

Number of Lane Changes Determined by Splashover Effects in Loop Detector Counts

Victor L. Knoop, R. Eddie Wilson, Christine Buisson and Bart van Arem, *Member, IEEE*

Abstract—Lane changes are important in quantifying traffic, both for operational and planning purposes. Traditional in-lane loop detectors do not count lane changes, hence historically, traffic engineers have estimated them using other data sources. This paper provides a method to estimate the number of lane changes based on observations of “straddling” vehicles that are detected simultaneously by the loops in adjacent lanes. In the data considered here, such “straddles” correspond almost always to vehicles that are in the process of changing lane. However, many lane-changes take place between detector sites and hence do not result in straddles. The methods developed here estimate the probability distribution for the number of lane changes given an observed number of straddles. The efficacy of this approach depends on calibration issues and on the size of the aggregation period. In the evaluation study presented here the results are good: the proposed method gives the number of lane changes with approximately 10% error, even though the number of lane changes per aggregation period varies by a factor of 10 over time.

I. INTRODUCTION

Lane changes play a crucial role in freeway traffic management and operations. They are relevant to four areas of traffic engineering: operations, control, modelling and planning, and safety. Firstly, at the operational level, they are claimed to influence capacity (e.g., [1]), or even induce stop-and-go waves (e.g., [2]). Furthermore, they cause synchronisation of speed between lanes: for instance, stop-and-go waves may develop in one lane, but after a short while, the speed reduction spreads across the carriageway. Secondly, in traffic control, a lane change ban can improve traffic conditions, e.g., [3]. Systems to influence the use of lanes may use road-based or in-vehicle devices [4]. Figure 1 shows the architecture of the control cycle, and where measuring lane changes comes into play. Thirdly, in modelling, measurements of lane changes are needed for lane changing models. Up to now, as a consequence of the lack of measurement methods, lane changing models are often calibrated by other (weaker) types of data, for example by the distribution of traffic counts over lanes [5]. For planning purposes, lane-changing rates can be used to estimate local OD matrices and weaving patterns through a sequence of successive merges and diverges. Finally, lane changing has a consequence for safety [6].

For all these areas, it is necessary to measure the frequency and location of lane changes on a given stretch of motorway.

Victor L. Knoop (v.l.knoop@tudelft.nl) and Bart van Arem are with the department of Transport and Planning, Delft University of Technology, Stevinweg 1, 2628 CN Delft, the Netherlands

Christine Buisson is with IFFSTAR and R. Eddie Wilson is with Southampton University

Manuscript received March 5, 2012

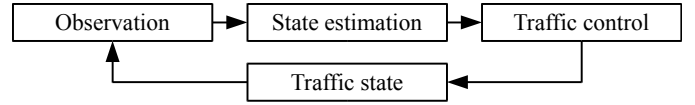


Fig. 1. The traffic control cycle

Once the occurrence of lane changes and their influence is known, control measures can be designed to (statically or dynamically) mitigate any undesired effects.

Loop detectors are the most often used device to collect traffic information. However, despite their importance, lane changes are not observable in standard stationary detector data, due to the local nature of the measurement devices. This article presents a novel methodology to extract the number of lane changes from loop detectors.

The basis of the method is to study what fraction of vehicles is detected simultaneously by the loops in *two adjacent lanes* — an effect known as “splashover” of the signal. A vehicle which thus activates two detectors will be called “straddling”. The chief idea is that straddles usually correspond to vehicles that are in the process of changing lane when they drive over the detector, and so they may be used to estimate the frequency of lane changes.

Unfortunately, only a small fraction of the lane changes that occur within a given (extended) road section will result in straddles at any given detector, because most lane changes will take place between detector sites. Thus straddles may be modeled as a stochastic process — that is, a given lane change results in a straddle with probability p that must be calibrated for the given road section and detector set-up. Following calibration, the estimation of the number of lane changes from a given number of observed straddles thus involves uncertainty and our method develops rigorous error bars for this process in the form of posterior probability distributions.

The paper is organized as follows. Firstly, section II gives a brief overview of local data sources and the state-of-the-art in their use in the estimation of lane changes. Then, section III gives a detailed account of the data that is used in this study. Briefly: we have access to an inductive loop data set which detects straddling vehicles and which has captured and stored full individual vehicle data. By means which we shall describe, full vehicle trajectories (and hence lane changes) are also observable and so the correlation between straddles and lane changes can be derived in detail. This is the purpose of section IV, which also shows calibrates the procedure.

Section V then derives an operational tool which is used to estimate the number of lane-changes in hold-out data that was

not used in the calibration step. Depending on the length of the aggregation interval, excellent performance may be achieved. Section VI then discusses issues for further development, for example, the dependence of p on the speed of the traffic and the distance between detector sites. Finally, conclusions are presented in section VII. In particular, we discuss the potential for applications to less ideal detector environments than that considered here.

II. LITERATURE REVIEW

Inductive loops are the most commonly used way to measure freeway traffic. At each detector site, one usually has an ‘in-lane’ arrangement with one (single or double) loop aligned in each lane of the road. In a good layout, the loops are sufficiently wide relative to the width of their respective lanes that very few vehicles can drive between the loops and avoid detection. However, some vehicles will trigger the loops in two adjacent lanes, and such a double count is usually considered erroneous [7]. Coifman [8] analyses such error rates and how these vary among different detectors. Magnetometers are rather similar in operation to loops, except that many more vehicles pass undetected [9].

To observe lane-changing directly, one needs individual vehicle trajectory data, and in the past 5-10 years this has been developed using video footage, for example in the NGSIM project [10], or in the Netherlands using a helicopter [11]. However, the analysis of video footage is not completely automatic and the extraction of trajectories from it is time consuming [12]–[14]. Hence even if video footage were available on-line, it could not presently be used in operational methods for counting lane changes.

It seems that the theory of lane changes may have begun in 1952, hypothesizing a dependency of their rate on macroscopic density [15]. There are also microscopic theories of lane changing, involving local criteria (e.g., gap acceptance) for a driver to change lane e.g., [16]. More recent work, for instance by Laval and Daganzo [17] shows the effects of lane changing: the voids caused by a lane change can lead to a capacity drop. McDonald and Brackstone [18] were among the first to develop systematic empirical observations of lane-changes. This was achieved by manual analysis of video images from a three-lane freeway. They found that “it is extremely difficult to validate the (lane change) model over a wide range of flows and conditions”. Later, a very large empirical data collection was carried out showing the fine details of individual lane changes, including their reasons and their duration [19]. Many studies (e.g., [20]) have studied the impact of lane changes at the level of platoons.

Most recently, the challenge left by MacDonald and Brackstone [18] was taken up [21] and a relationship between the number of lane changes and traffic density was established, using the same data source as the present article. All theories need to be tested with data and the present article develops a novel methodology for its collection.

III. DESCRIPTION OF THE DATA USED IN THIS PAPER

Our aim in this article is to show how the frequency of lane changing on highways can be estimated using inductive loop

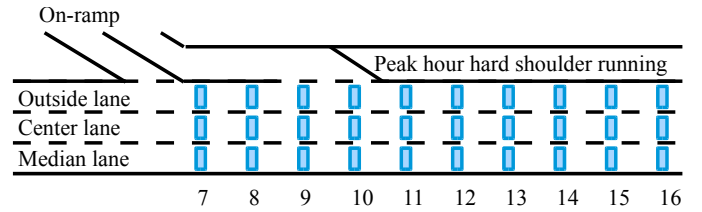


Fig. 2. Geometry of the downstream end of the instrumented section showing the location of double loop detector sites. Note that in the UK, slower vehicles drive on the left. HS denotes the hard shoulder (emergency lane) and ATM (Active Traffic Management) denotes hard shoulder which can be activated as an ordinary running lane in peak periods. The on-ramp has both a right-hand *forced merge* lane and a slower left-hand lane typically used by heavy goods vehicles (HGVs). When ATM is turned off, HGVs are required to join the main carriageway (lanes 1–3) between sites 8 and 10. In this study we try to reduce the effect of the on-ramp as far as possible and focus on the downstream section between sites 14 and 16.

detector data from a single detector site. The key requirement is that at this site, each lane of the highway is equipped with a (double) inductive loop detector and that there is some facility for identifying when a single vehicle overlaps (or *straddles*) the detectors in adjacent lanes. These conditions are realized in a recent study on the M42 motorway in the United Kingdom. Moreover, this site is equipped with a very dense coverage of loop detectors, enabling us to reconstruct individual vehicle trajectories. Thus lane changes can be observed independently of straddles and thus the correlation between lane-changing and straddles can be analyzed. For other sections with less dense loop detector data an alternative way to find the ground truth for lane changes (e.g., video [22]) is required for the calibration step. However, provided this calibration can be achieved, sparsity of the detector coverage in itself does not break our procedure. Firstly section III-A gives general background details of the M42 study and then section III-B gives further details of the inductive loop hardware.

A. Details of the instrumented section

The data used for this analysis come from the M42 motorway near Birmingham, United Kingdom. An approximately 1 mile section was equipped with detectors every 100 meters, at which individual vehicle data was recorded. Due to space restrictions, we only show the downstream part of the section (figure 2). At each site, the time and lane number of each passing vehicle are recorded, in addition to estimates of the vehicle’s length and speed.

The detector sites are separated by only 100m nominally, which gives travel times of 3-5 seconds in free flow conditions. In this short time interval, vehicles do not often decelerate or accelerate too markedly between sites, so it becomes possible to re-identify vehicles from site to site, and thus in effect follow them down the highway (an idea pioneered by [23]). Throughout the paper, re-identification is used to mean the identification of a vehicle at a more downstream site, as opposed to straddling, which means identifying the same vehicle in two lanes at the same site. Whereas straddling is crucial in applying the method presented in the paper, the re-identification, discussed in this section, is used to develop a calibration set for lane changes.

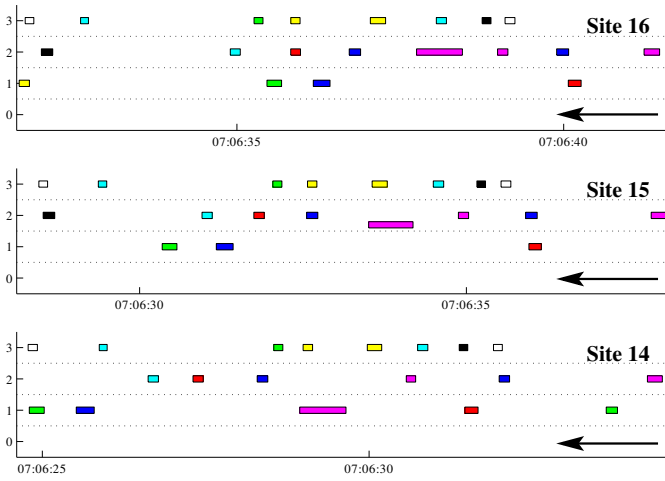


Fig. 3. Visualization of a short section of Individual Vehicle Data from three inductive loop sites labeled according to figure 2. The horizontal time axis is shifted from site to site to account for the average travel time between sites. The coloring of each vehicle helps indicate the re-identification between sites as obtained by automated algorithms. In this example, we observe a heavy goods vehicle (colored purple in the figure) pulling out from the outside lane towards the middle lane between sites 14 and 15, which has been *straddling* at site 15.

The basic idea of re-identification is to exploit the estimated arrival time at a loop, vehicle speed, headway and length to match vehicles. For further details, see [24], [25]. In very congested or strongly dynamic traffic conditions, however, the re-identification process fails, since the travel time between two sites increases, leading to uncertainty in the matching process. The present study is therefore restricted to fluid traffic conditions (speeds of over 20 m/s = 44 mph).

Figure 3 illustrates the re-identification process and the identification of lane changes in particular. One drawback of this approach is that the error rate of the re-identification algorithms is not known exactly. However, as part of a commercial trial, a small sample was manually cross-checked with video data and this process indicated an error rate in free flow conditions of order 1 in 1000 [26].

We should emphasize that the restriction on speed relates only to the limitations of the re-identification process, and not to the overall method that we propose. In other detector set-ups, the lane-changing data needed for our methods' calibration step would probably be derived by other means, e.g., from small samples of video data for which speed restriction would not apply. Of course, there is the independent problem that the straddling rate parameter may depend on the speed of vehicles, and we return to this question in section VI.

B. Details of the inductive loop hardware

Loop detectors as considered here are based on the principle of electromagnetic induction. The inductance of coils of wire buried in the road surface is altered by the proximity of (metallic) vehicles which may thus be detected by signal processing electronics in the road-side outstations. In standard installations, the auxiliary electronics detects the times at which the inductance rises and falls below given threshold

levels, corresponding roughly to the passage of a vehicle's front and rear end. (Of course, the precise timing is dependent on the threshold levels that are set, the material make-up of the bottom of the vehicle, and its clearance from the ground.) For single loop installations, this permits the direct counting of vehicles and the measurement of occupancy. However, in double inductive loop installations as we have here, two loops are installed which are separated longitudinally by several meters. Consequently each passing vehicle gives an "on" and "off" time at each of the two loops and these four times may be processed via divided differences to estimate the vehicle's length and speed.

However, the loops in our instrumented section are enhanced with the advanced IDRIS [26] signal processing system which also takes into account the inductance signature of each passing vehicle: that is the inductance profile of each passing vehicle is captured as a fine-resolution function of time. In particular, this inductance signature can be used to help re-identification and also to classify vehicle type (since axles show up very clearly in the signature). When there is a signal at the detectors in two adjacent lanes, the IDRIS system uses a proprietary algorithm to compare the two signatures and determine whether they are due to two separate vehicles, or to one single vehicle that *straddles* the two detectors. This effect is also known as *splashover* and methodologies are known to identify these straddles and to correct for double counting [7]. However, we will use this information to estimate the lateral position of each vehicle on the road.

In data thus obtained from the IDRIS system, each vehicle record has in addition to its lane data a straddling flag which takes one of three values:

- *no straddle*: the vehicle is in the center of its lane.
- *straddle left*: the vehicle is overlapping partially with the lane to its left.
- *straddle right*: the vehicle is overlapping partially with the lane to its right.

The idea is that a vehicle changing lane to the right, from lane i to lane $i + 1$, should pass through the following sequence of lateral positions: (1) no straddle in lane i , (2) straddle right in lane i , (3) straddle left in lane $i + 1$, (4) no straddle in lane $i + 1$. However, the duration of a lane change may be relatively short compared with the longitudinal spacing between detector sites. Consequently, many lane changes may not be detected by the straddle flag.

Finally, note that the sensitivity with which straddles are detected depends on many factors specific to the characteristics of the given road section and the inductive loop installation. Wider, longer or shallower loops (or narrower lanes) are likely to increase the detection rate of straddles, which will also depend on the signal processing electronics and algorithms, of which IDRIS is just one such possible system. Even for IDRIS installations, there are tolerances which can be set so as to affect the detection rate for straddles. In consequence, for any given implementation, there is no substitute for calibrating the straddling process with some "ground truth" data on the lane-changing rate.

IV. CALIBRATING THE STRADDLING PROCESS

We now show how to calibrate the relationship between straddling and lane changes. For concreteness, we work with the data set introduced in section III, where counts of lane changes may be obtained via the re-identification of individual vehicle data. In other implementations, a small set of observations of lane changes must be obtained by other means.

Since our main purpose is the illustration of the method, rather than an exhaustive study, we focus on the straddles recorded at a single detector, namely number 15 as labelled in figure 2. The re-identification through sites 14, 15 and 16 is then used to determine whether vehicles change lane. We thus correlate lane changes as detected in a section of approximately 200 meters in length with straddles as detected at its midpoint. In choosing the most downstream portion of the instrumented section, we simplify matters a little by minimizing the impact of the on-ramp which features further upstream. However, the analysis that follows can be worked through for any 3 consecutive sites in the data set. The main requirement is that at the straddling detector site the number of lane changes is representative of the number of lane changes of the studied section.

Table 4 counts the possible combinations of straddling and lane changing as derived from a single day of data (31 October 2008). To further simplify matters we focus on vehicles which are recorded in the middle lane at site 15, of which there are 22,505 in total on the day in question, after approximately 5,600 slow vehicles (slower than 20 m/s) are removed due to re-identification issues. The middle lane is the most interesting since vehicles change lanes both to the left and right, both in and out of the lane. However, the calibration procedure may be performed rather easily for the other lanes too.

The lower half of figure 4 displays the ratios of interest. Let us discuss the cases with straddling. Firstly, no vehicles at all straddle to the ‘wrong’ side (i.e., first straddling right and then changing lanes to the left downstream, and the 3 other related combinations). Likewise, very few vehicles give rise to straddles without changing lane — in fact, only 6 out of the 21,326 (21,320+3+3) cars that go straight on do so. Therefore, in this set-up it is a good modelling assumption to suppose that *all* straddles result from lane changes. However, in other implementations, the model may require a non-zero rate parameter for false positives. Finally, note that at this site, left (resp. right) straddles correspond more or less equally to upstream changes to the right (resp. left) and downstream changes to the left (resp. right). Thus the direction of the straddle cannot be used to determine the direction of the lane change. For the remainder of the paper, we will thus simplify matters by counting only three cases: (i) not changing lanes and no straddle, (ii) changing lanes and no straddle and (iii) changing lanes with a straddle.

These observations lead us to model the occurrence of straddles as a stochastic process based on lane changes. Specifically, we shall assume that each lane change gives rise to a straddle with probability p , which we shall assume for simplicity is constant and independent of factors such as the speed and length of each vehicle — we return to this

TABLE I
OBSERVATIONS AND RELATIONSHIPS DERIVED FROM FIGURE 4

	Nr. of obs.	Percentage	Comment
Total number of vehicles observed on lane 2 site 15	22505		
Number of vehicles observed also on lane 2 on site 14 and 16 (did not change lane neither between 14 and 15 nor between 15 and 16)	21326	21326/22505 = 94.8%	Inverse of lane change rate: 95% does not change lane; a little over 5% changes lane in 200m
Number of vehicles observed to straddle, but are also on lane 2 on site 14 and 16 (did not change lanes either between 14 and 15 nor between 15 and 16)	6	6/21326 = 0.028%	false positives: less than 3 per 10000
Number of vehicles observed on lane 2 on site 15 and 16 and either on lane 1 or on lane 3 on site 14 (change lane upstream of site 15)	526	526/22505 = 2.3%	2.3% of the total number of all vehicles changed lanes in the 100 meters between site 14 and site 15
Number of vehicles observed on lane 2 on site 15 and 16 and either on lane 1 or on lane 3 on site 14 (change lane before site 15) and made a straddle on site 15	88	88/526 = 17%	Straddle rate: 17% of the vehicles that did change lane upstream of site 15 generated a straddle on site 15
Number of vehicles observed on lane 2 on site 14 and 15 and either on lane 1 or on lane 3 on site 16 (change lane downstream of site 14)	653	653/22505 = 2.9%	2.9% of the total number of vehicles changed lanes in the 100 meters between site 15 and site 16
Number of vehicles observed on lane 2 on site 14 and 15 and either on lane 1 or on lane 3 on site 16 (change lane after site 15) and made a straddle on site 15	104	104/653 = 15.9%	Straddle rate: 16% of the vehicles that changed lanes downstream of site 15 generated a straddle on site 15

point in section VI. Thus assuming that straddling events are independent, the probability of obtaining s straddles from lc lane changes is thus given by the binomial distribution

$$P(N_S = s) = \binom{lc}{s} p^s (1-p)^{lc-s}, \quad (1)$$

where N_S denotes the stochastic variable for the number of straddles. Note that $(1-p)$ is the “failure” rate, that is the probability with which a lane change does not trigger a straddle. For any given implementation, the only goal of the calibration procedure, is to find p . The distribution in equation 1 implies the intuitive maximum likelihood estimator

$$p = N_S / N_{LC}, \quad (2)$$

where N_S and N_{LC} are the total number of observations of straddles and lane changes respectively in the calibration data. But more importantly, the maximum likelihood apparatus also provides confidence intervals for p . Using the data from 31 October 2008, we obtain $p = 0.162$ with a 95% confidence interval of about ± 0.02 . If one could use two days of data

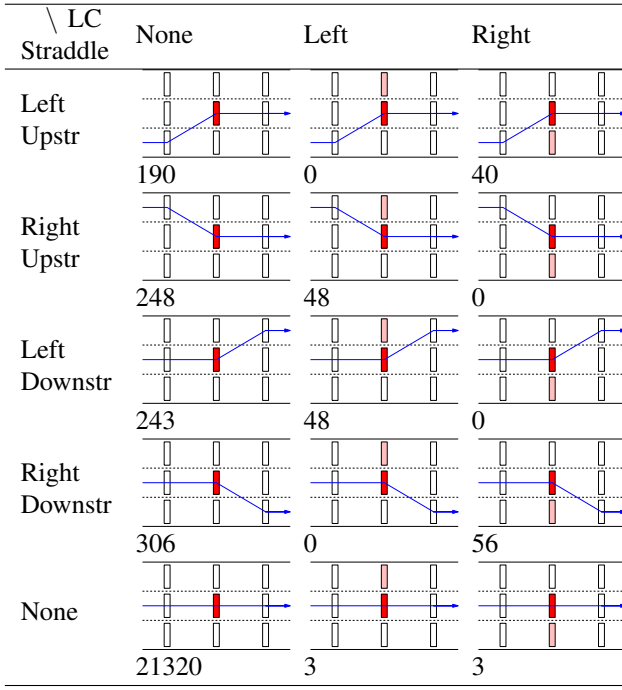


Fig. 4. The different possibilities of straddling and lane changing (upstream, “upstr”, and downstream, “downstr”) at site 15. The figure shows three lanes at sites 14, 15 and 16. The vehicles which are studied are the vehicles which are detected at site 15 at the middle lane (principally - sometimes some splashover effects), hence this detector is colored red. The movement of the vehicle is indicated with an arrow. The splashover effects in another lane, caused by a straddling vehicle, are indicated by pink shading. See also table I.

in the calibration, one would expect a confidence interval of $\simeq \pm 0.02/\sqrt{2} \simeq \pm 0.014$, by the asymptotic normality of the estimator.

Of course, in our particular implementation we have the luxury of a surfeit of data with which to perform the calibration and to test methods based upon it. To this end we took a particularly clean stretch of data (with few equipment outages etc.) from the period 1 October to 30 November 2008 inclusive. As before, we discard periods with slow moving vehicles (so that the re-identification is robust) and periods where there are equipment outages. Two thirds of the two-month data set is then selected at random on a day-by-day basis for calibration purposes, whilst the remaining one third is held out for validation (see section V). In this case, the maximum likelihood estimator gives $p \simeq 0.164$, with 95% confidence bounds of 0.161 and 0.167 — which are probably much tighter than one could expect to attain in fresh operational applications, owing to the large volume of calibration data available here.

Finally, we may aggregate the calibration data in intervals of 60 minutes — which was considered the shortest interval for which there were statistically robust numbers of straddles and lane changes to work with. Intervals where 20% or more of the vehicles had a speed of less than 20 m/s were considered (possibly partly) congested and therefore, due to re-identification issues, were discarded. However, even after this selection, more than 80% of the data is still remaining. We then record the number of straddles, N_S , and the number of lane changes, N_{LC} , that were recorded in the aggregation

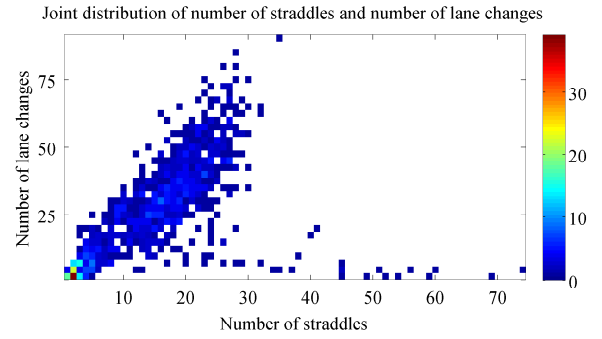


Fig. 5. The joint distribution of the number of straddles and the number of lane changes in the calibration set. The colour denotes frequency of one hour aggregation intervals as recorded in the 40 day calibration set. Vertical sections through this figure imply a probability distribution for the number of lane changes given an observed number of straddles.

intervals that remain.

The chief output of the aggregation is the joint distribution of figure 5, displaying the observed frequency of different (N_S, N_{LC}) pairs. The scatter in this figure emphasises the stochastic relationship between straddles and lane changes. In the next section we will consider how to use the data in this figure to estimate the number of lane changes from an observed number of straddles.

V. RECOVERING THE NUMBER OF LANE CHANGES FROM THE STRADDLES

We now consider how to estimate the number of lane changes in operational data. We suppose that the calibration step is complete, and to illustrate our technique, we use the one third portion of hold-out (validation) data left over from section IV (20 days in total), which we aggregate into 60 minute intervals. There are two goals:

- 1) Estimate the probability distribution for the number of lane changes in an aggregation period with a given observed number of straddles.
- 2) Show the accuracy of this estimate.

The first goal is achieved by the application of Bayes' rule

$$P(N_{LC}|N_S = S_0) = P(N_S|N_{LC}) \frac{P(N_{LC})}{P(N_S)} \quad (3)$$

For the prior $P(N_{LC})$ we use $P(N_{LC} = n) = (1/p)P(N_S = (p*n))$, where $P(N_S)$ is the experimental probability density function. $P(N_S|N_{LC})$ is a binomial function with success rate p . The remainder of the section illustrates how this rule is applied in practice.

The key idea is that the required distribution of the number of lane changes (goal 1) can be read from the matrix describing the joint distribution of straddles and lane changes (see figure 6a). This matrix shows how often each combination of straddles and lane changes occurs. Once this matrix is estimated, one row of the matrix will represent the distribution of the number of lane changes given an observed number of straddles.

The first step in this procedure to estimate the total distribution of lane changes (see section V-A). The second step is to

divide this distribution over different values for the number of straddles (section V-B). Section V-C shows the normalization which is required to estimate the number of aggregation intervals in which a specific number of lane changes is given. This last point is only relevant to check the accuracy of the estimation (goal 2), which we address in section V-D, by comparing with the true lane change count in the validation data.

A. From the measurement of the number of straddles to the distribution of the number of lane changes

The starting point for the method is a set of data for which the number of straddles is measured – here, we will use the validation set which is also used to test the accuracy of the method (goal 2). We will assume that the number of aggregation intervals is large enough to have a good distribution function for the number of straddles, which is required to test the accuracy of the estimate. We applied the method to the 1/3 hold-out data which was kept separate from the remainder of the data set for this purpose; as for the calibration set, we retain only uncongested data.

The first step is to obtain an overall distribution for the number of lane changes (see figure 6a). This is based on the observed distribution of straddles per aggregation interval (60 minutes) in the validation set, P_S (see figure 7a). The idea is to stretch this distribution in order to arrive at the distribution of lane changes, since the average number of straddles per lane change is known from the calibration step ($p = 0.164$).

Since the bin for each straddle count corresponds to several bins of lane change counts, the distribution function has to be interpolated to give the lane change distribution. This means that for each integer value lc of the number of lane changes, the corresponding average number of straddles is calculated from $s = p * lc$. The value s is generally non-integer, and we call its nearest integer neighbors s_+ and s_- with $s_- \leq s \leq s_+$. The estimated probability density function of lane changes $P_{LC}(lc)$ is then the weighted combination of the measured probability density functions of the straddling in s_+ and s_- , where the weights depend on the distance between s and s_- , and s and s_+ respectively. We have

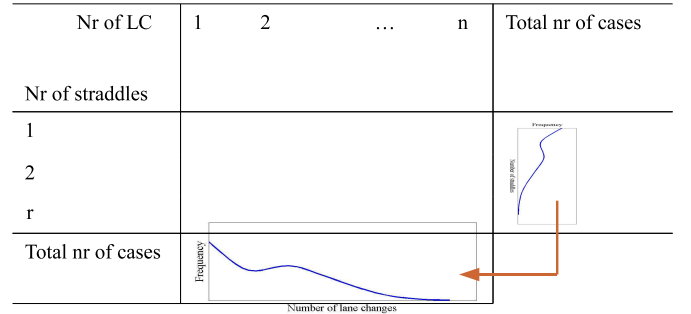
$$P_{LC}(lc) = (s_+ - s)P_S(s_-) + (s - s_-)P_S(s_+) \quad (4)$$

In case s happens to be an integer number, $s_- = s_+ = s$ and $P_{LC}(lc) = P_S(s)$.

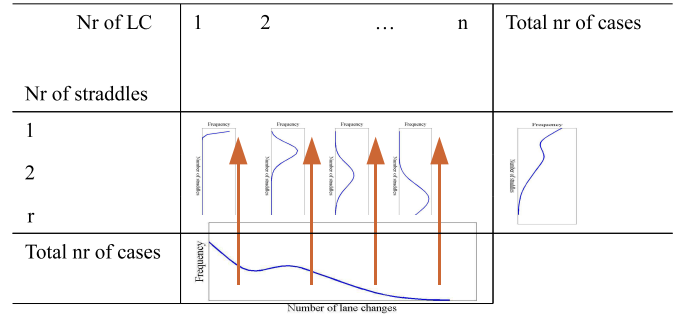
The function P_{LC} finally needs to be normalized so that

$$\sum_{lc=0}^{\infty} P_{LC}(lc) = 1 \quad (5)$$

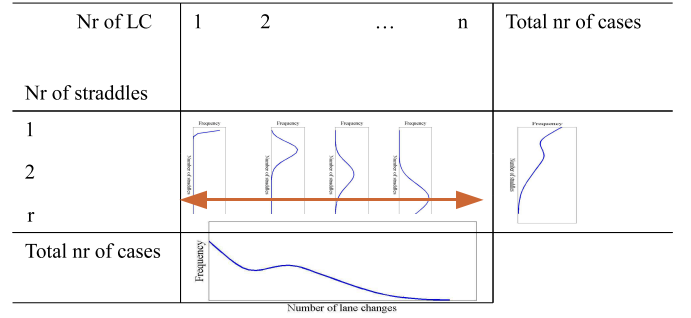
The accuracy of the result can be tested, since for the calibration set the number of lane changes is also measured. The estimated probability distribution function and the measured probability distribution function are shown in figure 7b, which shows that the estimation has the same pattern as the measured number of lane changes, but shows less noise.



(a) Step 1: the measured distribution of straddles is stretched (using $1/p$) to obtain an estimate for the total lane change distribution



(b) Step 2: the aggregation periods in each of the bins of equal number of lane changes are distributed over different numbers of straddling. To this end, a binomial distribution is used.

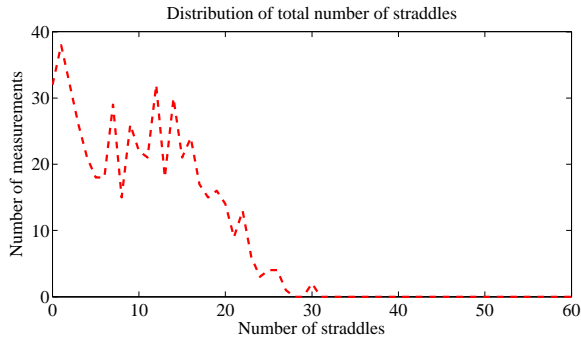


(c) Step 3: multiply the distribution by a scale factor such that sum of the obtained numbers in each of the bins of equal number of lane changes (for the same number of cars straddling) add up to the measured number of straddles.

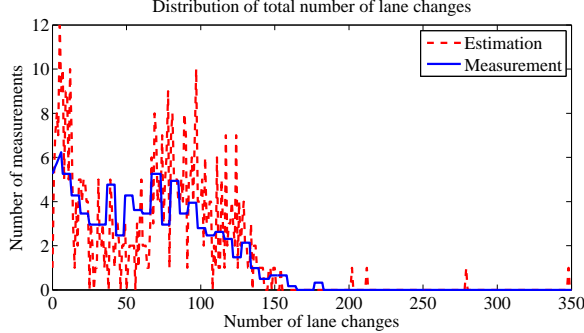
Fig. 6. The process of finding the distribution of lane changes. The basic idea is to construct a matrix which shows how often each combination of lane changing and straddles is measured.

B. From the distribution of lane changes to the joint distribution of the number of lane changes and straddles

The distribution function for the number of lane changes needs to be split up for different numbers of straddles. Only then can we find the distribution for the number of lane changes given the number of straddles in an aggregation interval (goal 1). This is achieved as follows. For each of the bins of the total distribution function of the number of lane changes lc , the distribution of the number of straddles is given by theory, namely equation 1. Once the number of lane changes is known (i.e., the column number) as well as the parameter of the binomial distribution (p from the calibration), this matrix can be filled, column by column. This spreads the distribution of the number of lane changes to a two-dimensional (joint) distribution showing the coincident



(a) Measured distribution of the number of straddles



(b) Estimated and measured distribution of the number of lane changes

Fig. 7. Estimation of the distribution of the total number of lane changes based on the distribution of the total number of straddles

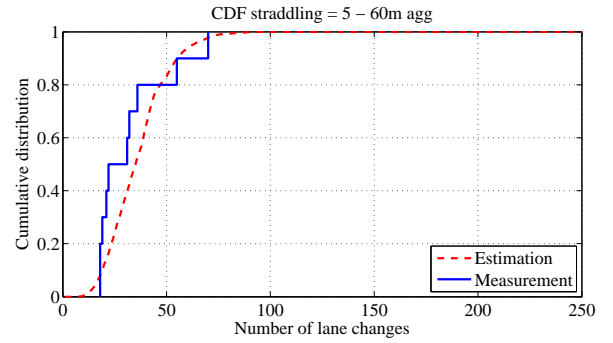
probability between the number of lane changes and the number of straddles. Figure 6b shows this graphically.

C. Normalization of the distribution function for the number of lane changes for a given number of straddles

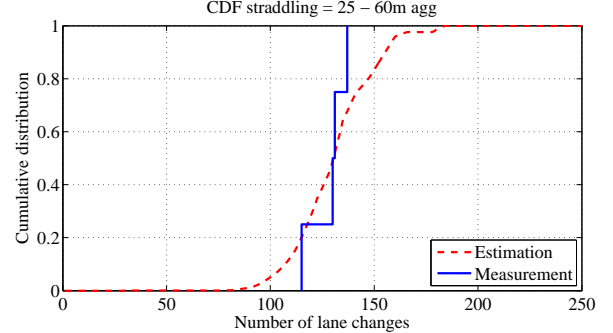
One row of the matrix indicates the distribution of the number of lane changes for a certain value of straddling. This now has to be checked for accuracy (goal 2 of this section). However, the sum of the aggregation periods in the row with s straddles is generally not equal to the measured number of aggregation periods with s straddles. Therefore, all frequencies in the row with s straddles are proportionally rescaled such that the total number of aggregation periods with s straddles matches the number of observations. Then, the resulting numbers are the expected frequencies of lane changing in the set of aggregation intervals with that number of straddles. Figure 6c shows this process graphically.

D. Results

We now present results to evaluate the accuracy of the method described in sections V-A to V-C (goal 2 as mentioned at the introduction of section V). The available data (see section III) has been split into two parts on a day-by-day basis. Two thirds of the data has been used for the calibration process as described in section IV. The method to find the number of lane changes from the number of straddles has been applied to the remaining 1/3 of the data, which is the validation set. In this section, we only study the number of lane changes between loop 14 and 16, and relate that to straddling at loop



(a) 5 cars straddling



(b) 25 cars straddling

Fig. 8. The estimated distribution of lane changes and the measured distribution for the validation set displayed by cumulative curves

15. As mentioned in section III, all data (in both data sets) where the average speed is below 20 m/s is removed due to re-identification problems. Section VI will discuss the impacts of these restrictions.

Our method leads to an estimation of the number of lane changes based on the number of straddles. This estimation is a distribution. To compare the distribution with the data (from the validation set), we gathered the aggregation periods in groups with the same number of straddles per aggregation period. This grouping was across the day and over all days in the validation set. The estimated distribution is the same for all aggregation periods in the group, since it is based only on the number of straddles. This estimated distribution can be compared with an empirical distribution of lane changes, namely the distribution of actual lane changes for all aggregation periods in the group. This process is shown in figure 8, both for the group with 5 straddles and for the group with 25 straddles. Here, 60 minutes is chosen as the length of the aggregation interval. The results are quite accurate. For the group with 5 straddles, the estimated distribution follows the empirical distribution closely. There are only 4 aggregation intervals which have 25 straddles. The empirical distribution therefore has only 4 “steps”, but even in this extreme situation, the observations are well aligned with the estimated distribution (figure 8b).

The most relevant quantity is the mean number of lane changes for a given number of straddles, indicated by $\langle N_{LC}(N_S = S_0) \rangle$. We compare this quantity with the mean expected number of lane changes for a given number of straddles, namely $\langle \widetilde{N}_{LC}(N_S = S_0) \rangle$. The relative error

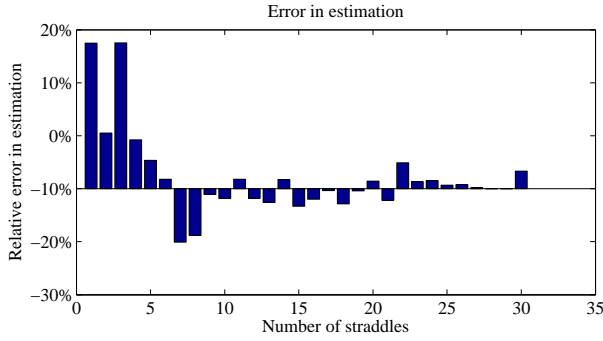


Fig. 9. The relative error of the estimation of the validation set, calculated by the difference of the mean of the estimation and the mean of the calculation divided by the measured number of lane changes. A negative relative error means an underestimation of the number of lane changes.

$\frac{\langle \widetilde{N_{LC}} \rangle - \langle N_{LC} \rangle}{\langle \widetilde{N_{LC}} \rangle}$ between the two is shown in figure 9 for all values of straddle counts. Even for the cases with very low numbers of straddle observations (see figure 7), the estimation is quite good (relative error never more than 60%). The relative error is mostly below 10%. This is good, since the number of lane changes in each aggregation interval varies by more than a factor 10.

VI. FACTORS IN THE GENERALISATION TO OTHER IMPLEMENTATIONS

Section IV was concerned with the calibration of p , which is the quotient relating the number of straddles to the number of lane changes. As we have already discussed, this quotient will depend on a number of site-specific and implementation-specific factors, for example:

- the width of lanes;
- the width of detectors;
- the sensitivity of the detector (incorporating both hardware sensitivity and algorithms / software thresholds);
- the width of vehicles;
- the “length” of the lane change manoeuvre — in the sense of the longitudinal distance that is covered from mid-lane position in the original lane to mid-lane position in the receiving lane — which may in turn depend on a vehicle’s speed;

and of course the length D_{section} of the road section considered, which so far we have taken to be 200 meters (namely, the distance from site 14 to site 16 in our implementation, see figure 2). The factors listed above may be rolled-up in to a single distance D_S , defined to be the longitudinal distance during a lane change through which the given vehicle would trigger a straddle, if a loop were present to detect it. Assuming a uniform distribution of lane changes over the length of the road, one obtains intuitively

$$p = \frac{D_S}{D_{\text{section}}} \quad (6)$$

and we will verify this formula with data in section VI-B. However, in section VI-A we will first examine the effect of vehicle speed on D_S .

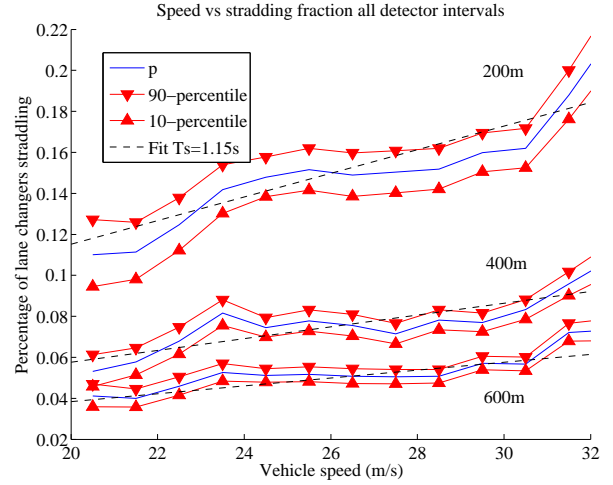


Fig. 10. The values of p as a function of the speed, and the fit with a constant straddling time per lane change

A. Speed of the vehicles

It is reasonable to assume that lane changes have a constant duration in time, of the order of several seconds [19], [27]. However, the straddling time T_{strad} , during which a straddle might be detected, is shorter than this owing to the sensitivity factors we have discussed, but a first order estimate is a constant straddling time. We may re-write equation 6 in the form

$$p = \frac{D_S}{D_{\text{section}}} = \frac{v T_{\text{strad}}}{D_{\text{section}}} = \beta v, \quad (7)$$

in which $\beta := \frac{T_{\text{strad}}}{D_{\text{section}}}$. We will now analyse the effects of speed on p , and thereby find β , and thus relate it to a straddling time.

Due to re-identification problems, cars driving under 20 m/s were discarded from the sample, but the speed dependency from 20 m/s and up to the speed limit (70 miles per hour, 31 m/s) will be analysed to check the assumption of a constant straddling time. To this end, we grouped vehicles in bins with similar speed (bin width=1 m/s), and calculated the straddling quotient p for the lane-changing vehicles using the same method as in section IV (a maximum likelihood estimator). The results, including the confidence intervals are plotted in figure 10 (see part labeled “200 m”). The figure indicates that the assumption of a constant straddling time gives a reasonable fit to data. When fitting only to the 200m section, we find $T_{\text{strad}} \simeq 1.12$ seconds. This fit lends some credence to the application of our process to vehicles that are slower than 20 m/s, for which calibration was not possible in our data (due to the low re-identification rate).

Furthermore, it is remarkable that at approximately 25 m/s the straddling rate is higher than the fitted line. This is probably caused by trucks which have a longer lane change time [27], and are wider, which means that they cause a straddle during a larger part of the lane change maneuver.

B. Distance between measurement sites

Equation 6 postulates the effect of the length D_{section} of the road section upon the rate parameter p . We now check this formula with the M42 data. To this end, we repeat the above analysis for longer road sections, by using non-consecutive loops to define the extent of the section considered. See figure 2. To neglect the effect of the on-ramp, we will prefer to stay as far downstream as possible, so that detector 10 will be the furthest upstream detector that we use. The approach (as before) will be to select three sites, with the most upstream and downstream used (via re-identification) to observe lane-changes, and with the site at the midpoint used to detect straddles. Note that only testing requires information about the ends of the section. A road section with loops every 600 meters can be split up in sections of 600 meters, with a loop in the middle of each section. In our case, with calibration using the individual loop data, we need loops on the end points as well. Table II shows that for section lengths 200-600 meters, the fitted values for T_{strad} are almost identical, hence verifying equation 6.

TABLE II
THE COMBINATION OF DETECTORS USED

start site	end site	straddle site	length of stretch	T_{strad} (fit)	p
14	16	15	200m	1.12s	0.164(\pm 0.003)
12	16	14	400m	1.14s	0.078(\pm 0.002)
10	16	13	600m	1.19s	0.061(\pm 0.001)

Finally figure 10 displays the different lines for $p(v)$, as well as the lines assuming a constant $T_{\text{strad}} = 1.12$ s for all detector intervals. In summary, equation 6 can be used to extrapolate the performance of our method for the longer road sections that must be used when detector coverage is relatively sparse. The penalty (of course) is that p is smaller, so longer aggregation intervals are required for statistical robustness. Table II also gives the estimated p for different section lengths, estimated regardless of the speed, as well as the accuracy of the estimate using one day of data. As the section length increases, p decreases, and the confidence intervals of p decrease with them. In fact, the relative confidence interval for the value of p is the same (with 95% certainty within 4% of the estimated value).

Similarly, given a certain number of straddles at a detector, the expected number of lane changes is higher for low p , as is the confidence interval. We test this by applying the methodology of section V on the p values from table II and a very conservative prior estimate for a uniform distribution of the number of lane changes. The confidence interval which is required to include 65% of the likelihood (comparable to one standard deviation for a normal distribution) is approximately 6% of the most likely number of lane changes. This holds for all values of p : for high p , the number of lane changes, as well as the confidence interval are higher, but the relative error is constant.

VII. DISCUSSION AND CONCLUSION

This paper proposed a method to find the number of lane changes in a traffic stream by exploiting loop detector data.

From the number of simultaneous activations of loops in adjacent lanes an estimate of the number of lane changes can be made. The method does not result in a deterministic value given the number of straddles observed in any one given period. A solution is to group different periods to get an average, and compare that to other averages. For instance, all periods where a stop-and-go wave starts can be grouped, and the number of straddles, and thus lane changes, in this group can be analysed. This way, traffic operations can be better described, and modelled. When grouping over time, the method may be used to give information about the OD matrix since one can measure the crossing flows for a weaving section with two in-links and two out-links.

The proposed methodology requires a site-specific calibration. This could be achieved by individual vehicle data from closely spaced loop detectors, as in this study, but there are alternatives, for instance counting the number of lane changes (manually, or using a video) and comparing this with the straddling measured in the same time period. This procedure gives an initial set to calibrate the relationship between straddling on a detector and the number of lane changes, and it will also provide the confidence intervals. Afterwards, straddling on detectors can be used as a (calibrated and validated) estimation for the number of lane changes. The more data that is available, the more accurate the estimation will be.

The paper calibrated this method for one detector site and checked the validity of the estimation for the same site. The method was able to estimate the number of lane changes on average within 10% off the measured value. A sensitivity analysis showed that the method can also be applied for sections with longer inter-detector distances. It showed that the straddling time remains equal at approximately 1.1 second per lane change. This in turn means that the straddling distance, which determines the straddling rate, depends on the speed. In case the speed in the validation set and the calibration set differs, one has to take this variability into account.

However, future studies needs to reveal to what extent this measure is site-specific. The straddling duration will depend on the width between the detectors, the width of the vehicle, the trajectory a driver chooses and the detector characteristics. These factors are can be country-specific. The first two elements can be found in the literature or design guidelines (per country); for the third one, one has to make additional assumptions on human behaviour.

The same methodology can be applied to different data sources, so long as the lateral position of the vehicle is recorded at one site. For instance, one can think of a camera pointing vertically downwards from a gantry, revealing whether a vehicle is in-between lanes. The method might also be used with magnetometers, which should be laterally placed close enough that no vehicle can pass in-between two magnetometers, and during a lane-change, each vehicle can be measured by two magnetometers at the same time.

ACKNOWLEDGEMENT

Research supported by the Netherlands Organisation for Scientific Research (“How traffic jams start”) and HTAS

SPITS, COST Action MULTITUDE, Foundation Next Generation Infrastructures and the UK Engineering and Physical Services Research Council (EP/E055567/1). Thanks to IDRIS Diamond Consulting Services [26] and the Highways Agency for providing the data.

REFERENCES

- [1] W. L. Jin, "Macroscopic characteristics of lane-changing traffic," in *89th Annual Meeting of the Transportation Research Board*, 2010.
- [2] S. Ahn and M. Cassidy, "Freeway traffic oscillations and vehicle lane-change maneuvers," in *Proceedings of the International Symposium of Transportation and Traffic Theory*, R. E. Allsop, M. G. H. Bell, and B. G. Heydecker, Eds. Elsevier, Amsterdam, 2007, pp. 691–710.
- [3] M. J. Cassidy, C. F. Daganzo, K. Jang, and K. Chung, *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer, 2009, ch. Spatiotemporal Effects of Segregating Different Vehicle Classes on Different Lanes, pp. 57–74.
- [4] R. Schubert, K. Schulze, and G. Wanielik, "Situation assessment for automatic lane-change maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 607–616, 2010.
- [5] C. Choudhury, M. Ben Akiva, T. Toledo, A. Rao, and G. Lee, "State dependence in lane changing models," in *Proceedings of the 17th International Symposium on Transportation and Traffic Theory*. London: Elsevier, 2007.
- [6] A. Amditis, E. Bertolazzi, M. Bimpas, F. Biral, P. Bosetti, M. Da Lio, L. Danielsson, A. Gallione, H. Lind, A. Saroldi, and A. Sjögren, "A holistic approach to the integration of safety applications: The insafes subproject within the european framework programme 6 integrating project prevent," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 554–566, 2010.
- [7] H. Lee and B. Coifman, "Algorithm to identify splashover errors at freeway loop detectors," in *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, 2010.
- [8] B. Coifman, "Vehicle level evaluation of loop detectors and the remote traffic microwave sensor," *Journal of Transportation Engineering*, vol. 132, no. 3, pp. 213–226, 2006.
- [9] J. C. Medina, A. Hajbabaie, and R. F. Benekohal, "Wireless magnetometers for traffic data collection at signalized intersections," in *Proceedings of the Summer Meeting of the Traffic Flow Committee of the Transportation Research Board*, Annecy, France, 2010.
- [10] NGSIM, "Home of the next generation simulation community," www.ngsim.fhwa.dot.gov/, 2008. [Online]. Available: www.ngsim.fhwa.dot.gov/
- [11] S. Hoogendoorn, H. J. Van Zuylen, M. Schreuder, B. G. H. Gorte, and M. Vosselman, "Microscopic traffic data collection by remote sensing," *Transportation Research Record: Journal of the Transportation Research Board No. 1885*, pp. 121 – 128, 2003.
- [12] A. Angel, M. Hickman, and P. Mirchandani, "Methods of Analyzing Traffic Imagery Collected From Aerial Platforms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 2, pp. 99 – 107, 2003.
- [13] L. Zhang, V. Kovvali, N. Clark, D. Sallman, and V. Alexiadis, "NGSIM-VIDEO User's Manual," US Department of Transportation, Federal Highway Administration, Tech. Rep. FHWA-HOP-07-009, 2007.
- [14] V. L. Knoop, S. P. Hoogendoorn, and H. J. Van Zuylen, "Processing Traffic Data collected by Remote Sensing," *Transportation Research Records, Journal of the Transportation Research Board*, vol. 2129, pp. 55–63, 2009.
- [15] J. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers, Part II*, vol. I, no. 2, pp. 325–378, 1952.
- [16] G. P. Gipps, "A model for the structure of lane-changing decisions," *Transportation Research Part B: Methodological*, vol. 20, pp. 403–414, 1986.
- [17] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transportation Research Part B: Methodological*, vol. 40, no. 3, pp. 251–264, 2006, doi: 10.1016/j.trb.2005.04.003.
- [18] McDonald, M., M. Brackstone, and D. Jeffery, "Simulation of lane usage characteristics on 3 lane motorways," in *Proceedings of the 27th ISATA Conference*, Germany, 1994, pp. 365–72.
- [19] E. C. B. Olsen, S. E. Lee, and M. J. ; Wierwille, Walter W.; Goodman, "Analysis of distribution, frequency, and duration of naturalistic lane changes," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 5, pp. 1789–1793, 2002.
- [20] A. Duret, S. Ahn, and C. Buisson, "Passing rates to measure relaxation and impact of lane-changing in queue," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 4, pp. 285–297, 2011.
- [21] V. Knoop, S. P. Hoogendoorn, S. Yasuhiro, and C. Buisson, "Quantifying the number of lane changes in traffic: Empirical analysis," in *Proceedings of the 91th Annual Meeting of the Transportation Research Board*, Washington D.C., 22–26 January 2012.
- [22] H. S.-H. Cheng, H.-Y., "Intelligent highway traffic surveillance with self-diagnosis abilities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1462–1472, 2011.
- [23] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," *Transportation Research Part A*, vol. 36, p. 899917, 2002.
- [24] R. E. Wilson, "From inductance loops to vehicle trajectories," *Transportation Research Circular*, vol. E-C149, pp. 134–143, 2011.
- [25] —, <http://www.enm.bris.ac.uk/trafficdata/>, <http://www.enm.bris.ac.uk/trafficdata/>.
- [26] IDRIS Consulting, <http://www.idris-technology.com/>, 2010.
- [27] S. Moridpour, M. Sarvi, and G. Rose, "Modeling the lane-changing execution of multiclass vehicles under heavy traffic conditions," *Transportation Research Record*, no. 2161, pp. 11–19, 2010.

Victor L. Knoop received his Master's degree in Physics from Leiden University (2005), and his PhD degree from Delft University of Technology (2009) on the effects of incidents on driving behaviour and traffic congestion.. After a post-doc at the university of Lyon (2009–2010) on lane changing, Victor Knoop is now again with the Delft University of Technology. His main research interest is the interaction between microscopic and macroscopic traffic flow phenomena.

R. Eddie Wilson moved to the Transportation Research Group in October 2010 as Professor of Modelling and Simulation. His background is in Applied Mathematics with a focus in particular in engineering and industrial applications. Eddie has collaborated with over 20 companies and government departments and is a leading member of the ESGI (European Study Groups with Industry) community. He serves on the EPSRC Strategic Advisory Team for Mathematics, and the scientific committee of the Smith Institute for Industrial Mathematics and Systems Engineering. He is a Fellow of the Institute for Mathematics and its Applications

Christine Buisson Traffic flow researcher since 1993, she is now Senior Researcher of IFSTTAR (formerly INRETS). She also teaches at ENTPE (French technical University in civil and environmental engineering). Her area of interest consists in macro- and micro-modeling of traffic flow and experimental validation of models. She is presently responsible of a French traffic data collection project MOCOPo (Measuring and modelling traffic Congestion and Pollution). She is member of Traffic Flow Theory and Characteristics committee (TRB, National Academy of Sciences, USA).

Bart van Arem holds a Master's degree (1986) and PhD degree (1990) in Applied Mathematics, speciality queueing theory at the University of Twente, the Netherlands. He has worked as a researcher and program manager at TNO between 1992 and 2009 on Intelligent Transport Systems, in which he has been active in various national and international projects. His interest focuses on Advanced Driver Assistance systems. Since 2009 he is a full professor Transport & Planning at Delft University of Technology, focusing on the impact of Intelligent Transport Systems on mobility.