Investigation on car-following heterogeneity and its impacts on traffic safety and sustainability

Yao, Xue; Yan, Qiruo; Sun, Zhanbo; Calvert, Simeon C.; Hoogendoorn, Serge P.

DOI
10.1080/23249935.2024.2407077

Publication date
2024

Document Version
Final published version

Published in
Transportmetrica A: Transport Science

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ttra21

Investigation on car-following heterogeneity and its impacts on traffic safety and sustainability

Xue Yao, Qiruo Yan, Zhanbo Sun, Simeon C. Calvert & Serge P. Hoogendoorn

To cite this article: Xue Yao, Qiruo Yan, Zhanbo Sun, Simeon C. Calvert & Serge P. Hoogendoorn (21 Sep 2024): Investigation on car-following heterogeneity and its impacts on traffic safety and sustainability, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2024.2407077

To link to this article: https://doi.org/10.1080/23249935.2024.2407077

Published online: 21 Sep 2024.

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2024.2407077

Investigation on car-following heterogeneity and its impacts on traffic safety and sustainability

Xue Yao[a], Qiruo Yan[b], Zhanbo Sun[b], Simeon C. Calvert[a] and Serge P. Hoogendoorn[a]

[a]Department of Transport & Planning, Delft University of Technology, Delft, The Netherlands;
[b]School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, People's Republic of China

ABSTRACT
This study proposes a general framework to investigate car-following heterogeneity and its impacts on traffic safety and sustainability. The framework incorporates rigorous driving style classification using a semi-supervised learning technique and a micro-simulation process that includes 66 fine-grained traffic scenarios exhibiting varying degrees of heterogeneity. Validated using two distinct real-world datasets reveals the superiority of S3VM-based classifiers over traditional SVM classifiers in driving style classification. Simulation results show that an increase in driving aggressiveness is correlated with higher safety issues and greater environmental impacts. Further elucidation of these impacts from the mechanism of underlying characteristics of driving behaviour and traffic flow dynamics indicates that less aggressive drivers can lead to the formation of vehicular platoons, thereby encouraging more aggressive drivers to adopt a milder driving style. Importantly, the formation of these platoons is influenced by both the proportion and spatial distribution of less aggressive vehicles. The proposed approach promises advantages in reducing the negative impacts of driving heterogeneity, thus benefiting Intelligent Transportation Systems (ITS) by improving traffic safety and sustainability.

ARTICLE HISTORY
Received 23 August 2023
Accepted 14 September 2024

KEYWORDS
Car-following; driving heterogeneity identification; micro-simulation; traffic safety; sustainability

1. Introduction

Driving heterogeneity constitutes a vital subject of investigation across various domains, including human-centred vehicle control systems, intelligent transportation systems, road safety, and environmental management. Studies have revealed that the variability in driving behaviours can lead to traffic externalities such as manifesting traffic hysteresis (D. Chen et al. 2014), causing more traffic accidents (Y. Chen, Wang, and Lu 2023), fuel consumption and emissions (Lárusdóttir and Ulfarsson 2015; Makridis et al. 2020; Yao et al. 2024). For example, reaction time and sensitivity to stimuli are directly associated with rear-end collisions, thus contributing to traffic accidents (J. Zhang, Wang, and Lu 2019; Y. Zhang and Talebpour 2024). Also, extreme driving actions, such as over-speeding, excessive acceleration, and sudden stops, are more fuel-intensive (Haque and Abas 2018), thereby increasing

CONTACT Xue Yao x.yao-3@tudelft.nl Department of Transport & Planning, Delft University of Technology, Stevinweg 1, 2628 CN, Delft, The Netherlands

emissions and energy consumption (Lárusdóttir and Ulfarsson 2015; Sun et al. 2024a). These findings underscore the necessity to investigate the impacts of driving heterogeneity on traffic flow performance and to understand them from the underlying mechanisms of driving behaviour, which can help develop measures to improve traffic safety and environmental sustainability.

According to Ossen and Hoogendoorn (2011), driving heterogeneity is defined as the difference between driving behaviours of driver/vehicle combinations under comparable conditions. In literature, driving heterogeneity is usually identified from observed driving behaviours and formulated as a classification problem with the output being categorical (discrete scales vary between two and eight levels) or numerical (score of 0 to 10). Vehicular variables such as velocity, acceleration, and braking have been widely used to identify drivers' different driving styles (Fadhloun and Rakha 2020; Yao, Calvert, and Hoogendoorn 2023). Calibration of car-following model parameters is another prevalent way to characterise different driving behaviours (Makridis et al. 2023; Shang and Stern 2020). Utilising these trajectory-related variables, computational models such as machine learning (ML) techniques are applied to characterise driving behaviour into distinct groups and infer specific driving profiles. Supervised and unsupervised learning are commonly used ML techniques for this purpose. For example, supervised learning classifiers such as Support Vector Machine (SVM), Random Forest (RF), K Nearest Neighbours (KNN), and Multilayer Perceptron (MLP) were employed to differentiate normal and aggressive driving styles, achieving accuracy as high as 91.7% (Xue et al. 2019). A k-means clustering algorithm was implemented to recognise driving profiles into usual, harsh, and eco-driving based on speed and acceleration data (Adamidis, Mantouka, and Vlahogianni 2020). Note that manually labelling of data for supervised learning is usually time-intensive and can introduce bias, and solely relying on unlabelled data in unsupervised learning may lead to unpredictable outcomes. Semi-supervised learning techniques offer a promising solution to overcome these challenges. They train classifiers to identify driving heterogeneity based on both labelled and unlabelled driving data, which can capture more characteristics of driving data and uncover heterogeneity (Bennett and Demiriz 1998). W. Wang et al. (2017) classified driving behaviour as aggressive and normal styles using both SVM and S3VM and revealed the superiority of the S3VM-based models over SVM-based models.

Based on the identification of driving heterogeneity, many efforts have been made to investigate the impact of driving styles on traffic performance by reproducing heterogeneous driving behaviours in microsimulation. Some studies developed stochastic car-following models by adding time-varying random noise (e.g. white noise) or distributions to deterministic models (Kesting, Treiber, and Helbing 2010; Suarez et al. 2022). For example, Zheng et al. (2023) developed a parsimonious enhanced Newell's car-following model incorporating the stochastic reaction time and the fluctuation around the vehicle's desired speed subject to the mean reversion process. The Rakha-Pasumarthy-Adjerid (RPA) car-following model has shown its capability to generate realistic VSP distributions and estimate fuel consumption and emissions (Fadhloun and Rakha 2020; J. Wang, Rakha, and Fadhloun 2017). Additionally, both variations in driving styles and vehicle characteristics are introduced in the Microsimulation Free-flow aCceleration (MFC) model, which has been utilised to simulate heterogeneous driving behaviour and estimate fuel consumptions and emissions (Makridis et al. 2020). Based on developed microsimulation methods, the impacts of car-following heterogeneity on traffic flow performance have been revealed. A study reported that promoting more stable driving styles during car-following can potentially mitigate the risk of rear-end collisions (J. Zhang, Wang, and Lu 2019). On the contrary, aggressive driving styles are associated with higher levels of speed variability, elevated engine revolutions, and a greater likelihood of road accidents compared to other driving styles (Haque and Abas 2018; Sun et al. 2024b). Additionally, aggressive driving increased fuel costs significantly (Szumska and Jurecki 2020), by more than 20% (Bakhit, Said, and Radwan 2015) and by 25% in urban areas (Fontaras, Zacharof, and Ciuffo 2017).

Despite the extensive research on heterogeneous driving behaviour modelling and the investigation of the impacts on traffic safety and sustainability, there are still some aspects that need to be explored. Usually, driving behaviours are diverse, and the corresponding driving heterogeneity identification and classified driving behaviour modelling require rigorous analyses, serving as an important preparation for traffic simulation. Furthermore, heterogeneous traffic flow needs more fine-grained traffic scenarios to demonstrate its diversity rather than a small number of representative fixed driving style ratios. Meanwhile, there is a need for nuanced investigation into how these influences happen from the mechanism of underlying characteristics of driving behaviour and traffic flow.

To bridge these research gaps, we propose a general micro-simulation approach to evaluate the impacts of car-following heterogeneity on traffic safety and sustainability. The novel contributions of this study are threefold: (i) A semi-supervised learning method, i.e. multi-classification S3VM, is developed to facilitate rigorous driving style classification and classified car-following behaviour modelling. (ii) Heterogeneous traffic flow is refined by 66 distinct traffic scenarios with varying degrees of heterogeneity, which allows a more nuanced examination of the impact of different driving styles and their proportion changes on traffic safety, fuel consumption and emissions. (iii) The impacts caused by car-following heterogeneity are elucidated from the mechanism of underlying characteristics of driving behaviour and traffic flow dynamics.

The remainder of the paper is organised as follows. Section 2 introduces the methodology for assessing the impacts of car-following heterogeneity on traffic flow. Section 3 presents experimental settings, including driving style classification and micro-simulation setup. Results and discussions are provided in Section 4, and conclusions with main findings and future research are finally presented in Section 5.

2. Methodology

This section outlines the methodology for assessing the impacts of car-following heterogeneity on traffic performance using a micro-simulation approach. The process is illustrated in Figure 1. Data plays an important role in ML-based driving heterogeneity identification, thus being prepared as the first step of this methodology. Utilising extracted data of car-following pairs, a multi-class semi-supervised Support Vector Machine (S3VM) is developed to classify drivers into different driving styles. Based on the classification results, 66 refined traffic flow scenarios representing varying degrees of heterogeneity are established

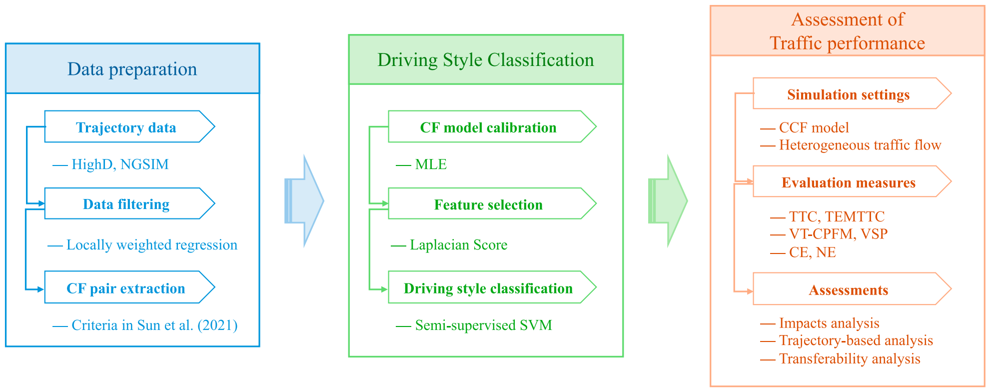


Figure 1. Overview of the proposed methodology.

in micro-simulation, where various indicators are adopted to estimate traffic safety, and fuel consumption and emissions.

2.1. Data preparation

The HighD and NGSIM datasets are key and commonly used datasets for highway driving behaviour analysis and are utilised to evaluate the proposed methodology in this study. The HighD dataset captures vehicle trajectories at 25 fps (one frame per 0.04 seconds) through a high-resolution drone-mounted camera, which is notable for its ability to capture a wide range of behaviours and interactions among vehicles from a bird's-eye view (Krajewski et al. 2018). Compared to the NGSIM dataset, HighD has higher granularity with longer recorded duration, driven distance and driven time. Therefore, we adopt HighD as the primary dataset for evaluation and NGSIM for transferability analysis of findings.

It is acknowledged the presence of errors and noise in original trajectory datasets could potentially impact the accuracy of microscopic studies (Aghabayk, Sarvi, and Young 2016). Thus, we conduct data smoothing and filtering according to reference Sun et al. (2021) before the extraction of car-following segments. This car-following pairs extraction process applies several criteria aimed at excluding congestion and free flow conditions, filtering minimum driving during, etc (Sun et al. 2021). Finally, 2744 and 1097 car-following trajectory pairs are extracted from HighD and NGSIM datasets, respectively, with each lasting a minimum of 30 seconds.

2.2. Driving style classification

2.2.1. Car-following model calibration

Car-following models capture the dynamics of longitudinal interactions between two adjacent vehicles navigating in the same lane without overtaking, aiming to simulate and understand driving behaviours in diverse traffic conditions. From a physics point of view, a car-following model used for microscopic traffic simulation should be as simple as possible (Treiber, Hennecke, and Helbing 2000b). According to Treiber, Hennecke, and Helbing (2000b), several criteria should be checked when choosing a microscopic traffic model. First, the parameters of traffic models should be intuitive and easy to calibrate, and the corresponding values should be realistic, facilitating the replication of realistic driving behaviour in simulation. Additionally, the model should be capable of representing a variety of traffic conditions, including congestion and hysteresis effects, to ensure simulations accurately mirror real-world traffic dynamics. It's also crucial that the model prevents vehicle collisions and supports efficient numerical simulation.

The time-continuous Intelligent Driver Model (IDM) model is highlighted for its simplicity and effectiveness in simulating realistic acceleration profiles and behaviours in nearly all single-lane traffic scenarios (Treiber and Kesting 2013), without leading to accidents. This physical model has interpretable parameters, promising advantages compared to machine learning-based models. To this end, IDM is widely used for traffic stability analysis and oscillation analysis in the literature and is capable of satisfactorily reproducing many characteristics of traffic flow (Ngoduy 2013, 2015; Shang and Stern 2021; Treiber and Kesting 2017). The acceleration assumed in the IDM is a continuous function of the velocity v, the gap s, and the velocity difference Δv between the following vehicle and its leading vehicle, as illustrated in Equation (1).

$$a(t_{i+1}) = a_0 \left[1 - \left(\frac{v(t_i)}{v_0} \right)^{\delta} \right] - \left(\frac{s^*(v(t_i), \Delta v)}{s(t_i)} \right)^2 \tag{1}$$

$$s^*\,(v(t_i), \Delta v) = s_0 + \max \left(0, v(t_{i+1})\,T + \frac{v(t_i)\Delta v}{2\sqrt{a_0\, b_0}} \right) \tag{2}$$

where a_0 is the maximum acceleration/deceleration of the follower, δ is the acceleration index, which is set the value of 4 to reduce additional safety issues (Treiber, Hennecke, and Helbing 2000a), v_0 is the desired speed and s_0 is the minimum distance gap. $s^*(v(t_i), \Delta v)$ means the desired gap, which is a function of $v(t_i)$ and Δv as shown in Equation (2), in which T is the safety time gap and b_0 is the comfortable deceleration.

Prior research has identified some limitations of the IDM, such as its inability to reproduce high-dimensional oscillations (Treiber and Kesting 2017). Huang et al. (2018) proposed a stochastic IDM that can qualitatively replicate the concave oscillatory growth pattern. The model adopts white noise to represent driver uncertainty, as shown in Equation (3). We utilise this stochastic IDM model to investigate the impacts of car-following heterogeneity on traffic flow performance.

$$v\,(t_{i+1}) = \max \left[\min\,(v\,(t_i) + a(t_{i+1})\Delta t + \text{noise}, v_0)\,, 0 \right] \tag{3}$$

where noise $=$ norm$(0, \sqrt{Q\Delta t})$, Q denotes the noise strength.

Model calibration was conducted at the individual vehicle level in this study, which enables the assignment of specific parameters to each driver indicative of his/her driving behaviour. The calibration of the car-following model involved identifying the most suitable parameter set to minimise the error between simulated and measured trajectories. Thereafter, traffic flow will be simulated using a mixture of drivers with their personalised parameters for car-following behaviour rather than using group average value. Recognising observed limitations of commonly optimised algorithms such as genetic algorithms and random parameter algorithms, Maximum Likelihood Estimation (MLE) is applied for model calibration in this study. Details about the MLE-based calibration procedure were described in Hoogendoorn and Hoogendoorn (2010).

2.2.2. Feature selection

Most car-following models usually have more than five parameters. High-dimensional features easily lead to feature redundancy, and irrelevant features usually negatively affect training efficiency and accuracy. We adopt the Laplacian Score (LS) for dimensionality reduction due to its ability to identify features that represent the data's underlying structure. The idea of LS is to evaluate features according to their locality-preserving power, e.g. two data points are more likely the same type if they are close to each other. In classification problems, the local geometric structure of the data space is more important than the global structure. The Laplacian Score for each feature is calculated based on how much the feature values change within local neighbourhoods. Features that have small variations within neighbourhoods but significant variations between different neighbourhoods are considered more important. For more information about LS, please refer to He, Cai, and Niyogi (2005). Additionally, other commonly used unsupervised feature selection methods, i.e. Principal Component analysis (PCA) and t-distributed stochastic neighbourhood embedding (t-SNE) are employed for the same feature selection task to verify the effectiveness of LS.

2.2.3. Semi-supervised driving style classification

Based on the selected features, we develop a semi-supervised approach, i.e. a multi-class semi-supervised support vector machine (S3VM), to classify driving style during car-following. S3VM is constructed using a mixture of labelled data (the training set) and unlabelled data (the working set), in which class labels in the training set are assigned to the working set to construct the 'best' SVM (He, Cai, and Niyogi 2005). The process of multi-class S3VM can be divided into multiple binary classification problems and solved using the 'one-vs.-one' or 'one-vs.-rest' strategy (Sun and Ban 2018). For the three-class problem in this paper, the 'one-vs.-one' approach can use a smaller size of classifiers and less computational time to get similar results as the 'one-vs.-rest' approach. Thus, the 'one-vs.-one' approach is adopted in this study. The dataset $\mathcal{S}$ of S3VM consists of labelled and unlabelled data points: $\{\mathcal{S}^{(l)}, \mathcal{S}^{(u)}\} = \{(x_1, y_1^d), \ldots, (x_n, y_n^d), x_{n+1}, \ldots, x_{n+m}\}$. where $\mathcal{S}^{(l)}$ and $\mathcal{S}^{(u)}$ denote the sets of labelled and unlabelled data, respectively. $y_i^d \in \{-1, 0, 1\}$ indicates whether $x_i \in \mathcal{S}^{(l)}$ belongs to the dth class, where $d \in D$. D is the total number of classes. Here, the -1, 0, and 1 represent Aggressive, Normal, and Mild driving styles, respectively.

With labelled and unlabelled data, the goal of our classification task is to train a K binary-class classifier: $f^d(x|w^d, b^d) = \langle w^d, x\rangle + b$, where $w^d \in \mathcal{R}^n$ is the desired hyperplane parameter vector for class d and b^d, is the bias term. To facilitate better performance, we apply a linear kernel and a nonlinear kernel function (i.e. Gaussian radial basis function kernel) separately for S3VMs. According to Gieseke et al. (2014), for class d, the optimal solution for f^d can be found with parameter vector $\alpha^d \in \mathcal{R}^{n+m}$,

$$f^d(x|\alpha^d) = \sum_{i=1}^{n} \alpha_i^d k(x_i, x) \tag{4}$$

2.3. Assessment of traffic flow performance

In micro-simulation, several indicators are employed to estimate traffic safety, fuel consumption and emissions. Specifically, traffic safety is assessed by Time to Collision (TTC) and

Time Exposed Modified Time to Collision (TEMTTC). Fuel consumption is evaluated by the Virginia Tech Comprehensive Power-Based Fuel Consumption Model (VT-CPFM) and Vehicle Specific Power (VSP), and emissions of heterogeneous traffic flow are Int Panis, Broekx, and Liu (2006).

2.3.1. Measures of traffic safety

Time-to-collision (TTC) is defined as the time when the speed of the objective vehicle is greater than its leading vehicle, the objective vehicle keeps the original driving state and does not take the corresponding deceleration behaviour until the two vehicles collide. It has been widely used in traffic safety evaluations. The formula is shown as follows:

$$\mathrm{TTC}_i(t) = \begin{cases} \dfrac{x_{i-1}(t) - x_i(t) - l_{i-1}}{v_i(t) - v_{i-1}(t)}, & \text{if } \forall v_i(t) > v_{i-1}(t) \\ \infty, & \text{if } \forall v_i(t) \le v_{i-1}(t) \end{cases} \tag{5}$$

where $x_i(t)$ and $v_i(t)$ are the location and speed of vehicle i at time t, respectively. $x_{i-1}(t)$ and $v_{i-1}(t)$ denote the location and speed for vehicle $i-1$; l_{i-1} represents the length of vehicle $i-1$.

A threshold value is selected to distinguish between safe and dangerous traffic conditions in the literature. According to Minderhoud and Bovy (2001), the time-exposed time-to-collision (TET) was proposed based on the TTC, which can be determined by Equation (6).

$$\mathrm{TET}^* = \sum_{i=2}^{N}\sum_{t=0}^{T} \delta_i(t)\cdot\Delta t, \quad \delta_i(t) = \begin{cases} 0, & \text{Otherwise} \\ 1, & 0 \le \mathrm{TTC}_i(t) \le \mathrm{TTC}^* \end{cases} \tag{6}$$

where T is the total observation time (denotes simulation time here) and Δt is the interval time. TTC^* represents the safety threshold whose value varies from 1 to 3 s (Li et al. 2014); $\delta_i(t)$ is a 0–1 variable. When $\mathrm{TTC}_i(t)$ is less than TTC^*, $\delta_i(t)$ is equal to 1; otherwise, it is 0. For $N(i = 2 \ldots N)$ drivers in the observation section, the total TET^* is calculated by Equation (6). The smaller the value of the TET^*, the higher level the of traffic safety. TTC^* is set as 2s in this paper.

The Modified time to collision (MTTC) is another metric that calculates the time required for the following vehicle to collide with a leading vehicle maintaining constant movement characteristics. Generally, an MTTC below 1.5 s is deemed unsafe. TEMTTC represents the accumulated time of unsafe MTTC experienced by each vehicle in the traffic flow (Yu et al. 2023). The calculation is expressed as follows:

$$\mathrm{MTTC}_i(t) = \frac{\Delta v_i(t) \pm \sqrt{\Delta v_i(t)^2 + 2\Delta a_i(t)\left[x_{i-1}(t) - x_i(t) - l_{i-1}\right]}}{\Delta a_i(t)} \tag{7}$$

$$\mathrm{TEMTTC} = \sum_{i=2}^{N}\sum_{t=0}^{T} \varsigma_i(t)\cdot\Delta t, \quad \varsigma_i(t) = \begin{cases} 0, & \text{Otherwise} \\ 1, & \mathrm{MTTC}_i(t) < 1.5\,\text{s} \end{cases} \tag{8}$$

where $\mathrm{MTTC}_i(t)$ is determined by (i) if both of MTTC terms are positive, the minimum of them is considered the final value of MTTC; and (ii) if one is positive while the other is negative, the positive one is considered the final value of MTTC; the Boolean variable $\varsigma_i(t)$ takes a value of 1 if the condition is unsafe and 0 otherwise.

2.3.2. Measures of traffic sustainability

The Virginia Tech Comprehensive Power-based Fuel Consumption Model (VT-CPFM) is a well-regarded model designed to estimate the fuel consumption of vehicles based on powertrain dynamics and vehicle-specific characteristics. It can be easily implemented in systems that require the use of a microscopic-level fuel consumption model (Hu et al. 2022), thus being employed to calculate fuel consumption in this study. The fuel consumption rate $F_{C,i}$ (g/s) is modelled in the form of a quadratic polynomial as:

$$F_{C,i} = \begin{cases} \theta_{0,i} + \theta_{1,i} P_{d,i} + \theta_{2,i} P_{d,i}^2, & \text{if } P_{d,i} \geq 0, \\ \theta_{0,i}, & \text{if } P_{d,i} < 0, \end{cases} \tag{9}$$

where $\theta_{(\cdot),i}$ are constant coefficients, and $\theta_{0,i} = 0.54, \theta_{1,i} = 0.06, \theta_{2,i} = 0.00017$; $P_{d,i}$ (kW) is the power output of the vehicle driveline and it calculated based on the vehicle velocity and acceleration:

$$P_{d,i} = \frac{\left(ma_i(t) + C_{A,i} v_i(t)^2 + mgf_{r,i}\right) v_i(t)}{\eta_{T,i}} \tag{10}$$

where m is the vehicle mass (1500 kg); $C_{A,i}$ is the coefficients of aerodynamic drag (0.4 kg/m); g is the gravitational acceleration (9.8 m/s^2); $f_{r,i}$ is the rolling resistance (0.015 kg/m); $\eta_{T,i}$ is the mechanical efficiency of the drivenline (0.8).

Vehicle specific power model (VSP) refers to the instantaneous power of a vehicle per unit mass and combines the driving characteristics of vehicles such as speed and acceleration, and road characteristics such as the road gradient (Song and Yu 2009). It has been widely used for fuel consumption modelling, because both the power for overcoming aerodynamic drag and rolling resistance and for the kinetic and potential energy of the vehicle are taken into account in VSP, by doing this the relationship between VSP and fuel consumption can be explained physically. The calculation of the VSP is shown in Equation (11).

$$\text{VSP} = 0.132 \cdot v + 1.1 \cdot v \cdot a + 0.0003202 \cdot v^3 \tag{11}$$

Vehicle Specific Power (VSP) is usually clustered into bins at certain intervals, and traffic emissions are estimated by average-speed-based VSP distribution within each bin (Frey, Rouphail, and Zhai 2006). The categorisation of VSP into 1KW/t intervals is detailed in Equation (12).

$$\text{VSP bin} = n, \forall : \text{VSP} \in [n - 0.5, n + 0.5) \tag{12}$$

where n is the VSP number.

Then, a nonlinear multivariate regression model is utilised to model instantaneous traffic emission by considering both average speed and other aspects of vehicle operation such as acceleration and deceleration (Int Panis, Broekx, and Liu 2006), which is expressed by:

$$E_i = \frac{\sum_{t=1}^{T} \max\left\{0, f_1 + f_2 v_i(t) + f_3 v_i(t)^2 + f_4 a_i(t) + f_5 a_i(t)^2 + f_6 v_i(t) a_i(t)\right\}}{T}. \tag{13}$$

$$E = \frac{\sum_{i=1}^{N} E_i}{N}. \tag{14}$$

where E_i is the traffic emission (g/s) of vehicle i; E is the average traffic emission of the entire traffic flow; the calibrated values of model parameters f_1, f_2, f_3, f_4, f_5, f_6 for petrol cars, as outlined in Int Panis, Broekx, and Liu (2006), are employed in this study. Specifically, for CO_2 emissions $f_1 = 5.54e^{-1}$, $f_2 = 1.61e^{-1}$, $f_3 = -2.89e^{-3}$, $f_4 = 2.66e^{-1}$, $f_5 = 5.11e^{-1}$, $f_6 = 1.83e^{-1}$; for NO_x emissions, model parameters associated with vehicle acceleration. when $a_i(t) \geq -0.5\,\mathrm{m/s}^2$, $f_1 = 6.19e^{-4}$, $f_2 = 8.00e^{-5}$, $f_3 = -4.03e^{-6}$, $f_4 = -4.13e^{-4}$, $f_5 = 3.80e^{-4}$, $f_6 = 1.77e^{-4}$; otherwise, $f_1 = 2.17e^{-4}$, $f_2 = f_3 = f_4 = f_5 = f_6 = 0$.

3. Experiments

In this section, classified-car-following models are first established according to driving style classification. Then the micro-simulation setup for estimating heterogeneous traffic performance is introduced. A preliminary experiment is conducted to justify the CCF models' capability to reproduce spatiotemporal traffic flow patterns before formal simulations start.

3.1. Classified car-following models establishment

Table 1 gives an overview of the bounds of the stochastic IDM model calibration. Following the methodology introduced in Section 2, the calibrated behavioural parameters are prepared for the subsequent driving style identification.

Feature selection is then conducted based on the calibrated car-following model parameters and the results of the score array are $[a_0 : -0.0831, b_0 : -0.0638, s_0 : -0.0573, T : -0.0768, v_0 : -0.1237]$. Considering that features with lower Laplacian scores are more important, v_0 emerged as the most important feature, followed by a_0 and others, indicating the greater importance of v_0 and a_0 in describing driving behaviour. Thus, we manually label driving styles for semi-supervised classification according to v_0 and a_0. Usually, drivers exhibiting higher desired speeds and greater acceleration are considered as a more aggressive driving style (W. Wang et al. 2017). This empirical knowledge enables the identification of drivers with distinct characteristics, forming the foundation for our redefining three driving styles: Aggressive, Normal, and Mild. Specifically, drivers with high designed speeds and accelerations (e.g. above the 75th percentile) were categorised as an Aggressive driving style. Conversely, drivers with lower values in these parameters (e.g. below the 25th percentile) were classified as a Mild driving style. Those with median values are labelled as a Normal driving style. Based on these criteria, a total of 295 driving samples were prelabelled, which comprise 99, 99, and 97 samples for Aggressive, Normal, and Mild styles, respectively.

A linear and a Gaussian radial basis function (RBF) kernel and obtain satisfactory results and reduce computational complexity when using SVMs. All the calibrated model parameters were divided into three disjoint parts: labelled dataset $S^{(l)} = \{x_i\}_{i=1}^{n}$, unlabelled dataset $S^{(u)} = \{x_j\}_{j=1}^{m}$, and test dataset $S^{(t)} = \{x_i\}_{i=1}^{r}$, where $n = 295$, $m = 2449$, $r \leq n$. Insufficient

Table 1. Summary of model parameters and their estimates.

Model	Parameter (unit)	Short description	Bounds	Mean	Median	Std
IDM	a_0 (m/s^2)	Max desired acceleration of follower	[0.1,5]	1.18	1.04	0.81
	v_0 (m/s)	Desired speed of follower	[10,40]	29.76	30.04	4.99
	s_0 (m)	Gap at standstill	[0.1,6]	2.22	2.17	0.97
	T (s)	Desired time headway of follower	[0.1,5]	1.72	1.67	0.94
	b_0 (m/s^2)	Comfortable deceleration of follower	[0.1,5]	1.45	1.33	0.92
	Q	Noise strength	[0.01,1]	0.50	0.51	0.28

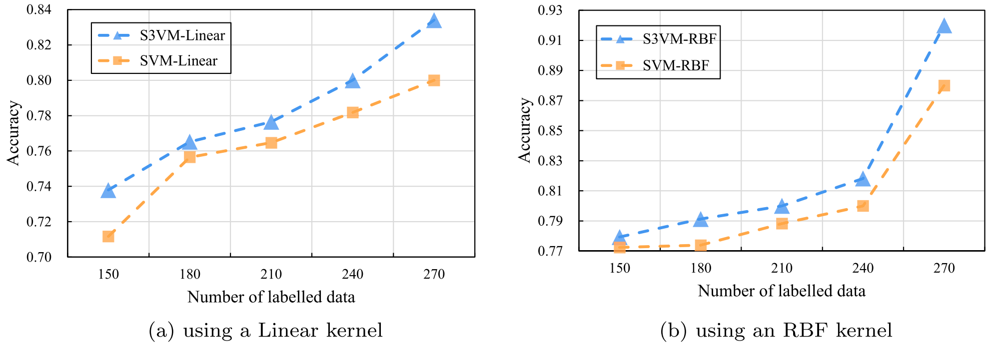


Figure 2. Classification accuracy of S3VM and SVM for multi-classification. (a) using a Linear kernel and (b) using an RBF kernel.

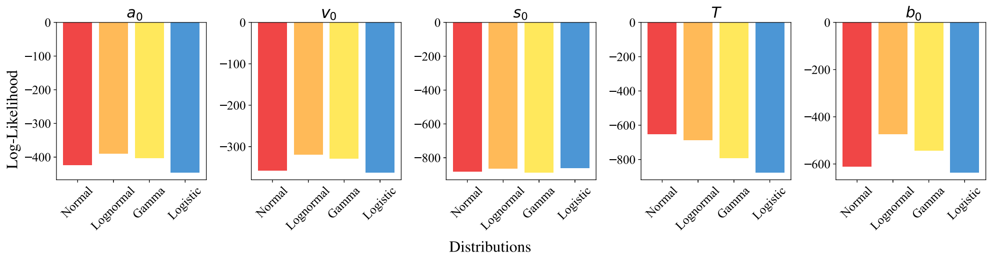


Figure 3. Comparison of fitted distributions of CF model parameters.

labelled data for training may not take full advantage of unlabelled data to properly reflect the reality of various driving styles. Hence, to further demonstrate and evaluate the S3VM's capability of exploiting the available data, comparative experiments were conducted using different amounts of labelled datasets $\mathcal{S}_P^{(l)}$, where $P \in \{150, 180, 210, 240, 270\}$. Classification results of SVM and S3VM are illustrated in Figure 2 where the vertical axis represents the multi-classification accuracy and the horizontal axis denotes the varied amount of labelled training data. In both Figure 2(a,b), the blue dotted lines with triangles consistently exceed the yellow dotted lines with squares, indicating S3VM generally outperforms SVM across both kernel types. The overall advantage of S3VM ranges from 1.74%-8.00% when employing a linear kernel, and from 1.82% to 8.60% with an RBF kernel. For instance, with 210 labelled data points, SVM achieved accuracies of 76.47% with a linear kernel and 78.82% with an RBF kernel. In contrast, S3VM attained accuracies of 77.65% and 80% with the respective kernels. Notably, the advantage of S3VM is more pronounced with a larger dataset. For example, with 150 labelled samples using an RBF kernel, the accuracy improvement of S3VM over SVM increases from 77.24% to 77.93%. This advantage becomes more significant with 270 labelled samples, which is with S3VM achieving 92% accuracy compared to 88% for SVM. Classification results with the highest accuracy are used for subsequent analyses.

The Classified Car-following (CCF) model has been proposed to personalise car-following models for heterogeneous drivers (Sun et al. 2021). We follow their methodology to establish CCF models for drivers in our micro-simulation. Four widely used parametric distributions, namely Normal, Lognormal, Gamma, and Logistic, are employed to determine the

Table 2. Classified IDM model parameters.

Driving styles	a_0	v_0	s_0	T	b_0	Q
Aggressive	1.93	33.55	1.91	1.35	1.14	0.373
Normal	1.04	29.45	2.02	1.48	1.04	0.370
Mild	0.44	27.21	2.00	1.36	1.08	0.376

probabilistic distributions of car-following model parameters for each driving style. The best-fitting distribution was chosen by comparing the goodness of fit of all parametric distributions. As shown in Figure 3, the Lognormal distribution outperforms the other distributions. Thus, the mean value (μ) of the Lognormal distribution is utilised to determine IDM model parameters for each driving style, as outlined in Table 2.

3.2. Micro-simulation setup

The micro-simulation is developed by MATLAB R2021b, which was executed on an Apple M1 Pro MacBook. The traffic volume is set as 1600 veh/h, and the vehicles are generated based on the negative exponential distribution $e^{-\lambda t}$. 30 vehicles were counted on the simulated road section.

The proportions of drivers with Aggressive, Mild, and Normal driving styles in the simulated traffic flow are denoted by p_a, p_m, and $p_n = 1 - p_a - p_m$, respectively. To thoroughly assess the effects of various driving styles and their proportional changes on traffic safety and sustainability, 66 fine-grained traffic scenarios are considered. The shares of each driving style (p_a, p_m and p_n) range from 0% to 100% in 10% increments. Acknowledging that drivers with identical driving styles may still exhibit individual differences, we introduce variability in the Car-Following (CF) model parameters in alignment with the values given in Table 2. To avoid extreme values resulting from generating parameters directly from distributions, we use Equation (15) to vary car-following parameters. By doing this, each driver has their personalised IDM model parameters meanwhile following a particular driving style. It is assumed that drivers maintain a consistent driving style throughout the simulation period (Mohammadnazar, Arvin, and Khattak 2021). Each traffic scenario is simulated with 100 seeds to enhance the reliability of the results. The final result for each traffic flow scenario is determined by calculating the mean of all repeated simulations under this setting.

$$PM_{ij}^{\text{sim}} = 0.9PM_{ij} + 0.1 \times \frac{1}{PM_{ij}\sigma_{ij}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(PM_{ij})-\mu_{ij}}{\sigma_{ij}}\right)^2} \tag{15}$$

where PM_{ij}^{sim} denotes the CCF model parameter used in simulation, $i = \{a_0, v_0, s_0, T, b_0, Q\}$, $j = \{\text{Aggressive, Normal, Mild}\}$. PM_{ij} follows the values shown in Table 2.

3.3. Preliminary simulation

Preliminary experiments were conducted to examine the traffic flow patterns generated by the CCF models. The spatiotemporal patterns of traffic flow with 20% aggressive and 80% normal drivers are shown in Figure 4. Stripes structures are observed, illustrating the

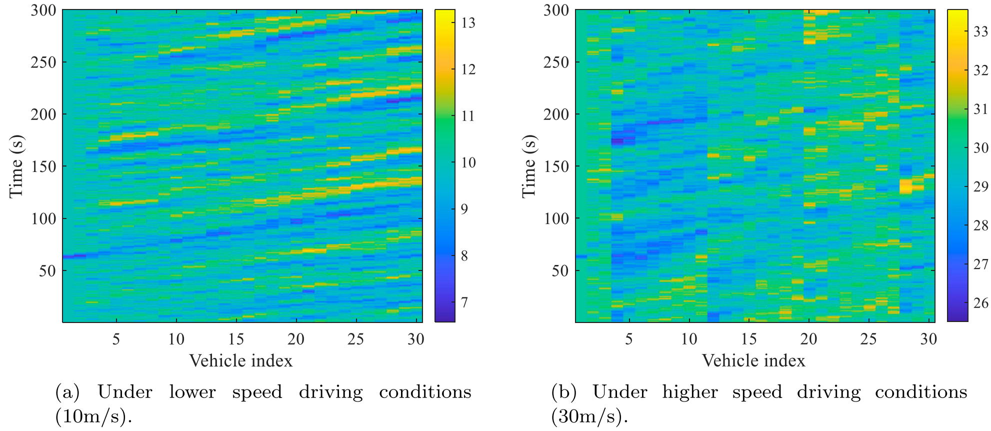


Figure 4. Spatiotemporal patterns of traffic flow under different speed settings. (a) Under lower speed driving conditions (10 m/s) and (b) Under higher speed driving conditions (30 m/s).

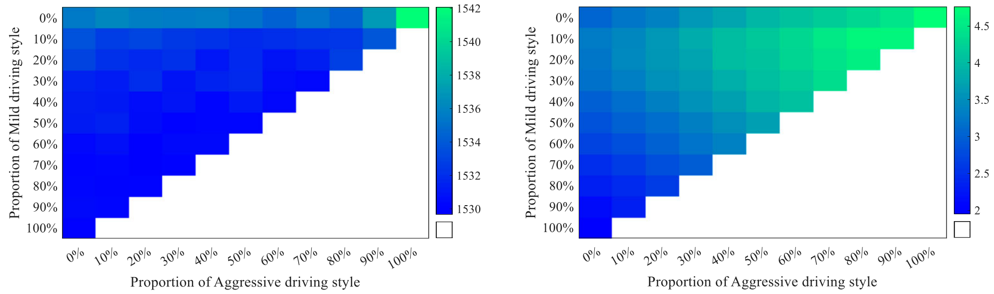
propagation, growth, dissipation, and merging of disturbances. In high-speed driving conditions (30 m/s) such as the HighD dataset, see Figure 4(b), fewer disturbances are observed compared to low-speed driving conditions such as the NGSIM dataset (see Figure 4(a)). This is due to the diminished influence of white noise on oscillation evolution on higher-speed traffic flow (Treiber and Kesting 2017). These results demonstrate that the stochastic car-following model can effectively replicate typical spatiotemporal traffic flow patterns.

4. Results and discussion

Based on the identification of car-following heterogeneity and experimental settings, micro-simulation results are presented in this section. The impacts of variations in driving styles on traffic safety, fuel consumption and emissions are analysed from the mechanism of underlying driving behaviours in heterogeneous traffic flow. And the proposed methodology is then verified by transferability analysis using a different dataset.

4.1. Impacts of CF heterogeneity

Statistics of the aforementioned traffic estimation indicators for 66 simulated traffic scenarios are presented in Figures 5–6. Indicators representing traffic safety, fuel consumption, and emissions are represented by blue, green and red colour sets, respectively. Within each colour set, a darker hue signifies a lower indicator value, such as fewer hazardous incidents or reduced emissions. Figure 5(a,b) present the results of TTC and TEMTTC, the scenario located in the upper right corner displays the lightest shade of blue, indicating that a traffic flow with 100% aggressive drivers displays a very low level of traffic safety among all heterogeneous traffic scenarios. The shift from light blue towards dark blue is observed from the top right to the bottom left, indicating that a decrease in the number of Aggressive drivers and an increase in the number of Mild drivers improves traffic safety. The safest traffic conditions are achieved with a composition of 100% Mild drivers. Results of fuel consumption and emissions across all simulated traffic scenarios are presented in Figure 6(a–d).



(a) Average TTC. (b) Average TEMTTC.

Figure 5. Statistics of traffic safety indicators in 66 traffic scenarios. (a) Average TTC and (b) Average TEMTTC.

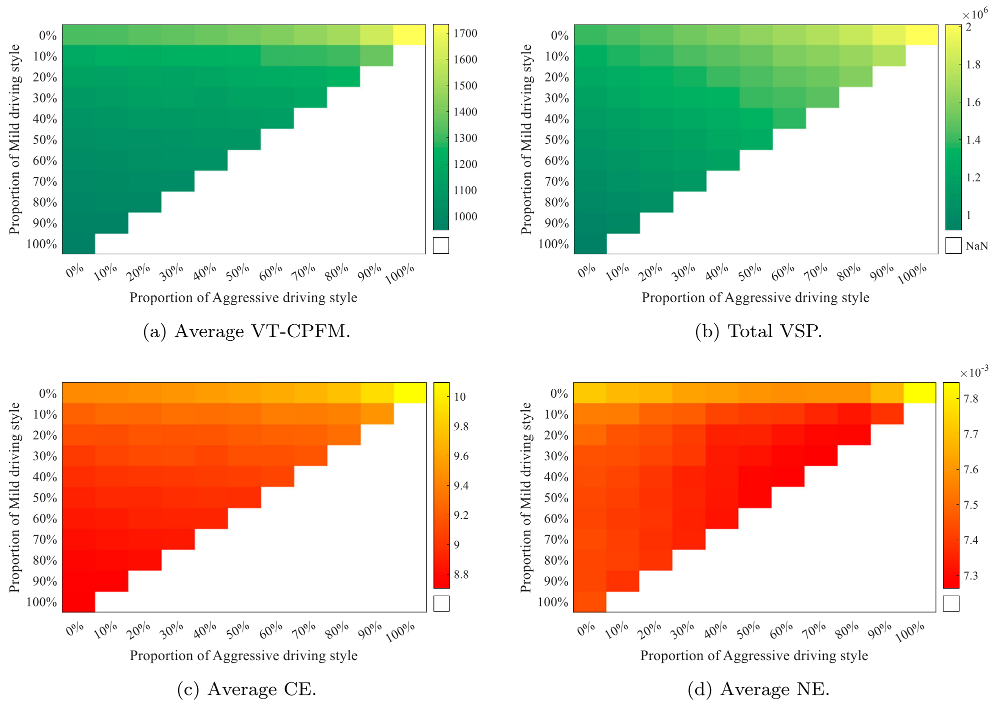


Figure 6. Statistics of traffic sustainability indicators in 66 traffic scenarios. (a) Average VT-CPFM. (b) Total VSP. (c) Average CE and (d) Average NE.

The colour gradient from the darkest in the bottom left corner to the lightest in the top right corner signifies a reduction in traffic sustainability. These results indicate that an increase in the share of Aggressive drivers and a decrease in Mild drivers in traffic flow leads to higher fuel consumption and emissions.

Overall, this analysis indicates a general trend that traffic safety and sustainability improve as the proportion of milder drivers rises and more aggressive drivers decrease, while the trend does not follow a strictly linear pattern. Take Figure 5(b) as an example, the traffic scenario with 80% Aggressive drivers and 20% Normal drivers is associated with fewer safety issues than one with 80% Aggressive drivers, 10% Normal drivers. and 10%

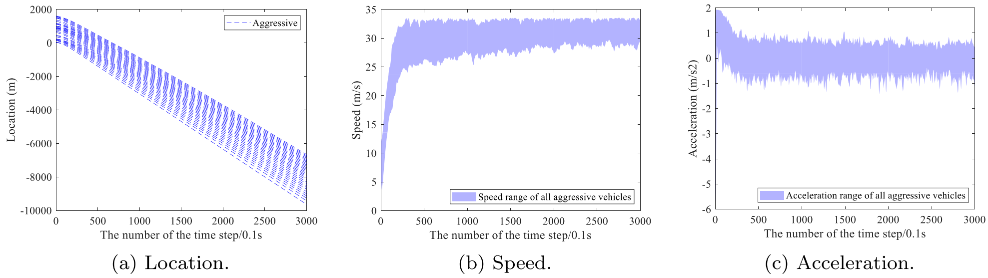


Figure 7. Vehicle trajectories of traffic flow with 100% Aggressive drivers. (a) Location. (b) Speed and (c) Acceleration.

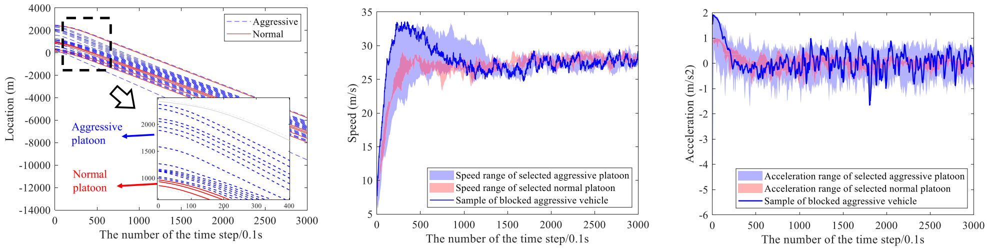
Mild drivers. This observation suggests the complexity of traffic dynamics and a need for a nuanced understanding of how various driving styles and their mixed proportions affect traffic safety and sustainability.

4.2. Trajectory-based analysis

To further understand the diversity of safety and sustainability levels caused by CF heterogeneity, we examined individual driving trajectories by diving into two-mixed (with two different driving styles) and three-mixed (with three different driving styles) heterogeneous traffic flows.

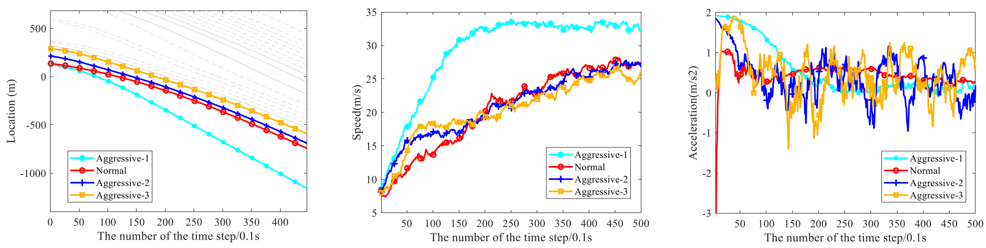
4.2.1. Analysis on two-mixed heterogeneous flow

Figures 7–13 shows the vehicle trajectories when different proportions of Normal drivers are introduced into an Aggressive style traffic flow. Trajectories of Normal and Aggressive drivers are represented by solid red lines and blue dotted lines, respectively. In Figure 7(a), despite the stochastic nature of the vehicles, they maintain stable and uniform spacing due to their consistent driving style within homogeneous traffic flow, leading to a smoother traffic flow pattern with little disruption. Similar observations are demonstrated in speed and acceleration diagrams shown in Figure 7(b,c). The introduction of Normal drivers disrupts this uniformity and leads to the formation of platoons, and these platoons are formed in traffic flow with Normal-style vehicles as the leader. In a traffic flow with 80% aggressive drivers shown in Figure 8(a), several platoons are observed which are led by the Normal vehicles, as red lines show. Figure 8(b) shows speed diagrams where light red and light blue represent the vehicle platoons of Normal and Aggressive, respectively. The dark blue line represents the lead vehicle in the Aggressive platoon, which is the first follower of a Normal vehicle. Notice that the speed of the Aggressive follower decreases over time, eventually aligning with the speed of the Normal platoon. This occurs because the Normal leader inhibits the Aggressive follower from maintaining a higher speed. Consequently, other Aggressive vehicles in the platoon, led by this impeded Aggressive vehicle, also reduce their speeds, resulting in the entire platoon adopting a Normal driving style. A similar trend is evident in Figure 8(c) where all platoons ultimately exhibit similar acceleration, despite Aggressive vehicles displaying significant fluctuations in acceleration. This variability of accelerations potentially leads to increased fuel emissions and higher traffic risks, elucidating the lower safety and sustainability caused by aggressive drivers.



(a) Location. (b) Speed. (c) Acceleration.

Figure 8. Vehicle trajectories of traffic flow with 80% Aggressive drivers and 20% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

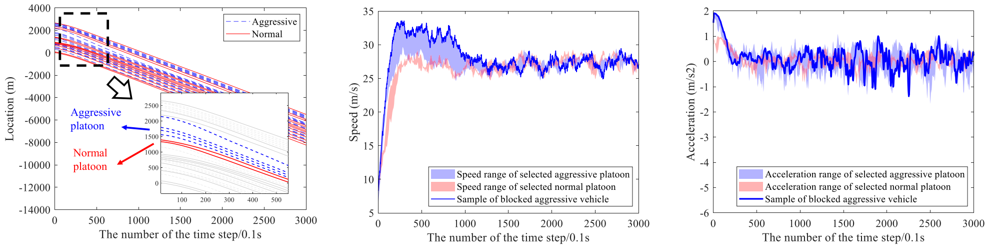


(a) Location. (b) Speed (c) Acceleration.

Figure 9. Quantitative analysis of vehicle trajectories (80% Aggressive and 20% Normal drivers). (a) Location. (b) Speed and (c) Acceleration.

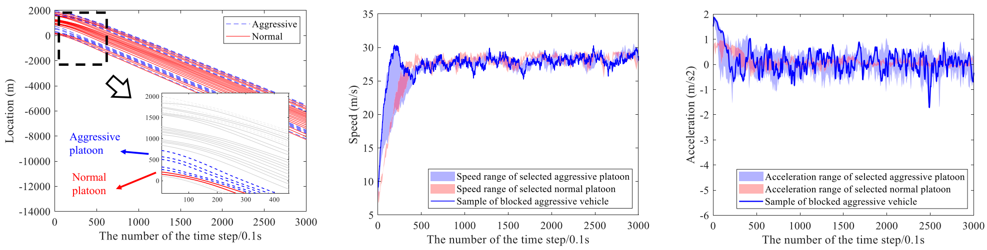
For more detailed analysis, trajectories of four exemplified vehicles in a traffic scenario with 80% Aggressive drivers and 20% Normal drivers are visualised in Figure 9. The red line with circles denotes a Normal vehicle, and lines with other colours and signs represent Aggressive vehicles. Initially, three vehicles with Aggressive styles exhibited maximum accelerations of $1.92\,\mathrm{m/s^2}$, $1.81\,\mathrm{m/s^2}$, and $1.85\,\mathrm{m/s^2}$, respectively, as the cyan line with asterisks (Aggressive-1), the blue line with plus (Aggressive-2), and the orange line with squares (Aggressive-3), as depicted in Figure 9(c). The Normal vehicle, denoted by the red line with circles, has the maximum acceleration of $0.98\,\mathrm{m/s^2}$. Based on originally preset accelerations, speeds of all vehicles increase as the simulation progresses, see Figure 9(b). During the initial 0–10 s, the speeds of Aggressive vehicles surpass that of the Normal vehicle. After 10 s, the speeds of Aggressive-1 continue to increase, stabilising at 33.6 m/s, whereas Aggressive-2 and Aggressive-3 flatten out and show a similar increase trend to that of Normal vehicle, reaching speeds of 27.6 m/s and 27.1 m/s, respectively. To this end, Aggressive-1 can behave as a real Aggressive driving style and distance itself from the following Normal vehicle. In contrast, Aggressive-2 and Aggressive-3 exhibit a Normal driving style due to the blocking of their Normal leader. Here, vehicles that are original with an Aggressive and Normal style are denoted as A_A and N_N, respectively, while Aggressive vehicles that exhibit as Normal due to the block of their leaders are represented as A_N.

In traffic scenarios with a decreasing proportion of Aggressive drivers and an increasing share of Normal drivers, the platoon formation dynamically shifts. In Figures 11(a)–12(a) where there are more Normal drivers than Aggressive drivers, these Aggressive vehicles appear at the end of each platoon. This occurs as Normal vehicles, with their lower



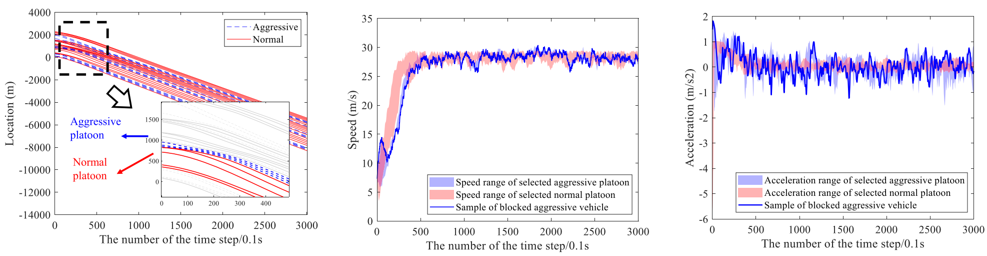
(a) Location. (b) Speed. (c) Acceleration.

Figure 10. Vehicle trajectories of traffic flow with 60% Aggressive drivers and 40% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.



(a) Location. (b) Speed. (c) Acceleration.

Figure 11. Vehicle trajectories of traffic flow with 40% Aggressive drivers and 60% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

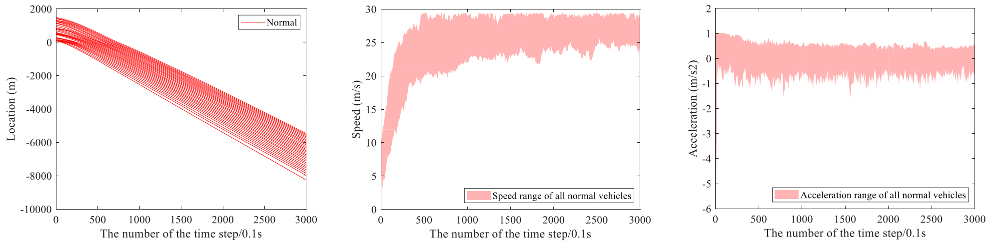


(a) Location. (b) Speed. (c) Acceleration.

Figure 12. Vehicle trajectories of traffic flow with 20% Aggressive drivers and 80% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

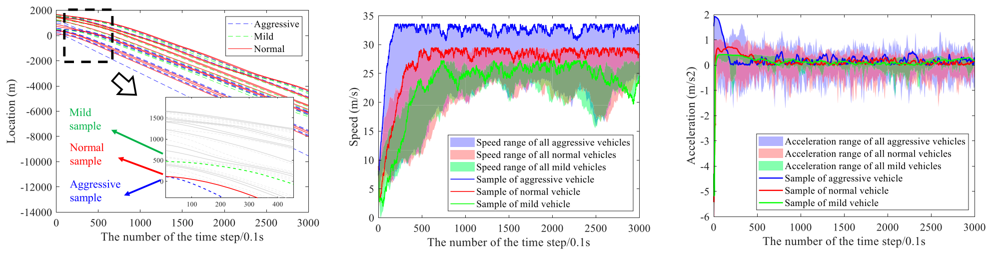
acceleration and speed, naturally fall behind their Aggressive leaders and eventually form platoons leading by themselves. Since these traffic flows have a high proportion of Normal drivers, most vehicles in platoons are with an N_N style rather than A_N. Moreover, a homogeneous traffic flow with 100% Normal drivers demonstrates a consistent and smooth pattern, mirroring the uniform behaviour seen in a pure Aggressive traffic flow, see Figure 13. Similar findings are observed through analyses of two-mixed traffic flow scenarios comprising Mild and Normal driving styles, which are presented in Appendix.

In summary, in a two-mixed traffic flow scenario, vehicles with lower aggressiveness (Mild in comparison to Normal, or Normal in contrast to Aggressive) tend to lead the formation of vehicular platoons. This occurs as they naturally fall behind their more aggressive leaders. Meanwhile, vehicles with higher aggressiveness are observed to adopt a driving



(a) Location. (b) Speed. (c) Acceleration.

Figure 13. Vehicle trajectories of traffic flow with 100% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.



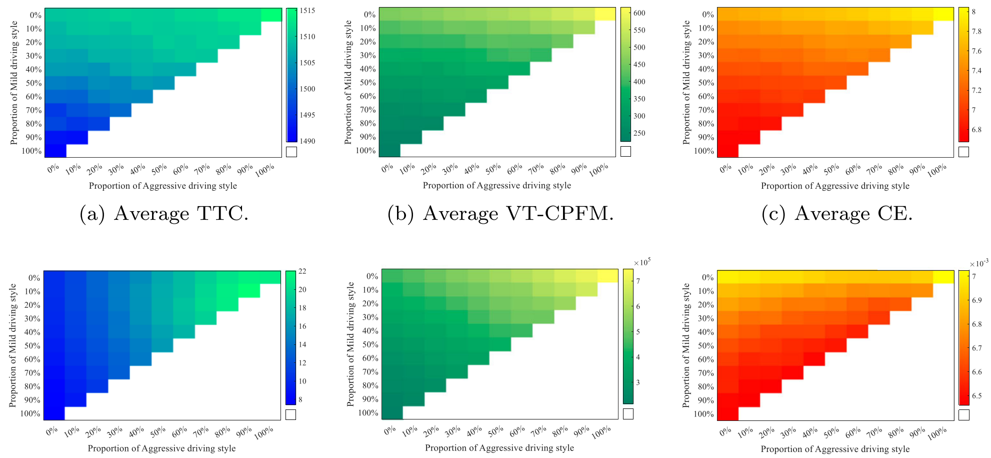
(a) Location. (b) Speed. (c) Acceleration.

Figure 14. Traffic flow with 30% Aggressive, 30% Mild drivers, and 40% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

style associated with lower aggressiveness due to being impeded by their less aggressive leaders. Importantly, the formation of these vehicular platoons is influenced not just by the proportion of less aggressive vehicles within the traffic, but also by their spatial positioning within the traffic flow. These findings explain observations that the negative effects of an aggressive driving style – or conversely, the beneficial impacts of a mild driving style – on traffic safety and sustainability do not escalate strictly linearly with their proportion changes.

4.2.2. Analysis on three-mixed traffic flow

In traffic flows comprising three-mixed driving styles, the dynamics of driving behaviour become increasingly complex. Figure 14 shows a heterogeneous traffic flow with 30% Aggressive, 30% Mild drivers, and 40% Normal drivers, in which Normal drivers are represented by red solid lines, and blue and green dash lines denote Aggressive drivers and Mild drivers, respectively. The Normal vehicle falls behind its Aggressive leader and creates a large spacing with its Mild follower. The three vehicles shown as blue, red, and green lines in Figure 14(b) exhibit their original driving styles (Aggressive, Normal, and Mild), largely because they are not impeded by vehicles of lesser aggressiveness ahead of them, enabling them to achieve the desired speed typical of their respective driving styles. This is similarly reflected in their acceleration profiles, as seen in Figure 14(c), where fluctuations remain minimal due to the lack of hindrance. Beyond these three initial vehicles, drivers with less aggressive styles tend to obstruct those who are more aggressive, compelling them to adopt similar, less aggressive behaviours in terms of acceleration and



(a) Average TTC. (b) Average VT-CPFM. (c) Average CE.

(d) Average TEMTTC. (e) Total VSP. (f) Average NE.

Figure 15. Statistics of traffic safety and sustainability indicators in 66 traffic scenarios (using NGSIM dataset). (a) Average TTC. (b) Average VT-CPFM. (c) Average CE. (d) Average TEMTTC. (e) Total VSP and (f) Average NE.

Table 3. Impact of heterogeneous driving style on traffic flow with the NGSIM dataset.

Proportions	TTC	TEMTTC	CE	NE	VSP	VT-CPFM
0%,100%,0%	1.51×10^3	10.30	7.57	6.99×10^{-3}	4.49×10^5	4.56×10^2
100%,0%,0%	0.39%	102.99%	6.23%	0.51%	66.37%	34.80%
0%,0%,100%	-1.30%	-27.88%	-11.86%	-7.40%	-52.02%	-50.72%
60%,20%,20%	0.01%	79.46%	-1.73%	-4.51%	17.41%	-7.96%
20%,60%,20%	-0.08%	24.77%	-2.99%	-3.26%	-4.22%	-13.67%
20%,20%,60%	-0.47%	14.38%	-7.75%	-6.17%	-26.42%	-34.37%
30%,30%,40%	-0.13%	35.67%	-4.23%	-4.56%	-5.99%	-19.30%

speed. This results in the formation of vehicle platoons that are predominantly led by drivers of lower aggressiveness, either Mild or Normal. These platoons have similar acceleration but with different fluctuations, as shown in Figure 14(c). The original Aggressive platoons exhibit larger acceleration variations, which aligns with the observations of two-mixed traffic flow.

4.3. *Transferability analysis*

To validate the proposed approach and corresponding findings, we conducted a transferability analysis using the NGSIM-I80 dataset. The simulation outcomes are presented in Figure 15. The colour shifting in all indicators of traffic safety and sustainability suggests that an increase in the proportion of Aggressive drivers correlates with a reduction in traffic safety, as well as an increase in fuel consumption and emissions. Conversely, the presence of Mild drivers results in a diminishing effect on both traffic safety and environmental issues. Such findings are aligned with the outcomes obtained from the HighD dataset.

Several representative traffic scenarios with different driving style proportions are further examined, with results shown in Tables 3–4. The three proportions represent

Table 4. Impact of heterogeneous driving style on traffic flow with the HighD dataset.

Proportions	TTC	TEMTTC	CE	NE	VSP	VT-CPFM
0%,100%,0%	1.53×10^3	3.05	9.45	7.72×10^{-3}	1.40×10^6	1.32×10^3
100%,0%,0%	0.42%	56.12%	6.81%	1.64%	43.17%	31.35%
0%,0%,100%	-0.37%	-36.04%	-7.89%	-3.55%	-34.56%	-28.23%
60%,20%,20%	-0.30%	40.57%	-2.34%	-5.13%	5.84%	-8.52%
20%,60%,20%	-0.24%	17.96%	-2.94%	-3.59%	-4.77%	-11.31%
20%,20%,60%	-0.38%	-0.65%	-5.59%	-4.25%	-18.43%	-20.58%
30%,40%,30%	-0.30%	19.86%	-3.38%	-4.13%	-5.38%	-12.74%

Aggressive, Normal and Mild driving styles in a certain traffic flow scenario. The traffic scenario with 100% Normal drivers serves as a baseline, with its indicators presented in absolute values. Relative changes in traffic safety, fuel consumption, and emissions for other traffic scenarios are calculated against this baseline, with positive and negative changes indicated in shades of red and blue, respectively; and darker shades signify larger magnitude changes. For example, the traffic scenario with 100% Aggressive drivers showed a notable increase in fuel consumption – specifically, a 31.35% rise according to VT-CPFM compared to the 100% Normal driver traffic scenario. Conversely, scenarios involving 100% Mild drivers demonstrated notable improvements in reducing safety issues, fuel consumption, and emissions. Even though in a three-mixed traffic flow, such as with proportions of 20% Aggressive, 20% Normal, and 60% Mild driving styles, scenarios incorporating Mild drivers consistently led to decreased safety risks, fuel consumption, and emissions when compared to a purely Normal driving scenario, despite the inclusion of Aggressive drivers.

Comparing the results in Tables 3–4 reveals that changes in the proportions of driving styles have a more pronounced effect on traffic performance in the NGSIM dataset. For example, a 100% Mild traffic flow in the NGSIM dataset improved emission by 11.86% and reduced fuel consumption by 52.02%, whereas in the HighD dataset, the improvements were only 7.89% and 34.56% HighD dataset, respectively. Similar findings can be observed from other driving style combinations such as 20%, 20%, and 60%. This discrepancy is attributed to the NGSIM dataset being collected under relatively congested traffic conditions, which increase the observable variations in driving behaviours compared to the HighD dataset.

5. Concluding remarks

This paper proposes a general framework to investigate car-following heterogeneity and its impacts on traffic safety, fuel consumption and emissions. The framework incorporates a rigorous driving style classification using a multi-class S3VM classifier and a micro-simulation process with 66 fine-grained heterogeneous traffic scenarios. The impacts of car-following heterogeneity on traffic flow performance are further elucidated from the mechanism of underlying characteristics of driving behaviour. The key findings of this research are summarised below:

(i) S3VM vs. SVM performance: Driving styles classification reveals that S3VM-based classifiers notably outperform traditional SVM classifiers in driving style

classification, with accuracy improvements up to 8.60%. This enhancement is particularly significant when utilising an RBF kernel over a Linear kernel.
(ii) Aggressiveness impacts: Less aggressive drivers can lead to the formation of vehicular platoons, thereby encouraging more aggressive drivers to adopt a milder driving style. Importantly, the formation of these platoons is influenced by both the proportion and spatial distribution of less aggressive vehicles, which makes the correlation between less aggressive driving and improvements in safety, fuel consumption, and emissions more complex.
(iii) Diversity under traffic conditions: The impact of driving style diversity on traffic performance is more pronounced in congested traffic conditions.

This study promises potential benefits for Intelligent Transportation Systems (ITS) by improving traffic safety and sustainability. The limitation of this study is that only inter-driving heterogeneity is considered, where each driver maintains a consistent driving style throughout the simulation. Future studies can improve the simulation by considering both inter- and intra-driving style heterogeneity. For instance, identifying driving heterogeneity through the underlying mechanisms of driving behaviour using primitive driving patterns (W. Wang et al. 2017) or *Action patterns* (Yao, Calvert, and Hoogendoorn 2024), thereby allowing for varied driving styles over time in micro-simulations. Additionally, external factors such as driving environment and internal factors at driver psychology level are promisingly considered to improve driving style classification.

Acknowledgments

The authors confirm their contribution to the paper as follows: study conception and design: Xue Yao, Zhanbo Sun; data collection: Xue Yao, Qiruo Yan; analysis and interpretation of results: Xue Yao, Qiruo Yan; draft manuscript preparation: Xue Yao, Qiruo Yan, Simeon C. Calvert, Serge P. Hoogendoorn, and Zhanbo Sun. Supervision: Zhanbo Sun, Serge P. Hoogendoorn, Simeon, C. Calvert. All authors reviewed the results and approved the final version of the manuscript.

Funding

The work was performed with funding from National Natural Science Foundation of China via grant 52072316 and 52302418, the Fundamental Research Funds for the Central Universities via grant 2682023CX047 and supported by Delft University of Technology.

Disclosure statement

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

References

Adamidis, Filippos K., Eleni G. Mantouka, and Eleni I. Vlahogianni. 2020. "Effects of Controlling Aggressive Driving Behavior on Network-Wide Traffic Flow and Emissions." *International Journal of Transportation Science and Technology* 9 (3): 263–276. https://doi.org/10.1016/j.ijtst.2020.05.003.

Aghabayk, Kayvan, Majid Sarvi, and William Young. 2016. "Including Heavy Vehicles in a Car-Following Model: Modelling, Calibrating and Validating." *Journal of Advanced Transportation* 50 (7): 1432–1446. https://doi.org/10.1002/atr.v50.7.

Bakhit, Peter, Dalia Said, and Laila Radwan. 2015. "Impact of Acceleration Aggressiveness on Fuel Consumption Using Comprehensive Power Based Fuel Consumption Model." *Civ Environ Res* 7 (3): 148–156.

Bennett, Kristin, and Ayhan Demiriz. 1998. "Semi-Supervised Support Vector Machines." *Advances in Neural Information Processing Systems* 11: 368–373.

Chen, Danjue, Soyoung Ahn, Jorge Laval, and Zuduo Zheng. 2014. "On the Periodicity of Traffic Oscillations and Capacity Drop: The Role of Driver Characteristics." *Transportation Research Part B: Methodological* 59:117–136. https://doi.org/10.1016/j.trb.2013.11.005.

Chen, Yao, Ke Wang, and Jian John Lu. 2023. "Feature Selection for Driving Style and Skill Clustering Using Naturalistic Driving Data and Driving Behavior Questionnaire." *Accident Analysis & Prevention*185:107022. https://doi.org/10.1016/j.aap.2023.107022.

Fadhloun, Karim, and Hesham Rakha. 2020. "A Novel Vehicle Dynamics and Human Behavior Car-Following Model: Model Development and Preliminary Testing." *International Journal of Transportation Science and Technology* 9 (1): 14–28. https://doi.org/10.1016/j.ijtst.2019.05.004.

Fontaras, Georgios, Nikiforos-Georgios Zacharof, and Biagio Ciuffo. 2017. "Fuel Consumption and CO2 Emissions From Passenger Cars in Europe–Laboratory Versus Real-World Emissions." *Progress in Energy and Combustion Science* 60:97–131. https://doi.org/10.1016/j.pecs.2016.12.004.

Frey, H. C.hristopher, Nagui M Rouphail, and Haibo Zhai. 2006. "Speed-and Facility-Specific Emission Estimates for on-Road Light-Duty Vehicles on the Basis of Real-World Speed Profiles." *Transportation Research Record* 1987 (1): 128–137. https://doi.org/10.1177/0361198106198700114.

Gieseke, Fabian, Antti Airola, Tapio Pahikkala, and Oliver Kramer. 2014. "Fast and Simple Gradient-Based Optimization for Semi-Supervised Support Vector Machines." *Neurocomputing*123:23–32. https://doi.org/10.1016/j.neucom.2012.12.056.

Haque, Fatima, and Mohd Azman Abas. 2018. "Review of Driving Behavior Towards Fuel Consumption and Road Safety." *Jurnal Mekanikal* 41 (6): 59–68.

He, Xiaofei, Deng Cai, and Partha Niyogi. 2005. "Laplacian Score for Feature Selection." *Advances in Neural Information Processing Systems* 18: 1–8.

Hoogendoorn, Serge P., and Raymond Hoogendoorn. 2010. "Generic Calibration Framework for Joint Estimation of Car-Following Models by Using Microscopic Data." *Transportation Research Record* 2188 (1): 37–45. https://doi.org/10.3141/2188-05.

Hu, Manjiang, Chongkang Li, Yougang Bian, Hui Zhang, Zhaobo Qin, and Biao Xu. 2022. "Fuel Economy-Oriented Vehicle Platoon Control Using Economic Model Predictive Control." *IEEE Transactions on Intelligent Transportation Systems* 23 (11): 20836–20849. https://doi.org/10.1109/TITS.2022.3183090.

Huang, Yong-Xian, Rui Jiang, H. M. Zhang, Mao-Bin Hu, Jun-Fang Tian, Bin Jia, and Zi-You Gao. 2018. "Experimental Study and Modeling of Car-Following Behavior Under High Speed Situation." *Transportation Research Part C: Emerging Technologies* 97:194–215. https://doi.org/10.1016/j.trc.2018.10.022.

Int Panis, Luc, Steven Broekx, and Ronghui Liu. 2006. "Modelling Instantaneous Traffic Emission and the Influence of Traffic Speed Limits." *Science of The Total Environment* 371 (1–3): 270–285. https://doi.org/10.1016/j.scitotenv.2006.08.017.

Kesting, Arne, Martin Treiber, and Dirk Helbing. 2010. "Enhanced Intelligent Driver Model to Access the Impact of Driving Strategies on Traffic Capacity." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (1928): 4585–4605. https://doi.org/10.1098/rsta.2010.0084.

Krajewski, Robert, Julian Bock, Laurent Kloeker, and Lutz Eckstein. 2018. "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems." 2118–2125.

Lárusdóttir, Elísabet Björney, and Gudmundur F. Ulfarsson. 2015. "Effect of Driving Behavior and Vehicle Characteristics on Energy Consumption of Road Vehicles Running on Alternative Energy Sources." *International Journal of Sustainable Transportation* 9 (8): 592–601. https://doi.org/10.1080/15568318.2013.843737.

Li, Zhibin, Seongchae Ahn, Koohong Chung, David R. Ragland, Wei Wang, and Jeong Whon Yu. 2014. “Surrogate Safety Measure for Evaluating Rear-End Collision Risk Related to Kinematic Waves Near Freeway Recurrent Bottlenecks.” *Accident Analysis & Prevention* 64:52–61. https://doi.org/10.1016/j.aap.2013.11.003.

Makridis, Michail A., Aikaterini Anesiadou, Konstantinos Mattas, Georgios Fontaras, and Biagio Ciuffo. 2023. “Characterising Driver Heterogeneity Within Stochastic Traffic Simulation.” *Transportmetrica B: Transport Dynamics* 11 (1): 725–743.

Makridis, Michail, Ludovic Leclercq, Biagio Ciuffo, Georgios Fontaras, and Konstantinos Mattas. 2020. “Formalizing the Heterogeneity of the Vehicle-Driver System to Reproduce Traffic Oscillations.” *Transportation Research Part C: Emerging Technologies* 120:102803. https://doi.org/10.1016/j.trc.2020.102803.

Minderhoud, Michiel M., and Piet H. L. Bovy. 2001. “Extended Time-to-collision Measures for Road Traffic Safety Assessment.” *Accident Analysis & Prevention* 33 (1): 89–97. https://doi.org/10.1016/S0001-4575(00)00019-1.

Mohammadnazar, Amin, Ramin Arvin, and Asad J. Khattak. 2021. “Classifying Travelers’ Driving Style Using Basic Safety Messages Generated by Connected Vehicles: Application of Unsupervised Machine Learning.” *Transportation Research Part C: Emerging Technologies* 122:102917. https://doi.org/10.1016/j.trc.2020.102917.

Ngoduy, Dong. 2013. “Analytical Studies on the Instabilities of Heterogeneous Intelligent Traffic Flow.” *Communications in Nonlinear Science and Numerical Simulation* 18 (10): 2699–2706. https://doi.org/10.1016/j.cnsns.2013.02.018.

Ngoduy, Dong. 2015. “Linear Stability of a Generalized Multi-Anticipative Car Following Model with Time Delays.” *Communications in Nonlinear Science and Numerical Simulation* 22 (1–3): 420–426. https://doi.org/10.1016/j.cnsns.2014.08.019.

Ossen, Saskia, and Serge P. Hoogendoorn. 2011. “Heterogeneity in Car-Following Behavior: Theory and Empirics.” *Transportation Research Part C: Emerging Technologies* 19 (2): 182–195. https://doi.org/10.1016/j.trc.2010.05.006.

Shang, Mingfeng, and Raphael Stern. 2020. “Calibrating Heterogeneous Car-Following Models for Human Drivers in Oscillatory Traffic Conditions.” In *2020 Forum on Integrated and Sustainable Transportation Systems (FISTS)*, 101–106. IEEE.

Shang, Mingfeng, and Raphael E. Stern. 2021. “Impacts of Commercially Available Adaptive Cruise Control Vehicles on Highway Stability and Throughput.” *Transportation Research Part C: Emerging Technologies* 122:102897. https://doi.org/10.1016/j.trc.2020.102897.

Song, Guohua, and Lei Yu. 2009. “Estimation of Fuel Efficiency of Road Traffic by Characterization of Vehicle-Specific Power and Speed Based on Floating Car Data.” *Transportation Research Record* 2139 (1): 11–20. https://doi.org/10.3141/2139-02.

Suarez, Jaime, Michail Makridis, Aikaterini Anesiadou, Dimitrios Komnos, Biagio Ciuffo, and Georgios Fontaras. 2022. “Benchmarking the Driver Acceleration Impact on Vehicle Energy Consumption and CO2 Emissions.” *Transportation Research Part D: Transport and Environment* 107:103282. https://doi.org/10.1016/j.trd.2022.103282.

Sun, Zhanbo, and Xuegang Ban. 2018. “Identifying Multiclass Vehicles Using Global Positioning System Data.” *Journal of Intelligent Transportation Systems* 22 (1): 1–9. https://doi.org/10.1080/15472450.2017.1358623.

Sun, Zhanbo, Qiruo Yan, Yafei Liu, Zhijian Fu, and Lei Yang. 2024a. “Fundamental Diagram and Stability Analysis of Mixed Traffic Considering Heterogeneous Car-Following Behaviors and Platoon Factors.” *Intelligent Transportation Infrastructure* 3:liae010. https://doi.org/10.1093/iti/liae010.

Sun, Zhanbo, Qiruo Yan, Yafei Liu, Zhijian Fu, and Lei Yang. 2024b. “Fundamental Diagram and Stability Analysis of Mixed Traffic Considering Heterogeneous Car-Following Behaviors and Platoon Factors.” *Intelligent Transportation Infrastructure* 3 (July): liae010. https://doi.org/10.1093/iti/liae010.

Sun, Zhanbo, Xue Yao, Ziye Qin, Peitong Zhang, and Ze Yang. 2021. “Modeling Car-Following Heterogeneities by Considering Leader–follower Compositions and Driving Style Differences.” *Transportation Research Record* 2675 (11): 851–864. https://doi.org/10.1177/03611981211020006.

Szumska, Emilia M., and Rafał Jurecki. 2020. "The Effect of Aggressive Driving on Vehicle Parameters." *Energies* 13 (24): 6675. https://doi.org/10.3390/en13246675.

Treiber, Martin, Ansgar Hennecke, and Dirk Helbing. 2000a. "Congested Traffic States in Empirical Observations and Microscopic Simulations." *Physical Review E* 62 (2): 1805–1824. https://doi.org/10.1103/PhysRevE.62.1805.

Treiber, Martin, Ansgar Hennecke, and Dirk Helbing. 2000b. "Microscopic Simulation of Congested Traffic." In *Traffic and Granular Flow'99: Social, Traffic, and Granular Dynamics*, 365–376. Springer.

Treiber, Martin, and Arne Kesting. 2013. "Car-Following Models Based on Driving Strategies." *Traffic Flow Dynamics: Data, Models and Simulation* 6: 181–204. https://doi.org/10.1007/978-3-642-32460-4.

Treiber, Martin, and Arne Kesting. 2017. "The Intelligent Driver Model with Stochasticity-New Insights Into Traffic Flow Oscillations." *Transportation Research Procedia* 23:174–187. https://doi.org/10.1016/j.trpro.2017.05.011.

Wang, Jinghui, Hesham A. Rakha, and Karim Fadhloun. 2017. "Validation of the Rakha-Pasumarthy-Adjerid Car-Following Model for Vehicle Fuel Consumption and Emission Estimation Applications." *Transportation Research Part D: Transport and Environment* 55:246–261. https://doi.org/10.1016/j.trd.2017.06.030.

Wang, Wenshuo, Junqiang Xi, Alexandre Chong, and Lin Li. 2017. "Driving Style Classification Using a Semisupervised Support Vector Machine." *IEEE Transactions on Human-Machine Systems* 47 (5): 650–660. https://doi.org/10.1109/THMS.2017.2736948.

Xue, Qingwen, Ke Wang, Jian John Lu, and Yujie Liu. 2019. "Rapid Driving Style Recognition in Car-Following Using Machine Learning and Vehicle Trajectory Data." *Journal of Advanced Transportation* 2019 (1): 9085238.

Yao, Xue, Simeon C. Calvert, and Serge P. Hoogendoorn. 2023. "Identification of Driving Heterogeneity Using Action-Chains." In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 6001–6006. IEEE.

Yao, Xue, Simeon C. Calvert, and Serge P. Hoogendoorn. 2024. "Driving Pattern Interpretation Based on Action Phases Clustering." arXiv preprint arXiv: 2407.17518.

Yao, Xue, Zhaocheng Du, Zhanbo Sun, Simeon C. Calvert, and Ang Ji. 2024. "Cooperative Lane-Changing in Mixed Traffic: A Deep Reinforcement Learning Approach." *Transportmetrica A: Transport Science*, (4): 1–23. https://doi.org/10.1080/23249935.2024.2343048.

Yu, Weijie, Xuedong Hua, Dong Ngoduy, and Wei Wang. 2023. "On the Assessment of the Dynamic Platoon and Information Flow Topology on Mixed Traffic Flow Under Connected Environment." *Transportation Research Part C: Emerging Technologies* 154:104265. https://doi.org/10.1016/j.trc.2023.104265.

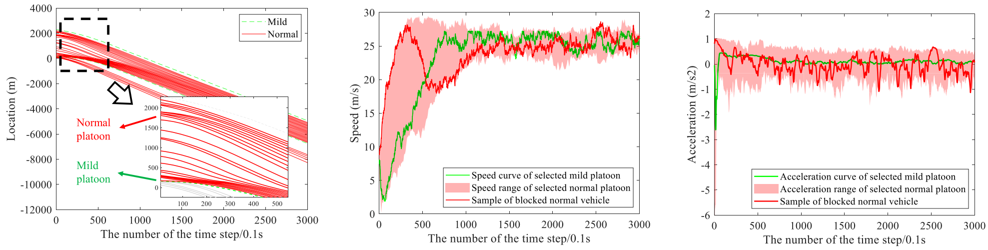
Zhang, Yanlin, and Alireza Talebpour. 2024. "Characterizing Human–automated Vehicle Interactions: An Investigation Into Car-Following Behavior." *Transportation Research Record* 2678 (5): 812–826. https://doi.org/10.1177/03611981231192999.

Zhang, Junjie, Yunpeng Wang, and Guangquan Lu. 2019. "Impact of Heterogeneity of Car-Following Behavior on Rear-End Crash Risk." *Accident Analysis & Prevention* 125:275–289. https://doi.org/10.1016/j.aap.2019.02.018.

Zheng, Shi-Teng, Rui Jiang, Bin Jia, Junfang Tian, Marouane Bouadi, Michail A. Makridis, and Anastasios Kouvelas. 2023. "A Parsimonious Enhanced Newell's Model for Accurate Reproduction of Driver and Traffic Dynamics." *Transportation Research Part C: Emerging Technologies* 154:104276. https://doi.org/10.1016/j.trc.2023.104276.

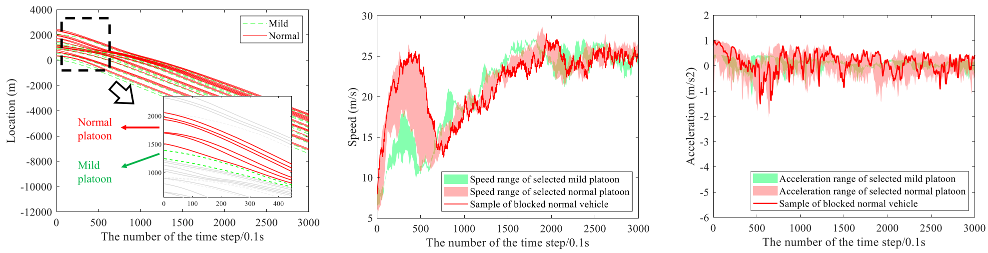
Appendix. Analysis of mild driving

Figures A1–A5 illustrates vehicle trajectories in heterogeneous traffic flow with different proportions of Mild drivers. When Normal drivers are involved in Mild traffic flow, vehicle trajectories are no longer neatly aligned and some platoons are formed, see Figure A1. These platoons are formed in traffic flow with Mild-style vehicles as the leader. For example, as the zooming-in trajectory diagram shows in Figure A1(a), several platoons are led by Mild vehicles, see the green dashed lines. This is because Mild vehicles have smaller acceleration and speed compared to Normal vehicles as settings, which makes a



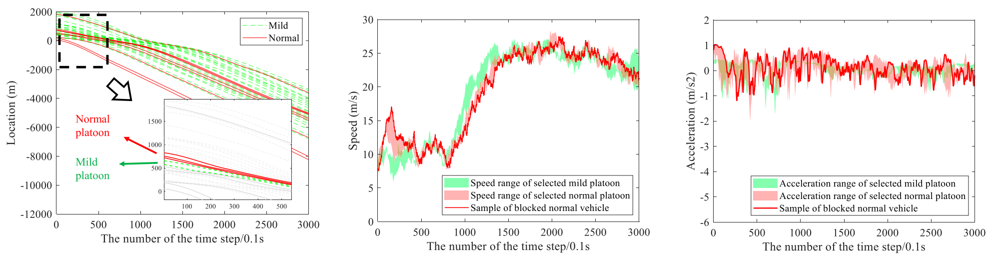
(a) Location. (b) Speed. (c) Acceleration.

Figure A1. Vehicle trajectories of traffic flow with 20% Mild drivers and 80% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.



(a) Location. (b) Speed. (c) Acceleration.

Figure A2. Vehicle trajectories of traffic flow with 40% Mild drivers and 60% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

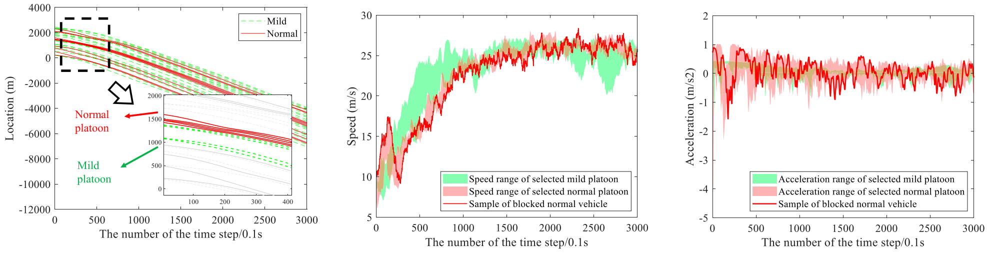


(a) Location. (b) Speed. (c) Acceleration.

Figure A3. Vehicle trajectories of traffic flow with 60% Mild drivers and 40% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.

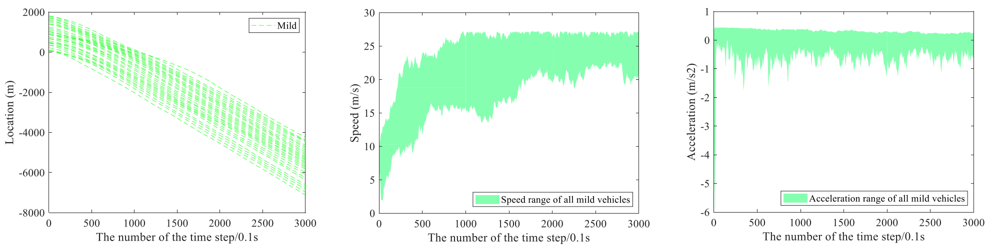
follower Mild vehicle fall behind its Normal leaders. Meanwhile, these Mild vehicles block their Normal followers and force them to behave in a Mild driving style with smaller acceleration and speed as well, see Figure A1(b,c). Specifically, the Mild leader notably slows down the Normal followers, eventually causing the Normal vehicles to adopt a mild driving style, as seen in the red region of Figure A1(b). The overall acceleration of Normal vehicles aligns with a mild driving style, whereas it still exhibits larger fluctuations compared to those in mild platoons, as shown in Figure A1(c).

When the proportions of Normal drivers decrease and Mild drivers increase, as shown in Figures A2–A4, the formation of platoons in traffic flow changes. In Figure A4(a) where there are more Mild drivers than Normal drivers, these Normal vehicles appear at the end of each platoon. This is because a Mild vehicle which has small acceleration and speed falls behind its Normal leader, forming a platoon leading by itself. Since these traffic flows have a high proportion of Mild drivers, most



(a) Location. (b) Speed. (c) Acceleration.

Figure A4. Vehicle trajectories of traffic flow with 80% Mild drivers and 20% Normal drivers. (a) Location. (b) Speed and (c) Acceleration.



(a) Location. (b) Speed. (c) Acceleration.

Figure A5. Vehicle trajectories of traffic flow with 100% Mild drivers. (a) Location. (b) Speed and (c) Acceleration.

vehicles in platoons are with an M_M style rather than N_M. Moreover, vehicle trajectories in a homogeneous traffic flow with 100% Mild drivers show a consistent and smooth trend, which is similar to a pure Normal traffic flow, see Figure A5.