JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2014; **48**:1000–1016
Published online 19 July 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.1241

Empirical assessment of urban traffic congestion

Andy H.F. Chow[1,4]*, Alex Santacreu[2], Ioannis Tsapakis[3], Garavig Tanasaranond[4] and Tao Cheng[4]

[1]*Centre for Transport Studies, University College London, London WC1E 6BT, U.K.*
[2]*Transport for London, 197 Blackfriars Road, London SE1 8NT, U.K.*
[3]*Texas Transportation Institute, Texas A&M University, San Antonio, TX 78213, U.S.A.*
[4]*SpaceTime Lab, Civil, Environmental and Geomatic Engineering, University College London, London WC1E 6BT, U.K.*

SUMMARY

This paper presents an empirical assessment of urban traffic congestion in Central London, UK. Compared with freeways or motorways, urban networks are relatively less studied because of its complexity and availability of required traffic data. This paper introduces the use of automatic number plate recognition technology to analyze the characteristic of urban traffic congestion in Central London. We also present the use of linear regression to diagnose the observed congestion and attribute them to different causes. In particular, we distinguish the observed congestion into two main components: one due to recurrent factors and the other due to nonrecurrent factors. The methodologies are illustrated through a case study of Central London Area. It is found that about 15% of the observed congestion in the region is due to nonrecurrent factors such as accidents, roadwork, special events, and strikes. Given the significance of London, the study will be valuable for transport policy evaluation and appraisal in other global cities. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS: automatic number plate recognition (ANPR); London Traffic Information System (LTIS); visualization; linear regression; congestion pie

1. INTRODUCTION

Understanding the nature and causes of urban congestion is a prerequisite for deriving transport policies and management plans. Urban congestion is a complex phenomenon that may be due to a number of factors apart from excess traffic volume [1]. As an illustration, it is found over the last decade that congestion has been increasing in Central London despite the falling traffic levels. Possible reasons for this may include the removal of road network capacity for general traffic by an increase in utility and development works and for other policy initiatives targeted at road safety, public transport, cyclist, and pedestrian priority measures and urban realm improvements, and others [2]. In general, factors causing congestion can be broadly classified into two categories: recurrent and nonrecurrent. Recurrent congestion refers to congestion that happens on a regular (e.g., daily) basis. Causes of recurrent congestion include excess traffic, physical capacity limitation, and operations of networks (e.g., traffic signal control). Causes of nonrecurrent congestion include traffic incidents such as accidents, vehicular breakdowns, police checks, closures due to roadwork, special events such as sport games and strikes, and adverse weather.

Previous studies on congestion assessment have been focusing on freeway traffic (see for example, PeMS [3] in California and MIDAS [4] in England), and Kwon, *et al*. [5] provided a brief review on relevant studies. Nevertheless, relatively less research has been carried out on urban network, which is

*Correspondence to: Andy Chow, Centre for Transport Studies, University College London, London WC1E 6BT, U.K.
E-mail: andy.chow@ucl.ac.uk

due to the lack of required traffic data and complexity of urban network traffic. Recently, the increasing availability of data from different sources has enabled us to carry out more comprehensive research on urban network traffic. Robinson [6] and Krishnamoorthy [7] developed various models for estimating and predicting journey times in urban network using automatic number plate recognition (ANPR) system and loop detector data. Nevertheless, these studies focus on journey times although they do not provide overall assessment of network performance. In fact, journey times alone cannot be a system-wide measure of road network performance. Journey time along a link has to be associated with the corresponding traffic volume on that link to provide a reliable and fair system measure.

This paper presents an empirical approach to investigate the nature and causes of urban congestion. Traffic data that we use include journey time estimates from ANPR system; Oyster smart card counts, and incident and weather records. With the traffic data and information, we first derive a system-wide performance measure, which we call flow-weighted total journey time (FTJT). We then introduce the use of visualization to retrieve the spatiotemporal feature of urban traffic. A regression-based method is proposed to diagnose the observed journey times and attribute them to different causes. In particular, we distinguish observed congestion into two main categories: one due to recurrent factors and the other due to nonrecurrent factors. The methodologies are illustrated through a case study of Greater London Area. It is found that about 15% of the observed congestion in the region is due to nonrecurrent factors such as accidents, roadwork, special events, and strikes. The study presented herein will be useful for transport policy evaluation and appraisal.

The paper is organized as follows: Section 2 introduces the traffic data that we use in this study. Section 3 presents a regression-based approach for diagnosing the urban congestion and attributing it to a range of factors. Finally, Section 4 gives some concluding remarks.

2. URBAN TRAFFIC DATA—CENTRAL LONDON, UK

Figure 1 shows the principal London road network. The network consists of two subnetworks: Transport for London (TfL) Road Network and Borough Principal Road Network. The former is a 600 km set of main roads in the Greater London Area, and TfL is the authority managing the network. The latter comprises 1100 km of the most important London roads managed by local boroughs. The total length of London road network is about 15 000 km [8]. Hence, the combination of the two accounts for more than 11% of London roads. On a weekday, between 07:00 and 19:00, the principal London road network carries 35 million vehicle-kilometers, which accounts for 54% of the London traffic.

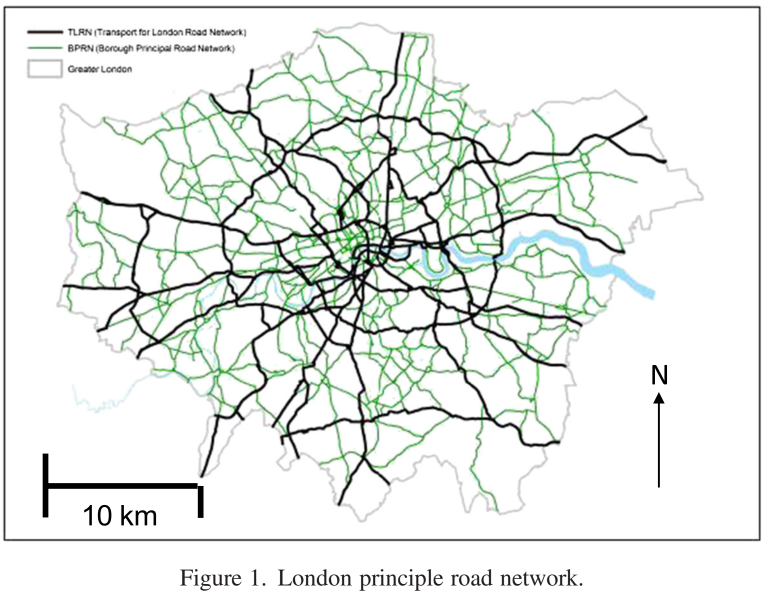


Figure 1. London principle road network.

DOI: 10.1002/atr

Several transport policies in London, including congestion charge and low emission zone schemes, are enforced by using the ANPR technology. Under the ANPR system, there are around 500 camera sites in Greater London Area [8]. The plate numbers of vehicles passing the ANPR cameras are recognized and recorded along with the corresponding times, which are used to decide whether the vehicles detected have paid the charge. The journey times of vehicles between two ANPR camera sites are then estimated by matching the license plate numbers. The journey times are processed and stored in 5-minute averages. The journey time data can be used to derive various performance metrics such as speeds, degree of saturation, journey time reliability, and impacts of major events including strikes [9] and Olympics [10]. This enables the road operator, TfL, to calculate the benefits and costs associated with different policies or control plans to improve the day-to-day operation of the network.

Nevertheless, it is noted that various errors may arise in matching the license plate numbers because of various reasons such as misreading of license plates, vehicles stopping en route, and vehicles taking unusual long route between the two camera locations [11]. Consequently, a set of data filtering and processing rules is adopted to improve the journey time estimation. For example, the overtaking rules described in Robinson and Polak [11] are used to eliminate the data noise caused by camera errors and delivery vehicles stopping along the route. Information from the Driver and Vehicle Licensing Agency is used to eliminate the data related to unauthorized vehicles on the bus lanes. In some occasions, data may be missing over some time intervals because there is no sample (e.g., no vehicles can be matched during the time interval) or failure of hardware system. A range of patching algorithms will be used to impute the missing data in those circumstances. The associated ANPR journey time data is flagged with a code referring to the type of patching mechanism that is applied. This type or level of patching ranges from 0 (best) to 3 (worst) as follows:

(0) all travel times are derived from field observations, and no patching is applied;
(1) observation is missing for only one time interval (5 minutes), and the missing value is taken as the value measured at the previous time interval;
(2) observations are missing for 2–6 time intervals (10–30 minutes), the missing values during those 2–6 time intervals are derived by using a linear interpolation between the two observations made at the start and end of the missing period;
(3) observations are missing for more than 6 time intervals (30 minutes). Missing data are patched by historical data measured during the same time period. For example, say data are missing during 07:00–07:45 on a Monday, then the missing data will be replaced with historical *measured* data during 07:00–07:45 on a previous Monday.

In this study, a set of 270 road sections, which are called the *core links* by TfL, is selected from the principal network (Figure 2). These core links consist of 90% of the Level-A arterial roads, which are called *A-road*, including major avenues such as City Road (A501), Western Avenue (A40), Park Lane, A13 Avenue, and Blackwall Tunnel (A12). These roads are under the most scrutiny and provide the best quality journey time estimates from the ANPR system. The journey time data adopted in this study are collected from this set of links. Moreover, we only include sections lying within the Central London area after considering the difference in travel pattern between Central London and Outer London as revealed in Figure 4. To ensure the best data quality, the patching level of each selected road sections does not exceed Level 2 (i.e., there are no more than six consecutive time intervals with missing data). In fact, 70% of the selected links only have a patching Level 1.

With the journey time data, we can obtain a picture of traffic pattern at different spatiotemporal scales through visualization where we can identify the locations and times of activation of the bottlenecks or pinch-points in urban networks. Pinch-points refer to locations where congestion starts and from where congestion spills over to other locations. A pinch-point is usually due to a number of reasons such as junctions, lane drops or closures, grade changes, curvature, poor traffic control, and occurrence of incidents. Some algorithms have been proposed to identify pinch-points on motorways (see for example, [12]). Contour plot of traffic density (e.g., [13]) is another useful tool to identify locations of pinch-points and congestion along a single route of travel over a space–time domain.

DOI: 10.1002/atr

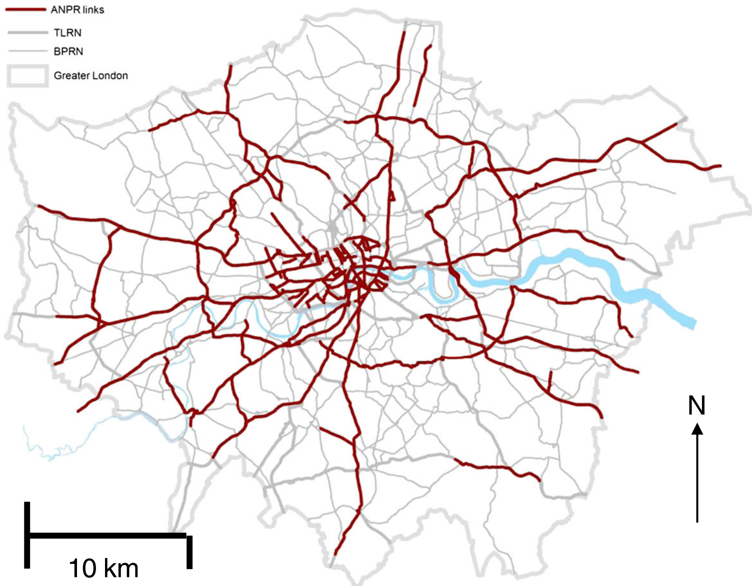


Figure 2. Core automatic number plate recognition links.

In this paper, we produce a thematic map [14] to visualize the traffic pattern over the core links on the London road network based upon the ANPR journey times. Figure 3 shows the thematic maps of journey times in Central London. The maps are generated by traffic data collected on 24 November 2010 (Wednesday) at three different times: 08:00, 10:00, and 17:00. The color scale on the maps represents the unit journey time (in min/km), which is the reciprocal of travel speed. Regions showing a high unit journey time (i.e., low speed) correspond to the congested regions. From the thematic maps, we can observe the spatiotemporal feature and movement of congestion over the area. We should note that spacings between pairs of ANPA cameras is typically over 1 km. As a result, ANPR data does not capture fine spatial feature of traffic. Currently, we are exploring the use of various statistical techniques to incorporate additional information (e.g., probe vehicle data) to refine the spatial granularity. Details of the development of the relevant techniques can be referred to Qiu, *et al*. [15], Bolbol, *et al*. [16], and Chow, *et al*. [17].

Analysis of the journey time date also provide us with important insights on travel behavior and management policies. Figure 4 shows the speed profiles on the London congestion charging zone during weekdays (Figure 4a) and weekends (Figure 4b). The broken lines are the weekday and weekend speed averages derived from ANPR journey time data collected in the zone from 29 July to 14 August 2011; the solid lines are the speed averages derived from data collected in the zone from 27 July to 12 August 2012. We highlight that the period from 27 July to 12 August 2012 is when the Olympics was being held in London. Comparing the difference between Figures 4a and 4b, we first note that the difference between weekday and weekend pattern is a clear speed drop during 06:00–20:00 due to traffic of commuters or other commercial activities. We also notice that the speed profiles show less variation during weekend although the congestion charging scheme is not in operation during weekends. It is also interesting to note that traffic was actually traveling faster during the Olympics period. It was indeed due to the management policies operated by the London transport authorities that include active traffic management, provision of alternative modes (e.g., subway), and restriction on roadwork (see [2]). From other empirical studies on Oyster (a smart card system adopted in London for public transport transactions), we also note a significant change in travel behavior that led to the improvement of road network performance during the Olympics period (see [10]).

3. DIAGNOSIS OF URBAN CONGESTION

In the previous section, we note that the performance of urban road network can be monitored by innovation applications of technology. We also notice that there is a potential for improving urban

DOI: 10.1002/atr

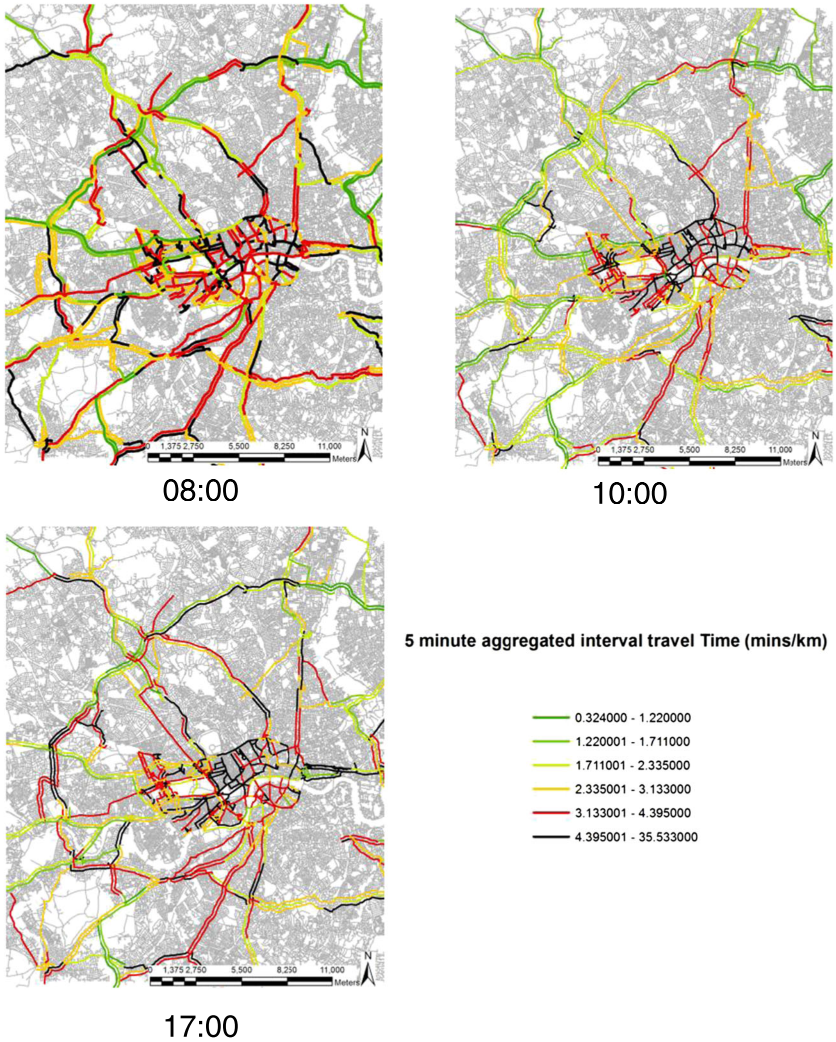


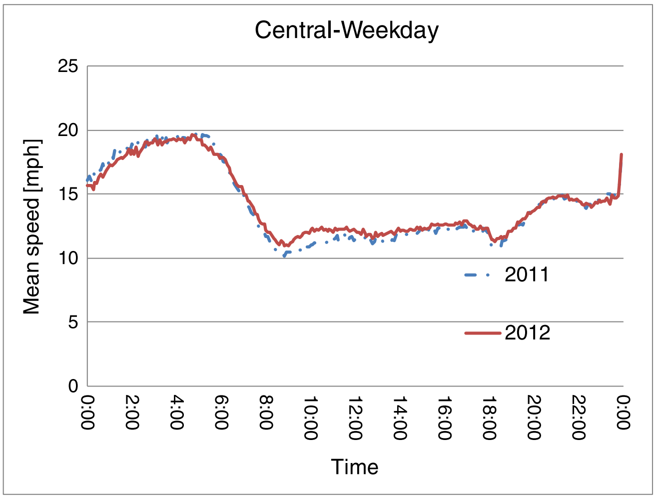
Figure 3. Thematic map of journey times in Central London area.

networks, even one as busy as Central London, through appropriate transport policies. An important step to deploy an appropriate policy is to understand the causes of the observed congestion. The paper presents a regression-based method to diagnose observed congestion and attribute them into different causes.

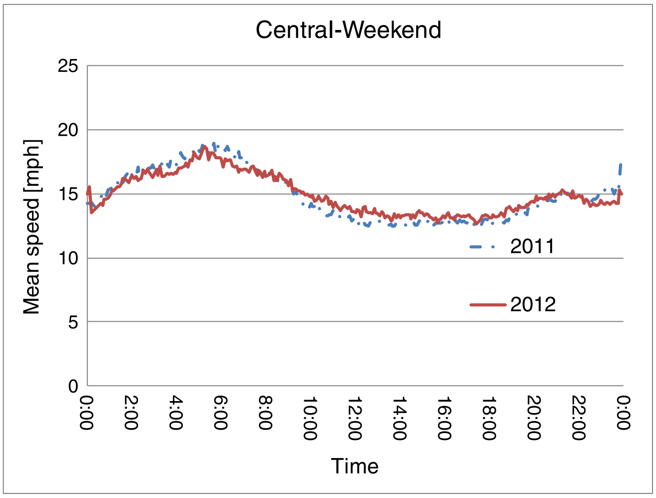
3.1. Measure of congestion—flow-weighted total journey times

We first need an indicator to quantify the degree of congestion. In this study, we propose to associate the journey times estimated by ANPR in the network with the corresponding link traffic volume to derive a system-wide measure. Due to the deployment and configuration of the urban detectors, we do not have direct measurements of traffic volumes on the ANPR links. Nevertheless, UK Department for Transport (DfT) carries out hourly manual counts on major roads in Greater London Area every 2 years. Based on the DfT's manual hourly counts, TfL derives a flow measure called annual average peak flow (AAPF) for each ANPR link. Although the flow information is inferred from a relatively small sample size (a 12-hour period from 07:00–19:00 over a two-year span), they are very accurate and provide an excellent granularity both in space and time.

DOI: 10.1002/atr



a) Weekday



b) Weekend

Figure 4. Journey time profiles in Central London in Summers 2011 and 2012.

With the AAPF and ANPR journey time data, we derive a system-wide measure of congestion that we call FTJT (unit in veh-hrs), $Y_{total}(d)$, for the entire Central London road network on each day d as follows:

$$Y_{total}(d) = \sum_{i=1}^{270} \sum_{t=7}^{18} \alpha_i(t) JT_i(t, d) \tag{1}$$

where t is the time index in hour. For example, $t = 7$ presents the period from 07:00–07:59, and so on. The subscript i is the index for the ANPR links in Central London area that we consider, where $i = 1,2,\ldots,270$. The notation $\alpha_i(t)$ denotes the AAPF obtained from the DfT's biannual survey on link i during the hour t; $JT_i(t,d)$ is the average ANPR journey time through link i during the hour t on day d. Consequently, $Y_{total}(d)$ is a weighted sum of journey times collected from the 270 selected major road sections in Central London area on a given day d.

As an illustration, Figures 5 and 6 show respectively the daily and monthly FTJT variations during the years 2009 and 2010. Each box represents the statistics over the corresponding dates or months. The central bars of the boxes refer to the median values, and the heights of the boxes represent the

DOI: 10.1002/atr

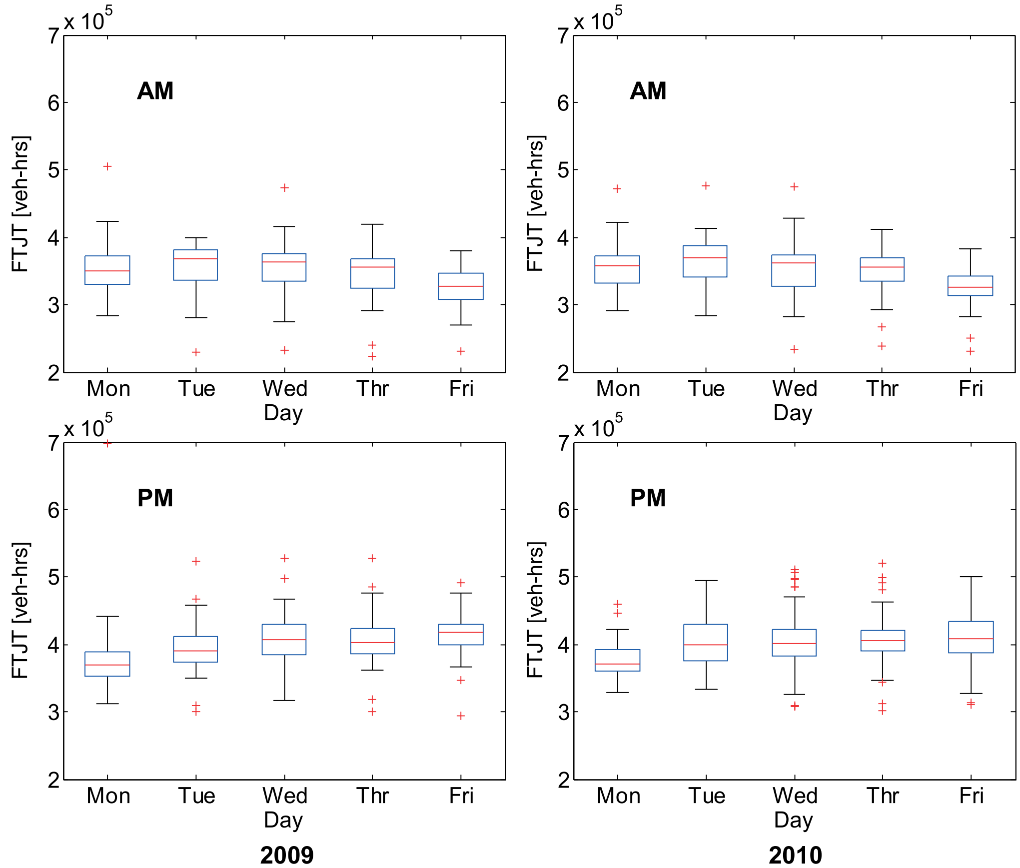


Figure 5. Daily variations of the flow-weighted total journey time [veh-hrs] over week.

interquartile range (i.e., from 25th to the 75th percentiles) of the data. The whiskers extend to the furthest data point that is no more than 1.5 times the interquartile range from the box. Any data point outside of the whiskers is plotted individually as the "+" sign. In Figures 5 and 6, "AM" refers to time period 07:00–10:00 and "PM" is 16:00–19:00. Transport for London identifies these two periods as the morning peak and evening peak in London, respectively.

Some interesting features noted from the Figures 5 and 6 include the following:

- the daily and monthly variations are similar for the 2 years. This suggests that there is no significant change in travel behavior and network operation across the 2 years;
- afternoon peak shows a higher FTJT than morning peak, which is due to more traffic on the road during 16:00–19:00 than 07:00–10:00;
- August shows the lowest average FTJT for both years. It may be due to the reduction in traffic volume because of the schools' summer vacation;
- December shows the highest FTJT variability, which may be due to complicated travel behavior in that month (the Christmas month).

3.2. Causes of congestion

Given the journey time measures, this study attempts to attribute them into a range of causes. The factors and information that we consider in the present study include the following.

3.2.1. Travel demand

Demand for travel is a major factor determining journey times and congestion on the road. Road traffic volume is a widely used indicator of travel demand, although we do not have reliable and systematic measure of traffic flow in our urban network. Nevertheless, we have access to Oyster count data of bus

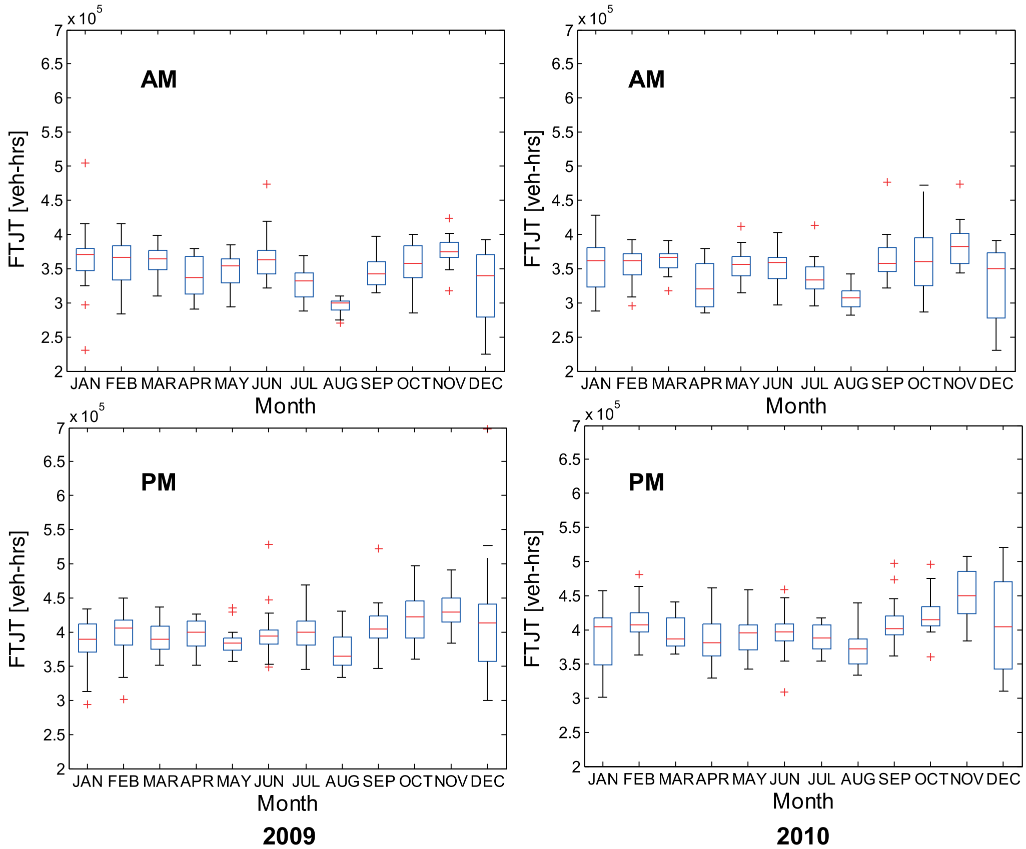


Figure 6. Monthly variations of the flow-weighted total journey time [veh-hrs] over year.

users. The dataset records the number of transactions made by Oyster cards on buses on a day. Bus ridership accounts for about 16% of the road trips made in Central London area (as of 2012), and this share is steady over the years [2]. Given the high penetration rate of Oyster, the bus Oyster count is used here as a proxy variable for travel demand. For example, a day associated with a high value of bus Oyster count implies that day should also be heavy with travel demand.

3.2.2. Incidents

In addition to travel demand, we also have access to various incident logs through the London Streets Traffic Control Centre (LSTCC). The LSTCC is monitoring and managing congestion on the London road network with over 1300 traffic cameras to detect incidents. Real-time information from the LSTCC is shared with members of the transport industry and the media through the London Transport Information System (LTIS). The LTIS database contains the following information for use in our analysis:

- Accidents
  - In general, it records the total duration (measured in [Event-Hrs]) affected by all accidents on a given day.
  - We further classify this category into "Moderate Accidents" and "Serious Accidents"
  - "Moderate Accidents" is the total duration affected by all accidents whose severity levels are classified as "moderate" or less.
  - "Serious Accidents" is the total duration affected by all accidents whose severity levels are classified as "serious" or more.
- Breakdowns
  - The total time duration affected by vehicular breakdowns on each day.

DOI: 10.1002/atr

- Obstruction
  - This category is used to record the duration affected by blockages or disruptions apart from broken vehicles. For example, fallen down tree or control devices' (e.g., traffic signals) failure.
- Police security checks
  - The total durations of police security checks on each day
- Special events
  - The numbers and durations of events happened on each day. Examples of events include filming, car race, carnival, State Opening of Parliament, demonstrations, and football matches.
- Roadwork
  - The information of road works and utility works is provided by Transport for London. The log consists of the numbers and total durations affected by different types of roadwork.

3.2.3. Weather

The weather information is obtained from the MET office through the British Atmospheric Data Centre. The weather information includes precipitation (measured in mm) and whether there is snow on a particular day. Note that the MET office also provides temperature information, although that is not used in our study.

3.2.4. Other factors

Considering its significance (see [9]), we also keep a record of tube (London underground) strikes and include that in our analysis.

3.3. Regression analysis

Given the data, this study proposes a regression analysis approach to decompose the observed congestion into various components and attribute them to associated factors.

3.3.1. Variables

The congestion is measured in terms of FTJT, which is denoted as the dependent variable "Y" in our regression model.

With the data and information presented in previous sections, we generate the following set of explanatory variables $\mathbf{X}(d)$ on each day d:

$X_{oyster}(d)$ is total number of bus oyster counts recorded on day d;
$X_{acc\_\mathrm{mod}}(d)$ is total duration (Event-Hrs) of "moderate" accidents on day d;
$X_{acc\_sns}(d)$ is total duration (Event-Hrs) of "serious" accidents on day d;
$X_{brk}(d)$ is total duration (Event-Hrs) affected by vehicular breakdowns on day d;
$X_{obs}(d)$ is total duration (Event-Hrs) affected by road obstruction on day d;
$X_{roadwork}(d)$ is the total duration (Event-Hrs) of road work on day d;
$X_{event}(d)$ is the total duration (Event-Hrs) of special events on day d;
$X_{police}(d)$ is the total duration (Event-Hrs) of police security checks on day d;
$X_{rain}(d)$ is the total precipitation (mm) measured on day d;
$X_{strike}(d)$ is a 0–1 indicator that equals to 1 if tube strike on day d; 0 otherwise;
$X_{snow}(d)$ is a 0–1 indicator that equals to 1 if snow on day d; 0 otherwise.

We adopt data collected on regular days (i.e., weekdays excluding holidays) in years 2009 and 2010. Considering the differences in the origin–destination trips made and the spatial distribution of traffic in morning and afternoon periods, the regression model is calibrated separately for the AM (07:00–10:00) and PM (16:00–19:00) periods. We do not calibrate the model differently for different years as we do not expect and actually observe any significant difference between years. Moreover, considering the significant difference between travel pattern between Central London and Outer London, also the lack

DOI: 10.1002/atr

of data in outer area, this study will be confined to Central London area, which includes boroughs of City of London, Camden, Chelsea, Hackney, Islington, Kensington, Lambeth, Southwark, Tower Hamlets, and Westminster.

3.3.2. Model

We then formulate a linear regression model in which congestion observed in the area on a day d is a function of a number of factors as follows:

$$Y_{total}(d) = \beta_0 + \boldsymbol{\beta}^T \cdot \mathbf{X}(d) + \varepsilon(d), \tag{2}$$

where

$Y_{total}(d)$ is the total observed FTJT in Central London area on day d;
$\mathbf{X}(d)$ is a column vector of the aforementioned explanatory variables;
$\boldsymbol{\beta}$ is a column vector, which has the same dimension as $\mathbf{X}(d)$, of the model parameters for the associated explanatory variables. The superscript "T" on $\boldsymbol{\beta}$ represents a transposed vector;
β_0 is the value of the intercept of the regression model. This is regarded as the portion of congestion that is due to the "unexplained" factors in the model;
$\varepsilon(d)$ is an error term following normal distribution with mean zero and a finite variance.

3.3.3. Regression results

The model parameters are estimated by using least-square method, and the results are summarized in Table I. Here are some basic information for reading the table.

- The third column (coeff.) shows the estimated values of β_0 (intercept), $\boldsymbol{\beta}$ (coefficients of different factors). The intercept β_0 is the amount of FTJT that to be observed when all explanatory variables are zero. This component is the portion of FTJT that is not represented by the factors that we include. We reckon this portion of FTJT as the *base* FTJT, which are journey times that are there anyhow the network is operated.

Table I. Regression model results.

Scenario	Factors	Coeff. (β)	Std. Error	t-value	p-value	R^2
AM	(Intercept)	211 343.4	6338.1	33.345	0.000 ***	0.61
	Bus Oyster counts	0.0628	0.0020	31.528	0.000 ***	
	Moderate Accidents	3102.3	695.2	4.463	0.000 ***	
	Serious Accidents	6558.2	1313.8	4.992	0.000 ***	
	Vehicular breakdowns	4402.3	967.7	4.549	0.000 ***	
	Obstruction	149.2	317.0	0.471	0.633	
	Roadwork	161.4	74.2	2.176	0.003 **	
	Special events	499.8	165.6	3.017	0.001 ***	
	Police checks	−955.2	1381.8	−0.691	0.492	
	Strike	99972.6	23429.8	4.267	0.000 ***	
	Snow	−11232.4	8126.5	−1.382	0.169	
	Precipitation	1502.3	5142.9	0.292	0.789	
PM	(Intercept)	227837.3	8124.9	28.042	0.000 ***	0.62
	Bus Oyster counts	0.0759	0.0026	29.216	0.000 ***	
	Moderate Accidents	2742.4	695.7	3.942	0.000 **	
	Serious Accidents	6050.1	1015.2	5.960	0.000 **	
	Vehicular breakdowns	3393.9	1085.4	3.127	0.000 ***	
	Obstruction	161.8	329.9	0.490	0.629	
	Roadwork	132.0	56.5	2.339	0.018 *	
	Special events	848.0	147.0	5.770	0.000 ***	
	Police checks	1255.5	1162.1	1.080	0.338	
	Strike	67268.1	18276.4	3.681	0.002 **	
	Snow	−17042.3	6884.7	−2.475	0.032 *	
	Precipitation	−112.1	1705.2	−0.066	0.945	

- The Standard Error (SE) of the parameter is the calculated as the square root of the ratio of the parameter variance to the associated sample size, which is equal to 261 (days) for all scenarios. The t-value and p-value reflect whether the parameters are significantly different from zero. The t-value is determined as the ratio of the parameter's estimate (i.e., Column 3) to the associated standard error (i.e., Column 4). The p-value of a parameter is the probability that the true value of that parameter is different from zero. Factors in the table being marked with "***", "**", "*" or "+" are regarded as *significant* factors, which refer those with a true value that is unlikely to be zero. The notation ("***", "**", "*" and "+") respectively represent a p-value between 0 and 0.001, between 0.001 and 0.01, between 0.01 and 0.05, and between 0.05 and 0.1.
- The R-square, which lies between 0 and 1, represents the goodness-of-fit, that is, how reasonable the regression model as a representation of the observed congestion. In general, the higher value of R-square, the better the regression model explains the variations of FTJT with respect to the explanatory variables.

From the table, we observe the following:

- The R-square is about 60% for both AM and PM scenarios. This is not very impressive, and the error is due to the nonlinear characteristics of congestion with respect to the factors.
- The intercept β_0, *Bus Oyster counts*, *Accidents* (both serious and moderate), *Breakdowns*, *Roadwork*, *Events*, and *Strike* are all significant factors for the observed FTJT in both scenarios. These factors are associated with a p-value < 0.005, and even <0.001 in some cases.
- *Bus Oyster counts* is recognized as a strongly significant factor. This is not surprising as the number of oyster transactions directly related to the demand for travel, which is also directly related to the level of journey times observed.
- The coefficient associated with *Serious Accidents* is generally more than twice as the one of *Moderate Accidents*. This implies that the marginal effect of a *Serious Accident* to the observed journey times is more than twice as that of a *Moderate Accident*.
- The events of *Snow* and *Strike* are rather rare as we only had 24 snow days and 3 strike days during 2009–2010. Nevertheless, *Strike* is statistically significant (with p-value ~ 0.002) in both scenarios, whereas *Snow* is only significant (with a p-value ~ 0.03) in the afternoon scenario. However, interestingly, the *Snow* variable is associated with a negative coefficient. An explanation is that snow reduced the traffic volume on the road and hence brought down the FTJT.
- Rather surprisingly, precipitation (rain) is not statistically significant. This may due to the general weather conditions in London where drivers are used to drive in wet.

3.3.4. Congestion pie

We remove all insignificant variables (i.e., those with a p-value > 0.05) and rerun the estimations until all the remaining factors are significant (with p-value > 0.05). We also remove the *Snow* factor although it has a p-value < 0.05. It is because the *Snow* factor is associated with a negative coefficient, which implies that the factor is not contributing to the observed congestion.

Given the regression model (2) and the calibrated coefficient $\boldsymbol{\beta}$, the total observed FTJT, Y, can be decomposed into the following components:

$$Y_i(d) = \beta_i X_i(d) \tag{3}$$

where $Y_i(d)$ is regarded as the portion of $Y_{total}(d)$ due to factor i on day d. We can also derive the average values of each Y_i over multiple days (e.g., a year) as

$$\overline{Y}_i = \beta_i \overline{X}_i \tag{4}$$

where $\overline{Y}_i$ and $\overline{X}_i$ are respectively the average values of Y_i and X_i over the days of interest.

Furthermore, the component associated with the intercept term is calculated as

$$\beta_0 = \overline{Y}_{total} - \sum_i \beta_i \overline{X}_i \tag{5}$$

and it is reckoned as the base FTJT as discussed.

Table II shows the estimates of the FTJT components using the method presented earlier. Following previous regression analysis, the (significant) factors i considered herein include the following: bus oyster counts (which we interpret as travel demand), moderate and serious accidents, vehicular breakdowns, roadwork, special events (e.g., sport games and carnival), strikes, and the portion of nominal FTJT (i.e., the intercept).

We consider the following four scenarios: 2009 AM, 2009 PM, 2010 AM, and 2010 PM. We note that the "base" FTJT accounts for around 60% in all scenarios, which suggests that 60% of the observed FTJT is due to normal network operation. Moreover, the *demand* factor accounts for 25% of the observed F-JT in the morning; 30% in the afternoon. The higher proportion in the afternoon is due to the higher traffic volume in the afternoon. We classify the portions of journey times due to base operations and travel demand as *recurrent congestion*, as they will be observed daily. The recurrent factors account for 85% of the observed journey times in our analysis.

Hence, there is about 15% of the observed journey times is due to the *nonrecurrent factors*. The nonrecurrent factors include accidents, vehicular breakdowns, roadwork, special events, and strikes. These are factors that occur occasionally and perhaps unexpectedly. This suggests that 15% of the urban congestion in London could be managed or reduced if we have better management of incidents or events.

Table II. Diagnosis of flow-weighted journey times.

Scenario	Factors	Coeff. (β)	Average Value	FTJT contribution	% of FTJT
2009 AM	Base FTJT (intercept)	217 606.6	N/A	217 606.6	62.29
	Demand (oyster counts)	0.0640	1 334 551	85 398.5	24.45
	Moderate accidents	3042.2	2.7162	8263.2	2.37
	Serious accidents	6759.9	0.5976	4039.7	1.16
	Vehicular breakdowns	4308.6	1.3695	5900.7	1.69
	Roadwork	147.2	119.1362	17 539.0	5.02
	Special events	523.1	18.7523	9809.4	2.81
	Strike	99 564.7	0.0079	786.6	0.23
2009 PM	Base FTJT (intercept)	228 616.6	N/A	228 616.6	56.63
	Demand (oyster counts)	0.0747	1 631 117	121 893.1	30.19
	Moderate accidents	2767.1	3.5597	9850.0	2.44
	Serious accidents	6112.0	0.8543	5221.5	1.29
	Vehicular breakdowns	3378.4	1.8205	6150.4	1.52
	Roadwork	121.5	122.0277	14827.6	3.67
	Special events	843.0	19.6733	16 585.0	4.11
	Strike	70 116.3	0.0079	553.9	0.14
2010 AM	Base FTJT (intercept)	217 606.6	N/A	217 606.6	62.40
	Demand (oyster counts)	0.0640	1 362 875	87210.9	25.01
	Moderate accidents	3042.2	3.3053	10 055.4	2.88
	Serious accidents	6759.9	0.7367	4980.0	1.43
	Vehicular breakdowns	4308.6	1.9322	8325.1	2.39
	Roadwork	147.2	88.9313	13 092.3	3.75
	Special events	523.1	13.5103	7067.3	2.03
	Strike	99 564.7	0.0040	393.5	0.11
2010 PM	Base FTJT (intercept)	228 616.6	N/A	228 616.6	56.91
	Demand (oyster counts)	0.0747	1 665 736	124 480.2	30.99
	Moderate accidents	2767.1	4.0836	11 299.7	2.81
	Serious accidents	6112.0	1.0206	6237.9	1.55
	Vehicular breakdowns	3378.4	2.2475	7593.1	1.89
	Roadwork	121.5	91.4531	11 112.5	2.77
	Special events	843.0	14.3416	12 090.3	3.01
	Strike	70 116.3	0.0040	277.1	0.07

DOI: 10.1002/atr

Table III. Correlation between explanatory variables.

AM								
	Bus oyster counts	Moderate accidents	Serious accidents	Breakdowns	Roadwork	Special events	Strike	Weather (rain)
Bus Oyster counts	1.00							
Moderate accidents	0.15	1.00						
Serious accidents	0.15	0.13	1.00					
Breakdowns	0.11	0.18	0.26	1.00				
Roadwork	−0.05	0.29	0.28	−0.06	1.00			
Special events	−0.11	0.04	−0.10	−0.05	−0.27	1.00		
Strike	−0.09	0.12	0.09	0.00	−0.10	−0.05	1.00	
Weather (rain)	0.03	−0.05	−0.03	0.00	0.07	−0.04	0.02	1.00
PM								
	Bus oyster counts	Moderate accidents	Serious accidents	Breakdowns	Roadwork	Special events	Strike	Weather (rain)
Bus oyster counts	1.00							
Moderate accidents	0.17	1.00						
Serious accidents	0.12	0.18	1.00					
Breakdowns	0.10	0.19	0.13	1.00				
Roadwork	0.01	0.24	0.17	−0.05	1.00			
Special events	0.03	0.00	−0.01	0.00	−0.15	1.00		
Strike	0.09	0.05	0.03	−0.05	0.02	−0.01	1.00	
Weather (rain)	0.04	0.02	0.02	−0.04	0.01	−0.04	−0.04	1.00

DOI: 10.1002/atr

We also observe that there is more roadwork conducted in Year 2009, hence roadwork accounts for a higher proportion of F-JT in that year than 2010. Finally, although it is statistically significant, the effect of "Strikes" is almost negligible in the entire picture. It is because there were only three "Strikes" events in the 2 years (525 days). The present study does not provide much insight on the effect of tube strike. Detailed analysis of tube strike on London road traffic is reported in Tsapakis, *et al*. [9].

When using multiple regression, an important statistical phenomenon called *multicollinearity* should be investigated. Multicollinearity refers to the correlation among the explanatory variables. This is particular important for our study. Say, one may expect the number of serious accidents to be correlated with the number of moderate accidents. Should that be the case, it will be inappropriate to claim in our previous section that "the impact of severe accidents is more than twice as important as that of the moderate accidents." Likewise, there may be a certain degree of correlation between the weather condition and the number of accidents occurred. Again, should that be the case, it will be inappropriate to claim that weather does not have an effect on the degree of congestion observed.

There are a number of tests in statistics to investigate the degree of multicollinearity. This study adopts a simple method by looking the correlation between each pair of explanatory variables. Table III shows the correlations among the explanatory variables, which are shown to be significant in Table I. Note that we also include the *Weather* factor. It is because one may expect *Weather* can be correlated with the occurrence of incidents. We conduct the analysis separately for AM and PM scenarios. Table III shows that the degree of correlation among explanatory variables is generally not strong in which most values of correlation are less than 20%. As we observe from Table III, *Serious Accidents* is positively correlated with *moderate accidents*, but the largest correlation value is only 19% (PM scenario) which is not strong. The reason is that we are not looking at individual links but indeed the area of Central London as a whole. The accidents occurred may indeed be located at two links that are far apart and hardly have any interaction with each other. For a similar reason, the factors of *Accidents*, *Roadwork*, and *Breakdown* are also positively correlated, although the largest value of correlation that we observe among all cases is only 28%, which is not really significant from a statistical perspective. Apart from the incidents, there exists some positive correlation between *Bus Oyster Count*, *Accidents*, and

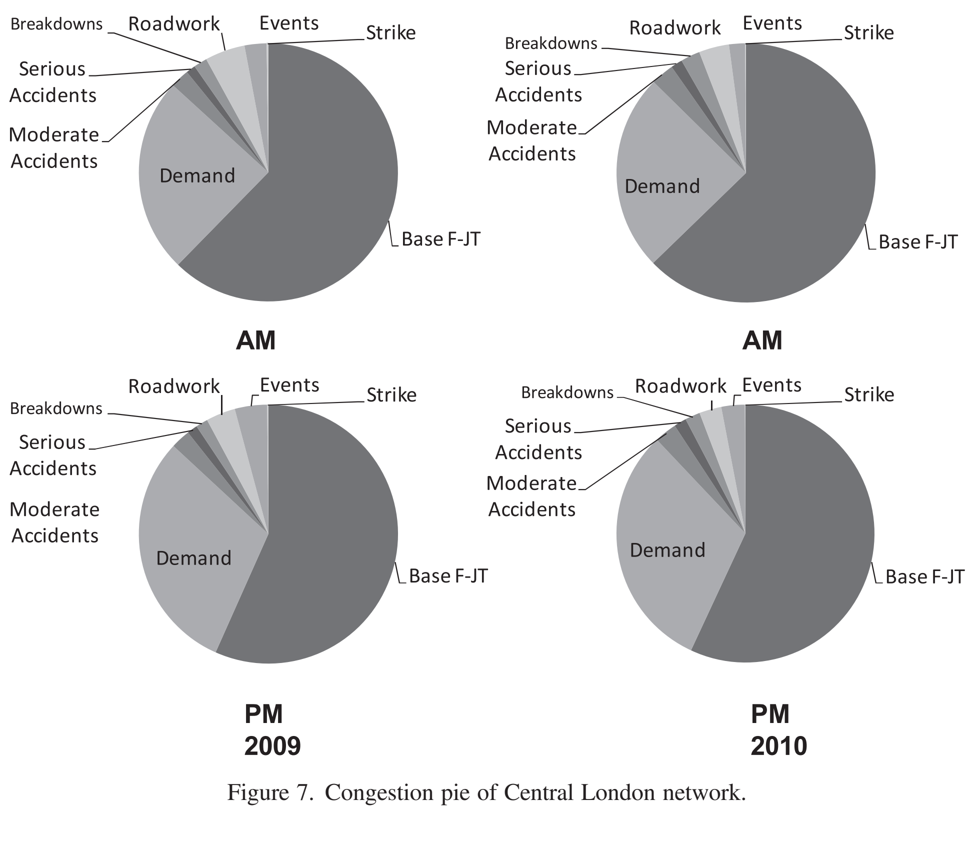


Figure 7. Congestion pie of Central London network.

DOI: 10.1002/atr

Breakdown. This is due to the increased likelihood of accidents and vehicular breakdown during busy periods. Interestingly, we do find that the numbers of roadwork and special events are negatively correlated. This may suggest that there is some coordination between roadwork and events in order to reduce the traffic impact. Overall, we do observe certain level of correlation among the explanatory variables. Nevertheless, from a statistical perspective, the level of correlation is not high, with the maximum value of correlation being less than 30%. This supports the shares of congestion computed in Table II to be reasonable estimates.

Finally, Figure 7 presents the FTJT components graphically in the form of a pie chart for all four scenarios. Kwon and Varaiya [13] regard this as a *congestion pie chart* that is a useful planning tool for designing and evaluating transport policies.

4. CONCLUDING REMARKS

The gateway to mitigating traffic congestion is to understand the nature and the causes of it. This paper presents an empirical assessment of urban congestion in Central London. The proposed methodology should be applicable to cities elsewhere apart from London with supply of relevant data. In this study, we measure congestion in terms of journey times that are obtained through ANPR system in London. Feature of urban congestion is analyzed through different visualization and statistical techniques. We also show how the journey time data can gain insight into travel behavior and impacts of special events such as strikes and Olympics. We then present a regression-based approach to diagnose the observed journey times and attribute them into different factors. Factors considered herein include demand for travel, accidents, vehicular breakdowns, roadwork, weather (precipitation and snow), and strikes. With the regression technique, we can identify the significant factors that will have an impact on the journey times. We also observe that 25%–30% of the observed journey times is due to travel demand, whereas 15% of the observed journey times in Central London is due to the nonrecurrent factors such as accidents, roadwork, special events, and strikes. This provides insights on how we should make investment on the transport infrastructure. For example, Figure 7 suggests that demand management could reduce about 25% of the observed congestion, whereas an effective incident management system could potentially reduce 15% of the congestion. Finally, the paper constructs the *congestion pie* for different scenarios that show graphically the portions of different types of congestion. The congestion pies and diagnosis method proposed in the study will be useful tools for authorities to derive and evaluate various transport policies and operational plans. The implication on policy and management can be for the Central London area where the data mainly come from. For example, the findings herein can suggest what should be the amount of roadwork that we can have in Central London considering its contribution (3–5% of the travel times) to the congestion in the area. Moreover, the findings can also suggest the level of investment that we should make on incident management and accident prevention, considering their traffic impact (about 7–8% of the travel times).

This study looks from a macroscopic perspective at the relationship between observed congestion and various factors. Nevertheless, we are aware of the heterogeneity of traffic pattern in different link types. This is the reason why we only include road sections (as stated in Section 2) in Central London as we are aware of the difference in traffic pattern between Central and Outer London areas. Nevertheless, we agree that further work needs to be performed, say exploring the use of more advanced statistical techniques such as the geographically weighted regression, to take spatiotemporal details into account and generate more specific and accurate design on policy and management.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referees for the constructive comments. An earlier version of this paper was presented at Transportation Research Board Annual Meeting 2012, Washington DC. Comments received are gratefully appreciated, and they lead to the redevelopment of the analysis and revision of results presented herein. The study is carried out under the STANDARD project that is funded by the UK Engineering and Physical Sciences Research Council under Research Grant EP/G023212/1. The authors would like to thank Transport for London for providing the traffic data and Andy Emmonds and Jonathan Turner for the constructive comments. The contents of this

report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not reflect the official views or policies of Transport for London or any other organizations.

5. LIST OF SYMBOLS AND ABBREVIATIONS

5.1. Symbols

$\alpha_i(t)$	Annual Average Peak Flow on link i during hour t
$JT_i(t, d)$	Average journey time through link i during hour t on day d
$Y_{total}(d)$	Total observed FTJT in Central London area on day d
$X_{oyster}(d)$	Total number of bus oyster counts recorded on day d
$X_{acc_mod}(d)$	Total duration (Event-Hrs) of moderate accidents on day d
$X_{acc_sns}(d)$	Total duration (Event-Hrs) of serious accidents on day d
$X_{brk}(d)$	Total duration (Event-Hrs) affected by vehicular breakdowns on day d
$X_{obs}(d)$	Total duration (Event-Hrs) affected by road obstruction on day d
$X_{roadwork}(d)$	Total duration (Event-Hrs) of road work on day d
$X_{event}(d)$	Total duration (Event-Hrs) of special events on day d
$X_{police}(d)$	Total duration (Event-Hrs) of police security checks on day d
$X_{rain}(d)$	Total precipitation (mm) measured on day d
$X_{strike}(d)$	An 0-1 indicator which equals 1 if tube strike on day d; 0 otherwise
$X_{snow}(d)$	An 0-1 indicator which equals 1 if snow on day d; 0 otherwise
$\mathrm{X}(d)$	Column vector of the explanatory variables X listed above
β	Column vector of the regression model parameters
$\varepsilon(d)$	Regression model error

5.2. Abbreviations

AAPF	Annual Average Peak Flow
ANPR	Automatic Number Plate Recondition
DVLA	Driver and Vehicle Licensing Agency
FTJT	Flow-weighted Total Journey Time
LSTTC	London Streets Traffic Control Centre
LTIS	London Traffic Information System
MET office	United Kingdom's National Weather Service
MIDAS	Motorway Incident Detection and Automated Signalling
PeMS	Performance Measurement System
TfL	Transport for London

REFERENCES

1. Federal Highway Administration (FHWA). Traffic congestion and reliability – trends and advanced strategies for congestion mitigation. Final report prepared by Cambridge Systematics and Texas Transportation Institute, 2005.
2. Transport for London (TfL). Travel in London – Report 5. 2012. URL www.tfl.gov.uk/assets/downloads/corporate/travel-in-london-report-5.pdf [Accessed on January 2013].
3. PeMS. PeMS homepage. URL http://pems.dot.ca.gov/ [accessed on January 2013].
4. MIDAS. Motorway incident detection and automated signalling (MIDAS) system. URL http://www.midasinfo.co.uk/ [accessed on January 2013].
5. Kwon J, Mauch M, Varaiya P. The components of congestion: delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Proceedings of the 85th Annual Meeting Transportation Research Board*, Washington, DC, USA, 22–26 January 2006.
6. Robinson S. The development and application of an urban link travel time model using data derived from inductive loop detectors. *Doctoral thesis*, *Imperial College London*, 2005.
7. Krishnamoorthy R. Travel time estimation and forecasting on urban roads. Doctoral thesis, Imperial College London, 2008.

DOI: 10.1002/atr

8. Transport for London (TfL). What do I need to know about the central London Congestion Charge camera system?, Technical report, Transport for London. 2011. www.tfl.gov.uk/assets/downloads/CC-Cameras.pdf [accessed on January 2013].
9. Tsapakis I, Turner J, Cheng T, Heydecker BG, Emmonds A, Bolbol A. Effects of tube strikes on journey times in the transport network of London. *Transportation Research Record* 2012; **2274**:84–92.
10. Chow AHF. London traffic during Olympics 2012. 2012. *Presentation at UCL public event*: *Transport and the Olympic Legacy*, 14 September, University College London, UK. URL www.youtube.com/watch?v=bx5U1Sd7BmE [accessed on January 2013].
11. Robinson S, Polak J. Overtaking rule method for the cleaning of matched license plate data. *ASCE Journal of Transportation Engineering* 2006; **132**(8):609–617.
12. Chen C, Skabardonis A, Varaiya P. Systematic identification of freeway bottlenecks. *Transportation Research Record* 2004; **1867**:46–52.
13. Kwon J, Varaiya P. The congestion pie: delay from collisions, potential ramp metering gain, and excess demand. *Proceedings of the 84th Annual Meeting Transportation Research Board*, Washington, DC, USA, 9–13 January 2005.
14. Cheng T, Tanaksaranond G, Emmonds A, Sonoiki D. Mutliscale visualization of inbound and outbound traffic delays in London. *The Cartographic Journal* 2010; **47**(4):323–329.
15. Qiu T, Lu X, Chow AHF, Shladover S. Estimation of freeway traffic density with loop detector and probe vehicle data. *Transportation Research Record* 2010; **2178**:21–29.
16. Bolbol A, Tsapakis I, Cheng T, Chow AHF. Sample size calculation for studying transportation modes from GPS data. *Procedia - Social and Behavioral Sciences* 2012; **48**:3040–3050.
17. Chow AHF, Scarinci R, Heydecker BG. Analysis of adaptive data fusion algorithm for urban network application. *Proceedings of the 92nd Annual meeting of the Transportation Research Board*, Washington, DC, USA, January 2013; 13–17.

DOI: 10.1002/atr