Data-Centric Engineering (2020), **xx**:xx 1–17
doi:10.1017/xxxx

DATA-CENTRIC ENGINEERING

RESEARCH ARTICLE

The use of location-based data to improve traffic forecasting in cities: a case study on volume-delay curves for London

Gerard Casey[1]*, Bingyu Zhao[2] and Kenichi Soga[2]

[1]Arup, London, UK. [2]Department of Civil and Environmental Engineering, University of California, Berkeley, USA.
*Corresponding author. E-mail: gerard.casey@arup.com

(Received xx xxx xxxx)

Keywords: Big-data crowd-sourced real-time traffic GPS modelling

Abstract

In many cities across the world traffic congestion has reached chronic levels. Despite many technological disruptions one of the most fundamental and widely used functions within traffic modelling, the volume delay function, has seen little in the way of change since it was developed in the 1960's. Traditionally macroscopic methods have been employed to relate traffic volume to vehicular journey time. These idealised functions attempt to generalise different aspects of an individual road's characteristics in order to create usable functions that do not require a large range of survey requiring inputs. The general nature of these functions enables their ease of use and gives widespread applicability. However, their ability to give results that consider the individual characteristics of a road (i.e. geometry, the presence of traffic furniture, road quality and the surrounding environment) reduces. This research investigates the use of two different data sources (crowd sourced location informed journey times and automated traffic counters) for the established models. The crowd sourced GPS informed journey times offer data that is context specific to a location. Automated traffic counters enable the harvesting of traffic count data over similarly fine temporal resolutions. By combining these two sources for different road types in London, new context specific volume-delay and speed-saturation functions can be generated. This method shows promise in some locations with the generation of robust functions. In other locations it highlights the need to better understand other influencing factors, such as the presence of on road parking or weather events.

Impact Statement

Volume delay curves are widely used in traffic analysis. They form a critical part of the traffic assignment step in 4 step modelling. More accurate representations of traffic behaviour under congestion using the data sources illustrated here may permit for a better understanding of real world context specific behaviour.

1. Introduction

When the vehicular demand for a road increases or exceeds the ability to supply, a journey time-delay is incurred as a result of the congestion generated. Relationship of traffic volume to time-delay has historically been simplified to macroscopic principles and used in traffic studies and planning (add references). The most widely adopted functional representation of such a volume-delay relationship is probably the Bureau of Public Roads (BPR) relationship proposed in the 1960s. Since then, the BPR curve coefficients have been calibrated by local traffic agencies across the world to suit the local needs. However, due to the difficulty in collecting the volume and delay information extensively (e.g., through surveys), the volume-delay functions used by the traffic engineers are often not updated frequently or

are too general (e.g., one set of coefficients for all roads in the city, not being able to consider the road furniture or on-street parking, etc.).

This research aims to ease the volume-delay curve calibration efforts by incorporating emerging data sources, mainly crowdsourced travel time information from location and routing service providers, such as Google Maps. Specifically, this paper investigates the pairing of two data sources for this purpose: (1) the crowd sourced location device informed journey times from Google Maps, and (2) the traffic count data from the Automated Traffic Counter (ATC) system in the Greater London Area (GLA). The feasibility and performance of obtaining site-specific volume-delay information based on these new data sources will be assessed in this study. In Section 2, relevant literature will be reviewed, including various functional forms of the volume-delay relationships, as well as some applications of such relationships in transport research and planning. Next, in Sections 3 and 4, the two data sources, namely the hourly traffic volume and traffic speed, are introduced. These data are used for estimating the road characteristics (free flow travel time and capacity) and calibrating the volume-delay curve coefficients. Sections 5 and 6 and explains the steps for data cleaning and model building. In particular, three models are compared in terms of their performances in capturing the delay effect at different traffic congestion states. This is followed by an extensive discussion in Section 7 on potential factors that could lead to the huge variability in the volume-delay relationship as seen in the real world, which is nearly impossible to be fully accounted for and certainly insufficient when representing this relationship using a single function for the whole city. As a result, it is important to understand the errors in prediction from the observations. Through the use of fine resolution data in the form of automated traffic counters and location device informed journey times on a range of roads, the new data-driven method for calibrating the volume-delay relationship is shown to be simple but effective in satisfying some of the identified limitations in the current methods in this case study in the GLA.

2. Literature review

Volume-delay functions, as the name indicates, connects two fundamental traffic parameters with mathematical expressions. The input parameter, volume, expresses the level of traffic demand and the output, delay, indicates the deterioration of speed as the demand increases. The actual trend between volume and delay is a property of the road, related to the speed limit, width, geometry, road furniture, etc. Even though the volume-delay relationships are highly context specific, there exist some widely accepted functional forms to model them. The most widely used function is the BPR curve, which was developed in the United States in the 1960's. Its simple mathematical form and input requirements are attributed to its widespread use (Skabardonis and Dowling, 1997):

$$t = t_0 \times (1 + \alpha(\frac{v}{v_c})^{\beta}) \tag{1}$$

where t_0 is the free flow speed on the road link; v_c is the road capacity (vehicles per unit time); α and β are calibration coefficients. v and t are the traffic flow and travel time to be modeled.

The BPR function was created by fitting a polynomial equation to uncongested freeway data from the 1950's and thus does not reflect modern operating conditions (Skabardonis and Dowling, 1997). As a result, many different organisations have adapted the BPR curve with various local empirical and/or simulated data to better suit the local road conditions (Mtoi and Moses, 2014; IRAWAN et al., 2010). The work of (Kurth et al., 1996) focuses on obtaining refined, free flow time and capacity (t_0 and v_c) for each road based on guidelines from the Highway Capacity Manual (HCM, 1994 version), rather than getting t_0 and v_c from traditional lookup tables with only a few categories. This is found to produce more accurate traffic speed estimations. (Kucharski and Drabicki, 2017) jointly estimated both the calibration coefficients α, β and the road characteristics t_0, v_c based on loop detector data. They suggested to transform the volume-delay relationship to speed-density relationship for regression, as the latter remains monotonic in congested cases. However, the statistical analysis of the work seems dubious as they used the normal R-squared to compare nonlinear model and inconsistent way in calculating

the density. In the case of London, TfL have calibrated the BPR function with exhibited traffic counts and thus defined $\alpha = 1.0$ and $\beta = 2.0$ for the area (Transport for London (TfL), 2010). In general, calibrated volume-delay functions are found to fit local observations better.

Apart from the BPR function, there are other alternative forms of the volume-delay relationships that have been proposed over the years, as nicely summarised in (Mtoi and Moses, 2014). Davidson Davidson (1966) proposed a general purpose travel-time formula in 1966 and this method has undergone numerous modifications since it was first proposed (Mtoi and Moses, 2014). It has exhibited a closer match to actual volume counts and has a stronger theoretical base than the BPR (Rose et al., 1989). Among the modifications of the Davidson function since it was first proposed (e.g (Tisato, 1991), (Akçelik, 1991)), the mostly widely used is the Akcelik functions. The Akcelik method is a time-dependent modification of the Davidson model which uses the coordinate transformation technique in an attempt to overcome the conceptual and calibration issues with the Davidson method (Akçelik, 1991). It has illustrated good results in certain road types, tolls roads and signalized arterials (Mtoi and Moses, 2014). Spiess (Spiess, 1990) proposed the conical method in 1990. It attempted to overcome some of the BPR limitations at both the upper and lower bounds through the use of hyperbolic conical sections whilst maintaining a similar form to the BPR and thus enabled a direct transfer of parameters. These alternative formulations are also adopted in practice or research.

As for their applications, volume-delay functions are typically associated with utility (e.g., time cost) estimation in regional, static or semi-dynamic traffic assignments, or informing route choices for agent-based modeling(Çolak et al., 2016). Because it is differentiable and convex, it is particular suitable for optimization-based traffic assignment, such as the assignment that satisfies the Wardrop's equilibrium. Despite their wide use in research and practice, the volume-delay functions are not lack of criticism. For example, it is possible to obtain traffic volume-to-capacity results much higher than one, which is unthinkable in reality (Chiu et al., 2011; Nie et al., 2004). Closely associated is the problem that the volume-delay functions can only model the hypocritical branch of the traffic fundamental diagram (flow monotonically increase with travel time and density on a link), but not the hypercritical branch (when a road is congested to a certain level, the flow will decrease despite the increase in density and travel time). These limitations means that volume-delay functions cannot model traffic phenomena such as spillbacks, wave propagation, gridlocks, etc., in which case a dynamic model is called for (Lo and Szeto, 2005; Chiu et al., 2011). Despite these shortcomings, volume-delay functions remain to be a very useful tool for large scale analysis and quick estimation. Carefully calibrated volume-relationships frequently show good match with the data (Kurth et al., 1996; IRAWAN et al., 2010).

To continuously take advantage the simple, intuitive expression of the volume-delay relationship in traffic analysis and seeing its promise in even larger scale studies in the future, it is crucial to maintain up-to-date and local specific function coefficients, which has proven to be a nontrivial task. The literature has identified the need to inform these functions with empirical and context specific data (Rose et al., 1989), (Spiess, 1990) but recognised the difficulty and cost associated with collecting such empirical data as being prohibitive(Rose et al., 1989). There have been studies that incorporate different forms of field sensor data for such calibration (Mtoi and Moses, 2014; Neuhold and Fellendorf, 2014; Kucharski and Drabicki, 2017). However, the data used in these studies has been generated specifically for that application and requires specific hardware and/or software for use. Recent innovations in the information and communication technologies have led to the increased use of real-time crowd-sourced data feeds in transport modelling. This includes applications of location data for emissions estimations (Hirschmann et al., 2010), building origin and destination matrices (Toole et al., 2015) and general urban traffic management applications (Artikis et al., 2014). This research investigates the use of new, real-time crowd-sourced data feeds that have a wider spatial spread and not generated specifically for this application. These sources may be used to consider some of these previously ignored characteristics and create temporally and spatially dynamic volume, speed and saturation relationships. Such data sources can harvest data at a finer resolution over a longer (even indefinite) period of time giving a far greater understanding of the temporal variations and trends exhibited on road infrastructure.

The work of this paper attempts to combine new data sources in order to create usable functions that do not require a large range of survey inputs. The general and transparent methodology of harvesting crowdsourced data enables its easy depolyment to multiple sites. As a result, highly localised relationships can be obtained for a large number of road links, reflecting individual characteristics of the road (i.e. geometry, the presence of traffic furniture, road quality and the surrounding land use). Such differing characteristics can result in very different vehicular behaviour on roads that may be considered similar by traditional approaches.

3. Traffic volume inputs: automated traffic counters

Automated Traffic Counters (ATCs) are magnetic induction loops located in the road surface. The passing of a vehicle results in an electromagnetic signal. The ATCs in the Greater London Area count every vehicle which passes over the inductions loop. The data used here was harvested over a period from the 29th February to the 21st March 2016.

There are 37 Department for Transport (DfT) ATC locations distributed in the Greater London Area (Figure 1). Among these data collecting locations, 34 roads have ATCs operate in both directions, while 3 roads have ATCs operate only in one direction. The ATC locations provide traffic counting information for a range of different DfT defined road classes (Table 1). Guidance on the road classification system in the UK is published by the DfT (UK Department for Transport (DfT), 2012a).

Table 1. ATC locations by road class

Road class	Trunk (Motorway or A)	Principle (A)	B	C	Unclassified	Total
Count	6	16	3	3	9	37

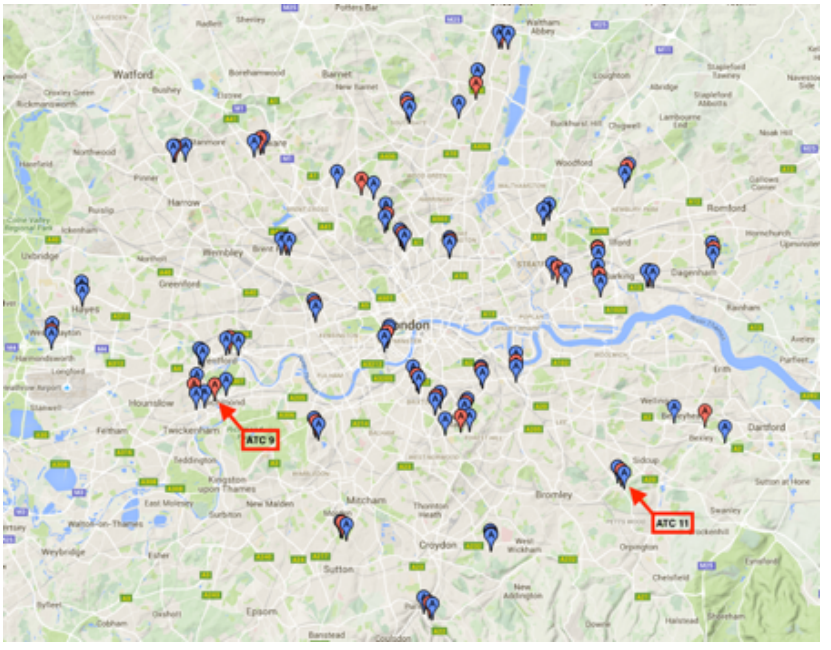

Figure 1. ATC Locations in Greater London Area (Google, 2017) with DfT data. The red flag illustrates the location of the ATC itself. The blue flags illustrate the origin and destination locations specified in order to harvest journey time information.

The raw ATC data contains individual records for each vehicle that passed, including the speed of the vehicle as well as the exact date and time (accurate to second) when passing the counter. Over the test period of 3 weeks, there were approximately 4.5 million records of vehicles at individual points. For example, one of such record on the first day of data collection at ATC Site 11 (labelled in Figure 1) reads:

Site: 11, Direction: Northbound, Date: 2016-02-27, Time: 00:00:28, Speed: 37.

In order to obtain the traffic volumes per hour (i.e., traffic flow), individual vehicle records from the ATCs are grouped by hour. The total number of records in a group is then taken as the hourly traffic volume passing through the measurement site. A Python (Van Rossum and Drake, 2003) script using the Pandas (McKinney et al., 2011) data library was used for this processing. First, a unique identifier is formed by concatenating the location ID and the directionality of the ATC. For example, the northbound detector at ATC location 11 is identified as site "11N". The time stamp is rounded up to the next hour in order to quantify the hourly traffic volume up to that hour. This results in an output dataset which features traffic counts per hour (volume) for a site, direction and date, a sample of which is shown in Table 2.

Table 2. Processed ATC data record sample

Combined ID	ATC ID	Direction	Date	Hour	Hourly Traffic Volume
9S	9	Southbound	2016-02-27	13	1206
9S	9	Southbound	2016-02-27	14	1222
9S	9	Southbound	2016-02-27	15	1408
9S	9	Southbound	2016-02-27	16	1604

Figure 2 shows an illustration of the variations in the hourly traffic volume at ATC location 11 (northbound and southbound). Location 11 is situated on the Royal Parade road (A208) with the northbound direction leading to central London and the southbound direction leaving from central London. It is clear that, during weekdays (March 7th - 11th, 2016), the northbound direction exhibit higher peak during the morning period (commuting trips into the city), while the southbound direction carries more traffic leaving the city during the evening peak. On weekends (March 12th and 13th, 2016), there is only one peak which also starts later than the morning peak on weekdays. This variation in traffic volume is in accordance with the general understanding of the distribution of traffic loads throughout the day.

4. Crowd sourced journey times

Link travel time, or equivalently the inverse of vehicle speed on a link, can be collected using several methods. For example, the ATC data presented in the previous section have vehicle speed record. However, the ATC speed measurements are instantaneous and may not be suitable to calculate the average travel time across the road link, as required by many traffic simulation studies. An alternative method to infer the travel time across the link based on real-time, crowdsourced location information gathered from mobile phone users. Mobile phones with location service on can enable the harvesting of fine resolution temporal and spatial position data, which is determined from the GPS, cell tower triangulation, WiFi SSID mapping, Bluetooth and other technologies, either in isolation or in tandem. Such data hold a great deal of promise due to the range of possible uses it has in the transportation sector (Zheng et al., 2010). The anonymous crowd-sourced collection of this data can be used to derive road traffic conditions (congested or smooth flowing roads) (Barth, 2009). Such data collection method is well suited to areas of high density, high travel demand and high mobile phone uptake, such as an urban area. This information has been widely used by service providers to inform their users of the optimum path to

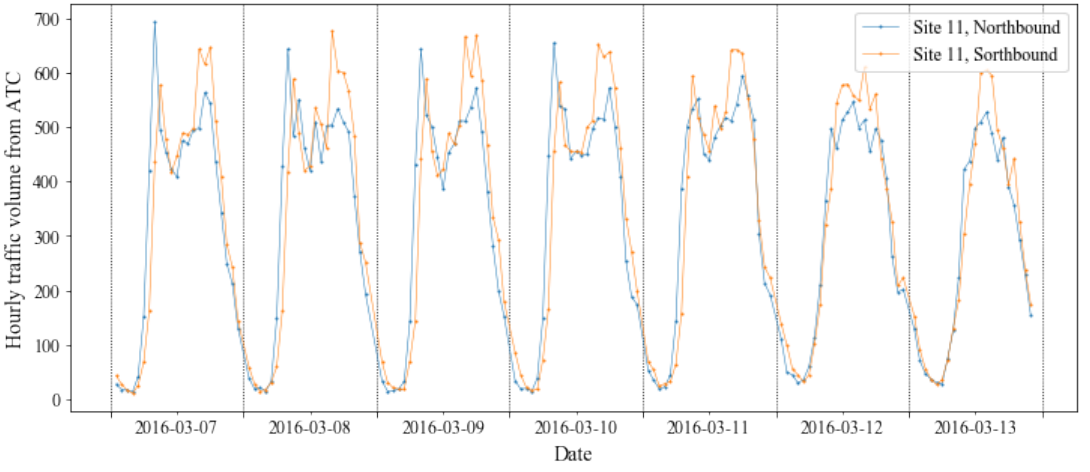

Figure 2. Hourly traffic volume distribution for Site 11 (March 7th-13th, 2016).

avoid the traffic, for example, Apple Maps (*apple.com/ios/maps*), Bing Maps (*bing.com/map*), Google Maps (*google.com/maps*) and TomTom (*tomtom.com*). Individuals with access to these services can then make an informed decision on route choice for a given mode or even a mode decision on how to get from their starting point to a desired destination with the lowest time cost, etc. For example, Google's traffic condition maps can be visually inspected online. Their colour coded scale gives a qualitative representation for traffic conditions on the roads where Google have sufficient data, as is shown in Figure 3.

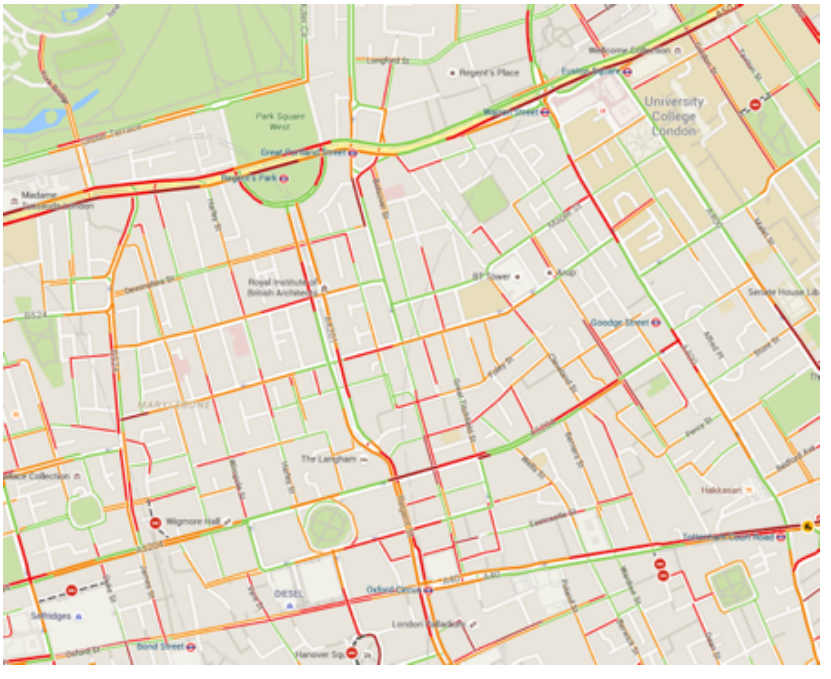

Figure 3. Google Maps Traffic Layer, Camden/Soho/Marylebone/Mayfair area of London (Google, 2017).

In this research, the traffic condition information used to generate route suggestions is requested in batch from the Google Maps Directions Application Protocol Interface (API) (Google, 2016). Dependent on their personal settings, an Android user or other Operating System (OS) user, with the free Google Maps mobile app installed on their location enabled phone send anonymous data to Google. Such data is personally and commercially sensitive and so post-processing is carried out by Google in order to ensure that no-one user's movements can be isolated from the flows. The Google's Directions API used in this paper is a service that calculates travel time and routing directions between given origins and destinations using a Hypertext Transfer Protocol (HTTP) request (Google, 2016). The use of a HTTP request allows for scheduled and bulk harvesting of journey information between given origin and destination pairs. The aim of this research is to combine crowd-sourced location device informed journey times with traffic counts from the DfT ATC network. As such, the first step is to specify origins and destinations for the Google Directions HTTP request that would provide a journey time for the traffic counts at defined ATC locations. There is a need to convert from EPSG:27700 (British National Grid), as provided by the DfT for ATC locations, to the EPSG:4326 (WGS84) as used by Google Maps. In order to find a journey time along a road where an ATC locates, an origin and destination location definition process is required. In deciding on origin and destination pairs a balance must be made between having sufficient distance in order to get meaningful journey time results and having sufficiently short distance so as to not include undesired results such as detours. Figure 4 shows the manually defined origin and destination points for ATC location 6 Eastbound, the A205 Dulwich Common SE21 in the Borough of Southwark. Automating the definition of these origin and destination pairs is challenging as simply taking the start/end of a given road often gave a route that is too long and distorted by other traffic, outside of the ATC consideration. Other methods based upon an idealised distance between points and the density of junctions is deemed too complex and not durable. For bidirectional ATC locations it is often not possible to define the Eastbound route as the inverse of the Westbound route as Google distinguish between different sides of the road, resulting in a route which involves a detour to safely navigate to the correct orientation. Thus a manual process is employed to visually inspect each location, the surrounding context and decide on the most appropriate origin and destination locations for querying for the travel time. Once this manual process is complete, a list of origin and destination pairs is produced containing ATC metadata that allows for the pairing of Google travel time results to its corresponding ATC traffic volume data.

Figure 4. ATC 6 Eastbound with defined origin and destination points (Google, 2017).

During the information request process, an HTTP request is made in Python to Google's servers with the origin, destination, mode (driving) and the specified departure time. In order to harvest real-time data that is informed by location device information at each hour, a cron scheduler is used to run the same origin and destination pairs repeatedly. In response to such HTTP request, Google returns results in the JavaScript Object Notation format (*json.org*), a lightweight data-interchange format. In the returned JSON results, the $duration_in_traffic$ field is taken as it states Google's estimation of journey time between the requested origin and destination. In the next section, these travel time data will be paired with ATC vehicle counts to produce volume-delay relationships.

Figure 5 shows a subset of the journey time data collected. Specifically, it shows the Google's estimation of travel time along the Royal Parade, between Manor Park Road and Bromley Road in Southeast

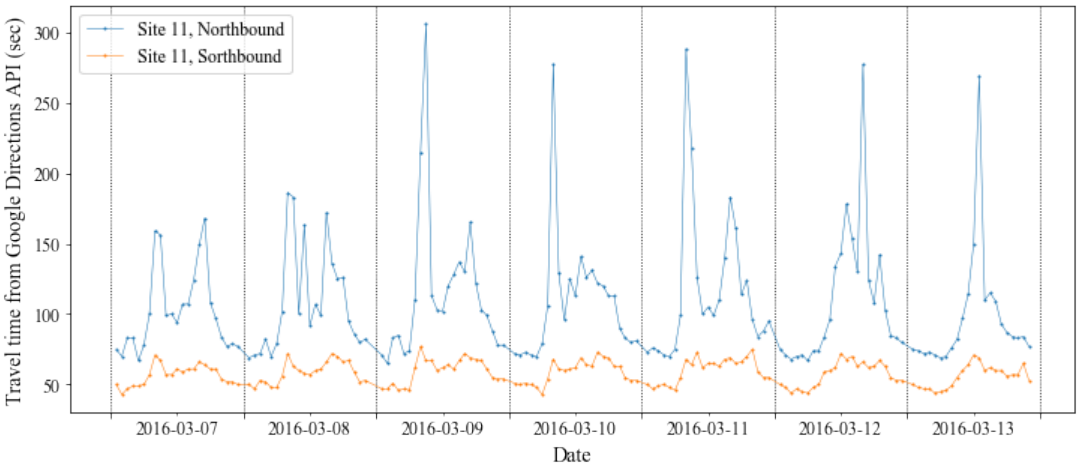

Figure 5. Journey time distribution for Site 67 (March 7th-13th, 2016).

London (561 metres in length). ATC 11 locates on the north end of this stretch of the road. Apart from the bimodal distribution of journey time with peaks associated with the morning (to work) and evening (leaving work) periods during weekdays and monopeak on weekends (Mullick and Ray, 2012), it is interesting to see a much higher travel time in the northbound direction at all times. A visual inspection of the corresponding road shows that there are on-street parking spots in the northbound lane. Also, the ending point of the northbound lane intersects with a more major road (Bromley Road, A222) than southbound (Manor Park Road, B264). Both factors could contribute to the unexpected slowness in the northbound direction.

5. Data cleaning and site selection

The volume delay relationship at different traffic conditions at each site can be constructed by combining the Google journey time and ATC vehicle count data. Figure 6 shows an example of such data for site 9S, 11N and 39N. It can be seen that there are a few outlier points with high travel times at moderate traffic flow rates. These abnormal delays could occur due to external reasons (e.g., weather and road incidents), or as mentioned in the literature review, being the hyper-congested cases where flow reduces with increase in delay (especially the cluster of points with volume between 400 and 600 for site 11N). The hyper-congested regime is not modeled by the volume-delay function. As shown in the end of this paper, they are instead regarded and quantified as residuals of the volume-delay curve. Regardless, data cleaning needs to be carried out to remove points falling too far away from the continuous volume-delay relationship. Specifically, a manually tuned DBSCAN algorithm is used to remove points in low density areas in the plots (Brownlee, 2020; scikit-learn developers, 2019). The DBSCAN algorithm has two parameters, ϵ, which sets the distance between points to be considered as a cluster and *minPoints*, the minimum number of points required to be considered as a core point in the cluster. Considering the horizontal and vertical axes have different scale, the original data are first normalized to the maximum observed volume/travel time, so that the data in the transformed axes are all within 0 to 1. It was found that by specifying $\epsilon = 0.1$ and $minPoints = 5$, it offers a good balance between removing outliers and keeping data points close to the main cluster. In fact, the data cleaning is kept to minimum to accommodate any possible variations that may occur in the real life observations. As a result, only 66 points are removed from the 17,745 observations.

Also, by plotting the journey time-traffic counts data for all the 39 original sites, it was immediately spotted that some sites do not have valid data (no or very low variations in observed travel time through

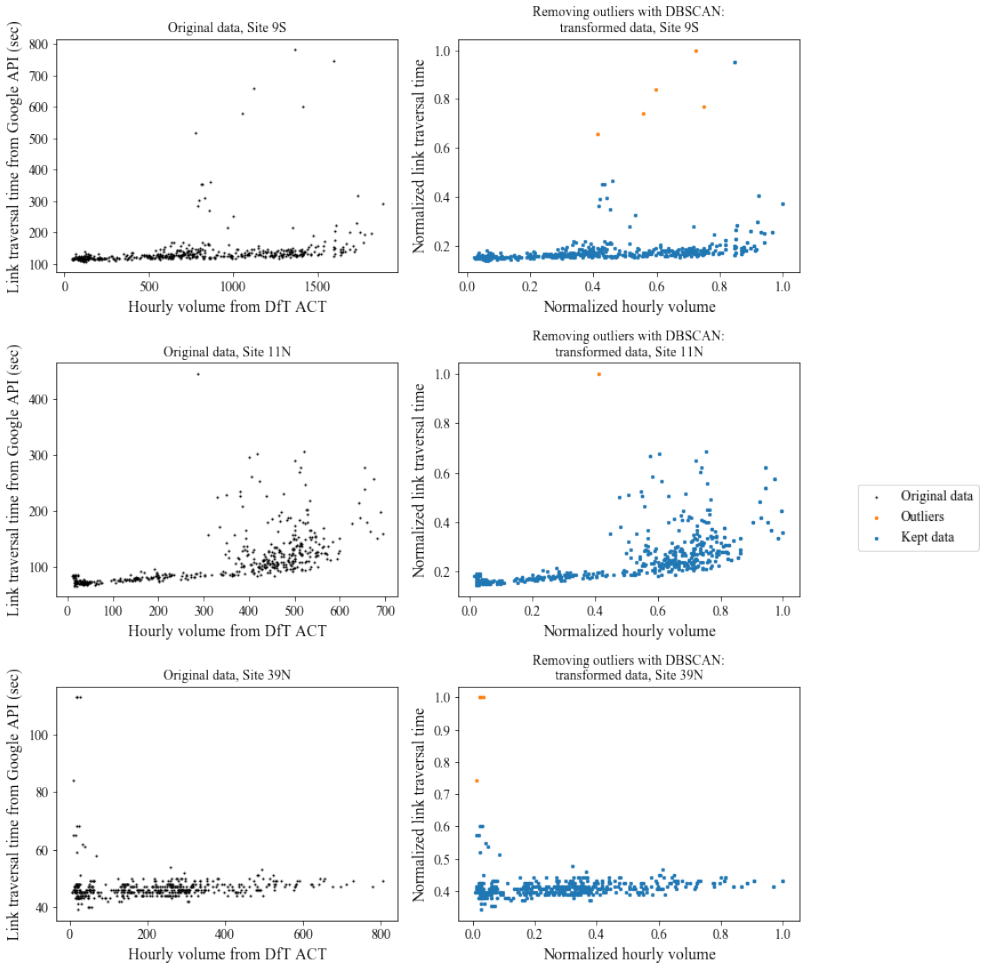

Figure 6. Removing outliers from the volume-delay scatter plot..

the observation period). These sites are subsequently removed from analysis and it leaves 28 sites remaining for further analysis.

6. Analysis and model building

The combination of Google data and DfT ATC vehicle count data enables various relationships to be assessed and defined. Based on these real world datasets, the goal of this section is to derive data-driven parameters/relationships related to the traffic analysis, including (1) free flow travel time, or its inverse, the free flow speed; (2) road capacity; and (3) volume-delay relationship.

6.1. Free flow travel time

The free flow travel time usually corresponds to the time for a vehicle to pass through a road link when no other vehicles present, such as the travel time experienced in the early morning. In research studies and engineering practice, it is usually taken as the time to go through a link when travelling at the designated speed limit, sometimes multiplied by a slow-down factor around 1.3 or so to reflect the minor delays caused by stopping at intersections, etc Çolak et al. (2016). For example, the TfL

defines the free flow journey time using time delay coefficients for different road types, speed limits, link lengths, widths, gradients, traffic junctions and so on (UK Department for Transport (DfT), 2002).

In this study, it is recognised that using agency recommended free flow speed (design value) based on road type frequently ignores the impact of localised, irregular factors, for instance, a curve or pothole on the road. Rather than making assumptions about the free flow journey time, it was estimated from the Google journey time data collected over the period of the study. For each site, the 5th percentile (fastest) travel time among all observation points is used to represent the observed free flow travel time of that particular site. Figure 7(a) shows the contrast between the observed free flow travel time and the design value according to the speed limit. Each point represents a specific site in the dataset. On average, the observed free flow time is 39% longer than the travel time based on the design value.

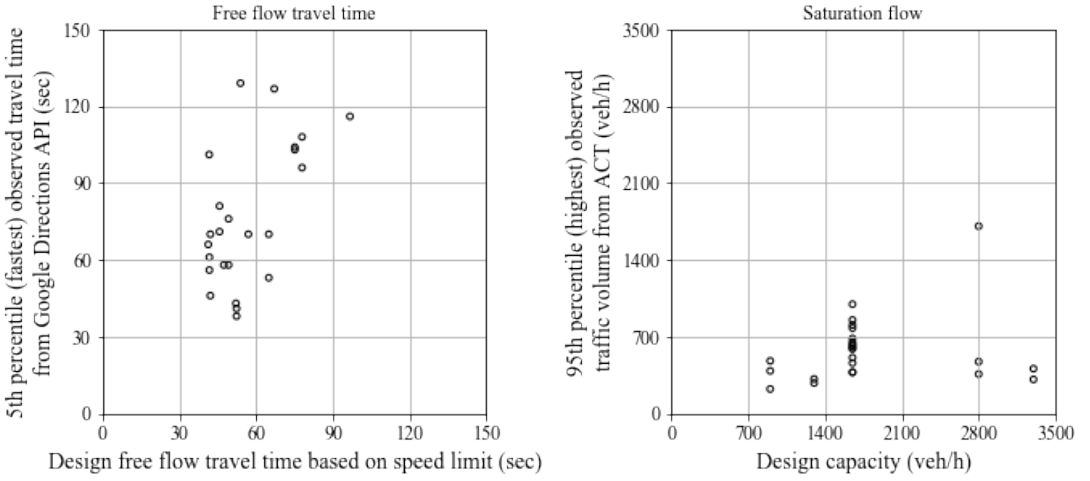

Figure 7. The design and observed free flow speed (left) and saturation flow (right) at 24 study sites..

6.2. Road capacity

Compared to the free flow travel time, road link capacity is a less clear quantity to define. For example, the UK Design Manual for Roads and Bridges (DMRB) define capacity vaguely as - "the maximum sustainable flow of traffic passing in 1 hour, under favourable road and traffic conditions (The UK Highways Agency, 1999)." The DMRB also provides look up tables that feature traffic capacities for a range of road types, road widths and number of lanes. A manual survey of satellite imagery was carried out to assess the lane count, estimate the road width for each of the ATC locations and thus provide an estimated capacity by this DMRB definition. However, such a method was deemed unacceptable due to the uncertainty in what constitutes *favourable road and traffic conditions*.

Instead, a more nuanced definition from Spiess (Spiess, 1990) is adopted, which defines capacity as the volume at which congested speed is half the free flow speed. In the paired Google and ATC data, since there is hardly any observation when the speed is exactly half of the free flow speed, a more relaxed definition of capacity is subsequently utilised. The capacity is taken as the highest 95th percentile hourly flow among all observations where the travel time falls between 1.8 to 2.2 times the free flow speed. As a result, each individual road is given a capacity attribution. This capacity attribution is generally lower than that from those which are more qualitative, as can be seen in Figure 7 (b).

6.3. Volume-delay relationship

Based on observations of road link travel time and hourly traffic counts presented above, the useful volume-delay relationship can be constructed. The frequently used BPR curve adopts a power function to describe the relationship between traffic saturation and time delay (Equation 1). The TfL recommends to use $\alpha = 1.0$ and $\beta = 2.0$. While t_0 and v_c are often obtained from the speed limit or road capacity guidelines. In this study, the power-form volume-delay function is kept due to its numerical advantages (differentiable and convex), as well as its widespread usage among the traffic community. As a result, the aim of this analysis is to propose alternative, site specific function parameters obtained from real data and present the variability of the observed and calculated time-delay at different congestion levels. A summary of the three models used for comparison is given in Table 3. The design curve follows existing parameter recommendations. For the first data informed alternative formulation, the design value of t_0 and v_c are replaced by observed values described previously in this section. For the second data informed alternative, not only t_0 and v_c are replaced by observed values, α and β are also obtained using regression based on the real world data collected.

Table 3. Summary of volume-delay relationships tested.

Function name	Parameters
Design curve	$\alpha = 1.0$, $\beta = 2.0$ t_0: road length divided by speed limit. v_c: DMRB definition.
Data informed 1	$\alpha = 1.0$, $\beta = 2.0$ t_0: 5th percentile link travel time (fastest) in observations. v_c: 95th percentile traffic volume (largest) for observations with travel time between 1.8 to 2.2 $\times t_0$.
Data informed 2	α, β: fitted value using nonlinear least square regression. t_0: 5th percentile link travel time (fastest) in observations. v_c: 95th percentile traffic volume (largest) for observations with travel time between 1.8 to 2.2 $\times t_0$.

It is obvious from Figure 6 that the observed relationships between the volume and delay are site-specific, non-linear and heteroskedastic (variance of the data increases with the independent variable). All these are expected from the traffic perspective. In particular, the nonlinearity means that the travel time increases more steeply as the hourly traffic volume approaches the capacity. The heteroskedasticity implies more variations in vehicle speed when the road is congested due to interactions with other vehicles. Usually, to deal with such data, a nonlinear model needs to be used (which is already the case for the BPR curve). To deal with heteroskedasticity, the usual techniques include data transformation or using Weighted Least Square (WLS) method. However, it is not a requirement if the goal is to estimate regression parameters, as the Ordinary Least Square (OLS) method also produces unbiased, though inefficient, estimation of the coefficients. Standard errors of the estimation parameters are not used, so the OLS method is deemed sufficient for obtaining parameters for the data informed function 2. Figure 8 shows the fitting results of all candidate functions for three representative sites. It can be seen that the design curve fits the poorest to the data. By simply substituting the free flow time t_0 and capacity v_c in the design curve, the fitting of the data driven model 1 improves significantly from the visual inspection. Quantitative evaluations of the performance will be given in the Section 6.4. The data driven model 2 (green curve in Figure 8) fits the data more closely than the data driven model 1. However, in certain cases, such as in Figure 8 (c), a β value less than one is obtained due to the influence of large numbers of observation points in the middle regime (probably the hyper-congestive conditions),

making the volume-delay curve unreasonably concave. Visually, data driven model 1 is a more realistic approach to improve the performance of volume-delay curves. (Note: log(t/v)-log(t0) = k1 * log(v) - k2 * log(cap) almost completely removes the nonlinearity and heteroskedasticity.)

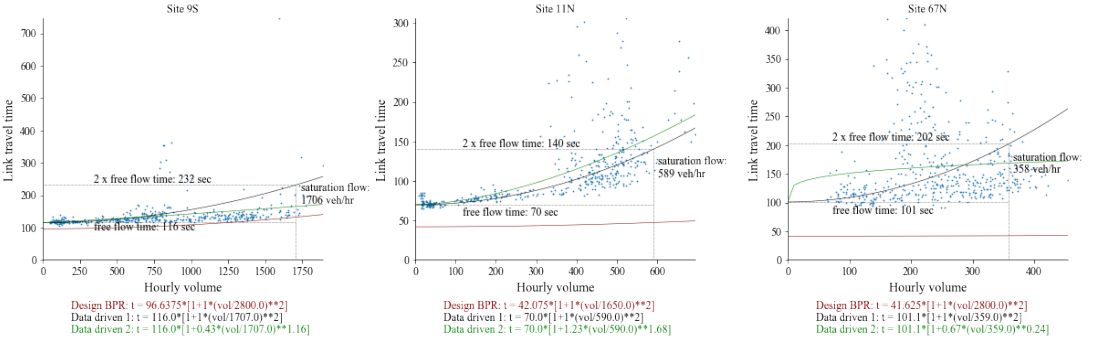

Figure 8. The fitting three different models in the observed data at two sites..

6.4. Variation quantification

As it can be seen in previous scatter plots of the volume-delay data, real world data are certainly too noisy to be represented by a single curve. There have been second-order models that include the acceleration effect, which can reproduce the variations in travel time around the same traffic flow conditions to a certain degree. However, well-calibrated first-order models that only include the speed term (no time derivatives of it), such as the BPR curve, are still very useful for certain tasks, including the demand estimation due to new infrastructure and other static and semi-static traffic simulations at a large scale. After fitting three candidate model to the data, results will be given here regarding the performances of the different types of models.

Mean Absolute Error (MAE) is used here as a model performance metric. It has the same unit (second) as the dependent variable and is less sensitive to outliers than the frequently used Root Mean Squared Error (RMSE). As the variations in the observed travel time are uneven and significantly larger when the hourly traffic volume approaches the capacity, MAE is given to different subsets of the data, divided based on the quartile of the hourly traffic volume. Figure 9 shows the MAE in terms of travel time (a-c) and speed (d-e) estimations of each road (indicated by line color), grouped by the traffic volume as a quartile of the estimated capacity. For example, Q1 means all data points where the hourly traffic volume (x-axis) are within 0-25% of the estimated capacity. Q5 are for those points where the hourly traffic volume exceeds the estimated capacity. It can be seen that, in terms of travel time or speed, the original model produces large errors. In fact, for the sites with the highest errors in Figure 9 (a) and (c), they correspond to site 67N and 67S, the former of which is also presented in Figure 8 (c). While for the two data driven models, the errors of the data driven function 2 appear to be less, but not much lower compared to the data driven function 1.

In fact, the above results not only indicate the performance of model, but also give engineers quantitative evidence on what amounts of variations to be expected when using the simple volume-delay function curves in their estimation. For example, without the site-specific knowledge, the speed estimations using the volume-delay curve based on the design value can be 10-30 km/h off (Figure 9 (a)). While by just fixing the speed and capacity estimations alone (data driven function 1), the speed estimation produces less than 5 km/h for low to medium traffic volume case, and usually less than 10 km/h when the volume is close to the capacity (Figure 9 (b)). If using calibrated coefficients α and β as in the

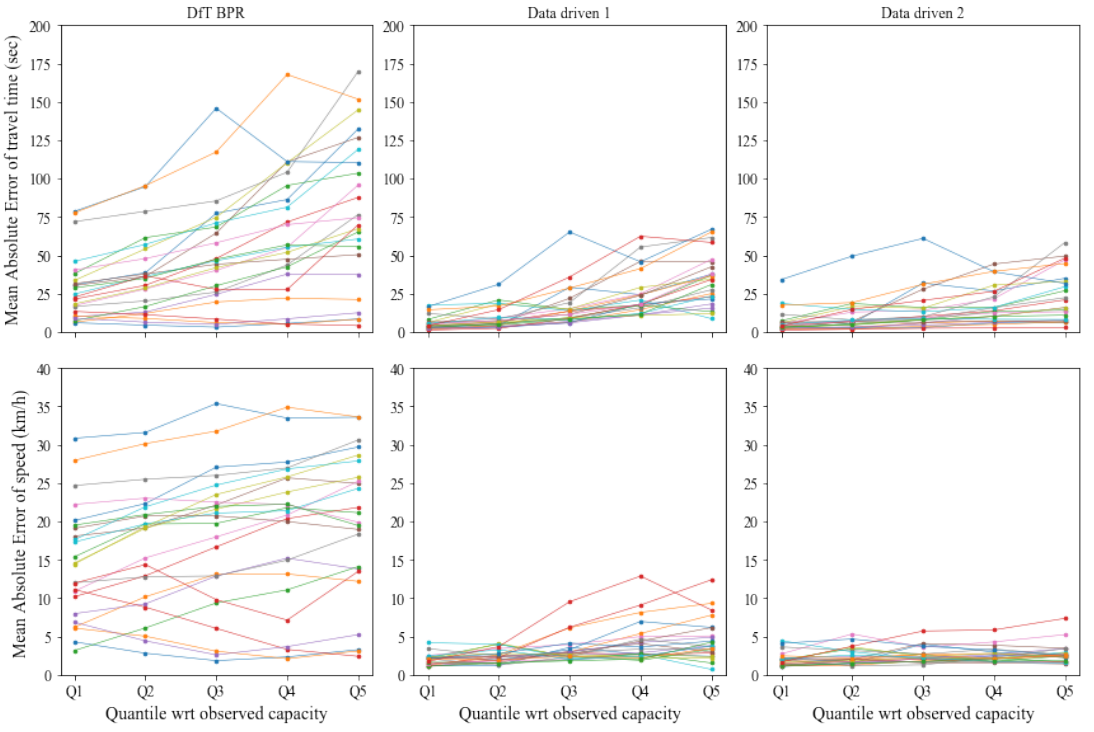

Figure 9. MAE for three model types. There are 24 lines in total, with each line representing a particular site..

data driven function 2, the speed estimations are almost always less than 5 km/h away from the observations (Figure 9 (c)). But it should be noted that calibrating α and β without constraints may produce concave volume-delay curves as shown in Figure 8 (c).

7. Discussion

In many cases, the plots generated in this study presented empirical data that did not match the conventional macroscopic understanding, as epitomised by the BPR functions. For each direction and each individual site, three function fits are generated to create context specific saturation delay and saturation speed functions. The fitting performances and residual variations are shown in Figures 8 and 9. It is clear that the observed volume-delay relationship is highly scattered, and there must be external unknown factors that have not been included in the candidate models.

7.1. Example scenario & possible factors

Figure 10 (a) presents the observed volume and delay scatter plot for ATC location 66 in both directions. It is obvious that the northbound direction consistently has higher delay compared to the southbound direction, even if the volumes are the same. In an attempt to better understand and possibly explain this distinct directional behaviour, satellite photography and street level photography of the road is assessed. Figure 11 shows a satellite image of ATC location 66 obtained from Google Maps (Google, 2017). The red marker gives the location of the ATC counter itself and the two blue markers illustrate the origin/destination location (dependent on direction) submitted to the Google Directions API request. From this satellite imagery and street view imagery possibly explanatory factors can be identified:

1. The southbound lane features on road parking;

2. The southbound lane features a large bus stop and taxi lay-by (for Leytonstone Station) and a smaller bus stop lay-by.

The larger bus stop and taxi lay-by serving Leytonstone is a significant geometric feature that is likely to have heavy impacts southbound traffic as buses/taxis leave and enter from both directions on the road. The second smaller bus stop lay-by and on-road parking may also have an impact, albeit in a smaller way to the traffic speed in the southbound direction. Such factors may explain the exhibited differences from the location device informed journey data and thus permit their inclusion in an informal way.

7.2. Challenging scenarios

Inspections of the volume-delay scatter plots reveal a subset of roads that are not suitable for the regression analysis. For example, as previously explained in Section 5, there are 11 sites that are eventually removed from the analysis due to invalid data. Among the remaining 28 sites, many still have outliers even after the data cleaning process (e.g., Figure 6). Figure 10 (b) shows another "abnormal" example, particularly with the northbound direction: the delay has a very large variation when the hourly volume is around 200 vehicles, far from the capacity (about 400 vehicles per hour). This probably shows the hyper-congested stage of the traffic, where the density has exceeded the optimum value and flow decreases despite continuous increase in the delay. Such phenomena cannot be correctly modeled by the monotone volume-delay function and should be dealt with using dynamic models if the hyper-congested stage is crucial in the analysis (which is usually the case for short-term predictions, such as evacuations). The errors of using the monotonic volume-delay function in face of hyper-congested real world observations are reflected in the MAE quantification as shown in Figure 9. Apart from these reasons, external factors could also affect the sensibility of the data, for example, in a road traffic collision or road flooding as a result of high level of precipitation in a short period of time.

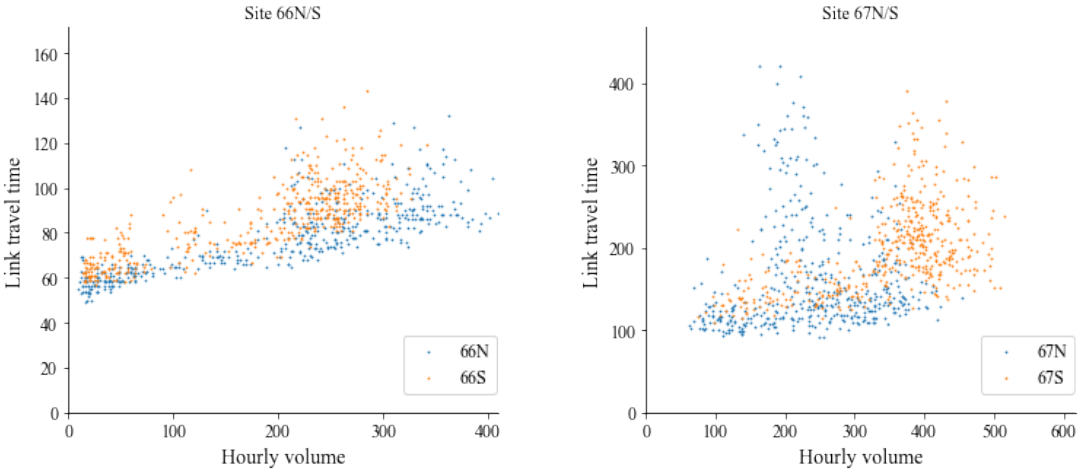

Figure 10. Challenging scenarios: volume and delay observations at ATC site 66 (a) and 67 (b).

7.3. Influencing factors

In Section 7.1 the existence of public transport infrastructure and on-road parking were identified as being potentially explanations of the distinct southbound northbound behaviour on that road. A range of other possible factors have been identified, which may help explain the unexplained variations:

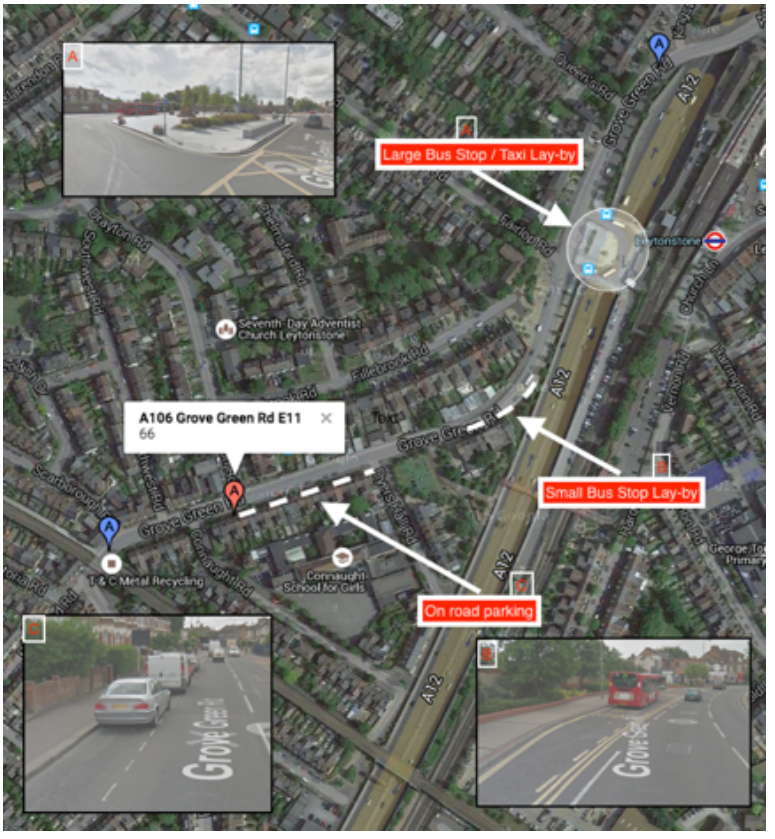

Figure 11. Satellite & Street View images of ATC 66 (Google, 2017).

1. Vehicle type
   Further work is planned to assess the impact of the vehicle type. In this paper each vehicle has been attributed equally despite the clear distinction between the interactions of a car and another car compared to that of a lorry to a car (Vap et al., 2007). A wide variety of vehicles mixes are exhibited on different road types and such differences should be considered. The use of statistical vehicle mix sampled by road type (UK Department for Transport (DfT), 2012b) was discounted for this study due to the small sample size of such statistics compared to the resolution of the data used here. The use of traffic cameras with number plate recognition and sufficient privileges to the Driver and Vehicle Licensing Agency database would enable the disaggregation of vehicle type at a similar spatial and temporal resolution to the Google Directions and DfT ATC data presented here.
2. Weather
   Weather events may impact on journey times by impacting on the performance of the vehicle, the performance of the road and/or the performance of the driver. Ongoing research at the University of Cambridge is combining a large dataset of Google Directions journey times with data from the UK's Met Office NIMROD precipitation dataset (The UK Met Office, 2003) in order to assess the relationship between these variables. At the microscopic level it is known that an increase in precipitation increases journey times as a result of increased risk and the resulting decrease in vehicle speeds to compensate for this (Mashros et al., 2014).
3. Road incidents
   Road works and road traffic collisions can lead to decreased or even zero capacity on a road link, resulting in increased saturation and thus impacting on journey time. Depending on the warning before such an event, the vehicle traffic may have the ability to adjust to this information, resulting

in a greater distribution of traffic, leading to mediated travel times. Alternatively, an accident may occur and not enable any warning to be given to other road users until they are committed to their route choice, resulting in large journey time increases, perhaps explaining extreme phenomena. A method incorporating different accident and road works databases with Google Directions data is currently being investigated.

4. Road geometry, type & land use
   Different road layouts may result in an increase in the complexity of vehicle interactions. For example, the curvature of a corner and the road surface quality will impact on the speed of a vehicle. The surrounding land use will likely also impact, adding safety concerns (for example a school or leisure centre) again impacting on vehicle speed. The inclusion of such factors poses many challenges, the size and complexity of the data plus the uncertainty and variability in how drivers react to the data. In Figure 11 a series of geometric factors are displayed as part of an attempt to explain different behaviour on the same road dependent on direction. The factors discussed there, on-road parking and bus lay-bys may be quantitatively captured using machine vision and data sources such as Google Street View.

8. Conclusions

A range of context specific saturation volume-delay curves for a range of different locations and road types have been generated. Specific examples have been presented here for discussion and all generated functions and plots are available for inspection here. In the most practical sense some of these functions may now be used in the traffic assignment stage of the traditional four step model. In some cases, the data presents clear evidence that unknown factors, such as those listed in Section 7.3, have a significant impact and warrant further investigation. In these cases the derived functions and indeed any standardised function have been shown to deviate significantly from empirical data and as such their use should be considered with care. Overall, there are significant collective correlations between the ATC traffic count data and that of the Google journey times, across a range of sites, presenting evidence which goes some way to validate the Google data and illustrate the collective value of this method. The data used here shows promise in considering the tangible factors which impact on road performance, such as local geometry, bus stops and so on, but that have historically been too challenging to be considered. These data sources have longevity, exist at close to real-time and in the case of the Google Directions data, relatively low cost with little or no capital expenditure required for its harvesting. These methods may be employed over a long time horizon and at a finer temporal resolution in order to better understand the temporal and spatial trends as well as the influencing factors such as sporting occasions and weather events. It can also be used to do real-time vehicle emissions estimations and modelling, as is being investigated presently. The automation of this method over longer time horizons may lead to explanations for the issues discussed previously and highlight areas that require investigation in order to better understand the performance of road infrastructure.

Acknowledgments. We are grateful to both Google and the Department for Transport for access to data.

Competing interests. None

Data availability statement. Data - https://github.com/gac55/delay-curves

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author contributions. Gerard Casey and Kenichi Soga conceptualised the project and methodology. Gerard Casey did the analysis and writing with Kenichi Soga reviewing and editing.

References

Akçelik, R. (1991). Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. *Australian Road Research*, 21(3).

Artikis, A., Weidlich, M., Schnitzler, F., Boutsis, I., Liebig, T., Piatkowski, N., Bockermann, C., Morik, K., Kalogeraki, V., Marecek, J., et al. (2014). Heterogeneous stream processing and crowdsourcing for urban traffic management. In *EDBT*, volume 14, pages 712–723.

Barth, D. (2009). The bright side of sitting in traffic: Crowdsourcing road congestion data. *Google official blog*.

Brownlee, J. (2020). 10 Clustering Algorithms With Python. https://machinelearningmastery.com/clustering-algorithms-with-python/. [Online; accessed 1-July-2020].

Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., and Hicks, J. (2011). Dynamic traffic assignment: A primer. *Dynamic Traffic Assignment: A Primer*.

Çolak, S., Lima, A., and González, M. C. (2016). Understanding congested travel in urban areas. *Nature communications*, 7(1):1–8.

Davidson, K. (1966). A flow travel time relationship for use in transportation planning. In *Australian Road Research Board (ARRB) Conference, 3rd, 1966, Sydney*, volume 3.

Google (2016). Google directions api documentation.

Google (2017). Google Maps.

Hirschmann, K., Zallinger, M., Fellendorf, M., and Hausberger, S. (2010). A new method to calculate emissions with simulated traffic conditions. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 33–38. IEEE.

IRAWAN, M. Z., SUMI, T., and Munawar, A. (2010). Implementation of the 1997 indonesian highway capacity manual (mkji) volume delay function. *Journal of the Eastern Asia Society for Transportation Studies*, 8:350–360.

Kucharski, R. and Drabicki, A. (2017). Estimating macroscopic volume delay functions with the traffic density derived from measured speeds and flows. *Journal of Advanced Transportation*, 2017.

Kurth, D. L., Van den Hout, A., and Ives, B. (1996). Implementation of highway capacity manual–based volume-delay functions in regional traffic assignment process. *Transportation research record*, 1556(1):27–36.

Lo, H. K. and Szeto, W. Y. (2005). Road pricing modeling for hyper-congestion. *Transportation Research Part A: Policy and Practice*, 39(7-9):705–722.

Mashros, N., Ben-Edigbe, J., Hassan, S. A., Hassan, N. A., and Yunus, N. Z. M. (2014). Impact of rainfall condition on traffic flow and speed: a case study in johor and terengganu. *Jurnal Teknologi*, 70(4):65–69.

McKinney, W. et al. (2011). pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9).

Mtoi, E. T. and Moses, R. (2014). Calibration and evaluation of link congestion functions: applying intrinsic sensitivity of link speed as a practical consideration to heterogeneous facility types within urban network. *Journal of Transportation Technologies*.

Mullick, A. and Ray, A. K. (2012). Dynamics of bimodality in vehicular traffic flows. *arXiv preprint arXiv:1205.2314*.

Neuhold, R. and Fellendorf, M. (2014). Volume delay functions based on stochastic capacity. *Transportation research record*, 2421(1):93–102.

Nie, Y., Zhang, H., and Lee, D.-H. (2004). Models and algorithms for the traffic assignment problem with link capacity constraints. *Transportation Research Part B: Methodological*, 38(4):285–312.

Rose, G., Taylor, M. A., and Tisato, P. (1989). Estimating travel time functions for urban roads: options and issues. *Transportation Planning and Technology*, 14(1):63–82.

scikit-learn developers (2019). Clustering: DBSCAN. https://scikit-learn.org/stable/modules/clustering.html#dbscan. [Online; accessed 1-July-2020].

Skabardonis, A. and Dowling, R. (1997). Improved speed-flow relationships for planning applications. *Transportation research record*, 1572(1):18–23.

Spiess, H. (1990). Conical volume-delay functions. *Transportation Science*, 24(2):153–158.

The UK Highways Agency (1999). The design manual for roads and bridges (volume 5) assessment and preparation of road schemes.

The UK Met Office (2003). Met office rain radar data from the nimrod system.

Tisato, P. (1991). Suggestions for an improved davidson travel time function. *Australian road research*, 21(2).

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., and González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162–177.

Transport for London (TfL) (2010). Traffic modelling guidelines. version 3.0.

UK Department for Transport (DfT) (2002). Cost benefit analysis manual. volume 13: Economic assessment of road schemes.

UK Department for Transport (DfT) (2012a). Guidance on road classification and the primary route network.

UK Department for Transport (DfT) (2012b). National travel survey: 2012.

Van Rossum, G. and Drake, F. L. (2003). Python language reference manual.

Vap, D., Sun, C., et al. (2007). Investigating large truck-passenger vehicle interactions.

Zheng, Y., Chen, Y., Li, Q., Xie, X., and Ma, W.-Y. (2010). Understanding transportation modes based on gps data for web applications. *ACM Transactions on the Web (TWEB)*, 4(1):1–36.