The Value of Calibration and Validation of Probabilistic Discretionary Lane-Change Models

V. L. Knoop PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
+31 15 278 8413
v.l.knoop@tudelft.nl

C. Buisson PhD
Senior Researcher
Université de Lyon, F-69622, Lyon, France
IFSTTAR, LICIT, F-69675 Bron
ENTPE, LICIT, F-69518 Vaux en Velin
christine.buisson@ifsttar.fr

October 24, 2013

Word count:

nr of words in abstract	149
nr of words (including abstract and references)	5750
nr of figures& tables	7* 250 = 1750
total	7500

Submitted to the 93th Annual Meeting of the Transportation Research Board

ABSTRACT

Lane changes are important in traffic flow operations. They cause differences in flow over lanes and they may determine the onset of congestion. Therefore, it is relevant to have a lane change model for which the predictive power is confirmed. This article calibrates and validates a probabilistic discretionary lane change model, both microscopically and macroscopically. This is done using traffic flow variables related to lane changing, which was possible due to a large data set where lane changes have been recorded. It is shown the model can be claimed calibrated with accurate parameter estimates using a maximum likelihood method following the current state of the practice. However, careful interpretation of the validation shows that the resulting model is not at all representing reality. It therefore is concluded that lane change models must be validated before they can be used, and the quality should be assessed by physically sound measures.

1 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 needs 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 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. Section 2.1 will show some definitions used in literature for validation. The main question discussed in this paper is whether a validated model is necessarily good. A subquestion discussed in that process is how can one define the quality of a stochastic model.

This paper will show that even a model which is calibrated and validated – and considered valid according to the used methods – 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 takes 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 to get an understanding of the quality of a stochastic model.

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 3) and an example lane change model which we will use in this paper (section 4). The calibration and validation are discussed in section 5 and 6 respectively. The paper ends with a discussion on the topic of validation (section 7) and conclusions in section 8.

2 LITERATURE REVIEW

This part first discusses the process of calibration and validation in general and then, section 2.2 discusses discretionary LC (lane change) models and calibration and validation thereof.

2.1 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 scientific 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. The collected data have to be divided 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 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 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.

We will hereafter illustrate that satisfy requirements in definition 1 and 2 is not sufficient to also satisfy requirement 3 or 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.

2.2 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.

(12) presents 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 integrated model (13) a LC decision depends on the possible longitudinal accelerations according to the longitudinal model. They use a maximum likelihood for the chosen path as objective function in calibration (14). No validation is carried out.

In the *Intelligent Driver Model* (15) a LC is valued according to 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. (16) 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.

(17) present another way of LC 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 LCs, where vehicles take two places (at two lanes) in the traffic stream, is shown by (18), who give a microscopic interpretation of the process. In (19) 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.

This overview shows that there is no standard in calibration and validation of discretionary LC models. For probabilistic models, the method of maximizing the log-likelihood is used most often, and mathematically the most sound, which is why we adopt this method in this paper. Moreover, only some models are calibrated and just two models are validated. This paper will show that even after calibration a validation is required. It might even happen that a calibrated model is still performing badly.

3 DATA COLLECTION AND TREATMENT

This section discusses the data: their availability and treatment (section 3.1), as well as the basic traffic characteristics (section 3.2).

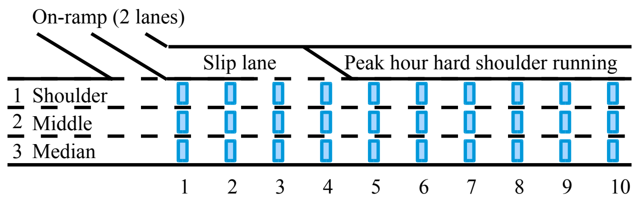
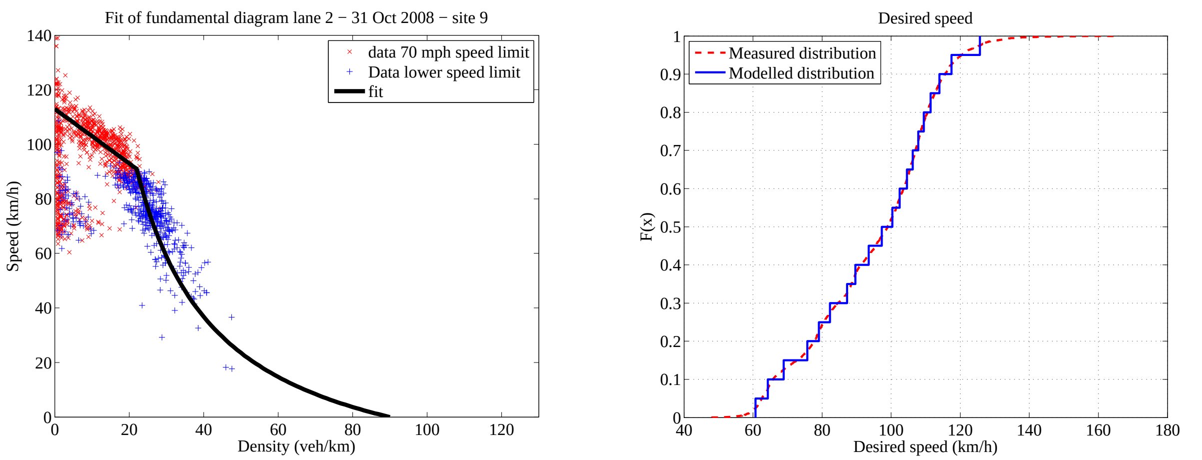


FIGURE 1 The layout of the motorway and the detectors. The detectors are placed at approximately 100 meter distances. Note that in the United Kingdom, people drive at the left hand side of the road. That means, left to right in this figure. The next off ramp is 4.5 kms further downstream.



(a) The fundamental diagram in the speed-density plane for the middle lane for a one-minute aggregation time

(b) The distribution of the desired speed and the modelled distribution

FIGURE 2 Characteristics of the data

3.1 Data availability

The data used are individual loop data from the M42 motorway in the UK near Birmingham (20). The loops are placed approximately 100 meters apart, and for all vehicles the passing time, the lane, the speed and their 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 and only LCs from the middle to the median lanes are considered.

The detailed data per vehicle allows re-identification from one site to the next. If a vehicle is re-identified at the further downstream site at another lane, a LC has taken place. It is assumed that no driver makes a LC from his lane and back to his original lane within 100 meters. The re-identification works very good in uncongested conditions. In fact, a re-identification rate of more than 999 out of 1000 vehicles is obtained (personal communication 21, but the re-identification breaks completely in congestion. 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. During two month, individual vehicle data are collected. Aggregated in periods of 5 minutes, we have – excluding the periods in which there is a dynamic speed limit or congestion – a set of 5020 time periods.

3.2 Traffic characteristics

Figure 2a 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.

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. Under that assumption, the distribution of speeds of vehicles which match the conditions is the free speed distribution. 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 2b also shows the distribution and the approximation.

4 MODELLING

In this paper we present a simple model to show the effects of calibration and validation. Earlier work showed that (22) – counter-intuitively – the denser the traffic in the target lane, the more LC towards that lane. There is no model which predicts so. The goal of this paper is not to test existing models. Instead, the aim of this paper is to present the relevance of calibration. Hence we propose a simple model, which can be used in a microscopic and macroscopic form, and is based on principles which seems plausible.

The probability to change lanes is seen here as the product of three elements; these model steps are more or less aligned with the fundamentals laid out by (23)

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 LC rate by multiplying them. We realize that the model only represents a part of the motivation of changing lanes (α), and there is another part unexplained. Later we will find that the unexplained part is approximately 10%. The combined model reads:

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

The remainder of the section quantifies the elements and shows how this is combined into one microscopic (section 4.4) or macroscopic (section 4.5) model.

4.1 Interest in improving speed

First of all, only drivers who drive at lower speeds than their desired speed are likely to change lanes voluntarily. We quantify this by the "virtual time to collision", the ratio between the current distance headway h and the difference between the desired speed v_d and the leader's speed v_l:

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

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.

For each combination of headway, desired speed, and speed of the leader a virtual time to collision can be calculated. If this virtual time to collision is smaller than a threshold value (κ_0), a driver is assumed to have interest in improving speeds. Hence, we model:

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

If all these all 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, and hence a distribution for the desired speed has to be assumed for every car. The fact that this is a *distribution* means that ϕ is a stochastic variable.

4.2 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 between 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 modelled a probability of 1 to change lanes in case the speed in the adjacent lane is double the speed in the current lane.

4.3 Possibility 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 LC. 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 (24). The minimum gap, g_{min}, is modelled as a minimum gap in 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 3 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 gap in the adjacent lane (g_{nf}) is larger than the minimum gap:

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

Note that for an individual vehicle this is fully observable, and deterministic. However, the minimal gap g_0 is a parameter which should be calibrated.

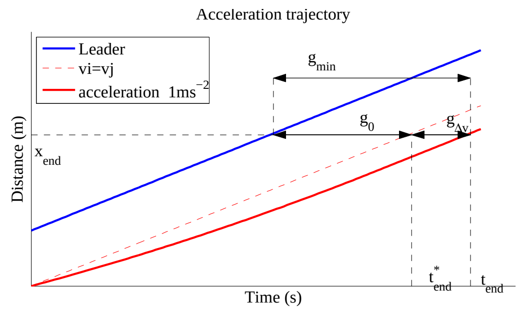


FIGURE 3 Illustation of the minimum gap

4.4 Combining towards lane change rates

Mathematically, Λ indicates a probability to change lanes *per consideration*, which, by definition, has a duration τ. 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 only one LC in maximum can take place during this period (later on we will see that τ is in the order of 15 seconds.) Substituting ϕ (equation 3), χ (4), and ϕ (7) in equation 8 gives:

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

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

4.5 Macroscopic description

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

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

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. Parameter τ denotes the time between two successive decision moments to change lanes. Furthermore, the number of LCs 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.

5 CALIBRATION

In this section we discuss the calibration. We first present the methods, for microscopic and macroscopic calibration, and then the results (section 5.2).

5.1 Calibration methodology

Both the microscopic and the macroscopic model require in principle all parameters, which can be optimized per vehicle. However, for the microscopic model we use in the test case, some variables are intrinsically microscopic. Both 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 LC considerations, τ, and the part of lane changing which can be explained by the model, α.

5.1.1 Microscopic calibration

This section presents how the probabilistic LC model is calibrated and validated. The steps are shown graphically in Figure 4a. Note that the base element is *the probability of a LC*, derived from equation 9:

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

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

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

- κ_{virt}, the time to collision if the following vehicle drove its desired speed,
- v_{new}, the speed of the new leader,
- v_{c}, the current speed,
- g_{nf}, time gap to the new follower,
- τ, the time between two LC considerations.

For each of the vehicles it is now possible to express the probability to find a LC according to the model, only dependent on the parameters τ, α, κ_0, and g_0.

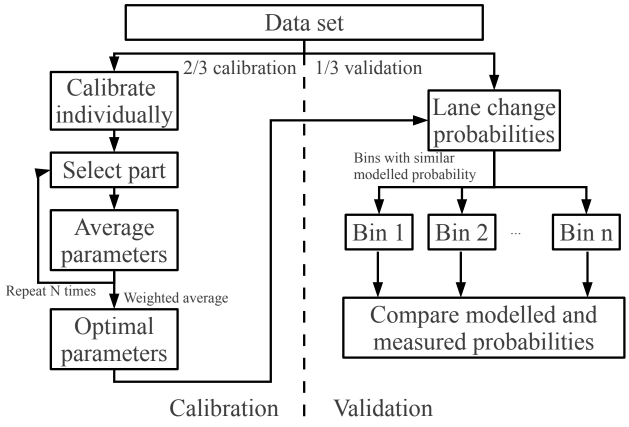
Note that we aim to validate the model for the same section and conditions. We hence split the data randomly into two parts: a calibration part (2/3) and a validation part which we keep aside for validation purposes later on. This is an example of *holdout validation* (2). For the calibration data, we 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 LC model is probabilistic, it is considered optimal if the likelihood that the model predicts the observed value is maximum. This likelihood is expressed as:

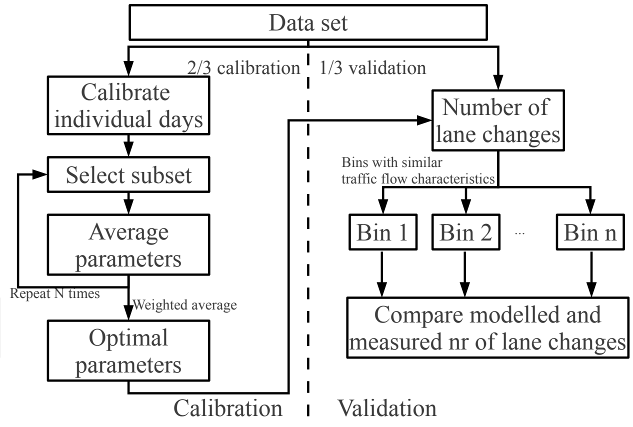
$$L(\kappa_{\text{virt}}, v_{\text{new}}, v_{\text{c}}, g_{\text{nf}}) = \prod_{c\in\mathcal{G}} P_{\text{change}}, c \times \prod_{s\in\mathcal{S}} P_{\text{no change}}, s \tag{13}$$

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

$$\begin{aligned} \mathcal{L}(\kappa_{\text{virt}}, v_{\text{new}}, v_{\text{c}}, g_{\text{nf}}) &= \log(L) = \log\left(\prod_{c\in\mathcal{G}} P_{\text{change}}, c \times \prod_{s\in\mathcal{S}} P_{\text{no change}}, s\right) \\ &= \sum_{c\in\mathcal{G}} \log\left(P_{\text{change}}, c\right) + \sum_{s\in\mathcal{S}} \log\left(P_{\text{no change}}, s\right) \end{aligned} \tag{14}$$



(a) Microscopic



(b) Macroscopic

FIGURE 4 An overview of the steps in calibration and validation

Now, parameters values τ^* α^*, κ_0^*, and ${g_0}^*$ are defined such that they maximize the likelihood:

$$\{\tau^*, \alpha^*, \kappa_0^*, {g_0}^*\} = \underset{\Delta T, \alpha, \kappa_0, g_0}{\arg\max} \left(\mathcal{L}(\kappa_{\text{virt}}, v_{\text{new}}, v_{\text{c}}, g_{\text{nf}})\right) \tag{15}$$

To get not only the parameter values but also the variability, we perform this optimization procedure repeatedly for different days, all from the calibration set. The set of parameters τ^*, α^*, κ_0^* and ${g_0}^*$ satisfying equation 15 are determined, using *Matlab*'s function *fminsearch*. Now, with probability 2/3 a day is included in the calibration for an iteration. This will give an average value for the parameter in that iteration. This is repeated 500 times (N in figure 4a), 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).

5.1.2 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 LCs as in the microscopic description. For this application, the model predicting the number of LCs 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. 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.

A graphical view of the calibration and validation process is given by Figure 4b. In analogy with the procedure described for the microscopic calibration, we define a likelihood function indicating how good the probabilistic model performs.

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

Calibration consists of finding these values for the parameters which maximize equation 16. In our case, we use *fminsearch* in *Matlab* to find τ and α.

TABLE 1 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		

5.2 Calibration results

The model has been calibrated on a microscopic scale and a macroscopic scale. Results of both are presented in table 1. 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 for different subset of the calibration set is relatively low, so one might conclude the estimates are accurate, and not dependent on the specific data set or optimization method.

5.2.1 Microscopic calibration results

The data set was classed per day, and changing subsets of the microscopic calibration results for each day were taken. Average results are shown in Table 1. 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 LCs 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 LC).

5.2.2 Macroscopic calibration results

As explained in section 5.1.2, 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 data. It was expected that the distribution of τ and α of would smaller since there is averaging over different drivers and vehicles. That means that extreme values are compensated and the average value has a narrower distribution.

6 VALIDATION

This section describes first the microscopic validation and then the macroscopic validation.

6.1 Microscopic validation

6.1.1 Microscopic validation methodology

For the validation, for each vehicle the probability of a LC 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 LC is similar. We choose bin edges based on the percentile values for the LC probability to ensure that each bin has the same number of observations.

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

6.1.2 Microscopic validation results

Figure 5a 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 LC rate of approximately 5% is measured, independently of the predicted lane change rate over the range of predicted LC 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.

6.2 Macroscopic validation

6.2.1 Probabilistic macroscopic validation

Validation of a LC model on macroscopic data is difficult and has the risk that the resulting macroscopic processes are in fact the result of a different underlying microscopic process than modelled. The best one can do is showing that the number of LCs is correct for certain traffic conditions.

For the macroscopic validation we use the data, hold back from the total set at the calibration, being 1474 time periods of 5 minutes.

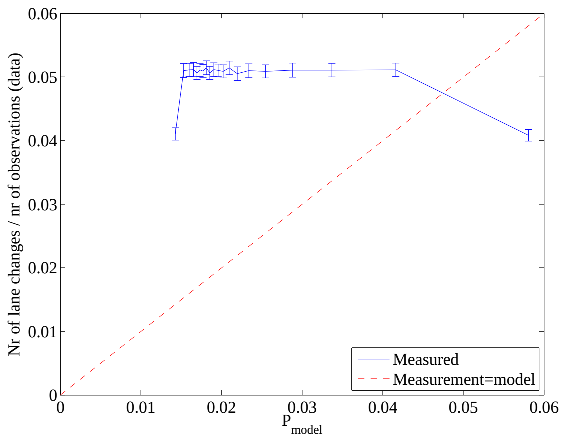
To do so, we consider bins with similar probability of changing lanes. Since this depends on the traffic density on lanes (and, via these, of the speeds on the lanes), we can create bins for a specified density in the origin lane (indicated k_i) and in the target lane (indicated k_j). This should give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

Figure 5b 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 LCs. The same test can 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.

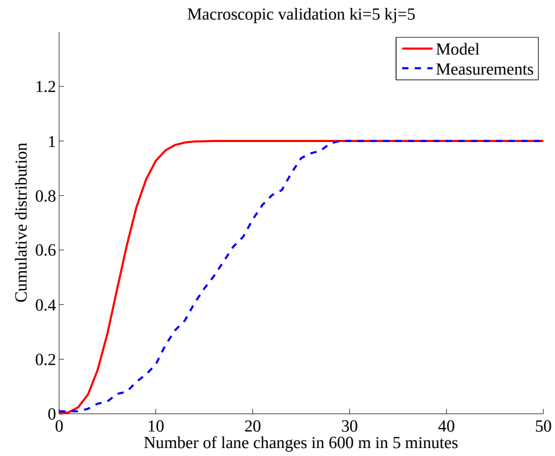
6.2.2 Use of validation to improve the model

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 LCs. To visualize 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 LCs (for the model) or the mean number of LCs (for the data).

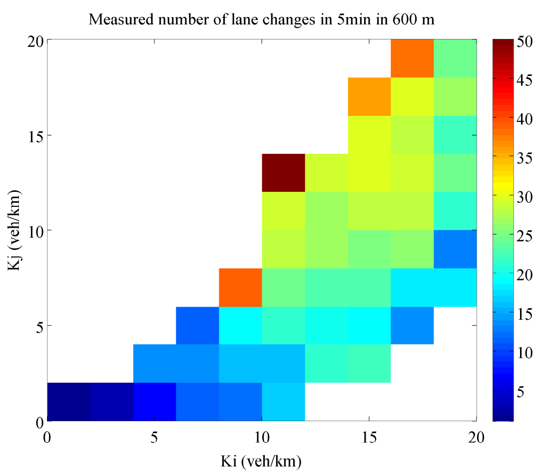
Figure 5c shows the measured number of LCs averaged per bin. This can be compared with the prediction for this average made with the model and the calibrated parameters, shown in Figure 5d. This can be used to improve the model: apparently, and counter-intuitively, an increase of the density in the target lane will lead to an increase of the number of LCs towards that lane (22). Improving the LC model is outside the scope of this paper; we do want to show, though, that by performing this validation one knows which influencing factors should be more (or less) included in the model.



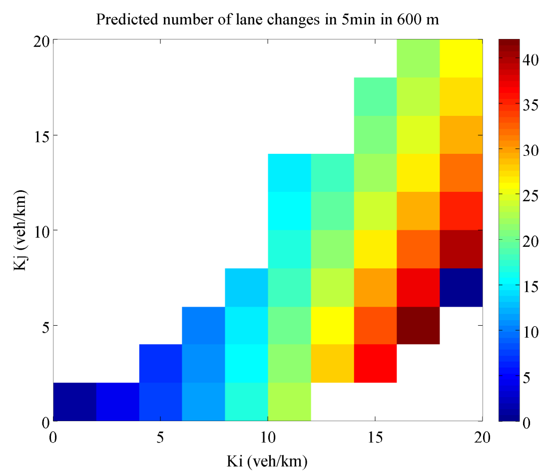
(a) Microscopic validation results



(b) Distribution of LCs 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



(c) Measured number of LCs



(d) Predicted number of LCs by the calibrated macroscopic model

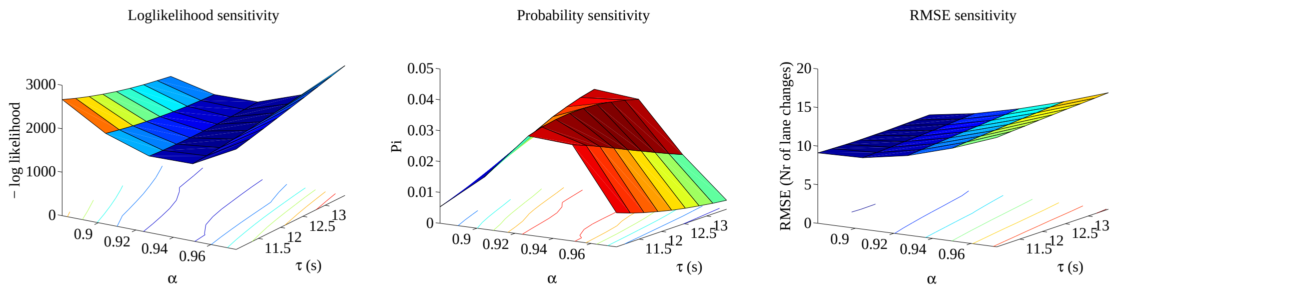
FIGURE 5 Validation results

7 DISCUSSION

The results show that the lane change model can be calibrated, but the model does not predict the right results. Note however, that according to the definitions in section 2.1 the model is validated. The same problems arise for the 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 first possibility for the bad validation results could be the split between the calibration and the validation part was unfortunate. We split the data in three parts, and repeated all analyses three times, for each of the three data parts being the validation set (and the other two being the calibration set). The results were almost identical, so this is not the case.

Internal variations per bin give an idea of the accuracy of the fit. The proposed model has an RMSE of 11 (both for the validation and the calibration set). We can separate this error to the intrinsic error, i.e. the variation per bin, and the model error. The standard deviation of the number of lane changes in each bin



(a) loglikelihood

(b) Probability

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

FIGURE 6 Quality of the fit as function of small variations in the parameters

is 5.8 lane changes (on typically 10-30 lane changes). If the model was perfect, the root mean square error (RMSE) of lane changes per bin would have this value. The RMSE of the averages per bin is approximately 9.5 lane change – this shows how far off the model is in predicting. This value is similar for the calibration and validation set, showing that the calibration is well done, and the model is not over-fitted.

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 is relevant (25) Figure 6a shows the error function, being the the log-likelihood. The maximum 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.

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:

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

Using the concavity of the log function, it can be proven that relaxing the assumption of equal probabilities will increase the mean probability. Figure 6b 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%, or 1 out of 28 cases. A zero-order model can be defined which randomly chooses a number of lane changes between 0 and the maximum observed number of lane changes in a time interval (51). This model would statistically be right in 1 out of 51 cases. The mode of the observed number of lane changes is 0. A model predicting no lane changes at all (which would obviously be wrong) would be right in 5.8% of the aggregation intervals, i.e. 1 out of 17 cases. So in terms of likelihood, this model outperforms the model proposed in this paper.

One could also analyse how the RMSE changes if the parameters change. In fact, this is not the measure of performance used in the optimization, so the RMSE of the averages is not necessarily minimum at the found parameter set. Figure 6c shows that it is indeed not minimum. Only this measure of performance 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 value. 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.

8 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 the 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, there is a need to follow a purposeful definition for validation. Describing the outcome of a validation as likelihood might capture the stochastic effects, but lacks physical interpretation. For practical applications the qualities of models should be prescribed in quality in terms of quantities which have a physical interpretation, and by which one can judge the performance of the model. Future research should show the quality of the prediction of other lane change models.

Acknowledgement This research was sponsored by the Dutch the Netherlands Organisation for Scientific Research (NWO, project: there is plenty of room in the other lane). The authors are grateful to use the dataset from IDRIS consultancy supplied by R. Eddie Wilson (University of Bristol).

REFERENCES

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

[3] Dowling, R., A. Skabardonis, and A. Vassili, *Traffic analysis toobox – volume III: Guidelines for Applying Traffic Microsimulation Software*. Federal Highway Administration, 2004.

[4] *www.multitude-project.org*, ????

[5] Brackstone, M., 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] Michalaka, D., Y. Yin, and D. Hale, Simulating High-Occupancy-Toll Lane Operations. In *Proceedings of the 92nd Annual Meeting of the Transporatation Reserach Board*, 2013.

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

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

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

[10] Paez, T. L., Introduction to Model Validation. *Proceedings of the 2009 IMAC*, 2009.

[11] Marczak, F., 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. Accepted for publication, 2013.

[12] Choudhury, C., 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*, Elsevier, London, 2007.

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

[14] Toledo, T., H. N. Koutsopoulos, and M. Ben-Akiva, Estimation of an integrated driving behavior

model. *Transportation Research Part C: Emerging Technologies*, Vol. 17, No. 4, 2009, pp. 365 – 380.

[15] Kesting, A., M. Treiber, and D. Helbing, General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record: Journal of the Transportation ResearchBoard, No. 1999*, 2007, pp. 86–94.

[16] Schakel, W. J., V. Knoop, and B. Van Arem, LMRS: Integrated Lane Change Model with Relaxation and Synchronization. *Transportation Research Record*, Vol. 2316, 2012, pp. 47–57.

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

[18] Jin, W. L., Macroscopic Characteristics of Lane-Changing Traffic. In *89th Annual Meeting of the Transportation Research Board*, 2010.

[19] Jin, W. L., A kinematic wave theory of lane-changing traffic flow. *Transportation Research Part B: Methodological*, Vol. 44, No. 8-9, 2010, pp. 1001 – 1021.

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

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

[22] Knoop, V. L., 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, 2012, pp. 31–41.

[23] Gipps, G. P., A Model for the Structure of Lane-Changing Descisions. *Tranportation Research Part B: Methodological*, Vol. 20, 1986, pp. 403–414.

[24] Duret, A., J. Bouffier, and C. Buisson, Onset of Congestion Due to Low-Speed Merging Maneuvers Within a Free-Flow Traffic Stream: Analytical Solution. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. Accepted for publication, 2010.

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