Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks

Berk Anbaroglu [a,*], Benjamin Heydecker [b,1], Tao Cheng [c,2]

[a] Department of Geomatics Engineering, Hacettepe University, 06800 Beytepe, Ankara, Turkey
[b] Centre for Transport Studies, 210 Chadwick Building, University College London, Gower Street, London WC1E 6BT, United Kingdom
[c] SpaceTimeLab, Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom

A B S T R A C T

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 urban road networks. In this paper we propose 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 as those LJTs that are greater than a threshold. The threshold is calculated by multiplying the expected LJTs with a congestion factor. To evaluate the effectiveness of the proposed NRC detection method, we propose two novel criteria. 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 associated with incidents. The proposed NRC detection methodology is tested for London's urban road network. The optimum value of the congestion factor is determined by sensitivity analysis by using a Weighted Product Model (WPM). It is found out those LJTs that are at least 40% higher than their expected values should belong to an NRC; as such NRCs are found to maintain the best balance between the proposed evaluation criteria.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Most of the traffic operation centres and researchers classify traffic congestion into two types: Recurrent Congestion (RC) and Non-Recurrent Congestion (NRC) (Dowling et al., 2004; OECD, 2007, p. 14; Ozbay and Kachroo, 1999, p. 1; 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. It can also

---

* Corresponding author. Tel.: +90 (312) 297 69 90.
    E-mail addresses: banbar@hacettepe.edu.tr (B. Anbaroglu), b.heydecker@ucl.ac.uk (B. Heydecker), tao.cheng@ucl.ac.uk (T. Cheng).
  [1] Tel.: +44 (0)20 7679 1553.
  [2] Tel.: +44 (0)20 7679 2738.

occur due to planned engineering works, special events (e.g. football matches or concerts) or inclement weather (FHWA, 2012a; Kwon et al., 2006).

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 child-care 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 as planned events like engineering works or social events on could be effectively managed (Hallenbeck et al., 2003a).

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 have been developed to measure NRC, which limits their usefulness to detect NRC events due to the following two reasons. First, substantial amount of data including travel demand, traffic capacity, incidents and weather is required to measure NRC. However, 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. Second, 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.

Detecting NRC events (hereafter NRCs) on an urban road network governs two main challenges. First, evaluating the detected NRCs is difficult due to the absence of ground-truth data. In other words, there is lack of information regarding whether or not a traffic observation belongs to an NRC in reality. This challenge could be addressed by using an incident data set, as traffic incidents 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 been previously been reported that many large delays cannot be explained by an incident data set (Hallenbeck et al., 2003a). Second, detecting NRCs accurately (i.e. without missing to detect any NRC and also not including day-to-day variations in traffic to belong to an NRC) is difficult due to the heterogeneous nature of an urban road network. An urban road network is heterogeneous in terms of link lengths, which may 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. This paper solely focuses on the first challenge.

The aim of this paper is to develop an NRC detection methodology that supports the accurate detection of NRCs on large urban road networks. In order to achieve this aim, three objectives are identified. First is to develop an NRC detection method for a large urban road network. Second is to represent each NRC in its evolution, so that the spatial and temporal extent of the detected NRCs could be identified. Third is to determine criteria to evaluate different NRC detection methods. The methodology is based on Link Journey Time (LJT) data, where an LJT is the estimated journey time through a link at an established time interval. The proposed NRC detection method does not make an assumption regarding the cause of an NRC, and aims to detect NRCs regardless their cause.

This paper is organised as follows. Section 2 discusses the relevant literature on NRC detection, and spatio-temporal clustering. Section 3 describes the proposed NRC detection method. Section 4 describes the proposed criteria to evaluate the performance of an NRC detection method. Section 5 applies the proposed NRC detection methodology on London's urban road network. Section 6 presents a discussion of the proposed NRC detection methodology emphasizing on the first challenge of developing an NRC detection methodology (i.e. evaluating an NRC detection method in the absence of ground-truth data). Finally, Section 7 concludes this paper by stating the limitations of the proposed methodology and the future research directions.

2. NRC detection and spatio-temporal clustering

The importance of NRC detection has been recognised by most traffic operation centres, yet research on this topic is recent. Even though there is lack of research on NRC detection, it is closely linked with incident detection. This is because an incident is usually assumed to cause an NRC. Thus, it is reasonable to consider that detection of an incident necessitates detection of an NRC. Therefore this section explores to what extent Automatic Incident Detection (AID) methods can be used for NRC detection. This section also identifies the relation between spatio-temporal clustering and NRC detection.

Automatic Incident Detection (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 a road network. Research efforts have resulted in a substantial number of AID methods and several surveys have been written to summarise these methods (Martin et al., 2001; Parkany and Xie, 2005; Weil et al., 1998).

Most of these AID methods have been developed for motorways (freeways), where inductive loops are commonly used to collect occupancy data. Therefore, implementing the existing AID methods for NRC detection on large urban road networks poses two limitations. Firstly, installing inductive loops across a large urban road network is not economically feasible, as inductive loops operate on a short section of a road (Hall, 2001). Secondly, a different theoretical consideration from that used for the well-developed motorway networks is required to understand traffic flows in an urban road network (Kerner, 2004, p. 582). This is because, analysis on an urban road network is more complicated due to the irregular interruptions (e.g. traffic lights, pedestrian crossings or side-street parking) and that the traffic flow is not conserved on urban road

networks (Lee et al., 2011; Ozbay and Kachroo, 1999, p. 79). Therefore, new techniques should be investigated for NRC detection on large urban road networks.

Traffic congestion has previously 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. Therefore, empirical research has investigated how spatio-temporal clustering can be used to detect congestion patterns in an urban road network. For example, Li et al. (2007) clustered road segments based on the density of common traffic they share. They have experimented with simulated trajectory data for this purpose. However, their research cannot differentiate whether the detected congestion patterns are due to RC or NRC.

3. Spatio-temporal clustering to detect NRCs

A novel NRC detection method is proposed in this section to detect NRCs on a large urban network based on spatio-temporal clustering of substantially high LJTs. The aim of the proposed method 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 detected NRC can be quantified based on its impact. The methodological framework is illustrated in Fig. 1.

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 paper should be described, which is Link Journey Time (LJT). The wide usage of Automatic Number Plate Recognition (ANPR) cameras for traffic enforcement purposes allows traffic operation centres to estimate aggregate LJT from the vehicle journey times that are obtained by matching the readings of ANPR cameras. Although the procedure to estimate LJTs is not within the scope of this paper, a brief explanation of this
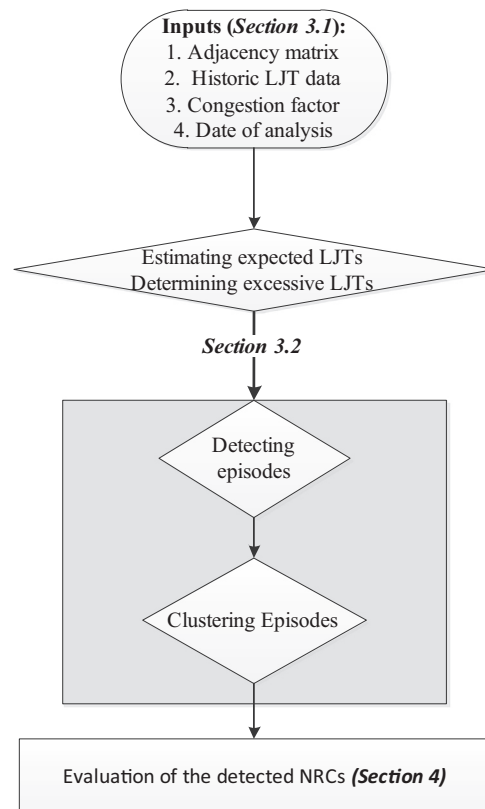

Fig. 1. Spatio-temporal clustering to detect NRCs based on LJT data.

process could be found in Robinson and Polak (2006). Because LJT is a prominent traffic data type in network-wide analysis, this paper uses LJT as the main traffic data type for NRC detection, where NRCs are considered to be a cluster of substantially 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, $n_a(\text{from})$ and $n_a(\text{to})$, where traffic on link a is flowing from $n_a(\text{from})$ to $n_a(\text{to})$.

The four main inputs to the proposed NRC detection method are described as follows:

Adjacency matrix (**M**) is the mathematical representation of a geographical road network. It is a $|\mathbf{A}| \times |\mathbf{A}|$ matrix, where $|\mathbf{A}|$ corresponds to the number of links. The value M(a, b) denotes the *adjacency relationship* between the two links, a and b. This paper 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; 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, M(a, a) = 1, $\forall a \in \mathbf{A}$.

The main reason to define the adjacency relation as aforementioned 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.

Historic LJT data are past LJT estimates that can be used for statistical analysis like estimating the expected 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 $c \geqslant 1$. '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). Therefore, determining the value of the 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. By this way, the NRCs that occurred within the analysed date would be reported to traffic operators.

Using these inputs, two important measures can be determined:

- *Expected LJT* is the mean travel time on a given link a at time interval t under normal traffic conditions. It is denoted as $\bar{y}_a(t)$ and measured in minutes. Expected LJTs 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., 2003b; Varaiya, 2007).
- *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 in Eq. (1):

$$\delta_a(t) = \begin{cases} y_a(t) - \bar{y}_a(t), & \text{if } y_a(t) > (c \times \bar{y}_a(t)), c \geqslant 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

The special case of excessive LJTs occurs when $c = 1$, in which case this paper uses the term '*delay*' instead of excessive LJT. Excessive LJTs are the main component of an NRC, because such LJTs are substantially higher than their expected values. The simplifying assumption in determining excessive LJTs is that the congestion factor is the same value for all the links and times. Although this assumption may not completely capture the heterogeneous nature of an urban road network, it is in-line with the operational uses of most of the traffic operation centres.

3.2. Clustering episodes for NRC detection

This section explains the proposed spatio-temporal clustering method to detect NRCs based on LJT data. Substantially high LJTs are clustered to detect NRCs. The proposed method consists of two steps; detecting episodes and clustering episodes, which are discussed in Sections 3.2.1 and 3.2.2 respectively. In Section 3.2.3, the process to determine the evolution of an NRC is introduced, which allows a researcher to identify the spatial and temporal extent of an NRC event; and an example is illustrated to clarify the concepts discussed in this section.

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 intervals indexed by m and n, as shown in Eq. (2) as follows:

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

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, where the episode begins and ends with the time intervals indexed by *m* and *n* respectively. The number of episodes on a link varies from zero where no LJT is excessive to *T*/2 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 $d(e\{a\}) = (n - m + 1) \times \tau$, where $m, n \in e\{a\}$ and τ is the time interval to estimate two consecutive LJTs 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 $s(e\{a\}) = \sum_{t=m}^{n} \delta_a(t)$, 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 thresholds defined by the high, medium and low congestion factors are denoted as 'High c', 'Medium c' and 'Low c' respectively in Fig. 2. Having defined these thresholds, after applying the definition of an episode, all the episodes that occurred on a link would be identified. If the threshold is defined by using the medium congestion factor, then four episodes are identified (denoted with *e1*, *e2*, *e3* and *e4*), as shown in Fig. 2.

Two observations can be made regarding the detected episodes:

- 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. Therefore, the detected episodes depend on the value of the congestion factor, which is usually determined by domain knowledge.
- 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 Fig. 2 is based on a single link. However, an urban road network usually comprises many links, and it is possible for an NRC event to span multiple links. Therefore, an NRC detection method should be able to detect NRCs that may span multiple links.

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 high enough to belong to an episode depends on the congestion factor. Therefore, having determined a congestion factor, an episode could be considered to be *similar* to an NRC event. A major indication of an NRC spanning multiple links, would be to observe episodes on adjacent links at the same times. This is because the reduction in the capacity of a link (e.g. due to an incident) could cause a vehicular queue to grow towards its upstream link. These vehicular queues would be represented by several episodes occurring on adjacent links, and consequently an NRC could be described as a cluster of these episodes.

In order to detect NRCs, *spatio-temporally overlapping* episodes are clustered. To clarify the concept of 'spatio-temporally overlapping', suppose that 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:
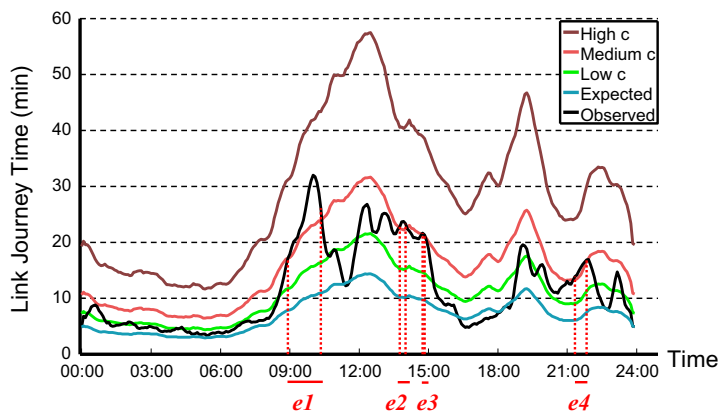

Fig. 2. Detection of episodes using the medium congestion factor.

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

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 pseudocode of the 'Clustering Episodes' method is depicted in Fig. 3. Set $\mathbf{E}$ denotes all of the episodes that are detected on the entire road network across the study period by applying the Eq. (1). Set $\mathbf{Z}$ denotes the NRCs, where an NRC contains a cluster of spatio-temporally overlapping episodes. The pseudocode illustrated in Fig. 3 is achieved by using the 'pseudocode' Latex package (Kreher and Stinson, 1999).

The set of NRCs (i.e. $\mathbf{Z}$) 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 an 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 Fig. 3 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 in terms of its *lifetime* and *severity*. Lifetime is the temporal extent of an NRC, and it represents the time period that an NRC contains at least one link. 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.
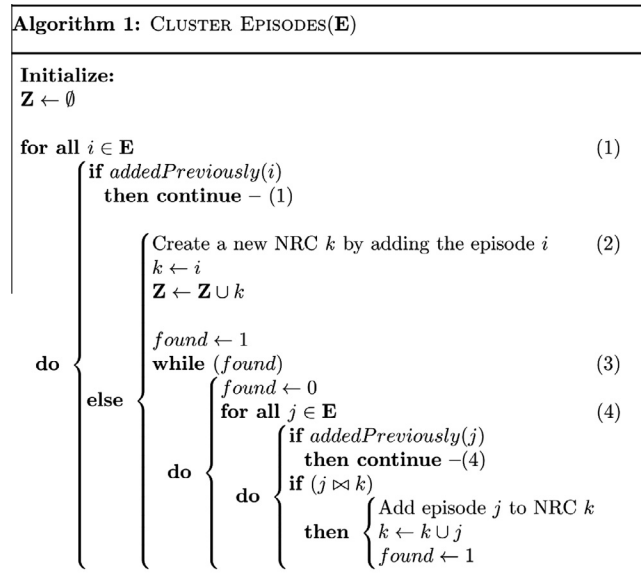
Algorithm 1: CLUSTER EPISODES(**E**)

Initialize:
$\mathbf{Z} \leftarrow \emptyset$

for all $i \in \mathbf{E}$ (1)
  if $addedPreviously(i)$
    then continue $-$ (1)

do
  else
    Create a new NRC k by adding the episode i (2)
    $k \leftarrow i$
    $\mathbf{Z} \leftarrow \mathbf{Z} \cup k$

    $found \leftarrow 1$
    while $(found)$ (3)
    do
      $found \leftarrow 0$
      for all $j \in \mathbf{E}$ (4)
      do
        if $addedPreviously(j)$
          then continue $-$(4)
        if $(j \bowtie k)$
          then
            Add episode j to NRC k
            $k \leftarrow k \cup j$
            $found \leftarrow 1$

Fig. 3. Pseudocode of the 'Clustering Episodes' method to detect NRCs.

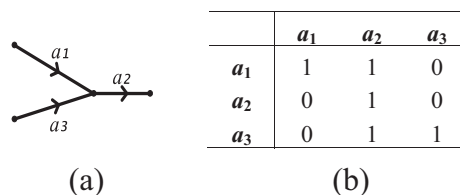|	a_1	a_2	a_3
a_1	1	1	0
a_2	0	1	0
a_3	0	1	1

(a)   (b)

Fig. 4. A simple road network (a) and its adjacency matrix (b).
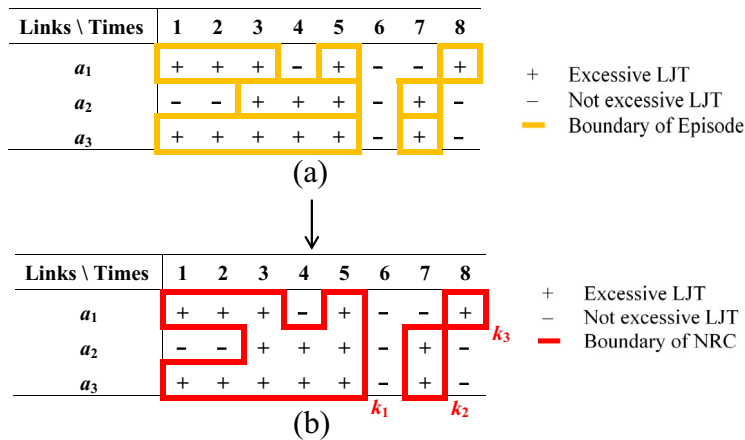
Fig. 5. Clustering episodes to detect NRCs.

To clarify the aforementioned concepts, suppose a three-link network as shown in Fig. 4(a), where the direction of traffic flow is indicated with an arrow. The adjacency matrix of this network is illustrated in Fig. 4(b).

Suppose that eight LJTs are estimated for these three links, and those which are excessive are illustrated with a ' + ' sign in Fig. 5. 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 Fig. 5(a). Three NRCs are found by clustering these episodes based on the Clustering Episodes method described in Fig. 3. The boundaries of these NRCs are highlighted with red in Fig. 5(b).

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 twelve excessive LJTs, whereas the third NRC contains only one excessive LJT. As this paper 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 Fig. 3, as it has a concave shape.

Representing the detected NRCs in their evolution would allow a researcher to quantify their lifetime and severity. By this way, each detected NRC could be quantified. The evolutions of the aforementioned NRCs are illustrated in Table 1.

The lifetime of the first NRC is between time intervals one and five; and the life-times of second and third NRCs are time intervals 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 would be $\delta_{a_2}(7) + \delta_{a_3}(7)$.

4. Evaluating the detected NRCs

The NRC detection method proposed in Section 3 to detect NRCs based on LJT data do not make any assumption regarding the severity or duration of an NRC event. The method uses excess LJTs as the main indicator of an NRC. Naturally, different values of congestion factor would lead to different episodes, hence, different NRCs. In order to determine the most suitable congestion factor, it is necessary to evaluate the detected NRCs. Therefore, the aim of this section is to seek for an answer to the question '*how can the performance of an NRC detection method be evaluated?*'

Table 1
Evolution of NRC clusters detected in Fig. 5(b).

NRC ID	Evolution	
	Time Interval (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

The traditional approach to evaluate the detected NRCs would be to compare them with a ground-truth data (i.e. data indicating whether or not the estimated LJTs do truly belong to an NRC). Lack of ground-truth data is a common issue in many clustering tasks (Dubes and Jain, 1976). Therefore, researchers often propose new criteria to evaluate clustering methods considering the uses of the designated clustering task (Luxburg et al., 2012). Consequently, two criteria that complement each other are proposed within this section. These two criteria are designed to consider the practical uses of an NRC detection method. The first detects NRCs that result in substantial impact that would take a traffic operator's attention. The second relates the detected NRCs with incidents. These two criteria are referred to as the 'High-Confidence Episodes' and the 'Localisation Index' respectively.

4.1. High-confidence episodes

This evaluation criterion assesses to what extent an NRC detection method detects 'High-Confidence' episodes (e^*s). 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 time intervals indexed by m and n in Eq. (3) as follows;

$$e^*\{a\} = \left\{ [mn] \left| \begin{array}{ll} \delta_a(t) > 0 \; \forall t \in [mn]; m, n \in \{1, 2, \ldots, T\} & \text{and} \\ \delta_a(m-1) = \delta_a(n+1) = 0 & \text{and} \\ n - m + 1 \geqslant t^{\min} \end{array} \right. \right\} \tag{3}$$

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 differentiates an episode with an e^*. Specifically, an e^* should occur for a minimum duration, which is denoted by $t^{\min}$. The continuous time of the minimum duration is $t^{\min} \times \tau$ minutes, where τ is the LJT estimation interval. The main reason for proposing this evaluation criterion is to satisfy a practical requirement, as very high LJTs that last a minimum duration demand attention from the traffic operators. Because an e^* commands this attention, an NRC detection method should be able to detect the e^* s that occurred on an analysed date.

Once the values of the unknown parameters of an e^* are determined by historical analysis (described in Appendix A) or by domain knowledge; the LJTs that belong to the detected NRCs and e^*s could be compared by using a Venn diagram as shown in Fig. 6. The number of LJTs that belong to both the e s and the detected NRCs is denoted with 'TP' corresponding to True Positives. The number of LJTs that do not belong to the e^*s, but belong to the detected NRCs is denoted with 'FP' corresponding to False Positives. The number of LJTs that belong to the e^*s but not the detected NRCs is denoted with 'FN' corresponding to False Negatives. Finally, the number of LJTs that neither belong to the e^*s nor the detected NRCs is denoted with 'TN' corresponding to True Negatives.

From this comparison, two important measures could be estimated. False Alarm Rate (FAR) is the proportion of all LJTs that are detected to belong to an NRC but do not belong to an e^*. Formally, FAR = FP/(TP + FP). False Negative Rate (FNR) is the proportion of all LJTs that belong to e^* which were not included within the detected NRCs to all LJTs that belong to e^*. Formally, FNR = FN/(TP + FN). These two measures assess to what extent an NRC detection method identified and missed e^*s, which are considered to be the episodes that should ideally be detected. Of these two measures, FNR is the critical one as it determines the proportion of e^*s that are missed. Therefore, an ideal NRC detection method should have low FNR. On the other hand, it is possible for an NRC to last less than the minimum duration defined in the definition of an e^*. Hence, an ideal NRC detection model should have some FAR, but its value is unknown due to the lack of ground-truth data.

4.2. Localisation index

The reported NRCs should ideally be associated with the reported incidents. In order to associate 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 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 lifetime, 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 Fig. 7. An incident occurs at link $a4$ and its effect propagates towards its upstream links during the analysis period. The excessive LJTs are illustrated with a '+' sign in Fig. 7, and the detected NRC is highlighted in red.

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 Fig. 7, there is only one connected component between times one and five.
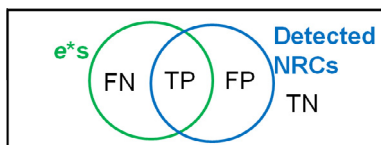

Fig. 6. Comparison of the detected NRCs with the high-confidence episodes (e^*s).

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

Traffic flow direction ↓

+ Excessive LJT
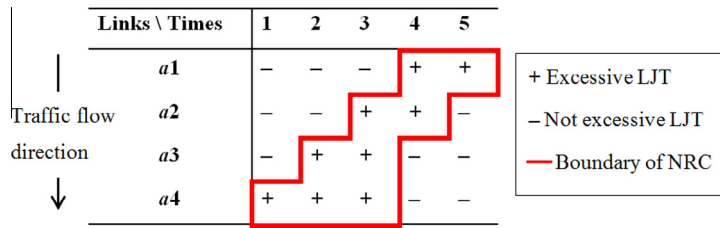− Not excessive LJT
— Boundary of NRC

Fig. 7. Propagation of the effect of an incident towards upstream links.

However, the links that are included within the connected component change over time. For example, at time one the connected component consists of only link $a4$, whereas at time three 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. Because the detected NRC consist of a single connected component throughout its lifetime, it is possible to relate it with an incident.

On the other hand, for the NRC illustrated in Fig. 8, 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.

Calculation of the 'Localisation Index' involves 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 average number of connected components that the kth NRC exhibits throughout its lifetime. For the NRCs illustrated in Figs. 7 and 8, this number would be 1.0 (5/5) and 1.7 (5/3) respectively.
2. The maximum average number of connected components for all the detected NRCs is the Localisation Index.

The higher the Localisation Index is the harder to relate the detected NRCs with the reported incidents. 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. By this way, an NRC detection method that considers day-to-day variations in traffic to belong to an NRC is penalised. 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 regarding the congestion formation on road 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 regarding 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).

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 km (TfL, 2013) and a road network of 13,800 km (TfL, 2012). 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 case study to investigate NRCs in an urban road network.
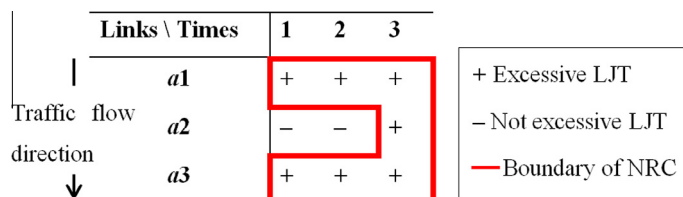
Links \ Times	1	2	3
a1	+	+	+
a2	−	−	+
a3	+	+	+

Traffic flow direction ↓

+ Excessive LJT
− Not excessive LJT
— Boundary of NRC

Fig. 8. An NRC whose spatial localisation is more difficult than the one illustrated in Fig. 7.

A link is usually defined as the route between two Automatic Number Plate Recognition (ANPR) cameras with the highest traffic flow. Two links may spatially overlap with each other due to the lack of ANPR cameras at every junction. However, including spatially overlapping links within the analysis is not feasible, as it creates redundancies, which unnecessarily increases the run time. Hence, Transport for London (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 data quality regarding the estimated LJTs. Therefore, when reporting the performance of road networks, TfL relies on data collected from core links. There are 1,261 ANPR links in total, 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 Fig. 9. 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, we have also used the core ANPR network. Hereafter 'link' indicates a 'core link'.

ANPR links display different lengths and characteristics. Link lengths vary from 175 to 13,487 m, and the total length of the network is around 1250 km. 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.

Traffic operators commonly use expected LJTs for road network performance monitoring. This paper uses the expected LJTs that are used in TfL. By this way, it is aimed to have consistency between the results obtained within the analysis of this paper and the practice. TfL classifies the days of a week into two: weekdays (Monday to Friday) and weekends (Saturday and Sunday). 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, approximately two months (March and April 2009) of data are used to estimate LJTs. The estimated expected LJTs are currently being used at TfL, as domain experts consider them to reflect the normal (recurrent) traffic conditions.

The analysis in this paper is conducted on the weekdays of October 2010. A tube strike occurred on 4 October, which increased LJTs considerably across the entire road network (Tsapakis et al., 2013). In order to preserve the integrity of the analysis, 4 October has been removed. Hence, a total of 20 weekdays are analysed. The analysis has been conducted between 07:00 and 19:00, as this time interval covers the AM/Inter/PM peak periods in London (TfL, 2010, p. 268). There are a total of 145 LJTs within the analysis period, as each LJT is estimated every τ = 5 min (12 h × 12 LJTs/hour + 1, where one is added as the analysis period is inclusive of 07:00 and 19:00).

The main parameter for using the Clustering Episodes method (denoted as 'CE') is the congestion factor. Five different congestion factors are tested (i.e. 1.2, 1.4, 1.6, 1.8 and 2.0) in accordance with the values which are currently being used at TfL (2010, p. 95). 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 traffic operators have a fixed allocated time to analyse the detected NRCs, the higher the number of NRCs, the less time to investigate each NRC. On the other hand, the lower the number of NRCs is the higher the likelihood to miss some NRCs. In addition, run-time of these different models is also compared in order to demonstrate the practical feasibility of the proposed method to detect NRCs (analyses are conducted on a workstation running Windows 7 with a 12 GB RAM and 3.2 GHz processor; and the code is written in Matlab R2012a.
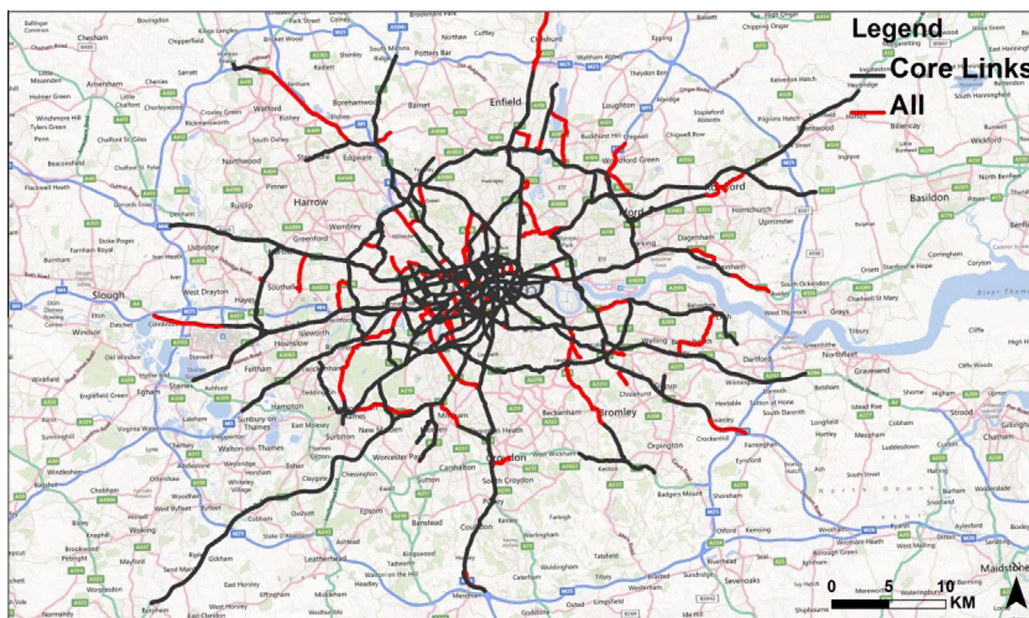

Fig. 9. London's ANPR network.

The code would be supplied to interested readers upon request). Finally, these models are compared based on the evaluation criteria described in Section 4. In the results, the number associated with each Clustering Episodes ('CE') model denotes the congestion factor used to determine whether or not an LJT is excessive in that model.

5.1. Number of detected NRCs

The boxplot of the numbers of NRCs detected by the five different CE models are illustrated in Fig. 10. The horizontal axis represents different CE models, whereas the vertical axis represents the number of NRCs. The red horizontal lines in the boxes refer to the median value of the number of NRCs on the analysed dates, while the upper and lower boundaries of the boxes indicate the upper and lower quartiles (i.e. 75th and 25th percentiles, and denoted as Q3 and Q1) respectively. The black lines vertical to the boxes (i.e. whiskers) show the maximum and minimal values that are not outliers. Outliers are indicated with a ' + ' sign. The value of an outlier exceeds 1.5 times the interquartile range (i.e. Q3–Q1). The arbitrary value of 1.5 has been commonly used to identify outliers since the advent of boxplots (Tukey, 1977).

The main outcome is that decreasing the congestion factor leads to 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, as the likelihood to include day-to-day variations to belong to an NRC increases by decreasing the congestion factor. Another interesting finding is that even when the most conservative model is considered (i.e. CE = 2.0), there are approximately 350 NRCs within 07:00 and 19:00. Large number of NRCs that occur within a day indicates the necessity of a dedicated team in a traffic operation centre to investigate these NRCs.

5.2. Run-time

The boxplot of the run-times of the five CE models are illustrated in Fig. 11, where the horizontal axis represents the CE models. Vertical axis represents the natural logarithm of the run-times in seconds in order to improve the legibility of the results.

Results indicate that the conservative models are much faster than the liberal models. This is reasonable as there would be fewer episodes to cluster in conservative methods; hence, faster run time. The results indicate the practical feasibility of the proposed method, as the CE $\leqslant$ 1.6 models usually complete the analysis less than one minute (i.e. $e^4 < 60$). Only when the most liberal method (CE = 1.2) is used, the average run-time increases to approximately one hour.

5.3. Evaluating the performance of the detected NRCs

This section evaluates the detected NRCs based on the two criteria proposed in Section 4. Specifically, 'high-confidence episodes' and 'Localisation Index' are evaluated for the aforementioned NRC detection models.

5.3.1. High-confidence episodes

The empirical evidence provided in Appendix A 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 duration of 25 min. High-confidence episodes are considered to be sufficient to attract a traffic operator's attention, as all the high-confidence episodes would account to the majority of the total delay. Having determined all e^*s, this performance criterion assesses to what extent an NRC detection model detected e^*s. The more an NRC detection model detects e^*s (i.e. low FNR), the better it is.

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
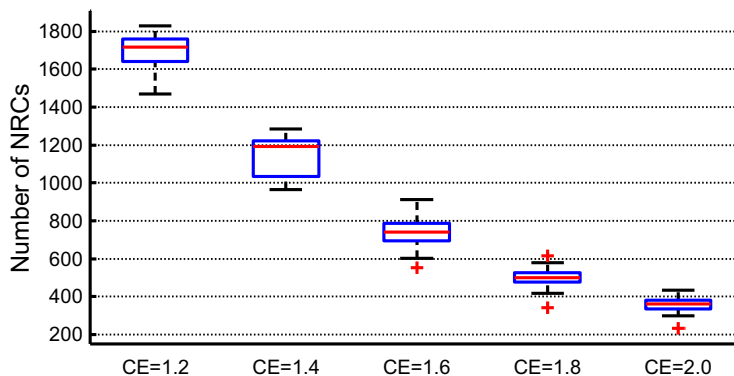

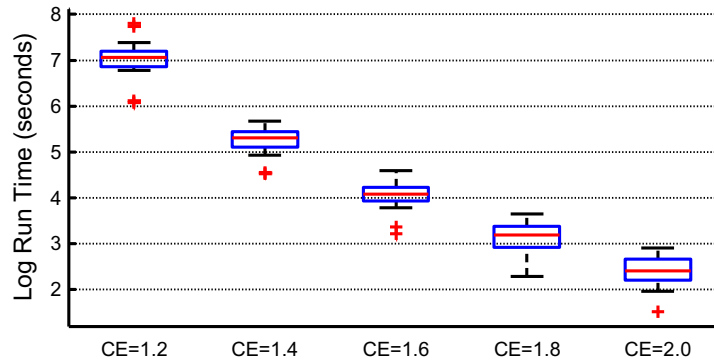**Fig. 10.** Boxplot of the number of detected NRCs.

Fig. 11. Boxplot of the natural logarithm of the run-times of CE models.

NRC detection model are calculated for all the analysed dates, and these values are then averaged and the results are illustrated in Fig. 12.

The best models regarding the detection of e^*s are the CE models that use a congestion factor of less than or 1.4 (which is the congestion factor used to define an e^*). As can be seen in Fig. 12, 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 increases the amount of e^*s that are missed. For example, when the congestion factor is set to 2.0, approximately 62% of e^*s are missed.

5.3.2. Localisation index

Localisation Index values for all the CE models have been calculated for all the analysed dates by applying the procedure described in Section 4.2. The box-plot of the Localisation Index values is illustrated in Fig. 13.

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 to consider day-to-day traffic variations 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 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' for the model CE = 2.0 is 3.51, which is actually detected as an outlier amongst the other dates. On the other hand, the maximum 'Localisation Index' for the model CE = 1.8 is 3.37. This highlights the possibility that the additional excessive LJTs might merge previously distinct connected components. To clarify, consider
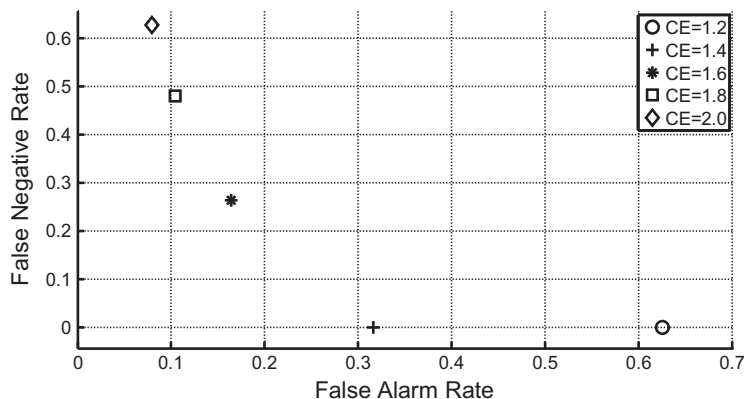

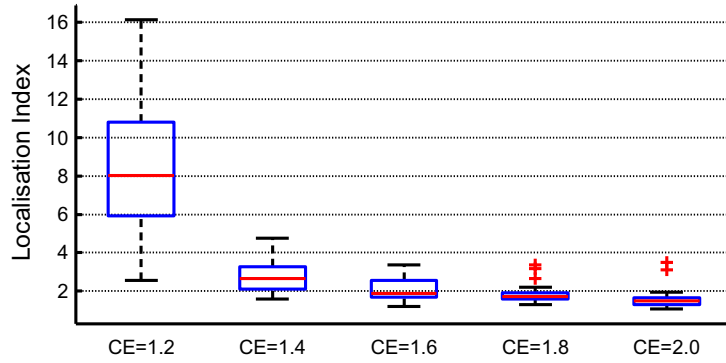**Fig. 12.** Performance of CE models when detecting high-confidence episodes.

Fig. 13. Boxplot of the Localisation Index values for different CE models.

the example illustrated in Fig. 8. If the LJT at link $a2$ at time 2 becomes excessive when the congestion factor is decreased, then the average number of connected components of the NRC will decrease. The results, however, indicate that this is a rare situation, and most of the time liberal ways of detecting NRCs lead to a higher Localisation Index.

5.3.3. Overall comparison of NRC detection methods

The previous two subsections investigated the performance of the different NRC detection models. As the congestion factor increases, the performance of the model degrades regarding the detection of high-confidence episodes; whereas the performance increases regarding the Localisation Index. Therefore, in order to determine the most suitable congestion factor, these two conflicting criteria should be combined into one criterion.

Multi-Attribute Decision Making (MADM, also referred to as multi-criteria decision making) provides the necessary theoretical background to compare different models (also referred to as 'alternatives') based on two or more, often conflicting, criteria. 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). This is a serious limitation, because it prevents a researcher to correctly identify the strengths and limitations of different NRC detection models with respect to each other. Consequently, identifying 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 models. For a given criterion, 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. The calculated ratios are multiplied to obtain the Final Score, which is illustrated by the Eq. (4) as follows:

$$\text{Final Score}(S_K, S_L) = \prod_{j \in \{e^*, \text{Localisation Index}\}} (S_{Kj}/S_{Lj})^{w_j}, \text{where} \sum_j w_j = 1 \tag{4}$$

where S_K and S_L are the two NRC detection models and j denotes a criterion which can be either the False-Negative Rate in e^* or the Localisation Index. S_{Kj} and S_{Lj} denote the values of the jth criterion of Kth and Lth NRC detection models respectively, and w_j is the weight (i.e. the relative importance) of criterion 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 criteria, 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 none of the values of a criterion could be zero. This assumption, however, is not satisfied within this papers' context, because the FNR values of the two Clustering Episode models (i.e. CE $\leqslant 1.4$) is zero. Recent studies explicitly 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 paper relies on WPM to compare different NRC detection models with the following adjustment: all the models' FNR values are incremented starting from 0.01 to 0.1 for all the analysed dates, by considering equal weights for both of the proposed criterion (i.e. $w_{e^*} = w_{\text{Localisatation Index}} = 0.5$). The median of the calculated Final Scores for the analysed 20 weekdays of October 2010 are illustrated in Fig. 14.

The following conclusions can be drawn by analysing the Final Scores. Firstly, the most suitable congestion factor is found to be approximately 1.4, as the CE = 1.4 model resulted the least Final Score amongst the other models. This suggests that whenever an estimated LJT is higher than 40% of its expected value, then it should belong to an NRC. The difference between
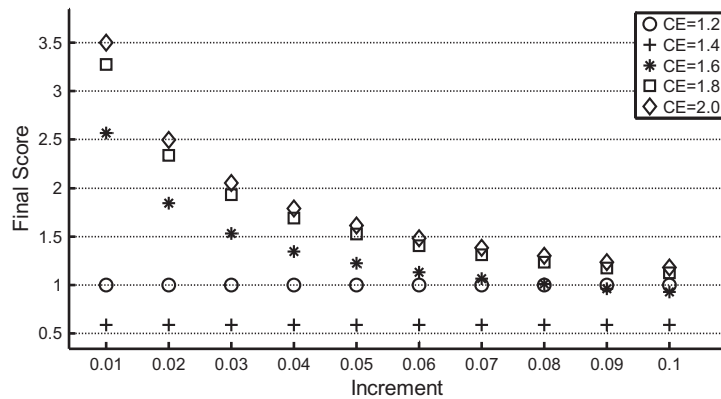
Fig. 14. Median Final Score values for the analysed NRC detection models with different increments to FNR values.

CE = 1.4 model and the rest decrease when the incremented value increases. This is a theoretically plausible outcome because whenever the same increment is added both to the numerator and denominator, the resulting ratio would be closer to one. In other words, the relative effect of FNR on the Final Scores would gradually decrease with the incremented amount. Due to this very reason, CE = 1.2 performs better than CE = 1.6 for increment values less than 0.08. When the incremented value is greater than 0.08, CE = 1.6 performs better than CE = 1.2 because the Localisation Index becomes relatively more important in the calculation of the Final Score. Note that, the Final Scores of CE = 1.2 is always one, because it is considered to be the reference model in which all the models are compared with respect to CE = 1.2. However, the ranking of the models would be the same regardless the chosen reference model or if we add another model (e.g. CE = 1.7) to the existing models as the Final Scores are calculated based on WPM.

6. Discussion

An important challenge of an NRC detection methodology is to evaluate a method in the absence of ground-truth data. Specifically, a researcher usually does not know the true spatial and temporal extent of NRCs. In order to understand why there is lack of ground-truth data, it is necessary to discuss the three main ways to obtain it:

- *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.
- *Using traffic simulations* simulations use traffic models to generate artificial traffic data. Traffic simulations, however, 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 separate data set* Research on Automatic Incident Detection (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 the ground-truth data may not be completely reliable.

Consequently, there is no feasible way of collecting the ground-truth data regarding the NRCs for a large urban road network. Nevertheless, the proposed two novel evaluation criteria have overcome this limitation, and demonstrated their effectiveness on a large urban road network as they complement each other. Specifically, liberal NRC detection models perform better regarding the detection of high-confidence episodes; whereas they perform poorly regarding the Localisation Index and vice versa for the conservative models. This outcome is theoretically plausible as the aim of NRC detection method is to detect all the NRCs that have occurred within a day without including day-to-day variations in traffic to belong to NRCs.

Even though this paper has only analysed London's urban road network, the proposed evaluation criteria are generic as long as the following requirements are met:

- Link Journey Time (LJT) data are available.
- The analysed road network could formally be described with an adjacency matrix (M) which is not a diagonal matrix. Formally, for the links a and b, M(a, b) = 1, $a \neq b \exists a, b \in \mathbf{A}$, where $\mathbf{A}$ denotes the set of links.
- An NRC could be observed on several adjacent links.

These requirements may not always be met in real-life scenarios. For example, it might be possible that a researcher or traffic operation centre has only access to data obtained from inductive loop detectors. In this case, because we do not have LJT data, it would not be possible to determine the performance of an NRC detection method regarding the detection of high-confidence episodes. Another case would be the situation where the adjacency matrix is represented by a diagonal matrix. This situation might occur when we are dealing with a sparse road network where links are not connected with each other. In this situation, Localisation Index criterion would be meaningless, because all the NRC detection methods would appear to perform the best regarding this criterion (i.e. the Localisation Index values would be one for all the NRC detection models). Nevertheless, these requirements are not very strict would likely be met in a traffic operation centre managing an urban road network. Therefore, the proposed NRC detection methodology could readily be adapted by a researcher or an urban traffic operation centre.

7. Conclusions

Accurate detection of NRCs in large urban road networks is gaining importance due to the growing necessity to manage urban road networks more effectively. This paper presents an initial attempt to address this necessity, and considers NRC detection as a task of spatio-temporal clustering, where substantially high LJTs are clustered. The proposed method clusters spatio-temporally overlapping episodes to detect NRCs. All of the detected NRCs are quantified based on their impact consisting of 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 characteristics are obtained by representing each detected NRC in its evolution. In this way, traffic operators could observe the dynamics of an NRC, including the development of an NRC on a road network and in time. In this way, detected NRCs could be quantified, hence their investigation could be prioritised.

The proposed NRC detection method relies on a single parameter, which is referred to as the 'congestion factor'. The usage of a single parameter to identify congestion levels is a common approach in urban traffic operation centres. Therefore, the proposed NRC detection method is compatible with the domain expertise. Different values of this parameter, however, would lead to different NRCs. Therefore, a formal process is required to evaluate the detected NRCs. Due to the lack of ground-truth data, two criteria are proposed considering the uses of NRC detection. 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 criteria. The analyses are conducted on London's urban road network. This road network consists of 424 links, and analysis on such a large urban road network is hardly matched in previous studies. The congestion factor values are varied between 1.2 and 2.0, and these values are determined based on the domain knowledge. The most suitable congestion factor is found to be approximately 1.4 by weighting both of the evaluation criteria equally in a Weighted Product Model (WPM). This finding suggests that, whenever an estimated LJT is at least 40 percent higher than its expected value, then it should belong to an NRC.

7.1. Limitations of the methodology

There are several limitations of the proposed NRC detection methodology. First, the proposed NRC detection method is indifferent to the variations in the expected LJTs, as it only considers the expected LJTs and the congestion factor to identify excessive LJTs. Considering a single congestion factor across the entire road network may lead to the detection of NRCs that do not truly have a physical significance. For example, day-to-day variations in short links (e.g. variations in the timings of a traffic signal) might be considered to belong to an NRC. Second, the way in which *spatio-temporally overlapping* episodes is defined may prevent a direct matching of the detected NRCs with incidents. Specifically, if two or more episodes are spatio-temporally overlapping, then they are assumed to belong to the same NRC, even though their underlying causes are different. Furthermore, 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 could be observed on both links individually that are *not* spatio-temporally overlapping. These episodes would be considered to belong to different NRCs, but actually these NRCs are due to a single cause.

There are also limitations regarding the proposed evaluation criteria. First, all high-confidence episodes are considered to be equal in terms of their contribution to the False-Negative Rate (FNR), regardless their severity. Specifically, missing two e^*s that have the same duration, but different severities causes the same effect towards the calculation of FNR. Second, the Localisation Index criterion is indifferent to the order of occurrence order of the number of connected components.

7.2. Future research

In order to explore the full potential of NRC detection on large urban road networks, this paper could be extended in the following directions. First, new NRC detection methods should be investigated that considers the statistical properties of the estimated LJTs. In this way, the heterogeneous nature of an urban road network (e.g. due to differences in link lengths or data quality) might be captured more effectively. This research direction would also necessitate investigating further evaluation criterion, which would be used to assess the extent to which an NRC detection method captured the heterogeneous nature of an urban road network. Second, the way in which the adjacency matrix is defined could be improved by either incorporating domain knowledge or by developing a more systematic approach that considers link lengths. Finally, it would be very interesting to apply the proposed methodology on different urban road networks and investigate how the results vary on different urban road networks. The vision of this future research direction is to develop a formal traffic theory to explain NRCs on an urban road network. In this way, NRCs could possibly be simulated on a large urban road network, which would allow researchers to evaluate the effectiveness of their NRC detection methods in a more comprehensive way.

Acknowledgements

Appendix A

This appendix is dedicated to explain the historical analysis conducted to determine the values of the unknown parameters within the definition of a high-confidence episode (e^*) as discussed in Section 4.1. Two parameters are required to define an e^*. The first is congestion factor c, which determines whether or not an LJT is excessive, and the second is the minimum duration ($t^{\min}$), which ensures that the episode lasted for a substantial amount of time.

A.1. 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 paper considers an LJT to be excessive enough to belong to an e^* if it is greater than or equal to its 95th percentile value. This is because 95th percentile values are commonly used in traffic science to indicate unusually high traffic data (Pu, 2011). Once the 95th percentile LJTs are determined, then the congestion factor that would classify them 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 95th percentile of LJTs using a historic data set. The 95th percentile LJT for link a at time interval t is denoted as $y_a^{95}(t)$, which is calculated by linear interpolation as follows (Matlab, 2013). Collect historical $y_a(t)$ and sort them in ascending order, and denote this data vector as **H**, where **H**(1) and **H**(h) denotes the lowest and highest $y_a(t)$ in the historical data respectively. Find the rank of the 95th percentile in **H** by, $r = (0.95 \times h) + 0.5$. Hereafter, $y_a^{95}(t)$ is calculated as the linearly interpolated value between **H**(r) and **H**($r + 1$), depending 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, $c_a^{95}(t) = y_a^{95}(t)/\bar{y}_a(t)$.
3. Repeat steps 1 and 2 for each LJT, $a \in \mathbf{A}$ and $t \in [1, 2, \ldots, T]$, where **A** denotes the set of links and T denotes the total number of LJTs within the analysis interval.
4. The median of the $c_a^{95}(t)$ values is considered to be the congestion factor that is used to define an e^*.

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 min, so the analysis interval contains T = 145 LJTs (12 h $\times$ 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 LJTs (i.e. h = 87). The box-plot of the estimated $c_a^{95}(t)$ values with respect to time is illustrated as shown in Fig. 15.

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 paper makes the simplifying 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
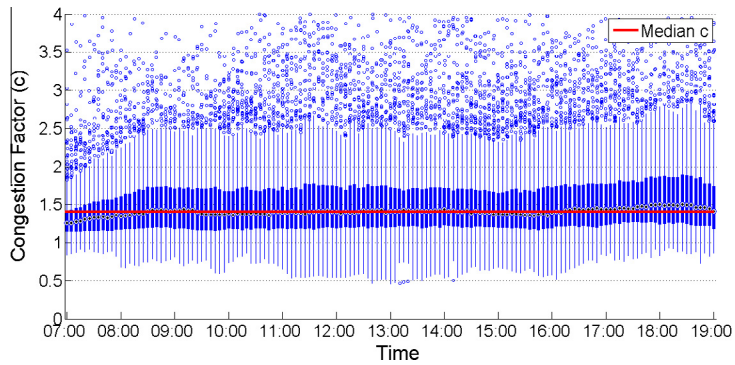
Fig. 15. Determining the congestion factor of a high-confidence episode.

is used instead of mean as it 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 found to be approximately 1.4. This suggests that whenever an LJT is 40 percent higher than its expected value, then it should be considered as excessive.

A.2. 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 (i.e. using a congestion factor of 1.4 to determine excessive LJTs) 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 $d(e_i) \geq d(e_{i+1})$, where $d(e_i)$ is the duration of the ith episode.
   Find the duration of the episode when the cumulative severities of the episodes explain the majority of the total delay. Formally, find $d(e_x)$, where $x = \min_x(\sum_{i=1}^{x} s(e_i)/\Psi) \geq 0.5$, and Ψ is the total delay observed during the analysis period and $s(e_i)$ is the total severity of the ith episode.
4. The $d(e_x)$ 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 $d(e_x)$ values for different days. The histogram of these $d(e_x)$ values for the analysis period is illustrated in Fig. 6.

The most noticeable outcome in Fig. 16. is the substantial decrease in the number of days with a $d(e_x)$ value greater than 7. These dates are highly congested; for example, the 25th May 2010, State Opening of Parliament, is one such date. However,
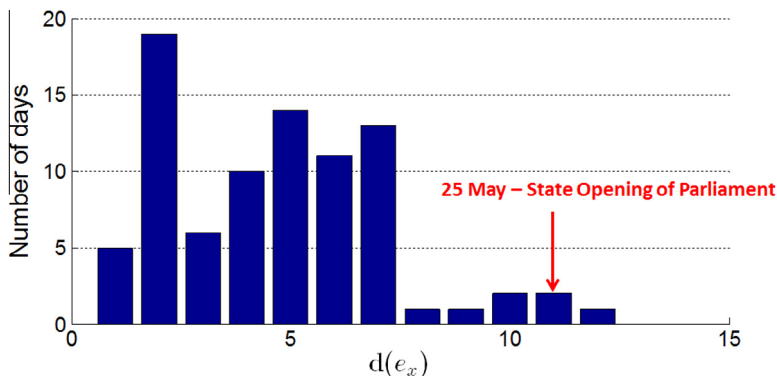

Fig. 16. Histogram of the days where episodes lasting $d(e_x)$ number of LJTs are sufficient to explain the majority of total delay.

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 ($t^{\min} = 5$). Since each LJT is estimated every five minutes, the minimum duration to define an e^* is determined to be $5 \times 5 = 25$ min.

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 min.

References

Bloomberg, L., Dale, J., 2000. Comparison of VISSIM and CORSIM Traffic Simulation Models on a Congested Network. Transp. Res. Rec. J. Transp. Res. Board 1727, 52–60. http://dx.doi.org/10.3141/1727-07.

British Library, 2013. The Rise of Cities [WWW Document]. URL: <http://www.bl.uk/learning/histcitizen/georgians/cities/riseofcities.html> (accessed 1.30.13).

Cheu, R.L., Ritchie, S.G., 1995. Automated detection of lane-blocking freeway incidents using artificial neural networks. Transp. Res. Part C Emerg. Technol. 3, 371–388. http://dx.doi.org/10.1016/0968-090X(95)00016-C.

Dowling, R., Skabardonis, A., Carroll, M., Wang, Z., 2004. Methodology for measuring recurrent and nonrecurrent traffic congestion. Transp. Res. Rec. J. Transp. Res. Board 1867, 60–68. http://dx.doi.org/10.3141/1867-08.

Dubes, R., Jain, A.K., 1976. Clustering techniques: the user's dilemma. Pattern Recognit. 8, 247–260. http://dx.doi.org/10.1016/0031-3203(76)90045-5.

Emmerink, R.H.M., Axhausen, K.W., Nijkamp, P., Rietveld, P., 1995. The potential of information provision in a simulated road transport network with non-recurrent congestion. Transp. Res. Part C: Emerg. Technol. 3, 293–309. http://dx.doi.org/10.1016/0968-090X(95)00012-8.

FHWA, 2012a. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation: Executive Summary [WWW Document]. URL: <http://www.ops.fhwa.dot.gov/congestion_report/executive_summary.htm> (accessed 10.29.12).

FHWA, 2012b. Travel Time Reliability: Making It There On Time, All The Time [WWW Document]. URL: <http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm#WhyisTTR> (accessed 8.30.12).

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. ACM, New York, NY, USA, pp. 33–42, doi:10.1145/312129.312190.

Hall, F.L., 2001. Traffic Stream Characteristics, in: Traffic Flow Theory: A State-of-the-Art Report. FHWA/TRB/ORNL.

Hallenbeck, M.E., Ishimaru, J.M., Nee, J., 2003a. Measurement of recurring versus non-recurring congestion –Technical Report. Washington State Dept. of Transportation.

Hallenbeck, M.E., Ishimaru, J.M., Nee, J., 2003b. Measurement of recurring versus non-recurring congestion – Final Report. Washington State Dept. of Transportation.

Han, L.D., May, A.D., 1989. automatic detection of traffic operational problems on Urban arterials. Res. Rep. Inst. Transp..

Helbing, D., 2001. Traffic and related self-driven many-particle systems. Rev. Mod. Phys. 73, 1067–1141. http://dx.doi.org/10.1103/RevModPhys. 73.1067.

Kerner, B.S., 2004. The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory, first ed. Springer.

Kerner, B.S., Konhäuser, P., 1994. Structure and parameters of clusters in traffic flow. Phys. Rev. E 50, 54–83. http://dx.doi.org/10.1103/PhysRevE.50.54.

Kreher, D.L., Stinson, D.R., 1999. Pseudocode: A LATEX Style File for Displaying Algorithms.

Kwon, J., Mauch, M., Varaiya, P., 2006. Components of congestion: delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. Transp. Res. Rec. 1959, 84–91. http://dx.doi.org/10.3141/1959-10.

Lee, W.H., Tseng, S.S., Shieh, J.L., Chen, H.H., 2011. Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location-based services. IEEE Trans. Intell. Transp. Syst. 12, 1047–1056. http://dx.doi.org/10.1109/TITS.2011.2144586.

Leskinen, P., Kangas, J., 2005. Rank reversals in multi-criteria decision analysis with statistical modelling of ratio-scale pairwise comparisons. J. Oper. Res. Soc. 56, 855–861. http://dx.doi.org/10.2307/4102186.

Li, X., Han, J., Lee, J.-G., Gonzalez, H., 2007. Traffic density-based discovery of hot routes in road networks. In: Papadias, D., Zhang, D., Kollios, G. (Eds.), Advances in Spatial and Temporal Databases, Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp. 441–459.

Long, J., Gao, Z., Orenstein, P., Ren, H., 2012. Control strategies for dispersing incident-based traffic jams in two-way grid networks. IEEE Trans. Intell. Transp. Syst. 13, 469–481. http://dx.doi.org/10.1109/TITS.2011.2171035.

Luxburg, U.von, Williamson, R.C., Guyon, I., 2012. Clustering: science or art? J. Mach. Learn. Res. – Proc. Track 27, 65–80.

Martin, P.T., Perrin, J., Hansen, B., 2001. Incident Detection Algorithm Evaluation. Utah Department of Transportation.

Matlab, 2013. Quantiles and Percentiles – MATLAB [WWW Document]. URL: <http://www.mathworks.co.uk/help/stats/quantiles-and-percentiles.html> (accessed 8.5.13).

Mela, K., Tiainen, T., Heinisuo, M., 2012. Comparative study of multiple criteria decision making methods for building design. Adv. Eng. Inform. 26, 716–726. http://dx.doi.org/10.1016/j.aei.2012.03.001.

Noland, R.B., Polak, J.W., 2002. Travel time variability: a review of theoretical and empirical issues. Transp. Rev. 22, 39–54. http://dx.doi.org/10.1080/01441640010022456.

OECD, 2007. Managing Urban Traffic Congestion. OECD Publishing.

Ozbay, K., Kachroo, P., 1999. Incident management in intelligent transportation systems. Fac. Publ. ECE.

Parkany, E., Xie, C., 2005. A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't.

Pu, W., 2011. Analytic relationships between travel time reliability measures. Transp. Res. Rec. J. Transp. Res. Board 2254, 122–130. http://dx.doi.org/10.3141/2254-13.

Robinson, S., Polak, J., 2006. Overtaking rule method for the cleaning of matched license-plate data. J. Transp. Eng. 132, 609–617. http://dx.doi.org/10.1061/(ASCE)0733-947X(2006) 132:8(609.

Sen, R., Maurya, A., Raman, B., Mehta, R., Kalyanaraman, R., Vankadhara, N., Roy, S., Sharma, P., 2012. Kyun queue: a sensor network system to monitor road traffic queues. In: Proceedings of the 10th A C M Conference on Embedded Network Sensor Systems, SenSys '12. ACM, New York, NY, USA, pp. 127–140, doi:10.1145/2426656.2426670.

Šingliar, T., Hauskrecht, M., 2009. Learning to detect incidents from noisily labeled data. Mach. Learn. 79, 335–354. http://dx.doi.org/10.1007/s10994-009-5141-7.

Skabardonis, A., Varaiya, P., Petty, K., 2003. Measuring recurrent and nonrecurrent traffic congestion. Transp. Res. Rec. J. Transp. Res. Board 1856, 118–124. http://dx.doi.org/10.3141/1856-12.

Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., Tadaki, S., Yukawa, S., 2008. Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam. New J. Phys. 10, 033001. http://dx.doi.org/10.1088/1367-2630/10/3/033001.

TfL, 2010. Travel in London Report 3.

TfL, 2012. London Streets [WWW Document]. URL: <http://www.tfl.gov.uk/assets/downloads/corporate/london-streets-factsheet.pdf> (accessed 1.30.13).

TfL, 2013. Key facts [WWW Document]. URL: <http://www.tfl.gov.uk/corporate/modesoftransport/londonunderground/1608.aspx> (accessed 1.30.13).

Thomas, T., van Berkum, E.C., 2009. Detection of incidents and events in urban networks. IET Intell. Transp. Syst. 3, 198–205. http://dx.doi.org/10.1049/iet-its:20080045.

Transport Canada, 2007. The Cost of Urban Congestion in Canada.

Triantaphyllou, E., Mann, S.H., 1989. An examination of the effectiveness of multi-dimensional decision-making methods: a decision-making paradox. Decis. Support Syst. 5, 303–312. http://dx.doi.org/10.1016/0167-9236(89)90037-7.

Tsapakis, I., Heydecker, B.G., Cheng, T., Anbaroglu, B., 2013. How tube strikes affect macroscopic and link travel times in London. Transp. Plan. Technol. 36, 109–129. http://dx.doi.org/10.1080/03081060.2012.745766.

Tukey, J.W., 1977. Exploratory Data Analysis, first ed. Pearson.

Varaiya, P., 2007. Finding and Analyzing True Effect of Non-recurrent Congestion on Mobility and Safety.

Weil, R., Wootton, J., García-Ortiz, A., 1998. Traffic incident detection: sensors and algorithms. Math. Comput. Model. 27, 257–291. http://dx.doi.org/10.1016/S0895-7177(98)00064-8.

Wright, C., Roberg-Orenstein, P., 1999. Simple models for traffic jams and congestion control. In: Proceedings of the ICE – Transport. Presented at the Proceedings of the Institution of Civil Engineers: Transport, pp. 123–130.

Yuan, F., Cheu, R.L., 2003. Incident detection using support vector machines. Transp. Res. Part C: Emerg. Technol. 11, 309–328. http://dx.doi.org/10.1016/S0968-090X(03)00020-2.

Zanakis, S.H., Solomon, A., Wishart, N., Dublish, S., 1998. Multi-attribute decision making: a simulation comparison of select methods. Eur. J. Oper. Res. 107, 507–529. http://dx.doi.org/10.1016/S0377-2217(97)00147-1.