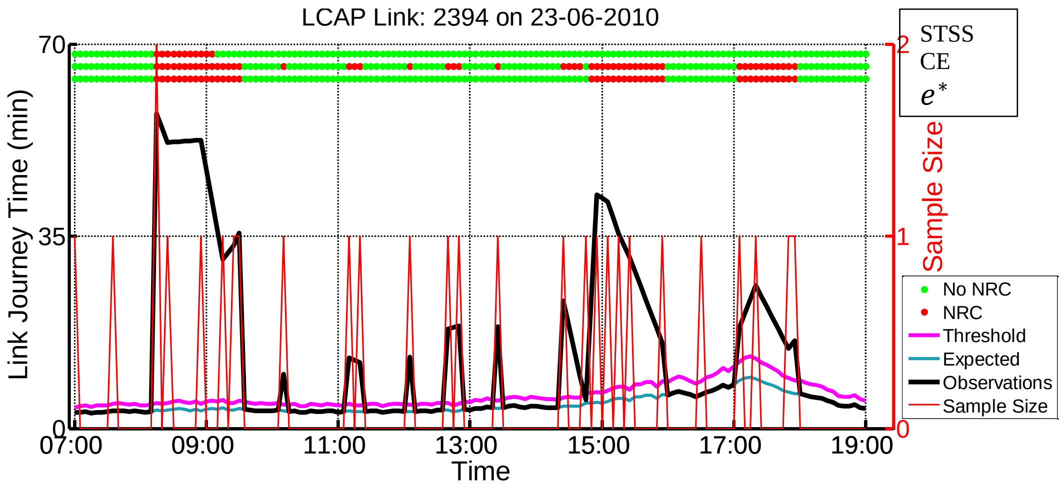


Figure 5-18 Unreliable LJT estimates wrongly reported to belong to an NRC

Most of the estimated LJTs are patched and the remaining LJTs are estimated based on a single vehicle's journey time. This increases the likelihood of the estimated LJT being erroneous, because the captured vehicle might have detoured. This makes it possible to observe such erratic increases in estimated LJTs. Note that no incident was reported on this link on 23 June 2010, which increases the likelihood that the substantially high

LJTs are due to sampling error. Furthermore, empirical investigation shows that the sample size used to estimate an LJT is usually very low on this link. Therefore, most estimated LJTs exhibit high variance. The high variance is captured effectively in STSS, as it reported only one episode on this link. On the other hand, CE and e^* are indifferent to such variations, and they detected nine and three episodes respectively. All of the detected episodes on this link are most likely to be false alarms (i.e. in reality they do not correspond to an NRC), because the estimated LJTs are not reliable and there was no reported incident which could explain such increases in LJTs.

It was previously noted that increasing the maximum temporal window size in STSS is not always preferable, and this link illustrates why this is the case. Specifically, if the maximum temporal window size is increased from four to six, STSS would detect two more episodes: one starting at approximately 15:00 and the other starting at approximately 17:00. Because all the episodes detected on this link are likely to be false alarms, it is better not to increase the maximum temporal window size more than four in STSS in this case; even though such a decision yield a better match with e^*s.

Short Links

The ANPR network is heterogeneous in terms of variations in the link lengths. Link lengths vary from 175 meters to approximately 13 kilometres. The shorter links exhibit high variability around their expected values, as the estimated LJTs are usually very low. The timing of a traffic light, for example, could easily substantially affect the estimated LJTs on short links. Therefore, it is interesting to understand the performance of NRC detection models on the shortest link of London's ANPR road network (i.e. link 2098), which has a length of 175 meters. The link is located at the Elephant and Castle roundabout in South London. The results regarding the analysis of this link are illustrated in Figure 5-19.

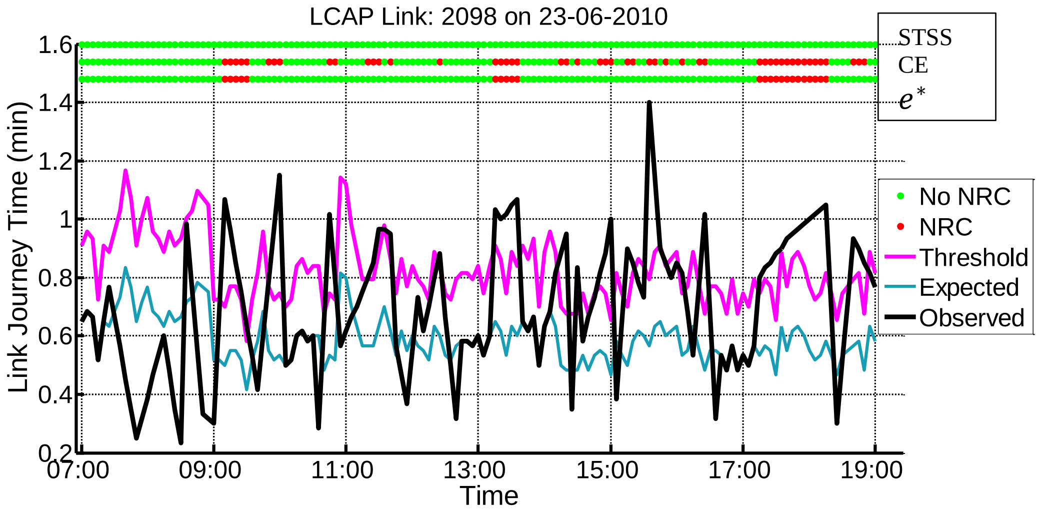


Figure 5-19 Detecting episodes on the shortest link of ANPR network on 23 June 2010

The advantages of STSS can be observed in this link, as it did not report any LJT as belonging to an NRC. Although some LJTs are excessive, this is probably due to day-to-day variations, as the delay is usually less than 30 seconds for most of the estimated LJTs. However, as shown in Figure 5-19, CE detected 17 episodes and e^* detected three episodes on this link. As there were no incidents reported on this link on 23 June 2010 and considering the amount of delay that the link exhibits at each LJT, these detected episodes are likely to be false alarms.

This finding adds further support to the fact that CE and e^* are not able to handle variations around the expected LJTs. This is because both approaches treat excessive LJTs as the main component of an NRC without considering whether the variation of the LJTs. The significance testing step in STSS ensures that variations in LJTs are taken into account. In this way, day-to-day variations due to link length are managed better in STSS than CE.

5.4.2. Determining the Impact of an Incident

Traffic operation centres are often interested in determining the impact of an incident. The impact of an incident can be observed as an NRC. The methods proposed in Chapter 3 could be used to detect such NRCs. Therefore, this part of the thesis demonstrates how the impact of an incident could be determined by using the proposed NRC detection models. Three NRC detection models are examined, and they are the same as discussed in the previous sub-section.

A traffic accident was recorded on 16 June 2010 that started at 10:04 and ended at 11:02. The accident is recorded as ‘serious’, which makes it even more interesting to investigate. The locations of the incident and links within the study area are shown in Figure 5-20. It should be highlighted that no other incident was recorded within the proximity of these links as can be seen from the figure. Therefore, this incident is the only known event that can have caused the increases in LJTs.

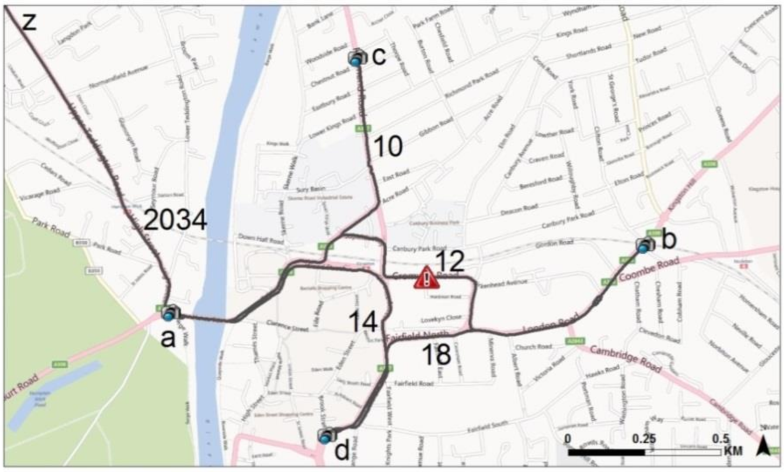


Figure 5-20 Serious incident at Cromwell Road on 16 June 2010

The characteristics of the five links that are proximal to the incident are summarised in Table 5-12.

Table 5-12 Characteristics of links that are proximal to the Cromwell Road incident on 16 June 2010

Link ID	**Traffic Flow Direction**	**Length (m)**	**Overlaps with**	**Adjacent of**
2034[2]	z – a	5035	–	10, 12, 2034
10	a – c	1346	12, 14	10
12	a – b	2095	10, 14, 18	12, 18
14	d – a	1398	10, 12, 18	10, 12, 14
18	b – d	1369	12, 14	14, 18

The ‘Overlaps with’ column indicates the links that spatially overlap with one another. The ‘Adjacent of’ column indicates the adjacencies of a link, where the traffic flow

[2] Link 2034 has not been shown to its complete spatial extent in order to make the rest of the study area more legible.

information is used to generate this information. The adjacency relationships are defined as described in section 3.1.

A spatial overlap relationship holds for different links in different contexts. For example, links 10 and 12 spatially overlap, and the traffic flow direction is the same on both links. It is likely that an incident occurring on either link would increase the estimated LJTs on the other as well. Thus, they may well be considered to be adjacent. On the other hand, links 12 and 18 also spatially overlap, but the traffic flow is in opposite directions. Thus, it is harder to think of these links as adjacent, as it is unlikely than an incident would affect both links. Therefore, this thesis does not consider spatially overlapping links to be adjacent. However, the developed NRC detection methods are capable of detecting that an NRC contains two spatially overlapping links, which is actually demonstrated by this example.

The incident that occurred on link 12 and the three aforementioned NRC detection models are used to determine its impact on LJTs. The estimated LJTs on 16 June 2010 are illustrated alongside the estimated LJTs on 15 June and 17 June in order to demonstrate that no other incident happened on these links that lasts for more than a day. The LJTs that are found to belong to an NRC are illustrated in Figure 5-21. The thresholds are calculated by multiplying the expected LJTs by the congestion factor that is used in both the CE and e^*model (i.e. 1.4).

Visual inspection of the estimated LJTs indicates the impact of the incident. Specifically, the episodes detected on links 10, 12 and 2034 can be associated with the presence of the incident. Links 10 and 12 had the highest increases in estimated LJTs, as the incident directly affected these links. It also appears that congestion built up due to the impact of the incident propagates to its link 2034.

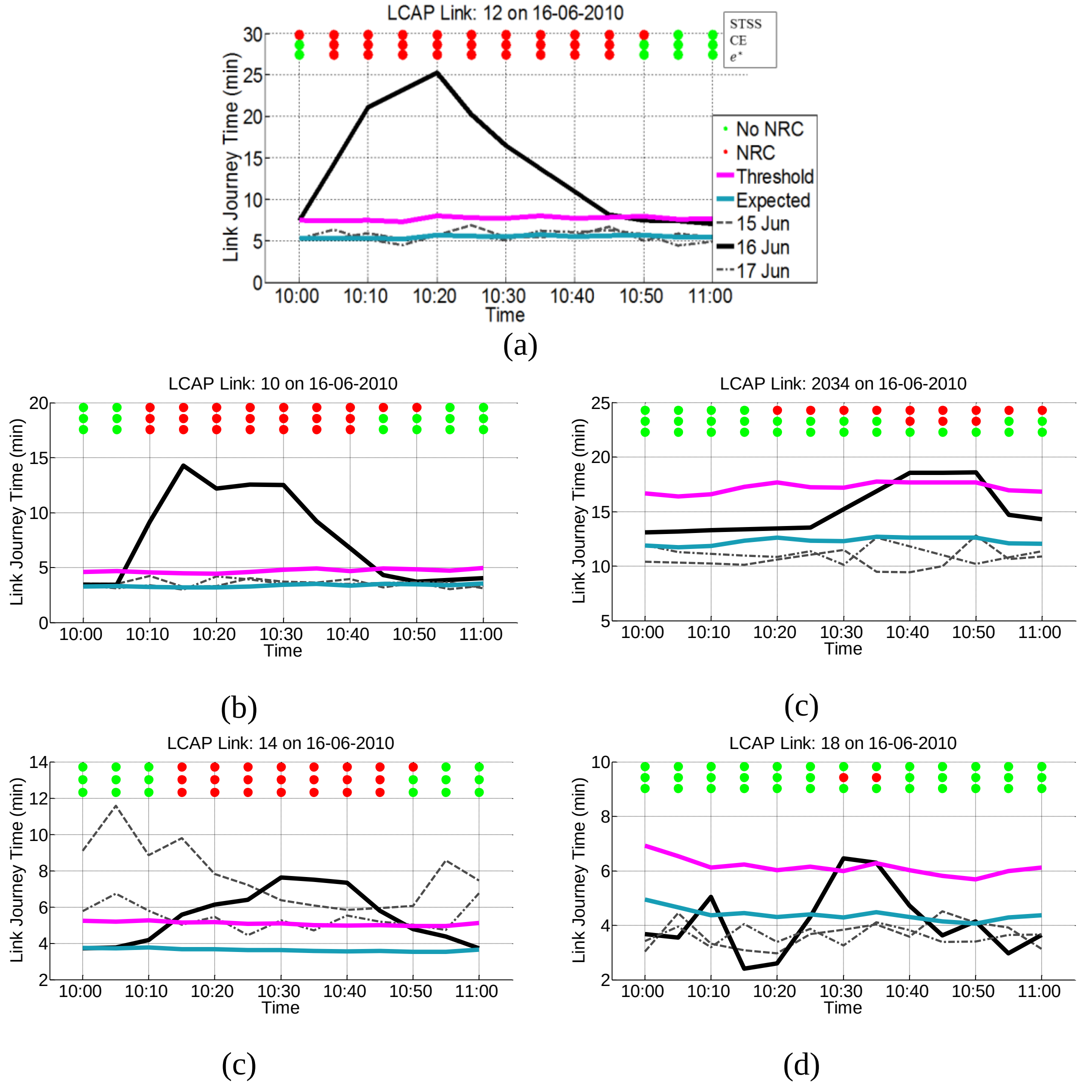


Figure 5-21 LJTs that are found to belong to an NRC on 16 June 2010 near Cromwell Road

The timings of the detected episodes also make sense. Specifically, the first LJT to appear to belong to the NRC occurred at 10:00 on link 12, which is the link where the incident happened. Ten minutes later, link 10 displayed its first LJT which belonged to the NRC because, during this 10 minute period, congestion propagated upstream of the incident towards the location where the two links diverged. Finally, the congestion propagated to link 2034 and STSS was able to capture the increase in the estimated LJTs at 10:20. Although it is more clear that the LJTs increased at 10:30 and this could be more easily attributed to the incident, the CE model experienced a delay in detecting this increase in LJTs and reported the first LJT as belonging to the NRC at 10:40. On the other hand, e^* did not detect any episode at all on link 2034 because the excessive LJTs lasted for only 15 minutes (i.e. three consecutive LJTs) when the congestion factor is set to 1.4.

Comprehension of the detected episodes on link 14 and 18 is more complicated. This is because the traffic flow direction is opposite to links 10 and 12. In addition, the previous and latter days also experienced similar increases in estimated LJTs. However, correspondence with a domain expert reveals that because link 2034 is congested, hence the location indicated by camera ‘a’, it could be possible that traffic signals could have adjusted to release the traffic building up on link 2034. This adjustment might have increased the journey times on link 14. On the other hand, link 18 is also flowing in the opposite direction to link 12, but the increased LJTs between 10:25 and 10:40 are specific to 16th June. This may also be explained by junctions that are common to both links 12 and 18 becoming congested, which might then have increased the estimated LJTs on link 18. Thus, it is also possible to associate the increased LJTs in link 18 with the incident on link 12. In this context, only CE was able to capture this increase as it detected an episode that lasted for 10 minutes.

These findings suggest that even though an incident affects two or more spatially overlapping links, it is still possible to detect the NRC on those links as long as they share a common upstream or downstream link. In this example, because the spatially overlapping links have a link that is either upstream or downstream to both, the effect of the NRC could be detected. This example also demonstrates that it is important to detect excessive LJTs that only last for a short period, as such increases could be attributed to an incident. This adds further support to the decision that NRC detection methods should preferably not make any assumption regarding the duration of an NRC.

5.5. Summary

This chapter illustrated the experimental results obtained from the analysis of NRCs on London’s urban road network. This chapter’s findings are summarised as follows.

The first section found the most suitable distribution to describe LJTs as lognormal distribution. Specifically, 63% and 88% of all LJTs are distributed according to lognormal distribution on raw and cleaned data respectively. This finding adds further support to literature suggesting the distribution of travel times is a lognormal distribution. This thesis used this finding to develop an expectation based STSS method to detect NRCs in section 3.3.

The second section determined the decisions regarding the experimental design. First, it is found that approximately 71.2% of all expected LJTs are the same regardless of the weekday. This finding allowed this thesis to consider all weekdays to be the same in its statistical analysis. Secondly, it is found that data quality is maximised between 07:00 and 19:00. Therefore, different NRC detection methods are compared during this period. Thirdly, anti-parallel links are found *not* to be adjacent, because 57.5% of the incidents that occur on either of the anti-parallel links do not affect the other anti-parallel link. Lastly, three types of days are determined on which NRC detection is applied. These days are identified based on travel demand, as follows: bank holidays, normal days and tube strike days. By analysing these three types of day, it is aimed to investigate the robustness of the proposed NRC detection methods under different travel demand conditions.

The third section compared the two NRC detection methods proposed in this thesis, where the analyses are conducted on three types of day: bank holidays, normal days (i.e. all October except for the tube strike day) and tube strike days that happened in 2010. The findings of this section are summarised as follows:

- Different CE models are compared by varying the congestion factor parameter. Congestion factor values varied between 1.2 and 2.0, which correspond to the operational values used in TfL. Increasing the congestion factor from 1.2 to 2.0 yields a more conservative NRC detection model. It is easier to relate the detected NRCs with incidents by conservative NRC detection models, as they have a low Localisation Index. However, they tend to miss e^*s, as they have high False Negative Rate (FNR) values. On the other hand, using a low congestion factor yields a liberal way of detecting NRCs, which are able to detect more e^*s, but such models are likely to consider day-to-day variations to belong to an NRC which makes it more difficult to relate the detected NRCs with incidents.
- Different STSS are compared by varying the maximum spatial and temporal window size parameters. In order to do this, all the STSS models used the lowest congestion factor that is used in CE, which is 1.2, and used a significance level of 0.05 to detect statistically significant STRs. It is observed that decreasing the maximum spatial and temporal window leads to a conservative model. This is because such models do not consider information coming from adjacent links/times. On the other hand, increasing the maximum spatial and temporal window size leads to more liberal ways to detect NRCs.

- Overall comparisons of the CE and STSS models are conducted in the third part. These comparisons involve two conflicting attributes (i.e. $e^*$and the Localisation Index) and 23 NRC detection models (i.e. five CE models and 18 STSS models). In order to avoid the rank-reversal problem that is common to many MADM methods, this thesis uses Weighted Product Model (Triantaphyllou and Mann, 1989). With an equal weighting of detecting high-confidence episodes and Localisation Index, it is found that the CE=1.4 model is preferable to most other models on all types of days. This finding could be related to the design of the e^*. Because, CE and e^* use the same idea to determine excessive LJTs and CE=1.4 uses the same congestion factor as e^*, the results might have favoured CE=1.4.

The fourth section investigated the effectiveness of exemplary NRC detection models, as well as high-confidence episodes on some specific exemplar links in order to demonstrate the effectiveness of NRC detection models in different situations. The section consists of two parts and the findings from these parts are summarised as follows:

- STSS is more useful to handle variations in the estimated LJTs. Specifically, day-to-day variations in short links and unreliable LJTs due to low-quality links are better handled by STSS. The exemplified situations also add further support to the decision that was made when designing the NRC detection methods; it is not reasonable for an NRC detection method to assume a minimum duration regarding the detected NRCs. Although such an assumption would substantially simplify NRC detection, it may not necessarily achieve what is intended because NRCs may last for a short time, but can still be attributed to an incident.

- Detection of NRCs could serve to determine the impact of incidents. Therefore, the second part focused on determining the impact of an incident and compared how different NRC detection models detect the impact it. An important aim of this part is to demonstrate that arbitrarily shaped NRCs could be detected by the proposed NRC detection methods.

6. DISCUSSION AND CONCLUSIONS

This chapter is devoted to summarising the major findings of this thesis. A brief overview of the thesis is presented in section 6.1, where the chapters are summarised. Contributions of this thesis are summarised in section 6.2 by linking them to the objectives of this thesis. The proposed NRC detection methods and evaluation criteria are critically assessed in section 6.3. Finally, directions for future research are provided in section 6.4.

6.1. Overview

Traffic congestion occurs when excess travel demand accumulates and surpasses traffic capacity (Varaiya, 2005). There are several components that interact with each other, and consequently effect travel demand and/or traffic capacity. Policy decisions, traffic management, weather and incidents are amongst the few notable components of traffic congestion. The interactions between these components form a system of organised complexity (Weaver, 1948) and make the research on traffic congestion a challenging and an interesting quest.

Recurrence of traffic congestion depends on the temporal scale on which the traffic data are analysed. It has already been argued that *recurrence* exists in most complex systems (e.g. climate) given a long duration for observations (Marwan et al., 2007). In traffic science, variations in weather and traffic volume amongst many others have been considered to be *recurring* (Lomax et al., 2003). Nevertheless, most traffic scientists and domain experts accept that traffic congestion can be divided into two: Recurrent Congestion (RC) and Non-Recurrent Congestion (NRC), in which the temporal scale to analyse traffic data is considered to be a 'day'. Therefore, RC exhibits a daily pattern and mainly caused by excess travel demand, inadequate traffic capacity or poor signal control. Commonly used terms like morning or afternoon peak periods correspond to RC.

The concept of NRC has been recognised for decades, but this recognition was mostly limited to the relationship between NRC and incidents. Specifically, an NRC is assumed to be caused by an incident. A vast amount of research conducted on Automatic Incident Detection (AID) led to the common usage of NRC within AID research. However, empirical findings suggest that many large delays cannot be explained via an incident data set (Hallenbeck et al., 2003b). Furthermore, it is theoretically shown and empirically observed that traffic congestion can emerge simply due to interactions between the drivers, causing what is referred to as ‘phantom jams’ or traffic congestion without a specific cause (Helbing, 2001; Kerner and Konhäuser, 1994; Sugiyama et al., 2008).

Most existing research has related NRC to incidents that occur on motorways. However, research on motorways usually deals with a linear road network. On the other hand, a growing number of traffic operation centres manage *urban* road networks containing *hundreds of links* that are linked with a graph structure. Therefore, it is of growing research interest to understand NRC in an urban road network consisting of hundreds of links, so that strategies can be developed to reduce NRC. These facts motivated this thesis to develop an NRC detection methodology to support accurate detection of NRCs on an urban road network.

Chapter 2 introduced the various aspects of research on NRC detection in an urban road network by reviewing the literature. These aspects are summarised as follows:

- The differences between motorways and urban road networks are highlighted in section 2.1.1. In an urban road network, traffic flow has regular and irregular interruptions. In addition, variations in link length and data quality lead an urban road network to be heterogeneous. These issues prevent knowledge that is mostly developed for motorways from being used for urban road networks; hence, making research on urban road networks difficult (Lee et al., 2011; Ozbay and Kachroo, 1999, p.79).
- The suitability of Link Journey Time (LJT) for research on a large road network is highlighted in section 2.2, as LJTs are estimated over a length of a road rather than a section of a road. Two main issues have been discussed within this section; the first is the estimation of an LJT. Although the procedure used to estimate an LJT is not within the scope of this thesis, an example is demonstrated to clarify the issues that arise during this process. The second is investigating the distribution of LJTs. The distribution of LJTs should satisfy two requirements: it should always be positive as no LJT could take a negative value and it should be positively skewed as most trips are completed within a narrow time interval due to the recurrent nature of traffic. Lognormal distribution is reported to satisfy these two requirements (Arezoumandi, 2011; Dandy and McBean, 1984; Hollander and Liu, 2008).

- Two main strategies to perform clustering are discussed in sections 2.4.1 and 2.4.2 respectively: clustering by using a similarity measure and clustering by significance testing. Similarity based clustering relies on a similarity measure that is defined between observations. Previous research has investigated the usage of similarity based clustering for traffic congestion detection on LJT data (Anbaroglu and Cheng, 2011) and trajectory data (Li et al., 2007; Ying et al., 2009), but these studies were not able to differentiate whether the detected congestion belongs to an RC or an NRC. Significance testing based clustering is often regarded as detecting clusters that are of statistical importance. Space-Time Scan Statistics (STSS) is regarded as the state-of-art method to detect statistically significant clusters. Previous studies on STSS have not explored the development of a significant cluster in space and time. Furthermore, common STSS approaches use overlapping space-time regions to scan the entire study area, which becomes computationally intractable as the study area becomes larger (Neill and Moore, 2004).

- Evaluation of the performance of an NRC detection method is discussed in section 2.5. Because this thesis considers NRC detection to be a topic under clustering, strategies to evaluate clusters are investigated. Two main strategies are identified, as internal and external criteria. Internal criteria rely only on the data where the clusters are found and various criteria are used to quantify the extent to which the detected clusters are well-separated and well-connected. External criteria rely on ground truth data, which is external data in which 'true' clusters are known. Ground truth data can be obtained in one of the three possible ways: the first uses simulations where a traffic model is used to generate traffic data, the second is domain experts manually labelling LJT data that belong to an NRC and the third is using incident data set, where a direct relationship between incidents and NRCs is assumed.

Chapter 3 proposed two novel NRC detection methods based on each of the clustering strategies – similarity based and significance testing based. The first method detects NRC by clustering spatio-temporally overlapping episodes, where an episode is defined as a maximal interval during which all LJTs are excessive on a link. This method is introduced in section 3.2. In order to quantify each NRC, section 3.2.3 described how the evolution of an NRC can be determined. In this way, each NRC is quantified by two characteristics: lifetime and severity. Lifetime is the time interval during which an NRC contains at least one link and severity is the total excess LJT that an NRC exhibits. The second method detects NRCs by using expectation based STSS. Firstly, the null and alternative hypotheses are identified according to the empirical evidence provided in section 5.1, which shows that LJTs can be modelled by lognormal distribution. Using the null and alternative hypotheses, a likelihood ratio function is derived. The derived likelihood ratio function allows a researcher to determine statistically significant Space-Time Regions (STRs), which are then clustered to detect NRCs. This method is described in detail in section 3.3.

Chapter 4 focused on evaluating the performance of NRC detection methods. As discussed previously, two main strategies can be used to evaluate the performance of an NRC detection method. The first uses external criteria, which requires the availability of ground truth data. Because there is a lack of research on simulating NRCs on a large urban road network consisting of hundreds of links and it is cost-prohibitive for domain experts to manually label LJT data, section 4.2 investigated whether an incident data set can be used as ground truth data to validate the detected NRCs. However, empirical findings demonstrate that the available incident data set is not suitable for this purpose. This is mainly because LJT data and incident data sets are generated by completely different priorities and resources. Although there is an overlap between these data sets, several reasons (e.g. unrecorded incidents, non-disruptive incidents or inactive cameras) prevent the incident data set from being used as ground truth data for the detected NRCs. Therefore, section 4.3 proposed two novel criteria considering the practical uses of an NRC detection method. The first is to detect NRCs that result in substantial impact. The second is to relate the detected NRCs with incidents. These two criteria are called high-confidence episodes and the Localisation Index respectively.

Chapter 5 analysed the performance of the proposed NRC detection methods on London’s urban road network, which consists of 424 links. Section 5.1 investigates the distribution of LJTs for both cleaned and raw LJT data. Section 5.2 determines the effect of weekdays on the estimated LJTs, the effective analysis period and the adjacency of anti-parallel links. Section 5.3 presents the results obtained by using the proposed NRC detection methods. The analysis focused on three types of day which correspond to the lowest, medium and highest travel demand. In this way, the robustness of the proposed NRC detection methods on different levels of travel demand is tested. Section 5.4 investigates the effectiveness of the proposed NRC detection methods on some real-life scenarios, which also provide useful information on the relative advantages and limitations of the proposed NRC detection methods.

6.2. Main Contributions

This thesis contributes to the literature by developing an NRC detection methodology to support the accurate detection of NRCs on large urban road networks by investigating on the possible answers to the following two questions:

- How can NRCs be detected on a large urban network?
- How can the performance of an NRC detection method be evaluated?

The formal investigation carried out within this thesis on these questions has resulted in several contributions. These contributions are described, in relation to the objectives stated in section 1.2, as follows:

Objective 1: *Develop an NRC detection methodology for a large urban road network.*

The effectiveness of using spatio-temporal clustering of substantially high Link Journey Time estimations (LJTs) for NRC detection has been demonstrated. In this respect, two novel methods have been proposed to detect NRCs based on LJT data. The first method clusters spatio-temporally overlapping episodes and is discussed in section 3.2. The main parameter of this method is the congestion factor, which determines whether an LJT is excessive or not. The second method clusters statistically significant Space-Time Regions (STRs) that are determined by using an expectation based Space-Time Scan Statistics (STSS), which is the state-of-art method to detect statistically significant clusters. This thesis illustrates, for the first time, how STSS can be used in traffic science for the purpose of NRC detection. Because STSS has been mostly investigated for disease outbreak detection, the following adjustments were proposed in order to use it for NRC detection:

- Consideration of all statistically significant STRs rather than only the most significant STR. By clustering the statistically significant STRs it became possible to observe the development of an NRC dynamically. In addition, the likelihood ratio value of only those STRs whose individual LJTs are excessive are evaluated. This adjustment reduced the number of STRs that should be evaluated; hence, increased the computational efficiency of the method.
- Derivation of a likelihood ratio function based on lognormal distribution. It has been shown that 63% and 88% of all LJTs are distributed according to lognormal

distribution on raw and cleaned data respectively. This finding is novel in the sense that no previous study has investigated the distribution of LJTs on a road network as large as that featured in this thesis. The existing literature on spatial or spatio-temporal scan statistics mainly uses Poisson or Bernoulli distribution to model the data. This adjustment led to understanding the performance of STSS when data is distributed according to lognormal distribution.

Objective 2: *Quantify the impact of an NRC event.*

In order to compare different NRC events (NRCs) and to estimate the impact of incidents, detected NRCs are quantified by two characteristics: lifetime and severity. Lifetime of an NRC is its temporal extent, and represents the time period that the NRC contains at least one link. Severity of an NRC is used to determine the impact of the NRC in terms of the total excess LJT it exhibits. These two quantifiable characteristics provide useful information to quantify the impact of an NRC event.

In order to quantify an NRC event, the evolution of that NRC is determined. In this way, at each time interval the links that belong to the NRC would be known. By representing each NRC event in its evolution, the two quantifiable characteristics (i.e. lifetime and severity) of the NRC could be determined.

Objective 3: *Understand the difficulties in evaluating the effectiveness of an NRC detection method by using an incident data set.*

In order to evaluate the effectiveness of an NRC detection method, the available incident data set has been investigated to determine whether it can be used as ground truth data. This investigation identified five reasons that make it difficult to use the incident data to validate the detected NRCs in a real-life context. These reasons are summarised as follows:

- An incident may occur on a road where no LJT data are collected, as LJT data are gathered using ANPR cameras, whereas incidents are recorded by surveilling CCTV cameras.
- Substantial increases in LJTs may not be explained by the incident data set, due to the subjective process of recording incidents.
- Recorded incidents may be non-disruptive; hence, they may not cause a substantial increase in LJTs.

- An incident may occur on links when no LJT are estimated, as ANPR cameras may be inactive during that period.
- An incident's beginning/end times encompass many uncertainties. This prevents the temporal alignment of incident and LJT data sets.

The empirical investigation carried out in thesis demonstrated that these issues are commonly observed in a real-life context. Hence, they prevent the available incident data set from being used as ground truth data in order to validate detected NRCs.

Objective 4: *Determine criteria to evaluate different NRC detection methods that consider high LJT as an indicator of NRC*.

The lack of ground truth data (i.e. the lack of information on whether an LJT belongs to an NRC or not) leads this thesis to propose novel criteria to evaluate the performance of NRC detection methods. Two criteria have been proposed by considering the applications of an NRC detection method. The first is to detect NRCs that result in substantial impact. The second is to be able to relate the detected NRCs with incidents. The terms 'High-confidence episode' and 'Localisation Index' are used to refer to these two applications.

A '*High-Confidence Episode*' (e^*) is an NRC on a single link that lasts for a minimum duration, during which all LJTs are excessive. This evaluation criterion assesses to what extent an NRC detection method successfully detects e^*s. The values of the two parameters had to be found to define an e^*. Firstly, a congestion factor has to be determined in order to establish the excessive LJTs. This was achieved by finding the average congestion factor that would classify the 95$^{\text{th}}$ percentile LJTs as excessive. The empirical analysis conducted in section 4.3.1 identified 1.4 to be a suitable value as the described congestion factor. Therefore, whenever an LJT is higher than 40% of its expected value, it is considered to be excessive. Secondly, a minimum duration has to be determined. This was achieved by detecting all episodes on all links and ranking them based on their duration. The cumulative severities based on the duration of the episodes are found, and the maximum duration that accounts for the majority of the delay was selected as the minimum duration. Historical analysis showed that episodes that last for at least five consecutive LJTs (corresponding to 25 minutes) is sufficient to account for the majority of total delay in London's urban road network. False-alarm rate (FAR) and false-negative rate (FNR) are calculated based by comparing the detected

NRCs and e^*s. The lower the FNR, the better the NRC detection model, as this means it missed fewer e^*s.

The '*Localisation Index*' assesses to what extent the detected NRCs can be related to incidents, which is calculated by finding the highest average number of connected components that an NRC exhibits throughout its lifetime. A connected component is a structure that contains any number of links with the condition that any two of these links could be connected to one another through the other links within the connected component. At each temporal interval during the lifetime of an NRC, the number of connected components is determined, and the highest average number of connected components is considered to be the 'Localisation Index' of the NRC detection model. The higher the Localisation Index, the harder it is to relate the detected NRCs with an incident. This is because there is a greater loss of spatial connectivity of the links that belong to an NRC, which decreases the possibility of relating a detected NRC with an incident.

The main advantage of the developed NRC detection methods and evaluation criteria is the extent to which domain knowledge and practical usage of is considered. For example, congestion factor values are determined by considering the values that are used in practice. This may initially appear to be 'non-scientific', however, there is research evidence addressing the necessity/importance of using domain knowledge in data mining methods. Effective usage of domain knowledge is actually considered to be one of the 10 preeminent challenges in data mining (Yang and Wu, 2006).

Finally, an urban road network consisting of 424 links is explored, a size that was hardly matched previously. This endeavour involved many challenges. For instance, there is a lack of research on representing the adjacency of overlapping and anti-parallel links. Section 3.1 introduced these issues, and discussed how they can be handled for NRC detection. In addition, such a real life network positioned this thesis to develop NRC detection methods that could be used in a traffic operation centre.

6.3. Critical Assessment of the Methodology

This section critically assesses the proposed methodology on detecting NRCs and evaluates the performance of the proposed NRC detection methods. The proposed NRC detection methodology consists of two methods and both require four inputs. The first is the adjacency matrix, which is the mathematical representation of the geographical road network based on the topological relationship between the links. The second is historic LJT data, which are used to determine the expected LJTs and the distribution of LJTs. The third is the congestion factor, which, together with the expected LJTs, determines whether an estimated LJT is excessive or not. Finally, a date is input on which the NRCs are to be detected. Both of the proposed NRC detection methods rely on these four inputs.

An important input is the adjacency matrix, which is the mathematical representation of a geographical road network. Defining an optimum adjacency relationship is important, as the spatial extent of an NRC may be overestimated if more links than the optimum number are included in the adjacency list of a link or all NRCs would consist of one link if all the topological relationships between links are ignored. However, there is usually very little formal guidance on such an optimum choice (Anselin, 2002). In addition, there is no common agreement on defining an optimum adjacency matrix. Further discussion on this topic is provided in Appendix A.

The difficulty in determining the optimal adjacency matrix has several aspects; firstly, in a real life road network, links may exhibit complex topological relationships. For example, it is difficult to determine the adjacency relationship between overlapping links, as they may occur in several different scenarios. These scenarios are illustrated in Figure 6-1, where ANPR cameras (i.e. nodes) are denoted with letters x, y, w and z and a link is represented by two ANPR cameras. The traffic flow on a given link starts from the node denoted with the first letter and ends at the node denoted with the second letter. For example, in Figure 6-1(a) there are two links xy and xz, and the traffic is flowing from x to y and x to z respectively. If there is no arrow, it means that on that section of the road network traffic is flowing both ways.

The scenarios illustrated in Figure 6-1(a-d) occur due to the partial coverage of the road network by ANPR cameras. If one more camera was located at the merge/diverge location of the links, then such overlapping links would not occur. On the other hand, situations illustrated in Figure 6-1(f-g) could happen in a city centre where the road network is dense and there are several main routes intersecting and diverging within a short distance.

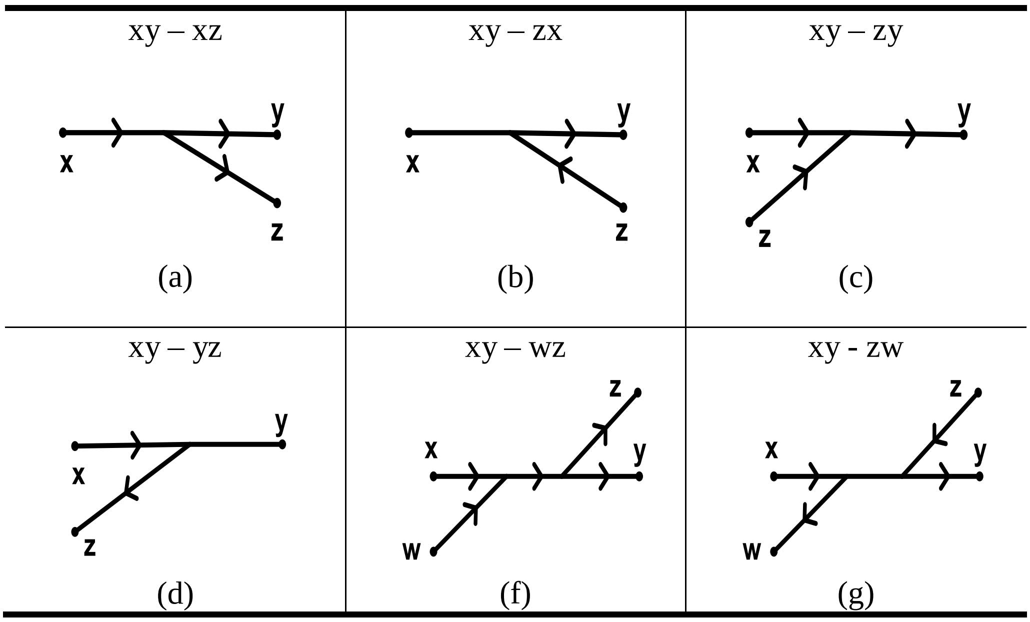


Figure 6-1 Different scenarios where two links overlap

The illustrated scenarios do take place in a real life network and these possibilities make it difficult to determine the adjacency of overlapping links. In order to simplify the analysis, this thesis assumed that two overlapping links are *not* considered to be adjacent if they do not possess the adjacency relationship. Two links exhibit adjacency relationship as long as the beginning of one link is the ending of the other link. Therefore, amongst the scenarios illustrated in Figure 6-1, only the links within Figure 6-1(b) exhibit adjacency relationship.

It is reasonable to expect that an incident occurring on the overlapping part of two links may actually increase the estimated LJTs on both links. In such situations, the spatial extent of a detected NRC could be underestimated. Note that; however, if the overlapping links share a common link that has adjacency relationship with both of them, then they can still belong to the same NRC as both of the proposed NRC detection methods can detect arbitrarily shaped NRCs. An example in which the detected NRC consists of overlapping links is demonstrated in section 5.4.2.

The second important input to the NRC detection methodology is the congestion factor, which is used to determine whether an LJT is excessive or not. If the estimated LJT is higher than a threshold that is defined by multiplying the congestion factor with the expected LJT, then the estimated LJT is excessive. Excessive LJTs are considered to be the main component of an NRC.

The main limitation to determine excessive LJTs is the reliance on a single congestion factor that is common for all links at all times. However, road network is heterogeneous in terms of the quality of links and link lengths. The quality of a link is associated with the reliability of the estimated LJTs. A low quality link may initially appear to consist of many excessive LJTs during an analysis period, but this is actually due to the fact that the LJTs on that link are not reliable. Link lengths also vary considerably due to the heterogeneous distribution of ANPR cameras. Day-to-day variations on short links may be considered to be excessive if the same congestion factor is used for detecting excessive LJTs on longer links. The substantial variation between the congestion factors that would classify 95th percentile LJTs as excessive in Figure 4-14 demonstrates the validity of this limitation.

6.3.1. Detecting NRCs by Spatio-Temporal Clustering

Two methods have been proposed to detect NRCs and these are a realisation of a spatio-temporal clustering strategy. The first method, Clustering Episodes, is the realisation of a similarity based clustering strategy. The second method, STSS, is the realisation of significance testing based cluster detection.

The first NRC detection method, Clustering Episodes (CE), defines a similarity function between the estimated LJTs and expected LJTs. This is a substantial difference between this thesis' approach and the traditional approach which would define a similarity function between the estimated LJTs only. The main reason for this modification is to focus on NRC detection. This method determines the maximal intervals on a link during which all LJTs are excessive which is called an episode. Two episodes spatio-temporally overlap if they occurred on adjacent links and they share a common time interval. Such spatio-temporally overlapping episodes are clustered to detect NRCs.

This method is simple as it only relies on one input: the congestion factor. If the congestion factor is set too high, then a 'true' NRC might be underestimated or may even be missed. If it is set too low, day-to-day variations in traffic may be considered to belong to an NRC, which leads to intractable growth of an NRC. Therefore, during the analysis, the congestion factor is varied between 1.2 and 2.0. These values are determined considering domain knowledge, and they are currently being used at Transport for London (TfL) to classify congestion events into discrete categories like moderate, serious or severe. Analyses conducted on bank holidays, normal days and tube strike days demonstrate that a congestion factor of 1.4 achieves a good overall performance, which is calculated by equally weighting the detection of e^*s and the Localisation Index.

Although CE is a simple and yet effective method used to detect NRCs, the main limitation of this method is that it is indifferent to variations in the expected LJTs, as it only considers the expected LJTs and the congestion factor to determine excessive LJTs. This leads to the detection of NRCs that do not truly have a physical significance. Analyses of the exemplar links in section 5.4.1 demonstrate that day-to-day variations in short links or erroneous LJTs due to a low sample size are considered to belong to an NRC when CE is used.

The second NRC detection method, STSS, uses the state-of-art statistical method to determine statistically significant clusters, which is space-time scan statistics. An expectation based STSS is developed and applied for the first time to detect NRCs. This method uses lognormal distribution to characterise each LJT and develops a likelihood ratio score accordingly. Space-Time Regions (STRs) are used to search for significant increases in LJTs. Unlike the traditional approach to searching for all possible STRs, the proposed method investigates STRs whose individual LJTs are excessive. If only a single LJT within an STR is not excessive, then that STR is not investigated for its significance. Once the significant STRs are detected, those which spatio-temporally overlap are clustered to detect NRCs.

The main advantage of STSS is that it considers the variations in expected LJTs when determining the likelihood ratio function, as well as the expected LJTs. In section 5.4.1 two exemplar links demonstrate this advantage in different ways. Firstly, short links exhibit higher variations in the estimated LJTs, as small changes may substantially affect the estimated LJTs (e.g. variations in traffic light timings). Secondly, longer, but unreliable links also exhibit high variations in the estimated LJTs. It is observed that STSS is capable of handling both types of variations to a certain extent depending on the maximum spatial and temporal window size values.

The main limitation of this method arises when deriving the likelihood ratio function. Specifically, LJTs are assumed to be independent of one another, and Monte Carlo simulations are generated accordingly. Because LJTs exhibit serial autocorrelation, the results obtained by increasing the maximum temporal window size (τ) lead to a more liberal way of detecting NRCs. This is because consideration of a greater number of consecutive LJTs on a link increases the likelihood that they will be detected as significant. In this way, more STRs are considered to be significant; hence, as τ is increased, more e^{*}s are detected. However, as the STSS models become more liberal by increasing τ, a higher Localisation Index is also observed indicating the increased difficulty to relating them with incidents.

Both of the proposed NRC detection methods have a common limitation, as they detect NRCs by clustering *spatio-temporally overlapping* episodes / significant STRs. This limitation has two aspects:

- If two or more episodes / significant STRs are spatio-temporally overlapping, then they are assumed to belong to the same NRC, even though their underlying causes are different. For example, if two accidents happen at the same time on adjacent links, the increased LJTs on both links would be governed within a single NRC. In this way, the quantification of either of the incidents by themselves becomes a difficult task as the detected NRC is influenced by two incidents.
- 'The *flows* on a road network are not considered. An incident happening at one link might show its effect at its adjacent link after a time lapse. During the elapsed time the adjacent link might become congested, whereas the link where the incident occurred becomes decongested. As a result, an episode / significant STR could be observed on both links individually that are *not* spatio-temporally overlapping. These episodes / significant STRs would be considered to belong to different NRCs, but actually these NRCs are due to a single cause.

Therefore, the proposed NRC detection methods do not guarantee any causal relationship between a detected NRC and a recorded incident.

6.3.2. Evaluating the Performance of NRC Detection Methods

Evaluating the performance of an NRC detection method is very difficult due to the lack of traffic theory that can be used to define an NRC in an urban road network context. In addition, ground truth data are not available, as there is lack of research on simulating NRCs on a large urban road network and it is cost prohibitive to manually label LJTs. Therefore, an incident data set has been investigated with regard to whether it can be used as ground truth data to evaluate detected NRCs. Reasons including unrecorded incidents, non-disruptive incidents or inactive cameras prevent the incident data set from being used as ground truth data for the detected NRCs.

This thesis proposed two evaluation criteria to assess to what extent an NRC detection model captures the two different uses of an NRC detection method.

The first use is to detect NRCs that caused substantial impact. In order to quantify what is meant by 'substantial impact', a 'high-confidence episode' (e^*) definition is given that relies on a multiplicative increase in the estimated LJTs that last for a minimum duration. Although this definition is easy to understand and implement, it has two main limitations; firstly, it favours liberal ways of detecting NRCs. For example, Clustering Episode models that use the same or lower congestion factor than that used by e^* resulted in the best performance, as they did not miss to detect any e^*. Secondly, all e^*s are considered to be equal in terms of their contribution to the False-Negative Rate (FNR), regardless of their severity. Specifically, missing two e^*s that have the same duration, but different severities causes the same level of impact towards the calculation of FNR.

The second use is to be able to relate the detected NRCs with incidents. In order to relate an incident with an NRC, the detected NRC should consist of links that are adjacent throughout its lifetime. The 'Localisation Index' is an evaluation criterion that determines the maximum average number of connected components that the detected NRCs exhibit throughout their lifetime. A limitation of this criterion is as follows. It is indifferent to the order of occurrence order of the number of connected components. Specifically, it does not consider the occurrence order of the connected components. An example of this situation is illustrated in Figure 4-18. Considering the temporal aspect in a more systematic way would improve the quality of Localisation Index criterion.

The overall evaluation of the proposed NRC detection methods rely on Weighted-Product Model (WPM), which is a type of Multi-Attribute Decision Making (MADM) method that does not suffer from rank-reversal (i.e. introduction of a new NRC detection model that is inferior to the existing models may alter the decision regarding the best-performing model). WPM relies on the pair-wise comparison of different NRC detection models with a ratio; hence, none of the attribute values should be zero. However, this is not the situation in this thesis as CE $\leq$ 1.4 models do not miss to detect any e^*; hence, have an FNR of zero. As there is no known solution to this limitation of WPM (Mela et al., 2012), this thesis considers the second smallest FNR value for those CE models.

6.4. Future Research

In order to explore the full potential of NRC detection on urban road networks, this thesis could be extended in the following directions:

- Incorporation of domain knowledge when defining the adjacency matrix remains a promising future research direction. This could probably be achieved in two ways; the first is determination of the adjacency matrix by domain experts. The second is determination of the adjacency matrix in real time. The first approach is simpler than the second, and the determined adjacency matrix will be a refined version of the adjacency matrix that is used in this thesis. However, this approach cannot adapt to the dynamics of the traffic situation, so an NRC detection method might still over/under estimate the severity of the 'true' NRC, which could ideally be related with a single incident. On the other hand, the second approach would require the development of semi-automatic NRC detection methods, which could assist a traffic operator to determine the 'true' extent of an NRC. Then again, such semi-automatic methods would be more complex, which might limit their common usage in traffic operation centres.
- In order to derive the likelihood ratio function used in expectation based STSS in an efficient manner (as discussed in section 3.3.2), LJTs are assumed to be independent of one another. However, LJTs usually exhibit serial autocorrelation. Therefore, it would be a very interesting research direction to consider the temporal correlations between LJTs and model STSS accordingly.
- Congestion factor parameter determines whether or not an LJT is excessive. This thesis assumed that the congestion factor parameter has a minimum value determined by domain knowledge (i.e. 1.2). Therefore, all the tested STSS models used this congestion factor to ensure that each LJT observation within an STR exhibits excess LJT. However, it would be intriguing to explore the impact of increasing the congestion factor when detecting NRCs using STSS. This research direction might be especially useful to prevent the intractable growth of NRCs on high travel demand days (i.e. high Localisation Index values), which are observed when tube strike days are analysed. In this way, contingency plans to reduce the impact of tube strikes could be developed more effectively.

- It would be fascinating to apply the ideas developed in the thesis to different urban road networks and explore how the results change. For example, determining to what extent the congestion factors analysed in this thesis apply to other urban road networks would be of value. The vision of this research direction is to develop a formal traffic theory to explain NRCs on an urban road network. Such a theory would be invaluable, as it would allow researchers to develop traffic simulations that generate NRCs on a large urban road network.
- Fusing probe-vehicle data (e.g. mobile phone data) with the proposed NRC detection methodology could potentially improve the accuracy of the detected NRCs. In this way, the current limitations regarding missing data, low quality LJTs or handling day-to-day variations on short links might be reduced.
- Further research is required to attain the full potential of the developed NRC detection methodology, which would include developing strategies to reduce NRC. The ideas developed within this thesis could be used as an input to advancing such strategies. For example, engineering companies could be charged according to the total impact they cause to the road network, which could be estimated by the proposed NRC detection methods.

APPENDIX A: SPATIAL AND SPATIO – TEMPORAL NEIGHBOURHOOD

Spatial and Spatio-Temporal Neighbourhood

The ubiquitous usage of the spatial and spatio-temporal neighbourhood (or context) throughout spatial and spatio-temporal clustering make this concept important. This appendix is dedicated to providing a more in-depth understanding of what is meant by spatial and spatio-temporal neighbourhood (denoted as SN and STN respectively) and their usage in spatial information sciences.

The spatial relationship between the observations is usually characterised by an adjacency matrix. The method of defining an adjacency or neighbourhood is crucial, as it poses an initial bias on the analysis; most of the time an adjacency matrix is determined by applying some ‘rules of thumb’. For example, two areal regions sharing a common boundary or two points that are within a certain distance to each other are assumed to be neighbours. More complicated adjacency scenarios can also be incorporated, like considering shared boundary length as a way to normalise the adjacency matrix. The decision to define adjacency might not necessarily reflect reality, and the resulting prejudice might lead to biased results. For example, in political science, two neighbouring countries might be completely different (economically and culturally) which will result in spatial heterogeneity (Anselin, 1995). Even though these countries might share a common border, in order to prevent a misleading interpretation of the results, it might be better if they were not viewed as adjacent in spatial analysis. In other words, ‘rules of thumb’ are generally defined considering only spatial homogeneity. However, spatial heterogeneity is also observed in spatial information science and most of the ‘rules of thumb’ cannot handle spatial heterogeneity (Griffith, 2006).

In order to provide a complete treatment on the topic, a brief historical progress of the concept is useful. The initial thought on adjacency was a binary relationship between two spatial regions. Thus, two spatial regions can either be adjacent or not and this depends only on whether they are spatially adjacent or not. This is the case in defining the adjacency matrix in Moran’s I (Moran, 1948) and Geary’c C (Geary, 1954). This rule constituted the first of the ‘rule of thumb’ to define adjacency between spatial regions. However, this proved to be an over-simplification of the issue and Dacey

(1965), as noted in Cliff and Ord (1981, p.15), points out that this way of representing spatial relationships might overlook the topological differences between different spatial units; this issue is referred to as '*topological invariance*'. As the creation of the adjacency matrix is based on some pre-defined rules (e.g. two spatial units are adjacent if they are contiguous), by altering the topology while considering these rules, one can generate completely different topologies with the same adjacency matrix. An example to illustrate this issue is presented in Figure A- 1, where there are four different topologies and if an adjacency is assumed to exist between two regions that share a border, then these different topologies will be represented by the same adjacency matrix.

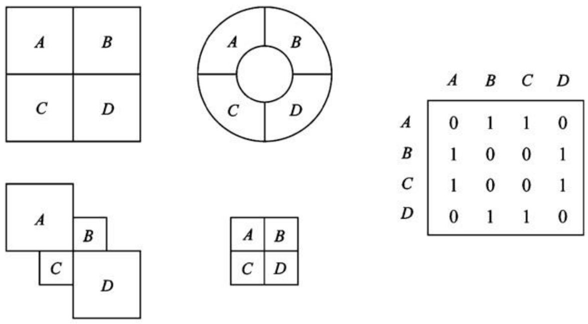


Figure A- 1 Different Topologies – Same Adjacency Matrix

Source: Getis (2009) with the permission of John Wiley and Sons

Cliff and Ord (1968) handle topological invariance by suggesting various types of spatial connections (not only spatial contiguity) via the use of flexible weighting schemes. It is stated that the only advantage of using a binary weight matrix is the ease of computation, and with the advance of computer power its relaxation will no longer be a limitation. However, it is clear that the way to define an adjacency matrix is a matter of practical consideration. For example, Getis and Aldstadt (2004) propose an adjacency matrix definition considering local variations via use of the Getis-Ord statistic and illustrate its effectiveness in detecting arbitrarily shaped neighbours. However, these models are complex and often result in arbitrarily shaped neighbourhood relations. Therefore, more research is required to demonstrate their effectiveness in practical applications.

Detection of the optimum spatial neighbourhood is an open-research area in spatial information science. This is because, when generating the adjacency matrix, most of the time spatial homogeneity is considered, and spatial heterogeneity is ignored. Unless there is a theory or strong belief that can be used to define the neighbourhood /

adjacency relationship, it might be useful to experiment with different 'rules of thumb'. Such a sensitivity analyses allow researchers to observe the effects of different adjacency matrices to describe the data.

Defining SN and STN for Different Spatial Data Forms

SN and STN concepts are widely used for all of the spatial data forms, including point, areal and network data. These data forms are used as the basis for various spatio-temporal data mining tasks including clustering, classification, prediction and anomaly detection. The main reason for researchers to use SN and STN is to model spatial correlation in modelling spatial or spatio-temporal phenomena, and hence reduce the error caused by ignoring such relations. This importance of SN and STN is either stated implicitly considering thematic (or semantic) relationships (Adam et al., 2004; Cheng and Li, 2006) or by homogeneity/heterogeneity relationships (Janeja et al., 2010). This section classifies spatial data into three main forms: *point*, *area* and *network*. The techniques used to define SN and STN according to these data forms are discussed in this section as follows.

There are three main techniques to define the spatial neighbourhood for *point data*. The first is to use a distance metric (e.g. Euclidean or Mahalanobis) and then define a threshold which will indicate whether or not two observations are neighbours. For example, Huang et al. (2006) and Celik et al. (2006) use Euclidean distance to create the spatial neighbourhood of each observation in order to detect the co-occurrent patterns. This approach, however, cannot handle the population differences between different regions. For example, a spatial neighbourhood within a dense region will contain more points than a sparse region and this might affect the integrity of the analysis. In order to overcome this limitation, the second way to define the spatial neighbourhood is to use *k*-closest observations. For example, Lu et al. (2003) use *k*-Nearest Neighbour (*k*-NN) based on Mahalanobis distance to find the spatial neighbourhood of an observation. In this case, however, the method of choosing the value of *k* makes the evaluation complicated and requires sensitivity analysis. The third way to define the spatial neighbourhood for point data is to transform the point data into polygons, which can be achieved by Delaunay Triangulation. Once the points are transformed to areal regions, those regions which share a common edge will be in their spatial neighbourhoods (Min-qi et al., 2008).

There are several ways to define the SN for *area data*. If the areas are in a regular structure, then this type of data form is called grid data. The neighbourhood strategies can be considered to be metaphors of chess pieces. Rook and queen neighbourhoods are the most common ways to define SN. For example, Yin and Collins (2007) define STNs as Rook neighbourhoods, combined with temporal neighbourhoods which cover the previous and subsequent time stamps. By using the STN of a pixel, they detect and track moving objects. Reynolds and Madden (1988) used Rook neighbourhood and suggest spatio-temporal autocorrelation metrics could be used to obtain STN, unless there is prior information on the data. In addition, the importance of weighting the neighbours' influence is mentioned. Rook neighbourhoods are also commonly used in digital terrain modelling analysis. Kidner (2003) stresses on the importance of choosing the correct neighbourhood size, which will lead to better interpolation in digital terrain modelling.

Another commonly used spatial neighbourhood strategy for grid data is Queen based spatial neighbourhoods, which are widely used in image/video analysis for various applications. Ng and Solo (1998) conduct research on optical flow estimation and the usage of a function is suggested to detect the optimal neighbourhood of pixels. Zhang et al. (2008) model the dynamically changing background in video sequence data and the STNs are defined by combining the previous time stamp and Queen based neighbourhoods. Similarly, Zhao et al. (2008) define STNs by combining five time stamps within 4×4 pixel spatial neighbourhoods. They model the background of a video on night times, where changes in lighting conditions are taken into account. However, not all of the SNs or STNs have a regular structure, as seen in rook and queen neighbourhoods. For example, Zhao and Billings (2006) use the spatial neighbourhood of a cell to extract cellular automata rules from observed patterns in which spatial neighbourhoods can appear in any arbitrary shape. Since the cells of the cellular automata evolve according to the rules defined for that neighbourhood, it is important to identify a correct neighbourhood.

If the areas do not have a regular structure, then they are called polygons, and if two polygons share a common edge they are thought to be SN. For example, Chawla and Sun (2006) use the SNs to detect spatial outliers by comparing the thematic domain observations of the polygon under investigation and its SN. Janeja and Atluri (2009) propose a three step procedure for defining the SN, which they then refine by using the interrelations between thematic domain observations and their distributions. Billard et al. (2007) use SN to predict an epidemic (i.e. mumps) in time across 12 states of the US.

The last spatial data form for which defining the SN and STN is important is *network data*. Even though network data is not as common as point or areal data in spatial information science, some important real life phenomena are represented in network data form, including road, communication and social networks. Road networks are a prominent type of network which are used in spatial information science. However, social networks, computer networks or power grid networks also have an intrinsic spatial dimension, as each object that is represented by these networks has a location. Networks can be described by two main components: nodes and edges. Nodes are connected via edges, and each edge starts and ends with a node. SNs are defined by graph connectivity, which means, depending on the research objective, either node or edge connectivity. For example, one might define adjacency between two nodes if they are connected via an edge. Similarly, an adjacency between two edges might also be defined if they share a common node. For example, Shekhar et al. (2001) define STN on road network data with the aim of detecting spatio-temporal outliers, and the neighbourhood relies on edges sharing a common node. Wang et al. (2007) take a similar approach by defining a spatial neighbourhood as spatial connectivity of the nodes and temporal neighbourhoods as a time series similarity between the two nodes to cluster traffic sensors. Cheng and Anbaroglu (2010) stress the importance of dynamic flows in an urban road network consisting of links. This means that whatever occurs on a link is expected to impact its adjacent link after a time lapse, due to the flowing nature of traffic, as vehicles move from a link to another link.

It is observed that most of the existing research assumes that STN is simply an extension of SN in time. In other words, SN is assumed to be static in time. However, if spatial phenomenon evolves with time, so should SN. Even though, the importance of SN has been recognised for over 40 years, the research investigating STN is relatively recent (McGuire et al., 2010). Recent findings suggest the importance of considering the dynamic nature of STN (Dubé and Legros, 2013).

APPENDIX B: QUALITY OF LJT ESTIMATES

The number of vehicles correctly captured at both ends of a link that is used to estimate an LJT is referred to as the ‘sample size’. The higher the sample size, the better the quality of the estimated LJT. Qualities of the estimated LJTs vary considerably on the road network and in time.

The aim of this appendix is twofold. The first aim is to demonstrate the importance of the quality of Link Journey Time (LJT) estimations. The second is to demonstrate the effect of congestion on the quality of estimated LJTs.

Importance of the Quality of LJTs

Previous studies agree that the quality (also called the ‘reliability’) of an LJT increases with the sample size (Srinivasan and Jovanis, 1996). A low sample size used to estimate an LJT may therefore pose a serious issue that could lead to a misleading conclusion.

In order to demonstrate the importance of sample size, the following analysis has been conducted. Average delay during an analysis period and is calculated on link a as: $\sum_t \delta_a(t)/|T|$, where $\sum_t \delta_a(t)$ denotes the total excess LJT that occurred within an analysis period on link a, and $|T|$ denotes the total number of LJTs within the analysis period. Because each LJT is estimated at five minute intervals, there are a total of $|T|$= 288 LJTs (i.e. 12 LJTs/hr × 24 hr/day) within a day. This analysis uses a congestion factor of one to determine whether or not an LJT is excessive, as it corresponds the calculation of delay (please refer to section 3.1 for the definitions of congestion factor and excessive LJT).

The average delay is calculated for every day of 2010 for the link illustrated in Figure B-1. The link is located in Central London, connecting Commercial Road with Bermondsey tube station and has a length of approximately 2.9 kilometres.

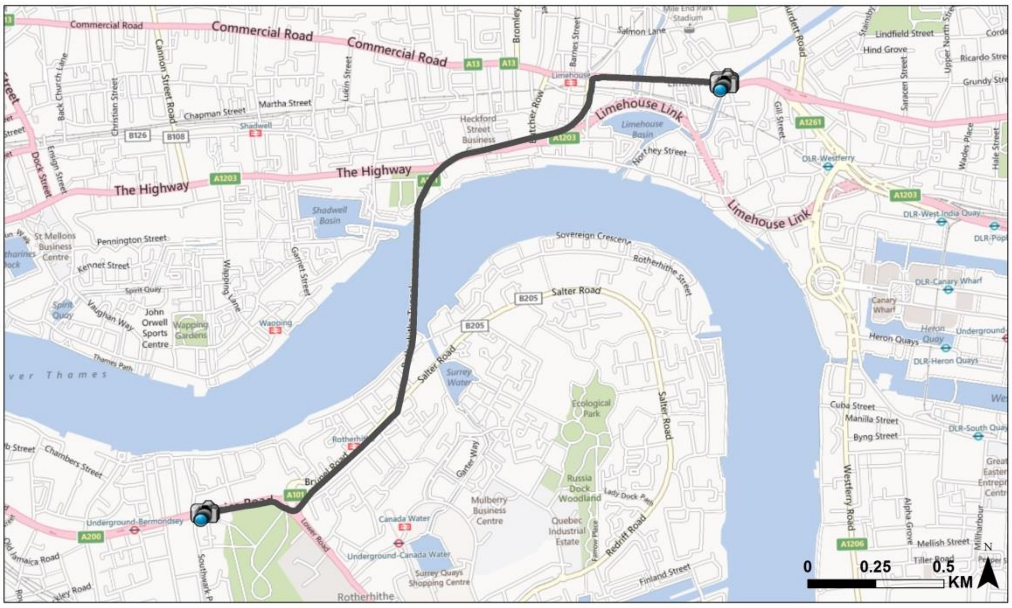


Figure B-1 Limehouse link tunnel (Link 2419)

As a daily analysis is being conducted, it is also important to demonstrate the impact of incidents that last for longer than one day (i.e. long-term incidents). In this way, the days with a high average delay can be associated with those incidents. Seven such incidents occurred on this link during 2010, and the details of these incidents are illustrated in Table B-1.

Table B-1 Start and end times of long-term incidents – Link 2419

Start Time	End Time	Event Category
25/01/2010 22:00 (Mon)	06/02/2010 06:57	Highway Maintenance
08/02/2010 10:00 (Mon)	13/02/2010 00:21	Highway Maintenance
13/02/2010 07:48 (Sat)	31/12/2011	TfL Street works
09/07/2010 21:00 (Fri)	12/07/2010 04:20	TfL Street works
16/07/2010 21:00 (Fri)	19/07/2010 01:02	TfL Street works
31/07/2010 08:00 (Sat)	01/08/2010 15:05	TfL Street works
15/10/2010 21:00 (Fri)	18/10/2010 01:27	TfL Street works

The common feature amongst these incidents is that all of the TfL Street works incidents began on a Friday or Saturday. The planned incidents are scheduled to cover the weekends, so that the impact of the incidents on the road network can be minimised. The average delay for all the days of 2010 are calculated and plotted in Figure B-2. There were no LJT data for the period between January 1st and March 29th due to a system failure in the ANPR cameras and this period is highlighted in blue. All Saturdays are indicated with a point and the days when a long-term incident began are indicated with a vertical red-line.

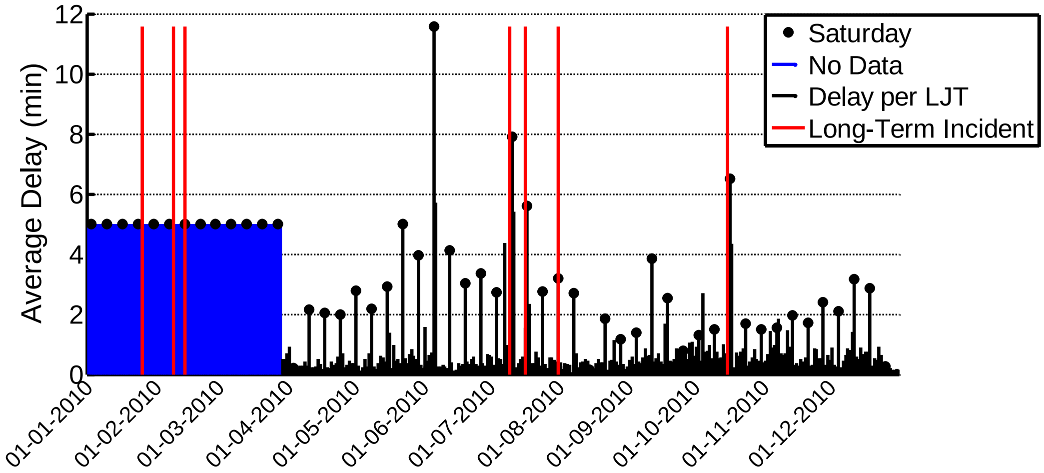


Figure B-2 Average delay over a year and the impact of long-term incidents

Although the linkage between most of the long term incidents and the increased average delay can be observed, there is another more interesting pattern: all Saturdays exhibit much higher average delay than the other days of the week. Therefore, the following question arises: '*Why do the highest average delay happen on Saturdays?*' This question makes even more sense when the nature of the day is considered, since less delay is expected on a Saturday due to a decrease in travel demand.

The analysis required to create Figure B-2 has been conducted with only one difference; instead of calculating the total delay over the whole day, the total average delay is calculated between 07:00 and 19:00. In other words, the analysis period is halved so as to only include the temporal interval which covers morning/inter/afternoon peak periods (TfL, 2010, p.268). The results are shown in Figure B-3.

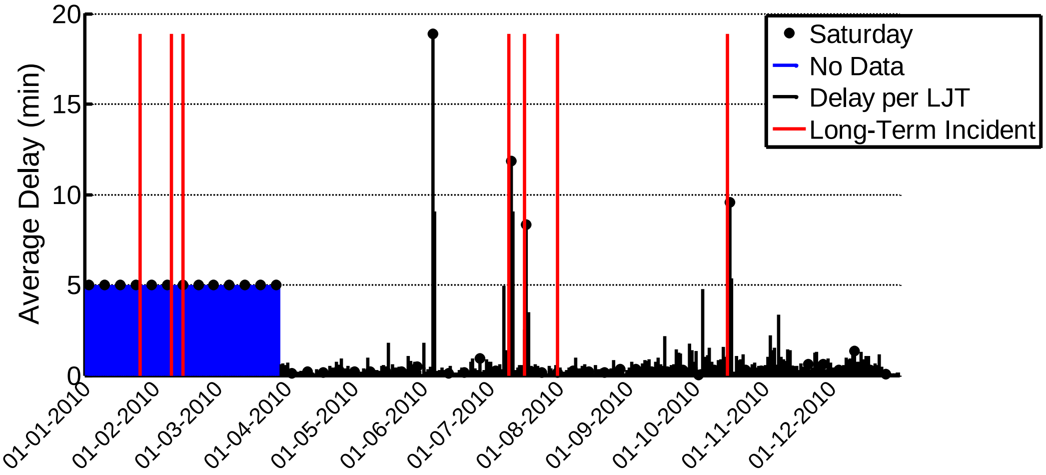


Figure B-3 Average delay over a year and the impact of long-term incidents for the period between 07:00 and 19:00

This figure provides a completely different outcome than that derived from Figure B-2, where the observed increase in the delay per LJT on Saturdays is diminished, which makes the impact of the long-term incidents much more visible. Now the question that must be answered is, '*What caused the average delay to drop radically on Saturdays when LJTs are analysed between 07:00 and 19:00 instead of over an entire day?*'

Sample sizes used to estimate each LJT for the aforementioned link were collected during 2010. The median sample size of all LJTs between 07:00 and 19:00 (day) is compared with the median sample size of all LJTs between 19:00 and 07:00 (night) on different days of week on the link. The calculated median sample sizes for different days of week during day and night periods are illustrated in Table B-2.

Table B-2 Median sample size on link 2419 during day/night periods

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Day	4	4	4	4	4	4	3
Night	1	1	1	1	1	**0**	2

As can be seen from Table B-2, the lowest sample size of the link was observed during Saturdays at night. It is also seen that night periods, regardless of the day of the week, display much lower sample sizes than day periods. As LJTs are estimated based on the journey times of only a few vehicles during night periods, there is an increased likelihood that LJTs are estimated based on implausible journey times. This finding suggests that the estimation of LJTs using a low sample size may lead a researcher to derive incorrect conclusions.

Effect of Congestion on Sample Size

The very occurrence of congestion may actually decrease the sample size used to estimate an LJT. A clear example of this situation is illustrated for link 2346 on 22nd August 2010 in Figure B-4. As can be seen, with the beginning of congestion around 16:00, the sample size decreases sharply from around 50 vehicles per LJT estimation to 3-4 vehicles per LJT estimation. Furthermore, during the congestion, the low sample size remains steady. The main reason for the reduced sample size during a congestion event is that vehicles move closely to one another, which prevents ANPR cameras from capturing the licence plates of those vehicles.

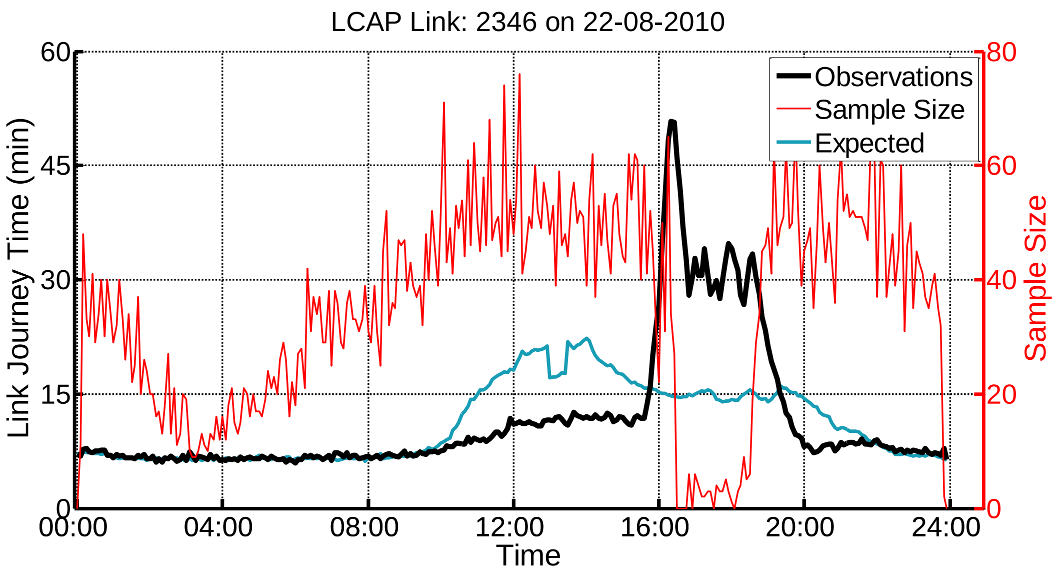


Figure B-4 An example illustrating the reduction in sample size during congestion

In order to prove that the increased LJTs between 16:00 and around 19:00 are not a mere outcome of the reduced sample size, the incident data set has been investigated. Indeed, the increased LJTs can be explained by an *accident* that was reported to start at 16:05 and end at 19:41.

This is an important empirical finding, as the very occurrence of congestion leads vehicles to move much more closely to one another. This decreases the effectiveness of ANPR cameras, as the occlusion between vehicles prevents ANPR cameras from capturing the licence plate of the vehicles. As a result, the number of vehicles that are successfully captured decreases.

APPENDIX C: ADDITIONAL RESULTS

This appendix is dedicated to illustrating the additional results that could not be covered in the main part of the thesis.

The original output which determines the congestion factor in high-confidence episodes is illustrated in Figure C- 1.

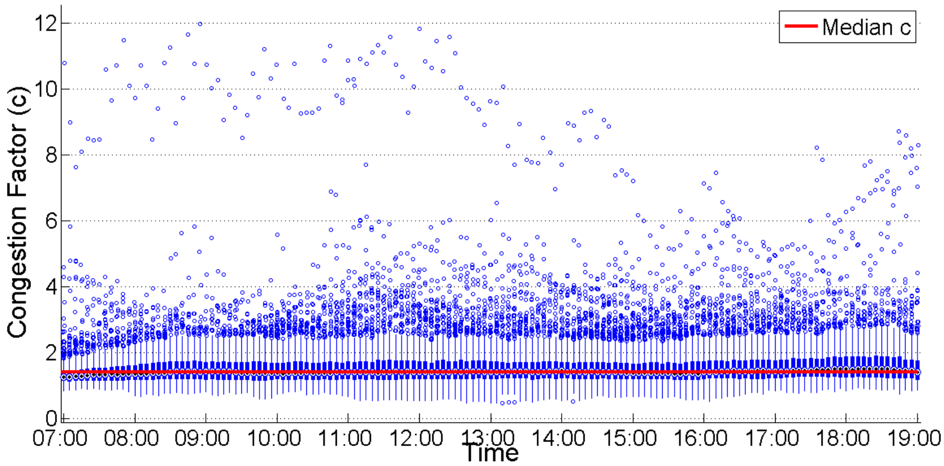


Figure C- 1 Determination of the congestion factor for high-confidence episodes

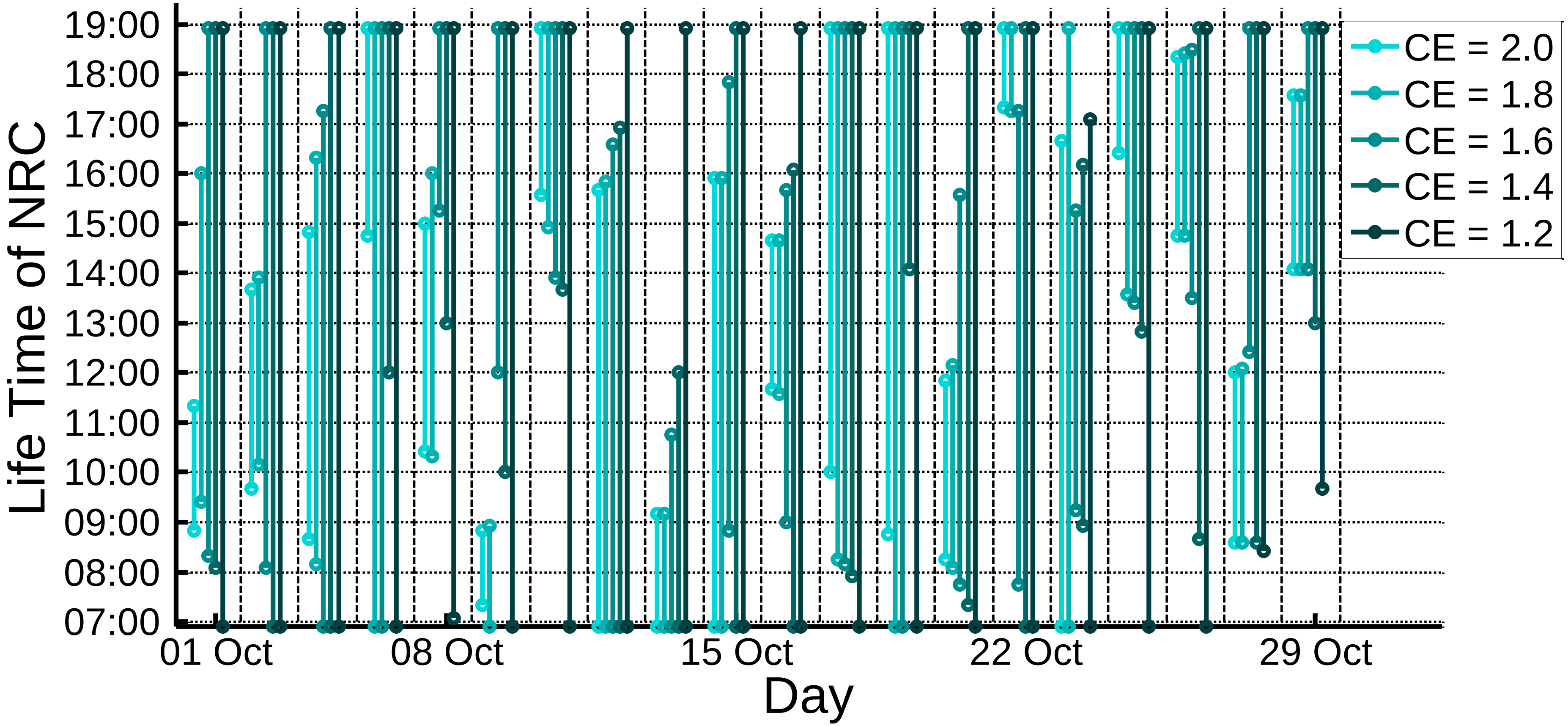
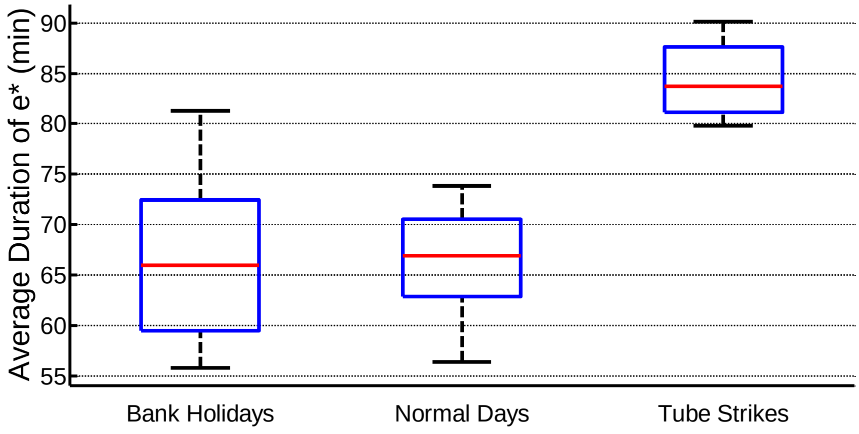
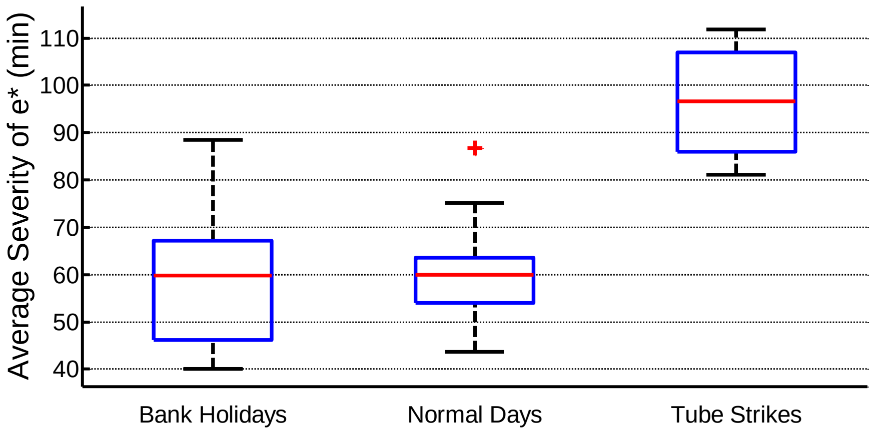


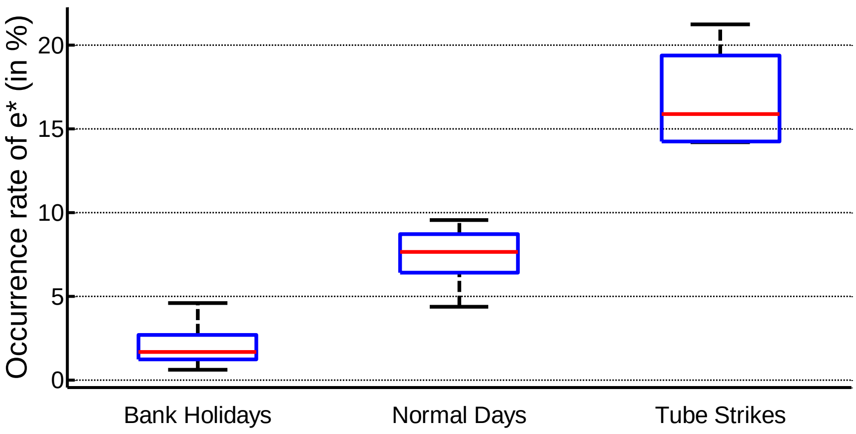
Figure C- 2 Lifetime of the most severe NRC on the weekdays of October 2010 (excluding the tube strike on 4th October)



(a)



(b)



(c)

Figure C- 3 Average duration (a) and severity (b) of high-confidence episodes and the percentage of LJTs that belong to high-confidence episodes, as detected by the expected LJTs used in STSS models (c)

APPENDIX D: OUTCOMES OF THIS THESIS

Book Chapters

- Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G. & Wang, J. (2013) **Spatio-Temporal Data Mining**. In: Fischer, Manfred M.; Nijkamp, Peter. (Eds). *Handbook of Regional Science*. ISBN 978-3-642-23429-3

Peer Reviewed Journal Publications

- Tsapakis, I., Heydecker, B.G., Cheng, T. & Anbaroglu, B. (2013) **How tube strikes affect macroscopic and link travel times in London**. *Transportation Planning and Technology*. 36(1), 109–129.

Peer Reviewed Conference Proceedings

- Anbaroglu, B., Cheng, T. (2011) **Where and when does traffic congestion begin and end? A spatio-temporal clustering approach to detect congestion**. In: *International Symposium on Spatial-Temporal Analysis and Data Mining*. ISPRS. London, UK.

- Cheng, T., Anbaroglu, B. (2010) **Defining Spatio-Temporal Neighbourhood of Network Data**. In: *Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science*. ISPRS. Hong-Kong, China.

- Cheng, T., Anbaroglu, B. (2010) **Spatio-Temporal Clustering of Road Network** Data. In: Fu Wang, Hepu Deng, Yang Gao, & Jingsheng Lei (eds.). *Artificial Intelligence and Computational Intelligence*. Lecture Notes in Computer Science vol. 6319, 116-123. Sanya, China.

- Cheng T., Anbaroglu B. (2009) **Methods on Defining Spatio-Temporal Neighbourhoods**. In 10th International Conference on GeoComputation, Sydney.

- Cheng T., Anbaroglu B. (2009) **Spatio-Temporal Outlier Detection in Environmental Data**. In Spatial and Temporal Reasoning for Ambient Intelligence Systems Workshop, Brest

Honours and Awards

- 2012 - Young European Arena of Research (YEAR 2012), 3rd in the pillar 'Complementarities of Transport Modes'.
- 2010 - International Society of Photogrammetry and Remote Sensing, **Travel Grant**

Professional Activities

- Reviewer: IEEE Transactions on Intelligent Transportation Systems

REFERENCES

Adam, N. R. et al. (2004) 'Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets', in *Proceedings of the 2004 ACM symposium on Applied computing*. 2004 Nicosia, Cyprus: ACM. pp. 576–583.

Adnan, M. et al. (2010) Towards Real-Time Geodemographics: Clustering Algorithm Performance for Large Multidimensional Spatial Databases. *Transactions in GIS*. 14 (3), 283–297.

Ahmed, S. A. & Cook, A. R. (1980) Time Series Models for Freeway Incident Detection. *Transportation Engineering Journal*. 106 (6), 731–745.

Ambroise, C. & Govaert, G. (1998) Convergence of an EM-type algorithm for spatial clustering. *Pattern Recognition Letters*. 19 (10), 919–927.

Anbaroglu, B. & Cheng, T. (2011) 'Where and when does traffic congestion begin and end? A spatio-temporal clustering approach to detect congestion', in *International Symposium on Spatial-Temporal Analysis and Data Mining*. 2011 ISPRS.

Ankerst, M. et al. (1999) OPTICS : ordering points to identify the clustering structure. *SIGMOD Rec.* 28 (2), 49–60.

Anselin, L. (1995) Local Indicators of Spatial Association-LISA. *Geographical Analysis*. 27 (2), 93–115.

Anselin, L. (2002) Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*. 27 (3), 247–267.

Arbelaitz, O. et al. (2013) An extensive comparative study of cluster validity indices. *Pattern Recognition*. 46 (1), 243–256.

Arezoumandi, M. (2011) Estimation of Travel Time Reliability for Freeways Using Mean and Standard Deviation of Travel Time. *Journal of Transportation Systems Engineering and Information Technology*. 11 (6), 74–84.

Ash, A. (1997) 'Incident detection in urban areas controlled by SCOOT', in *IEE Colloquium on Incident Detection and Management (Digest No: 1997/123)*. 1997 pp. 8/1 –8/5.

Assunção, R. et al. (2006) Fast detection of arbitrarily shaped disease clusters. *Statistics in Medicine*. 25 (5), 723–742.

Beevers, S. D. & Carslaw, D. C. (2005) The impact of congestion charging on vehicle emissions in London. *Atmospheric Environment*. 39 (1), 1–5.

Bell, M. G. H. & Cassir, C. (2001) *Reliability of Transport Network*. 1st edition. Research Studies Pre.

Bell, M. G. H. & Iida, Y. (1997) *Transportation Network Analysis*. 1st edition. Wiley.

Berkhin, P. (2006) 'A Survey of Clustering Data Mining Techniques', in *Grouping Multidimensional Data*. pp. 25–71. Available from: http://dx.doi.org/10.1007/3-540-28349-8_2 (Accessed 8 May 2010).

Bertini, R. L. (2006) '*You Are the Traffic Jam: Examination of Congestion Measures*', in 2006 . Available from: http://trid.trb.org/view.aspx?id=776881 (Accessed 29 October 2012).

Besag, J. & Newell, J. (1991) The Detection of Clusters in Rare Diseases. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 154 (1), 143–155.

Billard, L. et al. (2007) 'Modeling Spatial-Temporal Epidemics Using STBL Model', in *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*. 2007 pp. 629–633.

Birant, D. & Kut, A. (2007) ST-DBSCAN : An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* 60 (1), 208–221.

Black, P. W. R. (2010) *Sustainable Transportation: Problems and Solutions*. Guilford Press.

Bloomberg, L. & Dale, J. (2000) Comparison of VISSIM and CORSIM Traffic Simulation Models on a Congested Network. *Transportation Research Record: Journal of the Transportation Research Board*. 1727 (-1), 52–60.

British Library (2013) *The Rise of Cities* . Available from: http://www.bl.uk/learning/histcitizen/georgians/cities/riseofcities.html (Accessed 30 January 2013).

Browne, R. et al. (2005) Comparison and Analysis Tool for Automatic Incident Detection. *Transportation Research Record: Journal of the Transportation Research Board*. 1925 (-1), 58–65.

Brun, M. et al. (2007) Model-based evaluation of clustering validation measures. *Pattern Recognition*. 40 (3), 807–824.

Brunsdon, C. & Charlton, M. (2011) An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning B: Planning and Design*. 38 (2), 216–230.

Buch, N. et al. (2011) A Review of Computer Vision Techniques for the Analysis of Urban Traffic. *IEEE Transactions on Intelligent Transportation Systems*. 12 (3), 920–939.

Caldas de Castro, M. & Singer, B. H. (2006) Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association. *Geographical Analysis*. 38 (2), 180–208.

Celik, M. et al. (2006) 'Sustained Emerging Spatio-Temporal Co-occurrence Pattern Mining: A Summary of Results', in *Tools with Artificial Intelligence, 2006. ICTAI '06. 18th IEEE International Conference on*. 2006 pp. 106–115.

Chang, S.-L. et al. (2004) Automatic license plate recognition. *Intelligent Transportation Systems, IEEE Transactions on*. 5 (1), 42 – 53.

Chawla, S. & Sun, P. (2006) SLOM : a new measure for local spatial outliers. *Knowledge and Information Systems*. 9 (4), 412–429.

Chen, J. et al. (2008) 'Clustering Moving Objects in Spatial Networks', in *Advances in Databases: Concepts, Systems and Applications*. pp. 611–623. . Available from: http://dx.doi.org/10.1007/978-3-540-71703-4_52 (Accessed 6 October 2009).

Cheng, T. et al. (2012) Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*. 14 (4), 389–413.

Cheng, T. & Anbaroglu, B. (2010) 'Spatio-Temporal Clustering of Road Network Data', in Fu Wang et al. (eds.) *Artificial Intelligence and Computational Intelligence*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. pp. 116–123. Available from: http://dx.doi.org/10.1007/978-3-642-16530-6_15.

Cheng, T. & Li, Z. (2006) A Multiscale Approach for Spatio-Temporal Outlier Detection. *Transactions in GIS*. 10 (2), 253–263.

Cheu, R. L. & Ritchie, S. G. (1995) Automated detection of lane-blocking freeway incidents using artificial neural networks. *Transportation Research Part C: Emerging Technologies*. 3 (6), 371–388.

Choynowski, M. (1959) Maps Based on Probabilities. *Journal of the American Statistical Association*. 54 (286), 385–388.

Chu, L. et al. (2004) Using Microscopic Simulation to Evaluate Potential Intelligent Transportation System Strategies Under Nonrecurrent Congestion. *Transportation Research Record: Journal of the Transportation Research Board*. 1886, 76–84.

Cliff, A. D. & Ord, J. K. (1981) *Spatial Processes: Models and Applications*. Taylor & Francis.

Cliff, A. D. & Ord, J. K. (1968) *The problem of spatial autocorrelation*.

Cullip, M. & Hall, F. (1997) Incident Detection on an Arterial Roadway. *Transportation Research Record*. 1603 (1), 112–118.

D'Angelo, M. et al. (1999) Travel-Time Prediction for Freeway Corridors. *Transportation Research Record: Journal of the Transportation Research Board*. 1676 (-1), 184–191.

Dandy, G. & McBean, E. (1984) Variability of Individual Travel Time Components. *Journal of Transportation Engineering*. 110 (3), 340–356.

Dempster, A. P. et al. (1977) Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*. 39 (1), 1–38.

Deng, M. et al. (2011) A general method of spatio-temporal clustering analysis. *Science China Information Sciences*.

DfT (1999) *Design Manual For Roads And Bridges: Traffic Capacity of Urban Roads*. Available from:

http://www.dft.gov.uk/ha/standards/dmrb/vol5/section1/ta7999.pdf (Accessed 29 December 2012).

DfT (2012) *Managing, improving and investing in the road network* . Available from: https://www.gov.uk/government/policies/managing-improving-and-investing-in-the-road-network (Accessed 24 December 2012).

Dion, F. & Rakha, H. (2006) Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B: Methodological*. 40 (9), 745–766.

Dowling, R. et al. (2004) Methodology for Measuring Recurrent and Nonrecurrent Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board*. 1867 (-1), 60–68.

Dubé, J. & Legros, D. (2013) A spatio-temporal measure of spatial dependence: An example using real estate data*. *Papers in Regional Science*. 92 (1), 19–30.

Dubes, R. C. (1993) 'Cluster Analysis and Related Issues', in C. H. Chen et al. (eds.) *Handbook of Pattern Recognition and Computer Vision*. River Edge, NJ, USA: World Scientific Publishing Co., Inc. pp. 3–32. . Available from: http://dl.acm.org/citation.cfm?id=178866.178874 (Accessed 8 June 2012).

Dubes, R. & Jain, A. K. (1976) Clustering techniques: The user's dilemma. *Pattern Recognition*. 8 (4), 247–260.

Duczmal, L. et al. (2007) A genetic algorithm for irregularly shaped spatial scan statistics. *Computational Statistics & Data Analysis*. 52 (1), 43–52.

Eddington, R. (2006) *The Eddington Transport Study: The Case for Action*. . Available from: http://www.thepep.org/ClearingHouse/docfiles/Eddington.Transport.Study%20-%20Rod.pdf (Accessed 10 December 2012).

Ben-Elia, E. & Ettema, D. (2009) Carrots versus sticks: Rewarding commuters for avoiding the rush-hour - a study of willingness to participate. *Transport Policy*. 16 (2), 68–76.

Emmerink, R. H. M. et al. (1995) The potential of information provision in a simulated road transport network with non-recurrent congestion. *Transportation Research Part C: Emerging Technologies*. 3 (5), 293–309.

Ester, M. et al. (1996) *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. 2nd226–231.

Fei, X. et al. (2011) A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*. 19 (6), 1306–1318.

FHWA (2012a) *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation: Executive Summary* . Available from: http://www.ops.fhwa.dot.gov/congestion_report/executive_summary.htm (Accessed 29 October 2012).

FHWA (2010) *Traffic Incident Management Handbook Update*. . Available from: http://ops.fhwa.dot.gov/eto_tim_pse/publications/timhandbook/index.htm (Accessed 13 November 2012). Available from: http://ops.fhwa.dot.gov/eto_tim_pse/publications/timhandbook/index.htm (Accessed 13 November 2012).

FHWA (2012b) *Travel Time Reliability: Making It There On Time, All The Time* . Available from: http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm#WhyisTTR (Accessed 30 August 2012).

Geary, R. C. (1954) The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*. 5 (3), 115–146.

Bar-Gera, H. (2007) Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*. 15 (6), 380–391.

Geroliminis, N. & Daganzo, C. F. (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*. 42 (9), 759–770.

Getis, A. (2009) Spatial Weights Matrices. *Geographical Analysis*. 41 (4), 404–410.

Getis, A. & Aldstadt, J. (2004) Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis*. 36 (2), 90–104.

Goodwin, P. (2004) *The economic costs of road traffic congestion* . Available from: http://eprints.ucl.ac.uk/1259/ (Accessed 3 February 2011).

Goovaerts, P. & Jacquez, G. M. (2005) Detection of temporal changes in the spatial distribution of cancer rates using local Moran’s I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems*. 7 (1), 137–159.

Griffith, D. A. (2006) Hidden negative spatial autocorrelation. *Journal of Geographical Systems*. 8 (4), 335–355.

Guralnik, V. & Srivastava, J. (1999) 'Event detection from time series data', in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD ’99. . 1999 New York, NY, USA: ACM. pp. 33–42. . Available from: http://doi.acm.org/10.1145/312129.312190 (Accessed 19 March 2013).

Halkidi, M. et al. (2001) On Clustering Validation Techniques. *Journal of Intelligent Information Systems*. 17 (2), 107–145.

Hall, F. L. (2001) 'Traffic Stream Characteristics', in *Traffic Flow Theory: A State-of-the-art Report*. FHWA/TRB/ORNL.

Hallenbeck, M. E. et al. (2003a) *Measurement of recurring versus non-recurring congestion - Final Report*. . Available from: http://depts.washington.edu/trac/bulkdisk/pdf/568.2.pdf (Accessed 29 December 2012).

Hallenbeck, M. E. et al. (2003b) *Measurement of recurring versus non-recurring congestion - Technical Report*. . Available from: http://depts.washington.edu/trac/bulkdisk/pdf/568.1.pdf (Accessed 8 January 2013).

Han, J. & Kamber, M. (2006) *Data Mining: Concepts and Techniques*. 3rd edition. Morgan Kaufmann.

Han, L. D. & May, A. D. (1989) Automatic Detection of Traffic Operational Problems on Urban Arterials. *RESEARCH REPORT INSTITUTE OF TRANSPORTATION*. (8), Available from: http://trid.trb.org/view.aspx?id=318298 (Accessed 19 January 2012).

Hawas, Y. E. (2007) A fuzzy-based system for incident detection in urban street networks. *Transportation Research Part C: Emerging Technologies*. 15 (2), 69–95.

Haworth, J. & Cheng, T. (2012) Non-parametric regression for space-time forecasting under missing data. *Computers, Environment and Urban Systems*. 36 (6), 538–550.

Helbing, D. et al. (2007) Self-Organized Network Flows. *arXiv:physics/0702173*. . Available from: http://arxiv.org/abs/physics/0702173 (Accessed 26 September 2011).

Helbing, D. (2001) Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*. 73 (4), 1067–1141.

Hollander, Y. & Liu, R. (2008) Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies*. 16 (2), 212–231.

Huang, L. et al. (2009) Identifying Clusters of Active Transportation Using Spatial Scan Statistics. *American Journal of Preventive Medicine*. 37 (2), 157–166.

Huang, Y. et al. (2006) Mining Co-Location Patterns with Rare Events from Spatial Data Sets. *GeoInformatica*. 10 (3), 239–260.

Iyengar, V. S. (2004) 'On detecting space-time clusters', in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. . 2004 Seattle, WA, USA: ACM. pp. 587–592. . Available from: http://portal.acm.org/citation.cfm?id=1014124 (Accessed 26 October 2009).

Jain, A. K. et al. (1999) Data clustering: a review. *ACM Comput. Surv*. 31 (3), 264–323.

Jain, V. et al. (2012) 'Road traffic congestion in the developing world', in *Proceedings of the 2nd ACM Symposium on Computing for Development*. ACM DEV ’12. New York, NY, USA: ACM. pp. 11:1–11:10. Available from: http://doi.acm.org/10.1145/2160601.2160616 (Accessed 16 June 2013).

Janeja, V. et al. (2010) Spatial neighborhood based anomaly detection in sensor datasets. *Data Mining and Knowledge Discovery*. 20 (2), 221–258.

Janeja, V. P. & Atluri, V. (2009) Spatial outlier detection in heterogeneous neighborhoods. *Intelligent Data Analysis*. 13 (1), 85–107.

Jeong, Y.-S. et al. (2011) A wavelet-based freeway incident detection algorithm with adapting threshold parameters. *Transportation Research Part C: Emerging Technologies*. 19 (1), 1–19.

Karim, A. & Adeli, H. (2002) Comparison of Fuzzy-Wavelet Radial Basis Function Neural Network Freeway Incident Detection Model with California Algorithm. *Journal of Transportation Engineering*. 128 (1), 21–30.

Kerner, B. S. (2004) *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. 1st edition. Springer.

Kerner, B. S. & Konhäuser, P. (1994) Structure and parameters of clusters in traffic flow. *Physical Review E*. 50 (1), 54–83.

Kidner, D. B. (2003) Higher-order interpolation of regular grid digital elevation models. *International Journal of Remote Sensing*. 24 (14), 2981.

Kisilevich, S. et al. (2010) 'Spatio-temporal clustering', in Oded Maimon & Lior Rokach (eds.) *Handbook of Data Mining and Knowledge Discovery*. Springer US. pp. 855–874. . Available from: http://link.springer.com/chapter/10.1007/978-0-387-09823-4_44 (Accessed 30 November 2012).

Klein, L. A. (2001) *Sensor Technologies and Data Requirements for ITS*. Artech House.

Knox, E. G. (1964) The Detection of Space-Time Interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 13 (1), 25–30.

Kreher, D. L. & Stinson, D. R. (1999) *Pseudocode : A LATEX Style File for Displaying Algorithms*. . Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.8162 (Accessed 14 May 2013).

Kremer, H. et al. (2011) 'An effective evaluation measure for clustering on evolving data streams', in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD '11. . 2011 New York, NY, USA: ACM. pp. 868–876.

Kulldorff, M. et al. (2009) A Scan Statistic for Continuous Data Based on the Normal Probability Model. *International Journal of Health Geographics*. 858.

Kulldorff, M. et al. (2005) A Space - Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med*. 2 (3), e59.

Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics - Theory and Methods*. 26 (6), 1481.

Kulldorff, M. et al. (2006) An elliptic spatial scan statistic. *Statistics in medicine*. 25 (22), 3929–3943.

Kulldorff, M. et al. (1998) Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*. 88 (9), 1377–1380.

Kulldorff, M. (2001) Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 164 (1), 61–72.

Kulldorff, M. (2010) *SaTScan User Guide*. . Available from: http://www.satscan.org/cgi-bin/satscan/register.pl/Current%20Version:%20SaTScan%20v9.1.1%20released%20March%209%202011.?todo=process_userguide_download (Accessed 13 December 2012).

Kulldorff, M. (2006) Tests of Spatial Randomness Adjusted for an Inhomogeneity. *Journal of the American Statistical Association*. 101 (475), 1289–1305.

Kulldorff, M. & Nagarwalla, N. (1995) Spatial disease clusters: detection and inference. *Statistics in Medicine*. 14 (8), 799–810.

Kwon, J. et al. (2006) Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand. *Transportation Research Record*. 1959 (1), 84–91.

Lee, W. H. et al. (2011) Discovering Traffic Bottlenecks in an Urban Network by Spatiotemporal Data Mining on Location-Based Services. *IEEE Transactions on Intelligent Transportation Systems*. 12 (4), 1047–1056.

Leskinen, P. & Kangas, J. (2005) Rank Reversals in Multi-Criteria Decision Analysis with Statistical Modelling of Ratio-Scale Pairwise Comparisons. *The Journal of the Operational Research Society*. 56 (7), 855–861.

Li, X. et al. (2007) 'Traffic Density-Based Discovery of Hot Routes in Road Networks', in Dimitris Papadias et al. (eds.) *Advances in Spatial and Temporal Databases*. Lecture Notes in Computer Science. Springer Berlin Heidelberg. pp. 441–459. . Available from: http://link.springer.com/chapter/10.1007/978-3-540-73540-3_25 (Accessed 15 January 2013).

Lim, T.-S. & Loh, W.-Y. (1996) A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*. 22 (3), 287–301.

Lin, C.-R. et al. (2005) Dual clustering: integrating data clustering over optimization and constraint domains. *Knowledge and Data Engineering, IEEE Transactions on*. 17 (5), 628–637.

Liu, H. Q. et al. (2010) Non-local spatial spectral clustering for image segmentation. *Neurocomputing*. 74 (1-3), 461–471.

Liu, Y. et al. (2010) 'Understanding of Internal Clustering Validation Measures', in *2010 IEEE 10th International Conference on Data Mining (ICDM)*. . 2010 pp. 911–916.

Lomax, T. et al. (2003) *Selecting Travel Time Reliability Measures*. . Available from: http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/TTI-2003-3.pdf (Accessed 13 November 2012).

Long, J. et al. (2012) Control Strategies for Dispersing Incident-Based Traffic Jams in Two-Way Grid Networks. *IEEE Transactions on Intelligent Transportation Systems*. 13 (2), 469–481.

Lu, C.-T. et al. (2003) 'Detecting spatial outliers with multiple attributes', in *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*. 2003 pp. 122–128.

Luk, J. Y. K. et al. (2001) Characterization of incidents on an urban arterial road. *Journal of Advanced Transportation*. 35 (1), 67–92.

Luxburg, U. von et al. (2012) Clustering : Science or Art? *Journal of Machine Learning Research - Proceedings Track*. 2765–80.

Maciejewski, R. et al. (2010) A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*. 16 (2), 205–220.

Macqueen, J. (1967) 'Some methods of classification and analysis of multivariate observations', in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967 pp. 281–297.

Martin, P. T. et al. (2001) *Incident Detection Algorithm Evaluation*. . Available from: http://www.mountain-plains.org/pubs/pdf/MPC01-122.pdf (Accessed 26 December 2011).

Marwan, N. et al. (2007) Recurrence plots for the analysis of complex systems. *Physics Reports*. 438 (5-6), 237–329.

Matlab (2013) *Quantiles and Percentiles - MATLAB* . Available from: http://www.mathworks.co.uk/help/stats/quantiles-and-percentiles.html (Accessed 5 August 2013).

McGuire, M. et al. (2010) 'Spatiotemporal Neighborhood Discovery for Sensor Data', in Mohamed Gaber et al. (eds.) *Knowledge Discovery from Sensor Data*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. pp. 203–225. Available from: http://www.springerlink.com/content/mup2l52w01u74631/abstract/ (Accessed 19 July 2012).

Mela, K. et al. (2012) Comparative study of multiple criteria decision making methods for building design. *Advanced Engineering Informatics*. 26 (4), 716–726.

Milligan, G. W. (1981) A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*. 46 (2), 187–199.

Min-qi, Z. et al. (2008) 'An Algorithm for Spatial Outlier Detection Based on Delaunay Triangulation', in *Computational Intelligence and Security, 2008. CIS '08. International Conference on*. 2008 pp. 102–107.

Moran, P. A. P. (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*. 10 (2), 243–251.

Nakaya, T. & Yano, K. (2010) Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Transactions in GIS*. 14 (3), 223–239.

National Research Council (1994) *Curbing Gridlock: Peak-Period Fees to Relieve Traffic Congestion -- Special Report 242*. Available from: http://www.nap.edu/openbook.php?isbn=0309055040 (Accessed 3 January 2013).

Neill, D. B. (2009) An empirical comparison of spatial scan statistics for outbreak detection. *International Journal of Health Geographics*. 820.

Neill, D. B. (2006) *Detection of Spatial and Spatio-Temporal Clusters*. PhD thesis. Carnegie Mellon University.

Neill, D. B. (2008) Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*. 25 (3), 498–517.

Neill, D. B. & Moore, A. W. (2004) 'Rapid detection of significant spatial clusters', in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD ’04. New York, NY, USA: ACM. pp. 256–265.

Ng, A. Y. et al. (2001) 'On Spectral Clustering: Analysis and an algorithm', in *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*. 2001 MIT Press. pp. 849–856.

Ng, L. & Solo, V. (1998) 'Choosing the optimal neighbourhood size in optical flow problems with errors-in-variables modelling', in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. 1998 pp. 186–190 vol.2.

Noland, R. B. & Polak, J. W. (2002) Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*. 22 (1), 39–54.

OECD (2007) *Managing Urban Traffic Congestion*. Available from: http://dx.doi.org/10.1787/9789282101506-en (Accessed 30 November 2012).

Openshaw, S. et al. (1987) A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. *International journal of geographical information systems*. 1 (4), 335–358.

Openshaw, S. et al. (1988) Investigation of Leukaemia Clusters by use of a Geographical Analysis Machine. *The Lancet*. 331 (8580), 272–273.

Ordnance Survey (2013) *OS MasterMap Integrated Transport Network (ITN) Layer - intelligent geographical data*. Available from: http://www.ordnancesurvey.co.uk/oswebsite/products/os-mastermap/itn-layer/index.html (Accessed 28 January 2013).

Ozbay, K. & Kachroo, P. (1999) Incident management in intelligent transportation systems. *Faculty Publications (ECE)*. Available from: http://digitalcommons.library.unlv.edu/ece_fac_articles/103.

Parkany, E. & Xie, C. (2005) *A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't*. . Available from: http://trid.trb.org/view.aspx?id=754914 (Accessed 13 November 2012).

Patil, G. P. & Taillie, C. (2004) Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*. 11 (2), 183–197.

Payne, H. J. & Tignor, S. C. (1978) Freeway Incident-Detection Algorithms Based on Decision Trees with States. *Transportation Research Record*. (682), . . Available from: http://trid.trb.org/view.aspx?id=91809 (Accessed 26 December 2011).

Persaud, B. N. & Hall, F. L. (1989) Catastrophe theory and patterns in 30-second freeway traffic data - Implications for incident detection. *Transportation Research Part A: General*. 23 (2), 103–113.

Petty, K. et al. (2000) *A New Methodology for Evaluating Incident Detection Algorithms*. . Available from: http://escholarship.org/uc/item/35j0g3tm (Accessed 21 March 2013).

Pfeiffer, D. U. et al. (2008) *Spatial Analysis in Epidemiology*. 1st edition. Oxford University Press, USA.

Pline, J. L. (1992) *Traffic engineering handbook*. Englewood Cliffs, N.J.: Prentice-Hall.

Polus, A. (1979) A study of travel time and reliability on arterial routes. *Transportation*. 8 (2), 141–151.

Pu, W. (2011) Analytic Relationships Between Travel Time Reliability Measures. *Transportation Research Record: Journal of the Transportation Research Board*. 2254, 122–130.

Reynolds, K. M. & Madden, L. V. (1988) Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology*. 78 (2), 240–246.

Robinson, S. & Polak, J. (2006) Overtaking Rule Method for the Cleaning of Matched License-Plate Data. *Journal of Transportation Engineering*. 132 (8), 609–617.

Rogerson, P. & Yamada, I. (2009) *Statistical Detection and Surveillance of Geographic Clusters*. CRC Press.

Roy, P. & Abdulhai, B. (2003) GAID : Genetic Adaptive Incident Detection for Freeways. *Transportation Research Record: Journal of the Transportation Research Board*. 1856, 96–105.

Sansò, B. & Milot, L. (1999) Performability of a Congested Urban Transportation Network When Accident Information is Available. *Transportation Science*. 33 (1), 68–79.

SaTScan (2010) *SaTScan - Bibliography* . Available from: http://www.satscan.org/references.html (Accessed 3 December 2012).

Sen, R. et al. (2012) 'Kyun queue : a sensor network system to monitor road traffic queues', in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. SenSys '12. . 2012 New York, NY, USA: ACM. pp. 127–140.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Longman Higher Education.

Shehata, M. S. et al. (2008) Video-Based Automatic Incident Detection for Smart Roads: The Outdoor Environmental Challenges Regarding False Alarms. *IEEE Transactions on Intelligent Transportation Systems*. 9 (2), 349 –360.

Shekhar, S. et al. (2001) 'Detecting graph-based spatial outliers: algorithms and applications (a summary of results)', in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 2001 San Francisco, California: ACM. pp. 371–376.

Shekhar, S. et al. (2011) Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 1 (3), 193–214.

Šingliar, T. & Hauskrecht, M. (2009) Learning to detect incidents from noisily labeled data. *Machine Learning*. 79 (3), 335–354.

Sjöberg, J. et al. (1995) Nonlinear black-box modeling in system identification: a unified overview. *Automatica*. 31 (12), 1691–1724.

Skabardonis, A. et al. (2003) Measuring Recurrent and Nonrecurrent Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board*. 1856 (-1), 118–124.

Snow, J. (1854) *On the mode of communication of cholera*.

Srinivasan, K. & Jovanis, P. (1996) Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network. *Transportation Research Record: Journal of the Transportation Research Board*. 1537, 15–22.

Stephanedes, Y. J. & Chassiakos, A. P. (1993) Application of Filtering Techniques for Incident Detection. *Journal of Transportation Engineering*. 119 (1), 13–26.

Sugiyama, Y. et al. (2008) Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics*. 10 (3), 033001.

Sun, S. et al. (2006) A Bayesian Network Approach to Traffic Flow Forecasting. *Intelligent Transportation Systems, IEEE Transactions on*. 7 (1), 124–132.

Susilawati, S. et al. (2011) Distributions of travel time variability on urban roads. *Journal of Advanced Transportation*. Available from: http://onlinelibrary.wiley.com/doi/10.1002/atr.192/abstract (Accessed 26 January 2012).

Tango, T. & Takahashi, K. (2005) A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*. 4 (1), 11.

TfL (2013) *Key facts*. Available from: http://www.tfl.gov.uk/corporate/modesoftransport/londonunderground/1608.aspx (Accessed 30 January 2013).

TfL (2012a) *London Streets*. Available from: http://www.tfl.gov.uk/assets/downloads/corporate/london-streets-factsheet.pdf (Accessed 30 January 2013).

TfL (2012b) *Transport for London - Cameras* . Available from: http://www.tfl.gov.uk/termsandconditions/22246.aspx#page-link-where-does-tfl-operate-cctv-or-anpr-systems- (Accessed 30 November 2012).

TfL (2010) *Travel in London Report 3*. Available from: http://www.tfl.gov.uk/assets/downloads/travel-in-london-report-3.pdf (Accessed 25 September 2012).

Theodoridis, S. & Koutroumbas, K. (2006) *Pattern recognition*. Academic Press.

Thomas, T. & van Berkum, E. C. (2009) Detection of incidents and events in urban networks. *IET Intelligent Transport Systems*. 3 (2), 198 –205.

Transport Canada (2007) *The Cost of Urban Congestion in Canada*. . Available from: http://www.trb.org/Main/Blurbs/157230.aspx (Accessed 13 July 2013).

Treiber, M. et al. (2000) Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*. 62 (2), 1805–1824.

Triantaphyllou, E. & Mann, S. H. (1989) An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox. *Decision Support Systems*. 5 (3), 303–312.

Tsapakis, I. et al. (2013) How tube strikes affect macroscopic and link travel times in London. *Transportation Planning and Technology*. 36 (1), 109–129.

Tuia, D. et al. (2008) Scan statistics analysis of forest fire clusters. *Communications in Nonlinear Science and Numerical Simulation*. 13 (8), 1689–1694.

Tukey, J. W. (1977) *Exploratory Data Analysis*. 1st edition. Pearson.

Tung, A. K. H. et al. (2001) 'Spatial clustering in the presence of obstacles', in *Data Engineering, 2001. Proceedings. 17th International Conference on*. . 2001 pp. 359 –367.

Türkeş, M. & Tatlı, H. (2011) Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929-2007. *International Journal of Climatology*. 31 (14), 2055–2067.

Turner, S. M. et al. (1998) *Travel Time Data Collection Handbook*. . Available from: http://trid.trb.org/view.aspx?id=497690 (Accessed 12 November 2012).

USDOT (2006) *National Strategy to Reduce Congestion on America's Transportation Network.* Available from: http://www.trb.org/Main/Blurbs/157396.aspx (Accessed 24 December 2012).

Varaiya, P. (2007) Finding and Analyzing True Effect of Non-Recurrent Congestion on Mobility and Safety. *PATH Research Report*. Available from: http://trid.trb.org/view.aspx?id=836930 (Accessed 12 August 2013).

Varaiya, P. (2005) What We've Learned About Highway Congestion. *Access*. (27) . Available from: http://trid.trb.org/view.aspx?id=771535 (Accessed 26 December 2012).

Vendramin, L. et al. (2010) Relative clustering validity criteria: A comparative overview. *Statistical Analysis and Data Mining*. 3 (4), 209–235.

Wagstaff, K. et al. (2001) 'Constrained K-means Clustering with Background Knowledge', in *Proceedings of the Eighteenth International Conference on Machine Learning*. ICML '01. 2001 San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. pp. 577–584.

Wang, Y. et al. (2007) 'SPANBRE : An Efficient Hierarchical Clustering Algorithm for Spatial Data with Neighborhood Relations', in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*. 2007 pp. 665–669.

Wardrop, J. (1952) Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*. 1 (36), .

Weaver, W. (1948) Science and Complexity. *American Scientist*. 36 (4), 536–544.

Wei, L.-Y. & Peng, W.-C. (2009) Clustering Data Streams in Optimization and Geography Domains, in *Advances in Knowledge Discovery and Data Mining*. pp. 997–1005.

Weil, R. et al. (1998) Traffic incident detection: Sensors and algorithms. *Mathematical and Computer Modelling*. 27 (9-11), 257–291.

Wheeler, D. C. (2007) A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996 – 2003. *International Journal of Health Geographics*. 6 (1), 13.

Williams, B. M. & Guin, A. (2007) Traffic Management Center Use of Incident Detection Algorithms: Findings of a Nationwide Survey. *Intelligent Transportation Systems, IEEE Transactions on*. 8 (2), 351–358.

Wright, C. & Roberg-Orenstein, P. (1999) Simple Models for Traffic Jams and Congestion Control, in *Proceedings of the ICE - Transport*. August 1999 pp. 123–130. . Available from: http://trid.trb.org/view.aspx?id=512251 (Accessed 8 July 2013).

Wu, C.-H. et al. (2004) Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*. 5 (4), 276– 281.

Xu, R. & Wunsch, D. (2005) Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*. 16 (3), 645–678.

Yamada, I. et al. (2009) GeoSurveillance : a GIS-based system for the detection and monitoring of spatial clusters. *Journal of Geographical Systems*. 11 (2), 155–173.

Yang, Q. & Wu, X. (2006) 10 Challenging Problems in Data Mining Research. *International Journal of Information Technology & Decision Making*. 05 (04), 597–604.

Yang, Z. et al. (2009) 'Support Vector Machines for Incident Detection in Urban Signalized Arterial Street Networks', in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA '09. International Conference on*. . 2009 pp. 611–616.

Yeon, J. et al. (2009) Differences in Freeway Capacity by Day of the Week, Time of Day, and Segment Type. *Journal of Transportation Engineering*. 135 (7), 416–426.

Yin, Z. & Collins, R. (2007) Belief Propagation in a 3D Spatio-temporal MRF for Moving Object Detection, in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*. 2007 pp. 1–8.

Ying, X. et al. (2009) 'Cluster-Based Congestion Outlier Detection Method on Trajectory Data', in *Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09*. pp. 243 –247.

Yiu, M. L. & Mamoulis, N. (2004) 'Clustering objects on a spatial network', in *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*. Paris, France: ACM. pp. 443–454. .

Yuan, F. & Cheu, R. L. (2003) Incident detection using support vector machines. *Transportation Research Part C: Emerging Technologies*. 11 (3-4), 309–328.

Zanakis, S. H. et al. (1998) Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research*. 107 (3), 507–529.

Zhang, J. et al. (2011) Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*. 12 (4), 1624 –1639.

Zhang, S. et al. (2008) 'Dynamic background modeling and subtraction using spatio-temporal local binary patterns', in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. pp. 1556–1559.

Zhang, Z. et al. (2010) Spatial Scan Statistics Adjusted for Multiple Clusters. *Journal of Probability and Statistics*. 20101–11.

Zhao, Y. et al. (2008) 'Spatio-temporal patches for night background modeling by subspace learning', in *19th International Conference on Pattern Recognition*. pp. 1–4.

Zhao, Y. & Billings, S. A. (2006) Neighborhood detection using mutual information for the identification of cellular automata. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*. 36 (2), 473–479.