Analysis of Driver Behavior and Intervehicular Collision: A Data-Based Traffic Modeling and Simulation Approach

Lu Zhao, Nadir Farhi, Zoi Christoforou, Nadia Haddadou

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1068311, 20 pages
https://doi.org/10.1155/2022/1068311

Research Article

Analysis of Driver Behavior and Intervehicular Collision: A Data-Based Traffic Modeling and Simulation Approach

Lu Zhao,[1,2] **Nadir Farhi**,[1] **Zoi Christoforou**,[1,3] **and Nadia Haddadou**[2]

[1]*COSYS-GRETTIA, Univ Gustave Eiffel, IFSTTAR, F-77447 Marne-la-Vallée, Paris, France*
[2]*Renault, 1 Avenue du Golf, Guyancourt 78280, France*
[3]*University of Patras, Department of Civil Engineering, Panepistimioupoli Patron 26504, Patras, Greece*

Correspondence should be addressed to Lu Zhao; luzhao0513@gmail.com

Received 19 July 2021; Accepted 29 December 2021; Published 21 February 2022

Academic Editor: tiziana campisi

The emergence of intelligent connected vehicles (ICVs) is expected to contribute to resolving traffic congestion and safety problems; however, it is inevitable that ICV safety issues in mixed traffic (involving ICVs and human driven vehicles) will be a critical challenge. The numerical simulation of scenarios involving a mix of different driving profiles is expected to be an important safety assessment tool in the process of testing and validating ICVs, especially regarding extreme scenarios, including car collisions, which are rarely captured in real-world datasets. In this study, we propose a novel approach for car collision generation in numerical simulations based on the assumption that car collision occurrences are mostly associated with certain specific driver profiles. Using a dataset provided by the Next Generation Simulation (NGSIM) project, NGSIM 101 dataset, we identify three different driver profiles: aggressive, inattentive, and normal drivers. We then replicate car collision occurrences by varying the percentages of these three driver profiles in the simulated environment, allowing us to establish a relationship between driver profiles and car collision occurrences. We also investigate the severity of car collisions and classify them with respect to the driver profiles of the cars involved in the collisions. Our approach of replicating car collision occurrences in numerical simulations will facilitate the testing and validation of ICVs in the future, especially regarding the testing of ICV functionalities in dealing with traffic accidents.

1. Introduction

With the development of information and telecommunication technologies, along with the rapid growth of new energy vehicles (NEVs), intelligent connected vehicles (ICVs) have become an increasingly active research topic. ICVs are expected to reshape future mobility and contribute to mitigating road traffic congestion and safety problems [1]. However, a major challenge of ICVs is communicating with other vehicles and accurately recognizing the patterns of human driving behavior in mixed traffic [2]. To ensure the security of ICVs, they need to be tested by driving hundreds of millions of miles and need to be failure-free. Moreover, performing physical tests of ICVs is not only time consuming but can sometimes be dangerous as well. As an alternative to physical tests, traffic numerical simulation platforms can be used to create realistic traffic situations. In addition, traffic accidents can be generated in traffic simulations in order to test the functionality of ICVs in dealing with traffic accidents.

It is well known that traffic numerical simulation is established by using mathematical models to reconstruct road traffic [3] and can be performed on many scales (i.e., microscopic, mesoscopic, and macroscopic). Microscopic traffic simulators are particularly promising in the validation of ICV behavioral decision systems and work by replicating individually each vehicle behavior in detail in order to create an entire traffic environment. Various microscopic traffic simulators have been developed in recent years, such as SUMO [4], VISSIM [5], and AIMSUN [6]. However, among these simulators, vehicle collision generation mechanisms have not yet been described completely. In this study, we propose a novel approach for car collision generation in

numerical simulations based on the assumption that car collision occurrences are mostly associated with certain specific driver profiles.

As is well known, drivers have different driving patterns, and different driver profiles can be distinguished. Meiring and Myburgh [7] reviewed various methods of classification of driving styles, such as "normal," "aggressive," and "inattentive." Aggressive drivers are often characterized by risky behaviors with regards to velocity (such as abrupt instantaneous speed change, driving over the speed limit, excessive acceleration, or deceleration) and frequent lane changing behavior [8]. Inattentive drivers can be characterized by long reaction time and forced sudden lane changes with an abrupt deviation from normal behavior [7]. In our work, we focus on driver profiles only in terms of how they react when directly behind another car (the car-following behavior).

Several open datasets on vehicular trajectories are currently available and can be applied to studying and analyzing human driving behavior. Next Generation Simulation (NGSIM) has published four vehicle trajectory datasets (NGSIM [9]), which have been widely used in transportation research [10]. More recently, datasets on motion trajectory of traffic objects detected by autonomous vehicle sensors have been published (e.g., Waymo data [11], Argoverse [12], and nuScenes [13]). However, in these datasets, vehicle collisions are rarely observed. Nevertheless, the Strategic Highway Research Program 2 Naturalistic Driving Study (SHRP2-NDS) [14] provides a dataset where several accident situations are captured, yet seem unsatisfactory in terms of quantification of vehicle collision causes and impacts. Therefore, there is a lack of data on real-world vehicle collisions connected to human driving behavior, including a lack of both video data and vehicle trajectories in collision situations.

On the other hand, traditional road safety analyses are based on statistical methods which aim to understand the importance of geometric (road shape) and track (road type) characteristics for safety or to locate collision black-spots, based on crash datasets [15]. Crash datasets as mentioned in [16] provide only vehicle collision statistics, without involving car trajectories and driver profiles. A large body of the literature on road traffic safety explores the link between ambient traffic conditions and collision occurrences [17] and/or vehicle collision injury severity [18]. On highways specifically, the authors of [19] provide evidence that vehicle collisions have a specific relationship with traffic flow characteristics (volume, density, and speed). In [20], the authors identified the impact of general geometric parameters, weather, and traffic flow on different types of collision (rear-end, sideswipe, and multiple vehicle involved) based on French traffic and vehicle statistical data. Some other studies employed statistical and machine learning approaches, such as multivariate probit models [20] or support vector machine [21] to investigate the relationship between influencing traffic factors and vehicle collision occurrences.

Regarding vehicular collision involving unexpected human driving behavior, several previous works use traffic simulations. In [22], based on their own proposed models for three types of vehicle collision depending on vehicle interactions and maneuvers, the authors estimated the parameters of the models via simulation. In [23], the authors studied the impact of driving violations (driving over the speed limit, slow driving, and abrupt hard braking) on car collision occurrence for different traffic conditions through simulation. Moreover, in some other related works, the traffic conflicts technique (TCT) using surrogate safety measures based on vehicle trajectories has been proposed to assess traffic safety [24]. These indicators can be used for predicting potential collisions [15, 25, 26]. Nevertheless, most of the literature addresses road traffic safety without considering driving profiles. In our proposed approach, we would like to represent different driver profiles in traffic simulations and then increase the number of drivers with extreme driving profiles (aggressive and inattentive). Therefore, by the method of increasing the percentage of drivers with extreme profiles, car collisions could be generated in the traffic simulations.

In this study, we propose a novel approach for car collision (collision between cars) generation in numerical simulation by varying the percentages of different driver profiles in the traffic, aiming at establishing a relationship between driving profiles and car collision occurrences. The profiles are extracted from the NGSIM database and integrated in the traffic simulator SUMO [4, 27, 28] using the calibrated intelligent driver model (IDM), which is one of the most human-like car-following models [29, 30]. For the simulation, we used the IDM with an extension of driver reaction times. Therefore, our main goals in this study are to extract driver profiles from real driving data and replicate them using a microscopic traffic model, establish a method of reproducing realistic traffic simulation based on microscopic road traffic modeling in the SUMO simulator, generate car collisions by appropriately varying the percentages of the driver profiles, characterize the relationship between the car collision occurrences with the driver profiles, and observe the severity of the generated car collisions.

2. Materials and Methods

In this section, we present the relevant materials and the proposed method. The approach structure is shown in Figure 1. In this work, we study the driver profiles from car trajectories and propose the definition of an aggressive driver profile as a driver who always leaves short time headways with respect to the leading vehicle and an inattentive driver profile as a driver with particularly long reaction times compared to other drivers. We classify all drivers with intermediate values of reaction time and time headway as the normal profile. The selection and thresholds for the three driver profiles were based on the NGSIM 101 dataset [9]. After the classification of the driver profiles, the specific driver profiles (aggressive and inattentive) are represented using the calibrated IDM model (the existing IDM model with an extension of reaction time), and the road traffic is then simulated using SUMO ("Simulation of Urban MObility") [28]. The IDM model calibration is performed using a genetic algorithm to find the optimal set of IDM

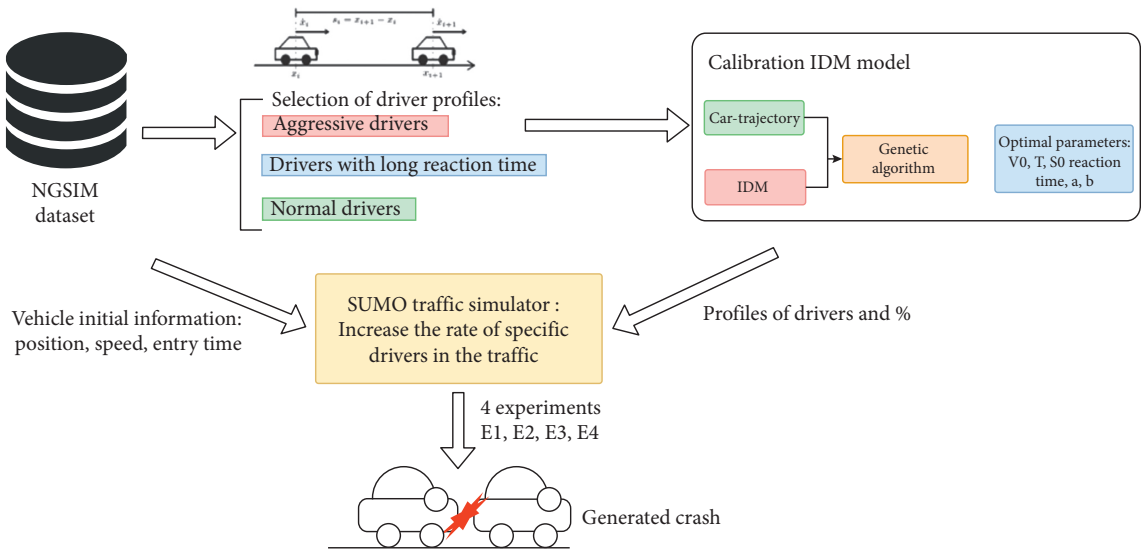

Figure 1: Steps of the approach.

parameters with an objective of minimizing the predefined error between real driver trajectory and the output of the IDM model. In essence, we artificially increase the percentage of drivers with extreme profiles (aggressive or inattentive) and then count and analyze the car collisions generated by traffic simulation. In the numerical simulation experiments, we propose 4 different experiments using different combinations of driver groups.

The remaining sections of this study are structured as follows: in the rest of this section, from Subsection 2.1 to 2.4, we present the necessary elements we predefined for our method of vehicular collision generation, including (1) the presentation of the NGSIM 101 dataset, (2) the proposition of different driver profiles, (3) SUMO traffic simulator and the choice of car-following model (IDM model), and (4) the method of IDM model calibration to present different driver profiles. In Sections 3– 5, we present the specific driver profiles in the NGSIM 101 dataset, the calibration of the IDM model for the specific drivers, and the simulation experiments description. Section 6 presents the results obtained from numerical simulations and the investigation of the relationship between generated car collisions and driver profiles. Section 7 presents the validation of the approach with a different part of data in the NGSIM 101 dataset. Section 8 deals with the severity of simulated car collisions with respect to different driving profiles. Finally, Section 9 summarizes major findings and provides recommendations for further research.

2.1. NGSIM 101 Dataset Description. NGSIM 101 [9] is an open dataset released by the US Federal Highway Administration (FHWA). On a highway section in Los Angeles, California, covering 640 meters of length, all vehicle trajectories are provided with a rate of 10 Hz, and all data were collected during the rush hour from 7 : 50 a.m. to 8 : 35 a.m. This section has 5 normal lanes and one auxiliary lane connecting an on-ramp and an off-ramp. In our study, we focused on the car trajectories only in the 5 main lanes. In addition, the whole NGSIM 101 dataset is divided into three subsets of 15 minutes, which cover the traffic from 7 : 50 a.m. to 8 : 05 a.m., from 8 : 05 a.m. to 8 : 15 a.m., and from 8 : 20 a.m. to 8 : 35 a.m. In Figure 2, we present traffic volumes and speeds during each of the three time periods. It can be observed that in the beginning of the first 15 minutes, the mean speed is between 10 and 15 m/s. After that, traffic becomes denser resulting in a sudden fall of the mean speed. In the second and third 15-minute time periods, traffic is more congested and the mean traffic speed varies between 5 and 10 m/s.

In this work, we applied the proposed approach to the first 15-minute dataset. The second 15-minute dataset was then used for method validation. In our approach, we are only interested in understanding driving profiles in terms of car-following behavior, without considering lane change or other behaviors. After processing the data of the first 15-minute time frame, we selected approximately 1500 car trajectories (drivers) from a total population of 1993 vehicle trajectories to extract specific driver profiles. The selection was based on the following condition: each car must always have a leading car whose trajectory is continuous for at least 40s. The same data processing was made for the second 15-minute of data and resulted in the selection of 1300 car trajectories out of a total of 1495.

2.2. Classification of Driver Profiles. A driver profile can be defined as the average driving behavior of a given driving class [31]. Driving behavior is related to a driver's skills, sociodemographic status (age, gender, and occupation) and current psychological state (fatigue and distraction). From vehicle trajectory data, a driver's characteristics can be illustrated by the speed, acceleration, jerk, and some other relevant traffic indicators, such as time headway (THW) and time to collision (TTC) [31]. THW and TTC are critical safety indicators for car-following behavior [32]. THW measures the time passed before reaching the

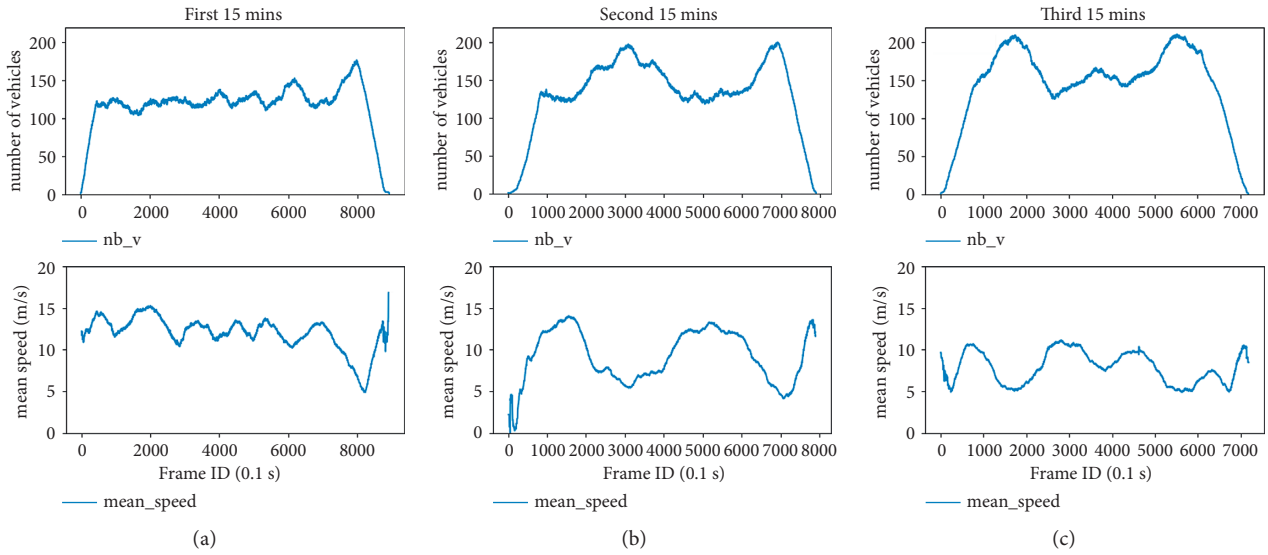

Figure 2: NGSIM 101 dataset traffic flow (number of vehicles in traffic and traffic mean speed for each time frame) for three 15-minute periods, respectively, (a) from 7 : 50 a.m. to 8 : 05 a.m., (b) from 8 : 05 a.m. to 8 : 15 a.m., and (c) from 8 : 20 a.m. to 8 : 35 a.m.

leading vehicle's position while running at current speed, while TTC is usually used for judging the moment to start braking and in the control of braking [33]. Several studies show that the distribution of THW is related to driving speed and also to traffic flow conditions. According to [34], a negative correlation between the car speed and THW can exist. In [35], it was shown that the speed and THW follow different distribution patterns under different traffic density levels. The THW can illustrate better the driver's characteristics in all conditions of car-following behavior than the TTC, which is only important for the braking behavior.

In this work, the classification of different driver profiles begins with the observations from the NGSIM data. Figure 3 shows the distribution of each car driver's maximum THW, mean THW, minimum THW, and standard deviation of THW, for the 1993 vehicle trajectories of the first 15 min NGSIM 101 dataset.

Figure 3 shows evidence of the heterogeneity of human driving profiles as the mean THW ranges from near 0s to 5s and the minimum value of THW ranges from near 0s to more than 3s. Based on these previous preliminary findings, we propose the definition of three profiles: (i) aggressive: shorter car time headway, (ii) inattentive: longer reaction time, and (iii) normal: for intermediate values of reaction time and car time headway. We notice that since the reaction time cannot be detected in the real driving data, we estimate it here, and we assume that the reaction time is approximate to the minimum THW value, which is the minimum safe time gap the driver estimates should be maintained with the leading vehicle during their whole trajectory (longer than 400 time steps, explained in Section 2.1). The definitions of the the two driver profiles (aggressive and inattentive) are formalized.

(1) Aggressive driver profile: A driver i is considered to be aggressive with respect to a threshold t^* on the time headway (THW) if

$$\overline{\mathrm{THW}}(i) = \frac{1}{T}\sum_{t}^{T}\mathrm{THW}(i,t) < t^*. \tag{1}$$

We will explain in Section 3 how we fixed the threshold t^* in our approach.

(2) Inattentive driver profile (drivers with long reaction time): A driver i is considered to be inattentive (with a long reaction time) with respect to a threshold $\tilde{t}$ on the time headway (THW) if

$$\min \mathrm{THW}(i) \coloneqq \min_{t}\mathrm{THW}(i,t) > \tilde{t}. \tag{2}$$

We will explain in Section 3 how we fixed the threshold $\tilde{t}$ in our approach.

(3) Normal driver profile: the drivers whose profiles are not aggressive or inattentive are called normal. They have intermediate values of reaction time and time headway.

2.3. SUMO Traffic Simulator and IDM Car-Following Model. SUMO ("Simulation of Urban MObility") is a microscopic open-source traffic simulator [4, 27, 28], widely used for traffic research. Microscopic traffic models generally include two driver behavior models: (a) car-following, corresponding to the behavior of a driver in reaction to the actions or stimulus of the leading vehicle; (b) lane change, corresponding to the behavior of lane changes that includes maneuvers of overtaking as well as insertion into a target lane. We limit our work to the car-following component. In addition, during the simulation, SUMO can detect physical collisions (front and back bumper meet or overlap) [36], and a collision is observed when a following vehicle collides into the rear-end of the proceeding vehicle. The collision counts become available at the end of each simulation.

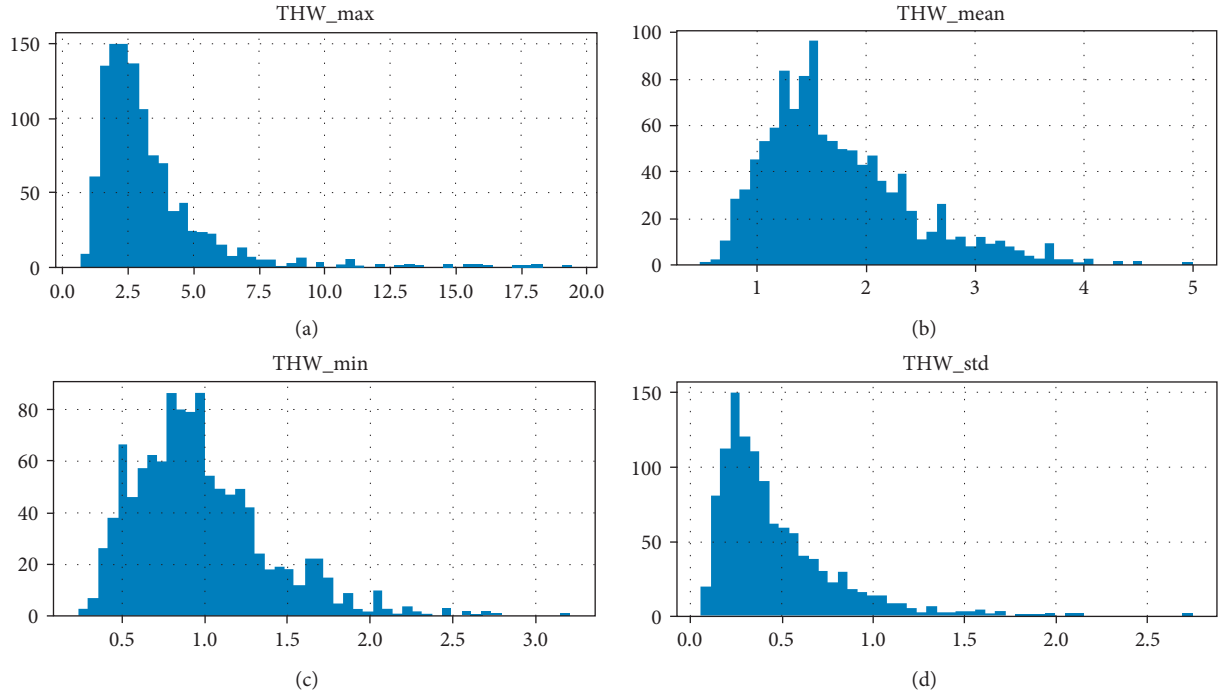

Figure 3: The distribution of (a) the maximum THW, (b) mean THW, (c) minimum THW, and (d) standard deviation of THW for each driver in the first 15 minutes of data in the NGSIM 101 dataset. The unit on the X-axis is seconds.

Many interesting reviews on car-following can be found in the literature [37–39]. We briefly present here some of the most commonly used. The GHR model [40] is a stimulus-response model in which drivers perform their acceleration/deceleration depending on car speeds, relative car speeds with respect to leading vehicles, intervehicle distances, and drivers' reaction time. Wiedemann's (1974) car-following model is used in the Vissim simulator and describes the psycho-physiological aspects of driving behavior in terms of four discrete driving regimes. This model considers different modes of operation, divided into "no reaction zone" (free-road), "closing in," "must decelerate," and "car-following" by human perceptual thresholds [38]. The Gipps model [41] and Krauss model [42] are based on safe distance. In the Gipps model, drivers update their car speeds with respect to keeping a minimum safe distance between themselves and the leading vehicle, in order to avoid collisions and in anticipation of the extreme case of a leading car braking suddenly. The Krauss model [42] is the car-following model by default in the SUMO simulator. It extends the Gipps model by modeling the imperfection of human drivers with stochastic terms of car speed. This property makes (or is supposed to make) the model more realistic. The IDM car-following model was first published in [43] which improved the initial results produced by Gipps' model. The acceleration is calculated as a function of the desired speed and the desired space headway. The IDM model is suitable for both free flow and congestion phases [38]. This attribute makes the model more efficient and facilitates its calibration with real driving data.

We chose here the IDM car-following model [43], which has been already implemented in SUMO, and it is shown by several research works that IDM is the most human-like car-following among certain compared models [29, 30]. Therefore, we briefly recall the model.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = a\left[1 - \left(\frac{V}{V_0}\right)^{\delta} - \left(\frac{s^*(V,\Delta V)}{s}\right)^2\right], \tag{3}$$

where

$$s^*(V,\Delta V) \coloneqq s_0 + \max\left(0, VT + \frac{V\Delta V}{2\sqrt{ab}}\right), \tag{4}$$

where $\mathrm{d}V/\mathrm{d}t$ is the vehicle acceleration, V is the vehicle speed, s is the intervehicular distance; ΔV is the relative speed with leading vehicle; a is the maximum acceleration, b is the maximum deceleration, V_0 is the desired speed, δ is the exponential parameter of speed, we fixed it at 4 referring to [43], s_0 is the minimal intervehicular distance with leading vehicle, and T is the desired time headway (desired THW).

The output of the IDM model is the acceleration $\mathrm{d}V/\mathrm{d}t$ given as a function of the influencing factors (the inputs of the model): the vehicles own speed V, the intervehicular distance s, and the relative speed ΔV with respect to the leading vehicle.

In addition, we notice that the original IDM car-following model does not include reaction time as a parameter [39]. Nevertheless, in a recent development of the SUMO simulator [28], the authors improved the work on reaction time modeling. They indicated that reaction time could be introduced into the driving model (car-following and lane change models) as an additional parameter. In the work we present here, we introduced a reaction time parameter into the IDM car-following model. In SUMO, the simulation step can be set to Δt, and the driver reaction time is a time delay to

make a decision both for updating acceleration and lane changing depending on the present state of traffic. This delay of decision time (reaction time) can be set as $(n * \Delta t)$, where n is an additional parameter which we determine by calibration.

2.4. Method of Calibration of the IDM Model. In order to simulate real driver behavior and driver profiles, the microscopic traffic model must be calibrated with real driving data. Two main types of calibration of car-following models exist: (1) estimation of driving model parameters accounting for the physical meanings of each parameter [44], in which parameters (such as speed maximum and acceleration maximum) can be extracted directly from vehicle trajectory data, such as the maximal value of speed in the trajectory and maximal value of acceleration in the trajectory, and (2) calibration of driving models, which can be constructed as an optimization problem. In the second approach, the objective function and the optimization method need to be selected, and the problem is performed to minimize the distance between the simulated vehicle trajectories by the model and real vehicle trajectories. Therefore, by this method, the optimal set of parameters can be found. In the methods of calibration by optimization, several mathematical optimization methods and algorithms such as Newton, Gauss-Newton, gradient descent, and Levenberg-Marquardt are presented for car-following model parameters optimization in [45]. The authors of [45] proved that genetic algorithms (GAs) are also effective to solve optimization problems for car-following model calibration. Many of the recent works on car-following model calibration used GA methods [29, 30], and their works show that the GA optimization method is efficient for car-following model calibration.

In this work, the parameters which need to be calibrated in the IDM model are as follows: a (maximal acceleration), b (maximal deceleration), V_0 (desired speed), s_0 (minimal intervehicular distance to leading vehicle), and T (desired THW), as well as the additional parameter n for the reaction time $(n * \Delta t)$. In addition, the parameter δ (exponential parameter of speed) is fixed at $\delta = 4$ as proposed in [43].

In our approach, the identification of drivers with extreme profiles is performed by calibrating the parameters of the IDM model, using a genetic algorithm referring to the work in [30]. The used GA method is the default model in Matlab 2020b [46], and for the algorithm description, it is presented in the book [47]. In general, the mean square error (MSE) is widely used for calibration of car-following models. Here, another error metric is used as in [30], root mean square percentage errors (RMSPE).

$$\mathrm{RMSPE}(\theta) = \sqrt{\frac{1}{P}\sum_{i=0}^{P}\frac{\left(x_i^{\mathrm{simul}}(\theta) - x_i^{\mathrm{data}}(\theta)\right)^2}{x_i^{\mathrm{data}}(\theta)^2}}, \tag{5}$$

where x_i^{data} (respectively, x_i^{simul}) denotes the car trajectory (positions in time) from data (respectively, from simulation) and where θ represents all the parameters of the considered model. The car longitudinal position is chosen as the target variable in the objective function.

The IDM model calculates car acceleration, as described in (6), with the driver's reaction time $(n * \Delta t)$. The reaction time is a delay of decision that driver takes for updating the acceleration with current situation with the leading vehicle. Car's speed and position can be calculated as follows in (7) and (8), based on the Euler method.

$$a^{\mathrm{simul}}(t + n\Delta t) = \mathrm{IDM}(\theta; \mathrm{variables}(t)), \tag{6}$$

$$v^{\mathrm{simul}}(t + \Delta t) = v^{\mathrm{simul}}(t) + a^{\mathrm{simul}}(t) * \Delta t, \tag{7}$$

$$x^{\mathrm{simul}}(t + \Delta t) = x^{\mathrm{simul}}(t) + v^{\mathrm{simul}}(t) * \Delta t. \tag{8}$$

In addition, for drivers with normal profile, we simulate them using an average profile. Their calibration of IDM parameters is extracted directly from the dataset by their physical meanings and combined with the default values in the SUMO simulator.

3. Specific Driver Profiles in the NGSIM Dataset

In a traffic dataset with a large number of trajectories, it is always possible to distinguish good drivers from risky ones. Following the definition of the two extreme driver profiles of Section 2.2, we consider the following four driver profile groups in the NGSIM dataset: group 1: the 2.5% most aggressive drivers, group 2: the second 2.5% most aggressive drivers (drivers ranked between 2.5% and 5% of mean THWs), group 3: the 2.5% most inattentive drivers, and group 4: the second 2.5% most inattentive drivers (drivers ranked between 2.5% and 5% in terms of inattentiveness). Mathematically, the four groups are defined as follows.

Let us denote the set $\mathcal{I}_{\mathrm{agg}}(t)$, the of drivers i whose $\overline{\mathrm{THW}}(i)$ is less than t:

$$\mathcal{I}_{\mathrm{agg}}(t) \coloneqq \left\{i, \overline{\mathrm{THW}}(i) < t\right\}. \tag{9}$$

Now, we denote by N the total number of drivers of the considered dataset and by $N_{\mathrm{agg}}(t)$ the cardinal (number of elements) of $\mathcal{I}_{\mathrm{agg}}(t)$:

$$N_{\mathrm{agg}}(t) \coloneqq \mathrm{Card}\left(\mathcal{I}_E(t)\right). \tag{10}$$

Finally, we denote by $p_{\mathrm{agg}}(t)$ the proportion of the number of drivers in $\mathcal{I}_{\mathrm{agg}}(t)$ with respect to the total number of drivers (N):

$$p_{\mathrm{agg}}(t) \coloneqq N_{\mathrm{agg}}\frac{(t)}{N}. \tag{11}$$

Let us now define t_1^* and t_2^* as follows:

$$\begin{aligned} t_1^* &= p_{\mathrm{agg}}^{-1}(0.025) \coloneqq \max\left\{t, p_{\mathrm{agg}}(t) \le 0.025\right\}, \\ t_2^* &= p_{\mathrm{agg}}^{-1}(0.05) \coloneqq \max\left\{t, p_{\mathrm{agg}}(t) \le 0.05\right\}. \end{aligned} \tag{12}$$

Groups 1 and 2 are then defined as follows:

$$\begin{aligned} \mathrm{Group\ 1} &= \mathcal{I}_{\mathrm{agg}}\left(t_1^*\right) = \mathcal{I}_{\mathrm{agg}}\left(p_{\mathrm{agg}}^{-1}(0.025)\right), \\ \mathrm{Group\ 2} &= \mathcal{I}_{\mathrm{agg}}\left(t_2^*\right) \backslash \mathcal{I}_{\mathrm{agg}}\left(t_1^*\right) \\ &= \mathcal{I}_{\mathrm{agg}}\left(p_{\mathrm{agg}}^{-1}(0.05)\right) \backslash \mathcal{I}_{\mathrm{agg}}\left(p_{\mathrm{agg}}^{-1}(0.025)\right). \end{aligned} \tag{13}$$

The groups 3 and 4 are defined similarly, that is,

$$\begin{aligned}
&\mathcal{I}_{\text{inatt}}(t) \coloneqq \{i,\ \min \text{THW}(i) > t\},\\
&N_{\text{inatt}}(t) \coloneqq \text{Card}\left(\mathcal{I}_{\text{inatt}}(t)\right),\\
&p_{\text{inatt}}(t) \coloneqq N_{\text{inatt}}\frac{(t)}{N},\\
&t_3^* = p_{\text{inatt}}^{-1}(0.025) \coloneqq \min\{t, p_{\text{inatt}}(t) \le 0.025\},\\
&t_4^* = p_{\text{inatt}}^{-1}(0.05) \coloneqq \min\{t, p_{\text{inatt}}(t) \le 0.05\},\\
&\text{Group 3} = \mathcal{I}_{\text{inatt}}\left(t_3^*\right) = \mathcal{I}_{\text{inatt}}\left(p_{\text{inatt}}^{-1}(0.025)\right),\\
&\text{Group 4} = \mathcal{I}_{\text{inatt}}\left(t_4^*\right) \backslash \mathcal{I}_{\text{inatt}}\left(t_3^*\right) = \mathcal{I}_{\text{inatt}}\left(p_{\text{inatt}}^{-1}(0.05)\right) \backslash \mathcal{I}_{\text{inatt}}\left(p_{\text{inatt}}^{-1}(0.025)\right).
\end{aligned} \tag{14}$$

Applying the definitions of groups 1, 2, 3, and 4 above, we obtain the thresholds of Table 1 on $\overline{\text{THW}}$ for groups 1 and 2 and on min THW for groups 3 and 4 (Figure 4).

4. IDM Model Calibration for Extreme Driver Profiles

As mentioned above, the calibration is considered as the optimization problem with the aim at minimizing the error between the simulated car trajectories by the IDM model and the real car trajectories from the dataset. We chose here to apply a genetic algorithm (the default algorithm in Matlab [48]) and to attempt to find the global optimum parameters of selected drivers. Moreover, we repeated the genetic algorithm 10 times for every driver trajectory, in order to find a global optimum solution for the parameters of the calibration.

In the application of the genetic algorithm, upper and lower bounds need to be set for all parameters in the IDM model. For the reaction time and desired THW, the bounds for these two parameters need to be set differently for the aggressive and inattentive driver profiles because the parameters' real distribution for these two profiles are different(Figure 4). Thus, we set that for aggressive drivers, the desired THW ranges from 0.1 to 4 s, while the reaction time is from 0.1 to 2 s, and for inattentive drivers, we set the desired THW ranges from 1 to 4 s, while the reaction time goes from $tr_i - 0.3$ to $tr_i + 0.3$ seconds, where tr_i is the value of min THW (i) (Section 2.2, where we approximate the driver's reaction time to be around tr_i). For other parameters, the bounds are the same for both aggressive and inattentive drivers as in [30]. Thus, the desired speed (V_0) ranges from 10 to 40 m/s, the minimal intervehicular distance s_0 ranges from 0.1 to 10 m, and the maximum acceleration and the maximum deceleration range from 0.1 to 5 m/s^2.

The results of calibration for the 4 specific groups of drivers are given in Appendix A. For the calibration on the car trajectories variable (car positions in time), we obtained for all 4 driver profile groups an RMPSE ranging in (1%, 2%). From the calibration results, given in Tables 2–5, we notice that group 1 driver profiles have shorter reaction times and desired headway times compared to group 2 driver profiles. On the other hand, group 3 driver profiles have longer reaction times and desired headway times than group 4 driver profiles. Moreover, the desired THW (parameter T of formula (4)) of attentive drivers is much larger than the one of aggressive drivers. This result of calibration of IDM parameters for aggressive drivers and inattentive drives is consistent with the findings from the dataset(Figure 4).

5. Numerical Simulation Experiments Setup

Before generating the collisions in traffic, we need to ensure that the NGSIM traffic can be simulated realistically. The reproduction of the NGSIM traffic data by numerical simulation consists on the creation of the road section, the selection of the microscopic traffic model (IDM car-following model plus the default lane changing model in SUMO), and the creation of vehicles. The road network of NGSIM is drawn manually using NETEDIT [49] provided by the SUMO simulator, which is presented as a road of 640 meters with 5 straight lanes. The creation of each vehicle is provided by its longitudinal origin position, longitudinal destination position, entering lane, and entering time (in seconds). This necessary information is extracted from the car trajectory data.

As what we supposed in Section 3, each extreme driver profiles have 2.5% in the initial traffic simulation. In Figure 5, we give car mean speed from the NGSIM real data and from numerical simulation in SUMO. This figure shows that the traffic simulation represents well the state of traffic (car speed) over time. This first simulation with initial driver profiles distribution in the traffic show that we can simulate the real NGSIM traffic in SUMO. Under this condition, we can begin the experiments for generating the collisions. In addition, during the simulation, SUMO can detect physical collisions (front and back bumper meet or overlap) [36]. We can get the collision counts at the end of each simulation. Four experiments of car collision simulation are proposed for different combinations of the two selected groups of drivers. Each experiment is carried out using different combinations of one group of aggressive drivers and one group of inattentive drivers. Driver profiles used in each experiment are given in Table 6, and the different groups of

Table 1: 4 groups' specific drivers and obtained threshold of profile definition.

Type of drivers	Indicator	Threshold
1st aggressive (group 1)	Mean THW	$\leq 0.82s$
2nd aggressive (group 2)	Mean THW	(0.82s, 0.94s)
1st inattentive (group 3)	min THW	$\geq 1.84s$
2nd inattentive (group 4)	min THW	(1.656s, 1.84s)

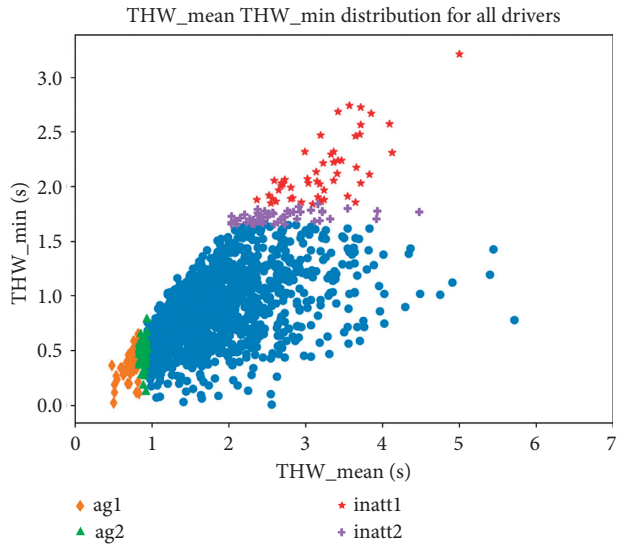

Figure 4: $\overline{\text{THW}}$ (THW_mean) and min THW (THW_min) distribution of all drivers in the first 15 min period data, group 1 in orange, group 2 in green, group 3 in red, and group 4 in purple. The remainder drivers (blue points) are considered as drivers with normal driving profile.

Table 2: Calibration results for the driver profiles of group 1 (2.5%, 1st aggressive drivers).

Parameters and error	Bounds	Mean	Std	Median	25%	75%
V0 (m/s)	(10, 40)	26.25	7.69	23.03	20.29	32.01
T (s)	(0.1, 4)	**1.08**	0.12	1.03	1.0	1.09
S0 (m)	(0.1, 10)	0.74	1.1	0.1	0.1	1.1
Reaction time (s)	(0.1, 2)	**0.45**	0.2	0.5	0.3	0.6
a (m/s^2)	(0.1, 5)	4.49	0.95	5.0	4.52	5.0
b (m/s^2)	(0.1, 5)	4.92	0.28	5.0	5.0	5.0
RMPSE position (%)		2.34	0.9	2.14	1.78	2.92

The values in bold indicate the errors.

Table 3: Calibration results for the driver profiles of group 2 (2.5%, 2nd aggressive drivers).

Parameters and error	Bounds	Mean	Std	Median	25%	75%
V0 (m/s)	(10, 40)	25.95	8.94	22.22	19.22	36.42
T (s)	(0.1, 4)	**1.21**	0.21	1.16	1.02	1.34
S0 (m)	(0.1, 10)	0.39	0.67	0.1	0.1	0.33
Reaction time (s)	(0.1, 2)	**0.55**	0.16	0.5	0.5	0.68
a (m/s^2)	(0.1, 5)	4.69	0.92	5.0	5.0	5.0
b (m/s^2)	(0.1, 5)	4.87	0.61	5.0	5.0	5.0
RMPSE position (%)		1.03	2.09	1.55	2.7	1.03

The values in bold indicate the errors.

driver profiles are described in Section 3. For generating the car collisions, we assume here that the car collisions can be generated in each simulation by increasing the number of aggressive and/or inattentive drivers in the traffic. The increasing of the percentages of extreme driver profiles is done artificially and randomly by replacing the normal drivers by the chosen extreme driver profiles. In each experiment, we reset numbers of simulations by using different rates of drivers associated to each extreme driver profile (2.5% in the origin dataset and then 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%, respectively, for aggressive drivers and inattentive drivers in each simulation). For each simulation, once the rates of the extreme driver profiles are fixed, all the other drivers are simulated as normal.

Table 4: Calibration results for the driver profiles of group 3 (2.5%, 1st inattentive drivers).

Parameters and error	Bounds	Mean	std	Median	25%	75%
V0 (m/s)	(10, 40)	22.33	8.81	19.8	15.75	24.16
T (s)	(1, 4)	**2.38**	0.71	2.43	1.8	2.83
S0 (m)	(0.1, 10)	3.8	3.84	1.94	0.1	7.56
Reaction time (s)	$(tr_i - 0.3, tr_i + 0.3)$	**2.16**	0.61	2.2	2.0	2.38
a (m/s^2)	(0.1, 5)	3.86	1.46	4.7	2.94	5.0
b (m/s^2)	(0.1, 5)	4.86	0.67	5.0	5.0	5.0
RMPSE position (%)		1.72	0.99	1.41	1.13	2.11

The values in bold indicate the errors.

Table 5: Calibration results for the driver profiles of group 4 (2.5%, 2nd inattentive drivers).

Parameters and error	Bounds	Mean	Std	Median	25%	75%
V0 (m/s)	(10, 40)	25.18	9.38	21.94	18.02	32.89
T (s)	(1, 4)	**1.98**	0.58	2.05	1.48	2.47
S0 (m)	(0.1, 10)	3.65	3.73	2.02	0.1	7.83
Reaction time (s)	$(tr_i - 0.3, tr_i + 0.3)$	**1.73**	0.49	1.8	1.72	1.9
a (m/s^2)	(0.1, 5)	3.99	1.64	5.0	3.59	5.0
b (m/s^2)	(0.1, 5)	4.85	0.6	5.0	5.0	5.0
RMPSE position (%)		1.68	0.9	1.43	1.15	1.89

The values in bold indicate the errors.

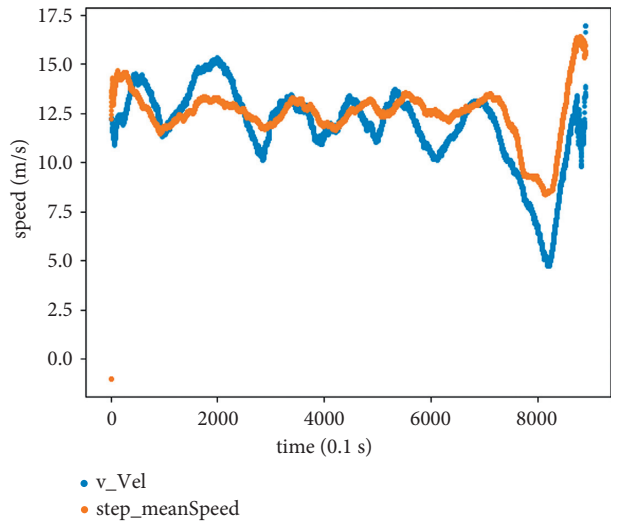

Figure 5: The average traffic flow speed from real data (blue) and the simulated average car speed (orange) in SUMO of the initial traffic for the first 15 minutes time frame, where 2.5% are aggressive drivers, simulated using their own IDM calibrated parameters, 2.5% are inattentive drivers, simulated using their own IDM calibrated parameters, and the rest are normal drivers in the traffic.

Table 6: Numerical simulation experiments.

E1	2.5% 1st aggressive drivers	2.5% 1st inattentive drivers	95% normal drivers
E2	2.5% 2nd aggressive drivers	2.5% 1st inattentive drivers	95% normal drivers
E3	2.5% 1st aggressive drivers	2.5% 2nd inattentive drivers	95% normal drivers
E4	2.5% 2nd aggressive drivers	2.5% 2nd inattentive drivers	95% normal drivers

6. The Results

In this section, we give the results of the numerical simulations. In Subsection 6.1, we provide the number of car collisions generated by numerical simulation, based on the first 15 minutes data, and, respectively, for the defined four experiments. In Subsection 6.2, we focus on the analysis of the car collisions obtained in the simulations of experiment 1 (E1). In addition, we show that the number of simulated car collisions can be written as a function of the rates of aggressive and inattentive driver profiles in traffic. In Section 7, we use another 15-minute time period data for the validation of our approach.

6.1. Number of Car Collisions by Numerical Simulation. We start with the implementation of experiment E1. We observed that the number of collisions occurred in each

simulation is various even when using the same rate of specific drivers. This is due to the random distribution of the specific (aggressive and inattentive) driver profiles over all the drivers. Furthermore, we observed the mean number of collisions converged after several times simulations for each case. Therefore, the presented collision counts in each case shown in the following tables are the average value over 10 simulations for each case. Moreover, the number of car collisions obtained by numerical simulation is given in Tables 7–10 for each experimentation (E1, E2, E3, and E4) respectively.

In Table 7, the simulated car collision counts are given for taking different rates of specific profiles in E1. The results indicate clearly that the number of intervehicular collision occurrences increases by increasing the number of aggressive and/or inattentive driver profiles. Similarly, from the results for E2, E3, and E4, given, respectively, in Tables 8–10, the number of intervehicular collision occurrences also increases by increasing the number of aggressive and/or inattentive driver profiles. It is shown that the number of obtained car collisions in these three experiments is lower than the simulated result in E1 (by comparing Tables 8–10 to Table 7). In addition, the number of car collisions obtained with E4 is the lowest compared to the other three experiments. This result shows that a different risk level is associated to each group of specific driver profiles. The first group of aggressive driver profiles could potentially cause more car collisions than the second group of aggressive driver profiles (comparing the results of E1 and E2). Similarly, the first group of inattentive driver profiles could cause more car collisions than the second group of inattentive driver profiles (comparing the results of E3 and E4). It is interesting to note that in all the four experiments, no collisions occurred when the percentage of extreme drivers was low. With the increase of the rate of each specific driver profile, the number of car collisions increases. However, this increase seems to be different from one experiment to another (Figure 6), where each line presents the number of car collisions in the simulation of each experiment with the condition that aggressive drivers and inattentive drivers have the same rate in the traffic. From Figure 6, we can observe clearly that E1, where we generated car collisions based on varying the percentage of the most aggressive and the most inattentive drivers, can generate more car collisions than the other three experiments. The E4 has probably the least number of car collisions, where we generated car collisions based on varying the percentage of the second aggressive and the second inattentive drivers.

6.2. Analysis of Car Collisions Obtained in Experiment 1 (E1). In this subsection, we focus on further analyzing the results obtained by E1 (Table 7). A car collision is observed in SUMO when a following vehicle (veh2) crashes into the rear-end of the proceeding vehicle (veh1). We can get the driver profile information of the two vehicles involved in each collision. Therefore, the distribution of rear-end car collisions produced by different driver profiles (aggressive, normal, and inattentive) is shown in Figure 7, and the distribution is shown in percentage. In Figure 7, the suffixes ag, long, and nor are used to indicate aggressive, inattentive, and normal driver profiles, respectively. We give the percentages of the number of car collisions caused by each pair of driver profiles rates regarding to all generated collisions. It is shown in Figure 7 that collisions generated by inattentive (veh1)-aggressive (veh2) represent the largest percentage, which is at 28.99% of the total number of simulated collisions.

Furthermore, Figure 8 shows the relationship between the number of simulated car collisions and the rates of aggressive and inattentive driver profiles, which presents in another way for the values of the result given in Table 7. From Figure 8, it seems that the sum of the two rates of aggressive and inattentive driver profiles is important in determining the number of car collisions. We can see in this figure that the contours (the different levels of numbers of collisions) have a regular relationship with the percentage of specific drivers, where the gap of different levels is similar and the contours have a similar shape. However, the contours are not linear which means that the sum of the rates of aggressive and inattentive driver profiles is important, but is not the unique parameter for the determination of the number of car collisions.

In order to better understand this relationship, we use the statistical regression method to assume that car collision counts (Y) are proportional to the percentage of aggressive drivers (x_1%) and the percentage of inattentive drivers (x_2%), which can be formulated as $Y = f(x_1, x_2)$.

We have first tried linear regression, and the result is shown in Appendix B, which shows that the assumption of linearity in not convincing. Moreover, we observe from Table 7 (red cases) that the number of car collisions generated with 30% of aggressive driver profiles and 30% of inattentive driver profiles is 60. Holding on this 60%, however, with 25% (also, respectively, 20%, 15%, 10%, 35%, 40%, 45%, and 50%) of aggressive driver profiles and 35% (also, respectively, 40%, 45%, 50%, 25%, 20%, 15%, and 10%) of inattentive driver profiles, the number of generated car collisions become less, although the sum of all the pairs of considered rates of aggressive and inattentive driver profiles is 60%. Therefore, it seems that the number of generated car collisions decreases with the increasing of the absolute difference between the rates of aggressive and inattentive driver profiles. In other terms, the number of generated car collisions increases with the uniformity of mixing both profiles (aggressive and inattentive). Based on this observation, we propose to add a quadratic term of the approximation of the relationship between the number of generated car collisions and the rates of aggressive and inattentive driver profiles.

$$\widehat{Y} = f(x_1, x_2) = \max\left(0, \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * (x_1 - x_2)^2\right), \tag{15}$$

where $\widehat{Y}$ is the estimation of car collision counts, x_1% is the percentage of aggressive drivers, and x_2% is the percentage of inattentive drivers. We set the lower bound of the function at zero, since the number of car collisions cannot be negative. We obtained the following coefficients: $\beta_0 = -23.35$, $\beta_1 = 1.4201$, $\beta_2 = 1.3602$, and $\beta_3 = -1.4377$, with RMSE = 4.864.

Table 7: Experiment 1: simulated car collision counts for different percentages of 1st aggressive and 1st inattentive drivers. For example, in the case that 25% for aggressive, 25% for inattentive, and the remaining 50% are normal drivers.

1st aggressive drivers (horizontal), 1st inattentive drivers (vertical)	2.5%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
2.5%	0	0	1	3	3	5	6	7.	11	11	13
5%	0	0	2	4	5	7	10	10	12	18	19
10%	4	4	7	10	12	16	21	26	29	38	43
15%	5	7	11	17	21	27	31	37	46	53	52
20%	8	11	16	18	24	31	38	46	56	64	69
25%	11	13	20	26	36	45	50	57	66	72	82
30%	12	16	22	30	39	49	60	62	75	87	92
35%	17	20	29	37	46	52	60	73	78	92	99
40%	19	22	28	35	50	55	65	76	88	96	106
45%	23	25	33	46	58	65	78	88	92	106	123
50%	25	32	42	52	62	74	87	95	105	118	124

Table 8: Experiment 2: simulated car collision counts for different percentages of 2nd aggressive and 1st inattentive drivers.

2nd aggressive drivers (horizontal), 1st inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0	0	1	3	4	5	7
5%_inatt	0	1	2	3	5	7	8
10%_inatt	4	5	5	9	13	16	17
20%_inatt	10	11	15	22	25	35	43
30%_inatt	12	14	17	24	32	39	52
40%_inatt	21	21	34	36	47	53	63
50%_inatt	27	29	37	48	61	69	80

Table 9: Experiment 3: simulated car collision counts for different percentages of 1st aggressive and 2nd inattentive drivers\enleadertwodots.

1st aggressive drivers (horizontal), 2nd inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0	0	2	4	4	8	10
5%_inatt	0	1	1	6	6	9	11
10%_inatt	2	2	2	9	15	18	20
20%_inatt	5	11	10	18	28	35	56
30%_inatt	11	11	18	25	38	54	69
40%_inatt	19	16	28	45	53	78	97
50%_inatt	24	29	32	55	73	82	101

Table 10: Experiment 4: simulated car collision counts for different percentages of 2nd aggressive and 2nd inattentive drivers.

2nd aggressive drivers (horizontal), 2nd inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0	0	1	2	2	4	8
5%_inatt	0	1	2	2	4	5	6
10%_inatt	2	1	2	4	6	10	14
20%_inatt	4	5	8	14	14	16	30
30%_inatt	8	12	12	16	28	30	42
40%_inatt	16	16	22	24	28	34	52
50%_inatt	25	20	26	38	42	52	58

By this result, first, the fact that $\beta_3 \neq 0$ confirms the significance of the effect of the absolute difference between the two rates of aggressive and inattentive driver profiles on the number of generated car collisions. Second, the fact that β_3 is negative confirms our hypotheses and shows the number of generated car collisions decreases with the increase of the absolute difference between the rates of aggressive and inattentive driver profiles.

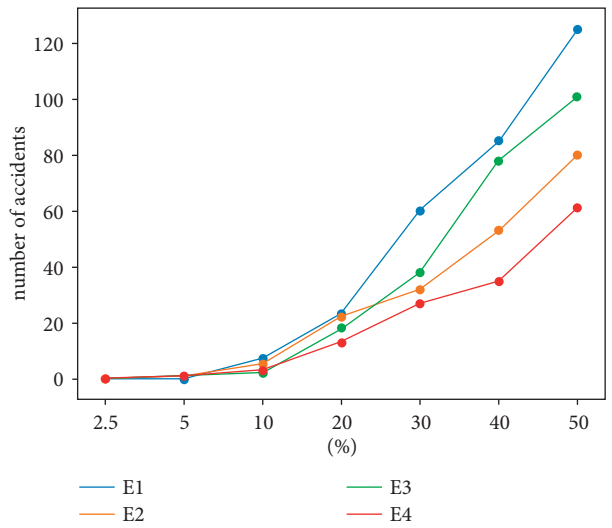

Figure 6: Simulated car collision counts for the four experiments, where aggressive and inattentive drivers have same rate (axis x) in traffic (based on the first 15-min data). For example, 2.5% (axis x) means each profile (aggressive, inattentive) takes 2.5% in the traffic.

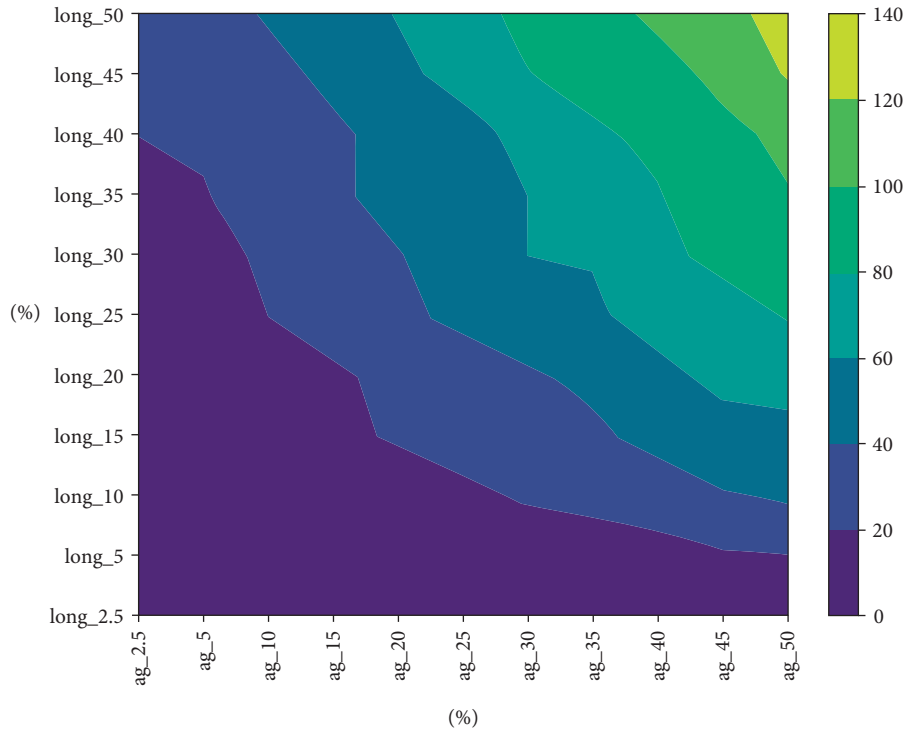

Figure 7: E1 : Simulated car -collision counts for different percentages of aggressive and inattentive drivers based on first 15-minute data. The suffixes ag and long are the aggressive and inattentive driver profile, respectively.

In Figure 9, the number of generated car collisions for each percentage of two drivers profiles is presented, as well as the surface obtained by the nonlinear regression function.

7. Validation of the Approach on the Second 15-Minute Time Period of Data

In this section, the same approach (Figure 1) for car collision generation is carried out based on the second 15-min time period of the NGSIM 101 dataset with the purpose of validation of the approach, where 1495 vehicles are registered in this period on the 5 normal traffic lanes. The thresholds on $\overline{\text{THW}}$ (mean THW) and min THW for each driver profile are given in Table 11. The simulation of traffic in SUMO for this period is shown in Figure 10. The simulated resulting car collisions, for the 4 experiments (E1-E4), are given in Tables 12–15.

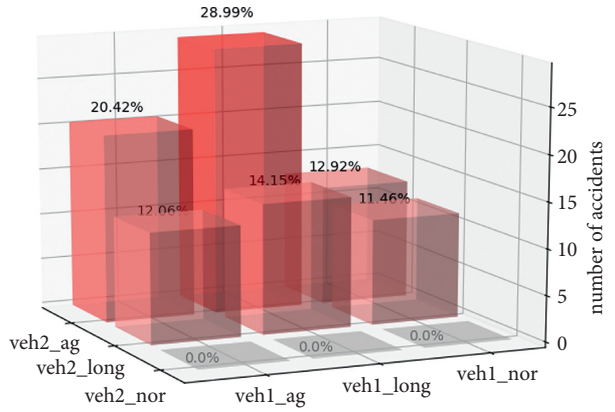

Figure 8: Distribution of involved driver profile types in simulated rear-end car -collisions (E1 based on first 15-minute data), veh1 indicates the leading vehicle, and veh2 indicates the following vehicle. The suffix ag, long, and nor are the aggressive, inattentive, and normal driver profiles, respectively.

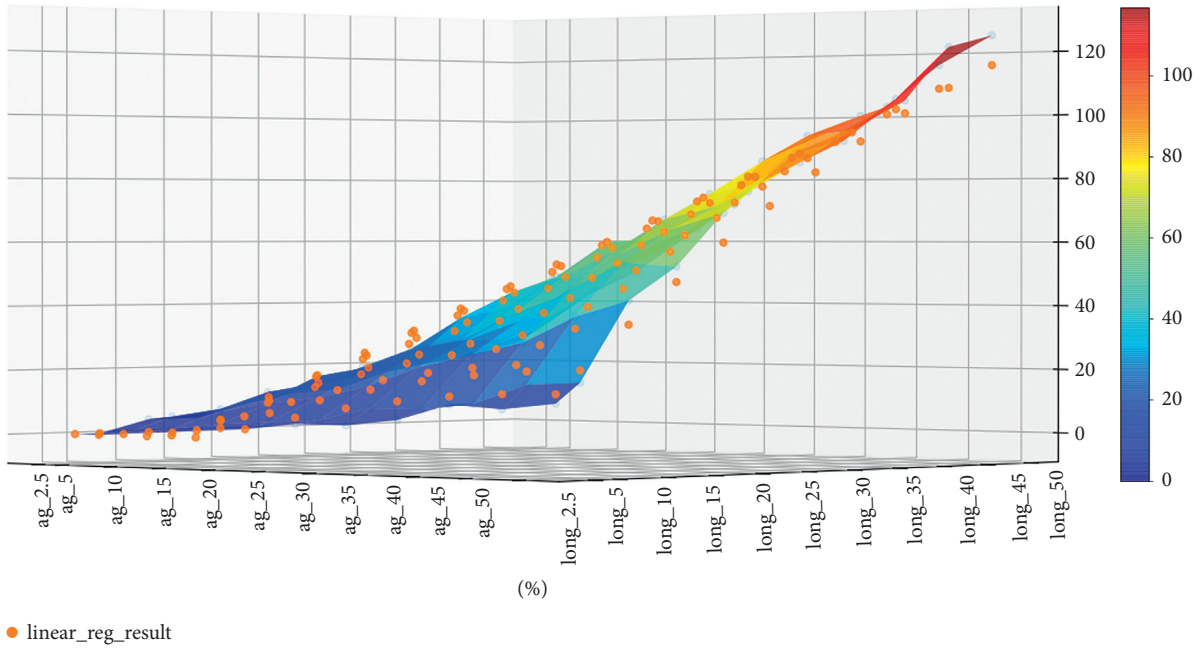

Figure 9: E1: nonlinear regression of the number of simulated car collisions. The suffixes ag and long are the aggressive and inattentive driver profile, respectively.

Table 11: Selected drivers and the associated threshold for the second 15-minute data.

Type of drivers	Indicator	Threshold
1st aggressive (group 1)	Mean THW	≤0.86*s*
2nd aggressive (group 2)	Mean THW	(0.86s, 0.98s)
1st inattentive (group 3)	min THW	≥1.99*s*
2nd inattentive (group 4)	min THW	(1.74s, 1.99s)

The result using the second 15-minute period data is close to the result using the first 15-minute period. The thresholds obtained from the four experiments are similar. The simulated car collisions using the second 15-minute data show the same trend as the first 15-minute data. Experiment E1 simulates the greatest number of car collisions, while E4 simulates the lowest number of car collisions (Tables 12–15). Furthermore, as shown in Figure 6 which shows the result for the experimentation based on the first 15-minute period data, Figure 11 shows similar results (number of generated crashes for the four experiments) for the experimentation based on the second 15-minute data. In Figure 11, each line presents the number of car collisions in the simulation resulted from one experiment. In Figure 11, experiment 1, where we generated car collisions based on varying the percentage of the most aggressive and the most inattentive

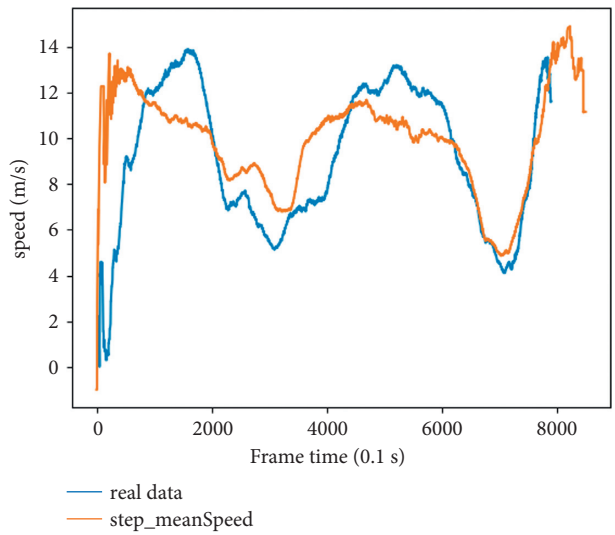

Figure 10: Average traffic flow speed of real data (blue) and simulation (orange) in SUMO for the second 15-min data.

Table 12: Experiment 1: simulated car collision counts by varying, respectively, the percentage of 1st aggressive and 1st inattentive drivers for second 15-min period.

1st aggressive drivers (horizontal), 1st inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0.0	2.0	4.0	9.0	11.0	17.0	20.0
5%_inatt	3.0	5.0	7.0	12.0	15.0	22.0	26.0
10%_inatt	11.0	14.0	16.0	21.0	30.0	34.0	39.0
20%_inatt	25.0	28.0	34.0	42.0	50.0	65.0	70.0
30%_inatt	34.0	39.0	45.0	51.0	56.0	68.0	79.0
40%_inatt	47.0	52.0	57.0	70.0	79.0	91.0	105.0
50%_inatt	58.0	60.0	68.0	79.0	82.0	102.0	115.0

Table 13: Experiment 2: simulated car collision counts by varying, respectively, the percentage of 2nd aggressive and 1st inattentive drivers over the second 15-minute period.

2st aggressive drivers (horizontal), 1st inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0.0	1.0	2.0	5.0	7.0	12.0	16.0
5%_inatt	4.0	5.0	6.0	9.0	13.0	15.0	19.0
10%_inatt	9.0	10.0	12.0	16.0	18.0	24.0	29.0
20%_inatt	22.0	25.0	26.0	32.0	37.0	41.0	44.0
30%_inatt	38.0	37.0	37.0	44.0	45.0	55.0	61.0
40%_inatt	44.0	48.0	52.0	56.0	61.0	67.0	78.0
50%_inatt	58.0	59.0	60.0	68.0	84.0	83.0	88.0

Table 14: Experiment 3: simulated car collision counts by varying, respectively, the percentage of 1st aggressive and 2nd inattentive drivers over the second 15-min period.

1st aggressive drivers (horizontal), 2nd inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0.0	1.0	4.0	10.0	12.0	19.0	20.0
5%_inatt	2.0	3.0	5.0	10.0	11.0	19.0	24.0
10%_inatt	6.0	7.0	8.0	14.0	18.0	26.0	31.0
20%_long	16.0	18.0	18.0	25.0	30.0	37.0	43.0
30%_inatt	22.0	26.0	29.0	33.0	38.0	47.0	54.0
40%_inatt	29.0	34.0	35.0	43.0	52.0	58.0	68.0
50%_inatt	33.0	33.0	40.0	50.0	53.0	69.0	72.0

drivers, can generate more car collisions than the other three experiments. Experiment 4 has probably the least number of car collisions, where we generated car collisions based on varying the percentage of the second group of aggressive drivers and the second group of inattentive drivers. These similar observations have been already obtained, as shown in Figure 6. In addition, the number of car collisions obtained in experiment E1 based on the second 15-min data is given

Table 15: Experiment 4: simulated car collisions number by varying, respectively, the percentage of 2nd aggressive drivers and 2nd inattentive drivers of second 15-min data.

2nd aggressive drivers (horizontal), 2nd inattentive drivers (vertical)	2.5%_ag	5%_ag	10%_ag	20%_ag	30%_ag	40%_ag	50%_ag
2.5%_inatt	0.0	1.0	2.0	7.0	9.0	10.0	15.0
5%_inatt	2.0	4.0	6.0	8.0	11.0	16.0	16.0
10%_inatt	5.0	6.0	7.0	10.0	12.0	17.0	19.0
20%_inatt	13.0	15.0	18.0	20.0	22.0	27.0	30.0
30%_inatt	20.0	22.0	23.0	26.0	31.0	35.0	38.0
40%_inatt	28.0	28.0	31.0	32.0	37.0	42.0	43.0
50%_inatt	30.0	32.0	32.0	37.0	40.0	45.0	46.0

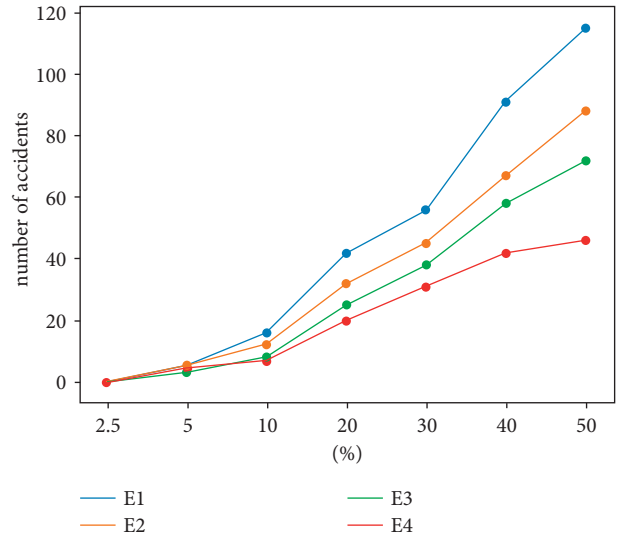

Figure 11: Simulated car collision counts for the four experiments, where aggressive and inattentive drivers have same rate (axis x) in traffic (based on the second 15-min data). For example, 2.5% (axis x) means each profile (aggressive, inattentive) takes 2.5% in the traffic.

in Table 12. We can see clearly that the results shown in Figure 12 are similar to the results shown in Figure 8. As the same, to further study on the relationship between the number of collision and the percentage of aggressive drivers and inattentive drivers in the whole traffic population, we applied to the same regression method for the first 15-min time period data. A nonlinear regression is performed (15), where we get the following coefficients: $\beta_0 = 0, -8.154, \beta_1 = 1.580, \beta_2 = 0.8502$, and $\beta_3 = -0.7575$, with an RMSE= 3.761.

8. Analysis of Crash Severity

All simulated car collisions can be further used to explore car collision severity. Several research efforts investigate the relationship of rear-end car collision severity to car speed. In [50], the authors reported that the critical impact speed was approximately 55 km/h for rear-end car collisions. In addition, in [51], the authors claim that the change of speed before and after car collision is a critical indicator for car collision severity outcomes.

In physics, the kinetic energy (KE) of an object is the energy that it possesses due to its speed (16). In the case of front-rear collision of two moving cars, the kinetic energy is related to the relative speed of the front vehicle and the rear car (17).

$$\mathrm{KE} = \frac{1}{2} m * v^2, \tag{16}$$

$$\mathrm{KE} \propto \Delta v. \tag{17}$$

Relative speed could thus be an important surrogate to indicate the severity of collision. In simulations with 50% drivers being aggressive and the other 50% being inattentive, 866 car collisions were generated (with 5 simulations). Among all car collisions, 220 (25.4%) concerned aggressive-aggressive couples of driver profiles, 408 (47.1%) concerned inattentive-aggressive, 107 (12.4%) concerned inattentive-inattentive, and 131 (15.1%) concerned aggressive-inattentive. In Figure 13, the distribution of relative speed for these 866 simulated collisions is shown. Car collisions involving inattentive (veh1)-aggressive (veh2) drivers accounted for the largest rate of all simulated collisions. All generated collisions have a relative speed below 50 km/h. Furthermore, collisions produced by two inattentive drivers are more severe, since its mean relative speed is at 23.27 km/h, which are more critical than the collision between other driver profiles. Car collision severity is much lighter between two aggressive drivers who have an average of relative speed at 15.7 km/h.

SCANeR studio [52, 53] is a driving simulation software, which also includes Bullet physics engine (a free and open-

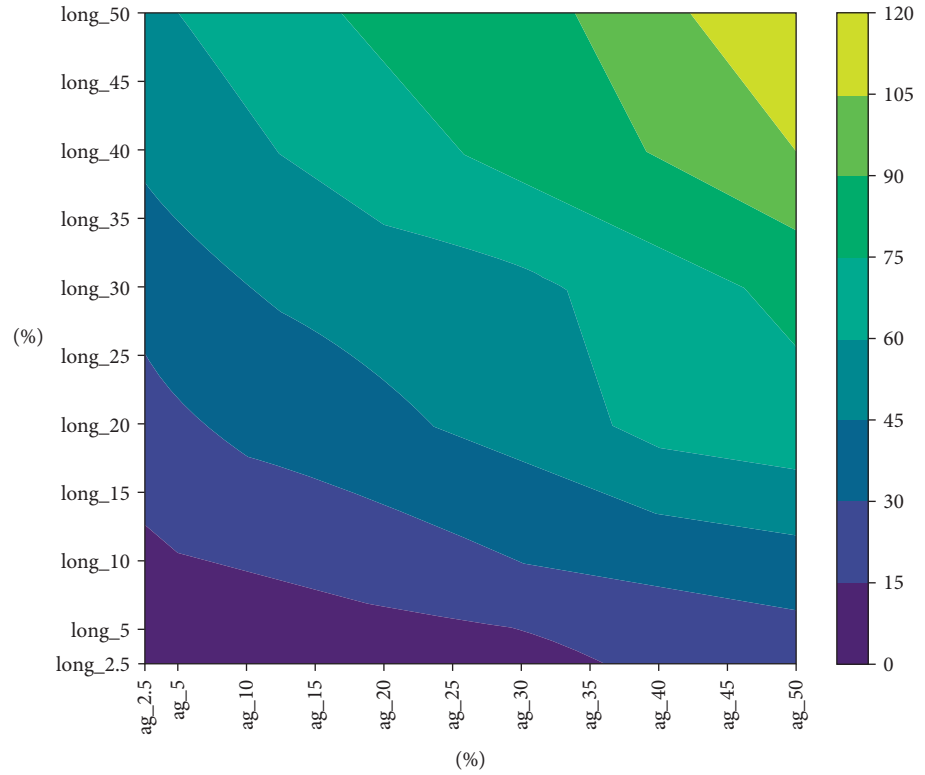

Figure 12: E1: simulated car collision counts for different percentages of aggressive and inattentive drivers based on second 15-minute data. The suffixes ag and long are the aggressive and inattentive driver profile, respectively.

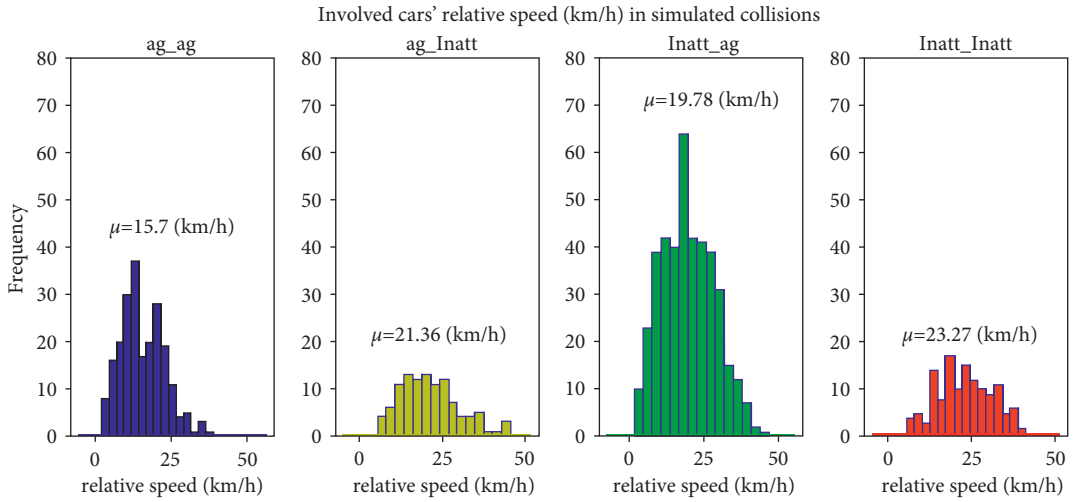

Figure 13: The distribution of relative speed for simulated car collisions (μ indicates the mean value for each subfigure). The suffixes ag and inatt are the aggressive and inattentive driver profile, respectively. It shows that the most car collisions are occurred between inattentive-aggressive drivers, and the most dangerous car collisions are produced between inattentive-inattentive drivers.

source software of simulation of collision detection, soft, and rigid body dynamics) and provides the opportunity to simulate vehicle car collisions with a vehicle physical model. For every simulated car collision in SUMO, the involved two vehicles' trajectory can be registered. Then, the car collision scene can be replayed in SCANeR Studio. Thus, the result of a car collision simulated in SUMO can be resimulated in SCANeR Studio (Figure 14). In Figure 15, the collision force for certain replayed car collisions is shown. Each point corresponds to one resimulated car collision by SCANeR. It seems that the force is a superlinear function of the relative speed (Figure 15). However, this fact needs more investigations and should be considered in our future research. The interest of this work will be to couple the SUMO traffic

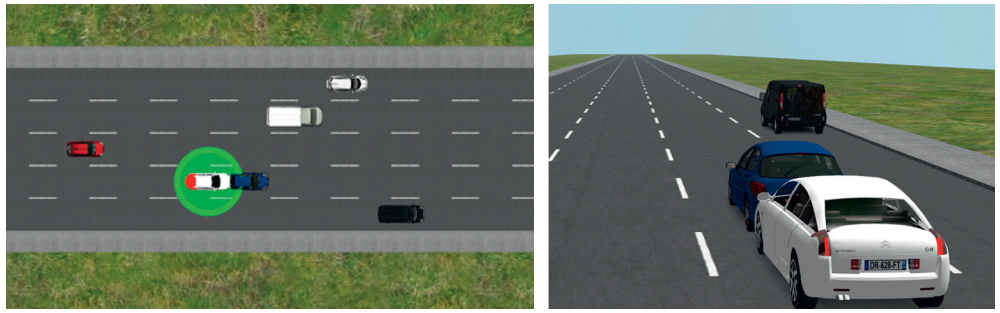

Figure 14: SCANeR screenshot for a collision.

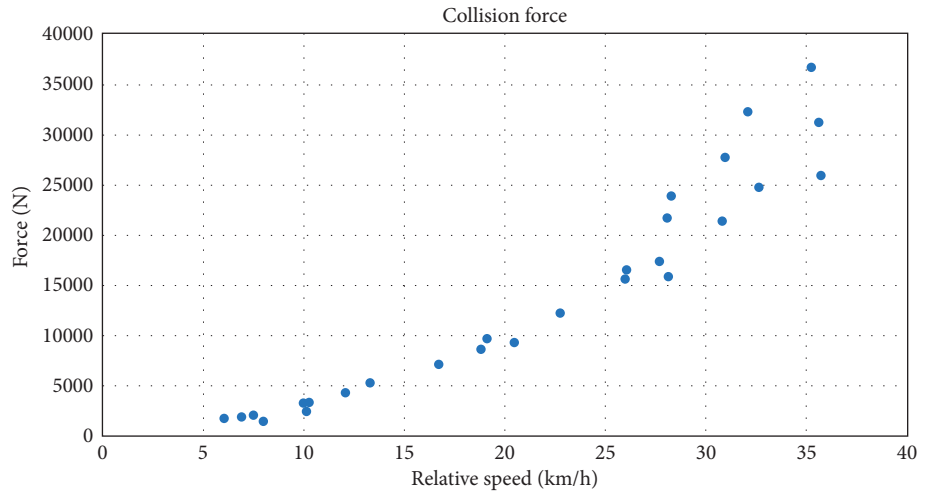

Figure 15: The collision force by SCANeR as a function of the relative speed (Y-axis for collision force (Newton) and X-axis for relative speed of collision (km/h)).

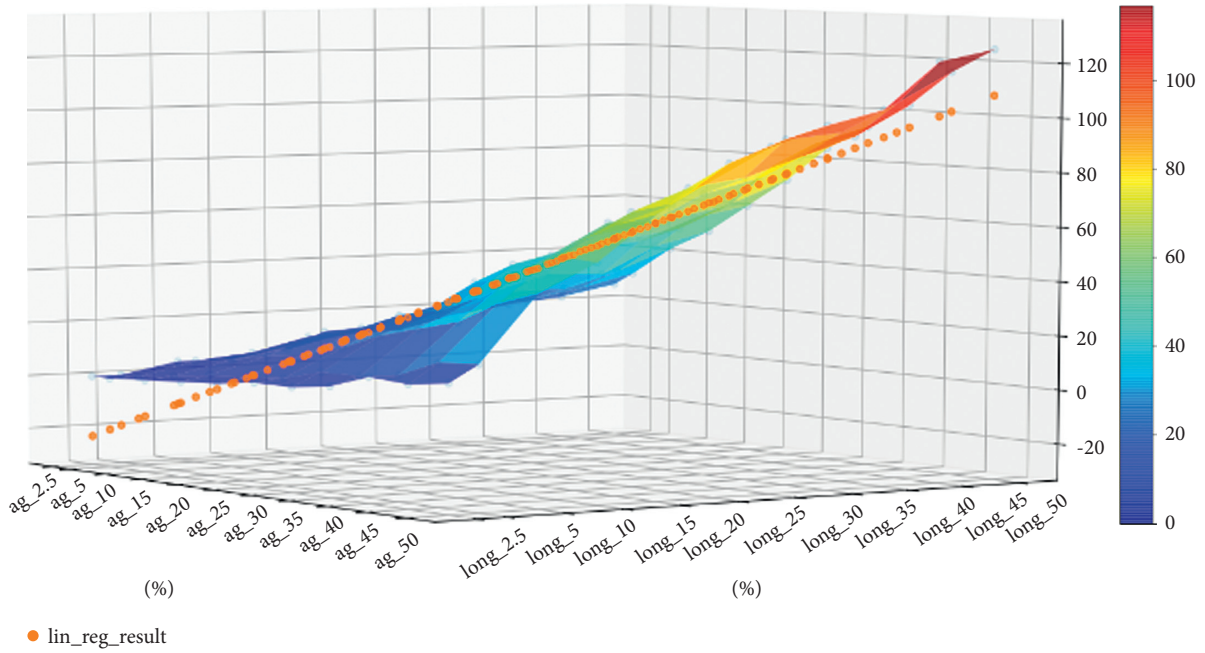

Figure 16: E1: linear regression of number of simulated car collisions. The suffixes ag and long are the aggressive and inattentive driver profile, respectively.

simulator with SCANeR studio (the immersive driving simulator) in order to build a high-performance system to test and validate autonomous vehicles.

9. Conclusion and Scope for Further Work

In this work, the main contributions consist of the following: extracting three proposed driver profiles based on real driving datasets, replicating these profiles in a simulated environment, and establishing a relationship between car collision occurrences and these different driver profiles by varying their percentages in the whole traffic via virtual simulation.

Based on the NGSIM 101 dataset, two specific driver profiles related to car collisions on road networks and a normal driver profile have been classified, defined as (i) aggressive drivers who keep short time headways with cars ahead of them, (ii) inattentive drivers with long reaction times, and (iii) normal drivers with intermediate values of reaction time and time headway. These three driver profiles have been simulated by using the intelligent driver model (IDM) car-following model, with an extension including driver reaction time. In order to represent the real driver profiles, the IDM model is calibrated using a genetic algorithm. Finally, by increasing the percentage of these two extreme driver profiles among all drivers in a virtual traffic simulation, we investigate the effect of these driver profiles on car collision occurrences.

The results of the numerical simulations show that the percentages of the aggressive and inattentive driver profiles over the whole driver population are determinant in the car collision occurrences and in the resulting severity outcomes. One of the important results we obtained in this work is the characterization of the relationship between the ratios of these two driver profiles over the whole driver population and the car collision occurrence count. We have also classified the car collisions that were generated and analyzed their severity, in particular with respect to the relative speed between the cars involved in the collisions. Another important result of this research is that the car collisions involving an aggressive driver in the leading vehicle and an inattentive driver following represent the most frequent collision occurrences, while collisions between two inattentive drivers were the most severe.

The safety validation of intelligent connected vehicles is essential and could be critical for their deployment. In order to complete the demonstration of the reliability of the system's safety, autonomous vehicles need to be driven for hundreds of millions of miles. However, the huge cost of such physical tests, combined with the inherent danger of testing situations where collisions can happen, makes the numerical simulation of scenarios mixing different driver profiles an important safety assessment tool for ICV testing. Furthermore, during the deployment phase of ICVs, the recognition of drivers' profiles should be considered in order to avoid collisions. As shown by our experiments, based on our approach, different traffic scenarios can be generated with different driver profiles in traffic simulations. For future research, this work is expected to greatly facilitate future ICV testing and validation for the car manufacturing industry via numerical traffic simulation.

In this work, we used the IDM model and the simulator SUMO to generate car collisions. The method is limited by calibration accuracy and model performance; therefore, the car-following model is not able to replicate 100% real human driving behavior. Regarding the perspectives for our proposed approach, tests with different car-following models need to be considered in further works. Moreover, this work is an experiment based on the NGSIM 101 dataset, and the profiles are extracted from this dataset. We also intend to work in the future with other datasets in order to confirm the derived driver profiles and/or derive new driver profiles, related to car collisions occurrences.

Appendix

A. Calibration Result for Specific Drivers

The results of calibration for 4 specific groups of driver in the 15-minute data of the NGSIM 101 dataset are given as follows: in Table 2 for drivers of group 1, in Table 3 for drivers of group 2, in Table 4 for drivers of group 3, and in Table 5 for drivers of group 4. These tables show the mean, median, standard deviation, 25% quantile, and 75% quantile, of each parameter and for each group drivers. The tables give also the bounds for the genetic algorithm.

B. Linear Regression of the Number of Simulated Car Collisions

In order to find the relationship between simulates car collisions number and the rate of specific profile drivers for the result given in Table 3, we obtain $\widehat{Y} = f(x_1, x_2) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2$, with $\beta_0 = -29.5569, \beta_1 = 1.4068$, and $\beta_2 = 1.3468$. The obtained root mean square error (RMSE) is 9.7838. In Figure 16, the number of car collisions for each percentage of two drivers profiles is presented, as well as the plane obtained by regression. This result allows us to estimate the proportion between aggressive and inattentive driver profiles generating the same number of car collisions:

$$\frac{\beta_2}{\beta_1} = \frac{13}{14} \approx 1. \tag{B.1}$$

Although the linear regression is satisfying with RMSE = 9.7838 and $R^2 = 0.9045$, we can see on Figure 1 that we need to still improve approximation.

Data Availability

This is a public dataset for traffic research. The homepage of this dataset is as follows: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167–181, 2015.

[2] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, 2018.

[3] Q. Chao, H. Bi, W. Li et al., "A survey on visual traffic simulation: models, evaluations, and applications in autonomous driving," in *Computer Graphics Forum*vol. 39, Wiley Online Library, 2019.

[4] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of sumo-simulation of urban mobility," *International Journal of Agile Systems and Management*, vol. 5, no. 3&4, 2012.

[5] F. Martin and P. Vortisch, "Microscopic traffic flow simulator vissim," in *Fundamentals of Traffic Simulation*, pp. 63–93, Springer, New York, NY, USA, 2010.

[6] J. Casas, J. L. Ferrer, D. Garcia, J. Perarnau, and A. Torday, "Traffic simulation with aimsun," in *Fundamentals of Traffic Simulation*, pp. 173–232, Springer, New York, NY, USA, 2010.

[7] G. A. M. Meiring and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, Article ID 30653, 2015.

[8] A. Ahmad, N. Alshabatat, and M. S. N. Al-Din, "Driving style recognition using fuzzy logic," in *Proceedings of the 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*, pp. 460–463, IEEE, Istanbul, Turkey, July 2012.

[9] V. Alexiadis, C. James, J. Halkias, R. Hranac, and G. McHale, "The next generation simulation program," *Institute of Transportation Engineers. ITEA Journal*, vol. 74, no. 8, p. 22, 2004.

[10] Z. He, "Research based on high-fidelity ngsim vehicle trajectory datasets: a review," pp. 1–33, Research Gate, 2017.

[11] P. Sun, H. Kretzschmar, X. Dotiwalla et al., "Scalability in perception for autonomous driving: waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2446–2454, Seattle, WA, USA, June 2020.

[12] M. F. Chang, J. Lambert, P. Sangkloy et al., "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8748–8757, Long Beach, CA, USA, June 2019.

[13] H. Caesar, V. Bankiti, A. H. Lang et al., "nuscenes: a multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Article ID 11621, Seattle, WA, USA, June 2020.

[14] S. Omar, N. Hawkins, Z. Hans et al., "Naturalistic Driving Study: Development of the Roadway Information Database," Report S2-S04A-RW-1, TRB Publication, Washington, D.C., USA, 2015.

[15] V. Astarita, C. Caliendo, V. P. Giofrè, and I. Russo, "Surrogate safety measures from traffic simulation: validation of safety indicators with intersection traffic crash data," *Sustainability*, vol. 12, no. 17, p. 6974, 2020.

[16] A. Abdulhafedh, "Road traffic crash data: an overview on sources, problems, and collection methods," *Journal of Transportation Technologies*, vol. 7, no. 2, pp. 206–219, 2017.

[17] X. Shi, Y. D. Wong, M. Z. F. Li, and C. Chai, "Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory," *Accident Analysis & Prevention*, vol. 117, pp. 346–356, 2018.

[18] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.

[19] H. H. Salem, N. Farhi, and J. P. Lebacque, "Risk index model building: application for field ramp metering safety evaluation on urban motorway traffic in paris," *Cybernetics and Information Technologies*, vol. 13, no. 4, pp. 33–41, 2013.

[20] Z. Christoforou, S. Cohen, and M. G. Karlaftis, "Identifying crash type propensity using real-time traffic data on freeways," *Journal of Safety Research*, vol. 42, no. 1, pp. 43–50, 2011.

[21] N. Dong, H. Huang, and L. Zheng, "Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects," *Accident Analysis & Prevention*, vol. 82, pp. 192–198, 2015.

[22] C. Lima Azevedo, J. L. Cardoso, and M. E. Ben-Akiva, "Probabilistic safety analysis using traffic microscopic simulation," 2018, https://arxiv.org/abs/1810.04776.

[23] J. Wang, Y. Kong, T. Fu, and J. Stipancic, "The impact of vehicle moving violations and freeway traffic flow on crash risk: an application of plugin development for microsimulation," *PLoS One*, vol. 12, no. 9, Article ID e0184564, 2017.

[24] L. Zheng, K. Ismail, and X. Meng, "Traffic conflict techniques for road safety analysis: open questions and some insights," *Canadian Journal of Civil Engineering*, vol. 41, no. 7, pp. 633–641, 2014.

[25] N. S. Kusumastutie and P. Rusmandani, "A brief review: traffic conflict techniques and the challenges of the studies in Indonesia," in *Proceedings of the MATEC Web of Conferences*, vol. 270, p. 3004, EDP Sciences, Bandung, Indonesia, November 2019.

[26] R. Jiang, S. Zhu, P. Wang, Q. C. Chen, Z. He, and S. Kuang, "In search of the consequence severity of traffic conflict," *Journal of Advanced Transportation*, vol. 2020, Article ID 9089817, 15 pages, 2020.

[27] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility: an overview," in *Proceedings of the SIMUL 2011, The Third International Conference on Advances in System Simulation*, ThinkMind, Barcelona, Spain, October 2011.

[28] P. Alvarez Lopez, M. Behrisch, L. Bieker-Walz et al., "Microscopic traffic simulation using sumo," in *Proceedings of the 2018 Twenty First International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, IEEE, Maui, HI, USA, November 2018.

[29] L. Vasconcelos, L. Neto, S. Santos, A. B. Silva, and Á. Seco, "Calibration of the gipps car-following model using trajectory data," *Transportation research procedia*, vol. 3, pp. 952–961, 2014.

[30] M. Zhu, X. Wang, A. Tarko, and S. E Fang, "Modeling car-following behavior on urban expressways in shanghai: a naturalistic driving study," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 425–445, 2018.

[31] T. H. Itkonen, J. Pekkanen, O. Lappi, I. Kosonen, T. Luttinen, and H. Summala, "Trade-off between jerk and time headway as an indicator of driving style," *PloS one*, vol. 12, no. 10, Article ID e0185856, 2017.

[32] I. W. Suweda, "Time headway analysis to determine the road capacity," *Jurnal Spektran*, vol. 4, no. 2, 2016.

[33] T. J. Ayres, L. Li, D. Schleuning, and D. Young, "Preferred time-headway of highway drivers," in *Proceedings of the ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585)*, pp. 826–829, IEEE, Oakland, CA, USA, August 2001.

[34] W. V. Winsum and A. Heino, "Choice of time-headway in car-following and the role of time-to-collision information in braking," *Ergonomics*, vol. 39, no. 4, pp. 579–592, 1996.
[35] A. K. Maurya, S. Das, S. Dey, and S. Nama, "Study on speed and time-headway distributions on two-lane bidirectional road in heterogeneous traffic condition," *Transportation Research Procedia*, vol. 17, pp. 428–437, 2016.
[36] SUMO, "Simulation of Collision in Sumo," 2019, https://sumo.dlr.de/docs/Simulation/Safety.html.
[37] M. Brackstone and M. McDonald, "Car-following: a historical review," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 2, no. 4, pp. 181–196, 1999.
[38] S. Mohammad and Z. Zheng, "Incorporating human-factors in car-following models: a review of recent developments and research needs," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 379–403, 2014.
[39] K. Aghabayk, M. Sarvi, and W. Young, "A state-of-the-art review of car-following models with particular considerations of heavy vehicles," *Transport Reviews*, vol. 35, no. 1, pp. 82–105, 2015.
[40] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic dynamics: studies in car following," *Operations Research*, vol. 6, no. 2, pp. 165–184, 1958.
[41] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, 1981.
[42] S. Krauß, "Microscopic modeling of traffic flow: investigation of collision free vehicle dynamics," PhD thesis, Dt. Zentrum für Luft-und Raumfahrt eV, Abt. Unternehmensorganisation und, German Aerospace Center (DLR), Cologne, Germany, 1998.
[43] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review*, vol. 62, no. 2, pp. 1805–1824, 2000.
[44] Z. Lu, T. Fu, L. Fu, S. Shiravi, and C. Jiang, "A video-based approach to calibrating car-following parameters in vissim for urban traffic," *International journal of transportation science and technology*, vol. 5, no. 1, pp. 1–9, 2016.
[45] M. Treiber and A. Kesting, "Microscopic calibration and validation of car-following models - a systematic approach," *Procedia - Social and Behavioral Sciences*, vol. 80, pp. 922–939, 2013.
[46] Matlab, "R2020b. Optimization Toolbox," 2020, https://fr.mathworks.com/help/gads/genetic-algorithm-options.html.
[47] S. S. Rao, *Engineering Optimization: Theory and Practice*, John Wiley & Sons, Hoboken, NJ, USA, 5 edition, 2020.
[48] Matlab, "Genetic Algorithm in Matlab R2019b," 2019, https://fr.mathworks.com/help/gads/genetic-algorithm-options.html.
[49] SUMO, "Netedit Documentation," 2019, https://sumo.dlr.de/docs/NETEDIT.html.
[50] C. Jurewicz, A. Sobhani, J. Woolley, J. Dutschke, and B. Corben, "Exploration of vehicle impact speed - injury severity relationships for application in safer road design," *Transportation research procedia*, vol. 14, pp. 4247–4256, 2016.
[51] R. Elvik, "A re-parameterisation of the power model of the relationship between the speed of traffic and the number of accidents and accident victims," *Accident Analysis & Prevention*, vol. 50, pp. 854–860, 2013.
[52] AVSimulation Scaner Studio Driving Simulator, https://www.avsimulation.com/scanerstudio/, 2019.
[53] A. Champion, R. Mandiau, C. Kolski, H. Alexandre, and A. Kemeny, "Traffic generation with the scaner© ii simulator: towards a multi-agent architecture," in *Proceedings of the Driving Simulation Conference*, pp. 311–324, Paris, France, July 1999.