Performance comparison of lane-changing models for merging scenarios in traffic simulation for driving simulators

Albert Solernou[1], Evangelos Paschalidis[1], Mohamed Hasan[2], He Wang[2], Gustav Markkula[1], and Richard Romano[1]

(1) Institute for Transport Studies, University of Leeds, United Kingdom, e-mail: {a.solernou, e.paschalidis, g.markkula, r.romano}@leeds.ac.uk
(2) School of Computing, University of Leeds, United Kingdom, e-mail: {m.hasan, h.e.wang}@leeds.ac.uk

Abstract - *Virtual intelligent traffic is essential for an immersive experience in a driving simulator, and crucial for obtaining reliable results in complex scenarios. However, current research has not focused on actors with human-like behaviour in driver behaviour models. The current paper presents results from a series of game theory based lane-changing models specifically designed for motorway merging situations. Different models were implemented in SmartActors, a newly developed microscopic traffic simulation platform and were tested in two different scenarios, one of which was a simulated road layout of the I-80 motorway, for which real traffic data is available from the NGSIM project. The results from the base merge scenario showed that trade-offs occurred between traffic and safety performance of the models as models that let to higher traffic throughput also let to higher number of crashes and unsafe merge behaviour. Results from the I-80 scenario showed that the road layout may affect the performance of the models, when compared to a base scenario with a shorter acceleration lane. As a conclusion, game theory models can be used for microsimulation applications, although in some cases additional restrictions to their original specification may need to be implemented to ensure safe performance.*

Keywords: *game theory, merging, mircosimulation, human-like*

Introduction

Virtual intelligent traffic is essential for an immersive experience in a driving simulator, and crucial for obtaining reliable results in complex scenarios. While microscopic traffic simulation has proven crucial for applications related to congestion analysis and policy making, commercial implementations do not focus on actors with human-like behaviour even though existing literature (Zheng, 2014) has highlighted the importance of capturing the accuracy of these models. A very extended approach is to use an engineering model, in the terms of Saifuzzaman and Zheng, 2014, determining the laws of motion of a vehicle using a function that depends of some stimulus. Alternatively, a human-based model would define an driving algorithm incorporating various human factors. However, most of the parameters that would be involved in this latter approach remain hidden or unobservable, thus hindering the development of new models.

In this regard, the game theory (GT) framework provides a fundamental framework for modelling human interactions, thus being a potential solution to providing responsive traffic in a simulator. The current paper aims to evaluate the suitability of a number of existing game theory approaches for creating ambient traffic. More specifically, our focus is on on-ramp merging models, in which a vehicle has to perform a mandatory lane change within the limitation of the acceleration lane length. Motorway merging is regarded as a major source of conflicts and breakdowns (Daamen, Loot, and Hoogendoorn, 2010, Liu and Hyman, 2012) for which a number of GT models have been developed during the last years as shown in Ji and Levinson, 2020 recent review.

With the aim of providing ambient traffic to Driving Simulators for this highly interactive scenario, we developed a new software platform, SmartActors, that we could connect to our Driving Simulator. We implemented three different GT highway merging models of increasing complexity on SmartActors, and this paper reports the evaluation of their performance in terms of traffic throughput, safety and similarity with human driving compared to the standard lane-change engineering model Gipps, 1986.

The rest of the paper is organised as follows: in Simulation Methods we present the models and their implementation, while the following two sections detail the simulated scenarios and the analysis methods. Sec. Results and Discussion demonstrates the performance of the different models in terms of traffic, conflicts, and similarity to the human driven NGSIM dataset. Finally, we summarise the works and discuss potential future directions.

Simulation Methods

Micro-traffic simulation is composed of two basic algorithms: a car following model and a lane changing

(LC) model. The following sections present the models that were used in this work.

Car following model

In order to benchmark a number LC models we consistently used the Gipps, 1981 car following model along all this work. This is a well-established crash-avoiding model that assigns a speed to every car as the slowest one from either a free-flow or a car-following regime. In order to introduce some variability in SmartActors, every car has an "ideal speed ratio" ranging from 0.9 to 1.1 that is randomly assigned, so that their free speed aim is set to this ratio times the speed of the lane.

Lane changing models

As previously mentioned, we took the Gipps, 1986 (Gipps86) model as a baseline. Our implementation was plain, in which the merging car looks for a gap on the main lane that is defined in terms of safety, without receiving any cooperation from the vehicles on the main lane.

All the theoretical games implemented are two player games, and they are referred as $\mathcal{M}$ (merge) and $\mathcal{L}$ (lag) although sometimes the state of the preceding vehicle ($\mathcal{P}$) is taken into account. On non-repetitive games, if $\mathcal{M}$ has decided to merge or overtake but $\mathcal{L}$ overtakes it, no new games are played for the pair unless the acceleration of $\mathcal{M}$ is negative reflecting the fact that the traffic ahead was stopping the car from performing its desired action. If the outcome of a certain game is for $\mathcal{M}$ to change lane, it starts looking for a time gap of 0.4 s upstream and 0.5 s downstream. These small values were chosen to reflect the human behaviour that has been measured in on-ramp merges (Marczak, Daamen, and Buisson, 2013). In order to more strictly evaluate the models, on games that are played only once the $\mathcal{M}$ vehicle was forced to find such gap before a threshold of 5 s. The same threshold would be applied to $\mathcal{L}$, before stopping any cooperation with $\mathcal{M}$. On repetitive games, no such threshold was used, reflecting the possibility of the vehicles to change their actions. Still, if $\mathcal{M}$ would arrive at the end of the lane deciding to merge or overtake, it would perform the LC manoeuvre even if dangerously.

The games were first played once $\mathcal{M}$ reached beginning of the acceleration lane, the maximum range for cars to be engaged in a game was set to 61 m (200 feet), as it is done in Liu, et al., 2007, while vehicles outside this range are assumed to be out of the interaction range (Toledo, 2002). Values for maximum acceleration (a_M), maximum deceleration (a_m) and comfortable acceleration (a_c) were consistently set across every simulation to 3.5, -4.6 and 3.0 m/s^2 respectively.

None of the original implementation of the models models that included simulation did specify the actual manoeuvre performed by the cars to change lane. In this work, we chose to approximate it as a Bézier curve, as it has been already proposed in Chen, et al., 2013 and Korzeniowski and Ślaski, 2016. As shown in Fig. 1, we plan a manoeuvre starting at the departure point and ending at the intersection point of the target lane with a straight line L that has an

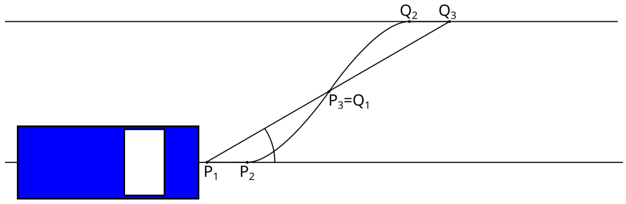

Figure 1: The lane change manoeuvre is performed as a composition of two Bézier curves P and Q, with control points P_1, P_2, P_3 and Q_1, Q_2 and Q_3.

inclination of 30^o with the current trajectory (Q_3 in the figure). Given the starting and the finishing points the planned manoeuvre is composed of two 2-degree Bézier curves. The first one has three control points: the first one at the starting of the manoeuvre, the next one 1 m behind in the current direction, and a third one halfway the previously calculated line L. The second Bézier curve starts at the end point of the first one, has a second control point 1 m before the destination, in this same direction, and the last control point at the end of the manoeuvre. The resulting piece wise curve is continuous and derivable and visually acceptable.

Liu07

The first of the three games implemented was proposed by Liu, et al., 2007 as a normal game that is played just once as soon as the players are within range. The aims of the two players are simple and intuitive: $\mathcal{M}$ aims to minimise the time spent on the acceleration lane, and $\mathcal{L}$ aims to minimise changes in speed. This is reflected in two actions per player: $\mathcal{M}$ can either change lane immediately, or wait for the next gap, while $\mathcal{L}$ can either ignore $\mathcal{M}$ (block), or yield. Finally, the game ensures that it leads to no deadlock: $\mathcal{M}$ will merge regardless of their decision if $\mathcal{L}$ does yield.

The payoffs functions of this game take as arguments the acceleration that both $\mathcal{L}$ and $\mathcal{M}$, would take in the different situations, and therefore we used similar values for the simulation. More specifically, the acceleration for $\mathcal{L}$ to block is explicitly stated in its payoff matrix, while to yield it was reduced to $0.3a_y$, and bound to a maximum of -0.5 m/s^2. Similarly, following the outcomes of the game, if $\mathcal{M}$ decides to change its acceleration is set to a_c or a_M depending on whether $\mathcal{L}$ yields or blocks, respectively. However, in the outcome of wait / block, the acceleration for $\mathcal{M}$ was chosen to be $-v/2x_{rd}$ (where x_{rd} is the remaining distance before the end of the acceleration lane) rather than the original fixed value. We also allow multiple rivals for $\mathcal{L}$ so that if a later game with a new $\mathcal{M}$ car would make $\mathcal{L}$ to yield, it would do it for the two cars, and if it this would make $\mathcal{L}$ to block, we would not allow it.

Yu18

The second game we implemented was proposed by Yu, Tseng, and Langari, 2018, and allows $\mathcal{M}$ to decide whether to merge or not, and the two players $\mathcal{M}$ and $\mathcal{L}$ to explicitly decide their acceleration. The game is played repeatedly at every 0.3 s, and in turns ($\mathcal{M}$ reacting to $\mathcal{L}$) and therefore it fully controls the two vehicles. The payoff functions involve time headway between $\mathcal{M}$ and $\mathcal{L}$ and the gap size, but not the distance to the end of the road.

In our implementation, we solve the game discretely at a resolution of 0.1 m/s^2, as suggested in the original paper. However, we used the values previously given for the a_M and a_m rather than the original values (-6 to 4 m/s^2). The model defines different aggressiveness for each vehicle, and an iterative perception model for this aggressiveness which is very interesting from a driving simulation perspective. However, our implementation was simpler in that we allow the different players to read the true aggressiveness value of their counterparts. Finally, when $\mathcal{L}$ is playing with multiple $\mathcal{M}$ vehicles, we assign to it the smallest of the chosen accelerations.

Kang20

Finally, we implemented the model of Kang and Rakha, 2020, which is a normal game in which $\mathcal{L}$ can either yield or block, and where $\mathcal{M}$ can wait, merge, or overtake. The game is played repeatedly every 0.5 s and the payoff functions are cumulatively updated with a weight factor of 1.4, which led to best results in the original work, giving more importance to the later decisions. Building on the game by Yu, Tseng, and Langari, 2018, the payoff functions account for the time headway and the time to collision, but also for the remaining length of the acceleration lane.

In our implementation, the accelerations assigned to $\mathcal{M}$ and $\mathcal{L}$ under the different set of strategies are slightly different to the ones shown in the original paper, as it omitted some details. Thus, if $\mathcal{L}$ blocks and if its ideal speed ratio is smaller than $\mathcal{P}$'s, then $\mathcal{L}$'s ideal speed ratio is increased to $\mathcal{P}$'s one. On the contrary, if $\mathcal{L}$ yields, it starts driving at Gipps' car following speed after $\mathcal{M}$. If $\mathcal{M}$ overtakes, its acceleration is set to a_M until it overtakes or reaches 1.4 times the speed of the lane. Alternatively, if $\mathcal{M}$ changes, a constant acceleration is set so that $\mathcal{M}$ targets the centre of the gap in 0.5 seconds, considering that $\mathcal{P}$ and $\mathcal{L}$ move at constant speed. This acceleration is again bound between a_m and a_M. If no $\mathcal{P}$ is present the acceleration is set to a_M. Finally, if $\mathcal{M}$ waits, a constant deceleration is applied so that it stops at the end of the lane. The accelerations for the two cars are updated every 0.5 s with the outcomes of the repeated game.

Traffic simulation

With all these elements, SmartActors run continuous simulations as a series of time steps, in which every actor would get a target acceleration within their vehicle limitations (a_m and a_M), the equations of motion integrated using a Verlet scheme, and the vehicles moved along their path by the correct amount. In order to solve the theoretical games, we used the Gambit library (McKelvey, McLennan, and Turocy, 2016) for normal games, and an efficient implementation of the backwards induction method to solve the only extended game.

Simulation Scenarios

This section presents the two different scenarios on which the lane-changing models were evaluated. On every simulation, the number of vehicles was kept constant, so that when a vehicle would reach the end of the route, a brand new vehicle would be created with an initial lane and a route assigned randomly, ready to be introduced to the simulation as soon as it was safe.

Base merge scenario

The first scenario refers to a 3-lane motorway (Fig. 2a) with an additional 40 m acceleration lane. Three different levels of traffic were considered, to study low, average and high traffic conditions. In the simulations this meant having 10, 15 or 20 vehicles occupying the road space at all times, with a target traffic speed set to 14 m/s for the motorway and to 9 m/s for the acceleration lane. While the total amount of vehicles is small, the traffic densities is significant. More specifically, the total length of road simulated is 0.2 miles, which results in traffic densities of 50, 75 and 100 vehicles per mile. The road density at capacity can be calculated using the NCHRP 825 method, as recommended by Margiotta and Washburn, 2017, and for this scenario it was found to be 66.7 vehicles per mile. Therefore, the level of traffic is suitable to stress the different models. Finally, for each one of the models and each traffic condition we ran 20 simulations, each simulation 60 minutes long. Running one-hour simulations on a Windows 10 machine with an Intel i7-6700 @ 3.40 GHz processor took 154 seconds for the Yu18 model, and less than 10 seconds for the rest of the models.

NGSIM I-80 scenario

The second scenario recreated a segment of the Interstate 80 (I-80) freeway, CA, USA (see Fig. 2b) for which vehicle data is available as part of the Next Generation SIMulation (NGSIM) project (Halkias and Colyar, 2006). The length of the road segment is approximately 500 meters (1650 feet) and it is composed of five regular lanes plus a high occupancy vehicle (HOV) lane. The full dataset consists of vehicle trajectories collected in three different time periods, namely, 4.00-4.15pm, 5.00-5.15pm, and 5.15-5.30pm. Although the NGSIM data was collected in 15-minute intervals, further inspection showed that total time in each of the three data was longer. Thus, the initial observations of all three data sets were removed to reduce the duration to 15 minutes.

In order to be able to compare the simulations with the human-driving data, we set the number of vehicles to reproduce the traffic conditions in the three data sets. Thus, the total numbers of vehicles in each time step was calculated in the raw NGSIM data sets, and a value close to the mean was later used in SmartActors, which resulted in 130, 165 and 160 vehicles respectively for the three time periods. Similarly, the target speed on the main lanes and the on-ramp traffic was set to the average speed observed in the NGSIM data sets. Specifically, the speed on the main lanes was set to 7.8, 5.5 and 4.8 m/s and to 6.2, 3.5 and 3 m/s for the acceleration lane on the 4.00-4.15pm, 5.00-5.15pm, and 5.15-5.30pm scenarios respectively. In terms of traffic density, the three scenarios have 66.3, 84.2 and 81.6 vehicles per mile respectively, again just and above the density at capacity (66.7 vehicles per mile).

As in the previous scenario, 20 simulations per model

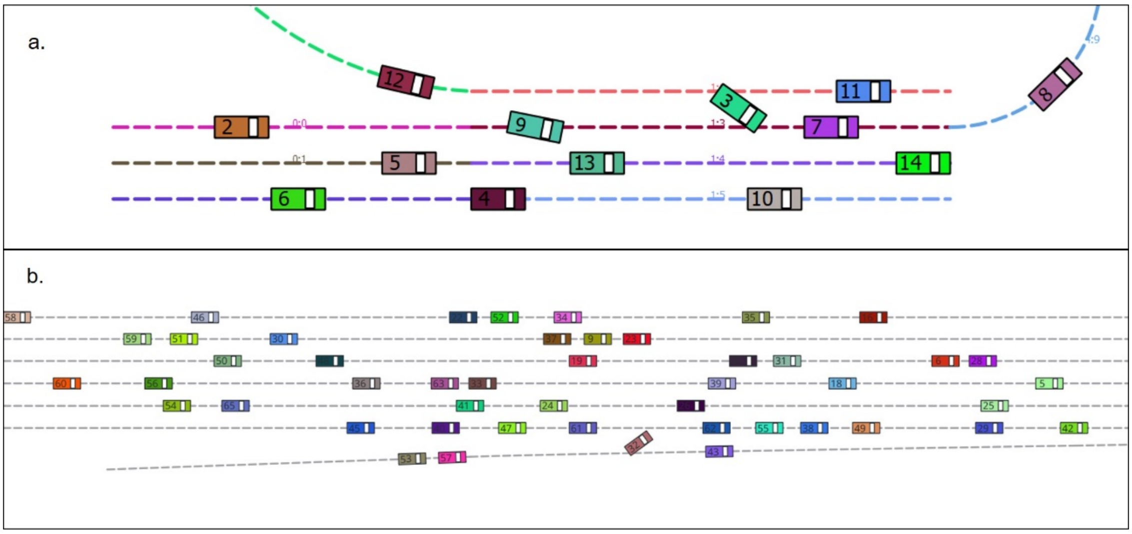

Figure 2: Simulation scenarios: (a) a 3-lane motorway scenario and (b) the I-80 scenario.

and traffic condition were also generated for the I-80 scenario. Each of the simulations was 20 minutes long, where the initial 5 minutes were used as a "warming up" period and discarded from the analysis. Running 20-minute simulations on a Windows 10 machine with an Intel i7-6700 @ 3.40 GHz processor took around 5.1 minutes for the Yu18 model, and less than 2 minutes for the rest of the models.

Analysis Methods

Traffic analysis

Traffic analysis was conducted considering the following series of variables: traffic volume, total vehicles per simulation, number of merges per simulation, time spent in simulation by the merging vehicles, time spent on the acceleration lane by the merging vehicles, and speed on the acceleration lane. For comparisons with the raw data, the Root Mean Square Error (RMSE) and Root Mean Square Percent Error (RMSPE) were used, as have been used consistently in similar studies (Choudhury, 2007). The total traffic was compared using the Geoffrey E. Havers (GEH) statistic. The traffic performance of a model is considered as sufficient when the GEH is below 5 for 85% of the simulated links.

Conflict analysis

The conflict-crash analysis was performed via the Surrogate Safety Assessment Model (SSAM) (Pu, Joshi, Energy, et al., 2008). The SSAM tool was developed by the Federal Highway Administration and is used for the detection of potential vehicle conflicts (situations close to collision). The tool uses two threshold values to define a conflict, based on time-to-collision (TTC) and post-encroachment time (PET), for which we used the default values of 1.5 s and 5 s respectively.

Moreover, SSAM is classifying conflicts as rear-end, lane-changing and crossing. Since there are no crossing manoeuvres on the simulations, and none was expected in the NGSIM dataset, these were treated as lane-changing. These lane-changing (merging) manoeuvres were the only ones considered in the analysis, so that one of the two vehicles involved in a conflict had to be placed on the acceleration lane at the beginning of the conflict event. Finally, the conflict severity analysis was conducted via TTC, PET and deceleration rate (DR) Surrogate Safety Measures.

Statistical analysis

In order to compare the different models, the variables considered for traffic and conflict analysis were analysed via statistical analysis. On those metrics with a low amount of observations, such as the total traffic/merges per simulation, counting data (number of conflicts and crashes), or non-normal data as TTC, PET and DR, the Kruskal-Wallis rank sum test was used. Given that this test provides an overall significance score, pairwise comparisons among the four models were also considered. To these, a Bonferroni correction was applied to the p-values to consider the effect of the total number of tests required. Those metrics with a high number of observations (time spend in acceleration lane, time merging vehicles spent in simulation, and speed in acceleration lane) were analysed with ANOVA. Again, a Bonferroni correction was applied to correct p-values in pairwise comparisons. In the following results sections, a summary the most relevant findings of this analysis is presented.

Results and Discussion

Base merge scenario

Traffic and conflict analysis

The results of the traffic analysis for the base merge scenarios are presented in Tab. 1. It can be seen that the GT models produced more traffic and merges compared to the Gipps86 model in all three scenarios. Moreover, Liu07 and Yu18 models resulted in considerably higher traffic performance, compared to

Table 1: Base scenario traffic analysis

		vehicles/h	merges/h	mean time in acceleration lane (s)	mean time in simulation (s)	mean speed in acceleration lane (m/s)
Low traffic (10 vehicles)	Gipps86	4882.5	1223.15	7.33	11.76	7.872
	Liu07	5128.1	1282	6.17	10.28	9.499
	Yu18	5154.6	1289.05	6.78	10.18	9.889
	Kang20	5010.9	1260.85	6.27	10.92	8.672
Average traffic (15 vehicles)	Gipps86	6396.2	1597.45	11.28	15.79	5.53
	Liu07	7613.4	1909.85	6.88	10.31	9.841
	Yu18	7695.2	1918.8	7.38	10.19	10.144
	Kang20	7192.4	1789.35	7.39	11.98	8.074
Heavy traffic (20 vehicles)	Gipps86	7276.3	1736.9	17.03	22.16	3.605
	Liu07	10029.2	2509.45	7.5	10.34	10.103
	Yu18	10147.0	2539.6	7.8	10.21	10.304
	Kang20	8656.7	2178.6	10.01	14.63	6.38

the Kang20 model in the 15 and 20-vehicle scenarios. The higher traffic throughput of the GT models is also reflected in a higher speed on the acceleration lane, and in a lower time spent on the acceleration lane and on the total simulation. The differences in the average values of these variables increases with traffic. Additional statistical analysis was conducted to investigate the significance of the differences of the traffic-related variables, for each one of the scenarios, showing that all pairwise p-values were significant ($p < 0.05$) except for the number of merges between the Liu07 and Yu18 models in the low traffic scenario.

The conflict analysis results can be read in Tab. 2. On every scenario, the GT models resulted in the occurrence of some crashes, while no crashes were observed using the Gipps86 model on any scenario. However, the latter finding should not be surprising as Gipps86 was designed to avoid any collision. Amongst the GT models, Kang20 was the one that led to significantly less crashes. Consistently, the observed lane-changing conflicts of the Liu07 and Yu18 models resulted in smaller TTC and PET values and also higher DR.

Additional statistical analysis was conducted to investigate the pairwise differences. First, on the low traffic scenario, the Liu07 and Yu18 models were not found to differ significantly on rear-end crashes and rear-end conflicts. Likewise, the Gipps86 and Kang20 models did not significantly differ regarding rear-end conflicts and DR. However, all other differences were significant. Secondly, on the average traffic scenario, conflict analysis showed that Liu07 and Yu18 models significantly differed only in the number of lane-changing conflicts. while all the several SSM metrics were significantly different. Lastly, on the high traffic scenario, the analysis revealed that all differences were significant apart from the rear-end conflicts between the Liu07 and Yu18 models. All differences regarding the SSM metrics were significant as well.

Overall, these results show a trade-off between traffic and safety performance does exist in traffic simulation. While, the Gipps86 model always resulted in lower traffic and merges, no crashes were observed during the application of the model. On the other hand, Liu07 and Yu18 models generated the highest traffic though leading to a significantly higher number of crashes, while the performance of the Kang20 model was somewhere between the aforementioned models.

I-80 simulation results

Traffic and conflict analysis

The traffic analysis results for the simulations on the I-80 scenarios presented in Tab. 3, showed different patterns when compared to the base merge scenario. In particular, the differences among different models were less obvious and there was not a dominant model in terms of traffic throughput. Further statistical analysis did not show significant differences regarding the number of merges in any of the I-80 scenarios, while in terms of total traffic, most of the significant differences were observed regarding the Yu18 model, which generated less traffic.

Breaking the results down to the three simulated traffic conditions, on the 4.00 - 4.15 pm scenario, vehicles spent less time in the acceleration lane under the Kang20 model while the highest speed was observed in the Yu18 model. Statistical analysis showed that all pairwise comparisons related to these variables were significant. However, no significant differences were observed regarding the total time spent in simulation. Similar patterns were also observed on the 5.00 - 5.15 pm scenario, regarding the traffic output and number of merges, as only the Yu18 model significantly differed compared to all other models, in terms of total traffic consistently. Regarding the remaining traffic related variables, all pairwise comparisons showed significant differences except for time spent in simulation in the Liu07 model when compared with the Gipps86 and Kang20 models. Lastly, on the 5.15 - 5.30 pm scenario, no significant differences were found regarding the number of merging manoeuvres except between the models Gipps86 and Yu18. Again, significant differences in total traffic were only found between Yu18 and the rest of the models. Finally, the analysis of the remaining traffic variables showed that every pairwise comparison was significant but total time spent in simulation for Gipps86 model with respect to the Kang20 and Liu07 models.

The different patterns found for the traffic analysis on the simulated NGSIM I-80 scenarios compared to the base scenario, are an indication that the performance of the models can be affected by external factors as the road layout or acceleration lane length. Moreover, the fact that no dominant model was found across all three scenarios, with non-significant differences in the number of merges, shows that the the differences on traffic levels affected the performance of all models equally.

On the contrary, the results of the conflict analysis

Table 2: Base scenario conflict analysis

		Conflict		Crash		TTC	PET	DR
		rear end	lane change	rear end	lane change			
Low traffic (10 vehicles)	Gipps86	7.2	3	0	0	1.352	1.305	-3.286
	Liu07	1.85	46.75	20.95	166.65	1.03	0.595	-3.847
	Yu18	0.9	18.15	20.25	185.15	0.804	0.365	-4.206
	Kang20	5.9	58.7	1.8	11.95	1.124	0.939	-3.43
Average traffic (15 vehicles)	Gipps86	27.7	95.2	0	0	1.345	1.347	-3.001
	Liu07	2.85	119.75	50.5	426.7	0.964	0.51	-3.917
	Yu18	2.25	62.8	47.2	431.95	0.843	0.348	-4.217
	Kang20	46.85	269.45	11.3	95.45	1.118	0.985	-3.16
High traffic (20 vehicles)	Gipps86	28.15	325.3	0	0	1.321	1.298	-2.348
	Liu07	5.1	25.25	86.6	735.9	0.939	0.436	-3.912
	Yu18	3.65	135.1	71.05	715.15	0.909	0.346	-4.154
	Kang20	59.85	510.2	41.75	226.45	1.144	1.12	-3.294

Table 3: I-80 scenario traffic analysis

		vehicles/h	merges/h	mean time in acceleration lane (s)	mean time in simulation (s)	mean speed in acceleration lane (m/s)
4.00 - 4.15 pm	Gipps86	1911.9	254.65	12.71	55.2	5.94
	Liu07	1905.15	257.7	14.4	54.85	6.09
	Yu18	1901.1	258.4	15.42	55.25	6.15
	Kang20	1912.3	257.95	11.98	55.04	5.8
5.00 - 5.15 pm	Gipps86	1774.2	191.95	22.71	82.57	3.377
	Liu07	1774.35	189.9	23.35	82.41	3.469
	Yu18	1739.55	187.65	27.71	87.3	3.403
	Kang20	1775.45	192.55	21.94	81.76	3.355
5.15 - 5.30 pm	Gipps86	1500.25	161.15	25.53	93.68	2.904
	Liu07	1496.55	157.7	26.72	94.36	2.989
	Yu18	1475.9	152.65	32.89	99.25	2.912
	Kang20	1498.1	157.8	24.22	93.24	2.9

displayed in Tab. 4, showed some similar patterns with the base merge scenario. However, in the I-80 scenarios, the Kang20 model performed much closer to the Gipps86 model as a few crashes were only observed in the 5.15 - 5.30 pm scenario. Moreover, these two models also resulted in fewer conflicts when compared to the Liu07 and Yu18 models across all scenarios. This unsafer behaviour of the two latter models is also reflected in the TTC, PET and DR values. Furthermore, statistical analysis showed that Gipps86 and Kang20 models never differed in terms of conflicts or SSM metrics, and similarly, no significant differences were found between Liu07 and Yu18 models in many pairwise comparisons.

Similarity with the NGSIM dataset

The performance of the models under study was further investigated with respect to their resemblance to the NGSIM data. As mentioned in the methodology section, the total traffic was evaluated via the GEH metric while the RMSE and RMSPE statistics were used for the remaining variables.

On the 4.00 - 4.15 pm scenario, the highest similarity with the number of merges was observed for the Gipps86 model. The Yu18 model had the highest similarity regarding the speed and time spent in the acceleration lane, followed by the Liu07 model. Yu18 also had the highest similarity in the total time spent in simulation followed by Gipps86 model.

On the 5.00 - 5.15 pm scenario, the closest proximity in terms of merges per hour was achieved by the Kang20 model, while the Liu07 model approximated better the speed in the acceleration lane. Finally, the time spent in the acceleration lane was again best approximated by the Yu18 model, while the total time in simulation was best approximated by the Kang20 model.

Finally, on the 5.15 - 5.30 scenario, the highest proximity of merges occurred in the Gipps86 model, as in the 4.00 - 4.15 pm one. The Liu07 model approximated better the speed spend in acceleration lane while, consistently with the other scenarios, Yu18 best reflected the time in acceleration lane. The total time spent in simulation was best captured by the Kang20 model.

In all three scenarios, the values of the GEH statistic were below 5 in all simulations.

A conflict analysis was also conducted for the observed NGSIM dataset. Given the low number of merging conflicts observed in the raw data, this analysis was performed considering all three datasets simultaneously in an aggregated approach. Although the NGSIM I-80 data is free of crashes, some crashes were observed because of wrong identification of the position of the vehicles in the raw data processing. Thus, two different approaches were followed namely (a) ignoring the crashes and (b) considering the crashes as conflicts.

The Spearman rank correlation coefficient was used to calculate the relationship between observed and simulated conflicts. The results are presented in Tab. 5. Since each observed traffic dataset corresponded to 20 simulations, the conflict values of the former were repeated accordingly to match the simulated data.

The data inaccuracies present in the NGSIM dataset had an impact on the analysis of the rear end conflicts. This is especially apparent in the analysis of the rear-end conflicts in which different signs and signif-

Table 4: I-80 scenario conflict analysis results

		Conflict		Crash		TTC	PET	DR
		rear end	lane change	rear end	lane change			
4.00 - 4.15 pm	Gipps86	0	2.4	0	0	1.365	1.3	-2.382
	Liu07	4.4	8.8	56.55	26.55	0.723	0.839	-3.643
	Yu18	4.75	13	61.05	23.9	0.6738	0.8354	-3.027
	Kang20	0	5.75	0	0	1.377	1.278	-1.715
5.00 - 5.15 pm	Gipps86	0	1.2	0	0	1.392	1.375	-0.799
	Liu07	13.25	9.7	14.45	10.55	0.871	1.253	-3.237
	Yu18	16.2	20	24.2	10.3	0.782	1.214	-3.777
	Kang20	0	1.85	0	0	1.362	1.435	-1.859
5.15 - 5.30 pm	Gipps86	0	0.35	0	0	1.4	1.457	-0.543
	Liu07	11.05	4.9	8.9	4	0.934	1.298	-3.193
	Yu18	8.25	13.15	34.75	2.4	0.822	1.25	-3.711
	Kang20	0.1	0.75	0.1	0.05	1.333	1.58	-2.249

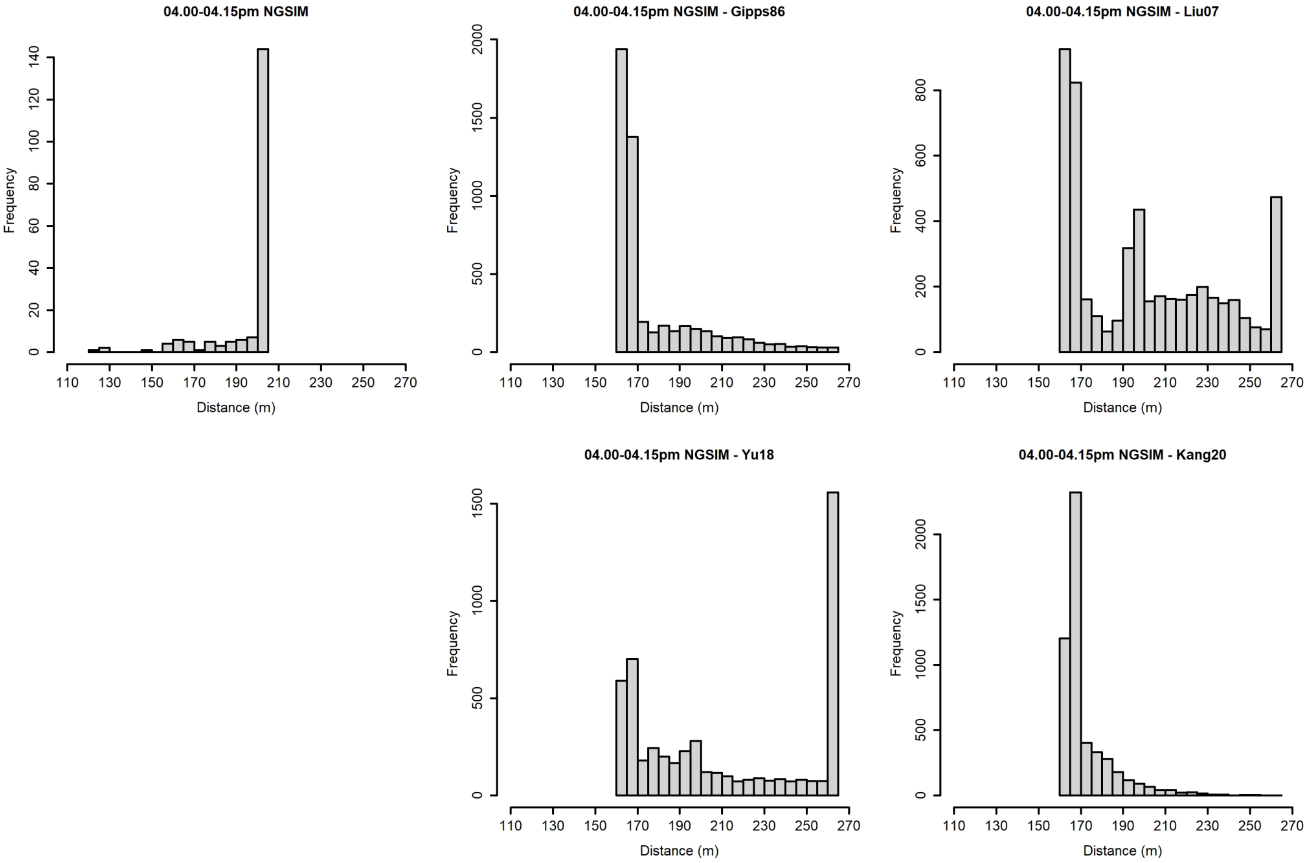

Figure 3: Merging locations for the NGSIM data and for the different simulation methods.

Table 5: Spearman rank correlation conflict analysis

Model	merge	rear-end	corrected merge	corrected rear-end
Gipps86	0.298*	-	0.298*	-
Kang20	0.272*	-0.263*	0.272*	0.227
Liu07	0.581**	-0.23	0.581**	0.573**
Yu18	0.557**	0.103	0.557**	0.331**

* significant at 0.05 level; ** significant at 0.01 level

icance levels were observed. The results based on the corrected rear end conflicts (considering crashes as conflicts) led to positive-only coefficients but only those of Liu07 and Yu18 models had moderately significant associations. On the other hand, the results of the lane-changing cases were not affected by false crashes. All correlation coefficients related to lane-changing conflicts were, indicating a positive association between the observed and simulated conflicts. This means that an increase in the number of conflicts in real traffic data is associated with increase in simulated conflicts. However, the significance of correlations was either weak or moderate. The Liu07 model approximated the number of conflicts most accurately while the remaining models presented in weaker correlation values. It should be noted though that the Liu07 model also resulted in many crashes, and therefore these results should be treated with caution.

The present analysis compared the simulation results with the observed traffic data via a series of aggregate traffic performance and safety measures. However, additional metrics could be considered to compare the behavioural representativeness of models with respect to their resulting “human-likeness”. For instance, Fig. 3 presents histograms of merging locations, defined as the position of a merging vehicle before arriving on the main lanes of the motorway.

In the NGSIM dataset, most of the vehicles used all the perceived acceleration lane and made a smooth merge. In the simulations however, vehicles attempted to merge as soon as they were allowed, thus a spike in the histograms is observed at around 160 m. If the vehicles failed to merge, the manoeuvre took place at a later point as is is apparent in the histograms. These are open issues that should be investigated in future research.

Conclusions

The paper presented and evaluated a series of motorway merging models implemented in SmartActors, a newly developed simulation platform aimed to create responsive ambient traffic. A series of simulations were run on two different road layouts, one with a short acceleration lane (base) and another based on the USA I-80 motorway, and under different traffic conditions to stress the capabilities of the models.

On the base merge scenario, clear patterns occurred regarding the performance of models. In particular, all game-theoretic models resulted in higher rates of crashes than the Gipps86 model for which no crash was observed. Consistently, the SSM analysis showed higher values for both TTC and PET using the Gipps86 model. The Kang20 model resulted in fewer crashes compared to the Liu07 and Yu18 models, though it also generated less traffic throughput and lower number of merges for higher traffic densities. This first scenario showed that trade-offs may exist between traffic and safety performance in simulation applications. Moreover, this analysis highlighted the distinction between fitting a model to a specific dataset and generating traffic using this model.

On the NGSIM I-80 based scenarios, the four models showed little difference on the traffic analysis, but only the Gipps86 and the game theoretic Kang20 model achieved acceptable safety performance although a few crashes occurred using the latter on the 5.15 - 5.30 pm I80 scenario. Moreover, their performance was comparable to the NGSIM traffic dataset.

In order to understand the differences in performance, it is worth remembering some fundamental details of the different GT models. Thus, although having intuitive payoffs, the Liu07 model is only played once, and thus if the outcomes are not correct, the risk of a crash is important. Yu18 is a repeated game, but it does not have the concept of a finite acceleration lane, and therefore the merging decision can take longer and result in a dangerous manoeuvre at the end of the acceleration lane as shown in Fig.3. Finally, the Kang20 model is a repeated game with memory that has the notion of a finite acceleration lane, with the extra advantage of being parameterised on the NGSIM dataset.

As a conclusion, some game theoretic models can be potentially used to represent traffic in simulation scenarios. The results of the current paper showed that their performance was close to both a traditional lane-changing model and real traffic observations. However, in order to improve their safety performance, it is important to encapsulate them in strong safety constraints. Future comparisons with high quality data are needed to account for individual behavioural differences such as merging position, and gap acceptance conditions.

References

Chen, J., Zhao, P., Mei, T., and Liang, H., 2013. Lane change path planning based on piecewise bezier curve for autonomous vehicle. In: *Proceedings of 2013 IEEE International Conference on Vehicular Electronics and Safety*. IEEE, pp. 17–22.

Choudhury, C. F., 2007. Modeling driving decisions with latent plans. PhD. Massachusetts Institute of Technology.

Daamen, W., Loot, M., and Hoogendoorn, S. P., 2010. Empirical Analysis of Merging Behavior at Freeway On-Ramp. *Transp. Res. Rec.*, 2188(1), pp. 108–118. ISSN: 0361-1981, 2169-4052.

Gipps, P. G., 1986. A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, 20(5), pp. 403–414.

Gipps, P. G., 1981. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2), pp. 105–111.

Halkias, J. and Colyar, J., 2006. NGSIM interstate 80 freeway dataset. *US Federal Highway Administration, FHWA-HRT-06-137, Washington, DC, USA*.

Ji, A. and Levinson, D., 2020. A review of game theory models of lane changing. *Transportmetrica A: transport science*, 16(3), pp. 1628–1647.

Kang, K. and Rakha, H. A., 2020. A Repeated Game Freeway Lane Changing Model. *Sensors*, 20(6), pp. 1554–1187.

Korzeniowski, D and Ślaski, G, 2016. Method of planning a reference trajectory of a single lane change manoeuver with Bezier curve. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 148. 1. IOP Publishing, p. 012012.

Liu, H. X., Xin, W., Adam, Z, and Ban, J., 2007. A game theoretical approach for modelling merging and yielding behaviour at freeway on-ramp sections. *Transportation and traffic theory*, 3, pp. 197–211.

Liu, R. and Hyman, G., 2012. Modelling motorway merge: The current practice in the UK and towards establishing general principles. *Transport Policy*, 24, pp. 199–210.

Marczak, F., Daamen, W., and Buisson, C., 2013. Key Variables of Merging Behaviour: Empirical Comparison between Two Sites and Assessment of Gap Acceptance Theory. *Procedia - Social and Behavioral Sciences*, 80, pp. 678–697. ISSN: 1877-0428.

Margiotta, R. and Washburn, S., 2017. *Simplified highway capacity calculation method for the highway performance monitoring system*. Tech. rep.

McKelvey, R. D., McLennan, A. M., and Turocy, T. L., 2016. *Gambit: Software Tools for Game Theory*. Tech. rep. Version 16.0.1. Available at: ¡http://gambit-project.org/¿.

Pu, L., Joshi, R., Energy, S., et al., 2008. *Surrogate Safety Assessment Model (SSAM)–software user manual*. Tech. rep. Turner-Fairbank Highway Research Center.

Saifuzzaman, M. and Zheng, Z., 2014. Incorporating human-factors in car-following models: A review of recent developments and research needs. *Transportation Research Part C: Emerging Technologies*, 48, pp. 379–403. ISSN: 0968-090X.

Toledo, T., 2002. Integrated driving behavior modeling. PhD. Massachusetts Institute of Technology.

Yu, H., Tseng, H. E., and Langari, R., 2018. A human-like game theory-based controller for automatic lane changing. *Transportation Research Part C: Emerging Technologies*, 88, pp. 140–158.

Zheng, Z., 2014. Recent developments and research needs in modeling lane changing. *Transportation Research Part B: Methodological*, 60, pp. 16–32. ISSN: 0191-2615.