



This is a repository copy of *Data-driven traffic assignment through density-based road-specific congestion function estimation*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/206888/>

Version: Published Version

Article:

Rocroft, A., Ramli, M.A.B. and Punzo, G. orcid.org/0000-0003-4246-9045 (2024) Data-driven traffic assignment through density-based road-specific congestion function estimation. *IEEE Access*, 12. pp. 192-205. ISSN 2169-3536

<https://doi.org/10.1109/ACCESS.2022.0122113>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Received 5 December 2023, accepted 19 December 2023, date of publication 25 December 2023,
date of current version 3 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346669

RESEARCH ARTICLE

Data-Driven Traffic Assignment Through Density-Based Road-Specific Congestion Function Estimation

ALEXANDER ROOCROFT^{1,2}, MUHAMAD AZFAR RAMLI², AND GIULIANO PUNZO¹

¹Department of Automatic Control and Systems Engineering, The University of Sheffield, S10 2TN Sheffield, U.K.

²Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632

Corresponding author: Alexander Roorcroft (aroorcroft1@sheffield.ac.uk)

This work was supported in part by The University of Sheffield; and in part by the Agency for Science, Technology and Research (A*STAR), through the A*STAR Research Attachment Program (ARAP).

ABSTRACT The ability to build accurate traffic assignment models on large-scale major road networks is essential for effective infrastructure planning. Static traffic assignment models often utilize standard formulations of congestion functions which suffer from various inaccuracies. Conversely, newer approaches in the literature rely on inverse optimisation to provide enhanced accuracy but incur significantly heavy computational costs. The work in this article develops density-based congestion function fitting in order to compute traffic assignment patterns. Computational efficiency makes the method amenable to be used on real-world networks at national scale. The methodology is applied on the motorway network connecting the primary metropolitan areas in England using Motorway Incident Detection and Automatic Signalling system data. The results demonstrate that the use of density-based congestion functions provides significant improvement in terms of computational runtime in the order of 11,000 times (22 secs vs 68 hours). Correspondingly, prediction error from this method (3.9 to 6.9% for time prediction and 10.4 to 10.7% for flow prediction) slightly outperforms the state-of-the-art Inv-Opt method (5.3 to 8.8% for time prediction and 10.5 to 11% for flow prediction). The increased accuracy provides greater confidence in modelling results for applications such as cost-benefit analysis and price of anarchy calculations.

INDEX TERMS Static traffic assignment, data-driven congestion functions, strategic road network, MIDAS.

I. INTRODUCTION

A. CONGESTION FUNCTIONS

Road networks are complex systems where drivers adopt selfish behaviours aimed at minimising their own travel costs, for instance time, disregarding the impact of such behaviours on the onset of congestion [1]. Macroscopic traffic assignment (TA) models have been used by transport planners to model driver behaviour since the middle of the twentieth century, however, the topic continues to attract significant attention from the research community [1], [2], [3]. In TA models the choice of congestion functions is a fundamental issue.

Congestion functions, also known as volume-delay functions, are smooth, monotonically increasing functions [4]

The associate editor coordinating the review of this manuscript and approving it for publication was Maurice J. Khabbaz¹.

modelling travel time in relation to traffic volumes and are used to produce flow assignment patterns through the solution of the Traffic Assignment Problem (TAP) [5].

1) ROAD-SPECIFIC FITTING

Transport agencies frequently use their own set of parameters for congestion functions, either area-wide but fitted to their specific network [6], or based on road type [7] that, once decided, tend to be rarely updated. Fitting functions to global road network data can improve TA model accuracy over using standard parameter values [8], [9], [10], although characteristics of specific road segments may get lost in the process. Evidence shows that the relationship between flow and travel time in congested conditions is highly dependent on road design [11].

Road-specific fitting to overcome this problem has seen the use of uncongested traffic flow data and travel time

from automatic detectors [12]. Yet this is only viable for non-congested traffic, where flow and travel time readings increase together. Such traffic flows are called hypo-critical, where the critical flow (i.e. capacity) is the threshold after which hyper-critical traffic occurs and travel time increases despite decreasing flows. As such, this approach struggles to represent hyper-critical congested traffic.

2) THE CHALLENGE OF HYPER-CRITICAL TRAFFIC

Fittings biased towards hypo-critical flows often lead to underestimated travel times for congested conditions [13]. It has been suggested that congestion functions should be treated merely as tools for the solution of the TAP for flow assignment patterns, as opposed to providing reliable estimates of travel time [14]. However, there are important reasons for having congestion functions that represent delays more accurately. For instance, the numerical convergence of the TAP is affected by the function's gradient. Also, travel time estimates are of key importance in testing policies concerned with vehicle speeds (e.g. air quality and emissions, cost-benefit analysis) [15]. Therefore, congestion functions that can more accurately represent travel time in all traffic regimes are necessary.

The delay predictions of congestion functions may be misleading in hyper-critical traffic as the functions are defined both above and below capacity flow, which is not physically possible. To avoid such misunderstanding, the functions can be thought to use flow demand to calculate the travel time of a road. This can be interpreted as the number of vehicles wishing to use the road, accepting that whenever the flow demand is greater than capacity significant delays are experienced [16].

3) ALTERNATIVE FITTING TECHNIQUES FOR HYPER-CRITICAL TRAFFIC

More accurate representations of the hyper-critical traffic regime have been obtained by estimating the non-observable flow demand (i.e. greater than capacity) from observable measurements [17]. These include incorporating queue-based theory into the estimation [15], using the queues measured by loop detectors at bottlenecks to provide the number of vehicles surplus to capacity that can be used to approximate flow demand. While promising, this approach has proven hard to implement, being difficult to measure traffic flow at every bottleneck point in non-stationary real-world traffic.

Alternatively, traffic density can be taken as a proxy for flow demand. By using a density-based approach to congestion function fitting, [13] introduced a more realistic estimation of congestion functions that can account for both hypo- and hyper-critical traffic regimes. Such a density-based approach has been found to lead to a lower error than the queue-based approach across different types of roads and areas [16].

Prominent works on hyper-critical fitting or road-specific functions are not often concerned with quantifying the

impact of these formulations on TAP solutions [12], [13] and are often limited to a small number of roads or traffic sensors. Furthermore, both [12] and [13] only investigated the suitability of the commonly used Bureau of Public Roads (BPR) formulation for these types of fitting.

The BPR formulation often uses a set of standard coefficients ($\alpha = 0.15, \beta = 4$) suggested at its inception [18]. Different candidate forms for congestion functions have been developed in an attempt to address the perceived shortcomings of BPR and to better incorporate road characteristics, these include Conical [4], Akçelik [11], and Exponential [15]. In countries such as England and the USA, highway authorities recommend using an empirically modified version of the Akçelik function, which depends on factors relating to road features [7], [19]. However, for long uninterrupted stretches of road such as motorways, the standard simple BPR formulation has been found to perform well and remains a popular choice [9], [20].

An open research question remains in identifying the most appropriate choice of function form for road-specific density-based fitting on a range of highways across a large-scale road network. This work aims to reveal some insight into this, how it compares to flow-based fittings, and the effect it has when used in a data-driven static TA model.

B. DATA-DRIVEN TRAFFIC ASSIGNMENT MODEL

1) MODEL INPUTS

In addressing the impact of congestion functions on the TA modelling, this work considers solely using loop detector data to build a fully data-driven model. A prominent example of this type of modelling approach is the static TA model used in [21] to effectively calculate flow patterns on the Eastern Massachusetts highway network. Similar models have also been applied to portions of the England Strategic Road Network (SRN) [22], [23].

The three key inputs of static TA models include 1) the road network topography; 2) the Origin-Destination (O-D) matrix of traffic demand within the network; 3) the congestion functions. These are used to produce the primary outputs of expected traffic flow and travel time for vehicles across each road segment for traffic assignment patterns. These traffic assignment patterns can include user-equilibrium and system-optimal.

2) TRAFFIC ASSIGNMENT PATTERNS

The user-equilibrium, also known as Wardrop equilibrium, is an assignment pattern where drivers pursue their selfish best route and no individual driver can improve their travel cost by unilaterally changing their routing. This flow pattern is frequently observed on real-world networks and it is commonly assumed that it matches the observed flows [21], [24]. System-optimal is the flow pattern under which the global cost of all drivers is minimised. To quantify the difference in cost between these routing patterns and highlight the loss of performance on the network, the price

of anarchy (POA) is used as a metric to compare the two [3], [25].

3) MACHINE LEARNING-BASED GENERALISED CONGESTION FUNCTION ESTIMATION

The model in [21] utilised a machine learning-based technique of fitting a generalized congestion function using Inverse-Optimisation (Inv-Opt) that fits only a single congestion function to all edges on a network specific to a particular time period of the day. Utilising a computationally intensive optimisation problem formulation, the method was only applied to a small 24-node network. This state-of-the-art method is the benchmark that the density-based road-specific congestion functions are compared against on a large network.

The investigated congestion functions are used in the calculation of the traffic flow pattern and within an O-D matrix congestion adjustment. The effect of using density-based fitting on the computation time and accuracy of recreating the observed flows is compared with machine learning-based Inv-Opt. The performance informs the suitability for modelling system-optimal flows.

C. SUMMARY OF CONTRIBUTION

This work utilizes the density-based fitting of congestion functions that are road-specific in order to improve the accuracy of TA model prediction and the calculation time over large highway networks. The techniques are applied to a sample subnetwork connecting the main metropolitan areas in England, using traffic speed, occupancy and flow count data obtained from the Motorway Incident Detection and Automatic Signalling (MIDAS) system used by National Highways on the National Traffic Information Service (NTIS) model of the SRN.

For the tested network, it is shown how the BPR is the best candidate function form when the individual road congestion functions are validated. Also, it is demonstrated that the incorporation of the proposed density-based BPR fitted congestion function compares favourably to other state-of-the-art methods for calculating user-equilibrium flows and associated travel times. At the same time, it remains computationally tractable and applicable to large networks. These results imply the approach is more suitable for use on similar real-world major road networks.

Specifically, the primary contributions of this work are:

- The identification of the functional form of congestion function that best captures the delay-flow demand relation on a large sample of road segments from the England SRN, comparing between the density-based and flow-based fitting approaches.
- The analysis of fitted road-specific congestion function parameters showing the advantages of density-based fitting over flow-based by allowing parameters to be fitted largely independently, which leads to more suitable functions for use in the TAP.

- The extension of methods to calculate TA traffic patterns by including a density-based road-specific congestion function fitting that is suited to large-scale real-world networks. This highlights the benefit of fitting road-specific parameters using density instead of using standard values.
- The benchmarking of the method against the current state-of-the-art for use on SRNs, resulting in a favourable comparison, especially in the trade-off between accuracy and computational time.

The contributions provide a more accurate and efficient data-driven static TA model that can be fitted to any time window of loop detector data and used for strategic planning. Whilst dynamic TA models have the advantage of modelling more complex phenomena such as spillbacks, in their current state they lack the convergence properties required to be practical in the context of strategic planning [26]. The ability to compute flow patterns for real-world large-scale transportation networks is necessary for testing strategic interventions, such as adaptive road charges, which aim to optimize traffic flows and enhance transportation efficiency. By improving congestion function estimation, this work advances the state of the art in transportation modeling, offering practical solutions with real-world applications.

The overall structure of this paper is summarized as follows. Section II introduces the network definition and mathematical notation used throughout the paper. Section III describes the overall methodology for creating a static TA model using network and loop detector data, and provides a summary description of the MIDAS and NTIS datasets used for the case study. In Section IV the main results are presented. Lastly, the paper is concluded with a discussion in Section V, followed by a short summary conclusion in Section VI.

II. PRELIMINARIES AND NOTATION

A. NOTATION

In this work, all the vectors are column vectors. For example, the column vector \mathbf{x} is written as $\mathbf{x} = \{x_i, \dots, x_{\dim(\mathbf{x})}\}$, where $\dim(\mathbf{x})$ is the dimension of \mathbf{x} . The work uses “prime” (e.g. \mathbf{x}') to denote the transpose of a matrix or vector. \mathbb{R}_+ denotes the set of all nonnegative real numbers. Matrix $\mathbf{Q} \geq \mathbf{0}$ or vector $\mathbf{x} \geq \mathbf{0}$ indicates that all entries of a matrix \mathbf{Q} or vector \mathbf{x} are nonnegative. Also, $|\mathcal{X}|$ represents the cardinality of a set \mathcal{X} .

B. NETWORK DEFINITION

The road network is modelled as a directed graph with a set of nodes \mathcal{V} and a set of edges \mathcal{A} . The nodes represent the interchanges of the road system and the edges are the roads connecting them. The model assumes the graph is strongly connected and is defined by the node-edge incidence matrix with $N \in \{0, 1, -1\}^{(|\mathcal{V}| \times |\mathcal{A}|)}$. On road networks in general and the England SRN in particular, there is a path between all pairs of nodes so the assumption is valid.

TABLE 1. Notation for network definition.

Symbol	Definition
\mathcal{V}	Set of Nodes
\mathcal{A}	Set of Edges
\mathcal{W}	Set of O-D Pairs
\mathcal{F}	Set of Feasible Flow Vectors
\mathcal{D}	Set of Hourly Observations
\mathbf{N}	Node-edge Incidence Matrix
x_a	Flow on Edge $a \in \mathcal{A}$

The demand for movement between O-D pairs on the network is represented by $d^{\mathbf{w}} \geq 0$ with $\mathbf{w} = (w_s, w_t)$ the O-D pair of nodes such that $\mathcal{W} = \{\mathbf{w}_i : \mathbf{w}_i = (w_{si}, w_{ti}), i = 1, \dots, |\mathcal{W}|\}$. $\mathbf{d}^{\mathbf{w}} \in \mathbb{R}^{|\mathcal{V}|}$ is a vector with all zeros except for a $-d^{\mathbf{w}}$ for node w_s and a $d^{\mathbf{w}}$ for node w_t .

Let \mathbf{x} be the vector of the total edge flow x_a on edge $a \in \mathcal{A}$. Then the set of feasible flow vectors \mathcal{F} is defined by:

$$\mathcal{F} \stackrel{\text{def}}{=} \{\mathbf{x} : \exists \mathbf{x}^{\mathbf{w}} \in \mathbb{R}_+^{|\mathcal{A}|} \text{ s.t. } \mathbf{x} = \sum_{\mathbf{w} \in \mathcal{W}} \mathbf{x}^{\mathbf{w}}, \mathbf{N}\mathbf{x}^{\mathbf{w}} = \mathbf{d}^{\mathbf{w}}, \forall \mathbf{w} \in \mathcal{W}\}$$

where $\mathbf{x}^{\mathbf{w}}$ indicates the flow vector attributed to O-D pair \mathbf{w} . This implies that the total flow vector \mathbf{x} is consistent with the demands $\mathbf{d}^{\mathbf{w}}$ between all O-D pairs.

\mathcal{D} is the set of hourly average observations in the measured data used in the congestion function fittings. A collection of the network variables is provided in Table 1.

III. METHODS AND DATA

The process to calculate the user-equilibrium traffic flow assignment from a TA model derived from raw loop detector data consists of four steps. After collecting the traffic data and associating them with the network model, a simplified, topographic representation of the motorway network is extracted. This is the network model on which the O-D matrix can be calculated as well as the congestion functions for the solution of the TAP, which is then addressed in the fourth and final step. This process is presented in Figure 1.

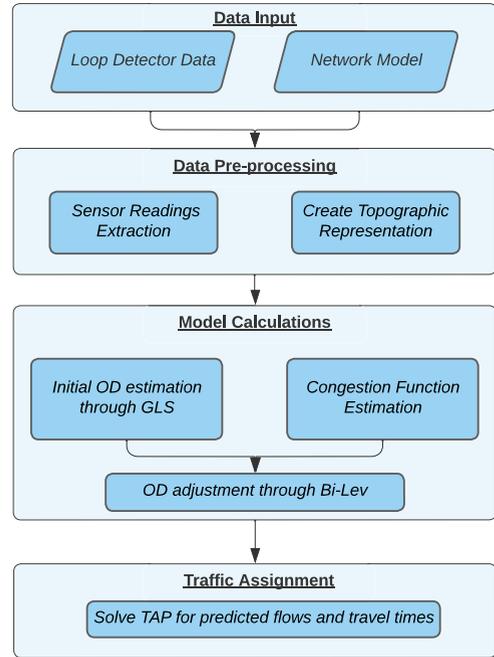
A. CONGESTION FUNCTIONS

Accurate congestion functions are key to TA models as they connect the travel time t_a to the vehicle flow demand \check{x}_a on edge $a \in \mathcal{A}$. Note the difference between flow demand \check{x}_a and flow x_a . Flow demand indicates the number of vehicles wishing to use the edge in a period of time. Under congested conditions, this may be greater than the flow ($\check{x}_a \geq x_a$), as flow cannot exceed capacity but flow demand can.

In the network model, congestion functions take the form:

$$t_a(\check{x}_a) = t_a^0 g\left(\frac{\check{x}_a}{m_a}\right), \quad (1)$$

where t_a^0 is the free-flow travel time of an edge $a \in \mathcal{A}$ and $g(\cdot)$, also known as the travel time multiplier, is a strictly increasing and continuously differentiable function dependent on the flow demand \check{x}_a divided by the flow capacity m_a of that edge $a \in \mathcal{A}$ (i.e. the saturation rate).


FIGURE 1. Procedure to obtain data-driven traffic assignment (TA) model on real-world road networks from loop detector data.

There are several main candidates for the form of the travel time multiplier $g(\cdot)$ that have been developed with the required traits for TA.

1) BPR

The first function, BPR, is consistent with Eq. 1 and widely used in TA models [24], [25]. In its more general form, it is:

$$t_a = t_a^0 \left(1 + \alpha \left(\frac{\check{x}_a}{m_a} \right)^\beta \right), \quad (2)$$

where the values of α and β are coefficients of the model commonly taken as 0.15 and 4, respectively [27]. In this work the form of BPR with these standard coefficients is known as *BPR – Standard*.

2) CONICAL

Secondly, conical is a function that was developed after BPR to address some of its perceived issues [4]. It was based on a simple mathematical derivation to meet the postulated requirements of well-behaved congestion functions. The function is:

$$t_a = t_a^0 \left(2 + \sqrt{\alpha^2 (1 - \check{x}_a/m_a)^2 + \beta^2} - \alpha (1 - \check{x}_a/m_a) - \beta \right), \quad (3)$$

where

$$\beta = \frac{2\alpha - 1}{2\alpha - 2},$$

and $\alpha > 1$.

3) AKÇELIK

Thirdly, Akçelik is a function form which has been used in practice to improve the BPR further by representing better the hyper-critical and hypo-critical flow regimes [11]. In previous works it has been shown to be more accurate in representing road facilities with signalised intersections [9]. The form of the function used in this work is a simplified version based on [15] and [28]:

$$t_a = t_a^0 + \left[0.25 \left(\frac{\check{x}_a}{m_a} - 1 \right) + \sqrt{\left(\frac{\check{x}_a}{m_a} - 1 \right)^2 + J \frac{\check{x}_a}{m_a}} \right], \quad (4)$$

where J is the delay parameter to be fitted, in this work it subsumes constants present in other formulations. In this equation, t_a is the average travel time per unit distance (hr/km) and t_a^0 is the free-flow travel time per unit distance (hr/km).

4) EXPONENTIAL

In other works, exponential functions are used as an alternative to BPR as they possess a similar shape [15], and in some studies they fit traffic with greater accuracy [3]. In this work, the form of exponential function used is:

$$t_a = t_a^0 e^{\alpha(\check{x}_a/m_a)^\beta}, \quad (5)$$

where the values of α and β are coefficients to be fitted.

B. DENSITY-BASED FITTING OF CONGESTION FUNCTIONS

To estimate the form of $g(\cdot)$ based on observed traffic, a density-based fitting of the congestion function can be used to calculate specific functions for each edge of the network (see [13] for details).

The nature of congestion on roads means that, past the onset of congestion, flows decrease with increasing travel time. The onset of congestion corresponds to the flow reaching capacity, the density at this point, $k_a(m_a)$, for an edge $a \in \mathcal{A}$ is called critical. It separates the hyper-critical ($k > k_a(m_a)$) from the hypo-critical ($k < k_a(m_a)$) regime. In the hyper-critical regime, the travel time increases with decreasing flow and the flow-to-capacity ratio does not exceed unit value, as the congestion function would instead indicate (Figure 2 (a)). This makes fitting a monotonic congestion function to flow-time data not possible. In contrast, when the non-dimensionalised travel time (t_a/t_a^0) is plotted against density, the hyper-critical region has travel times that increase as density increases so the congestion function can be fitted (Figure 2 (b)).

This method uses data to estimate the traffic density which can transform the congestion function to a form that a curve can be fitted and specific estimates of parameters obtained. The method assumes that the flow demand \check{x}_a of the congestion function is proportional to density k_a such that the

following mapping is assumed:

$$\check{x}_a = m_a \frac{k_a}{k_a(m_a)} \quad (6)$$

Flow is proportional to density in the hypo-critical region of the fundamental diagram (Figure 3), however in the hyper-critical it is not. The aim is to use density as a proxy for the number of vehicles wishing to use a road, \check{x}_a , assuming Eq. 6 holds for the hyper-critical region.

This mapping is applied to all the previously listed congestion function forms to fit them using density. In the example of BPR, the mapping transforms Eq. 2 into Eq. 7. The same α and β values can be used for both equations. The BPR expressed in terms of density is:

$$t_a = t_a^0 \left(1 + \alpha \left(\frac{k_a}{k_a(m_a)} \right)^\beta \right), \quad (7)$$

where, for an edge $a \in \mathcal{A}$, k_a is density of an edge $a \in \mathcal{A}$ and the critical density-at-capacity is $k_a(m_a)$. By using the relation of speed to edge length l and travel time ($v = l/t$), the BPR expression reformulated in terms of average speed and density is:

$$\hat{v}_a = \frac{v_a^0}{\left(1 + \alpha \left(\frac{k_a}{k_a(m_a)} \right)^\beta \right)}, \quad (8)$$

where, for an edge $a \in \mathcal{A}$, the free-flow speed is v_a^0 , and \hat{v}_a is the computed theoretical speed. The values of α and β can be fitted using a non-linear least-squares approach which computes the sum of squared difference between modelled and observed speeds in the set of observations \mathcal{D} :

$$\arg \min_{\alpha, \beta} \left(\sum_{d \in \mathcal{D}} |\hat{v}_{a,d} - v_{a,d}|^2 \right) \quad (9)$$

The data used in the estimation include all daytime measurements together. Night-time data are excluded as they rarely present congested flow conditions and so would bias the result. The per-minute observations are averaged with 60-min mean values to remove outliers to steady-state conditions. The fitting is only applied to edges with sufficient data in the hyper-critical congested region. Edges without congested data assume the standard values of α and β .

Estimating the traffic density using loop detector data is limited by the instantaneous time-mean measurements [29]. Accurate space-mean measurements of density are only practically available with aerial photography [30]. In this work, the measured vehicle occupancy is used to estimate the traffic density as is commonly done in practice through Eq. 10 [31], [32].

$$k = \frac{\rho}{L_v + L_s} \times C_{lanes}, \quad (10)$$

where ρ is the measured mean lane occupancy (the fraction of time the detector has a vehicle above). L_v is the mean length of vehicle and L_s is the length of sensor. C_{lanes} is the number of lanes for the road.

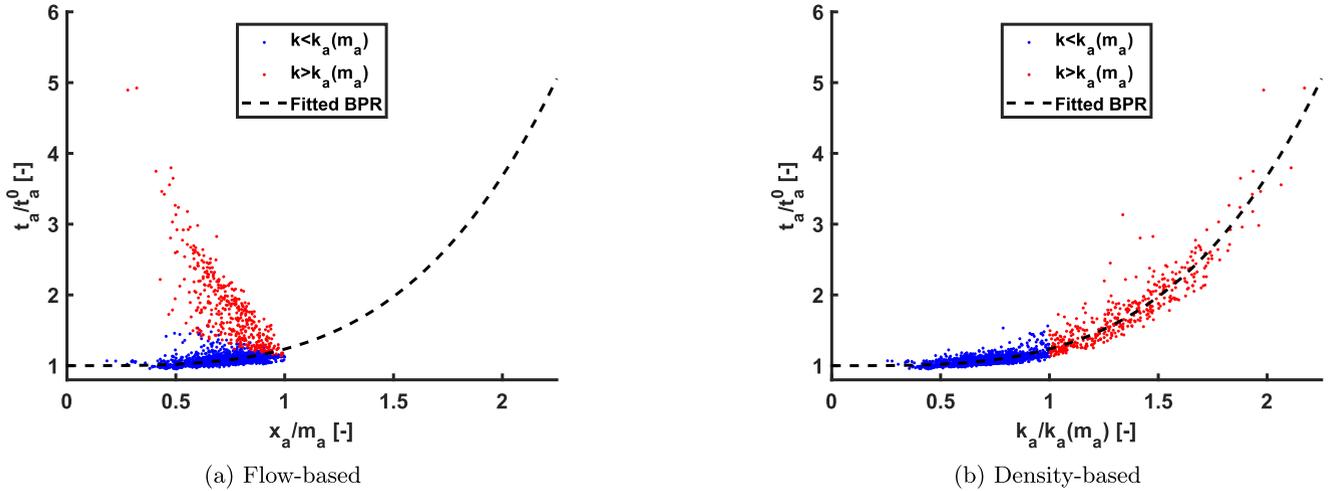


FIGURE 2. Example of hypo-critical (blue) and hyper-critical (red) observations for: (a) Hourly non-dimensional travel time (t_a/t_a^0) against hourly flow/capacity (x_a/m_a); (b) Hourly non-dimensional travel time (t_a/t_a^0) against hourly density/critical density ($k_a/k_a(m_a)$). The Bureau of Public Roads (BPR) function is fitted to the data in (b) and also plotted in (a) for comparison. The measurements are the hourly average traffic of an edge on the subnetwork on the weekdays selected for analysis between September 2018 and May 2019.

For the MIDAS system, the sensor length is 2m. The mean vehicle length is calculated by using the vehicle class-specific flow data to calculate a weighted average vehicle length for each minute recorded by the system.

During fitting, the authors in [13] suggested applying a weighting, as the number of data points recorded for hyper-critical congested flow ($k_a > k_a(m_a)$) is dwarfed by that of hypo-critical flow ($k_a < k_a(m_a)$). In this work, the values are not weighted as experiments with the data showed the effect to be limited on the TA results. The same authors also suggest including density-at-capacity and free-flow speed as variables to optimise in Eq. 9, however, the effect on the results of this approach was also limited so is not used. Time bin-specific congestion functions were considered, however, this reduced the number of data points and led to worse performance than combining all daytime measurements.

C. INVERSE-OPTIMISATION CONGESTION FUNCTION ESTIMATION

Acting here as a performance benchmark for the TA model, the Inverse-Optimization (Inv-Opt) of the TAP estimates a general function $g(\cdot)$ for all links $a \in \mathcal{A}$ for each analysis time period (e.g. AM 6 am - 10 am) [21]. Inv-Opt assumes the measurements of the time bin average flows on the roads are the user-equilibrium flows and solutions to the TAP for a specific congestion function and O-D demand matrix. It aims to find this congestion function such that the resulting calculated user-equilibrium flows are as close as possible to the observed measurements. It incorporates support vector regression with a polynomial kernel to produce a function of the form:

$$\hat{g}(x) = \sum_{i=0}^n \beta_i^* x^i = 1 + \sum_{i=1}^n \beta_i^* x^i, \quad (11)$$

where an optimal β^* is obtained by solving the Inv-Opt quadratic programming problem. n is a hyperparameter to be selected and is the order of the polynomial congestion function.

The implemented code is taken from [33] which includes an additional normalisation constraint to match the total cost of the fitted congestion function on all edges to that of a BPR with standard coefficients.

D. CALCULATING FREE-FLOW TRAVEL TIME, CAPACITY AND DENSITY-AT-CAPACITY

In this work, as in [34], the proposed congestion function estimation uses the maximum of the observed flows on an edge as its capacity. The density-at-capacity $k_a(m_a)$ is then obtained as the minimum density, across all the observations, corresponding to such a flow. This obtains the peak of the rising free-flow branch of the fundamental diagram (Figure 3). The NTIS-provided values of capacity are used for edges without sufficient congestion data for this estimation.

The free-flow travel time t_a^0 is obtained by taking the 95th percentile of the observed speeds [12], [21], [35] as the free-flow speed then converting it to the travel time through the edge length.

E. TRAFFIC ASSIGNMENT MODEL CALCULATIONS

1) ESTIMATING THE O-D DEMAND MATRIX

The Generalised Least Squares (GLS) method, together with the Bi-Level optimisation problem (BiLev) adjustment algorithm, is used to estimate the key input of the O-D demand matrix without a demand survey using only mean hourly flow counts. The method used in this work is similar to [21]. It is computationally limited by the size of network (i.e. number of O-D pairs) the optimisation can be applied to and this limits the size of network investigated in this work [23]. Furthermore, to obtain reliable demand estimates,

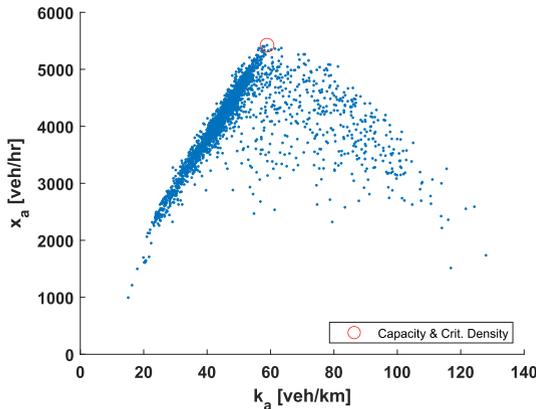


FIGURE 3. Example fundamental diagram for obtaining capacity m_a (which is taken as maximum observed flow rate x_a) denoted by the red circle in the plot and $k_a(m_a)$ (associated density-at-capacity) from the flow (x_a) vs. estimated density (k_a). The measurements are the observed hourly average traffic of an edge of the subnetwork on the weekdays selected for analysis between September 2018 and May 2019.

it needs a large enough sample of the mean hourly flow counts that depends on the size of the network. This limits how small the analysis window can be to obtain sufficient days of data.

To obtain different static demand profiles, the weekday flow data are grouped into time bins of distinct periods, AM: 6 am – 10 am, MD (midday): 10 am – 4 pm, PM: 4 pm – 8 pm. For each time bin, the mean hourly flow is calculated over the respective period. The time bins are chosen to encapsulate the demand for the morning and evening commuting traffic and the middle of the day period. These periods are of interest for the strategic analysis of congestion and they are similar to those used in other data-driven TA models [21]. By having three different time bins, it allows testing of the TA model performance of the alternative congestion function estimation methods in different demand scenarios.

2) FLOW PATTERN CALCULATION

The predicted user-equilibrium flow pattern can be calculated using the adjusted O-D demand matrix and evaluated congestion functions through the Frank-Wolfe algorithm. This is done with the following optimisation of the flow-based TAP [5]:

$$\min_{\mathbf{x} \in \mathcal{F}} T^{UE}(\mathbf{x}) = \sum_{a \in \mathcal{A}} \int_0^{x_a} t_a(s) ds \quad (12)$$

The Frank-Wolfe algorithm implemented uses a convergence criterion based on the size of relative gap between consecutive iterations [5]. A non-dimensional relative gap of $\epsilon = 10^{-5}$ is used for the convergence of the edge flows, as that has been shown to be sufficient in previous analysis [26].

The user-equilibrium flow pattern results from drivers pursuing their selfish best route and throughout this work it is assumed to match the observed flows as commonly done in other works [21], [24]. Stochastic and bounded rationality user-equilibria may provide more realistic driver behaviour

with increased computational complexity. However, deterministic user-equilibrium has been found to match observed traffic patterns well at the network scale [36]. The difference between the predicted user-equilibrium flow and travel time patterns compared to the observed flows and travel times is used as a measure of the suitability of a congestion function method in TA models.

For a flow pattern, the Total System Travel Time (TSTT) (i.e. total generalised network cost) for the network is calculated using:

$$C(\mathbf{x}) = \sum_{a \in \mathcal{A}} x_a t_a(x_a). \quad (13)$$

This quantity is a measure of the total cost for all drivers on the network for a time bin.

F. THE ENGLAND STRATEGIC ROAD NETWORK

1) RAW DATASET DESCRIPTION

This work uses traffic data obtained through the MIDAS system installed on key motorways of the England SRN. The dataset analysed includes nine months of weekdays from September 2018 to May 2019 (excluding public holidays).

Most of the traffic data from MIDAS is obtained through under-road inductive loops spaced approximately every 500m. At approximately 7000 site on the SRN, the system measures speed, flow, occupancy, and headway. Data is aggregated over 1-minute intervals and provided on a per-lane basis.

The NTIS Network and Asset Model contains information on the different systems National Highways uses to monitor traffic on the SRN. It contains information on the location of MIDAS sensors and geospatial information of the road junctions and motorways that can be converted into a topographic representation of the network [37]. Also, it contains information on the direction of travel, capacity and length of the associated weighted graph’s edges.

As a motorway system, it is assumed that the SRN does not have intersection control devices such as traffic lights [37], so the TA model does not need to account for associated delays at nodes. In general, TA models accounting for node delays have greater accuracy on road networks with such devices (e.g. urban areas) [1].

The central section of the SRN network was chosen for analysis (Figure 4). This portion encompasses the primary roads with relevant MIDAS sensor sites that connect major cities in England. It was selected due to its extensive coverage by the MIDAS system, facilitating the creation of a fully connected network suitable for a data-driven TA model. Additionally, its size was conducive to the O-D estimation techniques applied. By connecting some of England’s most heavily populated urban centers, such as London, Birmingham, Manchester, and Leeds, this subnetwork enables the exploration of a significant segment of SRN traffic, which in itself accounts for one-third of the total national vehicle miles travelled [38].

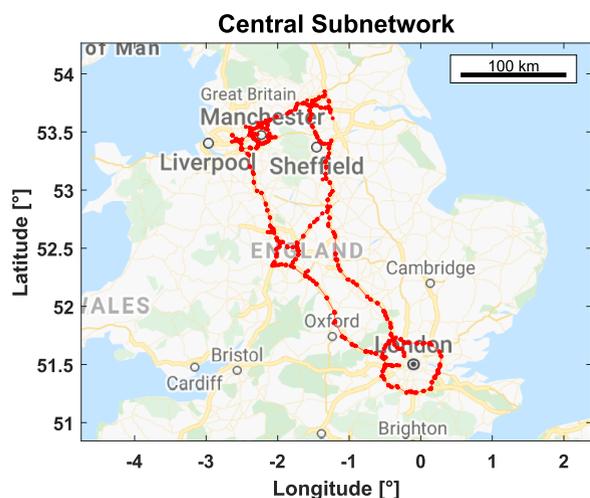


FIGURE 4. Graph Representation of the National Traffic Information Service (NTIS) model of the central England Strategic Road Network (SRN). Map underlay obtained from Google Maps [39].

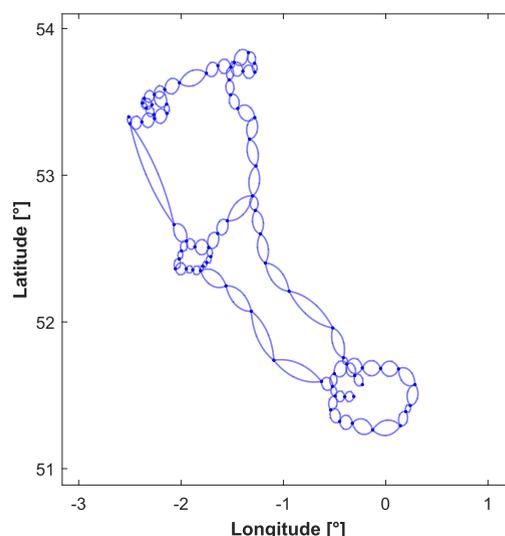


FIGURE 5. Topographic representation of the subnetwork of the main roads connecting the central England Strategic Road Network SRN.

2) NETWORK GRAPH TOPOGRAPHIC REPRESENTATION

The purpose of the TA model is to represent the overall flows of traffic around the network. It is not focused on how the vehicles navigate through the junctions between roads. For this reason, a processed degenerate arterial road topographic representation is created based on the NTIS road model. In this processed version of the NTIS model the junctions and interchanges are simplified to single supernodes and the carriageways connecting them are grouped as single superedges. After the process of node and edge combination, the supernodes and superedges that constitute the simplified topographic representation are referred to as its nodes and edges.

The topographic representation of the central subnetwork of the SRN is presented in Figure 5. It has 73 nodes, 156 edges and 5256 O-D pairs. Due to limitations on the MIDAS data and computational requirements, a number of minor roads and junctions are omitted.

The number of nodes and edges of the topographic representation is too many to compute Inv-Opt. The largest previously considered network for Inv-Opt was composed of 24 edges. As in previous works, to reduce computational difficulty the functions are fitted to the last 75 days of data on a further simplified network (see Appendix), which is then applied to the full topographic representation [21].

3) MIDAS DATA EXTRACTION

Using the NTIS dataset, available sensor data are extracted for the associated topographic edges. For each weekday time bin (AM, MD, PM), the mean hourly flow is calculated over the respective period. Data were omitted from dates which coincided with public holidays, days with major weather disruption (i.e. snow on 01/02/2019), and around the first and last week of the year due to unrepresentative holiday demand disruption (20/12/2018 - 07/01/2019).

Loop detector data can contain errors and needs to be processed before use [40]. To do this, the median flow

readings are used when multiple sensors are available on the same edge. This is to filter out erroneous readings, identified as those with a value of more than twice the median absolute deviation from the median. This assists the central tendency of measured flows to be resistant to individual sensors with faults and those that do not measure the main carriageway.

Due to reliability issues with loop detector data, the data available from the MIDAS sensors can change over the time window of analysis. The data available for each edge are assessed, and days of data that are missing measurements (i.e. zero flow) for any of the edges in any of the time bins are excluded from the final data set. If the sensors on a particular edge are consistently returning faulty readings, then the topographic graph is amended to absorb that edge into its neighbour. This is rare and usually occurs when a short edge is monitored by a single faulty sensor. The absorption of neighbouring edges omits intermediate junctions, reducing the accuracy of the topographic representation of the SRN; however, it produces an approximation for a test network suitable for the investigations comparing different congestion function estimation techniques.

From the period September 2018 to May 2019, the final number of used weekdays of data was 166 out of a possible 273 days, which is sufficient for the application of the OD demand estimation. The analysis performed on this data window captures an average trend in traffic on the network, useful for the high-level strategic analysis of structural congestion suited to static TA models.

The computation time for extracting and processing (all time bins) each day of MIDAS data was on average 12.5 mins run in parallel on a Dell PowerEdge C6320 with 2.4GHz Intel Xeon E5-2630 v3 CPU and 4GB RAM.

4) SELECTION OF EDGES TO APPLY ROAD-SPECIFIC FITTING

The edges that road-specific fitting is applied to are best limited to those with suitable data measurements to improve

accuracy. Those considered not suitable are: 1) edges without data in the hyper-critical regime (17 edges); 2) Edges with the peak of the flows missed by the sensor system (1 edge); 3) edges where it appears multiple speed limits have been in operation (1 edge). The edges can have a combination of the problems. In total, 18 (12%) of the edges are unsuitable, leaving 138 to apply the method to.

IV. RESULTS

A. CHOICE OF FUNCTION FOR DENSITY-BASED FITTING

1) ALTERNATIVE FUNCTIONS

In this section, the alternative forms for fitting the congestion functions are tested.

The fitting was applied to BPR, Conical, Akçelik and Exponential with the dependent variable trialled with density, all the flow and only the hypo-critical flow region (hypo-flow). As a non-linear least squares regression is used to fit the parameters, the goodness-of-fit is assessed using the Root Mean Square Error (RMSE) of the predicted speed values on each edge selected for fitting $a' \in \mathcal{A}$.

$$RMSE_{a'} = \sqrt{\frac{\sum_{i=1}^n (v_{i,a'}^{obs} - \hat{v}_{i,a'})^2}{n}}, \quad (14)$$

where the $v_{i,a'}^{obs}$ is the observed hourly average speed during time-bins AM, MD and PM. $\hat{v}_{i,a'}$ is the predicted average speed based on the observed hourly flow. $i \in \mathcal{D}$ is the hourly observation and n is the total number of observations.

The range of RMSE values for the fittings on the selected 138 edges of the sample subnetwork show that across the different approaches to fitting, density-based fitting has the lowest RMSE value (Figure 6). Within density-based, the best errors are found for the BPR and exponential function forms, with BPR slightly better. For flow and hypo-flow fittings, it can be seen that all forms perform similarly.

From these results, it can be concluded that BPR is the best choice of congestion function form within the tested functions as it consistently has a low error when applied systematically using density-based fitting. This finding is in line with previous research which has suggested that BPR is the best choice for uninterrupted flow facilities such as motorways [9].

BPR and exponential have a similar shape which explains why they perform similarly. Due to the slightly better result of BPR in Figure 6 (c) and the wide adoption of BPR in transport planning software it is BPR that is used for further analysis of density-based fitting.

B. BPR PARAMETER RANGE AND CORRELATION

The convergence of the TAP is determined by the range of values that are fitted to the BPR function. In Figure 7 it can be seen that the distribution of α and β are different when fitting with density, flow and hypo-flow. The flow-based fitting can be seen to have the largest spread of values. A large amount

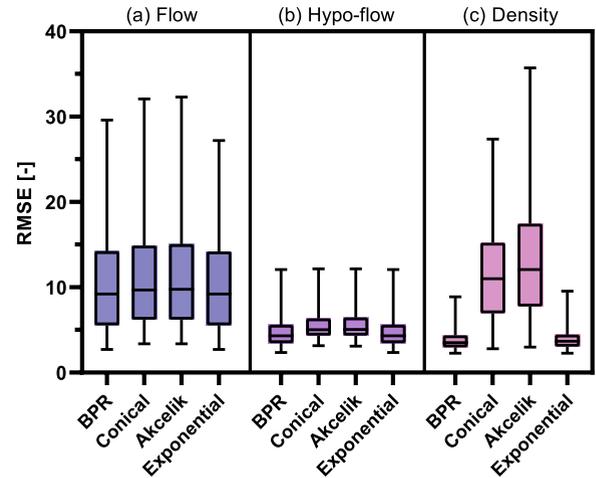


FIGURE 6. Distribution of Root Mean Square Error (RMSE) values with fitting different function forms to the edges of the subnetwork with different traffic variables: (a) Flow; (b) Hypocritical flow only; (c) Density. The whiskers of the boxplots represent the range, the box is the IQR and median. The predicted average speeds, based on the observed hourly flows, are compared with the observed hourly speeds using measurements from the weekdays selected for analysis between September 2018 and May 2019.

of the β values are close to 1 for the flow-based and hypo-flow-based fittings, this could have a negative impact on the convergence of TAP as the functions may not sufficiently encourage the redistribution of vehicles exceeding capacity on those edges. The fitting of β is limited to not going below 1, as this would cause convergence issues in the TAP. The β values for the density-based fitting are all greater than 2, which means they should sufficiently encourage the redistribution of vehicles exceeding capacity. Furthermore, they are in accordance with the values of β mostly used in practice, typically between 2 and 12 [41].

A key difference between the BPR parameters fitted with density and those fitted with flow is that the Pearson rank correlation coefficient shows that the values of α and β are essentially independent for density-based fitting ($r=0.009$), whereas for flow ($r=0.693$) and hypo-flow ($r=0.727$) fitting there is a large amount of positive correlation.

As can be seen in Figure 2, for fitting to flow and hypo-flow there are no data points with a saturation rate above 1 ($x_a/m_a > 1$), because flow cannot exceed capacity. This means there is an absence of information for fitting the region of the BPR where $\check{x}_a/m_a > 1$. This region has the most influence on β , the parameter which represents how quickly delays increase in hyper-critical conditions. The value of α has more influence representing delays in the hypo-critical regime $\check{x}_a/m_a \leq 1$. When fitting to flow and hypo-flow, both α and β are fitted to the same data points with saturation rates of less than 1 ($x_a/m_a < 1$), which leads to them exhibiting correlation. However, with its approximation of flow demand in the hyper-critical regime, density-based fitting does not have this problem and its parameters can be fitted with almost no correlation. The density-based fitting's β values can more accurately represent how flow demand increases in congested hyper-critical conditions.

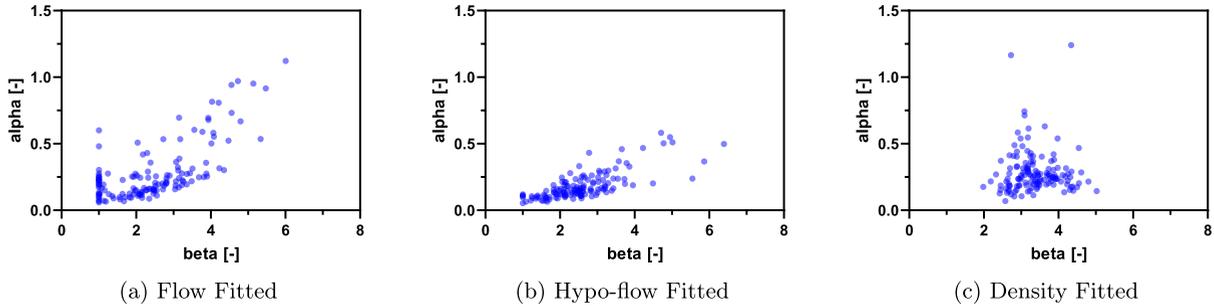


FIGURE 7. Distribution of Bureau of Public Roads (BPR) congestion function coefficient values fitted with different traffic variable: (a) Flow; (b) Hypo-flow; (c) Density. The fittings are applied to the observed hourly average traffic of each edge of the subnetwork on the weekdays selected for analysis between September 2018 and May 2019.

C. FITTED CAPACITY AND FREE-FLOW SPEED

In addition to the choice of congestion function form and parameters, the road capacity and free-flow speed (and travel time) have a strong impact on TA results.

On the subnetwork, the distribution of fitted capacities shows a wide range of values, from around 2700 to 8300, with the highest frequency of capacities between 4000 and 8000 (Figure 8 (a)). Compared to the capacity values provided by the NTIS model, which tend to group around either 6500 or 8500, there is a wider spread of values. On average the fitted values are 17% lower than the values provided in the NTIS model, which is not as large a difference as the 66% average decrease of fitted values from reference table values in [12].

The free-flow speed results show a spread between 90-120 km/h around the 113 km/h (70mph) speed limit (Figure 8 (b)). This result shows that using a free-flow travel time based on posted speed limit is inaccurate, as there is variation between the roads which could be dependent on road-specific features such as road condition, curvature and position in the network. Accuracy in estimating the resulting free-flow travel time is instrumental in the solution of the TAP.

D. USER-EQUILIBRIUM ASSIGNMENT PREDICTION

Relative errors in the flow and travel times of the user-equilibrium assignment prediction are used to evaluate the performance of the density-fitted BPR method of congestion function estimation (BPR-Density). A comparison is made using BPR with standard coefficients ($\alpha = 0.15$, $\beta = 4$) for all edges (BPR-Standard) and the benchmark of Inv-Opt previously used in this type of data-driven static TA model. The comparison is made on the subnetwork using MIDAS measurements from the period September 2018 to May 2019. For BPR-Density, the edges with the problematic data identified as not suitable for fitting assume the standard coefficients of $\alpha = 0.15$ and $\beta = 4$. All methods take the NTIS values of capacity for these edges.

The Absolute Percentage Errors (APE) are calculated as:

$$APE_{p,a}^t = \frac{|t_{p,a}^{user} - t_{p,a}^{obs}|}{t_{p,a}^{obs}}, \quad (15)$$

for travel time, while

$$APE_{p,a}^x = \frac{|x_{p,a}^{user} - x_{p,a}^{obs}|}{x_{p,a}^{obs}}, \quad (16)$$

is used for flows. For each time bin p and edge a , $x_{p,a}^{obs}$ is the observed flow and $t_{p,a}^{obs}$ is the travel time derived from observed speed. The values are the mean within each time bin over the fitting period. $t_{p,a}^{user}$ is the predicted travel time derived from the congestion function using $x_{p,a}^{user}$, which is the edge flow value predicted by the model through solving the user-equilibrium TAP with the calculated O-D matrix. The APE values for all time bins on all edges are grouped together to provide the sample, resulting in a $3 \times 156 = 468$ sample size that is used to create the boxplots.

The inaccuracy of the user-equilibrium assignment prediction is also assessed through the error in TSTT, C_{error} , such that:

$$C_{error}^p(x^{UE}) = \frac{\sum_{a \in \mathcal{A}} x_{p,a}^{user} t_{p,a}^{user} - \sum_{a \in \mathcal{A}} x_{p,a}^{obs} t_{p,a}^{obs}}{\sum_{a \in \mathcal{A}} x_{p,a}^{obs} t_{p,a}^{obs}}, \quad (17)$$

for time bin p . This calculation combines the errors in flow and time prediction and has particular relevance for analysis based on aggregate total system cost such as POA.

It was found that the user-equilibrium flows produced were similar for BPR-Density compared to Inv-Opt and BPR-Standard (Figure 9 (a)). Overall the errors for the methods in all time-bins are not statistically different when tested with a one-way ANOVA. The similarity of flow prediction error between methods can be expected, as the aim of the BiLev is to adjust the O-D matrix to make the resulting user-equilibrium flows match as closely to the observed flows, which they are all effectively able to do.

There is a difference in the results when comparing average edge travel time for the user-equilibrium flow pattern (Figure 9 (b)). The results of one-way ANOVA tests between the three methods show a statistically significant difference in performance. There is an improvement from using the edge-specific density-based BPR fittings compared to the network-wide values of Inv-Opt and the standard BPR values.

The same trends for flow and travel time accuracy are consistent within the time bin comparisons (Table 2), where it is evident that BPR-Density has the best performance in

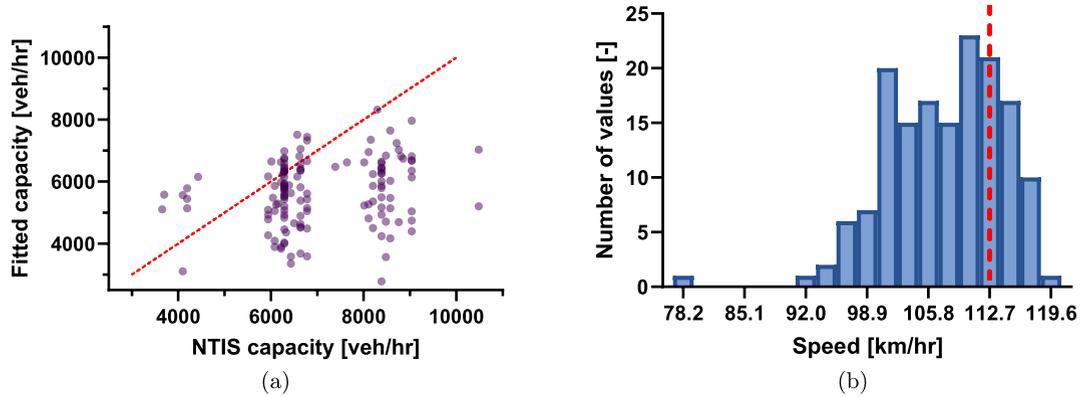


FIGURE 8. Fitted parameters for: (a) Comparison between fitted capacity against National Traffic Information Service (NTIS) capacity (points are semi-transparent, dashed line represent values where fitted capacity matches the capacity specified by the NTIS); (b) Histogram of free-flow speed (dashed line: posted speed limit). The data used to obtain the values are the observed hourly average traffic on the edges of the subnetwork for the weekdays selected for analysis between September 2018 and May 2019.

estimating the travel time. In Table 2, the mean error is the mean of the difference between measured and calculated values over all edges. It can be seen that the mean flow errors are all negative, as well as most of the mean time errors, indicating there is a systematic underestimation of TA results regardless of congestion function. BPR-Density has the lowest values of $C_{error}(\mathbf{x}^{ue})$ in all time-bins, indicating its superior total network cost prediction. This implies BPR-Density would be more accurate for use in cost-benefit analyses and POA calculations. The values of $C_{error}(\mathbf{x}^{ue})$ are negative in all cases indicating the models also systematically underestimate TSTT. The underestimation of the models is likely due to underestimates for the demand leading to less vehicles on the roads.

E. FUNCTION FITTING COMPUTATION TIME COMPARISON

BPR-Density compares particularly favourably to Inv-Opt when considering the computation time taken to estimate the congestion functions. Inv-Opt took 247420s for all three time-bins, BPR-Density took 22s in total for all the edges. The Inv-Opt results are for computing equations for all three time-bins but do not include the cross-validation to set the hyperparameters. Hence the difference in overall time is even greater.

The time complexity of Inv-Opt is influenced by the number of decision variables of the optimisation problem [42]. In the implemented version of the optimisation problem, the number of variables is of order $O(\mathcal{J}|(1 + |W||V|))$, where \mathcal{J} is the number of days of data used in the estimation. As the network is planar (edges cannot cross over one another), then $|A| \propto |V|$. Also, the number of O-D pairs is related to the number of nodes, $|W| = |V|(|V| - 1)$. Therefore, the complexity can be expressed in terms of the number of nodes, $O(\mathcal{J}|V|^3)$. The complexity of the optimisation problem and its computation time increases considerably with a larger network containing more nodes, and also more days of data used in the fitting. This contrasts with BPR-Density, which applies the low complexity non-linear least squares

regression to each edge of the network separately, so the computation time scales linearly with the number of edges $O(|A|)$, equivalent to $O(|V|)$. BPR-Density can be applied to the different edges independently in parallel so the computation time could be even less.

For Inv-Opt, the large increase in computation time as the network size increases makes the method impractical for large-scale national-level direct calculation, although the results suggest a simplified network may give a reasonable substitute in flow pattern accuracy.

The fitting of the congestion functions was performed on a Dell PowerEdge C6320 with 2.4GHz Intel Xeon E5-2630 v3 CPU and 24GB RAM.

V. DISCUSSION

From testing density-based fitting of congestion functions on a subnetwork of the England SRN it has been found that the best choice of function for use in that type of fitting is BPR. Despite BPR's perceived issues, in the case of unsignalised motorways, it fits the shape of the data better than other candidates. Alternative functions such as the exponential equation with a similar shape to BPR also perform well. The results also show the effect that density-based fitting has compared to flow-based and hypo-critical flow-based fitting. The density-based is able to fit the two parameters of BPR largely independently, whereas they are correlated when they are fitted with flow-based approaches. This is due to flow-based lacking information in the hyper-critical flow regime. The results can conclude that, at the network level, the density-based fitting of BPR is systematically superior for reliably obtaining the parameters.

In applying a density-based method, BPR-Density, for evaluating the congestion functions in a TA model, this work highlights its potential advantages over more conventional (i.e. BPR-Standard) and more computationally intensive machine learning-based methods (i.e. Inv-Opt). The results show that BPR-Density has a clear advantage over both BPR-Standard and Inv-Opt in the estimation of the travel time accuracy when applied in a TA model to predict the

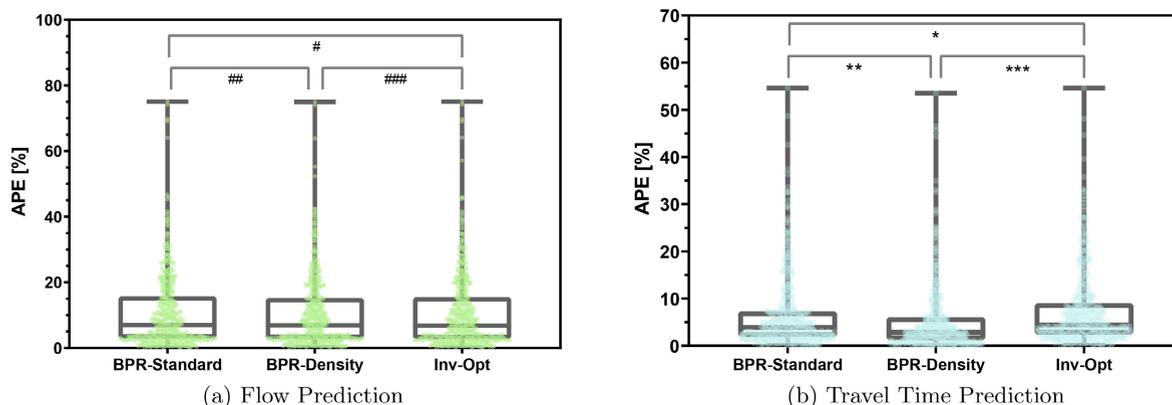


FIGURE 9. Absolute Percentage Error (APE) in prediction of (a) flow and (b) travel time for user-equilibrium assignment with alternative congestion function estimation methods. Results are for all time-bins and edges on the subnetwork, using data from the weekdays selected for analysis between September 2018 and May 2019. The points are the individual errors for each edge and time bin. The whiskers of the boxplots represent the range, the boxes are the IQR and median. The p-values of one-way ANOVA tests are, #: $p=0.998$, ##: $p=0.949$, ###: $p=0.966$, *: $p=0.237$, **: $p=0.035$, ***: $p=1.2e-4$.

TABLE 2. Time bin specific user-equilibrium prediction error statistics for all edges on the subnetwork during the analysis period September 2018 to May 2019. The mean errors refers to the mean across all the edges of the subnetwork.

		Mean Flow Error (veh/hr)	Mean Time Error (s)	Mean Flow APE (%)	Mean Time APE (%)	$C_{error}(x^{UE})$ (%)
AM	BPR-Standard	-197.8	-24.0	10.9	7.4	-7.8
	BPR-Density	-228.7	-10.2	10.7	5.2	-5.3
	Inv-Opt	-202.3	-27.4	11.0	8.0	-9.0
MD	BPR-Standard	-194.5	-12.5	11.0	4.3	-4.9
	BPR-Density	-191.9	+1.2	10.7	3.9	-1.4
	Inv-Opt	-199.2	-18.0	11.0	5.3	-6.5
PM	BPR-Standard	-186.5	-23.1	10.7	8.0	-7.7
	BPR-Density	-151.8	-8.9	10.4	6.9	-3.3
	Inv-Opt	-192.5	-28.0	10.5	8.8	-9.4

user-equilibrium assignment. This improved accuracy would have a positive impact on model applications, such as emissions models.

Compared to BPR-Density, Inv-Opt does have the advantage that it only requires flow data to fit the function, which could be useful if that is the only data type available. Also, while it was not possible to apply Inv-Opt to the full-size network, the functions fitted to the reduced network are of reasonable accuracy when applied to the more complex subnetwork. These advantages may be outweighed by the algorithmic complexity, translating into large differences in computation time. The results show a very large difference in time to compute the functions between Inv-Opt and BPR-Density, strongly suggesting the latter is more suited to the purpose of strategic transport planning on large SRNs where the functions may need to be updated regularly. It appears that the travel time for an individual edge is dependent on the physical features of the road, such as how it is connected to other roads. As these features often do not change throughout the day, BPR-Density’s superior performance suggests it may prove advantageous to fit a function based on road characteristics rather than the time of day.

The overall process of using loop detector traffic data to produce traffic assignments via data-derived congestion functions and O-D matrices could be used for other road networks with monitoring systems similar to this work, including many countries in Europe [43]. The techniques

could then be used to investigate and compare the POA on different road systems, as well as other high-level strategic analyses suited to static TA models.

A. FUTURE WORK

There are some shortcomings of the current study that could be addressed in future work.

By using alternative methods of obtaining the OD demand matrix for a large network, such as network modularity partitioning [23], or including alternative demand data sources (e.g. mobile phone [44]), the analysis could be applied to smaller windows of time than nine months, allowing the analysis of seasonal trends in traffic routing.

The limitations of the loop detectors of MIDAS reduce the accuracy of the topographic representation, limiting the fidelity of the traffic assignment results. Improvements to this could be made by augmenting the coverage of the loop detectors with other data, such as from road-side Bluetooth data [45]. Also, if such data sources can differentiate measurements for different vehicle types, this could allow BPR-Density to be applied to multi-class traffic (i.e. heavy trucks, motorcycles, etc.).

Possible improvements in the accuracy of density-based fitting may be sought by utilising better space-mean estimates of traffic density. Also, future work could look into the improvement of traffic assignment accuracy by using the density-based congestion function fittings in a fully

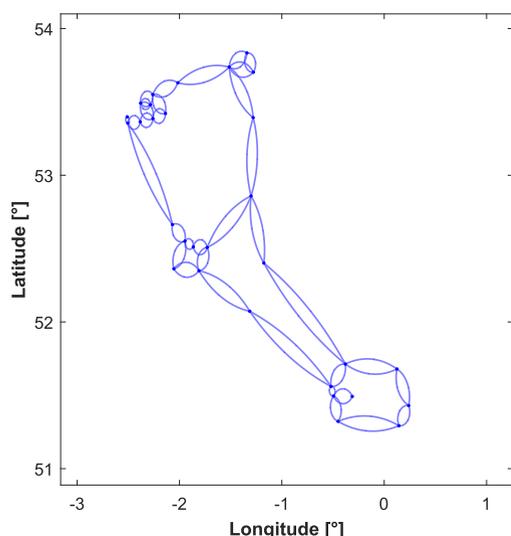


FIGURE 10. Simplified topographic representation of network for the subnetworks of the main roads connecting the central England Strategic Road Network. This has been used to fit the congestion functions using Inverse-Optimisation to ensure computational tractability.

density-based TAP formulation [46], rather than as a proxy for flow demand.

For the edges that do not have sufficient congested data, BPR-Density uses the standard coefficients and the NTIS capacity value. Further work could look into techniques to estimate such coefficients and capacities from those fitted to other edges on the network using other attributes (e.g. road grade).

VI. CONCLUSION

This paper develops the efficient computation of congestion functions specific to individual roads on a national highway network solely using loop detector data. The functions were fitted on a central subnetwork of the England SRN over a period of nine months and applied in a data-driven static TA model. The results indicate that this can be an effective approach in the solution of the TAP for use in strategic planning. There was an improvement in user-equilibrium travel time prediction accuracy for the road-specific BPR-Density fittings compared to the conventional BPR-Standard and the Inv-Opt method previously applied in such a type of TA model. Furthermore, the computational time of fitting the functions of BPR-Density was considerably lower than Inv-Opt.

VII. AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Alexander Roocroft, Muhamad Azfar Ramli, and Giuliano Punzo; data collection and coding: Alexander Roocroft; analysis and interpretation of results: Alexander Roocroft, Muhamad Azfar Ramli, and Giuliano Punzo; draft manuscript preparation: Alexander Roocroft, Muhamad Azfar Ramli, and Giuliano Punzo. All authors reviewed the results and approved the final version of the manuscript.

VIII. DATA AVAILABILITY

The data for the simplified topographic graph of the SRN and associated processed MIDAS traffic measurements are available at [47]. The raw data in this study have been provided by National Highways (England) via a data sharing agreement that does not allow further distribution. Requests for such data should be made to National Highways to whom the MIDAS and NTIS datasets used in this work belong.

APPENDIX

A. SIMPLIFIED SUBNETWORK FOR INVERSE-OPTIMISATION FUNCTION ESTIMATION

The method of Inv-Opt is applied to a simplified version of the central England SRN (Section III-F). This version (Fig 10) has many motorway junctions omitted as nodes. In total, it has 30 nodes, 70 edges and 870 O-D pairs.

ACKNOWLEDGMENT

The authors are thankful to National Highways and MWayComms for assistance in providing the raw data used in the article. For the purpose of open access, they applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

REFERENCES

- [1] M. C. J. Bliemer, M. P. H. Raadsen, L. J. N. Brederode, M. G. H. Bell, L. J. J. Wismans, and M. J. Smith, "Genetics of traffic assignment models for strategic transport planning," *Transp. Rev.*, vol. 37, no. 1, pp. 56–78, Jan. 2017.
- [2] Y. Pan, H. Zheng, J. Guo, and Y. Chen, "Modified volume-delay function based on traffic fundamental diagram: A practical calibration framework for estimating congested and uncongested conditions," *J. Transp. Eng., A, Syst.*, vol. 149, no. 11, 2023, Art. no. 04023112. [Online]. Available: <https://ascelibrary.org/doi/abs/10.1061/JTEPBS.TEENG-7903>
- [3] J. Qiao, D. de Jonge, D. Zhang, S. Simoff, C. Sierra, and B. Du, "Price of anarchy of traffic assignment with exponential cost functions," *Auton. Agents Multi-Agent Syst.*, vol. 37, no. 2, p. 42, Oct. 2023, doi: [10.1007/s10458-023-09625-6](https://doi.org/10.1007/s10458-023-09625-6).
- [4] H. Spiess, "Conical volume-delay functions," *Transp. Sci.*, vol. 24, no. 2, pp. 153–158, May 1990.
- [5] M. Patriksson, *The Traffic Assignment Problem: Models and Methods*, 2nd ed. Mineola, NY, USA: Dover, 2015.
- [6] *Traffic Modelling Guidelines. Version 3.0.*, Transp. London, London, U.K., 2010.
- [7] *HCM 2010: Highway Capacity Manual*, 5th ed., Transp. Res. Board, Nat. Res. Council, Washington, DC, USA, 2010.
- [8] L. V. Leong, "Effects of volume-delay function on time, speed and assigned volume in transportation planning process," *Int. J. Appl. Eng. Res.*, vol. 11, no. 13, pp. 8010–8018, 2016.
- [9] E. T. Mtoi and R. Moses, "Calibration and evaluation of link congestion functions: Applying intrinsic sensitivity of link speed as a practical consideration to heterogeneous facility types within urban network," *J. Transp. Technol.*, vol. 4, no. 2, pp. 141–149, 2014.
- [10] D. Nobel and S. Yagi, "Network assignment calibration of BPR function: A case study of metro manila, the Philippines," *J. East Asia Soc. Transp. Stud.*, vol. 12, pp. 598–615, Jan. 2017.
- [11] R. Akçelik, "Travel time functions for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function," *Austral. Road Res.*, vol. 21, pp. 49–59, Jan. 1991.
- [12] G. Casey, B. Zhao, K. Kumar, and K. Soga, "Context-specific volume-delay curves by combining crowd-sourced traffic data with automated traffic counters: A case study for London," *Data-Centric Eng.*, vol. 1, p. e18, Dec. 2020.
- [13] R. Kucharski and A. Drabicki, "Estimating macroscopic volume delay functions with the traffic density derived from measured speeds and flows," *J. Adv. Transp.*, vol. 2017, pp. 1–10, Jan. 2017.

- [14] G. Gentile, L. Meschini, and N. Papola, "Macroscopic arc performance models with capacity constraints for within-day dynamic traffic assignment," *Transp. Res. B, Methodol.*, vol. 39, no. 4, pp. 319–338, May 2005.
- [15] L. F. Huntsinger and N. M. Roupail, "Bottleneck and queuing analysis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2255, no. 1, pp. 117–124, Jan. 2011.
- [16] X. Wu and X. S. Zhou, "Characterization and calibration of volume-to-capacity ratio in volume-delay functions on freeways based on a queue analysis approach," in *Proc. Transp. Res. Board 100th Annu. Meeting*, Washington, DC, USA, Jan. 2021, p. 23.
- [17] R. Moses, E. Mtoi, H. McBean, and S. Ruegg, "Development of speed models for improving travel forecasting and highway performance evaluation," Florida Dept. Transp. (FDOT), Tallahassee, FL, USA, Final Rep. BDK83-977-14, 2013.
- [18] *Traffic Assignment Manual for Application With a Large, High Speed Computer*, U.S. Dept. Commerce, Urban Planning Division, U.S. Bur. Public Roads, Washington, DC, USA, 1964.
- [19] *TAG Unit M3.1 Highway Assignment Modelling*, U.K. Dept. Transp., London, U.K., 2014.
- [20] E. Qiu, N. Virdi, H. Grzybowska, and T. Waller, "Recalibration of the BPR function for the strategic modelling of connected and autonomous vehicles," *Transportmetrica B, Transp. Dyn.*, vol. 10, no. 1, pp. 779–800, Dec. 2022, doi: [10.1080/21680566.2022.2040063](https://doi.org/10.1080/21680566.2022.2040063).
- [21] J. Zhang, S. Pourazarm, C. G. Cassandras, and I. C. Paschalidis, "The price of anarchy in transportation networks: Data-driven evaluation and reduction strategies," *Proc. IEEE*, vol. 106, no. 4, pp. 538–553, Apr. 2018.
- [22] A. Roocroft, M. A. Ramli, and G. Punzo, "Efficient computation of optimal traffic assignment in nationwide highway networks from raw loop detector data," presented at the TRB 100th Annu. Meeting, Washington, DC, USA, Jan. 2021.
- [23] A. Roocroft, G. Punzo, and M. A. Ramli, "Flow count data-driven static traffic assignment models through network modularity partitioning," *Transportation*, Sep. 2023, doi: [10.1007/s11116-023-10416-x](https://doi.org/10.1007/s11116-023-10416-x).
- [24] L. D. Grange, C. Melo-Riquelme, C. Burgos, F. González, and S. Raveau, "Numerical bounds on the price of anarchy," *J. Adv. Transp.*, vol. 2017, pp. 1–9, Oct. 2017.
- [25] H. Youn, M. T. Gastner, and H. Jeong, "Price of anarchy in transportation networks: Efficiency and optimality control," *Phys. Rev. Lett.*, vol. 101, no. 12, pp. 1–4, Sep. 2008.
- [26] P. N. Patil, K. C. Ross, and S. D. Boyles, "Convergence behavior for traffic assignment characterization metrics," *Transportmetrica A, Transp. Sci.*, vol. 17, no. 4, pp. 1244–1271, Dec. 2021.
- [27] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*, 1st ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.
- [28] R. Dowling and A. Skabardonis, "Urban arterial speed-flow equations for travel demand models," in *Proc. Transp. Res. Board Conf.*, vol. 2, pp. 109–113, Jan. 2008.
- [29] V. Knoop, S. P. Hoogendoorn, and H. van Zuylen, "Empirical differences between time mean speed and space mean speed," in *Traffic Granular Flow '07*. Berlin, Germany: Springer, 2009, pp. 351–356.
- [30] B. Park, T. Kim, I. Yang, J. Heo, and B. Son, "A method for measuring accurate traffic density by aerial photography," *J. Adv. Transp.*, vol. 49, no. 4, pp. 568–580, Jun. 2015.
- [31] Y. Kim and F. L. Hall, "Relationships between occupancy and density reflecting average vehicle lengths," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1883, no. 1, pp. 85–93, Jan. 2004.
- [32] N. Gartner, C. Messer, and A. Rathi, *Traffic Flow Theory: A State-of-the-Art Report*. Washington, DC, USA: Federal Highway Administration, Jan. 2001.
- [33] J. Zhang, S. Pourazarm, C. Cassandras, and I. C. Paschalidis. (2017). *InverseVIsTraffic*. [Online]. Available: <https://github.com/jingzbu/InverseVIsTraffic>
- [34] G. Dervisoglu, G. Gomes, J. Kwon, A. Muralidharan, P. Varaiya, and R. Horowitz, "Automatic calibration of the fundamental diagram and empirical observations on capacity," in *Proc. Transp. Res. Board 88th Annu. Meeting*. Washington, DC, USA: TRB, Jan. 2009, pp. 1–14.
- [35] A. P. Silvano, H. N. Koutsopoulos, and H. Farah, "Free flow speed estimation: A probabilistic, latent approach. Impact of speed limit changes and road characteristics," *Transp. Res. A, Policy Pract.*, vol. 138, pp. 283–298, Aug. 2020.
- [36] M. Yildirimoglu and O. Kahraman, "Searching for empirical evidence on traffic equilibrium," *PLoS ONE*, vol. 13, no. 5, pp. 1–16, May 2018, doi: [10.1371/journal.pone.0196997](https://doi.org/10.1371/journal.pone.0196997).
- [37] *National Traffic Information Service DATEX II Service v10*, Highways England, National Highways, London, U.K., 2019.
- [38] N. Harris, "Strategic road network initial report," Nat. Highways, London, U.K., Tech. Rep., 2023. [Online]. Available: https://nationalhighways.co.uk/media/mjzibdr/cre22_0102-srm-initial-report-2025-2030_vn.pdf
- [39] Google Maps. (2021). *Map of Central England*. [Online]. Available: <https://www.google.com/maps/place/England,+U.K./>
- [40] V. L. Knoop and W. Daamen, "Automatic fitting procedure for the fundamental diagram," *Transportmetrica B, Transp. Dyn.*, vol. 5, no. 2, pp. 129–144, Apr. 2017.
- [41] M. B. Mahdi, A. K. Alrawi, and L. V. Leong, "Compatibility between delay functions and highway capacity manual on Iraqi highways," *Open Eng.*, vol. 12, no. 1, pp. 359–372, May 2022.
- [42] D. Bertsimas, V. Gupta, and I. C. Paschalidis, "Data-driven estimation in equilibrium using inverse optimization," *Math. Program.*, vol. 153, no. 2, pp. 595–633, Nov. 2015.
- [43] GraphHopper. (2021). *GraphHopper Open Traffic Collection*. [Online]. Available: <https://github.com/graphhopper/open-traffic-collection>
- [44] A. Dypvik Landmark, P. Arnesen, C.-J. Södersten, and O. A. Hjelkrem, "Mobile phone data in transportation research: Methods for benchmarking against other data sources," *Transportation*, vol. 48, no. 5, pp. 2883–2905, Oct. 2021.
- [45] A. R. Kulkarni, N. Kumar, and K. R. Rao, "Efficacy of bluetooth-based data collection for road traffic analysis and visualization using big data analytics," *Big Data Mining Anal.*, vol. 6, no. 2, pp. 139–153, Jun. 2023.
- [46] P. Kachroo and S. Sastry, "Traffic assignment using a density-based travel-time function for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1438–1447, May 2016.
- [47] A. Roocroft, M. Ramli, and G. Punzo. (2023). *MIDAS-on-SRN*. [Online]. Available: <https://github.com/AlexR2184/MIDAS-on-SRN>



ALEXANDER ROOCROFT received the M.Eng. degree in civil engineering from Imperial College London, in 2014, and the Ph.D. degree in transport engineering from The University of Sheffield, in 2023. During the Ph.D. degree, he undertook a research attachment with A*STAR, Singapore, from 2019 to 2022. His research interests include data-driven traffic models, road network analysis, and social routing schemes. In addition to his research activities, he was an Engineering Consultant in London, U.K., and Dubai, United Arab Emirates.



MUHAMAD AZFAR RAMLI received the B.Eng. and Ph.D. degrees in mechanical engineering from the National University of Singapore, in 2007 and 2012, respectively. He is currently a Senior Scientist and the Deputy Department Director of the Institute of High Performance Computing, Agency of Science, Technology and Research. His research interests include urban systems modeling and simulation and complex network analysis.



GIULIANO PUNZO graduated in aerospace engineering. He received the Ph.D. degree in swarm engineering. He worked in industry. After postdoctoral appointments with the University of Glasgow, the University of Strathclyde, and The University of Sheffield, in 2019, he joined the Department of Automatic Controls and Systems Engineering, The University of Sheffield, as a Lecturer. His research interests include dynamic networks, control theory, consensus, and complexity.