An analysis of overtaking and flow from ANPR data on single-carriageway two-lane roads

Frank Brooks-Tyreman and R. Eddie Wilson,
Department of Engineering Mathematics, University of Bristol

1 Introduction

On single-carriageway, two-lane (known in the UK as S1-type) roads, a significant lack of overtaking opportunities results in faster vehicles becoming trapped behind their slower counterparts. In this paper we will outline the techniques used to extract overtaking rates and flow from empirical data between Blair Atholl and Carrbridge, a 100km stretch of the A9 road, Scotland. In this paper we analyse ANPR data provided by Transport Scotland (TS) with the aims: i) to outline the techniques used to obtain estimates for both flow and overtaking rates; and ii) to suggest how we would extract an empirical value of the overtaking parameter α (units s^{-1}) used in our simulation model.

To measure road performance, and to consider whether a particular road is a candidate for dualling, transport modellers use three main metrics: i) average speed; ii) percentage time spent following (PTSF); and iii) overtaking rates. Average speed is the simplest metric to obtain; with site-to-site distances and site arrival times we are able to produce straight line trajectories with average speed equivalent to the trajectory gradient. PTSF is more difficult to extract from data (and is beyond the scope of this paper). Within the literature there is also some contention over the use of PTSF as a measure of road performance due to its near impossibility to measure empirically. Difficulty arises in defining when a vehicle is *following* — the Highway Capacity Manual (HCM 2010, 2010) uses a proxy measure of a critical time headway (3s) to determine whether a vehicle is following which has been disputed as over-estimated by numerous publications (see (Laval, 2006; Van As and Van Niekerk, 2004) and (Luttinen, 2001) for a review).

Overtaking rates are also difficult to obtain empirically with various procedures for obtaining accurate measures still being researched i.e., Bluetooth and anonymised automatic number plate recognition (ANPR) data. We received both Bluetooth and ANPR data from TS of vehicle orderings at multiple site locations on the A9. ANPR data has a higher recognition rate ($95\%+$ recognition (Patel et al., 2013)) for registering vehicles as they pass a site than Bluetooth — we therefore will only be analysing ANPR data in this paper. There are some obvious issues with using these fixed locations: i) most significantly, there are no ground truth trajectories — we have no knowledge of vehicle behaviour between sites i.e., there are opportunities for vehicles to leave the A9 and choose another route; and ii) there are sections of dual carriageway along the A9 which would drastically change vehicle orderings. Through this paper we will outline assumptions used to minimise the impact of these issues.

We have also built a parsimonious microscopic simulation model that describes driver behaviour on S1 roads. The model consists of two parts: a (kind of) first order longitudinal model which describes platoon dynamics; and an overtaking model which describes the manoeuvre, the decision making process and the propensity to overtake. The modern approach to car-following models is second-order (acceleration as the second derivative of displacement) and a function of external stimuli such as relative displacement and speed (see the IDM (Treiber and Kesting, 2013) and (Brackstone and McDonald, 1999) for a review). This approach has been incorporated in leading commercial microsimulation packages (AIMSUN, 2019; VISSIM, 2019; TSIS-CORSIM, 2019). Unlike these models, where congestion is the cause of flow-breakdown, we model flow-breakdown caused by platoon formation (allowing us to neglect finer details such as vehicles' acceleration). A limitation in our model is our overtaking model which neglects the *other side of the road* — we aim to use empirical data to extend our model to include both sides of the road. We believe this simulation model, despite its simplicity, can be used to help inform policy makers on whether a road is a candidate for dualling.

The paper is structured as follows: in section 2 we will describe the data set used and assumptions employed to clean the data; in section 3 we describe the techniques used to analyse the data to obtain flow and overtaking rates; in section 4 we present results; and in section 5 we discuss our modelling, results and outline areas of future work.

2 ANPR data

In this section we describe the data set provided by TS. A 100.67km stretch of road between Blair Atholl and Carrbridge along the A9 is considered with ANPR data collected at 5 locations (10 sites in total including both northbound and southbound directions) — see Table 1 for location information, and Fig. 1 for a map view.

Table 1: ANPR site locations, cumulative distance between sites and the corresponding segment lengths.

Site	Site Displacement (km)	Segment Length (km)
Blair Atholl (A9 East of B8079)	0	—
Dalwhinnie (A9 South of A889)	43.90	43.90
Kingussie (A9 West of Lynchat)	69.27	25.37
Aviemore (A9 at Loch Alvie)	82.84	13.57
Carrbridge (A9 West of A938)	100.67	17.83

The data was collected on the 18$^{\text{th}}$ March 2014 between 07.00 and 19.00. Of the 2014 layout of the A9, approximately 22km of the road was dualled (with the majority in the section between Blair Atholl and Dalwhinnie). Since then £3bn has been invested by the Scottish government to dual 80 miles of the A9 between Perth and Inverness by 2025 (Transport Scotland, 2018). As well as dual carriageway, roughly 8km of the road also disallows overtaking; thus only 67% of the road between the Blair Atholl and Carrbridge sites is strictly S1-type. Dualling is a significant factor in our model as it affords drivers more opportunities to overtake (safely) thus improving throughput of vehicles over the segment — we should therefore expect there to be some impact and skew in our outputs: overtaking rate and flow. Whilst one may expect the no-overtaking sections to impact on overtaking results; these sections form mainly in the lead up / run off from the dual carriageway sections or when there is a significant right-turning to a B road such as the Blair Atholl turnoff (B8079) where for safety reasons there would be no overtaking anyway.

Another factor to bear in mind is that the data set was collected before the A9 road was fitted with average speed cameras (October 2014). In January 2015 a report published by TS (Transport Scotland, 2015) indicated that the average speed cameras had reduced overtaking rates; resulting from private vehicle speed being greatly reduced and the increase in HGVs (heavy goods vehicle) speed from 40mph to 50mph.

The data set is moderately detailed (see a sample at a particular site in the southbound direction in Table 2) and includes:

- Site indexing from which we are able to locate vehicles appearing at consecutive sites.
- Vehicle indexing: in the form of an alphanumeric string (this is the anonymised ANPR data).
- Time stamps: the precise time a vehicle passes the ANPR infrastructure — using times we are able to build straight line trajectories. For a vehicle i, over a segment of length Δx appearing at consecutive sites m and $m+1$ at times t_m and t_{m+1} respectively, average speed $\bar{v}_i = \Delta x/(t_{m+1} - t_m)$.
- Vehicle classification: vehicles can be either: LVs — light vehicles; HVs — heavy (goods) vehicles; and PSVs — public service vehicles. We do not directly use this data in this paper however, our simulator model considers heterogeneous classes and using this data we can simulate a road with the correct vehicle class proportions.

We are also able to find averages (both spatial and temporal) for the fundamental variables (speed, flow and density) and build a basic picture of vehicle behaviour on the A9 between each site in both northbound and southbound directions.

Figure 1: Map of the section of the A9 between Blair Atholl and Carrbridge where the ANPR data has been collected. ANPR cameras at sites 1-5 took vehicle measurements in both the northbound and southbound directions.

There are limitations in this data set: we have information at only 5 particular sites along the A9; and the road is also non-homogeneous. It is therefore difficult to infer true vehicle behaviour between sites and there are multiple pitfalls including (but not limited to): i) A small chance that the ANPR camera has incorrectly read / missed a plate; ii) vehicles join / leave the road between sites and use alternate routes; iii) vehicles may re-overtake. We must therefore be careful in both our modelling practise, and in the accuracy that we attribute to the results.

To clean the data, i.e., remove spurious results such as unlikely journey times we delete vehicles whose average speed falls below a critical speed (we assume these vehicles do not make a direct trip between sites); the critical speed used in our analysis is 13m/s ($\approx$ 30mph, which is very low for a road with a 60mph speed limit — it is worth noting though, in March 2014, the speed limit for HGVs on the A9 was 40mph). We are unable to verify accurately whether these vehicles have overtaken / been overtaken so it seems prudent to remove them and reduce their impact.

We also make the clear choice that we only consider vehicles which appear at both sites when computing our overtaking rates (aware that we will be actively underestimating). The proxy measure for overtaking is using rank sorting analysis on two ordered sets — we are unable to order accurately, vehicles which do not appear at both sites (this will be explained further in the modelling section).

For flow data we also build artificial trajectories for vehicles which appear at only a single site (using the average speed of a road segment). We make this decision with the understanding that we will be slightly overestimating flow (we do not know precisely when or where single site trajectories enter / exit the road segment) but we justify this as considering only vehicles which appear at both sites significantly underestimates flows.

Table 2: Subsets of ANPR data in the form received from Transport Scotland. Both heavy vehicles (HVs) at site 2 reappear at site 4 in the same order. The light vehicle (LV) VTNSC79 appears at site 4 much earlier indicating overtaking has occurred.

Site	Plate	Class	Time
2	WXPXCS7	HV	07:00:04
	STNX53C	LV	07:00:06
	WXPXCS6	HV	07:00:09
	VTNSC79	LV	07:00:12
	6VPYQV4	LV	07:00:13
Site	Plate	Class	Time
3	VTNSC79	LV	07:27:13
	⋮	⋮	⋮
	WSPYYTU	HV	07:28:26
	WXPXCS7	HV	07:28:46
	WXPXCS6	HV	07:28:59
	W5PSWI7	LV	07:29:01

3 Modelling

In this section we will describe the modelling techniques used to obtain estimates for both time-averaged flow and overtaking rates illustrated in Fig. 2. Consider an area $A = [\Delta x \times \Delta t]$ where Δx is the length of the road segment between consecutive sites and Δt is the length of each time block. We consider $m = 1, \ldots, M$ road segments, with $M = 4$ (the number of segments between ANPR sites) — in future we may choose to consider segments of fixed lengths, but for the purpose of this paper the lengths chosen correspond to where vehicles are registered, important for modelling overtaking. Time block length is uniform with $l = 1, \ldots, L$ time blocks considered. The length of each block is bounded below by the length of time an average vehicle of constant speed $\bar{v}$ should take to traverse the road segment.

As previously mentioned, there are two types of trajectory to consider: i) full vehicle trajectories, for vehicles that appear at adjacent sites; and ii) artificially created trajectories obtained from vehicles that appear at only one site (by using the average speed for the road segment we are able to create a full trajectory). We take both trajectory types into consideration when counting flow, but only full trajectory data sets for the overtaking rate. We choose to use only full trajectories for overtaking rates as we know explicitly the orderings of arrival at both sites 1 and 2; by using artificial trajectories we are actively including fictitious overtaking counts. Whilst we understand that we are underestimating the number of overtakes, we do not have enough information about incomplete trajectories to accurately use them in our overtaking counts. We do however use both trajectory types for flow rates; we argue that it is better to overestimate the flows in each region A as we would be missing too large quantities of the general traffic make-up.

In order to find the time-averaged flows and speeds we take into account that not all trajectories are solely contained in the region A, with some trajectories overlapping with adjacent regions thus crossing the end points of A at $\{\tau_l, \tau_{l+1}\}$ — white markers in the figure identify the start / end points of vehicle trajectories in these situations. These new start / end points for each region provide us with vehicle orderings of vehicles entering ($\mathcal{D}_A^m$) and exiting ($\mathcal{D}_A^{m+1}$) the area A.

3.1 Edie's generalised definition of flow

Edie's generalised definitions of traffic variables (Edie, 1963) are useful for maintaining consistent results in both empirical data and modelling. They also remove the requirements of correctly averaging when using alternative methods such as in- / out-flows. To calculate flow q (veh/hr) consider $i = 1, \ldots, N$ vehicles on the road with some $n \leq N$ contained in a region $A = [\Delta x \times \Delta t]$, then each vehicle i is weighted by the distance d_i it travels inside A. Flow is given by

$$q_A = \sum_{i=1}^{n} \frac{d_i}{\Delta x \Delta t}. \tag{1}$$

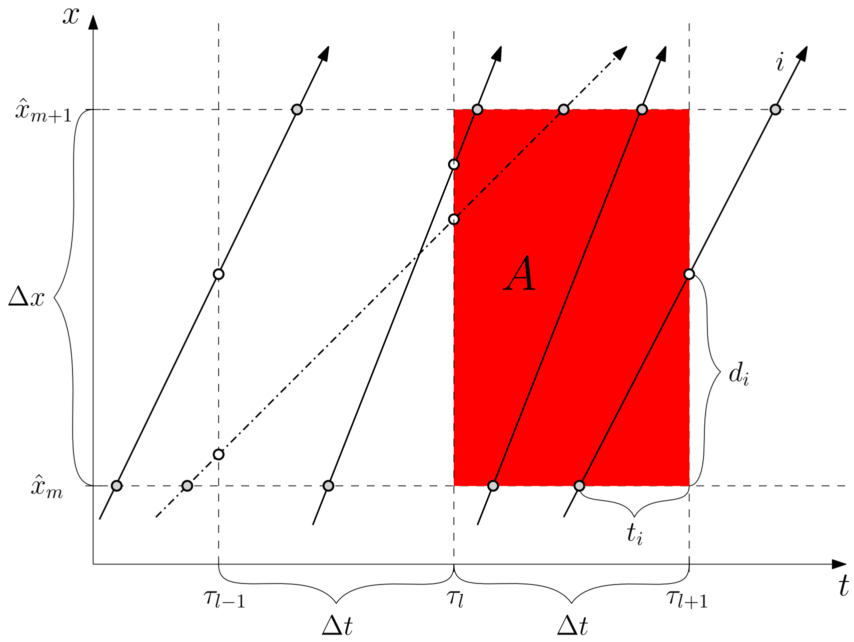

Figure 2: A trajectory plot illustrating differing vehicle trajectories used to obtain flow data and overtaking rates. Grey markers indicate vehicle trajectory start and end points. White markers are the assumed positions at the end points of each time block. Edie flows weighted by trajectory length d_i inside areas A.

Essentially each trajectory is weighted according to how far each vehicle travels in A. As we use full trajectories (either natural or artificial) we are afforded a simple check as to whether our flow counts through each area A are correct as $q_{\text{tot}} = \sum_i d_i / \Delta x \Delta t = N/\Delta t$.

3.2 Overtaking

There is a lack of ground truth when it comes to identifying the number of overtakes which occur between consecutive sites — there is not enough information to correctly infer where, when or how vehicles have made overtakes. We make two critical assumptions: i) due to overtaking being difficult on S1 roads, if a vehicles overtakes another, it is never re-overtaken by that same vehicle; ii) the order that vehicles arrive at site $m+1$ corresponds directly to the number of overtakes they make.

We are able to identify the number of overtakes using a sorting algorithm, with the number of swaps required to satisfy $\mathcal{D}_A^{m+1} = \mathcal{D}_A^m$ used a measure for the time-averaged overtaking rate (see Table 3). In our modelling process we use the *bubble sort* algorithm — bubble sort is chosen for two reasons: i) it is simple; and ii) the assumption of no re-overtakes is mirrored by the bubble sort algorithm which compares two adjacent pairs and swaps them if they are in the wrong order. We will now refer to sites m and $m+1$ as sites 1 and 2 respectively, without loss of generality.

At site 1, the vehicle orderings are given by the set of ordered pairs $\mathcal{D}_A^1 = \{(x,t) : i \in A,\ x \leq \hat{x}_2,\ t < \tau_{l+1}\}$ and at site 2, $\mathcal{D}_A^2 = \{(x,t) : i \in A,\ x \leq \hat{x}_2,\ \tau_l < t \leq \tau_{l+1}\}$. For each i we assume a straight-line trajectory given average speed v_i with which we can approximate the locations of each vehicle at $t = \{\tau_l, \tau_{l+1}\}$ — disregarding trajectories ending at site 2 before τ_{l+1}. Sets $\mathcal{D}_A^1$ and $\mathcal{D}_A^2$ provide an ordering of vehicles in A.

To obtain the sets $\mathcal{D}_A^1$ and $\mathcal{D}_A^2$ consider Fig. 3, where vehicles arriving at either τ_l (white markers), or $\hat{x}_m$ (grey) correspond to the order $\mathcal{D}_A^1$. To obtain $\mathcal{D}_A^2$, consider the grey markers at x_{m+1}; these are vehicles which complete their trajectories inside A, and thus they are ordered *temporally*; for those vehicles who have not left the road segment inside A, at time τ_{l+1} they are ordered *spatially*. We now have the indices according to entry at site 1 ordered according to space and time at site 2 given by the set $\mathcal{D}_A^2$. We can now sort $\mathcal{D}_A^2$ using the bubble sort algorithm and obtain the number of overtakes expected in the area A. For an example, see Table 3.

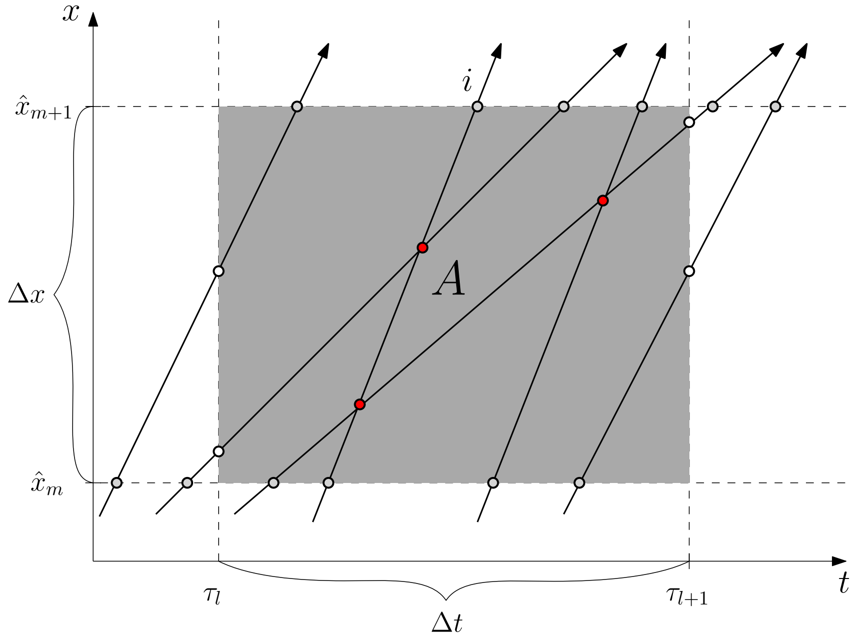

Figure 3: A trajectory plot illustrating how vehicle trajectories are used to obtain overtaking rates. Site 1 vehicles are ordered temporally (grey markers) when arriving at $t \in [\tau_l, \tau_{l+1})$ and then spatially (white markers) for vehicles at spatially for $t = \tau_l$. Site 2 vehicles are ordered temporally for $t \in (\tau_l, \tau_{l+1}]$ and spatially for $t = \tau_{l+1}$. A bubble sort algorithm is used and counts swaps (represented by the red markers for crossing trajectories).

Table 3: Fig. 3 example. Proxy measure for the number of overtakes using a bubble sort algorithm

Site 1 order	Site order 2	Overtakings
1	1	0
2	4	0
3	2	0
4	5	2
5	3	1
6	6	0
		total = 3

4 Results

In this section we present some preliminary results. The total number of overtakings in each time block are compared to the Edie flows to investigate whether there is any correlation.

In Fig. 4a we compare the northbound overtakings and directional flow. Whilst it is difficult to view a clear correlation between the overtaking rates and flow, there are some trends to notice: i) there is a delay between a spike in the number of overtakings and an increase in the flow — this suggests that when there are more overtakings, the flow in the corresponding direction experiences a temporary boost; ii) flow in the northbound direction in the morning is generally higher than in the evening, we see that the number of overtakes decreases in a corresponding manner.

In Fig. 4b we compare northbound overtakings with opposing directional flow. In contrast with northbound flow, southbound flow increases throughout the day. This result agrees with intuition that as flow in the opposing direction increases, there are fewer opportunities to overtake. Earlier in the day, the total number of overtakings is high corresponding to low flows in the opposing direction.

We can find the individual overtaking rate, overtakes per vehicle per second, quite easily. Using q_A from Eq. (1) and the average speed of a time block $\bar{v}_A$, we can obtain the density for each time block, $\rho_A = q_A/\bar{v}_A$ (from the fundamental equation of traffic flow). From the total number of overtakes we can

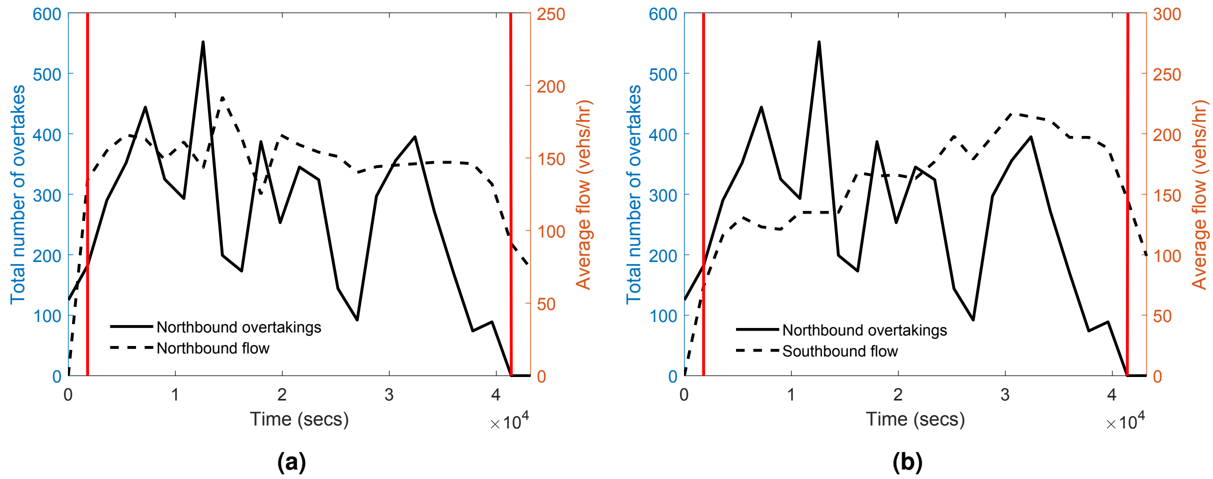

Figure 4: Plots for the total number of overtakings vs average flow over the total time set. We consider the road segment between sites 1 (Blair Atholl) and 2 (Dalwhinnie) for time blocks of length $\Delta t = 30$ mins. For reference, the red bars correspond to the first and last time blocks in the data. We generally disregard these due to skew in the data caused by the collection bounds. There should be a drop off in the number of overtakings caused by: in the morning, a lack of vehicles preceding those in the system, thus we do not have the data for these vehicles and whether or not they were overtaken; in the evening, no more vehicles are entering the system, and any of these who perform overtakings will be lost.

divide through by Δx and Δt to obtain the total number of overtakes per meter per second. Dividing through by density ρ_A we obtain the overtaking rate for an individual. From Fig. 4a we find that the average individual overtaking rate between $t = [1800, 3600]$s is 0.002(veh s)$^{-1}$ which corresponds to an individual vehicle making an overtake every 500 seconds (or just over 8 minutes); alternatively this can be interpreted as an overtake every 7-8 miles. Whilst these numbers are much smaller than the average number of lane changes made on a multi-lane highway (Knoop et al., 2012) they are reasonable due to the limited safe opportunities to overtake on an S1 road.

5 Conclusion and Discussions

We have analysed ANPR data between Blair Atholl and Carrbridge on the A9 road in Scotland to obtain overtaking rates and flows. From Edie's (Edie, 1963) classic definitions of generalised fundamental traffic variables we were able to obtain flow rates in a consistent way for road segments of length Δx m and time blocks of length Δt s.

Using a bubblesort algorithm as a proxy measure we were able to obtain approximations for the number of overtakes (total (per km per s), and at an individual level (per s)) for the entire system and for each time block of length Δt.

So far we have been unable to find an explicit, strong correlation between overtaking rates and both directional and opposing flows. We would have expected there to be a positive correlation between overtakings and flow in the direction of travel (as density increases, so do the number of overtakes (Wardrop, 1952)). We also expected a negative correlation between overtakings and opposing flow (more vehicles on the other side of the road would reduce the opportunities to overtake) — dependent on the local density of vehicles on the other side of the road. The more spread out vehicles are on the other side of the road, the less opportunities to overtake; with more vehicles trapped in platoons vehicles will experience longer sections of the road where overtaking is possible.

The next step is to be able to explain overtaking rate as a function of both directional and opposing flows.It may also be that if we can calculate average platoon size (or PTSF) from the data set then overtaking may be correlated with this. Finding this will allow us to improve our simulator model.

Currently, the simulator model we have built has a significant limitation; it neglects the effects of the other side of the road, namely, the impact that opposing flow has on overtaking opportunities. Using empirical data for S1 roads we aim to extend our model to include these factors. The overtaking rate function in our simulator model is given by

$$\alpha\left(1-\frac{v_i}{v_i^{\max}}\right),$$

where v_i is a vehicles current speed, $v_i^{\max}$ is a vehicles desired speed (speed when travelling unimpeded) and α (units s^{-1}) normalises the overtaking rate to a corresponding rate of overtaking. Our aim is to empirically approximate the value of α thus extending our model to be explicitly inclusive of opposing flow.

We also foresee an opportunity to use our simulator to provide a ground truth for vehicle behaviour between ANPR sites. We are limited in our data analysis currently due to only having straight line trajectories. With an α value obtained empirically we would be able to simulate driving behaviour between sites and be able to compare simulation results with real world data with the aim to obtain a ground truth for the number of overtakings.

References

AIMSUN (2019). `https://www.aimsun.com`.

Brackstone, M. and McDonald, M. (1999). Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4):181–196.

Edie, L. C. (1963). *Discussion of traffic stream measurements and definitions.* Port of New York Authority.

HCM 2010 (2010). Highway Capacity Manual. *Washington, DC: Transportation Research Board*.

Knoop, V., Hoogendoorn, S., Shiomi, Y., and Buisson, C. (2012). Quantifying the number of lane changes in traffic: Empirical analysis. *Transportation Research Record: Journal of the Transportation Research Board*, (2278):31–41.

Laval, J. A. (2006). A macroscopic theory of two-lane rural roads. *Transportation Research Part B: Methodological*, 40(10):937–944.

Luttinen, T. (2001). *Traffic flow on two-lane highways: an overview*. TL Consulting Engineers.

Patel, C., Shah, D., and Patel, A. (2013). Automatic number plate recognition system (ANPR): A survey. *International Journal of Computer Applications*, 69(9).

Transport Scotland (2015). Performance data for A9 and the average speed cameras published. https://www.transport.gov.scot/news/performance-data-for-a9-and-the-average-speed-cameras-published/.

Transport Scotland (2018). A9 dualling Perth to Inverness. `https://www.transport.gov.scot/projects/a9-dualling-perth-to-inverness/`.

Treiber, M. and Kesting, A. (2013). Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*.

TSIS-CORSIM (2019). `https://mctrans.ce.ufl.edu/mct/`.

Van As, S. and Van Niekerk, A. (2004). The operational analysis of two-lane rural highways. *SATC 2004*.

VISSIM (2019). `http://vision-traffic.ptvgroup.com`.

Wardrop, J. G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362.

Acknowledgements

Thanks to Stephen Cragg and Transport Scotland for providing the ANPR data for the A9 road Scotland used in this paper.