

Document Version

Final published version

Licence

CC BY

Citation (APA)

Hofbaur, F., Calvert, S., Fellendorf, M., & van Lint, H. (2026). Analysis of two rule-based lane change models utilising naturalistic vehicle trajectory data. *Transportmetrica B*, 14(1), Article 2718468.
<https://doi.org/10.1080/21680566.2026.2718468>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

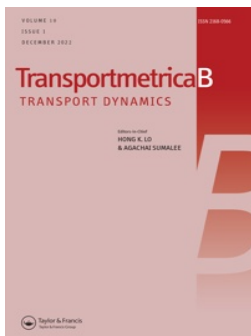
In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Analysis of two rule-based lane change models utilising naturalistic vehicle trajectory data

Felix Hofbaur, Simeon Calvert, Martin Fellendorf & Hans van Lint

To cite this article: Felix Hofbaur, Simeon Calvert, Martin Fellendorf & Hans van Lint (2026) Analysis of two rule-based lane change models utilising naturalistic vehicle trajectory data, Transportmetrica B: Transport Dynamics, 14:1, 2718468, DOI: [10.1080/21680566.2026.2718468](https://doi.org/10.1080/21680566.2026.2718468)

To link to this article: <https://doi.org/10.1080/21680566.2026.2718468>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 21 Aug 2026.



Submit your article to this journal [↗](#)



Article views: 257

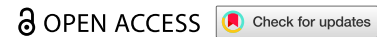


View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



Analysis of two rule-based lane change models utilising naturalistic vehicle trajectory data

Felix Hofbaur^a , Simeon Calvert^b, Martin Fellendorf^a and Hans van Lint^b

^aGraz University of Technology, Institute of Highway Engineering and Transport Planning, Graz, Austria; ^bDelft University of Technology, Department of Transport and Planning, Delft, the Netherlands

ABSTRACT

Lane change modelling is a key challenge in driver behaviour research, as lane-changing involves complex interactions with surrounding vehicles and significantly impacts traffic flow. Despite advancements in lane change modelling, calibration is often limited to macroscopic indicators, limiting microscopic performance assessment. Yet, micro-level accuracy is crucial, in particular for simulating mixed traffic and setting up realistic testing environments for automated driving functions. Hence, this study evaluates the micro-level performance of existing lane change models (*LMRS* and *MOBIL*) by utilising naturalistic trajectory data. Additionally, comprehensive data processing aspects are addressed, and trajectory-based model parameter optimisation is conducted, enabling manoeuvre-level performance evaluation. The results reveal that *MOBIL* struggles to capture anticipatory human driver behaviour. This issue arises because *MOBIL* perceives close but faster vehicles in the target lane as obstacles, whereas in reality, drivers often do not treat them as such. *LMRS*, on the other hand, shows difficulties with gap acceptance.

ARTICLE HISTORY

Received 7 May 2025
Accepted 18 July 2026

KEYWORDS

Human lane change behaviour; lane change models; model parameter optimisation; microscopic model evaluation; trajectory-based analyses

1. Introduction

Human drivers often exhibit anticipatory behaviour, adjusting their actions based on the expected movements of surrounding vehicles (Chen et al. 2023; Treiber, Kesting, and Helbing 2006). Others behave more myopically, which may appear less rational and can lead to unpredictable and sometimes risky driving. Especially during lane change manoeuvres, where multiple surrounding vehicles may be affected, myopic behaviour can lead to traffic disturbances and safety-critical situations (Ahn and Cassidy 2007; Zheng, Ahn, and Monsere 2010). The heterogeneity of driving styles, particularly the contrast between anticipatory and myopic behaviour, poses substantial challenges to accurately modelling human lane-changing. Current models often struggle to capture those phenomena. Capturing anticipatory behaviour requires accounting for subtle interactions with multiple vehicles, while modelling myopic behaviour demands the ability to reflect occasional irrationality.

In traffic flow simulation (TFS), lane-change models govern lateral behaviour, while car-following models control longitudinal motion. Early research on driver behaviour primarily focused on car-following (Brackstone and McDonald 1999; Brackstone, Sultan, and McDonald 2002), while lane-change modelling received less attention, partly because of the lack of suitable data (Brackstone, McDonald, and Wu 1998). Cross-sectional data proved inadequate for analysing lane-change decision-making because of the complex spatiotemporal dependencies involved (Kesting, Treiber, and Helbing 2007). Over the years, various models for lane change decision-making have been proposed, including but not limited to rule- and incentive-based models (Gipps 1986; Kesting, Treiber, and Helbing 2007; Schakel, Knoop, and van Arem 2012), utility theory-based models (Ahmed et al. 1996; Choudhury et al. 2007; Toledo, Koutsopoulos, and Ben-Akiva 2003), and models based on game theory (Ali et al. 2021; Talebpour, Mahmassani, and Hamdar 2015). In recent years, many data-driven models have emerged for the analysis of lane-changing behaviour (Wang et al. 2026), with a particular focus on lane-change intention

CONTACT Felix Hofbaur  felix.hofbaur@tugraz.at  Graz University of Technology, Institute of Highway Engineering and Transport Planning, Graz, Austria

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

prediction (Gao et al. 2022; Xie et al. 2019; Zhang et al. 2026). This increase in lane-change research has been partly driven by the availability of trajectory datasets, enabling more accurate analyses (2020; Li et al. 2025; Ma and Li 2023).

While numerous modelling approaches for lane-change decision-making exist, rule- and incentive-based models remain the most widely implemented in TFS environments. Of these, the *MOBIL* (*Minimising overall braking deceleration induced by lane changes*) (Kesting, Treiber, and Helbing 2007) and the *LMRS* (*Lane change model with relaxation and synchronisation*) (Schakel, Knoop, and van Arem 2012) are among the most frequently referenced and applied models in academic research. Behavioural models in general aim to capture micro-level driver behaviour as a basis for replicating overall traffic flow characteristics. However, calibration and performance evaluation are often limited to meso- and macroscopic indicators (e.g. the number of executed lane changes under certain traffic flow conditions). Depending on the application, these metrics may be insufficient and can lead to unsatisfactory results. This issue is particularly relevant, as these models, which were originally developed to represent human driver behaviour, are increasingly used in automated vehicle (AV) contexts. For example, the *LMRS* has been applied to model AV driving behaviour (Calvert, Schakel, and van Lint 2017), and the *MOBIL* is integrated into the widely used *highway-env* environment (Leurent 2018) for simulating surrounding traffic in AV controller development. In such applications, faithfully reproducing micro-level decision-making is often more important than matching aggregate traffic flow characteristics. More generally, Zheng (2014) emphasises that macroscopic traffic indicators are inadequate for evaluating microscopic models and highlights that the logic of the *MOBIL* model has yet to be empirically justified. Consequently, empirical evidence on how well existing lane-change models used in TFS capture observed microscopic driving behaviour remains limited.

In addition, Ali, Zheng, and Bliemer (2023) found that trajectory-based lane-change model evaluation has received inadequate attention, except for some probabilistic models. While they addressed data processing issues and examined model sensitivity to inappropriate data selection, they did not benchmark or evaluate model functionality – a task requiring carefully designed calibration. Building on this, we note that beyond the lack of empirical studies on microscopic performance, there is also a methodological gap in trajectory-based parameter optimisation for lane-change models. To the best of our knowledge, no systematic study has assessed and compared multiple optimisation methods for such models, although similar research exists for car-following models (Ciuffo, Punzo, and Montanino 2012; Punzo, Ciuffo, and Montanino 2012).

This paper addresses these gaps by presenting a comprehensive microscopic evaluation of two widely used rule-based lane-change models, *MOBIL* and *LMRS*, both implemented with the intelligent driver model (*IDM*), to also assess how longitudinal dynamics influence lane-change decision-making. We introduce new methods to process vehicle trajectory data and benchmark conceptually distinct optimisation approaches – Stochastic Gradient Descent (SGD), Genetic Algorithm (GA), and Bayesian Optimisation (BO) – for lane-change model parameter optimisation. This study uses the highD dataset (Krajewski et al. 2018), focusing on discretionary lane changes on open-road highways. In addition to quantitative performance metrics, we conduct in-depth analyses of correctly and incorrectly captured manoeuvres, offering insights into the strengths and limitations of the evaluated models and contributing to a deeper understanding of human driver behaviour.

The rest of this paper is organised as follows: The next Section 2 reviews related work regarding lane-change modelling, model calibration and parameter optimisation. We then present the methods (see Section 3). There, the utilised data is explained, the conducted data processing steps and the lane-change models, as well as the investigated optimisation algorithms, are described. This is followed by the results in Section 4. Finally, the findings are discussed in Section 5, followed by a conclusion (see Section 6) that summarises the work and outlines future research directions.

2. Related work

In this section, we briefly discuss previous work on model performance evaluation of existing lane-change models. Additionally, the literature addressing model calibration is discussed, focusing on but not limited to human lane-change behaviour and optimisation of lane-change model parameters. For a comprehensive review regarding modelling human lane-change behaviour and future research directions, the reader is referred to Rahman et al. (2013) and Zheng (2014).

In general, existing lane-change models have often been calibrated and evaluated using mesoscopic or macroscopic indicators or surrogate measures. This is particularly true for models implemented in TFS environments.

The MOBIL model was introduced by Kesting, Treiber, and Helbing (2007). Its functionality was evaluated through simulation experiments focusing on lane-change rates. The experiments were conducted on a two-lane road section, including a merging area, and various parameter settings were investigated. However, model performance was not evaluated using empirical data. The authors noted that empirical validation using highly resolved trajectory data should be addressed in future research. The initial calibration and validation of the *LMRS* model proposed by Schakel, Knoop, and van Arem (2012) was conducted using cross-sectional lane distribution data from a 9 km motorway stretch, without employing microscopic performance measures. In the study by Jin et al. (2019), 373 lane changes from the NGSIM dataset were extracted and analysed. A simple rule-based lane-change model was implemented and used as a baseline. The base model was further extended by adding a component that considers a longer planning horizon (up to 9 s). In both cases, model performance was evaluated based on the number of correctly executed lane changes.

While rule-based lane-change models have primarily been evaluated using aggregated data or surrogate measures, probabilistic lane-change models are often calibrated using vehicle trajectory data, from which microscopic indicators can be derived.

For example, Ahmed et al. (1996) and Choudhury (2007) proposed models for different aspects of human lane-change behaviour based on utility theory. Their model parameters were estimated using the NGSIM dataset, allowing the investigation of factors influencing lane-change decision-making. However, simulation-based validation of these models primarily relied on aggregated traffic measures (e.g. lane-specific speeds), while microscopic indicators were neglected.

Similarly, Pang et al. (2020) utilised the NGSIM dataset to estimate the parameters of a probabilistic lane-change model incorporating driver heterogeneity and memory effects. Their results indicated that prediction accuracy improved when longer observation time windows were considered.

An extensive investigation regarding the calibration and validation of discretionary lane-change behaviour is shown in Knoop and Buisson, 2014. The authors used cross-sectional data from a motorway stretch in the UK with a 100 m distance span between the loop detectors, which enabled the re-identification of vehicles among the cross-sections. A probabilistic lane-change model was proposed, and the modelled and measured lane-change probabilities were compared, which the authors defined as microscopic calibration. In addition, the number of lane changes was analysed.

Ali et al. (2019) evaluated a game-theoretic approach for modelling mandatory lane changing. Their evaluation framework explicitly included microscopic indicators directly related to human lane-change behaviour. Among these, confusion matrix indicators as well as position- and time-related error measures were considered. This modelling approach was later extended by Ali et al. (2021) into a complete lane-change decision modelling framework for both connected and traditional traffic environments, applying similar evaluation methods.

More recently, Ali, Zheng, and Bliemer (2023) investigated how data selection and preprocessing influence the calibration of lane-change models. They found that model performance is highly sensitive to the precise start location and time of the vehicle trajectories used. Specifically, model performances are overestimated by using trajectory data in which drivers had already initiated a lane-change manoeuvre. The authors also addressed the data imbalance between lane-keeping and lane-changing events and recommended maintaining a ratio close to 1:1, not exceeding 2:1, i.e. the number of lane-keeping instances should not be more than double the number of lane-change events during model calibration.

Overall, probabilistic lane-change models provide valuable behavioural insights into lane-change decision-making and its influencing factors. However, these models are rarely fully integrated into traffic flow simulation environments and are often tested only in simplified simulation scenarios. Nevertheless, they are typically microscopically justified using trajectory data, and their findings provide useful methodological insights for trajectory-based evaluation of other lane-change models.

With respect to driver model parameter optimisation, numerous methods exist, particularly for car-following models. Durrani, Lee, and Maoh (2016) calibrated the Wiedemann model using NGSIM vehicle pairs, Kesting and Treiber (2008) applied a genetic algorithm (GA) to calibrate the IDM, and Ciuffo, Punzo,

and Montanino (2012) systematically compared algorithms, finding that the GA was effective, while Downhill Simplex and Interior Point methods performed poorly for trajectory-based calibration. Punzo, Zheng, and Montanino (2021) provide further guidelines for trajectory-based car-following model calibration. To the best of our knowledge, no similar systematic comparison exists for lane-change models.

In summary, prior research reveals three gaps: (i) evaluations often rely on macroscopic or surrogate measures rather than direct microscopic metrics, particularly for rule-based models heavily used in TFS; (ii) trajectory-based calibration of rule-based lane-change models remains underdeveloped; and (iii) no systematic comparison of optimisation algorithms for such models exists, unlike for car-following.

This study addresses these gaps by performing a comprehensive microscopic evaluation of two widely referenced rule-based lane-change models – *MOBIL* (Kesting, Treiber, and Helbing 2007) and *LMRS* (Schakel, Knoop, and van Arem 2012) – chosen for their contrasting complexity, conceptual similarities, and frequent use in both research and automated vehicle testing environments. They are implemented together with the *IDM* to assess how longitudinal dynamics influence lane-change decisions. Secondly, we benchmark conceptually distinct optimisation algorithms: stochastic gradient descent (commonly used in data-driven models), the genetic algorithm (proven effective for car-following calibration), and Bayesian Optimisation (suited for non-differentiable, computationally expensive objectives). Finally, through an in-depth examination of specific manoeuvre types, this study offers insights into human lane-change behaviour, contributing to a deeper understanding of the phenomenon and serving as a foundation for further lane-change modelling.

3. Methodology and analyses

This study aims to evaluate and calibrate lane-change models in terms of micro-level driver behaviour. This requires the utilisation of high-resolution trajectory data. Owing to data imbalance issues between lane-keeping and lane-changing situations, careful data processing is necessary (see Section 3.1). With respect to data processing, we also present a new approach to determine the time instance of lane change initiation, aiming to prevent overestimation issues. In this paper, the *MOBIL* and *LMRS* models (see Section 3.2) are considered as representative examples, balancing simplicity with sufficient differences to provide meaningful comparative insights. For all model parameters, a preliminary sensitivity analysis (see Section 3.3) is conducted to identify the most influential parameters for subsequent optimisation (see Section 3.4). Using different optimisation algorithms, we aim to fit the model parameters to best align with the observed data. This allows us to evaluate how well the algorithms identify parameter sets with minimal loss. Finally, the optimised parameter sets enable a comprehensive evaluation of both the optimisation algorithms and the lane-change models. These analyses provide insights into microscopic human lane-change behaviour and the capability of the investigated models to replicate real-world lane-change behaviour. Figure 1 introduces notation for all vehicles involved in a lane change manoeuvre. The lane-changing vehicle (ego vehicle) is depicted in grey. The notation is based on the work of Choudhury et al. (2007).

3.1. Data processing

In this work, the highD dataset is used, which is a naturalistic trajectory dataset recorded by drones on German motorway segments (Krajewski et al. 2018).

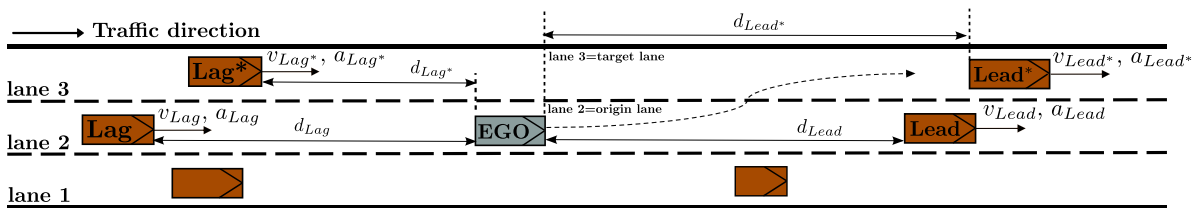


Figure 1. Discretionary lane change manoeuvre from lane 2 (origin lane) to lane 3 (target lane), showing the ego vehicle, the lead and lag vehicles in the origin and target lanes, and the associated variable.

The highD dataset contains 16.5 hours of trajectory data, representing approximately 110,500 vehicles. In total, it includes 60 separate recordings (*recId*), each capturing a varying number of vehicle trajectories (*vehId*). As described in Krajewski et al. (2018), post-processing was applied to the data to retrieve smooth positions, speeds, and accelerations in both the x- and y-directions.

The data covers traffic states ranging from dense to free-flow, which enhances the transferability of the analyses. It was collected from six different locations, primarily on open roads without merging and diverging areas. Consequently, the analyses focus on discretionary lane changes, i.e. primarily driven by speed advantages or local traffic conditions rather than mandatory route-related effects.

To visualise the trajectory data, the TraViA tool was heavily used, which is advantageous for checking the obtained results and for analysing specific scenarios (Siebinga 2021).

Regarding data processing, we first present a method to determine the time instance of lane change initialisation and describe the set of manoeuvres, which were selected for later model parameter optimisation. According to Ali, Zheng, and Bliemer (2023), it is crucial to extract the appropriate data when utilising vehicle trajectory data for evaluating lane-change models in order to avoid substantially overestimating model performance. The correct determination of the lane change initiation time is a prerequisite for excluding data points after the lane change starts. This study proposes a heuristic approach based on analysing the relative change of the lateral offset d_{lat} between a vehicle's centre point and its neighbouring lane markings (either to the left or right) depending on the lane change direction. By calculating the relative change (rc) in the lateral offset between two consecutive time instances (see formula 1), one can identify already small displacements, enabling early detection of lane-change manoeuvres.

$$rc(t = t_0) = \frac{d_{lat,t=t_0}}{d_{lat,t=t_0-1}} - 1 \quad (1)$$

To reduce noise and account for potential data inaccuracies with respect to the lateral positions, we applied an exponentially weighted moving average (EWMA) filter, using the *ewm*-function included in the Python pandas (2020) package. For differentiating between lane-keeping and lane-changing time instances, a threshold based on descriptive statistics was defined. This was feasible due to the fact that lane-keeping instances far outnumber lane-changing events. Additionally, relative lateral changes are much more pronounced during lane-changing compared to lane-keeping, meaning that lane-change events appear as statistical outliers. This facilitates the definition of a threshold to classify lane-changing and lane-keeping events. Since the ground-truth of the lane change initialisation is not available in the highD-data, evaluation is mainly limited to visual inspection of the recordings. The computed lane-change durations are, however, comparable to those reported by Toledo and Zohar (2007) and Li et al. (2023). In comparison to methods based on the lateral position, our method is independent on road geometry and can therefore also be applied to curved segments. The time series of two vehicles showing the relative change of the lateral offset between a vehicle's centre point and its neighbouring lane marking is presented in Figure 2. Both vehicles performed two lane changes. However, the second lane change shown in Figure 2(b) is not fully recorded,

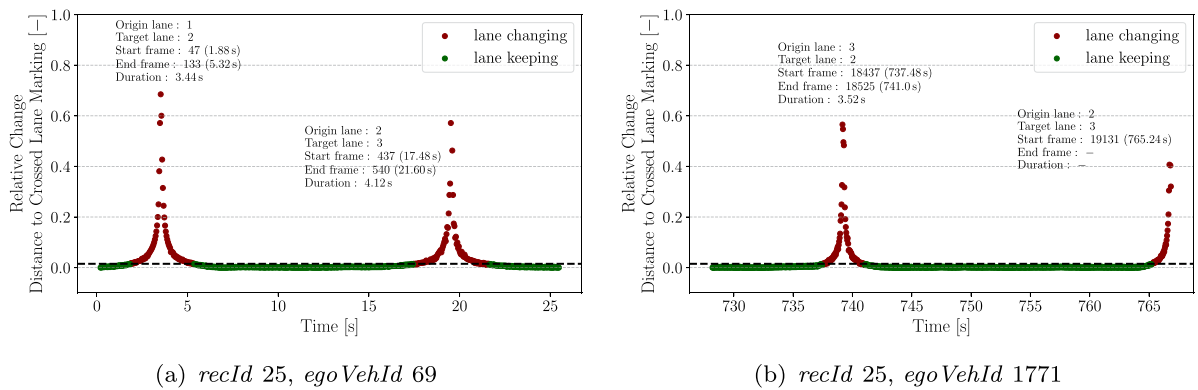


Figure 2. Determination of lane change initialisation time instance and differentiating between lane-changing and lane-keeping manoeuvres based on Formula 1.

as the vehicle left the drone's field of view (FoV) while the lane change was still ongoing. Such manoeuvres are not included in the later analyses as interactions with surrounding vehicles cannot be observed comprehensively. The threshold, represented by the dashed line, was set to 0.015. It should be noted that this value may need adjustment for datasets recorded with different levels of fidelity.

Besides determining the lane-change initiation time, it is crucial to extract appropriate data for lane-change model parameter optimisation. Unlike data-driven lane change intention models, which utilise multiple features over time (Xie et al. 2019; Zhang and Sun 2024), the models examined in this study (see Section 3.2) assess lane change decisions at each time step. They rely solely on features of the ego vehicle and its surrounding vehicles in both the origin and target lanes. Initial visual inspections revealed that lane changes often occur after a surrounding vehicle begins to execute a lane change toward the ego vehicle's origin lane. For instance, as shown in Figure 3, the ego vehicle performs a lane change in response to a leading vehicle in the adjacent right lane, which initiates a lane change and starts to drift laterally toward the ego vehicle's origin lane. In many trajectory datasets, such as the highD dataset, the IDs of surrounding vehicles are updated only when a vehicle's centre point crosses a neighbouring lane marking. At this point, a lane change is already ongoing, and affected vehicles may have already started to adapt their behaviour. To address this, we processed the data such that vehicles already performing a lane change toward either the ego vehicle's origin or target lane were included, ensuring that their impact is accurately captured in the analysis.

Using the updated trajectory data, all suitable lane-change manoeuvres were filtered, while also including potential false-positive situations by extracting keep-lane instances. Considering the remarks for data processing mentioned in Ali, Zheng, and Bliemer (2023), the following data were extracted for later model parameter optimisation: First, all time instances (see t_0 in Figure 3) of suitable lane-change manoeuvres were extracted. While filtering out lane changes that were not fully recorded within the drone's FoV, lane changes initiated at the very beginning or end of the drone's FoV were also excluded. This ensures that all relevant surrounding vehicles influencing the lane change decision are accurately recorded and represented in the data. The parameter ΔT_{FoV} (see Figure 3), introduced to account for the drone's field of view limitations, was set to 4 s. In addition to lane-changing events, two types of lane-keeping situations were considered. Following Ali, Zheng, and Bliemer (2023), a decision time window of 2 s was also assumed in this study. For each lane-changing event, we included the time instance 2 s before the lane change was initiated, during which the vehicle was still in a lane-keeping phase. This approach is intended to optimise model parameters in order to capture microscopic interactions and short-term dependencies effectively. While lane-change models should reflect microscopic interactions, they must also perform well on macroscopic characteristics, such as the number of executed lane changes. Therefore, time instances from keep-lane trajectories where no lane change occurred were also considered. For these keep-lane trajectories, it was verified whether a lane change to the left or right lane was permissible. Specifically, for potential left lane changes, only keep-lane trajectories with a leading vehicle present were used. This ensures that these potential lane change scenarios are realistic and avoid extracting random situations where a left lane change is highly unlikely.

To summarise, for each lane change event, two types of lane-keeping time instances were extracted: (i) the time instance 2 seconds before the lane change was initiated and (ii) an additional random time instance selected from a pre-filtered set of lane-keeping trajectories where a lane change was permissible but not executed. These three situations are illustrated in Figure 3. Including two lane-keeping time instances for each lane change event results in a 2:1 ratio, leading to an imbalanced dataset with more

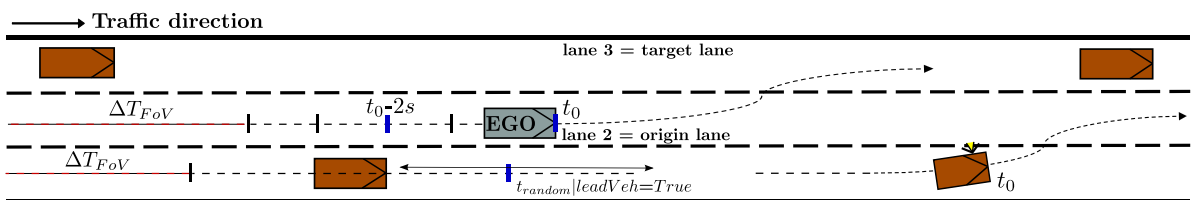


Figure 3. Processing of surrounding vehicle IDs for lane-change decision-making and extraction of lane-keeping and lane-changing situations for subsequent optimisation.

lane-keeping situations than lane-changing events. From the dataset, we extracted 2,337 lane changes to the right-hand side and 2,040 lane changes to the left-hand side, fulfilling the described criteria. To achieve a balanced dataset with respect to lane change directions, 2,040 lane changes from the set of right-hand side lane changes were randomly selected. The final dataset used in this study contains 4,080 lane-changing and 8,160 lane-keeping situations, resulting in 12,240 observations. While 80% of the data were used as training data, 20% were used as test data, when optimising the model parameters (see Section 3.4) of the *MOBIL* and *LMRS* lane-change model (see Section 3.2).

3.2. Lane-change model implementation

In this study, the *LMRS* and the *MOBIL* lane-change models were investigated. Both are incentive-based models and propose lane-changing rules for discretionary and mandatory lane changes. Explicit rules are only returned when the models are linked with a car-following model, which computes the acceleration values in the longitudinal direction for the ego vehicle and neighbouring vehicles involved in the lane-change decision. For both models, the *IDM* (Treiber, Hennecke, and Helbing 2000) in its original version was applied. The acceleration function of the *IDM* is given in the following formula 2, which consists of two main parts. One part compares the current speed v_{ego} to its desired speed $v_{des,ego}$, and the other part compares the desired distance s^* to the current distance headway between the ego and its lead vehicle d_{lead} . For the desired speed, we assigned the maximum measured speed for each vehicle. The acceleration function of the *IDM* is defined as:

$$a_{ego} = a \cdot \left(1 - \left(\frac{v_{ego}}{v_{des,ego}} \right)^\delta - \left(\frac{s^*(v_{ego}, \Delta v_{lead})}{d_{lead}} \right)^2 \right) \quad (2)$$

where a and δ are model parameters regarding the acceleration behaviour. As stated in Treiber and Kesting (2012), the desired distance is given by:

$$s^*(v_{ego}, \Delta v_{lead}) = s_0 + \max \left(0, v_{ego} \cdot T + \frac{v_{ego} \cdot \Delta v_{lead}}{2\sqrt{ab}} \right) \quad (3)$$

where s_0 and b reflect parameters for the minimum distance headway and comfortable deceleration, respectively. A short explanation of the relevant model parameters is shown in Table 1. For a detailed description of the investigated models and their parameters, we refer the reader to the corresponding publications: *IDM* (Treiber, Hennecke, and Helbing 2000), *LMRS* (Schakel et al. 2023; Schakel, Knoop, and van Arem 2012), and *MOBIL* (Kesting, Treiber, and Helbing 2007).

Table 1. Short description of the model parameters for the *LMRS* resp. of the *MOBIL* lane-change model and the *IDM* car-following model.

<i>IDM</i> car-following parameters			
s_0	Minimum stopping distance [m]		
T	Time gap [s]		
$v_{des,i}$	Desired speed of vehicle i [m/s] (derived from trajectory data)		
δ	Acceleration exponent [-]		
a	Acceleration parameter (increased parameter decreases desired distance in Equation 4)		
b	Comfortable deceleration parameter [m/s ²] (increased parameter decreases desired distance in Equation 4)		
Lane change parameters			
<i>LMRS</i>		<i>MOBIL</i>	
b_{safe}	Limit for safe deceleration [m/s ²]	b_{safe}	Limit for safe deceleration [m/s ²]
σ	Social-speed sensitivity [-]	p	Politeness factor [-]
T_{min}	Min. time gap during LC [s]	cf. T	Default <i>IDM</i> time gap [s]
T_{max}	Regular time gap during LC [s]	–	
d_{free}	Changing threshold [-]	Δa	Changing threshold [m/s ²]
cf. d_{free}	Asymmetry/Keep-Right term in <i>LMRS</i> equal to d_{free}	a_{bias}	Asymmetry/Keep-Right term [m/s ²]
x_0	anticipation distance [m]	–	–
v_{crit}	drivers ignore possible speed gain toward right lane if speed is higher than v_{crit} [km/h]	–	–
v_{gain}	full desire for speed incentive is experienced for a speed gain of v_{gain} [km/h]	–	–

3.2.1. LMRS

The initial version of the LMRS distinguishes between the following incentives to change lanes: to gain speed, to follow a route, and to keep right (given that a keep-right rule is applicable). This approach is extended in Schakel et al. (2023), where a fourth incentive is added for social-speed sensitivity. This additional component should reflect the social interaction between the ego vehicle and faster-following vehicles. Based on the basis of these four incentives, one can compute a total desire to change from lane i toward lane j :

$$d^{ij} = d_r^{ij} + \theta_v^{ij} \cdot (d_s^{ij} + d_b^{ij} + d_o^{ij}) \quad (4)$$

where d_r is the desire to follow a route, d_s is the desire to gain speed, d_b is the desire to keep right, and d_o is the desire to stay out of the way due to fast followers. The latter may increase the desire to change lane toward the slower lane or lowers the desire to change to the faster lane. θ_v is a model parameter that considers the extent to which a driver includes voluntary incentives in their lane change decision. In addition, the model includes relaxation and synchronisation. Relaxation is described as a phenomenon wherein drivers sometimes accept lower time gaps for a short time period. By applying small decelerations, the desired following headway is once again restored. If the acceleration of both the lane changer (ego) and the new lag vehicle in the target lane (lag*), is greater than a safe deceleration threshold, the gap is permitted:

$$\dot{v}^g \geq -b \cdot d^{ij} \quad (5)$$

with $g \in \{\text{ego}, \text{lag}^*\}$. If a gap is not accepted, drivers typically begin to synchronise their speed with the target lane in order to reduce speed differences and facilitate the lane-change manoeuvres. This behaviour is also incorporated into the LMRS model.

3.2.2. MOBIL

The MOBIL model formulates incentive criteria for both symmetric and asymmetric passing rules, enabling the consideration of driving regulations regarding lane usage, such as keep-right rules. In addition, the incentive criterion allows to include directly affected neighbouring vehicles. However, compared to the LMRS, varying incentives for performing a lane change are not distinguished. The incentive criterion for lane changing is formulated as follows:

$$\underbrace{\hat{a}_{ego} - a_{ego}}_{\text{ego vehicle}} + p \cdot \left(\underbrace{\hat{a}_{lag^*} - a_{lag^*}}_{\text{new lag vehicle}} + \underbrace{\hat{a}_{lag} - a_{lag}}_{\text{old lag vehicle}} \right) > \Delta a + a_{bias} \quad (6)$$

where the first two terms $\hat{a}_{ego} - a_{ego}$ reflect the advantage (utility) of the ego vehicle by performing a lane change. $\hat{a}_{ego}$ refers to the possible longitudinal acceleration in the target lane (new situation), whereas a_{ego} corresponds to the acceleration in the current lane (old situation). The third term, including the politeness factor p , considers the acceleration gain or loss of the lag vehicles in the current lane (lag) and in the target lane (lag*) due to the lane change of the ego vehicle. The politeness factor p varies between 0 and 1 and enables the modeling of purely egoistic ($p = 0$) to completely altruistic ($p = 1$) behaviour. The lane-changing threshold Δa hinders lane changes, where the advantage of lane changing is very low. Furthermore, a constant bias a_{bias} is added in the case of asymmetric behaviour (e.g. keep-right rule). In addition, a safety criterion is defined, preventing unsafe lane change situations, which can cause very high decelerations of the lag vehicle in the target lane. Hence, this deceleration must not exceed a given safe limit b_{safe} . In comparison to the LMRS, this criterion is not dependent on the desire to change lanes.

$$\hat{a}_{lag^*} \geq -b_{safe} \quad (7)$$

3.3. Sensitivity analyses

The results of the performed sensitivity analyses are shown below. These values were used as a basis for selecting the most influential parameters for the subsequent optimisation. The parameter selection ensures

comparability across different optimisation algorithms and between the two benchmarked lane-change models. This should also ease the optimisation procedure by reducing complexity without necessarily compromising performance, as demonstrated by Punzo, Montanino, and Ciuffo (2015) in their research on car-following model parameter optimisation for the *IDM*.

Since this is merely an intermediate step, we opted for a straightforward *One-At-A-Time* sensitivity analysis to identify the most influential parameters, acknowledging that this approach does not allow us to investigate the mutual impact of some parameters on the results. To assess parameter sensitivity, we iteratively varied each parameter within the range specified in Table 2, keeping all other parameters fixed. These constraints were also applied for later optimisation. They were defined with a broad range on purpose, allowing the optimisation procedure to explore extreme values and identify potential limitations of the models. To introduce variability, each parameter value was evaluated using 10 randomly drawn samples from the training data described in Section 3.1. Each data sample consisted of 1,024 instances of lane-changing and lane-keeping situations. For the sensitivity analyses, we evaluated the misclassification rate, i.e. the proportion of manoeuvres incorrectly classified by the models. To account for the data imbalance between lane-keeping events and lane-changing events, a weighted misclassification rate (wMR) was computed. Table 2 shows the corresponding minimum, maximum, and standard deviation values.

For the *LMRS* model, the most impactful parameters are the changing threshold d_{free} , the safe deceleration parameter b_{safe} , the v_{gain} parameter affecting the desire to gain speed and the anticipation distance x_0 . For the *MOBIL* model, the time gap parameter T , the safe deceleration parameter b_{safe} , the politeness factor p and the keep-right threshold a_{bias} have the greatest influence on the model results. The socio-speed sensitivity (σ) in the *LMRS* model and the lane-changing threshold (Δa) in the *MOBIL* model were also included for comparability reasons, as their counterparts in the other models are highly influential.

In total, seven parameters for the *LMRS/IDM* model and six parameters for the *MOBIL/IDM* model were included in the subsequent optimisation framework. These parameters, highlighted in bold in Table 2, represent the most influential factors driving model performance.

3.4. Model parameter optimisation

To optimise the parameters of the described lane-change models in order to best align with the measured data and observed behaviour, varying algorithms were applied: stochastic gradient descent, the genetic algorithm, and Bayesian optimisation. The algorithms were selected to represent fundamentally different optimisation paradigms: gradient-based, evolutionary, and probabilistic-based search. All three are mature, widely applied in optimisation research, and have been used in related model calibration tasks, ensuring both methodological diversity and comparability with previous studies.

Given the binary nature of the target variable, the misclassification rate was selected as the loss function for model evaluation during the optimisation process. Since the dataset used for optimising the model parameters is imbalanced with respect to lane-keeping and lane-changing events (see Section 3.1), a

Table 2. Results of OAT sensitivity analyses for the *LMRS* and *MOBIL* model.

<i>LMRS/IDM</i>						<i>MOBIL/IDM</i>					
Param.	Range	Def. value	wMR min	wMR max	STD	Param.	Range	Def. value	wMR min	wMR max	STD
s_0	[0,10]	2.0	0.354	0.414	0.017	s_0	[0,10]	2.0	0.317	0.332	0.004
T	[0.5,5]	1.0	0.355	0.412	0.017	T	[0.5,5]	1.0	0.317	0.417	0.037
δ	[0,10]	4.0	0.355	0.412	0.017	δ	[0,10]	4.0	0.316	0.320	0.001
a	[0.5,5]	1.0	0.328	0.433	0.024	a	[0.5,5]	1.0	0.317	0.340	0.008
b	[0.5,5]	1.5	0.351	0.422	0.018	b	[0.5,5]	1.5	0.315	0.329	0.004
b_{safe}	[-25,0]	-2.0	0.329	0.443	0.024	b_{safe}	[-25,0]	-2.0	0.279	0.389	0.032
σ	[0,1]	0	0.337	0.416	0.019	p	[0,1]	0.5	0.304	0.363	0.018
T_{min}	[0,5]	-0.25	0.355	0.414	0.017	—	—	—	—	—	—
T_{max}	[0,5]	0.25	0.355	0.413	0.017	—	—	—	—	—	—
d_{free}	[0,1]	0.25	0.313	0.448	0.030	Δa	[0,2]	0.1	0.313	0.344	0.009
cf. d_{free}	—	—	—	—	—	a_{bias}	[0,2]	0.3	0.301	0.350	0.016
x_0	[50,500]	250.0	0.352	0.461	0.023	—	—	—	—	—	—
v_{crit}	[0,10]	60.0	0.355	0.412	0.016	—	—	—	—	—	—
v_{gain}	[0,100]	20.0	0.346	0.453	0.024	—	—	—	—	—	—

weighted misclassification error rate is also computed, which enables a more meaningful interpretation of the results. In the following, the investigated optimisation algorithms will shortly be described.

Stochastic gradient descent (SGD) is a widely used local optimisation algorithm, particularly in the field of machine learning. It minimises a loss function by iteratively updating model parameters in the direction of the steepest descent, as determined by the gradient (Ruder 2016). The parameter update at each iteration n is given by:

$$\theta_{n+1} = \theta_n - \eta \nabla L(\theta_n | X_i) \quad (8)$$

where θ_{n+1} is the updated parameter set, η is the learning rate, and $\nabla L(\theta_n | X_i)$ represents the gradient of the loss function in iteration n with respect to the parameters for a given mini-batch sample i of the complete training dataset X . The learning rate, η , is a critical parameter that must be carefully defined (Deisenroth, Faisal, and Ong 2020). In typical machine learning tasks, feature normalisation ensures that features fall within a similar range, facilitating stable optimisation (Ruder 2016). However, for the investigated models, normalisation is not feasible because both the model parameters and the input features have physical meanings, and normalisation would distort the results. Using a uniform learning rate across all features could lead to inconsistent parameter updates: features with larger values would result in smaller updates, while features with smaller values could cause disproportionately large updates. To address this, varying learning rates tailored to each parameter's range were defined. Additionally, gradient clipping was applied to prevent excessively large updates, ensuring stable optimisation. SGD operates as a local optimiser, aiming to converge to a local optimum near the randomly initialised parameter set. To mitigate the risk of becoming stuck in weak local minima, stochasticity is added by computing gradients based on mini-batches randomly sampled from the training data. This approach introduces slight variations in the loss function across iterations, increasing the likelihood of escaping suboptimal solutions. In our implementation, the mini-batch size was set to 512, striking a balance between stochasticity and stability. Within one optimisation procedure, 1,000 epochs were computed. Usually, in one epoch, the entire training dataset is processed once, while during each epoch, multiple parameter updates are performed for every mini-batch. Finally, it should be noted that the *MOBIL* and *LMRS* models rely on discrete conditions. To enable gradient-based optimisation, we reformulated these conditions with smooth, differentiable surrogates. For example, a Boolean condition $g(d)$ returning 1 if $d < d_{\max}$ was replaced by a continuous gating function such as:

$$g(d) = \tanh(\max(0, k \cdot (d_{\max} - d))) \quad (9)$$

where the scaling parameter k controls the steepness of the transition. Similar combinations of tanh, sigmoid, and ReLU functions were applied for other conditions. This reformulation preserves the original decision logic while ensuring full differentiability for gradient-based optimisation.

The genetic algorithm (GA) is a global optimisation heuristic inspired by principles of natural selection (Goldberg 1989). We defined a population size of 25, with the initial population sampled randomly. Each individual (parameter set) was evaluated based on the misclassification rate. The five best individuals were carried over to the next generation to ensure the retention of high-performing solutions. Using roulette wheel selection, a pool of parent individuals was created, from which new individuals were generated via crossover by exchanging and recombining parameter values. Additionally, for GA, random mutation introduces diversity by altering some parameter values. Within one optimisation procedure, 150 generations, each consisting of 25 individuals, are computed and evaluated.

Bayesian optimisation (BO) is an iterative global optimisation algorithm, which consists of two components: (i) a statistical model for approximating the loss function and (ii) an acquisition function for deciding which parameter set sample and evaluate next (Frazier 2018). The acquisition function intends to balance exploration and exploitation of the parameter space (Christianson and Gramacy 2024). As a statistical model, most often and also in this study, a Gaussian Process ($\mathcal{GP}$) model is applied:

$$L(\theta | X) \sim \mathcal{GP}(\mu(\theta), k(\theta, \theta')) \quad (10)$$

meaning that the loss function with respect to the parameter vector and given training data is approximated via an $\mathcal{GP}$ model, where $\mu(\theta)$ is the mean function, representing the expected value of the loss function, and $k(\theta, \theta')$ denotes the covariance function, which is often referred to as the kernel function. In this study, a radial

basis function (RBF) kernel was employed to compute the covariance matrix, capturing the similarity between parameter sets. The $\mathcal{GP}$ model provides a Bayesian posterior probability distribution $p(L|\theta)$ that quantifies the uncertainty of the loss value for a given parameter set θ . Every time, when a new parameter set is evaluated, this posterior distribution is updated. For simplicity, the prior mean function $\mu(\theta)$ is typically assumed to be zero unless prior knowledge suggests otherwise. After conditioning on the observed data, which is done after every iteration, the posterior mean is updated and computed based on the covariance between the observed parameters and the new parameters, enabling the $\mathcal{GP}$ model to predict the loss and its uncertainty at unseen parameter settings. Based on the current posterior distribution, an acquisition function is then used to measure the value that would be generated by evaluating the loss function for a new parameter set θ . As an acquisition function, the expected improvement (EI) function was applied. As described in Jones, Schonlau, and Welch (1998), an improvement with respect to the loss function is given by:

$$I(\theta) = \max(L_{\min} - L(\theta), 0) \quad (11)$$

meaning that there is an improvement if the loss value of the current parameter vector θ is better than the current best loss value $L_{\min}$. As $L(\theta)$ is treated as a random variable in the $\mathcal{GP}$ model, we aim to compute the expected improvement, which can be written as:

$$EI(\theta) = \mathbb{E}[\max(L_{\min} - L(\theta), 0)] \quad (12)$$

A closed-form solution of this expectation can be found in Jones, Schonlau, and Welch (1998). The expected improvement algorithm then returns the parameter set for the next iteration with the largest expected improvement, given by:

$$\theta_{n+1} = \arg \max_{\theta} EI_n(\theta) \quad (13)$$

As for GA, 150 iterations were also performed for BO within one optimisation procedure. Compared to SGD, where parameters are updated for every mini-batch, and GA, which evaluates 25 individual parameter sets in each generation, BO evaluates fewer parameter sets overall. This is because BO becomes computationally expensive as the number of iterations increases; the covariance matrix required for updating the posterior probability distribution by using the $\mathcal{GP}$ model grows quadratically with each iteration.

We execute the full optimisation procedure 150 times for each of the described algorithms. This approach allows a comprehensive evaluation of both the lane-change models and the differences among the optimisation algorithms. The results are detailed in the following section.

4. Results

This section presents the main results. First, we compare the different optimisation techniques. Subsequently, representative parameter sets obtained during the optimisation procedure are used to evaluate and compare the model performance based on confusion matrix metrics. In addition, the results are analysed with respect to lane-changing behaviour, focusing on the desire to change lanes and gap acceptance characteristics.

4.1. Comparison optimisation results

In Table 3, descriptive statistics of the initialised and optimised loss values are presented for 150 executed optimisation procedures and for the various optimisation algorithms. The final evaluation was performed on the test data; however, the corresponding final loss values for the training data are also reported in brackets to indicate potential overfitting. The weighted misclassification rate is shown, which accounts for data imbalance with respect to lane-keeping and lane-changing manoeuvres. Hence, a loss value of 0.5 (50%) would most likely indicate that the lane-change model would never/always predict a lane change.

As the parameter sets were initialised randomly within the defined parameter space (see range in Table 2), there are only minor differences when comparing the statistical indicators of the loss values in the very beginning (init loss) of each optimisation run. Comparing those init loss values between the two lane-change models shows that the worst (maximum) loss values are rather similar, while the minimum

Table 3. Comparison of initial and optimised loss values (wMR) for different optimisation algorithms; $n = 150$.

		Init test loss (Init training loss)			Optimised test loss (optimised training loss)		
		SGD	GA	BO	SGD	GA	BO
<i>LMRS/IDM</i>	Minimum	0.322 (0.345)	0.357 (0.364)	0.299 (0.296)	0.255 (0.222)	0.265 (0.263)	0.258 (0.261)
	25%-quantile	0.410 (0.366)	0.427 (0.430)	0.431 (0.435)	0.263 (0.228)	0.298 (0.296)	0.271 (0.276)
	Median	0.429 (0.443)	0.439 (0.438)	0.440 (0.442)	0.269 (0.256)	0.311 (0.308)	0.279 (0.278)
	75%-quantile	0.438 (0.427)	0.444 (0.444)	0.444 (0.444)	0.428 (0.394)	0.326 (0.329)	0.287 (0.289)
	Maximum	0.444 (0.446)	0.445 (0.444)	0.445 (0.444)	0.444 (0.403)	0.373 (0.371)	0.326 (0.334)
	IQR _{0.75–0.25}	0.028	0.017	0.013	0.165	0.028	0.016
<i>MOBIL/IDM</i>	Minimum	0.223 (0.229)	0.222 (0.227)	0.227 (0.232)	0.223 (0.184)	0.215 (0.224)	0.210 (0.221)
	25%-quantile	0.277 (0.316)	0.302 (0.310)	0.315 (0.321)	0.231 (0.199)	0.233 (0.235)	0.218 (0.227)
	Median	0.310 (0.311)	0.356 (0.361)	0.372 (0.375)	0.233 (0.204)	0.239 (0.242)	0.221 (0.226)
	75%-quantile	0.383 (0.408)	0.400 (0.406)	0.403 (0.404)	0.234 (0.188)	0.251 (0.251)	0.223 (0.228)
	Maximum	0.444 (0.443)	0.445 (0.445)	0.442 (0.444)	0.443 (0.413)	0.314 (0.311)	0.230 (0.233)
	IQR _{0.75–0.25}	0.106	0.098	0.088	0.003	0.018	0.005

initialisation loss is much lower for the *MOBIL* model compared to the *LMRS* model. The optimised loss values on the right-hand side in Table 3 show the differences among the optimisation algorithms. For the *MOBIL* model, the best results can be achieved when utilising Bayesian Optimisation. First, the minimum achieved loss values ($=0.210$) are slightly lower compared to those of SGD and GA. This also applies to other statistical indicators, such as the 25%-quantile and the median value. In addition, the worst (maximum) achieved optimised loss value for BO is substantially lower than the maximum loss value for SGD and also GA. This means that in some rare cases, SGD does not manage to converge toward a parameter set with improved loss values. This phenomenon can also be seen for the *LMRS* model, where the 75%-quantile also has a relatively high loss value, meaning that there were even more optimisation runs, where SGD did not manage to improve the loss values. At the same time, the median and 25%-quantile loss values are lower for SGD compared to BO and GA. This implies that if SGD successfully converges to a minimum, then it outperforms BO and GA, despite being a local optimization algorithm. This, however, comes at the cost of overfitting, as indicated by lower loss values on the training data compared to the test data, which reflects weaker generalization. In contrast, BO and GA do not exhibit this behaviour, suggesting that their broader search strategies yield solutions that generalize more robustly.

Figure 4 presents the empirical probability density functions alongside the corresponding cumulative distribution functions for both the initialised (red) and optimised (blue) parameters. While the figure shows representative results for SGD and BO, a comprehensive overview covering all model parameters across different optimisation algorithms is provided in the appendix (see Appendix A and Appendix B). The red cumulative distribution functions show that the parameters were sampled from a uniform distribution among all optimisation algorithms, which guarantees comparability. In general, the results indicate that certain parameters tend to converge toward the boundaries of the defined constraints across different optimisation algorithms.

For both models, this effect is particularly pronounced for the time gap parameter T . After running the optimisation, the time gap parameter T is generally between 0.5 and 1.0 s, except for SGD, which occasionally yields larger values, as shown in Figure 4(a). As already described with respect to the results in Table 3, SGD sometimes does not manage to converge toward an area in the parameter space with reduced loss values. The result in Figure 4(a) indicates that when T is initialised with relatively high values, SGD struggles to escape that region in the parameter space, leading to consistently high loss values. Owing to the shape of the loss function in the region, the gradients are most likely very low. This leads only to few parameter updates, which hinders the algorithm to optimise toward an area with improved loss values.

Furthermore, the optimised values of the safe deceleration parameter b_{safe} show similar distribution patterns across the considered optimisation algorithms and lane-change models (see representative results in Figure 4(b)). In all cases, the values tend to converge toward relatively large negative values – up to -25 m/s^2 – which corresponds to the lower bound of the defined constraint. Such high negative parameter values are not realistic, as they would suggest extremely aggressive behaviour of the lane-changing vehicles. Moreover, such high decelerations could not be observed among all the analysed driving situations. This indicates that the models need to be parameterised aggressively, in order to depict the observed microscopic human lane-change behaviour. A notable difference between the two models can be found when analysing the parameters that account for the influence of surrounding vehicles on lane

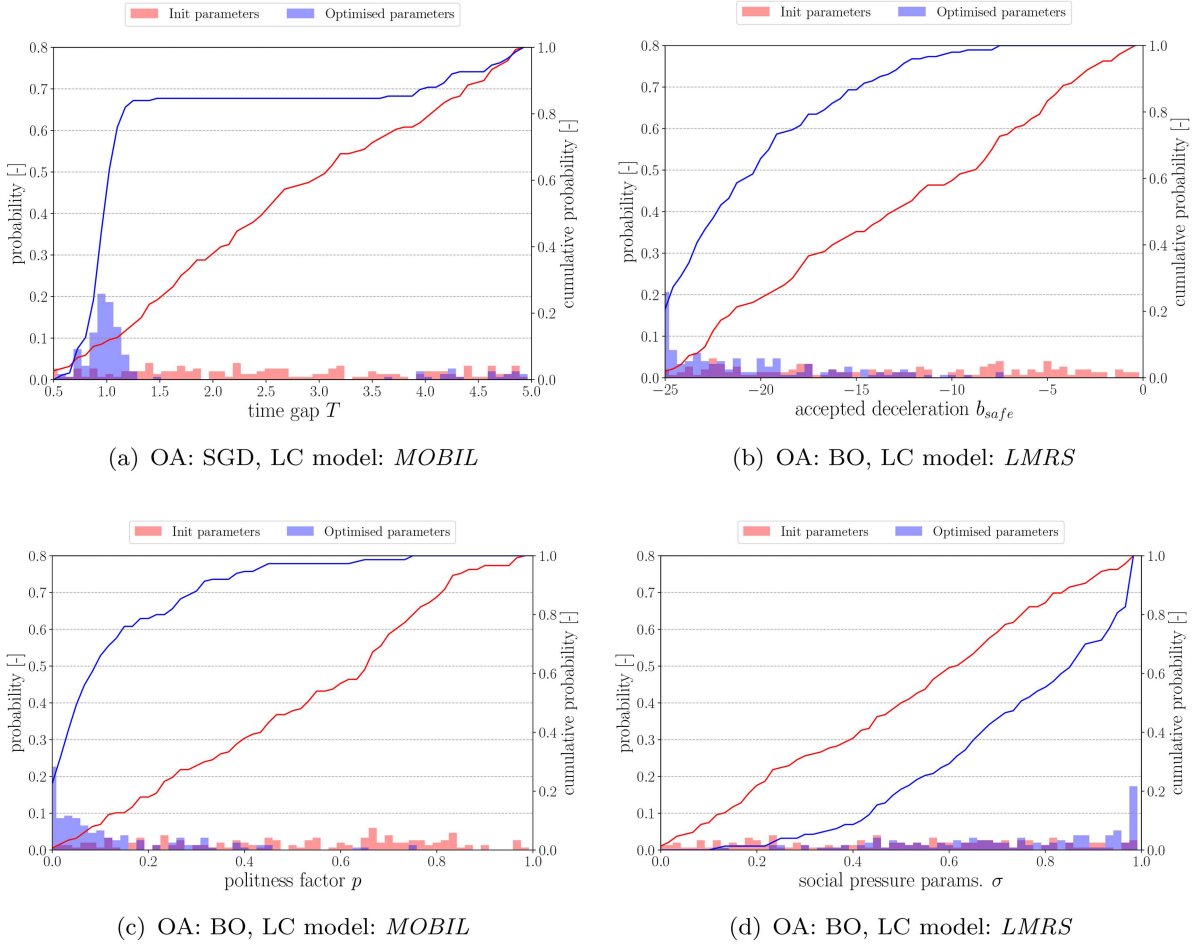


Figure 4. Exemplary distributions of initialised and optimised parameter values obtained for different optimisation algorithms (OA) and lane-change (LC) models.

Table 4. Representative *LMRS* and *MOBIL* parameter sets were used for further detailed evaluation.

<i>LMRS/IDM</i>						<i>MOBIL/IDM</i>					
Param.	PS 1	PS 2	PS 3	PS 4	PS 5	Param.	PS 1	PS 2	PS 3	PS 4	PS 5
T	1.00	0.89	0.52	0.75	0.50	T	1.00	2.64	0.93	0.75	0.50
a	1.00	2.96	4.68	1.00	1.00	a	1.00	1.75	0.82	1.00	1.00
b_{safe}	-2.0	-11.0	-22.7	-8.0	-16.0	b_{safe}	-2.0	-11.7	-21.1	-8.0	-16.0
σ	0.50	0.92	0.91	0.75	1.00	p	0.50	0.55	0.44	0.50	0.50
d_{free}	0.20	0.08	0.04	0.10	0.05	Δa	0.10	1.54	1.04	0.30	0.30
x₀	200.0	223.5	121.5	150.0	100.0	a_{bias}	0.30	0.73	1.70	0.50	1.00
v_{gain}	50.0	32.60	10.00	25.0	10.0	—	—	—	—	—	—

change decision-making. For the *LMRS* model, the optimised values of the social-speed sensitivity parameter σ tend to be higher (see Figure 4(d)). In contrast, the politeness factor p in the *MOBIL* model was predominantly reduced during optimisation (see Figure 4(c)).

4.2. Micro-level lane-change model evaluation

To gain deeper insights into the investigated lane-change models, we analyse the model output results for some representative parameter sets. These analyses aim to reveal the models' functionalities and potential limitations, enhancing the understanding of both human lane-change behaviour and its modelling. The investigated parameter sets, listed in Table 4, were selected as follows: the first parameter set (PS1) uses values proposed by Schakel et al. (2023) for the *LMRS* model and by Kesting, Treiber, and Helbing (2007) for

the *MOBIL* model. The second (PS2) and third (PS3) parameter sets were derived as representative sets from the optimisation procedure conducted via SGD. Specifically, optimisation runs that successfully converged to a local optimum were identified. For PS2, the results were filtered with loss values (wMR) equal to 0.35, and the median parameter values of the corresponding runs were computed. Similarly, PS3 was determined by filtering results with loss values equal to 0.25. As such, PS2 reflects a suboptimal intermediate result, whereas PS3 represents a parameter set with an optimised loss. Since extreme parameter values are frequently returned during optimisation and complicate the interpretability of the resulting parameter sets, two additional sets (PS4 and PS5) were defined heuristically. Specifically, the highly negative values observed for the b_{Safe} parameter were retained, while most other parameters were kept close to their default settings commonly used in TFS environments. In addition, selected parameters (T , Δa , v_{gain}) were adjusted to represent a rather aggressive driving style.

For all parameter sets, we computed how many manoeuvres would have been correctly executed by the lane-change models, distinguishing between lane changes toward the left and right hand-side. It should be noted that this evaluation is not intended to assess the performance of the optimisation algorithms themselves. Instead, our objective is to evaluate the micro-level validity of the lane-change models for a set of representative parameter sets (derived either from optimisation or defined heuristically). To increase the robustness and generalisability of this evaluation, we therefore used the full dataset (training and test data), rather than restricting the analysis to the test set only. Additionally, we emphasise that the model parameters were not optimised separately for left and right lane changes. However, in the subsequent analyses, we report the results by lane-change direction, which allows for the identification of potential direction-dependent model limitations. The results are presented in Table 5, with a distinction made between lane-keeping and lane-changing manoeuvres. As described in Section 3.1, two different lane-keeping situations were considered in the analyses, namely, a keep-lane time instance 2 seconds before the lane change was initiated and an additional random time instance selected from a pre-filtered set of lane-keeping trajectories. The percentage values for True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) situations, which are shown in Table 5, are related to the corresponding correct manoeuvre type. For computing the accuracy and F1 score, all manoeuvres in the dataset were considered. The accuracy (A) and the F1 score are given by the following formulas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

Table 5. Confusion matrix indicators for *LMRS* and *MOBIL*.

LC model	PS	LC dir.	$y_{\text{true}} = \text{LC}$		$y_{\text{true}} = \text{KL}$		$y_{\text{true}} = \text{KL2 sb. Lc}$		A	F1
			TP	FN	TN	FP	TN	FP		
<i>LMRS/IDM</i>	1	Left	16.57%	83.43%	99.46%	0.54%	85.74%	14.26%	67.3%	25.2%
		Right	0.54%	99.46%	99.90%	0.10%	99.41%	0.59%	66.6%	1.1%
	2	Left	44.02%	55.98%	95.49%	4.51%	68.53%	31.47%	69.3%	48.9%
		Right	33.58%	66.42%	96.42%	3.58%	83.68%	16.32%	71.2%	43.8%
	3	Left	76.96%	23.04%	88.82%	11.18%	52.75%	47.25%	72.8%	65.4%
		Right	55.00%	45.00%	89.80%	10.20%	68.43%	31.57%	71.1%	55.9%
	4	Left	50.20%	49.80%	94.90%	5.10%	63.68%	36.32%	69.6%	52.4%
		Right	37.40%	62.60%	95.29%	4.71%	77.79%	22.21%	70.2%	45.5%
	5	Left	67.21%	32.79%	89.41%	10.59%	55.64%	44.36%	70.8%	60.5%
		Right	57.70%	42.30%	86.37%	13.63%	65.15%	34.85%	69.7%	56.0%
<i>MOBIL/IDM</i>	1	Left	56.72%	43.28%	92.30%	7.70%	60.74%	39.26%	69.9%	55.7%
		Right	67.16%	32.84%	84.80%	15.20%	63.33%	36.67%	71.8%	61.3%
	2	Left	46.27%	53.73%	89.17%	10.83%	67.25%	32.75%	67.6%	48.7%
		Right	40.29%	59.71%	84.36%	15.64%	80.69%	19.31%	68.4%	46.0%
	3	Left	57.55%	42.45%	99.36%	0.64%	75.15%	24.85%	77.4%	62.9%
		Right	81.42%	18.58%	79.71%	20.29%	51.81%	48.19%	71.0%	65.2%
	4	Left	71.86%	28.19%	96.42%	3.58%	58.48%	41.57%	75.6%	66.2%
		Right	70.10%	29.90%	84.85%	15.20%	60.15%	39.85%	71.7%	62.3%
	5	Left	62.79%	37.25%	98.87%	1.13%	67.50%	32.55%	76.4%	63.9%
		Right	86.13%	13.87%	78.28%	21.76%	48.68%	51.32%	71.0%	66.5%

$$\text{Recall} = \frac{TP}{TP + FN} \quad (16)$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

The results in Table 5 show substantial differences among the lane-change models, the varying investigated parameter sets and also for the different lane-change directions. For *LMRS*, the results of PS 1 indicate that these parameters are not well-suited for realistically reflecting the observed lane-change behaviour. For PS1, the *MOBIL* model shows better results, particularly for right-hand lane changes, as 67.2% of the lane changes to the right would have been executed correctly. For *LMRS*, the optimised parameter set (PS3) resulted in the best results, as indicated by the F1 scores for the left and right lane changes. However, there is a difference regarding the lane change direction, as lane changes to the left-hand side do have a higher F1 score compared to lane changes to the right, mainly influenced by a higher share of True-Positives. For the *MOBIL* model, PS4 slightly outperforms the parameter sets PS3 and PS5. Moreover, PS4 is more balanced with respect to the lane change direction. For PS3 and PS5, there are more True-Positive right lane changes. At the same time, the number of False-Positive situations was also increased for both lane-keeping time instances 2 s before the lane change was initialised and also lane-keeping instances, which were randomly selected from a set of pre-filtered keep lane trajectories. The *LMRS* and *MOBIL* perform comparably well for lane changes to the left-hand side. For lane changes to the right, the *MOBIL* shows better results.

For both models, we observed that certain lane changes executed in reality would not have been suggested by the models. Table 6 presents the intermediate model results responsible for these unexecuted lane changes, offering valuable insight into these False-Negative situations. In some cases, the models would have suggested a lane change at an earlier or later time, but these cases – listed in the last row for the respective lane change direction – were excluded from this analysis. The results are categorised by lane-change model, parameter set, and lane-change direction. As outlined in Section 3.2, the *MOBIL* model requires two main criteria to be met for a lane change to occur: (i) the desire to change lanes (LcD_{Ego}) must be sufficiently high, and (ii) a suitable gap must exist toward the lag vehicle in the target lane (GA_{Lag}). For the *LMRS*, a third criterion is introduced: the required deceleration of the ego vehicle (GA_{Ego}) due to a leading vehicle on the target lane must also be limited. Consequently, the *LMRS* model has seven potential combinations of criteria that could prevent a lane change, compared to only three for the *MOBIL* model. The results in Table 6 indicate which set of criteria were not fulfilled and responsible for not executing a lane change.

For the *LMRS* model, left lane changes are most often not executed because the gap acceptance criteria for the lag vehicle and a sufficiently high desire to change lanes are not met. This number is notably reduced in PS3, where only 65 left lane changes were not executed because of these factors. Additionally, there are numerous cases where all conditions (LcD_{Ego} , GA_{Ego} , GA_{Lag}) were not fulfilled, likely occurring during dense traffic with many nearby surrounding vehicles. Across all parameter sets, a relatively small number of lane changes were not executed solely because of an unfulfilled gap acceptance criteria for the ego vehicle or insufficient desire to change lane. One possible explanation is the link between these two factors: a very close leading vehicle in the target lane not only reduces the desire to change lane but also lowers the likelihood of accepting the lead gap.

For right lane changes, the primary reasons for not executing a lane change are a low desire to change lane and the combination of this with an unfulfilled gap acceptance criterion for the ego vehicle. This suggests that the desire to change to the right lane cannot be captured well with the chosen parameter values. A similar pattern is observed in the *MOBIL* model, where right lane changes are primarily hindered by an insufficient desire to change lane toward the right-hand side. However, the absolute numbers are lower for the *MOBIL* model. One reason could be that *MOBIL* defines a separate bias parameter a_{bias} for keeping-right. Therefore, the absolute numbers are substantially decreased for PS3 and PS5, where relatively high a_{bias} parameter values are defined.

For left lane changes in the *MOBIL* model, both the gap acceptance criterion and the desire to change lanes play significant roles. In PS2, the b_{safe} parameter, which is mainly responsible for gap acceptance, was set to -11.7 m/s^2 . This value allows all the gaps to be accepted if the lag vehicle on the target lane is not

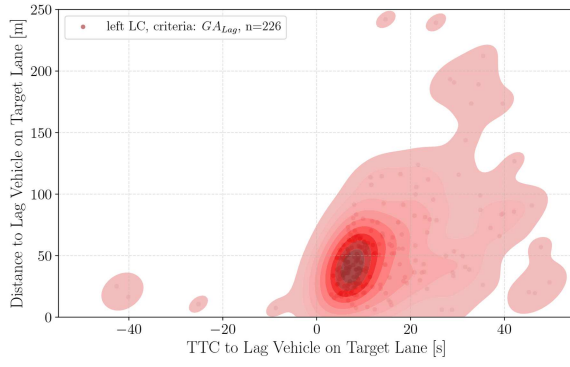
Table 6. Number of False-Negative manoeuvres per model criterion, which hindered the respective model from suggesting a lane change; GA_{Ego} corresponds to the gap acceptance criterion of the ego vehicle; GA_{Lag} corresponds to the gap acceptance criterion due to a lag vehicle on the target lane; LcD_{Ego} corresponds to an insufficient lane change desire.

Not fulfilled criterion		LMRS/IDM					MOBIL/IDM				
		PS 1	PS 2	PS 3	PS 4	PS 5	PS 1	PS 2	PS 3	PS 4	PS 5
Left LC	only GA_{Ego}	40	78	16	18	4	–	–	–	–	–
		2.7%	8.6%	7.0%	2.7%	1.3%					
	only GA_{Lag}	150	165	24	95	20	318	104	1	17	1
		10.1%	18.4%	10.6%	14.4%	6.7%	51.0%	14.6%	0.3%	5.6%	0.2%
	only LcD_{Ego}	122	12	18	16	17	103	284	507	231	403
		8.2%	1.3%	7.9%	2.4%	5.3%	16.5%	39.7%	98.6%	76.5%	98.8%
	GA_{Lag} , LcD_{Ego}	631	241	65	310	140	202	327	6	54	4
		42.5%	26.9%	28.6%	47.0%	46.7%	32.4%	45.7%	1.2%	17.9%	1.0%
	GA_{Lag} , GA_{Ego}	61	113	16	25	9	–	–	–	–	–
		4.1%	12.6%	7.0%	3.8%	3.0%					
Right LC	LcD_{Ego} , GA_{Ego}	94	24	13	26	18	–	–	–	–	–
		6.3%	2.7%	5.7%	3.9%	6.0%					
	LcD_{Ego} , GA_{Ego} , GA_{Lag}	387	263	75	170	94	–	–	–	–	–
		26.1%	29.4%	33.0%	25.8%	31.0%					
	Σ not executed FN	1485	896	227	660	302	623	715	514	302	408
		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Σ earlier/later exec. FN	217	246	243	356	367	260	381	352	273	352
	only GA_{Ego}	3	81	14	24	3	–	–	–	–	–
		0.1%	7.2%	1.9%	2.3%	0.4%					
Left LC	only GA_{Lag}	6	50	10	11	0	13	58	0	0	0
		0.3%	4.4%	1.4%	1.1%	0.0%	3.2%	6.8%	0.0%	0.0%	0.0%
	only LcD_{Ego}	541	148	154	221	130	351	401	183	348	118
		27.0%	13.2%	21.1%	21.5%	19.3%	85.6%	46.9%	97.3%	97.2%	97.5%
	GA_{Lag} , LcD_{Ego}	174	79	50	49	34	46	396	5	10	3
		8.7%	7.0%	6.9%	4.8%	5.0%	11.2%	46.3%	2.7%	2.8%	2.5%
	GA_{Lag} , GA_{Ego}	1	43	4	7	0	–	–	–	–	–
		0.0%	3.8%	0.5%	0.7%	0.0%					
	LcD_{Ego} , GA_{Ego}	896	468	358	551	388	–	–	–	–	–
		44.7%	41.6%	49.1%	53.7%	57.5%					
Right LC	LcD_{Ego} , GA_{Ego} , GA_{Lag}	383	255	139	164	120	–	–	–	–	–
		19.1%	22.7%	19.1%	16.0%	17.8%					
	Σ not executed FN	2004	1124	729	1027	675	410	855	188	358	121
		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Σ earlier/later exec. FN	25	231	189	250	188	260	363	191	252	162

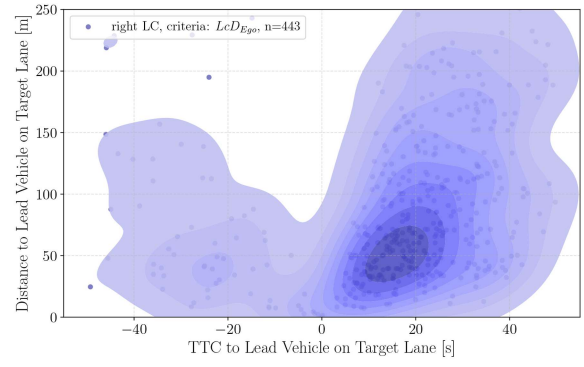
required to decelerate by more than 11.7 m/s^2 , as determined by the *IDM* car-following model. In reality, this corresponds to a very high deceleration, as modern emergency braking systems perform decelerations only up to 9 times m/s^2 . Despite this rather high deceleration parameter, there are still many manoeuvres where the lag gap is not accepted. This means that the *IDM* car-following model would suggest an even higher deceleration for the lag vehicle if the ego vehicle changes lanes. In this study, we did not explicitly analyse critical situations, but no events with such high decelerations were observed in the data.

For the *LMRS* model, the gap acceptance criterion toward the lag vehicle on the target lane is most often unmet for left lane changes, whereas for right lane changes, an insufficient desire to change lane is the primary reason for not suggesting a lane change. For the *MOBIL* model, the desire to change lanes is often too low, at least for parameter sets 3, 4, and 5, which prevents the execution of a lane change. This applies to both lane-change directions.

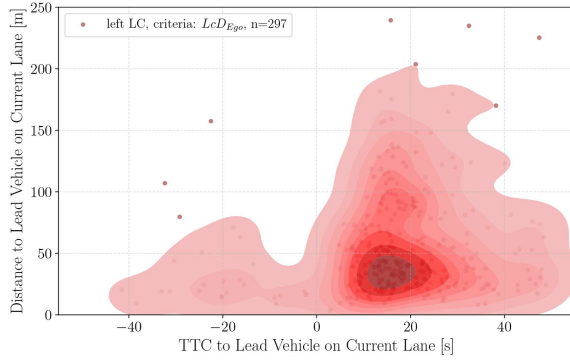
To obtain an intuition about those False-Negative situations and to foster an understanding of human lane-change behaviour, Figure 5 presents a scatter plot with a corresponding kernel-density estimation for the results of parameter set 5. The x-axis shows the time-to-collision (TTC), while the y-axis shows the distance to a surrounding vehicle, which mainly affects lane-change decision-making. Note that negative TTC values do not hold relevance in a traditional safety analysis context. A negative value does not indicate a past crash but rather that the leading vehicle is moving faster than the following vehicle. In such cases, no future conflict point exists, resulting in a negative TTC value. However, in this study, these values are included to illustrate relative vehicle interactions. For gap acceptance, the lag vehicle on the target lane is most important, as shown in the results in Table 6. For the desire to change lane, the lead vehicle on the



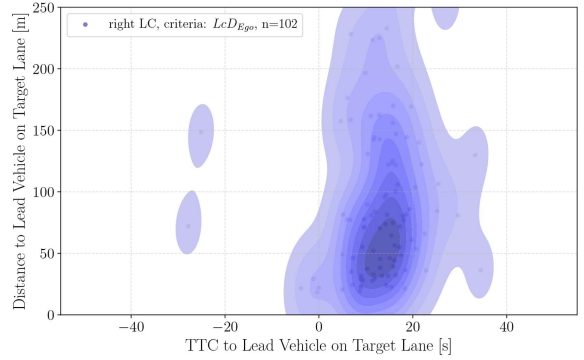
(a) LMRS, FN left LC, distance and TTC to left lag* vehicle



(b) LMRS, FN right LC, distance and TTC to right lead* vehicle



(c) MOBIL, FN left LC, distance and TTC to lead vehicle



(d) MOBIL, FN right LC, distance and TTC to right lead* vehicle

Figure 5. Kernel-density estimation plots for False-Negative lane-change manoeuvres, showing time-to-collision and distance to surrounding vehicles for features related to gap acceptance and the desire to change lane.

current lane plays a key role for left lane changes, whereas the lead vehicle on the target lane strongly impacts lane change decision-making toward the right lane.

Figure 5(a) illustrates the gap acceptance feature values for False-Negative lane changes at the time of lane change initiation. In most of these cases, the TTC between the lane-changing vehicle and the lag vehicle on the target lane ranged from 5 to 10 seconds, while the distances between the vehicles' front and rear bumpers were typically between 30 and 70 meters. These scenarios often occur in dense traffic, where both myopic and anticipatory driving behaviours were observed. Notably, some large gaps (>150 m) toward the lag vehicle were not accepted. This could only be observed for the *LMRS*, as the desire to change lane also affects gap acceptance characteristics (see formula 5). Hence, even with a high b_{safe} parameter value (as in PS5), a low desire to change lane to the left led to rejecting large gaps on the target lane. For left lane changes not correctly suggested by the *MOBIL* model, we examined feature values related to the lead vehicle on the current lane. In a few rare cases, a negative TTC was observed. This indicates that the lead vehicle was faster than the lane-changing vehicle, which most likely lowered the desire to change lane. In the majority of the manoeuvres, however, the lane-changing vehicle was faster and rather close to the lead vehicle, which most likely increased the desire to change lane. Two main reasons might explain why the desire to change lanes did not exceed the changing threshold Δa : (i) the presence of a close lag vehicle on the target lane that would be negatively impacted, or (ii) a close lead vehicle on the target lane that could hinder the ego vehicle's ability to accelerate after the lane change. The primary reason for not executing the lane change likely lies with the lead vehicle on the target lane, as the results obtained from parameter set 5 show only a few cases where the gap toward the lag vehicle was not accepted. This suggests that the *MOBIL* model struggles to anticipate the behaviour of surrounding vehicles, as these lane changes were executed in reality and are not observed as issues in the *LMRS* model. For right lane changes, feature values

related to the lead vehicle on the right lane were analysed for both the *LMRS* and *MOBIL* models. The results show clear differences between the models. In some cases, the *LMRS* model does not suggest a lane change to the right, even when the lead vehicle on the right lane is relatively far away and drives at a higher speed. Such situations are not observed for the *MOBIL* model, as the keep-right parameter a_{bias} provides sufficient motivation to change lanes. For the *LMRS* model, the keep-right bias is set to zero if the anticipated speed on the current lane is higher than on the right lane. Otherwise, the bias equals the changing threshold d_{free} , which was significantly reduced during optimisation to improve the performance for left lane changes. In most cases, right lane changes are not suggested when the lead vehicle on the target lane was close and travelling at a lower speed. This is intuitive, as changing to the right lane may not improve driving conditions under these circumstances. Additional analysis revealed that these situations most frequently occurred in congested traffic. Furthermore, some manoeuvres deviated from rational patterns, such as vehicles briefly changing to the right lane and then returning to the left after only a few seconds. These findings highlight the heterogeneity and complexity of human driver behaviour.

Having analysed False-Negative in very detail, we now focus on analysing True-Positive situations, where the models correctly predicted a lane change at the time of its initialisation. As shown in Table 5, the optimised model parameters not only increase the ratio of True-Positives but also result in a higher number of False-Positives, particularly for keep-lane situations extracted 2 seconds before the lane change initialisation. This suggests a time offset between when the models first predict a lane change and when the actual lane change is executed. Figure 6 presents the distribution of these time offsets for parameter sets 2 and 3, separated by lane-change models and distinguishing between left and right lane changes. The histogram includes offsets smaller than 4 seconds, which is equal to ΔT_{FoV} , which we defined during data processing to filter out lane changes recorded in the very beginning and end of the drone's FoV. This means that in the histogram situations are not considered, where the lane-change model would have most likely continuously executed a lane change from the very beginning. Those situations with a time offset greater than 4 s mainly occurred during free-flow traffic, where a relatively distant leading vehicle yielded a certain desire to

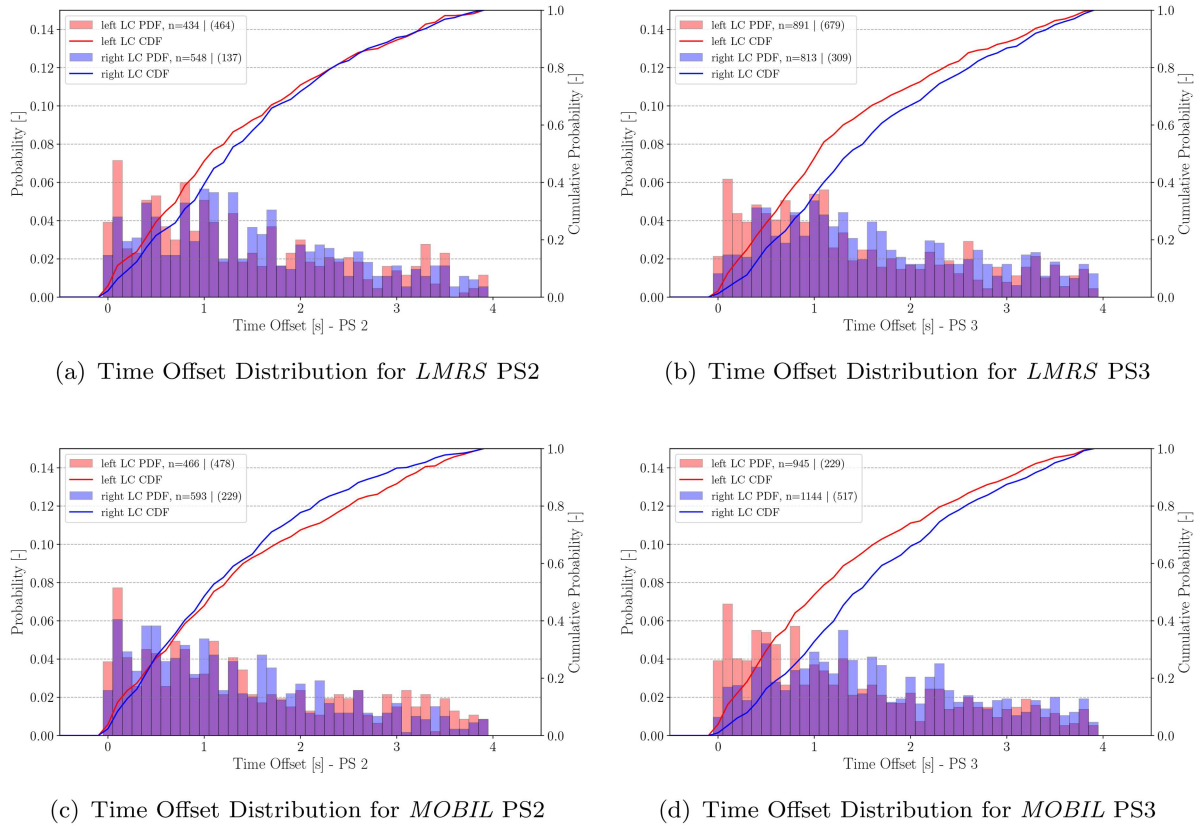


Figure 6. Time-offset distribution for true-positive lane-change manoeuvres with $t_{offset} < 4s$, numbers in brackets refer to situations where $t_{offset} > 4s$.

change lane, while there, where hardly any vehicles on the target lane. The number of those situations is indicated in brackets in the figure legend.

The distributions in Figure 6 show comparable results among the two investigated lane-change models. Comparing parameter sets PS2 (Figures 6a, c) and PS3 (Figures 6b, d), minimal improvements are observed for left lane changes, while a noticeable increase in time offset is found for right lane changes with PS3. For both *MOBIL* and *LMRS*, the number of True-Positives significantly increases when PS 3 is applied compared to PS2. However, both models face challenges in accurately replicating the time- and space-dependent microscopic behaviour of these lane change manoeuvres. In addition, the F1 values from Table 5 cannot be directly associated with lower or higher time and location offsets. While PS 3 captures a larger number of manoeuvres correctly than PS2, the corresponding time offsets relative to the actual lane change start do not differ substantially.

5. Discussion

The analyses in the previous results section revealed key insights, which are discussed in the following.

The first main finding relates to the performance of the investigated optimisation algorithms. For calibrating car-following models, GA proved to yield satisfactory results (Ciuffo, Punzo, and Montanino 2012; Kesting and Treiber 2008; Punzo, Ciuffo, and Montanino 2012), which does not necessarily hold for lane-change parameter calibration – at least not for the investigated models. For both models, BO was found to be more effective than the GA in searching the parameter space and returning optimised model parameters. The heuristic optimisation process in GA appears suboptimal for generating and sampling new parameter sets compared to BO. While this is less critical for simpler models such as *MOBIL*, the GA appears to be less suitable for optimising more complex models such as *LMRS*. The results also demonstrate that SGD performs well, given that the initialised parameter values are within a plausible range, enabling the algorithm to compute reasonable gradients and therefore to converge to a local minima. However, we found that parametrisation of SGD is rather challenging. A smaller mini-batch size introduced excessive stochasticity, preventing convergence, while careful tuning of learning rates was required to avoid gradient-related issues, such as exploding or vanishing gradients. To address these issues, applying the adaptive momentum estimation (ADAM) optimizer instead of SGD was also tested. ADAM is an extension of SGD that automatically adapts learning rates for individual parameters throughout the optimisation process. Despite this adaptive approach, ADAM did not lead to satisfactory results for the investigated models.

A second key aspect concerns differences in performance metrics compared to previous studies, as the optimised model parameters in this study yielded higher performance metrics for the *MOBIL* model compared to those presented by Ali, Zheng, and Bliemer (2023). For instance, the authors report an F1 score slightly above 50% for their 2:1 control case (lane-keeping to lane-changing situations). In contrast, the best F1 score achieved in this study was 64.6% (PS4, averaged for left and right lane changes). While incorporating lane-keeping situations from non-lane-changing trajectories positively impacts the F1 score, the improvements can also be attributed to the extensive optimisation procedure conducted in this paper. Note that this study did not account for different driver styles. Optimising and applying distinct parameter sets tailored to varying driving behaviours could potentially improve performance indicators. However, this approach requires a rigorous classification of driver styles for both lane-changing and lane-keeping scenarios, categorising them into groups such as aggressive versus non-aggressive or anticipatory versus myopic drivers.

The third key finding of this study offers critical insights into lane-change modelling. The conducted analyses indicate that the *LMRS* model effectively captures the desire to change lane, whereas the *MOBIL* model struggles in this regard. Specifically, the *LMRS* accurately models the desire to gain speed by considering the anticipated speeds on both the current and target lanes. In contrast, the *MOBIL* model relies on the *IDM* to compute the desire to change lane (see formula (6)), which limits its ability to reflect anticipatory driving behaviour. As a result, close but faster-moving vehicles on the target lane are often perceived as obstacles, an assumption that does not necessarily align with real-world driver behaviour. Despite its strengths in capturing the desire to change lane, the *LMRS* struggles to realistically model gap acceptance. This shortcoming arises because the model links gap acceptance directly to the desire to

change lane. While this connection appears intuitive – since a higher desire to change lane naturally leads to accepting smaller gaps – a relatively low desire can overly restrict gap acceptance. Refining this aspect by reducing the influence of a very low lane change desire on gap acceptance could enable more realistic lane-change modelling and yield more meaningful values for the required deceleration parameter b_{safe} .

Conversely, the *MOBIL* model performs better in capturing gap acceptance but occasionally fails to execute lane changes, even when a high safe deceleration parameter is specified. Such extreme decelerations were not observed in the empirical data, highlighting a discrepancy between the longitudinal behaviour predicted by the *IDM* and actual driver behaviour. Additionally, during optimisation, the *IDM* time gap parameter T frequently approaches the lower constraint boundary of 0.5 s. While this leads to improved lane-change modelling characteristics, applying such a low parameter value within a simulation environment would likely result in unrealistic longitudinal behaviour and traffic flow characteristics.

The results suggest a substantial dependency on the used car-following model. As varying car-following models were not investigated, it remains unclear how the results would differ when integrating alternative models, particularly those capable of better representing anticipatory behaviour.

To address these limitations in the future, it might also be required to consider alternative modelling approaches. One approach could involve defining a separate set of car-following parameters specifically for the lane-change model. However, this would require incorporating relaxation behaviour into the car-following model used to prevent abrupt decelerations as a result of lane changes suggested under a different parameter set. Alternatively, structural refinements to the lane-change models themselves may offer a more robust solution. For instance, in the *LMRS*, revising the link between lane-change desire and gap acceptance, refining the desire to keep-right calculation, or introducing an additional and independent keep-right bias parameter – similar to *MOBIL* – could improve performance.

These observations are based on a detailed microscopic evaluation, which remains rare in previous research. The findings underscore the importance of microscopic analyses in revealing the strengths and limitations of existing models, while we acknowledge the need to consider both microscopic and macroscopic lane-change indicators for a comprehensive evaluation.

6. Conclusion

This study focuses on the microscopic evaluation of existing lane-change models using naturalistic vehicle trajectory data. It introduces novel data processing methods and explores trajectory-based parameter optimisation by benchmarking three distinct optimisation algorithms (stochastic gradient descent, the genetic algorithm and Bayesian Optimisation). These efforts form the foundation for a detailed micro-level assessment of two widely used rule-based models – *LMRS* and *MOBIL* – which constitute the main contribution of this study. Additionally, the analysis provides valuable insights into human lane-change behaviour.

The results indicate that BO outperforms GA. Although GA has often been applied for trajectory-based car-following model calibration, it proved less effective for trajectory-based lane-change model calibration. SGD yielded effective optimisation results when initialised with reasonable parameter values. However, proper parameter tuning for the investigated models has proven to be challenging, and SGD is also prone to overfitting.

The investigated lane-change models showed comparable overall performance metrics. Detailed analyses, however, revealed that the *LMRS* was able to capture anticipatory human driver behaviour, which is not really the case for the *MOBIL* model. The logic of the *IDM* car-following model has a substantial impact on the *MOBIL* model, whereas its influence on the *LMRS* is less pronounced. Analyses of the gap-acceptance criterion for both models revealed a discrepancy between the empirically observed behaviour of the surrounding vehicles and the longitudinal behaviour predicted by the *IDM*. Despite this, the *MOBIL* model outperformed the *LMRS* in terms of gap acceptance and its ability to capture the desire to keep-right. Addressing these limitations in the *LMRS* could enhance its performance. First, the influence of the desire to change lane on gap acceptance should be constrained. Second, introducing a dedicated and independent parameter to account for potential keep-right rules or adapting the method of computing the desire to keep-right could further improve the model's capabilities.

The conducted analyses are based on the highD dataset, which covers a diverse range of traffic states. This diversity enhances the transferability of the identified model limitations, which appear to be largely related to the structural properties of the models rather than solely to the specific dataset used. Nevertheless, we acknowledge that the magnitude of microscopic performance and the observed limitations may vary across datasets owing to cultural and contextual differences in driving behaviour. Therefore, extending such analyses to urban environments and to datasets from different regions would be a valuable direction for future research. Moreover, the analyses focus on discretionary lane changes on open motorway segments, which introduces certain limitations because the upstream and downstream route context is not available in the dataset. Consequently, long-term route-related influences on lane-change decisions cannot be considered. This may affect the results of both models, particularly for right lane changes, as influences from still distant off-ramps are not captured. As the *LMRS* explicitly models route-related influences, it may be more sensitive to the absence of such long-term effects.

In addition, the analyses focused on the microscopic evaluation of existing lane-change models, while macroscopic indicators such as lane-change rates or traffic flow characteristics were not considered. Future work should therefore aim at a comprehensive evaluation that includes both microscopic indicators and macroscopic traffic flow properties.

As the analyses revealed relevant limitations, it may also be necessary to evaluate other lane-change models through microscopic analysis. This could include testing whether the proposed optimisation framework is applicable to other lane-change models. Finally, we underscore the importance of microscopic analyses in revealing potential model limitations, but also to gain further insights into human driver behaviour, which lays the basis for future driver behaviour modelling.

Author contributions

CRedit: **Felix Hofbaur:** Writing – Original Draft, Conceptualisation, Methodology, Software, Data curation, Formal analysis, Visualisation. **Simeon Calvert:** Writing – Review & Editing, Methodology, Validation, Supervision. **Martin Fellendorf:** Writing – Review & Editing, Methodology, Supervision, and Funding acquisition. **Hans van Lint:** Writing – Review & Editing, Methodology, Conceptualisation, Supervision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Usage of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to obtain suggestions for improving readability. After using this tool, the author reviewed and edited the content as needed and took full responsibility for the content of the published article.

Funding

This work received funding from FFG, the Austrian Research Promotion Agency (Grant No. 891143) in the course of the TRIDENT-project.

ORCID

Felix Hofbaur  <http://orcid.org/0000-0002-1566-6935>

References

- Ahmed, K., M. E. Ben-Akiva, H. N. Koutsopoulos, and R. G. Mishalani. 1996. "Models of Freeway Lane Changing and Gap Acceptance Behavior." In *Transportation and Traffic Theory*. Vol. 13. 501–515, Lyon: Pergamon.
- Ahn, S., and M. J. Cassidy. 2007. "Freeway Traffic Oscillations and Vehicle Lane- Change Maneuvers." In *Proceedings of the 17th International Symposium on Transportation 27 and Traffic Theory*. Elsevier. http://sueahn.weebly.com/uploads/5/0/5/2/5052784/2007_isttt.pdf

- Ali, Y., Z. Zheng, and M. C. Bliemer. 2023. "Calibrating Lane-Changing Models: Two Data-Related Issues and a General Method to Extract Appropriate Data." *Transportation Research Part C: Emerging Technologies* 152: 104182. <https://doi.org/10.1016/j.trc.2023.104182>.
- Ali, Y., Z. Zheng, Md. M. Haque, and M. Wang. 2019. "A Game Theory-Based Approach for Modelling Mandatory Lane-Changing Behaviour in a Connected Environment." *Transportation Research Part C: Emerging Technologies* 106: 220–242. <https://doi.org/10.1016/j.trc.2019.07.011>.
- Ali, Y., Z. Zheng, M. Md. Haque, M. Yildirimoglu, and S. Washington. 2021. "CLACD: A Complete LANE-Changing Decision Modeling Framework for the Connected and Traditional Environments." *Transportation Research Part C: Emerging Technologies* 128: 103162. <https://doi.org/10.1016/j.trc.2021.103162>.
- Brackstone, M., and M. McDonald. 1999. "Car-Following: A Historical Review." *Transportation Research Part F: Traffic Psychology and Behaviour* 2 (4): 181–196. [https://doi.org/10.1016/S1369-8478\(00\)00005-X](https://doi.org/10.1016/S1369-8478(00)00005-X).
- Brackstone, M., M. McDonald, and J. Wu. 1998. "Lane Changing on the Motorway: Factors Affecting Its Occurrence, and Their Implications." In *Ninth International Conference on Road Transport Information and Control* 160–164. IEE.
- Brackstone, M., B. Sultan, and M. McDonald. 2002. "Motorway Driver Behaviour: Studies on Car Following." *Transportation Research Part F: Traffic Psychology and Behaviour* 5 (1): 31–46. [https://doi.org/10.1016/S1369-8478\(02\)00004-9](https://doi.org/10.1016/S1369-8478(02)00004-9).
- Calvert, S. C., W. J. Schakel, and J. W. C. van Lint. 2017. "Will Automated Vehicles Negatively Impact Traffic Flow?" *Journal of Advanced Transportation* (1): 3082781. <https://doi.org/10.1155/2017/3082781>.
- Chen, K., V. L. Knoop, P. Liu, Z. Li, and Y. Wang. 2023. "How Gaps Are Created during Anticipation of Lane Changes." *Transportmetrica B: Transport Dynamics* 11(1): 958–978. <https://doi.org/10.1080/21680566.2022.2152129>.
- Choudhury, C. F. (2007) Modeling Driving Decisions with Latent Plans. PhD Thesis. Doctoral dissertation. <http://hdl.handle.net/1721.1/42220>.
- Choudhury, C., M. E. Ben-Akiva, T. Toledo, R. Anita, and G. Lee. 2007. "State Dependence in Lane Changing Models 1 PAPER: 196 State Dependence in Lane Changing Models." In *Proceedings of the 17th International Symposium on Transportation and Traffic Theory*. London: Elsevier.
- Christianson, R. B., and R. B. Gramacy. 2024. "Robust Expected Improvement for Bayesian Optimization." *IIEE Transactions* 56(12): 1294–1306. <https://doi.org/10.1080/24725854.2023.2275166>.
- Ciuffo, B., V. Punzo, and M. Montanino. 2012. *The calibration of traffic simulation models: Report on the assessment of different goodness of fit measures and optimization algorithms*. Luxembourg: Publications Office.
- Deisenroth, M. P., A. A. Faisal, and C. S. Ong. 2020. *Mathematics for machine learning*. Cambridge and New York, NY and Port Melbourne and New Delhi and Singapore: Cambridge University Press.
- Durrani, U., C. Lee, and H. Maoh. 2016. "Calibrating the Wiedemann'S Vehicle-Following Model Using Mixed Vehicle-Pair Interactions." *Transportation Research Part C: Emerging Technologies* 67: 227–242. <https://doi.org/10.1016/j.trc.2016.02.012>.
- Frazier, P. I. (2018) A Tutorial on Bayesian Optimization. <https://arxiv.org/abs/1807.02811>.
- Gao, J., Y. L. Murphey, J. Yi, and H. Zhu. 2022. "A Data-Driven Lane-Changing Behavior Detection System Based on Sequence Learning." *Transportmetrica B: Transport Dynamics* 10(1): 831–848. <https://doi.org/10.1080/21680566.2020.1782786>.
- Gipps, P. G. 1986. "A Model for the Structure of Lane-Changing Decisions." *Transportation Research Part B: Methodological* 20 (5): 403–414. [https://doi.org/10.1016/0191-2615\(86\)90012-3](https://doi.org/10.1016/0191-2615(86)90012-3).
- Goldberg, D. E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Repr. with corr ed. Reading, Mass.: Addison-Wesley.
- Jin, C.-J., V. L. Knoop, D. Li, L.-Y. Meng, and H. Wang. 2019. "Discretionary Lane-Changing Behavior: Empirical Validation for One Realistic Rule-Based Model." *Transportmetrica A: Transport Science* 15 (2): 244–262. <https://doi.org/10.1080/23249935.2018.1464526>.
- Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization* 13: 455–492. <https://doi.org/10.1023/A:1008306431147>.
- Kesting, A., and M. Treiber. 2008. "Calibrating Car-Following Models By Using Trajectory Data." *Transportation Research Record* 2088: 148–156. <https://doi.org/10.3141/2088-16>.
- Kesting, A., M. Treiber, and D. Helbing. 2007. "General Lane-Changing Model MOBIL for Car-Following Models." *Transportation Research Record: Journal of the Transportation Research Board* 1999 (1): 86–94. <https://doi.org/10.3141/1999-10>.
- Knoop, V. L. and C. Buisson. 2014. "Calibration and Validation of Probabilistic Discretionary Lane-Change Models." In *IEEE Transactions on Intelligent Transportation Systems* 1–10. IEEE. <https://doi.org/10.1109/TITS.2014.2340434>.
- Krajewski, R., J. Bock, L. Kloeker, and L. Eckstein. 2018. "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems." In *21st International Conference on Intelligent Transportation Systems (ITSC)* 2118–2125. IEEE. <https://doi.org/10.1109/ITSC.2018.8569552>.
- Leurent, E. 2018. An Environment for Autonomous Driving Decision-Making. <https://github.com/eleurent/highway-env>.
- Li, G., Z. Yang, Y. Pan, and J. Ma. 2023. "Analysing and Modelling of Discretionary Lane Change Duration Considering Driver Heterogeneity." *Transportmetrica B: Transport Dynamics* 11(1): 343–360. <https://doi.org/10.1080/21680566.2022.2067599>.
- Li, L., R. Jiang, Z. He, X. Chen, and X. Zhou. 2020. "Trajectory Data-Based Traffic Flow Studies: A Revisit." *Transportation Research Part C: Emerging Technologies* 114: 225–240. <https://doi.org/10.1016/j.trc.2020.02.016>.
- Li, J., J. Wang, X. Long, H. Zhang, and M. Xie. 2025. "Prediction of Following Vehicles' Longitudinal Game Behaviour in Lane-Changing Game Scenarios Considering Driving Style." *Transportmetrica B: Transport Dynamics* 13(1): 2553218. <https://doi.org/10.1080/21680566.2025.2553218>.

- Ma, C., and D. Li. 2023. "A Review of Vehicle Lane Change Research." *Physica A: Statistical Mechanics and its Applications* 626: 2023. 129060. [10.1016/j.physa.2023.129060](https://doi.org/10.1016/j.physa.2023.129060).
- pandas. 2020. pandas-dev/pandas: Pandas. <https://doi.org/10.5281/zenodo.3509134>.
- Pang, M.-Y., B. Jia, D.-F. Xie, and X.-G. Li. 2020. "A Probability Lane-Changing Model Considering Memory Effect and Driver Heterogeneity." *Transportmetrica B: Transport Dynamics* 8(1): 72–89. <https://doi.org/10.1080/21680566.2020.1715310>.
- Punzo, V., B. Ciuffo, and M. Montanino. 2012. "Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted?" *Transportation Research Record: Journal of the Transportation Research Board* 2315 (1): 11–24. <https://doi.org/10.3141/2315-02>.
- Punzo, V., M. Montanino, and B. Ciuffo. 2015. "Do We Really Need to Calibrate All the Parameters? Variance-Based Sensitivity Analysis to Simplify Microscopic Traffic Flow Models." *IEEE Transactions on Intelligent Transportation Systems* 16 (1): 184–193. <https://doi.org/10.1109/TITS.2014.2331453>.
- Punzo, V., Z. Zheng, and M. Montanino. 2021. "About Calibration of Car-Following Dynamics of Automated and Human-Driven Vehicles: Methodology, Guidelines and Codes." *Transportation Research Part C: Emerging Technologies* 128: 103165. <https://doi.org/10.1016/j.trc.2021.103165>.
- Rahman, M., M. Chowdhury, Y. Xie, and Y. He. 2013. "Review of Microscopic Lane-Changing Models and Future Research Opportunities." *IEEE Transactions on Intelligent Transportation Systems* 14 (4): 1942–1956. <https://doi.org/10.1109/TITS.2013.2272074>.
- Ruder, S. 2016. An overview of gradient descent optimization algorithms. CoRR abs/1609.04747. <http://arxiv.org/abs/1609.04747>.
- Schakel, W. J., V. L. Knoop, and B. van Arem. 2012. "Integrated Lane Change Model with Relaxation and Synchronization." *Transportation Research Record: Journal of the Transportation Research Board* 2316 (1): 47–57. <https://doi.org/10.3141/2316-06>.
- Schakel, W., V., Knoop, M., Keyvan-Ekbatani, and H., van Lint. 2023. Social Interactions on Multi-Lane Motorways: Towards a Theory of Impacts. <https://doi.org/10.20944/preprints202305.0193.v1>.
- Siebinga, O. 2021. "TraViA: A Traffic Data Visualization and Annotation Tool in Python." *Journal of Open Source Software* 6 (65): 3607. <https://doi.org/10.21105/joss.03607>.
- Talebpour, A., H. S. Mahmassani, and S. H. Hamdar. 2015. "Modeling Lane-Changing Behavior in a Connected Environment: A Game Theory Approach." *Transportation Research Procedia* 7: 420–440. <https://doi.org/10.1016/j.trpro.2015.06.022>.
- Toledo, T., and D. Zohar. 2007. "Modeling Duration of Lane Changes." *Transportation Research Record* 1999(1): 71–78. <https://doi.org/10.3141/1999-08>.
- Toledo, T., H. N. Koutsopoulos, and M. E. Ben-Akiva. 2003. "Modeling Integrated Lane-Changing Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1857 (1): 30–38. <https://doi.org/10.3141/1857-04>.
- Treiber, M., and A. Kesting. 2012. *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg. <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1590625>.
- Treiber, M., A. Hennecke, and D. Helbing. 2000. "Congested Traffic States in Empirical Observations and Microscopic Simulations." *Physical Review E* 62(2): 1805–1824. <https://doi.org/10.1103/physreve.62.1805>.
- Treiber, M., A. Kesting, and D. Helbing. 2006. "Delays, Inaccuracies and Anticipation in Microscopic Traffic Models." *Physica A: Statistical Mechanics and its Applications* 360(1): 71–88. <https://doi.org/10.1016/j.physa.2005.05.001>.
- Wang, Y., Y. Xiong, H. Hu, Y. Niu, S. shah, Q. Li, and L. Yang. 2026. "CHSAC: Reinforcement Learning-Based Lane-Changing Decision-Making Integrating Cognitive Hierarchy and Risk Awareness." *Transportmetrica B: Transport Dynamics* 14(1): 2677019. <https://doi.org/10.1080/21680566.2026.2677019>.
- Xie, D.-F., Z.-Z. Fang, B. Jia, and Z. He. 2019. "A Data-Driven Lane-Changing Model Based on Deep Learning." *Transportation Research Part C: Emerging Technologies* 106: 41–60. <https://doi.org/10.1016/j.trc.2019.07.002>.
- Zhang, C., and L. Sun. 2024. "Bayesian Calibration of the Intelligent Driver Model." *IEEE Transactions on Intelligent Transportation Systems* 25(8): 1–13. <https://doi.org/10.1109/tits.2024.3354102>.
- Zhang, H., T. Shao, P. Ranjitkar, Y. Wang, J. Li, Y. Cao, and Y. Huo. 2026. "Lane-Change Trajectory Prediction Model Based on Multi-Modal Fusion Input." *Transportmetrica B: Transport Dynamics* 14(1): 2663848. <https://doi.org/10.1080/21680566.2026.2663848>.
- Zheng, Z. 2014. "Recent Developments and Research Needs in Modeling Lane Changing." *Transportation Research Part B: Methodological* 60: 16–32. <https://doi.org/10.1016/j.trb.2013.11.009>.
- Zheng, Z., S. Ahn, and C. M. Monsere. 2010. "Impact of Traffic Oscillations on Freeway Crash Occurrences." *Accident; analysis and prevention* 42 (2): 626–636. <https://doi.org/10.1016/j.aap.2009.10.009>.

Appendices

Appendix A. Distribution of initial and optimised parameter values for the MOBIL model. n (optimisation runs) = 150

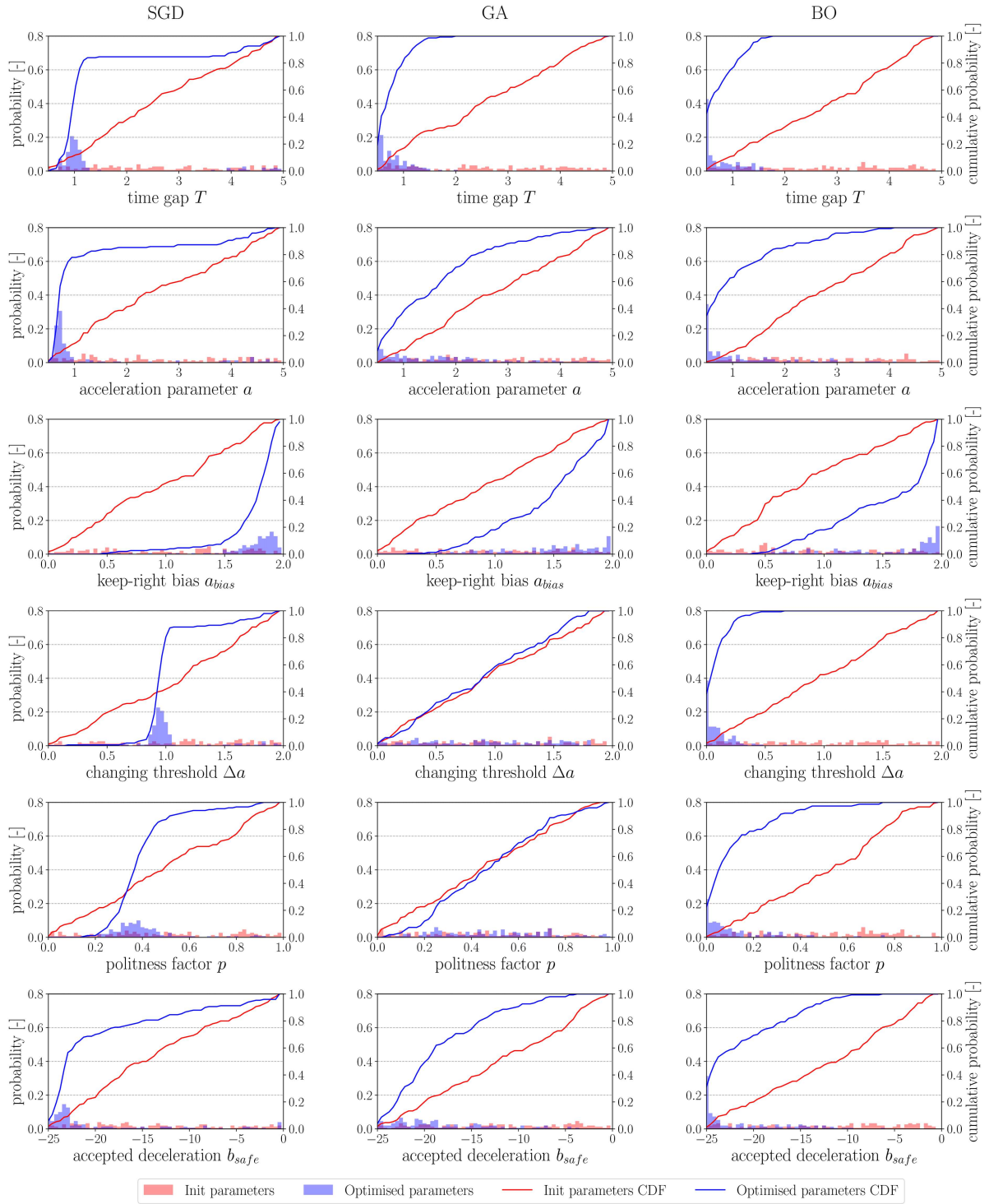


Figure A1. Distribution of initial and optimised parameter values for MOBIL model.

Appendix B. Distribution of initial and optimised parameter values for the LMRS model.
 n (optimisation runs) = 150

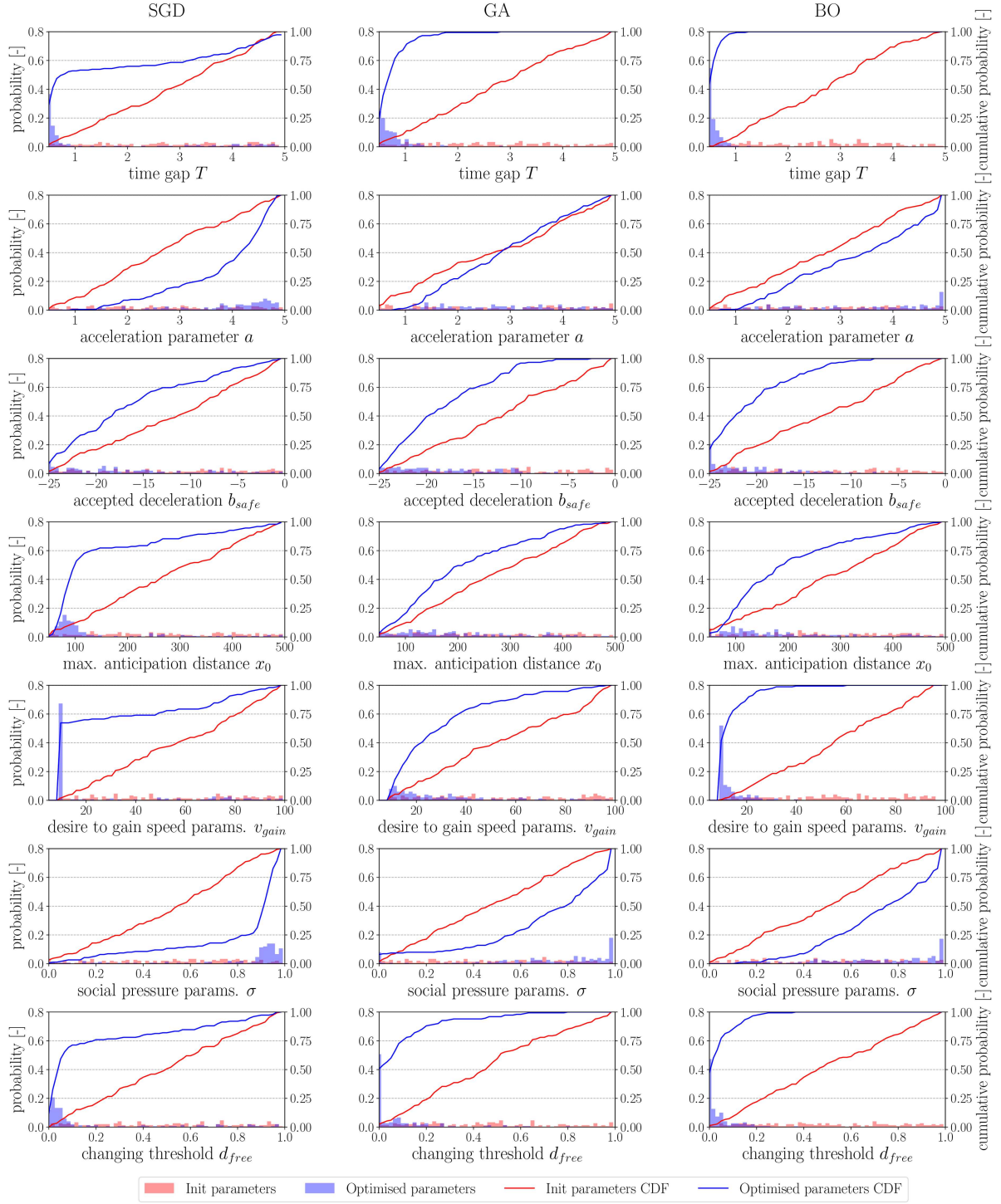


Figure B1. Distribution of initial and optimised parameter values for the LMRS model.