Calibration and validation of probabilistic discretionary lane-change models

Victor L. Knoop, and Christine Buisson

***Abstract*—Lane changes are important in traffic flow operations. They cause differences in flow over lanes and determine in some cases the start of congestion. Whereas calibration and validation are commonly used with car-following models, this is not common practice with lane-change models. Even then, it is not clear what calibration and validation entails for probabilistic lane change models. Therefore, this article reviews methodologies to calibrate and validate probabilistic lane change models, both microscopically and macroscopically. A likelihood is often used in calibration, but does not intuitively show the quality of the model. An example showed that it is possible to have the model calibrated and validated with accurate parameters all having the same error in the validation as in the calibration, but the quality of the model was still bad. Using a likelihood ensures the stochastic effects are well captured, but the conclusion is that for validation purposes one can better use a measure which has physical interpretation and which gives a value indicating the quality of the model for the purpose for which it needs to be used.**

I. Introduction

In the process of designing motorway and motorway traffic management measures, it is useful to have tools which can predict traffic operations. The modelled traffic process can be divided into two parts: longitudinal (speed) and lateral (lane changing) operations. Lane changes can form the basis of traffic jams, and therefore need to be modelled accurately [1].

Calibration and validation of traffic models is essential to have predictive value. Whereas calibration is "the estimation of parameters to maximize the models descriptive power(...)"[2], the general goal for validation is to show whether the calibrated model can be used for prediction. The purpose of the model can change (location, time of day, weather), and hence the range of conditions for validation changes. In practice, different approaches are used for validation.

Lane changing is a stochastic process, and especially for stochastic models the approach for validation is not trivial. Section II-A will show definitions used in literature for validation. The main question discussed in this paper is whether the log-likelihood value can be used to calibrate and validate the in order to achieve model the required quality of the model.

This paper will show the need of validation even for calibrated models. Moreover, it will show that even a model which is calibrated and validated – and considered valid according to some available definitions – can be bad. We will do so using a basic stochastic model for discretionary lane changes. Essential part in the analyses is that the model is stochastic, which means that the calibration and validation take place by means of the *likelihood* of the model predicting the lane changes right. The paper will show a way of giving a physical interpretation to the outcomes of a likelihood test, which makes it possible use log-likelihood to show the model is bad. A interpretation can also be given by the transferring the model predictions to a physical quantity like the number of lane changes, and determine the RMSE.

The rest of the paper is set-up as follows. We first discuss the literature on calibration and validation, as well as lane change models. Then, we present the data which we will use (section III) and an example lane change model which we will use in this paper (section IV). The calibration and validation are discussed in section V and VI respectively. The paper ends with a discussion on the topic of validation (section VII) and conclusions in section VIII.

II. Literature review

A. Theory of the calibration and validation process

In the past decade, scientific attention has been given to the need to calibrate a traffic simulation program before its results can be trusted. For instance, calibration is part of the US DOT guidelines [3] and the topic of the European research program Multitude [4].

In a recent survey [5] 77% of the 215 respondents declared that they performed a validation in their last simulation study. However, there is no clear and universally accepted definition of validation. The transportation community uses the following definitions:

1) Some authors evaluate the parameters resulting from a calibration procedure. A global match of those values with the ones reported in literature is taken as a proof of validation [6].
2) Some authors suggest to divide the collected data into two parts, a calibration part and a validation part, where the calibration part is used for estimating the parameters. If the errors in the validation part are comparable to the errors the in the calibration part, the model is considered valid [7]. For a deterministic case, this is being used in [8].
3) A more purposeful definition is given by [2]: "Validation is the process of determining the reliability of a model, i.e., the degree to which it is an accurate representation of the real world from the perspective of the intended uses." Similar definitions, led by the purpose of the

V.L. Knoop is with the Department of Transport& Planning, Delft University of Technology, the Netherlands, v.l.knoop@tudelft.nl, C. Buisson is with Université de Lyon, Laboratoire Ingénierie Circulation Transports LICIT (IFSTTAR/ENTPE), France.

Manuscript received June 2013, revised April 2014

model, are given by [9] and [10]. For the application of this definition, one needs to specify moreover what quantity needs to be in what range of values. Later, we will refer to this method as physical variable.

We will hereafter illustrate that satisfying requirements in definition 1 and 2 is not sufficient to also satisfy requirement 3 and guarantee that the model is sensible. The requirement of the validated model in this paper is that it represents traffic on a particular roadway stretch. In the validation we hence only have to consider the same roadway stretch.

B. Discretionary lane changing models and their calibration and validation

Classically, lane changes (LCs) are divided into two types, distinguishing mandatory LCs from discretionary LCs. Mandatory LCs originate from an obligation to change lane for example for merging or diverging manoeuvres. For the first case, the literature review is made recently and we direct the reader to [11]. This paper will focus on discretionary LCs.

For calibration and validation data are needed. Traditional traffic observations are made at one point, and that limits the possibility of observing the performed lane changes, although recent techniques allow to overcome this limitation [12].

[13] present a discretionary lane change model, where the desire to change lanes depends on the drivers' state. This many-parameter model is calibrated on the maximum likelihood to find the correct vehicle trajectory. It is the first model which is also validated, but based on the the travel time distribution of vehicles. Note this variable is only on the second order linked with the number of lane changes manoeuvres.

In the same line of reasoning, [14] present a model where lateral and longitudinal driving behaviour are combined. The lane change decision depends on the possible longitudinal accelerations according to the longitudinal model. Later, the model is extended with several parts [15] and the final *integrated model* is presented. They use a maximum likelihood for the chosen path as objective function in calibration [16]. No validation is carried out.

The *MOBIL* lane change model [17] assigns utility to a lane change based on the acceleration of the lane changer, but also that of the surrounding vehicles. The model weights the utilities of the drivers and decides on a lane change based on this combined utility. The model is proposed without a calibration or validation. [18] presents a microscopic lane change model where drivers also react on each other, but even take time to adapt to new leaders. This stochastic microscopic model is calibrated and validated on macroscopic quantities, being the lane distribution (i.e., which part of the traffic flow is in which lane) on a motorway section of more than 9 km.

A behavioural theory on lane choice is presented in [19]. He introduces two driver types: *slugs* and *rabbits*. He poses that *slugs* only drive on the shoulder lane(s), and *rabbits* only on the median lane(s), as long as these are faster than the shoulder lane(s). This has been validated qualitatively on a Japanese freeway data[20]. No automated model calibration is presented but the lane changes are studied, and cumulative counts are used.

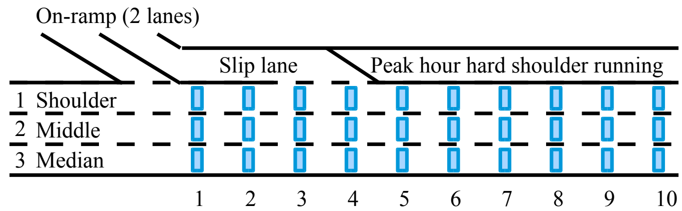


Figure 1: The layout of the motorway and the detectors. The detectors are placed at approximately 100 meter distances. Note that in the United Kingdom, people drive at the left hand side of the road. That means, left to right in this figure

[21] present another way of lane change modelling, using a macroscopic representation. The lane change rate, including relaxation, can be introduced into the macroscopic traffic equations. They show the validation thereof using cumulative counts.

The effect of lane changing, where vehicles take two places (at two lanes) in the traffic stream, is shown by [22], who give a microscopic interpretation of the process. In [23] this is translated into macroscopic traffic description. For both, a calibration has been carried out based on the shape of the fundamental diagram. This is also not directly linked to the number of lane changes.

As the overview of literature shows, there is no standard in calibration and validation of discretionary lane change models. For probabilistic models, the method of maximizing the log-likelihood is used most often, and mathematically the most sound, which is why we will adopt in this paper. Papers which use this method generally do not use the third definition for validation. Moreover, some models are calibrated, but not all. Only two models are validated. This paper will show that even after calibration a validation is required. It might even happen that a validated model is still performing badly.

III. Data collection and treatment

This section discusses the data: their availability and treatment (section III-A), as well as the basic traffic characteristics (section III-B).

A. Data availability

For the calibration, the larger the amount of data, the better. We therefore use an automated data collection method. The data we use are individual loop data from the M42 motorway in the UK near Birmingham [24]. The loops are placed approximately 100 meters apart, and for all vehicles the passing time, the lane, the speed and the length is recorded. It is a three lane motorway, and upstream of the section is a slip lane from an on ramp, see figure 1. To avoid the impact of the merging as much as possible, the article will focus on sites 5-10. To reduce this impact more, and also to reduce the impact of trucks, only the middle and median lanes are considered, and lane changes from the middle towards the median lane.

Since the passing times and the speeds, as well as the lengths and the lanes, are known for each vehicle at each site, the vehicles can be re-identified from one site to the next. If a vehicle is re-identified at the further downstream

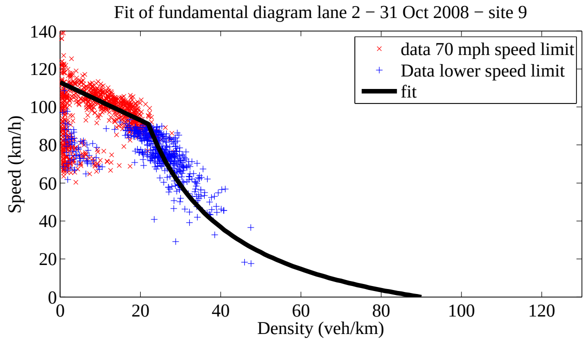


Figure 2: The fundamental diagram in the speed-density plane for the middle lane for a one-minute aggregation time

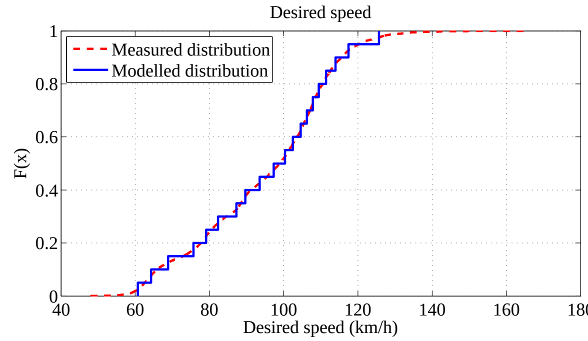


Figure 3: The distribution of the desired speed and the modelled distribution

site at another lane, a lane change has taken place. A very high re-identification rate of more than 999 out of 1000 vehicles is obtained (personal communication [25]), but the re-identification breaks completely in congestion. Then, the headways and speeds vary less over different, successive vehicles, and re-identification is impossible. Therefore, the data used in this study is limited to non-congested periods, here defined by speeds of 20 m/s (72 km/h) and higher.

In total, 2 months of individual vehicle data are available, recorded continuously (24/7). We aggregated them in periods of 5 minutes, and remove periods with dynamic speed limits or congestion. The resulting set consists of 5020 time periods of 5 minutes.

Since the vehicle is re-identified at the following detector, the time between the passing at two successive sites is known. Also, it is known whether the driver has performed a lane change between the two detectors from the re-identification at the downstream loop. It is assumed that no driver makes a lane change from his lane and back to his original lane within 100 meters.

B. Traffic characteristics

Figure 2 shows the fundamental diagrams for the middle lane. The blue crosses indicate the traffic operations for which there is a dynamically lowered speed limit, and which are discarded for the remainder of the paper. The red dots indicate the traffic operations for the other time periods, which are included. The free flow branches are fitted by hand to the data for which there was no dynamic speed limit; this is done for each lane separately. The congested branch has only points with a dynamic speed limit, and these fits will not be used in the remainder of the analyses.

The fact whether drivers are driving freely or not, depends on (1) the headway and leader's speed and (2) his/her own desired speed which is not directly observable. To overcome this limitation we consider that a driver is freely flowing if his/her distance headway is over a certain threshold (80 meters, corresponding to a density per lane of 12.5 veh/km). Note that to exclude congested regimes vehicles, we choose distance instead of the time headway here, since in congestion this time headway might be very large. The resulting distribution of desired speeds is presented in figure 3.

For reasons of computation, the distribution of desired speed is simplified and split into 20 key values, which each represent 5% of the free flow speed distribution. Figure 3 also shows how this approximation follows the real distribution.

IV. MODELLING

To illustrate our calibration and validation methodology, we need a discretionary lane change model to be tested against the data presented above. The sequel of this section will present a simple extension of a gap acceptance model. This simple model might not be optimal to catch all the (possibly underlying) processes. However, the aim of this model is to present the relevance of calibration, and not necessarily to represent the best model.

Due to limitations in the data set, we can focus our model on discretionary lane changes in the free flow states. We propose a simple model extending the concept of gap acceptance. The model bases the probability to change lanes on three elements:

1) The desire for a higher speed, represented by ϕ
2) The speed advantage in another lane, represented by χ
3) The availability of a gap, represented by ψ

In the end the elements are combined to a total modelled lane change probability by multiplying them. We realise that the model only represents a part of the motivation of changing lanes (α), and there is another part unexplained, which we will later on find is approximately 10%. Therefore, the combined model reads:

$$\Lambda = \alpha \cdot (\phi \cdot \chi \cdot \psi) + (1 - \alpha) \tag{1}$$

All elements ϕ, χ and ψ can be interpreted as a conditional lane change probability given the respective other two conditions are satisfied. In the remainder of this section, we will first introduce these elements one by one. In section IV-D the different parts of the lane change function are combined into one microscopic function which gives the overall probability for a vehicle to changes lanes. Section IV-E presents the macroscopic version of the same model.

A. Interest in improving speed

We focus this study on discretionary lane changes, from the slower lane (lane 2) toward the faster lane (lane 3). Therefore, we assume only drivers who drive at lower speeds than their desired speed are likely to change lanes voluntarily. Therefore we compute the ratio between the current distance headway and the difference between the desired speed and the leader's speed. We name this time "virtual time to collision" (κ_{virt}). For reasons of simplicity, we calculate the distance headway as the time headway (h) times the leader's speed (v_l), thereby ignoring the vehicle length.

$$\kappa_{\text{virt}} = \frac{hv_l}{v_d - v_l} \tag{2}$$

This differs from the regular time to collision in the speed of the follower, where we take the desired speed, v_d , rather than the actual speed. From the data we extract the combination of time headways and speed of the leader. By taking the combination of the two, we account for a possible correlation between time headway and speed of the leader.

For each combination of headway, desired speed, and the speed of the leader a virtual time to collision can now be calculated. If this virtual time to collision is smaller than a threshold value (named κ_0 hereafter), the ϕ value is positive. This κ_0 value is found by calibration for which the methodology is explained in section V-A. Hence, we model:

$$\phi(k) = P(\kappa_{\text{virt}} < \kappa_0) \tag{3}$$

If all these elements were known for a vehicle, this would be a discrete value. However, from measurements, even individual measurements, the desired speed of vehicles is not known, so the desired speed is assumed to follow a distribution, as indicated in Figure 3. *This distribution requires that

B. Interest in changing lanes

Drivers will only change lanes in case it is in their own benefit. So even if condition 1 is satisfied – they are stuck behind another vehicle, and are willing to pass that vehicle – lane changing is only useful if the speed in the other lane is higher. In the model, we consider the desire to change lanes to be dependent on the speed difference in the lanes, and assume:

$$\chi = \min\left\{\max\left\{\frac{v_{\text{new}} - v_{\text{c}}}{v_{\text{c}}}, 0\right\}, 1\right\} \tag{4}$$

This equation shows the relative advantage one has to change lanes from the current driving speed v_{c} to the speed in the adjacent lane v_{new}. In case the speed in the adjacent lane is lower, there is no benefit, and hence the desire to changes lanes is 0. For the sake of simplicity, we model a desire of 1 to change lanes in case the speed in the adjacent lane is double the speed in the current lane.

C. Opportunity to change lanes

Even if a driver wants to drive faster, and if the speed in the adjacent lane is higher, there is not a certain lane change. It is furthermore required that there is a sufficiently large gap in the adjacent lane. In fact, this is often the limiting factor in lane changing.

Quantitatively, the minimum time gap is determined as function of the speed difference, in which we follow [26]. The minimum gap, g_{min}, is modelled as a minimum gap in

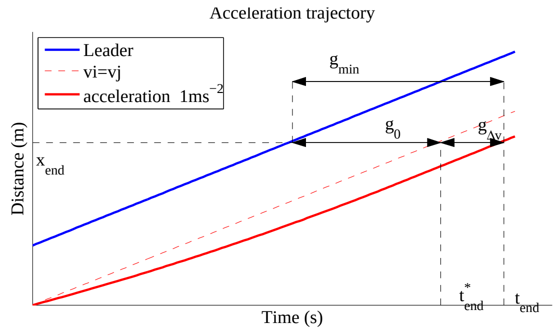


Figure 4: Illustation of the minimum gap

case there is no acceleration needed, g_0, and a dynamic part $g_{\Delta v}$, which accounts for the extra time needed to accelerate to the speed in the destination lane, v_j.

$$g_{\text{min}} = g_0 + g_{\Delta v} \tag{5}$$

Figure 4 shows this graphically. In this figure (t_{end},x_{end}) is the point in the time-space plane at which a driver reaches the speed of the target lane v_j, assuming a constant acceleration (value taken here is 1 m/s^2) from the moment he enters in point $(0, 0)$ with an initial speed v_i. t^*_{end} is the moment when this driver would have reached the position x_{end} if the initial speed had been v_j. The dynamic part of the time gap is now defined by the difference between those two instants:

$$g_{\Delta v} = t_{\text{end}} - t^*_{\text{end}} \tag{6}$$

ψ represents the probability that the upstream vehicle in the adjacent lane (the possible new follower) is far enough upstream. This gap is called g_{nf}. The gap acceptance ψ can therefore be expressed as

$$\psi = P(g_{\text{nf}} > g_{\text{min}}) \tag{7}$$

Note that for an individual vehicle this is fully observable, and there is no probabilistic element, and hence the value is 0 or 1 for an individual observation. However, the minimal gap g_0 is a parameter which should be calibrated.

D. Combining towards lane change rates

Mathematically, Λ indicates a probability to change lanes *per consideration of one driver*, which, by definition, has a duration τ (how often does a driver consider a possible lane change). There are $\Delta T/\tau$ considerations in the time interval ΔT, hence the lane probability that a driver does not change lanes in a time interval ΔT is

$$P(\text{No lane change in } \Delta T) = \prod_{i=1}^{\Delta T/\tau} (1 - \Lambda) = (1 - \Lambda)^{\Delta T/\tau} \tag{8}$$

Note that τ is short enough that typically not more than one lane change takes place during this period (later on we will see that τ is in the order of 15 seconds.)

In equation 8 we substitute the expressions for ϕ (equation

3), χ (4), and ϕ (7):

$$P(\text{No lane change in } \Delta T) = \cdots$$
$$\left(1 - \left(\alpha\left(\phi_{\kappa_0}(\kappa_{\text{virt}})\chi(v_{\text{new}}, v_{\text{c}})\psi_{g_0}(g_{\text{nf}})\right) + (1-\alpha)\right)\right)^{\Delta T/\tau} \tag{9}$$

This closed-form equation will be used to calibrate the parameters.

E. Macroscopic description

To get from a probability of a lane change P_{change} per unit of time to a probability of finding L_0 lane changes in a period of time, one uses a binomial distribution function:

$$P(L = L_0) = \binom{A}{L_0} (P_{\text{change}})^{L_0} (1 - P_{\text{change}})^{A-L_0} \tag{10}$$

In this equation A is the number of possible attempts to change lanes and can be expressed as follows:

$$A = k \cdot l \cdot \Delta T/\tau \tag{11}$$

In this equation, k denotes the density and l the length of the considered road section. A parameter in the function is τ, the time between two successive decision moments to change lanes. Furthermore, the number of lane changes depends on a parameter α, indicating how much of the lane changing is explained by the model, and two other intrinsic microscopic parameters, g_0 and κ_0.

V. Calibration

In this section we discuss the calibration. We first present the methods, for microscopic and macroscopic calibration, and then the results (section V-B).

A. Calibration methodology

Both the microscopic and the macroscopic model require in principle all parameters, which can be calibrated per vehicle. However, some variables are intrinsically microscopic. In particular, the critical gap (g_0) and the critical time to collision (κ_0) have to be calculated for individual vehicles. For the sake of simplicity, in the macroscopic calibration we will therefore fix the critical gap and the critical time to collision, and calibrate the time between two lane change considerations, τ, and the part of lane changing which can be explained by the model, α.

1) Microscopic calibration: This section presents how the probabilistic lane change model is calibrated and validated. The steps are shown graphically in figure 5. Note that the base element is *the probability of a lane change*.

From equation 9 we can derive the probability of a lane change:

$$P_{\text{change}} = P(\text{Lane change in } \Delta T) = 1 - P_{\text{no change}} = \tag{12}$$
$$1 - \left(1 - \left(\alpha\left(\phi_{\kappa_0}(\kappa_{\text{virt}})\chi(v_{\text{new}}, v_{\text{c}})\psi_{g_0}(g_{\text{nf}})\right) + (1-\alpha)\right)\right)^{\Delta T/\tau}$$

The goal of the calibration is to find the parameters τ, α, κ_0, and g_0. This is done by a maximisation of the (log-)likelihood.

For each vehicle passing at a detector site, the following variables can be measured for each vehicle individually:

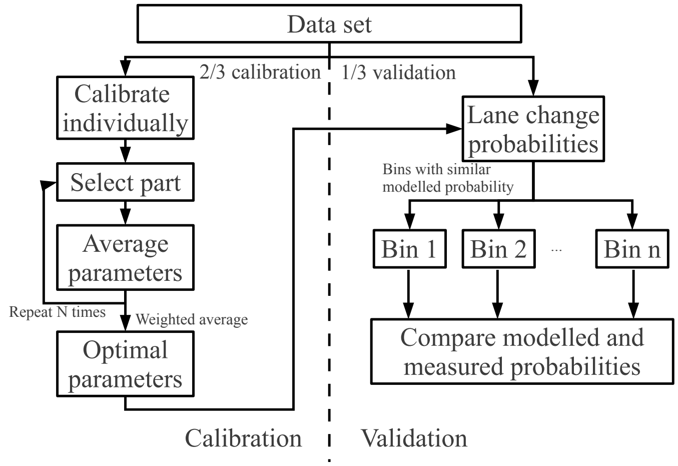


Figure 5: An overview of the steps in microscopic calibration and validation

- h, the headway between the vehicles
- v_{new}, the speed of the new leader
- v_{c}, the current speed
- g_{nf}, time gap to the new follower

Furthermore, we assume the free speed distribution, κ_{virt} from the data, figure 3. For each of the vehicles it is now possible to express the probability to find a lane change according to the model, only dependent on the parameters τ, α, κ_0, and g_0. These are the parameters which are calibrated.

Note that we aim to validate the model for the same section and conditions. We hence split the data randomly into two parts: a calibration part (2/3) and a validation part which we keep aside for validation purposes later on. This is an example of *holdout validation* [2]. For the calibration data, we list all vehicle passages at sites 4-9, later on identified at loops 5-10. These passages we split into two groups, those vehicles that have changed lanes ($\mathcal{G}$) and those that did not change and go straight on in the same lane ($\mathcal{S}$).

Since the lane change model is probabilistic, it is considered optimal if the likelihood that the model predicts the observed value is maximum. The likelihood is the probability of the occurrence of the observations as function of the parameters influencing these probabilities:

$$L(\tau, \alpha, \kappa_0, g_0) = \prod_{c \in \mathcal{G}} P_{\text{change}}, c \times \prod_{s \in \mathcal{S}} P_{\text{no change}}, s \tag{13}$$

For mathematical reasons, it is more useful to calculate the log-likelihood $\mathcal{L}$:

$$\begin{aligned}\mathcal{L}(\tau, \alpha, \kappa_0, g_0) &= \log(L) \\ &= \log\left(\prod_{c \in \mathcal{G}} P_{\text{change}}, c \times \prod_{s \in \mathcal{S}} P_{\text{no change}}, s\right) \\ &= \sum_{c \in \mathcal{G}} \log\left(P_{\text{change}}, c\right) + \sum_{s \in \mathcal{S}} \log\left(P_{\text{no change}}, s\right) \end{aligned} \tag{14}$$

Now, parameters values τ^* α^*, ${\kappa_0}^*$, and ${g_0}^*$ are found such that they maximise the log-likelihood:

$$\{\tau^*, \alpha^*, {\kappa_0}^*, {g_0}^*\} = \underset{\tau, \alpha, \kappa_0, g_0}{\arg\max}\left(\mathcal{L}(\tau, \alpha, \kappa_0, g_0)\right) \tag{15}$$

Table I: The results of the parameter calibration process

	microscopic		macroscopic	
Var	avg	stdev	avg	stdev
τ	16s	1.2s	12.1s	0.4s
α	0.915	0.006	0.927	0.004
κ_0	4.0s	0.4s		
g_0	4.0s	0.2s		

There are other techniques to find for instance the critical gap, but by using the maximum likelihood for all vehicles, we can optimize all parameters simultaneously.

To get not only the parameter values but also the variability, we perform this optimisation procedure repeatedly for different days, all from the calibration set. The set of parameters τ^*, α^*, $\kappa_0{}^*$ and $g_0{}^*$ satisfying equation 15 are determined, using *Matlab*'s function *fminsearch*. Now, we use 2/3 of the data (randomly chosen in each of the N iterations) in the calibration and hold back the remaining 1/3 of data for validation. This will give an average value for the parameter in that iteration. This is repeated 500 times (N in figure 5), which will give a distribution for the parameter values, which can be interpreted as the correctness of fit with those values (on this calibration set).

2) Macroscopic calibration: Rather than using the individual data and individual gaps, one might also use the *number of lane changes* (L) per time interval per road length as input for the calibration process. In this section we use the same model for the lane changes as in the microscopic description. For this application, the model predicting the number of lane changes has been adapted for aggregated input variables, and is described below.

For the sake of simplicity, the microscopic parameters, g_0 and κ_0, will not be calibrated on the macroscopic level and their values from the microscopic calibration will be used. Analogous to the microscopic methodology, we draw with 2/3 probability a calibration set from the data. This again is a holdout calibration [2]. Also analogous to the microscopic methodology, we repeatedly perform a calibration with a changing part of the calibration set to get a spread of the (calibrated) parameter values.

Figure 6 shows the process graphically. In analogy with the procedure described for the microscopic calibration, we choose a likelihood function as to indicate the performance of the probabilistic model:

$$L = \prod_{\text{all intervals}} P(L_{\text{measured}} = L_{\text{model}}) \qquad (16)$$

Calibration consists of finding these values for the parameters which maximise equation 16.

B. Calibration results

The model has been calibrated on a microscopic scale and a macroscopic scale. Results of both are presented in table I. Generally, it can be concluded that the microscopic and macroscopic calibration give more or less the same value. Furthermore, the standard deviation of the value is relatively low, so one might conclude the estimates are accurate, and not dependent on the specific data set or optimization method. Sections V-B1 and V-B2 discuss the results of the macroscopic and microscopic calibration in more detail.

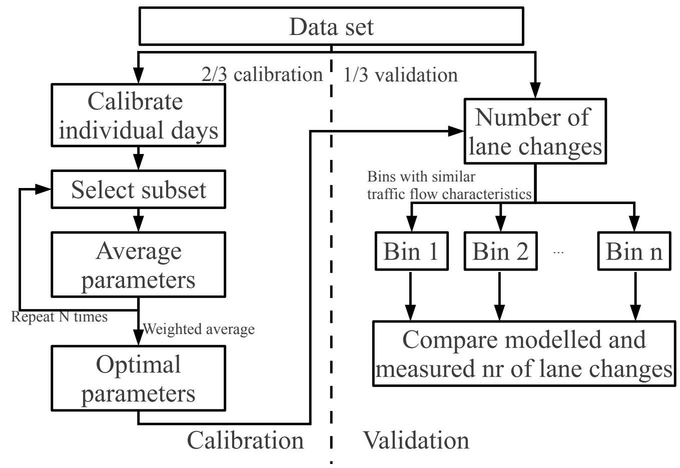


Figure 6: An overview of the steps in macroscopic calibration and validation

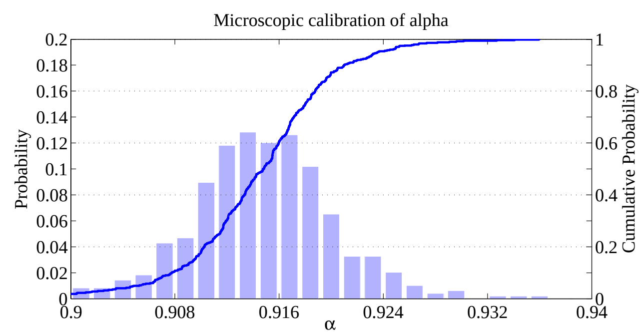


Figure 7: Calibration values for α

1) Microscopic calibration results: The data set was categorized per day, and changing subsets of the microscopic calibration results for each day were taken. Average results are shown in table I. Time between two considerations of lane changing, τ, varies from around 10s to 20s.

These values are in line with the expectations: one could expect the values to differ for different drivers, and all having about this order of magnitude. The fraction of the lane changes which can be described by the model, α, is relatively high, at approximately 90%. The critical time to collision is approximately 4 seconds. Also the critical gap for lane changing is approximately 4 seconds, which is well in line with a typical short headway of 2 seconds (a gap is divided into two headways after a lane change). The distribution of the parameter estimates is also available, for example for the parameter α as shown in figure 7.

2) Macroscopic calibration results: As explained in section V-A2, there is a changing subset used for a series of calibrations. For each series, τ and the part of the lane changing described by the model α are calibrated.

Comparing the values of the microscopic calibration, we find that α has almost exactly the same value, and the value for τ is at the upper bound of the values found in the microscopic

Table II: Validation results

Level	Calibration			Validation		
	$\mathcal{L}$	Nr	Norm.	$\mathcal{L}$	Nr	Norm.
Micro	9.7E4	2.1E6	4.6E-2	8.1E4	1.8E6	3.6E-2
Macro	2277	1559	1.46	1802	1214	1.48

data. It was expected that the width of the distribution of τ and α of would decrease since there is averaging over different drivers and vehicles. That means that extreme values are compensated and the average value has a narrower distribution.

VI. Validation

This section checks whether the model could be considered validated, and follows the three definitions mentioned in section II-A. Note that the model is validated using the 1/3 of the data which is hold back (holdout data), so the model can at best be validated for the specific site and conditions (UK, LC from the center to the median lane etc.).

A. Right value of the calibrated variables

Table I shows the calibration results for the microscopic and macroscopic model. The values are comparable. A time between two lane change considerations (τ) of approximately 15 seconds seems reasonable. Also the minimal (critical) gap (g_0) of 4 seconds in the other lane in order to change lanes is reasonable. Also that a drivers is motivated to change lanes once his time to collision becomes less than approximately 4 seconds (κ_0) is what one could expect. Moreover, the factor α indicates that the model can predict 92% (microscopic) or 93% (macroscopic) of the lane changes – which is impressive. Hence, the model can be called validated according to definition 1.

B. Comparable errors in the calibration and validation

To judge the validation according to definition 2, we need to consider the goodness of fit for the validation and the calibration. For that we use the log-likelihood as defined in equation 14 for microscopic model, or the log or the likelihood defined in equation 16 for the macroscopic model. Note that the log-likelihood scales with the number of observations. To normalise, we hence divide by the number of observations.

The results, given in table II show that the normalised likelihood for calibration and validation are similar, both for the microscopic model as for the macroscopic model separately. Hence, according to definition 2, the model is validated.

C. Validation using a physical variable

The third definition requires that a definition is given based on a physical variable, preferably related to the purpose of the future use of the model. Therefore, we cannot simply rely on the numbers produced in the calibration and validation process. Instead, a traffic interpretation needs to be given to the numbers, which is done in this section.

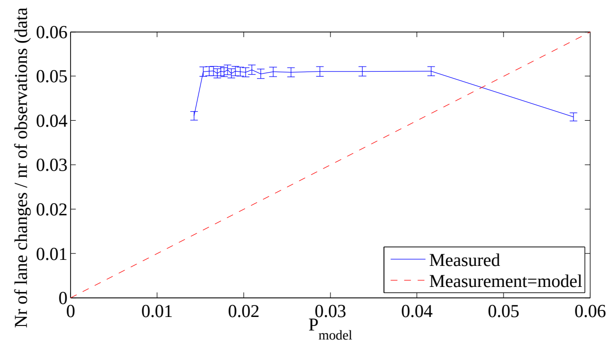


Figure 8: Microscopic validation results

1) Microscopic validation: For the validation, for each vehicle the probability of a lane change is calculated using equation 12. The essence of probabilistic validation is that all measurements from the validation set are now split up into bins for which the probability of a lane change is similar. We choose bin edges based on the percentile values for the lane change probability to ensure that each bin has the same number of observations.

For all vehicles, it is observed if they change lanes or not. Therefore, in each bin the average lane change rate, as well as the standard deviation thereof can be determined. This means that per bin, there is the predicted lane change rate as well as the observed lane change rate. The model is validated based on the way the modelled probabilities match the observed probabilities.

Figure 8 shows the validation results, which appear very bad. First of all, all modelled probabilities are always low ($< 10\%$), which means the model will never give a clear statement that a vehicle will change lane. In fact, in the data an almost constant lane change rate of approximately 5% is measured, independently of the predicted lane change rate over the range of predicted lane change rates of 1-5%. The model is clearly invalid on the microscopic scale. However, it cannot be seen in what conditions the model is wrong. To this end, a macroscopic validation is needed.

2) Macroscopic validation: In order to get an idea for the value of a log-likelihood, we convert this to an individual probability. In this conversion, we assume that the probabilities for predicting the right number of lane changes in an interval i (P_i) are equal for all intervals. Having n intervals, inverting equation 16 gives:

$$P_i = \sqrt[n]{L} = \sqrt[n]{\exp\mathcal{L}} = (\exp\mathcal{L})^{1/n} = \exp\mathcal{L}/n \qquad (17)$$

Using the concavity of the log function, it can be proven that relaxing the assumption of equal probabilities will increase the mean probability. Figure 10(b) shows how this average probability changes for variations in variables α or τ around the maximum.

Assuming equal probabilities, the probability that the model predicts the right number of lane changes is approximately 3.5% (using the values from table II). Considering that zero is the mode of the number of lane changes, one can propose a LC model predicting no LCs – which obviously is wrong. This would be right in 5.8% of the aggregation intervals. So in terms of how often the model predicts the right outcome, this model outperforms the model proposed in this paper, showing clearly that the model used here is not the best suited model to describe the process.

VII. Discussion

The model is calibrated and according to the first two definitions presented in section II-A, the model is validated. Nevertheless, the model does not predict the right results. This section discusses the reasons of this discrepancy. First, the measure of performance is addressed (section VII-A), and then calibration and validation issues are addressed (section VII-B). The same problems arise for macroscopic and microscopic analysis. Since the macroscopic description allows us to understand where the model makes errors, we will discuss the results only for the macroscopic analysis.

A. Using RMSE of number of lane changes

This section describes how the root mean square error of the number of lane changes can be used for an interpretation of the quality of the model. First, the methodology and results are presented, and then, in section VII-A2 it is shown how this can be used to improve the model.

1) Methodology and results: To do so, we consider bins with similar probability of changing lanes. Since this depends on the traffic density on lanes (and, via these, of the speeds on the lanes), we can create bins for a specified density in the origin lane (indicated k_i) and in the target lane (indicated k_j). This should give a binomial distribution of lane changes, given by equation 10. This distribution can be compared to the distribution in practice. If the model predicts a different number of lane changes than observed, or one finds that the dependency of the number of lane changes on the a traffic variable (e.g., lane density) is not correct, one must conclude the model is not valid.

Figure 9(a) shows the measured and modelled distribution for one combination of density in the origin lane and destination lane. It shows the model has an error in the order of 100% of the measured number of lane changes. The same test be done for the other combinations of density. To quantify the quality of in probabilistic terms, a statistical distance between the two distributions can be used – for instance a Kolmogorov-Smirnov test. Numerically, the results are presented as aggregate values in table III; this table will be further discussed in section VII-B.

2) Implications and use: improve the model: Whereas the microscopic validation only shows that the model is wrong, the macroscopic validation can help in improving the model. In particular, it can show for which traffic conditions there is an overestimation and for which conditions there is an underestimation of the number of lane changes. To visualise this, we keep the bins of a specified density in the origin lane and a density in the target lane. From this, we take the expected number of lane changes (for the model) or the mean number of lane changes (for the data).

Figure 9(b) shows the measured number of lane changes averaged per bin. This can be compared with the prediction for this average made with the model and the calibrated parameters, shown in figure 9(c). The data show that – counter-intuitively – the number of lane changes towards a particular lane increases with an increasing density in the target lane for a constant density in the origin lane. This finding has been observed at different sites [27]. Three traffic engineering explanations can be given (for a more detailed explanation, see [27]). (1) A void in the shoulder lane and a high density in the median lane means drivers are preparing to overtake a truck; if the driver does not change lanes, he might not be able to find a gap, so the higher the density in the target lane, the higher the urgency to change lanes. (2) Future traffic conditions are not included. (3) Lane changes towards a lane might be induced by lane changes from that lane (lane swapping), which increase with the density in the target lane.

This shows how the RMSE can show the weaknesses and can help finding better a model, namely by incorporating the above errors in the model. Reformulating an improved model is outside the scope of this paper.

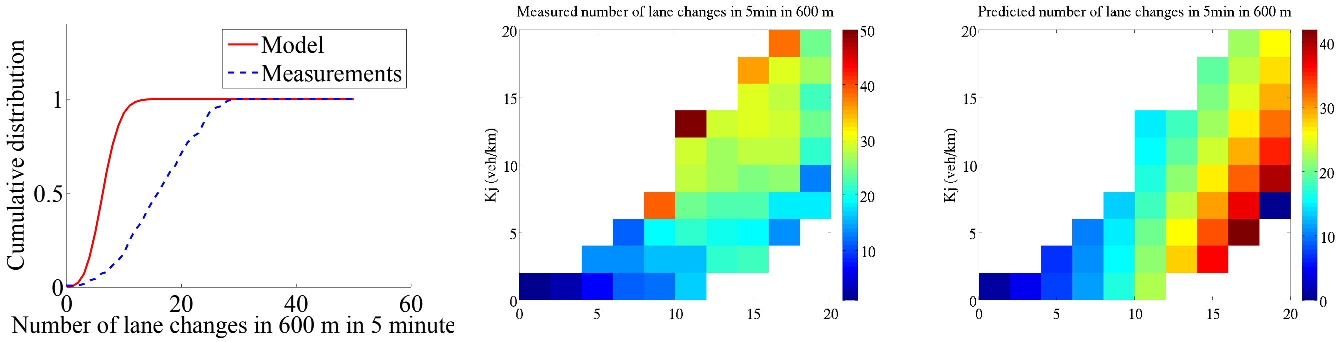
B. Calibration and validation issues

This section explores what technically could be other measures to look at rather than the model quality itself could be causes for the bad validation results. The section uses the RMSE as basis for the quality of the method, since this value is understandable.

1) Quality of the calibration: The most quantitative definition for validation (section II-A) is that a model is validated once the error in the validation set is comparable to the error in the calibration set. We thus calculate the RMSE for the calibration set (third and fifth column of table III) and have to compare them to the same values of the validation set (third and fifth column). These values are pairwise very similar, meaning the conditions for validation are met according to definition 2 in section II-A. Nonetheless, this is a pretty bad result.

2) Influence of the calibration and validation part: The calibration is performed of 2/3 of the total data set and the validation at the remaining 1/3 of the data set. There is a possibility that this was a particular separation of the data set. All days are assigned a group 1, 2 or 3 and the data from that day is assigned to the corresponding part. With the three parts, there are are three possibilities to make the division between calibration and validation sets. This analysis is called an *inverse cross-validation* in chapter 16 of [2]. All three analyses have been performed, and the errors are presented in the different lines of table III. These are very similar, showing that the separation in calibration and validation does not have an influence.

3) Variations per bin: We averaged the number of lane changes per aggregation period per bin with similar traffic conditions. To check the quality of a prediction model, one could compare the average number per bin with the average of the measurements in that bin. In this case, one compares figure 9(b) with figure 9(c), and compute the RMSE weighted for the number of observations. This is shown in table III in the third and fourth column. It shows that on average, the

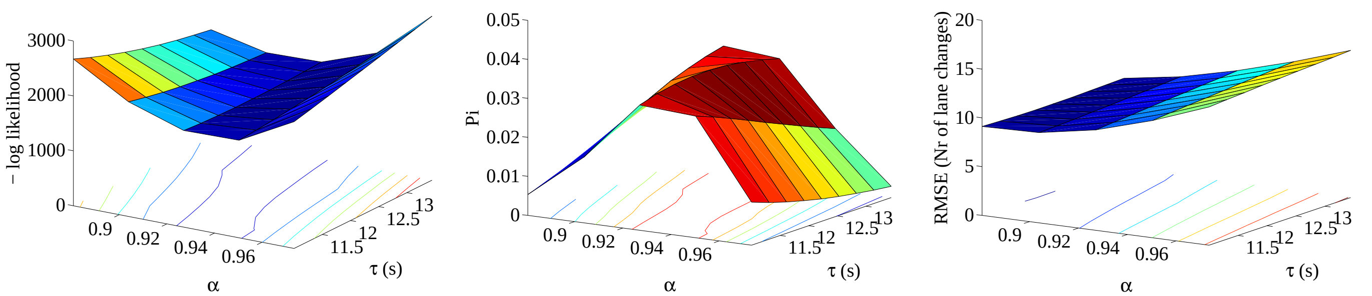


(a) Distribution of lane changes in bin 4< k_i <6 (density in the origin lane) veh/km/lane and 4< k_j <6 (density in the target lane) veh/km/lane

(b) Measured number of lane changes

(c) Predicted number of lane changes

Figure 9: Validation based on macroscopic data



(a) log-likelihood

(b) Probability

(c) Root mean square error of the number of lane changes compared to the expected number of lane changes

Figure 10: Quality of the fit as function of small variations in the parameters

Table III: The root mean square errors of the fits compared to the measured values expressed in lane changes in 5 min in 600 m

1	2	3	4	5	6
Set	Stdev	Errors in bin average		Individual errors	
		cal set	val set	cal set	val set
1	5.8	9.5	9.4	11.1	10.9
2	5.8	9.5	9.7	11.1	11.3
3	5.8	9.5	9.4	11.1	11.1

mean error is approximately 9 lane changes. This should be compared to a typical number of lane changes of 10-30.

Another measure of variation is the RMSE per bin: comparing the number of lane changes in an aggregation period with the predicted number of lane changes for these traffic conditions. In this case, the measurements are not averaged per bin. This must lead to a higher error, since all variations are still included. We find an RMSE of 11 lane changes (columns 4 and 5). This is twice the intrinsic variation of the number of lane changes per bin (standard deviation, see second column of table III), which is a mathematically lower limit to the RMSE (if and only if the values for each of the bins are equal). This indicates that the lane change model is not really accurate and the results – even for traffic conditions which are assumed equal – are not very good.

4) Sensitivity of the optimal solution: For the assessment of the quality of a fit, both the absolute error in the validation set as well as the sensitivity to the parameters are relevant [28]. Figure 10(a) shows the error function, being the log-likelihood. The optimum appears sharper because the likelihood is a product of different terms. A slight difference in each of the terms will appear magnified in the likelihood, and thus in the log-likelihood; in other words, the values are in an ordinal scale, but not in an interval scale.

One could also analyse how the RMSE changes if the parameters change. In fact, this is not the measure of performance used in the optimisation, so the RMSE of the averages is not necessarily minimum at the found parameter set. Figure 10(c) shows that it is indeed not minimum. It is a useful analysis nonetheless since it clearly shows that the minimum of the parameters is not really sharp, i.e. (1) even in the best fit, the RMSE is still high and (2) the RMSE does not sharply increase for a change in parameter values. In fact, the figure thus shows that the parameter values do not have a large impact on the quality of the predictions, hence the model calibration did not lead to a high quality model.

VIII. CONCLUSIONS

This paper illustrated the calibration and validation process for a probabilistic model for discretionary lane changes. It showed that it is possible to have the model calibrated, and validated in terms of having the same errors in the calibration set as in the validation set. Nevertheless, the quality of the model was poor, and the prediction in terms of behaviour of individual drivers or the number of lane changes was bad. In fact, the model was not able to capture the phenomena observed in traffic.

This paper thus shows that, contrary to current state of the art for discretionary probabilistic lane change models, it is best to use physically interpretable measures during validation. These measures can also help defining the minimum quality of the model. Describing the outcome of a validation as likelihood might capture the stochastic effects, but lacks this physical interpretation. Future research should show the quality of the prediction of other lane change models.

ACKNOWLEDGMENT

This research was sponsored by the Dutch the Netherlands Organisation for Scientific Research (NWO, projects: “how traffic jams start” and “there is plenty of room in the other lane”), as well as the COST Multitude action (STSM). The authors are grateful to use the dataset from IDRIS consultancy supplied by R. Eddie Wilson (University of Bristol).

REFERENCES

[1] S. Ahn and M. Cassidy, “Freeway traffic oscillations and vehicle lane-change maneuvers,” in *Proceedigs 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.

[2] M. Treiber and A. Kesting, *Traffic Flow Dynamics: Data, Models and Simulation*. Springer-Verlag Berlin Heidelberg, 2013.

[3] R. Dowling, A. Skabardonis, and A. Vassili, “Traffic analysis toobox – volume III: Guidelines for Applying Traffic Microsimulation Software,” Federal Highway Administration, Tech. Rep. FHWA-HRT-04-040, 2004.

[4] “www.multitude-project.org.”

[5] M. Brackstone, M. Montanino, W. Daamen, C. Buisson, and V. Punzo, “Use, calibration, and validation of traffic simulation models in practics: Results of a web-based survey,” in *Proceedings of the 91th Annual Meeting of the Transportation Research Board*, 2012.

[6] D. Michalaka, Y. Yin, and D. Hale, “Simulating high-occupancy-toll lane operations,” in *Proceedings of the 92nd Annual Meeting of the Transporatation Research Board*, 2013.

[7] A. Law and W. Kelton, *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York., 2000.

[8] X. Wu and H. Liu, “A shockwave profile model for traffic flow on congested urban roads,” *Transportation Research part B*, vol. 45, pp. 1768–1786, 2011.

[9] A. D. May, *Traffic Flow Fundamentals*. Englewood Cliffs: Prentice Hall, New Jersey, 1990.

[10] T. L. Paez, “Introduction to model validation,” *Proceedings of the 2009 IMAC*, 2009.

[11] F. Marczak, W. Daamen, and C. Buisson, “Key variables of merging behaviour: empirical comparison between two sites and assessment of gap acceptance theory.” *Transportation research Part C*, vol. 36, pp. 530–546, 2013.

[12] V. Knoop, C. Buisson, R. Wilson, and B. Van Arem, “Number of lane chanes determined by splashover effects in loop detector counts,” *IEEE Transactions on Intelligent Transport Systems*, vol. 13, pp. 1525–1534, 2012.

[13] 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.

[14] T. Toledo, H. N. Koutsopoulos, and M. E. Ben-Akiva, “Modeling integrated lane-changing behavior,” *Transportation Research Records, Journal of the Transportation Research Board*, vol. 1857, pp. 30–38, 2003.

[15] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, “Integrated driving behavior modeling,” *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 2, pp. 96 – 112, 2007.

[16] ——, “Estimation of an integrated driving behavior model,” *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 4, pp. 365 – 380, 2009.

[17] A. Kesting, M. Treiber, and D. Helbing, “General lane-changing model mobil for car-following models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, pp. 86–94, 2007.

[18] W. J. Schakel, V. Knoop, and B. Van Arem, “Lmrs: Integrated lane change model with relaxation and synchronization,” *Transportation Research Record*, vol. 2316, pp. 47–57, 2012.

[19] C. F. Daganzo, “A behavioral theory of multi-lane traffic flow. part i: Long homogeneous freeway sections,” *Transportation Research Part B: Methodological*, vol. 36, no. 2, pp. 131 – 158, 2002.

[20] A. D. Patire and M. J. Cassidy, “Lane changing patterns of bane and benefit: Observations of an uphill expressway,” *Transportation Research Part B*, vol. 45, pp. 656 – 666, 2011.

[21] J. A. Laval and L. Leclercq, “Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model,” *Transportation Research Part B: Methodological*, vol. 42, no. 6, pp. 511–522, 2008.

[22] W. L. Jin, “Macroscopic characteristics of lane-changing traffic,” in *89th Annual Meeting of the Transportation Research Board*, 2010.

[23] ——, “A kinematic wave theory of lane-changing traffic flow,” *Transportation Research Part B: Methodological*, vol. 44, no. 8-9, pp. 1001 – 1021, 2010.

[24] British Highway Agency, http://www.highways.gov.uk/knowledge/1334.aspx, 2010.

[25] IDRIS Consulting, http://www.idris-technology.com/, 2010.

[26] A. Duret, J. Bouffier, and C. Buisson, “Onset of congestion from low-speed merging maneuvers within free-flow traffic stream,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2188, no. 1, pp. 96–107, 2010.

[27] V. L. Knoop, S. P. Hoogendoorn, S. Yasuhiro, and C. Buisson, “Quantifying the number of lane changes in traffic: Empirical analysis,” *Tranportation Research Records, Journal of the Transportation Research Board*, vol. 2278, pp. 31–41, 2012.

[28] V. L. Knoop, H. J. Van Zuylen, and S. P. Hoogendoorn, *Proceeding of the International Symposium of Transportation and Traffic Theory*. Springer, New York, 2009, ch. Microscopic Traffic Behavior near Accidents.

Victor L. Knoop received his Masters 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 assistant professor with the Delft University of Technology. Traffic flow phenomena are his main research interest, as well as the interaction between microscopic and macroscopic levels of traffic description. He is a member of Traffic Flow Theory and Characteristics Committee of the Transportation Research Board.

Christine Buisson Traffic flow researcher since 1993, she is now Senior Researcher of IFSTTAR (formerly INRETS). She also teaches at ENTPE, member of Université de Lyon. 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 a member of Traffic Flow Theory and Characteristics Committee of the Transportation Research Board.