TRB Annual Meeting

Analyzing Car-Following Behavior Using an Empirical Prior Statistical Learning Framework

--Manuscript Draft--

Full Title:	Analyzing Car-Following Behavior Using an Empirical Prior Statistical Learning Framework
Abstract:	This study investigates the integration of traditional car-following models with machine learning techniques to analyze car-following behavior in the presence of lane-changing interactions on multi-lane road segments. Traditional car-following models, such as the Intelligent Driver Model (IDM), are effective with limited data and robust to noise but often fail to capture complex driving behaviors. In contrast, data-driven models like Gaussian Process Regression (GPR) can model intricate behaviors but require extensive, high-quality datasets. To leverage the strengths of both approaches, we propose a hybrid framework that combines IDM with GPR, using IDM as a prior to enhance GPR predictions. The analysis demonstrates that the hybrid model significantly improves the accuracy of acceleration predictions compared to IDM alone, effectively capturing the nuances of real-world driving scenarios. This approach mitigates the limitations of traditional models and reduces the dependency on large datasets. The findings underscore the potential of combining traditional and data-driven methods to improve traffic behavior simulations, offering promising applications in traffic management and autonomous driving. Future research will focus on extending this framework to other driving behaviors and integrating additional data sources for further enhancement.
Additional Information:	
Question	**Response**
The total word count limit is 7500 words including tables. Each table equals 250 words and must be included in your count. Papers exceeding the word limit may be rejected. My word count is:	5653
Manuscript Classifications:	Operations; Traffic Simulation ACP80; Calibration/Validation; Car-following; Vehicle Trajectory
Manuscript Number:	TRBAM-25-04257
Article Type:	Presentation
Order of Authors:	Kaitai Yang
	Yuan-Zheng Lei
	Xianfeng Yang

ANALYZING CAR-FOLLOWING BEHAVIOR USING AN EMPIRICAL PRIOR STATISTICAL LEARNING FRAMEWORK

Kaitai Yang
Ph.D. Candidate & Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Maryland, College Park, MD, U.S., 20742
Email: kaitai74@umd.edu

Yuan-Zheng Lei
Ph.D. Candidate & Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Maryland, College Park, MD, U.S., 20742
Email: yzlei@umd.edu

Xianfeng Terry Yang, Ph.D., Corresponding Author
Associate Professor
Department of Civil and Environmental Engineering
University of Maryland, College Park, MD, U.S., 20742
Email: xtyang@umd.edu

Word Count: 5153 words + 2 table(s) × 250 = 5653 words

Submission Date: August 1, 2024

ABSTRACT

This study investigates the integration of traditional car-following models with machine learning techniques to analyze car-following behavior in the presence of lane-changing interactions on multi-lane road segments. Traditional car-following models, such as the Intelligent Driver Model (IDM), are effective with limited data and robust to noise but often fail to capture complex driving behaviors. In contrast, data-driven models like Gaussian Process Regression (GPR) can model intricate behaviors but require extensive, high-quality datasets. To leverage the strengths of both approaches, we propose a hybrid framework that combines IDM with GPR, using IDM as a prior to enhance GPR predictions. The analysis demonstrates that the hybrid model significantly improves the accuracy of acceleration predictions compared to IDM alone, effectively capturing the nuances of real-world driving scenarios. This approach mitigates the limitations of traditional models and reduces the dependency on large datasets. The findings underscore the potential of combining traditional and data-driven methods to improve traffic behavior simulations, offering promising applications in traffic management and autonomous driving. Future research will focus on extending this framework to other driving behaviors and integrating additional data sources for further enhancement.

Keywords: Car-following model, Microscopic simulation

INTRODUCTION

Car-following behavior refers to the dynamics of a driver maintaining a safe and efficient distance behind a leading vehicle while traveling on a roadway. This behavior is influenced by various factors such as the relative speed and distance between the vehicles, the driver's reaction time, and their desire to maintain a comfortable and safe driving experience. Understanding car-following behavior is crucial for traffic flow analysis, safety assessments, and the development of advanced driver-assistance systems (ADAS). Researchers have developed numerous models over the years to simulate this behavior, incorporating different assumptions and approaches to capture the complexities of real-world driving. These models are essential for designing traffic management systems, improving road safety, and developing autonomous driving technologies.

The first car-following concepts were proposed in the early 1950s (*1*). Since then, different classic car-following models have been developed based on various assumptions. For example, the Gazis-Herman-Rothery (GHR) model employs a linear form that considers reaction time, a sensitivity parameter, and the speed difference to simulate the following vehicle's acceleration ((*2*), (*3*)). The primary advantage of this model is its simplicity. However, it has several disadvantages, such as assuming uniform driver behavior during both acceleration and deceleration phases, which overlooks potential variations in driver responses. Subsequently, (*4*) introduced a memory function into the linear GHR model to address the limitation of using a constant to react to speed differences. Moreover, (*5*) proposed an improved version of the model to simulate drivers' reactions to multiple vehicles ahead, rather than solely the leading vehicle. Another category of car-following models primarily focuses on the following distance rather than the speed difference as the input, whereas the GHR model and its extensions mainly consider the speed difference. The rationale behind this approach is the drivers' intent to maintain a safe distance from the leading vehicle. The most popular model based on this concept is Gipps' car-following model (*6*), which has been used by simulation models (*7*). According to this model, drivers will always opt for the lower speed between the free flow mode and the car-following mode, underpinning the assumption that drivers aim to ensure their vehicle can be safely stopped, even if the leading vehicle suddenly brakes. Bando's Optimal Velocity Model (OVM) was proposed with the notion that each driver has an optimal velocity based on their space headway and will constantly adapt their speed to reach this optimal velocity (*8*). However, a significant disadvantage of the OVM is its potential to generate unrealistic acceleration rates under certain circumstances (*9*). To address this issue, the Full Velocity Difference Model (FVDM) was introduced by incorporating a term to represent the driver's reaction to the speed difference with the leading vehicle (*10*). Recognizing that physical inputs alone do not fully capture driver behavior, another category of car-following models, named psycho-physical models, incorporates perceptions into their framework. A prominent example of this approach is Wiedemann's car-following model, widely adopted by simulation software (*11*), segregates driving scenarios into different regimes (free driving, following, closing in, and emergency braking) and develops the acceleration equation individually for each regime (*12*). While Wiedemann's model offers a nuanced understanding of driver behavior by considering psychological factors, it is not without drawbacks. One significant limitation is its reliance on predefined thresholds to transition between driving regimes, which can oversimplify or misrepresent the fluidity of human driving behavior. Also, the equations adopted for the different thresholds are undisclosed (*13*). The Intelligent Driver Model (IDM), proposed by (*14*), stands out among car-following models for its unique approach to simulating driver behavior. Unlike previous models, IDM incorporates both the driver's desired speed and desired following distance to determine the acceleration rate, factoring in the current speed, speed difference, minimum safe following distance, and maximum acceleration. This model has proven its versatility across various following regimes, offering smooth transitions between them (*15*). Moreover, each parameter in the IDM can be related to a separate factor of the driver and has a clear physical meaning.

With advancements in machine learning and the increasing availability of detailed vehicle trajectory data ((*16*), (*17*), (*18*), (*19*), (*20*)), attention has shifted towards data-driven car-following models, garnering interest from researchers worldwide ((*21*), (*22*), (*23*), (*24*), (*25*), (*26*)). A significant branch in this domain employs Artificial Neural Networks (ANN) to analyze car-following behavior, providing a novel approach to understanding and predicting driver actions. Like classical car-following models, these models use various inputs to predict the driver's acceleration, speed, or following distance at the next time stamp (*25*). By categorizing drivers into different driving modes, (*21*) utilized the following distance and the speed of the leading vehicle as inputs for the ANN to predict the speed of the following vehicle. Their simulations demonstrated that this model outperforms the Gipps-based model. However, it's noteworthy that their study was based on a dataset limited to only 300 seconds of real-world data for training, which might not suffice for comprehensive model training. (*27*) explored the use of speed, speed difference, and following distance to predict the following vehicle's acceleration rate, indicating that predictive accuracy highly depends on the training data utilized. Furthermore, (*28*) introduced the driver's reaction delay as an additional input and evaluated the model's predictive performance using the Next Generation Simulation (NGSIM) dataset (*16*), achieving satisfactory results. (*25*) proposed a recurrent neural network-based car-following model to examine its applicability in predicting oscillations, comparing its performance with the IDM and demonstrating improved efficacy. Compared to classical car-following models, given a sufficient supply of training data, data-driven car-following models possess merits such as higher prediction accuracy, the capability to learn complex behaviors, and the ability to continuously improve through ongoing data input. However, data-driven models are often criticized for their notable sensitivity to data quality, their lower ability to generalize, and their lack of transparency and interpretability, stemming from their nature as "black boxes" (*29*).

By recognizing the strengths and weaknesses of traditional car-following models and data-driven car-following models, researchers have sought to combine both approaches to leverage their respective advantages (*30*). (*31*) developed a physics-informed deep learning car-following model to predict the acceleration of the following vehicle. The primary concept involves first using traditional car-following models to generate collocation data, which is then combined with observed data. By incorporating the prediction errors from both the traditional car-following models and ANNs, the results demonstrate improved accuracy in performance. Later (*32*) proposed a physics-informed Transformer model for predicting vehicle's longitudinal trajectories.

The main idea of this type of physics-informed deep learning method involves changing the simulation problem into a prediction problem. The collocation data is first generated by the traditional car-following models, and then the collocation data and the observed data are split into training and testing sets. The model is trained by modifying the loss function of the deep learning model. However, the goal of analyzing car-following behavior should be focused on simulation rather than prediction. In simulation, only the initial state of the following vehicle should be taken as known, rather than taking the information of the following vehicle at any time stamp. Additionally, incorporating the car-following model into the loss function of the deep learning method does not always guarantee better performance, as the collocation data is purely generated without any observation. Moreover, the traditional car-following model often performs unsatisfactorily when dealing with lane-changing scenarios. Acknowledging this, this study proposes a new method to combine the car-following model with Gaussian Process Regression to analyze car-following behavior interacting with lane-changing behaviors. Instead of integrating the car-following model into the loss function, it is treated as the prior belief of the estimated acceleration. The performance of the proposed model under different car-following scenarios is compared with the traditional car-following model and pure GPR, showing that the proposed method outperforms both.

The remainder of the paper is organized as follows: Section 2 introduces the proposed methodology; Section 3 conducts the case study of the proposed model and demonstrates the performance using real-world data ; and finally, Section 4 concludes with a discussion on the proposed model and suggestions for future work.

METHODOLOGY

Deterministic model

Traditional car-following models exhibit significant drawbacks when interacting with lane-changing maneuvers. As shown in Figure 1, when vehicle $j-1$ cuts in, the gap between vehicle i and $i-1$ changes abruptly, leading to a sudden drop in the predicted acceleration for vehicle i. This deficiency stems from the inherent accident-free assumption of traditional car-following models. These models assume that the following vehicle will anticipate the worst-case scenario, where the leading vehicle suddenly brakes to a complete standstill. Similarly, for vehicle j when the leading vehicle abruptly leaves the current lane, the sudden increase in the gap can result in an unrealistically high predicted acceleration. For the proposed framework, the empirical deterministic model plays a crucial role in accurately predicting the longitudinal trajectory. In this study, the IDM is chosen to serve as the deterministic model for the following reasons:

1). Comprehensive Consideration of Driver Behavior: IDM incorporates both the driver's desired speed and desired following distance to determine the acceleration rate, taking into account factors such as current speed, speed difference, minimum safe following distance, and maximum acceleration. This model has demonstrated its versatility across various following regimes, offering smooth transitions between them.

2). Parameter Interpretability and Flexibility: Each parameter in the IDM can be associated with a distinct aspect of driver behavior and has a clear physical meaning. One of the advantages of the IDM is its ability to represent a wide range of driving styles by adjusting its parameter values.

3). Realistic Accelerating Behaviors: When simulating accelerating behavior, the IDM generates realistic acceleration rates than other car-following models, avoiding the unrealistic values that can sometimes be produced by alternative models.

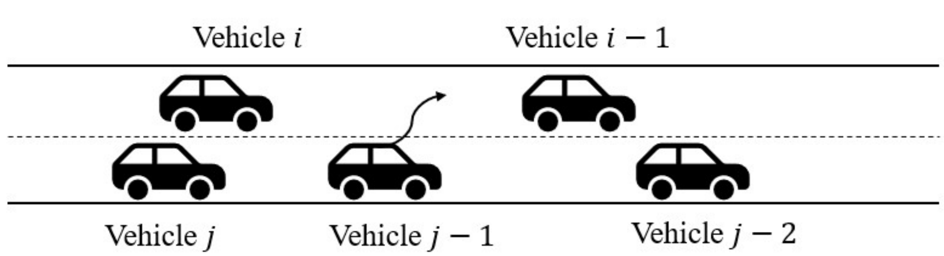


FIGURE 1: Illustration of the lane changing behaviors

The original IDM can be expressed as follows:

$$\frac{dv_n}{dt} = a_n[1-(\frac{v_n}{v_{n,0}})^{\delta} - (\frac{s^*(v_n,\Delta v_n)}{s_n})^2] \tag{1}$$

$$s^*(v_n,\Delta v_n) = s_0 + max(v_n T + \frac{v_n \Delta v_n}{2\sqrt{a_n b_n}}, 0) \tag{2}$$

$$\Delta v_n = v_n - v_{n-1} \tag{3}$$

$$S_n = x_{n-1} - l_{n-1} - x_n \tag{4}$$

The free accelerating term $a[1-(\frac{v_n}{v_0})^{\delta}]$, governs the acceleration of the vehicle. Here, a represents the maximum acceleration, and v_0 denotes the vehicle's desired speed. Given an unobstructed path for a stationary vehicle, the vehicle would first accelerate at the rate of a, and the acceleration gradually decreases as the speed increases. Such reduction is controlled by the exponent term δ, and the vehicle would not exceed its desired speed. In accordance with the IDM author's recommendation ((*15*)), this paper assigns the value of 4 to δ. On the other hand, the vehicle's decelerating process is regulated by the braking term $(\frac{s^*(v_n,\Delta v_n)}{s_n})^2$, where $s^*(v_n,\Delta v_n)$ represents vehicle's desired gap and s_n is the actual gap as shown by Equations 1 - 4. The term $s_0 + v_n T$ denotes the vehicle's desired following distance at the steady state. Here, the minimum gap, s_0, represents the space gap between the standstill vehicles, T is the time gap that the driver aims to maintain while in motion. The dynamic term $\frac{v_n \Delta v_n}{2\sqrt{ab}}$, symbolizes the driver's response to the speed difference Δv_n based on its own comfortable deceleration b.

Stochastic modeling

In this study, we employ Gaussian Process Regression (GPR) to model the underlying dynamics of car-following behavior. This section would introduce the GPR from function-space view (to understand the GPR from weight-space view, readers could refer to (*33*), A Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution. Specifically, a distribution over functions $f(\mathbf{x})$ are described by a Gaussian process. It is fully specified by its mean function $m(\mathbf{x})$and covariance function $k(\mathbf{x},\mathbf{x}')$ as shown below.

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{5}$$

$$k(\mathbf{x},\mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{6}$$

Thus, the Gaussian process can be expressed as:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')) \tag{7}$$

Generally, for a pure Gaussian Process Regression (GPR) problem, the mean function is set to zero, represents the prior belief of the value of $f(\mathbf{x})$. On the other hand, the covariance function determines the relationship between variables. Consider the observed data $(\mathbf{x}_i, f_i | i = 1,2,3,...,n)$ and the estimated mean function $\mathbf{f}$, the goal is to predicted new data $\mathbf{x}_*$ as $\mathbf{f}_*$. The joint distribution of the $\mathbf{f}$ and $\mathbf{f}_*$ is multivariate Gaussian distribution, which can be expressed as:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}) \\ m(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{nn} & \mathbf{K}_{n*} \\ \mathbf{K}_{*n} & \mathbf{K}_{**} \end{bmatrix} \right) \tag{8}$$

Knowing the joint probability of $\mathbf{f}$ and $\mathbf{f}_*$, the conditional distribution of $\mathbf{f}_*$ could be derived:

$$\mathbf{f}_* | \mathbf{f}, \mathbf{x}, \mathbf{x}_* \sim \mathcal{N}(\mathbf{K}_{*n}\mathbf{K}_{nn}^{-1}\mathbf{f}, \mathbf{K}_{**} - \mathbf{K}_{*n}\mathbf{K}_{nn}^{-1}\mathbf{K}_{*n}) \tag{9}$$

It should be noticed the $f(\mathbf{x})$ is a noise-free function. For a more realistic scenario, the noisy versions $y = f(\mathbf{x}) + \varepsilon$, where ε is a independent identically distributed Gaussian noise with variance

σ_n^2. The prior on the noisy observations now is:

$$\mathrm{cov}(\mathbf{y}) = \mathbf{K}_{nn} + \sigma_\varepsilon^2 \mathbf{I} \tag{10}$$

and Equation 10 now becomes:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}) \\ m(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{nn} + \sigma_\varepsilon^2 \mathbf{I} & \mathbf{K}_{n*} \\ \mathbf{K}_{*n} & \mathbf{K}_{**} \end{bmatrix} \right) \tag{11}$$

And the conditional probability of $\mathbf{y}$ over $\mathbf{f}_*$ becomes:

$$\mathbf{y}_* | \mathbf{y}, \mathbf{x}, \mathbf{x}_* \sim \mathcal{N}(\mathbf{y}_* | \mathbf{K}_{*n}[\mathbf{K}_{nn} + \sigma_\varepsilon^2 \mathbf{I}]^{-1}\mathbf{y}, \mathbf{K}_{**} - \mathbf{K}_{*n}[\mathbf{K}_{nn} + \sigma_\varepsilon^2 \mathbf{I}]^{-1}\mathbf{K}_{n*})) \tag{12}$$

Although GPR is a promising and widely-used non-parametric method that performs well across various tasks, it faces criticism when dealing with large datasets due to its computational complexity. The time complexity of GPR is $\mathcal{O}(n^3)$ and the storage requirement is $\mathcal{O}(n^2)$, where n is the number of samples. These limitations make it challenging to scale GPR to larger datasets.

Integration of deterministic and stochastic model

Notice that the input data used for calibrating the IDM only requires the variables: the following distance S_n, the speed difference Δv_n, and the speed of the following vehicle v_n. Consistent with the IDM, this study also adopts these three variables as inputs. Define the observed data set as $\mathbf{x} = \{x^i = (S_n^i, v_n^i, \Delta v_n^i) \mid i = 1, 2, \ldots, N\}$ and a set of scalars $\mathbf{y} = \{a^i \mid i = 1, 2, \ldots, N\}$, representing the observed acceleration rates of the following vehicle. The data $\mathbf{x}$ is first calibrated using $f_{IDM}(\mathbf{x} \mid \lambda)$ to obtain the optimal set of parameters λ. The mean function of the GPR is then set as $m(\mathbf{x}) = f_{IDM}(\mathbf{x} \mid \lambda)$, indicating that the IDM-predicted acceleration of the following vehicle represents the prior belief of the acceleration value in the GPR. After training the GPR, the performance of the proposed model is evaluated. Unlike previous work (*31*), this study does not use the observed information of the following vehicle as input for prediction. This approach mitigates the risk of overfitting and ensures a fair comparison with traditional car-following behavior models. The initial state $x^0 = (S_n^0, v_n^0, \Delta v_n^0)$ is set to the known condition, and the predicted acceleration $\widehat{a}^0$ is used to update the new state $\widehat{x^1} = (\widehat{S_n^1}, \widehat{v_n^1}, \widehat{\Delta v_n^1})$ for the next timestamp as shown by Equation 13 and 14 where Δt is set to 0.1 seconds in this study. In other words, for the prediction at time step $k+1$, the input is the simulated $\widehat{x^k} = (\widehat{S_n^k}, \widehat{v_n^k}, \widehat{\Delta v_n^k})$.

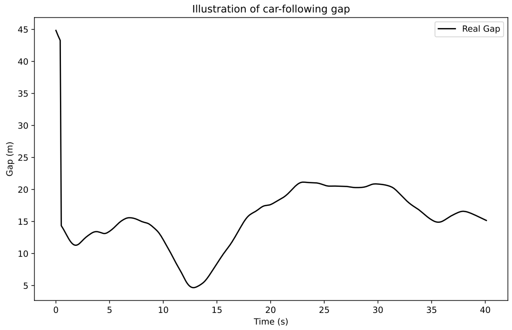
$$v_n(t + \Delta t) = v_n + a_n(t) \cdot \Delta t \tag{13}$$

$$x_n(t + \Delta t) = x_n(t) + v_n(t + \Delta t) \cdot \Delta t \tag{14}$$

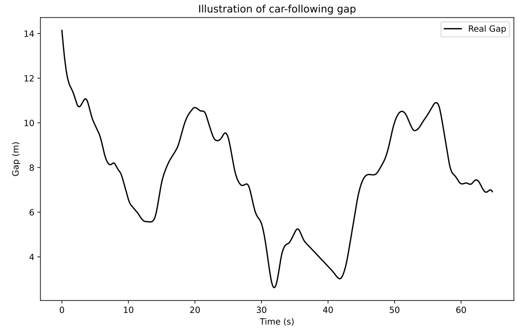
CASE STUDY

Experiment setup

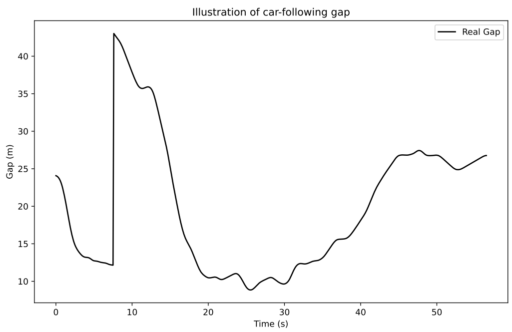
To validate the performance of the proposed framework, real-world vehicle trajectory data collected by the NGSIM was utilized. The detailed vehicle trajectory data was gathered on the eastbound I-80 in Emeryville, CA, on April 13, 2005. The study segment featured five freeway lanes and a high-occupancy vehicle (HOV) lane, spanning approximately 1,640 feet. Given the presence of errors in the original dataset, this study adopted the reconstructed version proposed by (*34*). This study aims to investigate car-following behavior in the context of lane-changing interactions. To this end, five vehicle trajectories were selected, encompassing various scenarios: car-following without lane-changing (vehicle 1593 and vehicle 1700), car-following with cutting-in (vehicle 797), car-following with the leading vehicle leaving the current lane (vehicle 1649), and car-following with both cutting in and leading vehicle leaving the current lane (vehicle 3222). The following gap of each target vehicle is shown in Figure 2.



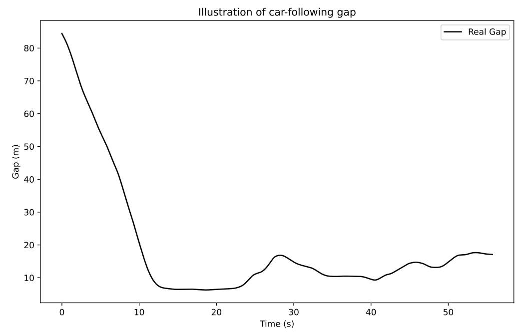
(a) Following gap of vehicle 797



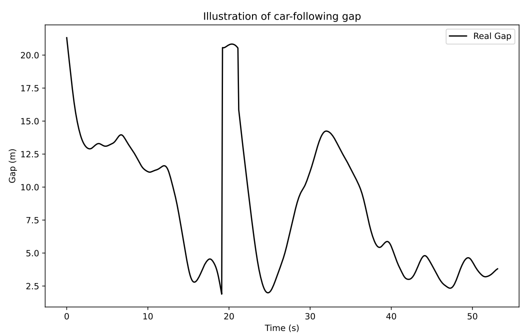
(b) Following gap of vehicle 1593



(c) Following gap of vehicle 1649



(d) Following gap of vehicle 1700



(e) Following gap of vehicle 3222

FIGURE 2: Illustration of the following gap of each target vehicle

Before testing the performance of the proposed model, each individual trajectory was first calibrated using the IDM. Various methods can be used to calibrate the IDM, such as maximum likelihood estimation (*35*) and least square estimation (*36*). Given the complexity of the IDM, this study adopts the Genetic Algorithm (GA) for calibration due to its heuristic nature and gradient-free approach (*37*). The parameters for executing the GA in this study are as follows: the GA runs for a maximum of 200 generations, with each generation consisting of a population of 100 individuals. The mutation rate is set to 0.05. On the other hand, it is essential to ensure that the calibrated parameters of the IDM remain within realistic bounds to maintain interpretability and prevent overfitting. The boundaries for these parameters are defined as follows: the time gap T s set between 0.1 and 3 seconds, and the minimum spacing s_0 is limited to 1 to 5 meters. The maximum acceleration a restricted to 0.1 to 4 m/s^2, which corresponds to a maximum acceleration rate of equivalent to 0-100km/h in 6 seconds, while the comfortable deceleration boundary is set at 0.1 to 9 m/s^2. The upper limit for the desired velocity v_0 is set to 33.6 m/s (120km/h). The lower limit must exceed the highest velocity observed in the dataset to avoid excessive deceleration due to the power of 4 applied in the term $(\frac{v_n}{v_0})^4$ This consideration is crucial because v_0 in the IDM is primarily designed for modeling acceleration rather than deceleration scenarios (*15*). By minimizing the error of following gaps, The calibrate parameters of each following vehicle is shown by Table 1. As mentioned before, the calibrated IDM model for each single trajectory is taken as the mean function of the GPR, and for the selection of the kernel function, after evaluating multiple available kernel functions, the study adopts the radial basis function (RBF) kernel as shown by Equation 15, where σ and θ are hyper-parameters. The goodness-of-fit (GoF) function and the measure of performance (MoP) in this study is the root mean square error (RMSE) of acceleration as shown by Equation 16.

$$k(\mathbf{x},\mathbf{x}') = \sigma^2 e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\theta^2}} \tag{15}$$

$$RMSE(a) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[a_i(t) - \widehat{a}_i(t)]^2} \tag{16}$$

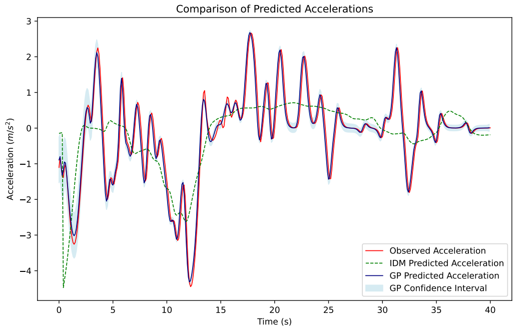
TABLE 1: Illustration of calibration result of IDM

Vehicle	v_0	T	s_0	a	b
797	31.79 m/s	1.08 s	4.96 m	0.77 m/s^2	2.09 m/s^2
1593	32.71 m/s	0.79 s	3.40 m	3.97 m/s^2	1.28 m/s^2
1649	29.51 m/s	1.10 s	5.00 m	0.40 m/s^2	1.25 m/s^2
1700	21.3 m/s	0.96 s	5.00 m	0.97 m/s^2	1.47 m/s^2
3222	31.14 m/s	0.99 s	1.25 m	4.00 m/s^2	8.99 m/s^2

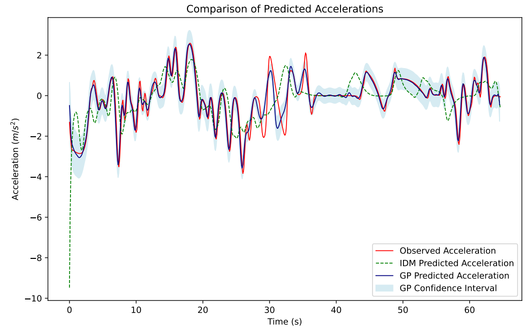
Results discussions

To compare the performance of the IDM, pure GPR with no empirical prior, and GPR with IDM prior, three groups of experiments were conducted. The corresponding RMSE values and the estimated accelerations and the hyperparameters of the proposed model are presented in Figure 3 and Table 2. In the plots, the red lines represent the real acceleration rate of each following vehicle, the green dashed lines depict the simulated acceleration by the IDM model, and the blue lines show the predicted acceleration simulated by the GPR with the IDM prior, while the light blue areas represent the 95% confidence intervals of the simulated accelerations. Across all trajectories, the observed acceleration fluctuates over time, capturing the dynamic nature of real-world vehicle behavior.

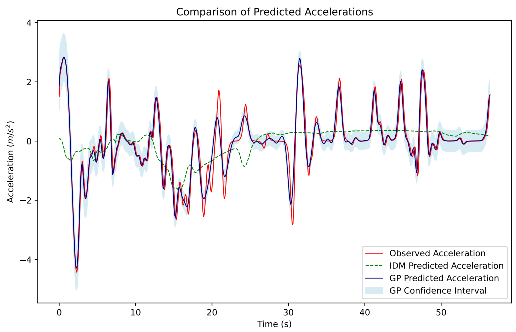
The analysis of the five trajectory plots comparing observed accelerations, IDM predicted accelerations, and GPR predicted accelerations, along with the GP confidence intervals, reveals several key insights. The trajectory of vehicle 797 depicts a cutting-in scenario where a vehicle from the adjacent lane cuts in and the leading vehicle switches lanes around 2 seconds. The target vehicle continues to follow the new vehicle until the end of the data. The plot shows that the IDM model predicts a drastic deceleration when the new vehicle cuts in. In contrast, the GPR with IDM prior provides a more accurate estimation, maintaining accuracy throughout the car-following



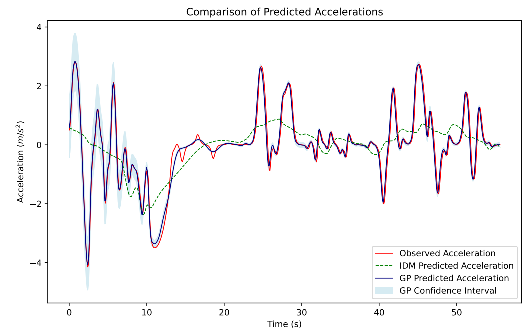
(a) Acceleration comparison for vehicle 797



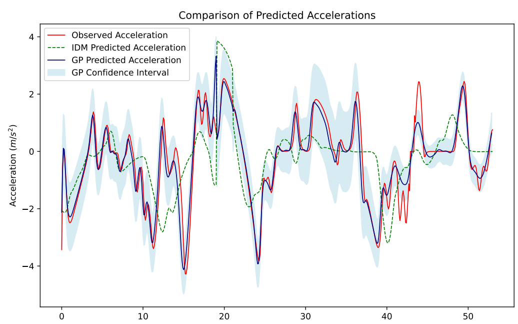
(b) Acceleration comparison for vehicle 1593



(c) Acceleration comparison for vehicle 1649



(d) Acceleration comparison for vehicle 1700



(e) Acceleration comparison for vehicle 3222

FIGURE 3: Comparison of estimated acceleration results

behavior simulation, whereas the IDM model fails to capture the fluctuations in acceleration. The RMSE values reflect this performance, with the IDM having an RMSE of 0.95, while the pure GPR and GPR with IDM prior have much lower RMSE values of 0.11 and 0.09, respectively. For vehicle 1593, although there are no lane-changing behaviors, the IDM model predicts a significant deceleration rate near -10 m/s^2 at the beginning, while the actual deceleration rate is around -3 m/s^2. The GPR with IDM prior performs better in this scenario, accurately reflecting the observed behavior. This is supported by the RMSE values, where the IDM has an RMSE of 1.02, and the pure GPR and GPR with IDM prior have RMSE values of 0.51 and 0.27, respectively. For vehicle 1649, where the leading vehicle leaves the current lane around 8 seconds, the IDM model does not show an increase in acceleration as the following gap increases, likely due to the estimated maximum acceleration being 0.40 m/s^2. Conversely, the GPR with IDM prior shows a better performance, accurately capturing the increase in acceleration. The RMSE values indicate this as well, with the IDM having an RMSE of 1.18, while the pure GPR shows a wore performance that IDM with RMSE of 2.68 ,while GPR with IDM prior have RMSE values 0.22, shows a better performance.For vehicle 1700, during the first 10 seconds of the simulation, both the IDM and GPR models fail to simulate the drastic fluctuations in acceleration. However, the GPR still outperforms the IDM by providing a closer match to the observed data. This is reflected in the RMSE values, where the IDM has an RMSE of 0.95, and the pure GPR and GPR with IDM prior have RMSE values of 0.51 and 0.15, respectively. Vehicle 3222 presents a more complex scenario where the leading vehicle leaves the current lane around 20 seconds, followed by another vehicle from the adjacent lane cutting in. The IDM model generates a sudden change from deceleration to maximum acceleration and then another drastic change back to deceleration. In contrast, the GPR with IDM prior avoids such abrupt transitions, offering a more realistic result. The RMSE values show this clearly, with the IDM having an RMSE of 1.29, while the pure GPR and GPR with IDM prior have RMSE values of 1.01 and 0.40, respectively.

TABLE 2: Results comparison

Vehicle	θ	σ^2	RMSE(a) (IDM)	RMSE(a) (pure GPR)	RMSE(a) (GPR with IDM prior)
797	3.06	1.03	0.95	0.11	0.09
1593	1.16	0.726	1.02	0.51	0.27
1649	2.99	1.81	1.18	2.68	0.22
1700	3.64	6.43	0.95	0.51	0.15
3222	3.41	2.64	1.29	1.01	0.40

The IDM predicted acceleration generally follows the observed trends but exhibits deviations at various points, indicating that while IDM captures overall behavior, it lacks the precision to accurately model complex interactions and abrupt changes in real driving scenarios. The traditional car-following model tends to show drastic switches in acceleration when dealing with lane-changing behavior. On the other hand, pure GPR can sometimes produce worse estimation results than IDM. In contrast, the predicted acceleration by GPR with IDM prior closely aligns with the observed acceleration across all trajectories, demonstrating the higher accuracy of the proposed method. The confidence intervals typically encompass the observed acceleration, suggesting reliable predictions and realistic uncertainty estimation. This consistency is evident in all five trajectories, where the proposed model consistently outperforms both the IDM model and pure GPR in terms of accuracy and robustness. Specifically, the proposed model's flexibility and capacity to incorporate both observed data and a physics-informed prior from IDM allow it to better adapt to variations in acceleration, capturing more intricate driving behaviors. This integration of data-driven approaches with traditional models proves beneficial, as it mitigates the risk of overfitting and ensures a fair comparison with traditional car-following models. Consequently, the GPR with IDM prior's superior performance highlights its potential advantages for applications in traffic management and autonomous driving, offering more precise and reliable predictions for vehicle accelerations. Overall, the comparative analysis of these trajectories underscores the effectiveness of the proposed model in modeling car-following behavior, demonstrating significant improvements over traditional IDM estimations.

CONCLUSION

This study investigates car-following behavior in conjunction with lane-changing interactions on a multi-lane road segment by combining a traditional car-following model with a machine learning method, specifically Gaussian Process Regression (GPR). Traditional car-following models, while robust to varying sampling rates and noise, often struggle to accurately replicate complex empirical driving behaviors. On the other hand, data-driven car-following models can capture these intricate behaviors but demand high-quality, large-scale datasets. To leverage the strengths of both approaches, this study proposes a new framework that integrates traditional car-following models with GPR. The proposed method utilizes the IDM as a prior in the GPR framework to enhance the accuracy of acceleration predictions. The analysis demonstrates that GPR, when informed by IDM, provides superior performance in predicting vehicle accelerations compared to using IDM and pure GPR alone. The results indicate that the GPR with IDM prior is more effective at modeling the complex interactions between car-following and lane-changing behaviors, capturing the nuances of real-world driving scenarios more accurately. This hybrid approach not only mitigates the limitations of traditional models but also reduces the dependency on large-scale, high-quality datasets typically required by purely data-driven methods.

Overall, the findings highlight the potential of combining traditional car-following models with advanced machine learning techniques to improve the accuracy and reliability of traffic behavior simulations. This integrated approach offers significant advantages for applications in traffic management and autonomous driving, providing more precise and robust predictions of vehicle dynamics. Future research could extend this framework to other aspects of driving behavior and explore the integration of additional data sources to further enhance model accuracy and applicability. The continued development of such hybrid models holds promise for advancing the field of traffic flow analysis and contributing to safer and more efficient transportation systems.

CREDIT

Kaitai Yang: Conceptualization, Methodology, Writing - original draft. **Yuan-Zheng Lei**: Conceptualization, Methodology, Writing - original draft. **Xianfeng Terry Yang**: Conceptualization, Methodology and Supervision.

ACKNOWLEDGEMENT

This research is supported by the award "OAC Core: Stochastic Simulation Platform for Assessing Safety Performance of Autonomous Vehicles in Winter Seasons (# 2234292)" which is funded by the National Science Foundation. The authors acknowledge the utilization of ChatGPT, a large language model, exclusively for enhancing the linguistic quality of the manuscript preparation process.

REFERENCES

1. Pipes, L. A., An operational analysis of traffic dynamics. *Journal of applied physics*, Vol. 24, No. 3, 1953, pp. 274–281.
2. Chandler, R. E., R. Herman, and E. W. Montroll, Traffic dynamics: studies in car following. *Operations research*, Vol. 6, No. 2, 1958, pp. 165–184.
3. Herman, R., E. W. Montroll, R. B. Potts, and R. W. Rothery, Traffic dynamics: analysis of stability in car following. *Operations research*, Vol. 7, No. 1, 1959, pp. 86–106.

4. Lee, G., A generalization of linear car-following theory. *Operations research*, Vol. 14, No. 4, 1966, pp. 595–606.
5. Herman, R., Car-following and steady state flow. In *Theory of Traffic Flow Symposium Proceedings, 1959*, 1959.
6. Gipps, P. G., A behavioural car-following model for computer simulation. *Transportation research part B: methodological*, Vol. 15, No. 2, 1981, pp. 105–111.
7. Barceló, J. and J. Casas, Dynamic network simulation with AIMSUN. In *Simulation approaches in transportation analysis: Recent advances and challenges*, Springer, 2005, pp. 57–98.
8. Bando, M., K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, Dynamical model of traffic congestion and numerical simulation. *Physical review E*, Vol. 51, No. 2, 1995, p. 1035.
9. Helbing, D. and B. Tilch, Generalized force model of traffic dynamics. *Physical review E*, Vol. 58, No. 1, 1998, p. 133.
10. Jiang, R., Q. Wu, and Z. Zhu, Full velocity difference model for a car-following theory. *Physical Review E*, Vol. 64, No. 1, 2001, p. 017101.
11. Fellendorf, M. and P. Vortisch, Microscopic traffic flow simulator VISSIM. *Fundamentals of traffic simulation*, 2010, pp. 63–93.
12. Wiedemann, R., Simulation des Strassenverkehrsflusses., 1974.
13. Saifuzzaman, M. 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, 2014, pp. 379–403.
14. Treiber, M., A. Hennecke, and D. Helbing, Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, Vol. 62, No. 2, 2000, p. 1805.
15. Treiber, M. and A. Kesting, Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*, 2013, pp. 983–1000.
16. Alexiadis, V., J. Colyar, J. Halkias, R. Hranac, and G. McHale, The next generation simulation program. *Institute of Transportation Engineers. ITE Journal*, Vol. 74, No. 8, 2004, p. 22.
17. Krajewski, R., J. Bock, L. Kloeker, and L. Eckstein, The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st international conference on intelligent transportation systems (ITSC)*, IEEE, 2018, pp. 2118–2125.
18. Bock, J., R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2020, pp. 1929–1934.
19. Zhan, W., L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, et al., Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*, 2019.
20. Zheng, O., M. Abdel-Aty, L. Yue, A. Abdelraouf, Z. Wang, and N. Mahmoud, Citysim: A drone-based vehicle trajectory dataset for safety-oriented research and digital twins. *Transportation Research Record*, 2022, p. 03611981231185768.
21. Panwai, S. and H. Dia, Neural agent car-following models. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 1, 2007, pp. 60–70.

22. Mathew, T. V. and K. Ravishankar, Neural network based vehicle-following model for mixed traffic conditions, 2012.
23. Papathanasopoulou, V. and C. Antoniou, Towards data-driven car-following models. *Transportation Research Part C: Emerging Technologies*, Vol. 55, 2015, pp. 496–509.
24. He, Z., L. Zheng, and W. Guan, A simple nonparametric car-following model driven by field data. *Transportation Research Part B: Methodological*, Vol. 80, 2015, pp. 185–201.
25. Zhou, M., X. Qu, and X. Li, A recurrent neural network based microscopic car following model to predict traffic oscillation. *Transportation research part C: emerging technologies*, Vol. 84, 2017, pp. 245–264.
26. Zhu, M., X. Wang, and Y. Wang, Human-like autonomous car-following model with deep reinforcement learning. *Transportation research part C: emerging technologies*, Vol. 97, 2018, pp. 348–368.
27. Chong, L., M. M. Abbas, and A. Medina, Simulation of driver behavior with agent-based back-propagation neural network. *Transportation Research Record*, Vol. 2249, No. 1, 2011, pp. 44–51.
28. Khodayari, A., A. Ghaffari, R. Kazemi, and R. Braunstingl, A modified car-following model based on a neural network model of the human driver effects. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 42, No. 6, 2012, pp. 1440–1449.
29. Wang, Z., Y. Shi, W. Tong, Z. Gu, and Q. Cheng, Car-following models for human-driven vehicles and autonomous vehicles: A systematic review. *Journal of transportation engineering, Part A: Systems*, Vol. 149, No. 8, 2023, p. 04023075.
30. Bahari, M., I. Nejjar, and A. Alahi, Injecting knowledge in data-driven vehicle trajectory predictors. *Transportation research part C: emerging technologies*, Vol. 128, 2021, p. 103010.
31. Mo, Z., R. Shi, and X. Di, A physics-informed deep learning paradigm for car-following models. *Transportation research part C: emerging technologies*, Vol. 130, 2021, p. 103240.
32. Geng, M., J. Li, Y. Xia, and X. M. Chen, A physics-informed Transformer model for vehicle trajectory prediction on highways. *Transportation research part C: emerging technologies*, Vol. 154, 2023, p. 104272.
33. Williams, C. K. and C. E. Rasmussen, *Gaussian processes for machine learning*, Vol. 2. MIT press Cambridge, MA, 2006.
34. Punzo, V., M. T. Borzacchiello, and B. Ciuffo, On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, 2011, pp. 1243–1262.
35. Treiber, M. and A. Kesting, Microscopic calibration and validation of car-following models–a systematic approach. *Procedia-Social and Behavioral Sciences*, Vol. 80, 2013, pp. 922–939.
36. Ossen, S. and S. P. Hoogendoorn, Car-following behavior analysis from microscopic trajectory data. *Transportation Research Record*, Vol. 1934, No. 1, 2005, pp. 13–21.
37. Punzo, V., Z. Zheng, and M. Montanino, About calibration of car-following dynamics of automated and human-driven vehicles: Methodology, guidelines and codes. *Transportation Research Part C: Emerging Technologies*, Vol. 128, 2021, p. 103165.