

A Review on Trajectory Datasets on Advanced Driver Assistance System Equipped-vehicles

Hang Zhou, Ke Ma*, Xiaopeng Li

Abstract—This paper presents a comprehensive review of trajectory datasets from vehicles equipped with Advanced Driver Assistance Systems, with the aim of precisely modeling the behavior of Autonomous Vehicles (AVs). This study emphasizes the importance of trajectory data in the development of AV models, especially in car-following scenarios. We introduce and evaluate several datasets: the OpenACC Dataset, the Connected & Autonomous Transportation Systems Laboratory Open Dataset, the Vanderbilt ACC Dataset, the Central Ohio Dataset, and the Waymo Open Dataset. Each dataset offers unique insights into AV behaviors, yet they share common challenges in terms of data availability, processing, and standardization. After a series of data cleaning, outlier removal, and statistical analysis, this paper transforms datasets of varied formats into a uniform standard, thereby improving their applicability for modeling AV car-following behavior. Key contributions of this study include: 1. the transformation of all datasets into a unified standard format, enhancing their utility for broad research applications; 2. a comparative analysis of these datasets, highlighting their distinct characteristics and implications for car-following model development; 3. the provision of guidelines for future data collection projects.

I. INTRODUCTION

The rapid development of Autonomous Vehicles (AVs) necessitates a more profound understanding of their behavior to ensure traffic safety and efficiency. This involves developing strong, real-world models that accurately capture the complex behaviors of AVs. Substantial advancements have been made in the development of models to analyze AV behavior in varied scenarios such as car-following, platooning, intersection navigation, and lane-changing maneuvers. These models demand fine-tuned parameters, underpinned by real-world empirical data.

Recently, a surge in autonomous driving datasets like BDD100K [1], Argoverse [2], [3], and Waymo Open Dataset [4] has been observed, with substantial reviews in the literature [5], [6]. Primarily, these datasets concentrate on perception data collected by cameras, LiDAR, radar, Inertial Measurement Units (IMU), and other sensors. Although these perception data can ensure the basic safety conditions and be used to develop simple dynamics of AV, they are not accurate enough for complex behavior models. This oversight constrains the progression of detailed AV-specific traffic flow analysis.

To address the need for accurate vehicle dynamics behavior modeling in future transportation construction and

operation, trajectory data enriched with Global Position System (GPS) information emerges as a critical resource. This data type, superior in accuracy to perception data, is indispensable for developing sophisticated AV dynamics models. Such models require data that encapsulate a wide array of scenarios (like rural road, highway, or intersection), cases (car-following, lane-changing, or merging), and states (low/high speed following, cruise control, or emergency deceleration), with precise measurements of vehicle position and speed. Consequently, leveraging GPS to gather extensive trajectory data presents an economically viable alternative to perception data.

However, the availability of trajectory data is currently quite limited. Compounding this issue is the challenge of processing trajectory data due to inconsistencies in acquisition methods, equipment used (different GPS models, and whether they carry IMUs), vehicles involved (brands, models, and years), road environments, and most importantly, data release standards. These differences hinder the development of uniform datasets essential for calibrating AV behavioral models, like traditional linear models (will be further explained below), and simulating authentic driving behavior. The current research highlights a shortage of datasets focusing on trajectory data. It also points out that the variety in data formats and collection methods, such as different sampling rates, adds complexity. This situation delineates a significant research need for a comprehensive review that summarizes and analyzes current trajectory datasets and provides guidelines for future trajectory data collection projects.

In addressing these challenges, this study undertakes a comprehensive review of existing open datasets that include trajectory data, particularly from vehicles equipped with the Advanced Driver Assistance System (ADAS). We focus on the dataset with car-following trajectories, since it is the most important model in the traffic follow analysis. Our objective is to transfer these datasets into a uniform, user-friendly format. This proposed standard can serve as a benchmark for future trajectory datasets. Our methodology encompasses the extraction of main features related to the car-following behavior, data cleaning, and the removal of outliers. Our research also includes a statistical analysis and the calibration of these datasets. Our key contributions are as follows:

- We summarize the existing open trajectory datasets on ADAS, which provide a comprehensive optional data list for the traffic flow analysis especially for the development of the car-following model.
- We transform three datasets, i.e., the OpenACC Dataset,

*Research supported by the U.S. National Science Foundation under Grant No.2313578. (corresponding author: Ke Ma). Hang Zhou, Ke Ma, and Xiaopeng Li are now with the Department of Civil and Environmental Engineering, Univ. of Wisconsin-Madison, Madison, WI, USA. (e-mails: kma62@wisc.edu).

the Connected & Autonomous Transportation Systems (CATS) Laboratory Open Dataset, and the Vanderbilt ACC Dataset, into a uniform and standardized format. This standard format is expected to provide a guideline to collect and publish trajectory data for future studies.

- Then, an initial data cleaning is conducted to fill in missing data and remove outlier and unstable data. The improved data has a higher qualify for car-following model development.
- In the experiments, we compare the statistical analysis and the calibration results of a linear car-following model to three selected datasets. Results show that the CATS Laboratory Open Dataset and the Vanderbilt ACC Dataset are suitable for middle speed model, while the OpenACC Dataset is better for low speed model.

The remainder of this paper is organized as follows: Section II reviews related works on existing AV trajectory datasets. Section III introduces our data processing method. Section IV describes the calibration results. Finally, conclusions are presented in Section V.

II. RELATED WORKS

The availability of high-quality and publicly accessible datasets is crucial for the development and validation of novel AV control models. Due to the combined efforts of research institutions and the industry, a significant number of datasets have been developed over recent decades. These datasets have been thoroughly reviewed and evaluated in academic literature, including works by entities like the IEEE Emerging Transportation Technology Testing (ET3) Committee. This section primarily focuses on reviewing datasets that feature extended car-following trajectories, specifically those exceeding 15 seconds in duration. Noteworthy datasets containing shorter or partial trajectories, such as the Argoverse Motion Forecasting Dataset [2], [3], are excluded from our review. Table I presents a compilation of existing representative datasets.

A. OpenACC Dataset

The OpenACC Dataset [8], collated by the European Commission, offers an insightful perspective on the behavior of commercial Adaptive Cruise Control (ACC) systems under diverse driving conditions. This dataset encompasses data from 28 vehicles, including internal combustion engine (ICE) vehicles, electric vehicles, and hybrid vehicles, with 22 of these vehicles being outfitted with cutting-edge commercial ACC systems. To gather comprehensive data, the collection team embarked on four experimental campaigns, aimed at rigorously testing vehicle dynamics in real-world conditions, as well as the functionality of ACC systems and car-following patterns. In the OpenACC dataset, data from these four campaigns are methodically organized into corresponding folders: Cherasco, Vicolungo, AstaZero, and ZalaZone. Each of these folders contains several CSV files, with each file documenting the trajectory of a vehicle platoon. In total, the dataset encompasses over 3000 kilometers of trajectory data, captured at a resolution of 10 Hz.



Fig. 1. AVs used to collected data in CATS dataset.

B. Connected & Autonomous Transportation Systems (CATS) Laboratory Open Dataset

The Connected and Autonomous Transportation Systems (CATS) Laboratory at the University of Wisconsin-Madison has collected the ADAS-equipped vehicle platooning dataset [11]. The CATS ADAS-equipped vehicle platooning dataset documents commercial ADAS-equipped vehicle platooning behavior with different headway settings. Three vehicles participated in the experiments. The leading vehicle (i.e., ICE vehicle) operated with cruise control, following a pre-designed speed profile. The following two vehicles (shown in Figure 1) operated with ACC. The data collection occurred on a segment of State Road 56 in Florida between Bruce B. Downs Boulevard (Blvd.) and Gall Blvd. when traffic was light. The data collection team equipped each participating vehicle with a GPS device to record the real-time speed and location.

C. Vanderbilt ACC Dataset

Vanderbilt University collected Vanderbilt ACC data [7], which included car-following data of commercial ADAS-equipped vehicles with ACC enabled. In total, the data collection team recorded 1,200 miles of data. Participating vehicles were a lab-developed ADAS-equipped vehicle (hybrid vehicles) and commercial ADAS-equipped vehicles (ICE and hybrid vehicles) of two makes. The data collection team installed GPS devices on the vehicles to collect the vehicle data. This dataset includes two parts. The first part is the two-vehicle tests, in which the leading vehicle drove in a designed speed profile using cruise control, and the ACC vehicle followed. The data collection team used the same leading vehicle across different runs and tested different following vehicles. The second part is the eight-vehicle platoon tests. The lab ADAS-equipped vehicle served as the leading vehicle, following a precise speed profile. Seven other identical ACC vehicles followed.

D. Central Ohio Dataset

The Central Ohio dataset [9], [10] captures naturalistic ADAS-equipped vehicle behavior on roadways in central Ohio. Two types of vehicles were used: readily identifiable ADAS-equipped vehicles and discreet ADAS-equipped vehicles. Both vehicles are Level 2 ADAS-equipped vehicles. The Ohio project collected and processed trajectories for adjacent vehicles in the traffic stream; thus, both vehicles are equipped with cameras and LiDAR to capture adjacent vehicle movements. For the readily identifiable ADAS-equipped vehicles,

TABLE I
AUTONOMOUS DRIVING DATASETS WITH TRAJECTORY DATA

Name	Main features	Size	Format	Roadway type
Vanderbilt ACC Dataset [7]	time, speed, spacing, acceleration	15 minutes continuous trajectory data.	All data stored in one CSV file. Each row contains one frame data for two vehicles.	highway
OpenACC Dataset [8]	time, speed, position, spacing	More than 3000 km trajectory data with 28 vehicles.	Including seven folders, each folder contains several CSV files. Each row contains one frame data for multiple vehicles.	highway, closed track
Central Ohio Dataset [9], [10]	time, speed, position, spacing, acceleration, lane	120 hours trajectory data, with intotal 4169447 rows.	All data stored in one CSV file. Each row contains one frame data for multiple vehicles.	highway, freeway
CATS Laboratory Open Dataset [11]	time, speed, position	Two 25 minutes continuous trajectory data.	Each vehicle's data is stored in one XLSX file. Each sheet contain one trajectory.	highway
Waymo Open Dataset processed by Hu et al. [12]	time, speed, position, acceleration, jerk	470 AV-related car-following pairs with about 20 seconds each.	All data stored in one CSV file. Each row contains one frame data for two vehicles	urban

the radar and LiDAR sensors are visible on the vehicles. For the discreet ADAS-equipped vehicle, most sensors are not visible or were installed to be disguised (e.g., on top of the luggage stored on the top of the car). The team collected data from three generalized testing scenarios: vehicle following on highway and arterial roadways, lane change on highway and arterial roadways, and intersection approach and departure. These three scenarios were evaluated in different traffic densities and involved either a single ADAS-equipped vehicle or multiple ADAS-equipped vehicles. The two types of data are published in [9] and [10], respectively. For the vehicle-following and lane-change scenarios on highways and freeways, the Ohio data collection team used Ohio's 33 Smart Mobility Corridor and Interstate 270 (I-270). The Ohio data collection team also collected data on urban arterial roadways, including Riverside Drive and Morse Road.

E. Waymo Open Dataset

The Waymo Open Dataset offers high-resolution sensor data, amassed using ADAS-equipped vehicles across various environments, objects, and weather conditions [13]. This dataset includes information from diverse locations such as San Francisco, Phoenix, Mountain View, Los Angeles, Detroit, and Seattle. It comprises two parts: a perception dataset and a motion dataset. As of the latest update in August 2023, the perception dataset, with high-resolution sensor data and labels, consists of 2030 segments, while the motion dataset contains object trajectories and corresponding 3D maps for 103354 segments. Each segment encompasses about 200 frames, with a 0.1-second interval between consecutive frames. However, it's important to note that these data do not directly include the relationships of car-following pairs.

Hu et al. [12] processed 1000 segments of the original Waymo Open Dataset released in 2019 and manually extracted car-following pairs from it. In the processed dataset, all crucial information pertinent to the driving behavior of AVs, surrounding vehicles, and other road users has been integrated into a single file. Rather than extract the car-following pairs through automatic methods, the authors manually extracted car-following pairs based on video analysis to ensure the quality and reliability of the data.

III. METHODS

A. The Framework for Data Processing

In our methodology, we implement a three-step process to restructure the datasets into a standardized, user-friendly format, specifically tailored for traffic flow research, with a particular focus on car following modeling in this paper. The first step involves the extraction of several critical features necessary for calibration, followed by the conversion of all datasets into a unified CSV format encompassing these features. The second step is dedicated to data cleaning, which involves the meticulous removal of noise, outliers, and extraneous information from the raw data. This is vital to ensure that the datasets accurately represent high-quality, real-world behavior. The final step encompasses the calibration of a linear model using these refined datasets, allowing us to conduct a comparative analysis of the results.

B. Uniform Data Format

The first process is to extract valuable features and transfer the data into a uniform format using CSV files. We aim to use the trajectory data to calibrate the car-following models. In this study, we focus on the linear car-following model, where the variables available as inputs to the upper-level controller include only the spacing between vehicles, relative velocity, and the velocity of the ACC vehicle itself. The linear model for the following vehicle is shown in Equation (1).

$$\hat{a}(t) = f_s \times s(t) + f_v \times v(t) + f_{\Delta v} \times \Delta v(t) + z \quad (1)$$

where $\hat{a}(t)$ is the predicted acceleration at time t , $s(t)$ is the spacing between the leading vehicle and the following vehicle at time t , $\Delta v(t)$ is the speed difference between the leading vehicle and the following vehicle at time t , $v(t)$ is the speed of the following vehicle at time t , $f_s, f_v, f_{\Delta v}, z$ are the parameters. Therefore, we obtain the main labels that are related to the model calibration in Table II.

In our approach to standardizing the GPS-based data, which typically includes vehicle speed and coordinates, we address the absence of spacing and acceleration data in some datasets. To achieve uniformity in data processing, we calculate the spacing by determining the distance between

TABLE II
SELECTED FEATURES IN THE DATA.

Labels	Description
Trajectory_ID	The ID of one car-following trajectory.
Time_Index	The time of the corresponding row. (s)
ID_LV	The ID of the leading vehicle.
Speed_LV	The speed of the leading vehicle. (m/s)
ID_FAV	The ID of the following AV.
Speed_FAV	The speed of the following AV. (m/s)
Acc_FAV	The acceleration of the following AV. (m/s ²)
Space_Gap	The spacing between the leading vehicle and the following vehicle. (m)
Speed_Diff	The speed difference of the leading vehicle and the following vehicle. (m/s)

the leading and following vehicles, derived from their coordinates. For the computation of acceleration, we employ the first-order difference of vehicle speed. By standardizing spacing and acceleration calculations, the data becomes more reliably comparative, facilitating better analysis and interpretation. Finally, we consolidate all trajectory data from each dataset into a single CSV format file.

C. Data Cleaning

The second step is data cleaning. Before model calibration, we encountered various issues with files, data, and labels within the datasets. Consequently, it became imperative to undertake a thorough cleansing and pre-processing of the initial dataset. The data-cleaning process is specifically designed to ensure the integrity and reliability of the data we undertake a data cleaning and pre-processing regimen, delineated by the ensuing steps:

- Step 1: Linear Interpolation for Missing Data. To address gaps in location data, linear interpolation techniques were employed to generate reasonable approximations. The interpolation step fills in data voids, thus augmenting the dataset's comprehensiveness and reducing the likelihood of errors in analyses dependent on location data.
- Step 2: Outlier Removal. All data points residing outside ± 3 standard deviations from the mean were systematically eliminated to isolate abnormal data. The exclusion of outliers enhances the dataset's credibility, making it more amenable to statistical analysis.
- Step 3: Omission of Unstable Periods. Periods characterized by unstable car-following, specifically those recorded at the commencement or conclusion of test runs, were deliberately omitted. This exclusion preserves the quality and consistency of the dataset, thereby upholding its analytical reliability.

Through this comprehensive cleaning and pre-processing regimen, we substantially enhance the dataset's quality, aligning it more closely with the stringent requirements of automated vehicle technology research.

D. Car-following Model Calibration

In this research, our primary focus is on the linear car-following model as described in (1). The calibration

of this model can be effectively resolved using the least squares method. However, it's important to recognize that the simplistic form of the linear model may not accurately capture the complexities of car-following behavior. In future work, we intend to extend our analysis to include non-linear models, such as the Intelligent Driver Model (IDM) [14], and compare their calibration results. This approach will allow us to explore the efficacy of various modeling techniques in replicating the nuances of car-following dynamics more accurately.

IV. EXPERIMENTS

A. Experiment Condition Setting

In this study, we process three datasets: the AstaZero dataset in the OpenACC Dataset, the CATS Laboratory Open Dataset, and the Vanderbilt ACC Dataset. For each of these datasets, we implement the complete data processing method proposed in Section III and compare the statistical results of the processed data with the calibration data.

For these three datasets, we manually identified the range of unstable periods through observation. Notably, in the OpenACC dataset, when the vehicle speed approaches 0 (indicating the vehicle is still in the start-up phase and not yet engaging in car-following behavior), there is minor fluctuation in the spacing. To mitigate the impact of this data, we removed entries where the leading and following vehicle speeds were less than 0.1 m/s. Additionally, we excluded data where the acceleration was greater than +5 or less than -5 m/s².

In each dataset, we calibrate the AVs with all their related trajectories. In total, seven AVs were calibrated: an ACC-equipped SUV 2019 from the Vanderbilt ACC Dataset, a Lincoln MKZ 2016 and a Lincoln MKZ 2017 from the CATS Laboratory Open Dataset, and an Audi A6, a BMW X5, a Mercedes A-Class, and a Tesla Model 3 from the OpenACC Dataset. For the calibration of the OpenACC Dataset, a 2 seconds response time delay of the input variables is introduced as discussed in [8]. It is important to note that in the experimental section of this study, all distance data are in meters, speed data are in meters per second, and acceleration data are in meters per second squared.

B. Statistical Results

The statistical analysis and the distribution of the main features for the processed datasets are detailed in Table III and Figure 2, where the notations are the same with Section III and a represents the acceleration of the following vehicle. The results reveal that the speed difference and the following vehicle's acceleration across the three datasets are relatively consistent, predominantly ranging from -2 m/s to 2 m/s and -1 m/s² to 1 m/s², respectively, with mean values near 0. However, significant disparities are observed in terms of spacing and the following vehicle's speed. For instance, the OpenACC Dataset exhibits spacing evenly distributed between 15m and 30m, averaging 24.56m, and a following vehicle's speed primarily between 12 m/s to 20 m/s, averaging 18.49 m/s. This distribution suggests an environment

TABLE III
STATISTICAL ANALYSIS RESULTS OF THE THREE DATASETS.

Statistical measures	Vanderbilt ACC Dataset				CATS Laboratory Open Dataset				OpenACC Dataset			
	s	v	Δv	a	s	v	Δv	a	s	v	Δv	a
max	56.35	34.34	2.07	0.57	61.04	25.41	1.95	0.79	62.15	33.16	3.00	2.41
min	16.95	25.67	-2.52	-0.62	20.71	20.58	-1.96	-0.80	5.18	3.47	-3.00	-2.42
mean	35.85	29.23	0.02	0.01	42.17	23.23	-0.01	-0.01	24.56	18.49	0.02	0.01
std	6.19	1.59	0.78	0.19	10.41	0.83	0.70	0.24	9.53	4.17	0.84	0.49
#sample	8752				3577				401411			
#trajectory	1 trajectory for each vehicle				7 trajectories for each vehicle				8 trajectories for each vehicle			

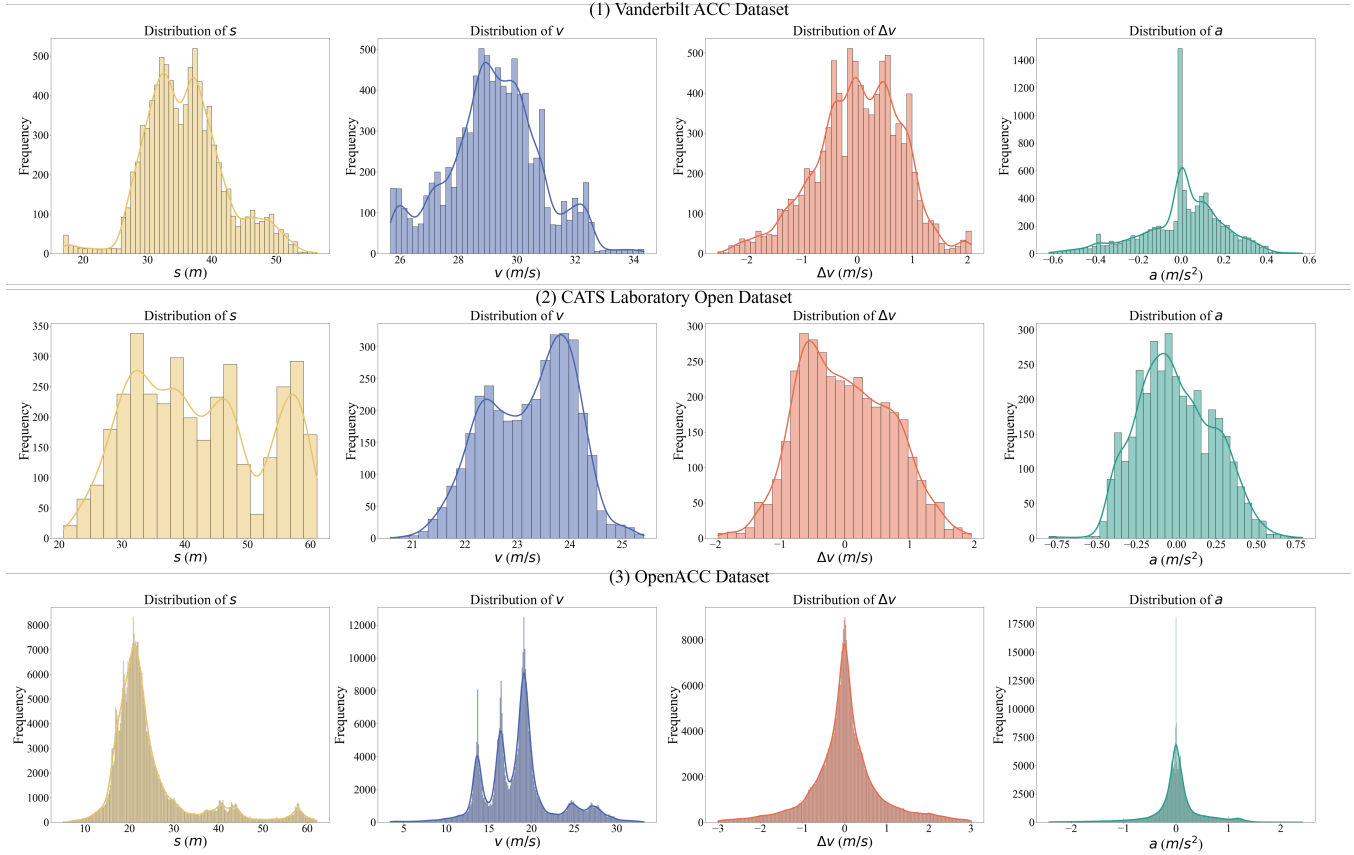


Fig. 2. Distributions of the main features for the three datasets.

characterized by relatively small spacing and lower speeds. Conversely, the Vanderbilt ACC Dataset predominantly features spacing between 26m and 50m, averaging 35.85 m, with the following vehicle's speed mostly in the 25 m/s to 32 m/s range, averaging 29.23m/s. The CATS Laboratory Open Dataset shows a uniform spacing distribution between 25 m and 60 m, averaging 42.17 m, and the following vehicle's speed concentrated between 21m/s and 25 m/s, averaging 23.23 m/s. Both the Vanderbilt ACC Dataset and CATS Laboratory Open Dataset indicate environments with moderate spacing and speed. These findings suggest that models calibrated from the OpenACC dataset are better suited for low-speed ranges, while those from the Vanderbilt ACC Dataset and CATS Laboratory Open Dataset are more applicable to medium-speed ranges. Moreover, in terms of dataset size, the OpenACC dataset, with over 400 thousand

samples, is substantially larger than the other two datasets, each comprising less than 10 thousand samples.

C. Calibration Results

The calibration results of the linear models are presented in Table IV. Here we use the R^2 to measure the accuracy of the linear model. From Table IV, it can be observed that the R^2 for the linear models of most vehicles range between 0.5 and 0.7, except the two CATS vehicles that are much higher than 0.7. This indicates that the linear models can reflect AV car-following behavior to a certain extent. However, due to the inherent limitations of the model structure and the complexity of the ACC systems, linear models struggle to perfectly simulate car-following behavior. Analyzing from the perspective of model parameters, parameters f_s and $f_{\Delta v}$ for all vehicles are positive. This implies that a larger

spacing and speed difference lead to a greater acceleration of the following vehicle. However, regarding parameter f_v , the values for the Vanderbilt SUV and OpenACC AVs are negative, whereas only the CATS AVs' parameter is positive. Moreover, the intercept term z in the CATS AVs model is very small, suggesting that the model predicts a negative acceleration value at the initial phase of the vehicles' start. Considering the range of the datasets' distribution, we infer that the linear model derived from the CATS dataset may only be valid in scenarios where the speed is predominantly above zero and is not applicable for low-speed conditions. Another possible reason for this discrepancy is that the CATS data was collected at a frequency of 1 Hz, which is a large time interval that may not accurately calibrate the model.

TABLE IV
LINEAR MODEL PARAMETERS FOR THE TEST VEHICLES.

Vehicle	R^2	f_s	f_v	$f_{\Delta v}$	z
Vanderbilt SUV	0.6561	0.0165	-0.0067	0.1532	-0.3921
Lincoln MKZ 2016	0.7422	0.0009	0.1733	0.3953	-4.0682
Lincoln MKZ 2017	0.8198	0.0012	0.1940	0.4022	-4.5568
Audi A6	0.6318	0.0038	-0.0010	0.4346	-0.0645
BMW X5	0.6318	0.0061	-0.0014	0.4838	-0.1214
Mercedes AClass	0.6556	0.0057	-0.0049	0.3910	-0.0447
Tesla Model3	0.7075	0.0036	-0.0019	0.5767	-0.0566

V. CONCLUSIONS

In this paper, we conducted an extensive review of five major trajectory datasets in the literature relevant to ADAS, focusing our analysis on three of them. Our study began with a detailed exposition of these datasets, encompassing their key features, data volume, formatting, and the environments in which the tests were conducted. We proposed a structured data processing methodology, which involves three critical steps: standardizing the data into a uniform format, data cleaning, and car-following model calibration. The experimental section of the paper delves into the statistical and calibration results. Our findings indicate that the CATS and Vanderbilt ACC datasets are more apt for models operating at medium speeds, whereas the OpenACC dataset is better suited for low-speed models. Future directions for this research include extending our data processing method to the remaining datasets. Additionally, we envisage undertaking a review and data processing of some trajectory datasets with short segments of less than 15 seconds that are designed for motion prediction.

REFERENCES

- [1] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.
- [2] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8748–8757.
- [3] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes *et al.*, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," *arXiv preprint arXiv:2301.00493*, 2023.

- [4] J. Mei, A. Z. Zhu, X. Yan, H. Yan, S. Qiao, L.-C. Chen, and H. Kretzschmar, "Waymo open dataset: Panoramic video panoptic segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 53–72.
- [5] M. Liu, E. Yurtsever, X. Zhou, J. Fossaert, Y. Cui, B. L. Zagar, and A. C. Knoll, "A survey on autonomous driving datasets: Data statistic, annotation, and outlook," *arXiv preprint arXiv:2401.01454*, 2024.
- [6] W. Liu, Q. Dong, P. Wang, G. Yang, L. Meng, Y. Song, Y. Shi, and Y. Xue, "A survey on autonomous driving datasets," in *2021 8th International Conference on Dependable Systems and Their Applications (DSA)*. IEEE, 2021, pp. 399–407.
- [7] Y. Wang, G. Gunter, M. Nice, and D. B. Work, "Estimating adaptive cruise control model parameters from on-board radar units," *arXiv preprint arXiv:1911.06454*, 2019.
- [8] M. Makridis, K. Mattas, A. Anesiadou, and B. Ciuffo, "Openacc. an open database of car-following experiments to study the properties of commercial acc systems," *Transportation research part C: emerging technologies*, vol. 125, p. 103047, 2021.
- [9] FHWA, "Advanced driver assistance system (adas)-equipped single-vehicle data for central ohio," 2023. [Online]. Available: https://data.transportation.gov/Automobiles/Advanced-Driver-Assistance-System-ADAS-Equipped-Si/iie8-uenj/about_data
- [10] —, "Advanced driver assistance system (adas)-equipped two-vehicle data for central ohio," 2023. [Online]. Available: https://data.transportation.gov/Automobiles/Advanced-Driver-Assistance-System-ADAS-Equipped-Tw/vhz2-exyi/about_data
- [11] X. Shi and X. Li, "Empirical study on car-following characteristics of commercial automated vehicles with different headway settings," *Transportation research part C: emerging technologies*, vol. 128, p. 103134, 2021.
- [12] X. Hu, Z. Zheng, D. Chen, X. Zhang, and J. Sun, "Processing, assessing, and enhancing the waymo autonomous vehicle open dataset for driving behavior research," *Transportation Research Part C: Emerging Technologies*, vol. 134, p. 103490, 2022.
- [13] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.
- [14] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.