Comparative Analysis of Synthetic Inertial Data Quality from CARLA and SUMO Driving Simulators

Author1[1], Author2[2], Author3[1]

[1] Address1

[2]Address2

email@email.com, email@email.com, email@email.com

Abstract. *CARLA and SUMO are widely adopted open-source simulators in autonomous driving and intelligent transportation systems. While neither was initially designed to generate high-fidelity inertial data such as that from an Inertial Measurement Unit (IMU), both have been employed in contexts where such data is critical. In this study, we evaluate the quality of inertial data produced by CARLA and SUMO through direct comparison with the naturalistic UAH-DriveSet dataset. Our analysis reveals that SUMO performs better in qualitative and quantitative assessments. To our knowledge, this is the first systematic comparison of its kind, highlighting limitations in current simulation tools and offering insights for autonomy-related studies using simulated inertial data.*

1. Introduction

The area of Driver Behavior Recognition and Autonomous Vehicles is gaining increasing attention as researchers become more interested in driving assistance, road safety, and autonomous decision making. Understanding events in real-time using inertial data, and even predicting them, is useful to help an AI-based system make better decisions, alert drivers in case of danger, assess how likely a driver is to cause accidents, and more. However, this area lacks a standard baseline: available datasets are scarce and often not public, data collection methods vary widely across studies, and the observed results cannot be easily compared.

To address these issues, some researchers have turned to synthetic inertial data from simulators as a cost-effective and safer alternative. While this approach enables larger datasets and controlled experimentation, the quality of the generated data remains largely unexamined, despite being critical for ensuring model performance in real-world applications.

Considering the need to assess the realism and limitations of inertial data from simulators as alternative data sources, this work focuses on evaluating the quality of such data in two of the most widely used platforms in autonomous driving and driver behavior recognition research: Simulation of Urban MObility (Eclipse SUMO) [Alvarez Lopez et al. 2018] and CARLA [Dosovitskiy et al. 2017]. To this end, we conduct a comprehensive comparison between these simulators and the UAH-DriveSet [Romera et al. 2016]. This well-known public dataset includes real inertial data from various driving behaviors along clearly defined routes.

To ensure a fair and consistent evaluation, the route used in the UAH-DriveSet was replicated in both simulators using OpenStreetMap data. CARLA's built-in profiles for

normal and aggressive driving were matched with equivalent configurations in SUMO. Inertial data collected from each source using the same type of virtual sensor is then compared using distribution similarity metrics to assess how closely the simulated data mirrors real-world conditions.

This paper presents the first systematic comparison of inertial data generated by CARLA and SUMO with real-world data from the UAH-DriveSet. The main contributions are: (1) a quantitative and qualitative evaluation of simulated versus real inertial signals; (2) an analysis of the impact of driving behavior parameters on the synthesized data; and (3) a reproducible framework for fair comparison across simulators using matched routes and behavior profiles. These findings highlight key limitations in current simulators and provide guidance for their use in research relying on inertial data.

2. Related Work

Developed in the early 2000s by the German Aerospace Centre, SUMO (Simulation of Urban MObility) was originally created to simulate traffic at the level of individual vehicles, and has been adopted by researchers worldwide to study mobility patterns across city and highway networks [Krajzewicz et al. 2002]. Years later, in 2017, CARLA (Car Learning to Act) was introduced as an open-source platform focused on autonomous driving [Dosovitskiy et al. 2017]. Built on a game engine, CARLA offers controlled urban environments and sensor configurations that support the testing of both traditional and learning-based driving systems. In Table 1, we compare characteristics of both simulators in their current versions, highlighting relevant aspects of driving simulation scenarios.

Table 1. Characteristics of SUMO and CARLA simulators

Characteristic	SUMO	CARLA
Simulation core	2D with no physics-engine	3D with Unreal Engine physics-engine
Allows for loaded maps	Yes	Yes
Allows for route definition	Yes	Yes
Allows adjusting vehicle behavior parameters	Via car-following and lane-changing models	Via agent scripting
Simulated noisy data	Not natively	Adjustable noise
Implements car-following moels	Yes, built-in models	Limited via traffic manager basic controls
IMU data available	Acceleration and angle	Acceleration, gyroscope and angle
Camera and LIDAR data available	No	Yes

Table 1 shows that although SUMO and CARLA were not specifically designed to simulate data that is typically captured in the real-world by Inertial Measurement Units (IMUs) through sensors such as accelerometers and gyroscopes, both simulators provide this information using different strategies. While SUMO lacks a physics engine and simulates sensor readings from programmed trajectories in 2D scenarios, CARLA enhances user expectations by incorporating an Unreal Engine®, which can simulate realistic physical interactions in tridimensional environments, including accelerations, rotations, and collisions. Given the inherent challenges of conducting real-world tests, it is unsurprising that researchers in autonomous driving adopt CARLA and SUMO to develop and validate their control algorithms.

As noted by Sarker and colleagues, simulators are commonly used to enhance the insights gained from ground-truth data by generating synthetic scenarios that would be impractical or dangerous to capture in real life [Sarker et al. 2025]. This highlights the increasing recognition that synthetic datasets can facilitate controlled, scalable, and

repeatable testing, thereby addressing significant gaps left by traditional data collection methods. For example, Chah et al. and Han et al. independently developed simulation-based datasets to tackle the risks, high costs, and limitations associated with capturing rare or hazardous scenarios in real-world environments [Chah et al. 2024, Han et al. 2024]. Chah et al. used the SUMO simulator to create a diverse database of driving behaviors based on actual urban layouts. Meanwhile, Han et al. developed the CARLA-Loc dataset using the CARLA simulator to assess SLAM algorithms under various and challenging conditions.

In some cases, simulators are part of the Software-in-the-Loop (SIL) test strategy, enabling early software validation through tests conducted on a controllable and repeatable virtual environment. For example, in [Stević et al. 2019], the authors developed ADAS (Advanced Driver-Assistance Systems) applications by employing ROS (Robotic Operating System) as a prototyping platform. They used sensing data provided by the CARLA simulator and integrated it with an autonomous platform, allowing a complete feedback loop, where commands from the autonomous platform resulted in simulated actions within CARLA, and simulated sensor data was fed back to the software. Similarly, in [Gómez-Huélamo et al. 2021], the CARLA simulator was adopted to validate a fully-autonomous driving architecture based on ROS, explicitly focusing on the behavioral decision-making layer. Specific challenging driving scenarios, such as *Stop*, *Pedestrian Crossing*, *Adaptative Cruise Control* (ACC), and *Unexpected Pedestrian*, were simulated to evaluate the architecture's performance.

Beyond the use of simulated data for validating control algorithms, synthetic datasets are also being used to directly train machine learning models as a means to address privacy and security concerns, but also as a self-supervised strategy to label data that would be hard to annotate in real-world data [Sarker et al. 2025]. For example, [Brahim et al. 2022] trains machine learning models, such as XGBoost and LSTMs, with simulated data from CARLA to classify driver behavior, systematically collecting diverse scenarios, including different road types, traffic situations, and weather conditions, while controlling environmental parameters like speed limits and weather. Similarly, [Gutiérrez-Moreno et al. 2022] leveraged CARLA to train their Deep Reinforcement Learning (DRL) agent for autonomous driving. Their work focused on tackling the complexity of autonomous driving at intersections, which are considered challenging scenarios due to factors like the uncertainty in surrounding vehicle behaviors and the variety of intersection types.

These studies demonstrate the versatility of simulation environments and the increasing use of synthetic data to address real-world challenges. However, none have fully validated their results against actual inertial data, raising concerns about the reliability and applicability of their findings. Our work highlights the importance of quality and realism in simulated inertial data for meaningful autonomy-related outcomes. By comparing outputs from CARLA and SUMO with a naturalistic dataset, we aim to assess the fidelity of these simulation tools, providing informed perspectives to guide future methodological choices in simulation-based research.

3. Method

In short, our method consists of reproducing a real-world driving trajectory from the UAH-DriveSet dataset – specifically, the *normal2-secondary* route – within both CARLA and SUMO simulators to enable a fair and controlled comparison between real versus simulated inertial data. This was achieved by converting OpenStreetMap (.osm) data into the respective input formats required by each simulator: *.xodr* for CARLA and *.net* for SUMO. As a baseline for natural variability, we also compared inertial data from two different drivers performing the same route in the real dataset.

Some simplifications and assumptions were made to ensure the simulations are as close as possible:

- Although the real data was collected from trips with different vehicles present, we assumed that no other vehicles were present in the simulation. Including other vehicles would require evaluating car-following models, which are not directly implemented in CARLA.
- We translated each available CARLA agent parameter to SUMO vehicles in order to perform equivalent simulations.
- CARLA provides adjustable parameters for Gaussian noise, and SUMO doesn't. Even if it could be done through post-processing of the SUMO data, we avoided including noise for direct comparison of what the simulators offer.

In the following sections, a description of the real dataset used and the procedures to set each of the two simulators are provided.

3.1. UAH-driveset

UAH-driveset is a dataset collected from the monitoring app DriveSafe. The application was run by six different drivers and vehicles, performing 3 different behaviors (normal, drowsy and aggressive) on two types of roads (motorway and secondary road), resulting in more than 500 minutes of naturalistic driving with its associated raw data and processed semantic information, together with the video recordings of the trips. The smartphone was placed at the windshield of the vehicle [Romera et al. 2016].

Each driver repeats pre-designated routes by simulating a series of different behaviors: normal, drowsy and aggressive driving. When driving normally, the tester is only told to drive as he usually does. When driving drowsy, the driver is told to simulate slight sleepiness, which normally results in sporadic unawareness of the road scene. When driving aggressively, the driver is told to push to the limit his aggressiveness (without putting the vehicle at risk), which normally results in impatience and brusqueness while driving. For simplicity, in this study we only consider normal and aggressive driving and we randomly selected samples from two drivers (driver 1 and driver 2) to analyze the divergence between them driving normally and aggressively. The variation in driving behaviors led to different durations for completing the same route. The selected samples yielded time series of 259 seconds (normal) and 200 seconds (aggressive) for driver 1, and 262 seconds (normal) and 237 seconds (aggressive) for driver 2.

Although the dataset includes both motorway and secondary routes, we focused solely on the secondary route due to its closer alignment with typical simulation use

cases—lower speeds and simpler lane structures. Additionally, limitations in CARLA's stability led us to use only a subsection of this route.

The relevant sensors provided in the dataset for this study are GPS and accelerometer. The GPS data is sampled at 1 Hz and includes speed, latitude, longitude, altitude, vertical and horizontal accuracy, course, and course variation. Accelerometer data was originally recorded at 100 Hz but is provided in the dataset at 10 Hz, as the dataset authors downsampled it by averaging every ten samples. This data includes both acceleration and gyroscope measurements along the three axes.

3.2. CARLA simulator

CARLA runs in a 3D environment with multiple sensors (camera, radar, IMU) and realistic physics and rendering. Creating true 3D maps of real-world locations requires manual reconstruction, so we converted OpenStreetMap (.osm) files to the OpenDRIVE (.xodr) format in order to reproduce the UAH-driveset routes. This comes at the cost of losing 3D detail (OSM is purely 2D) and limiting 3D sensor performance, though the remaining sensors still provided valuable data. This is a clear limitation of our work, but as will be seen in Section 4, the remaining sensors still provide valuable information.

To reproduce the route and driving behaviors from the dataset, we used CARLA's Agent navigation mode, which allows setting customized routes and changing driver behaviors, some of which are already predefined in the simulator. The parameter values for each behavior are shown in Table 2.

Table 2. CARLA agent parameters for normal and aggressive drivers. The parameter _max_speed_ is not default

Parameter	Normal	Aggressive	Description
max_speed	113	132	Maximum speed in km/h
speed_lim_dist	20	15	Target speed offset from the speed limit in km/h
speed_decrease	10	8	Speed reduction when approaching a slower vehicle in km/h
safety_time	3	3	Time-to-collision threshold in seconds
min_proximity_threshold	10	8	Min distance before avoiding other objects in meters
braking_distance	5	4	Distance to initiate emergency stop in meters
tailgate_counter	0	-1	Time before resuming tailgating in seconds

In Table 2, the only values that differ from the default values provided by CARLA are _max_speed_ and _speed_lim_dist_, which were changed to account for the maximum and mean speeds from the respective behavior in the UAH-driveset. For a 113 km/h maximum speed and a 93 km/h mean speed from the real dataset, for example, _max_speed_ is set to 113 and _speed_lim_dist_ to 20, the difference between said maximum and the mean. All other parameters are kept as defaults, since no background vehicles were present.

Simulations ran in synchronous mode, and we applied the same data preprocessing steps as the UAH-drive dataset: data were generated at 100Hz, followed by a downsampling to 10Hz. The sensor was also positioned at the windshield of the vehicle, yielding 407 seconds of normal driving data and 325 seconds of aggressive driving data. All recordings are available on _URL removed for blind review_.

3.3. SUMO simulator

SUMO doesn't support 3D sensors or cameras, providing only total acceleration and vehicle angle (rather than separate x and y components). To get the individual components of the accelerometer, a decomposition based on the angle was applied. Gyroscope data isn't directly available, but can be inferred from angle changes over time.

SUMO vehicles rely on base vehicle specifications plus car-following and lane-changing models, which can substantially alter traffic behavior, as shown in [Schrader et al. 2023]. In our single-vehicle, few-lane-change setup, the impact of these models is minimal, avoiding divergences, since they are not implemented by default in CARLA. The key parameters (Table 3) were taken from CARLA when available or left at SUMO's defaults to normal drivers, then slightly adjusted to simulate aggressive behavior.

Table 3. SUMO vehicle parameters for normal and aggressive drivers. Parameters translated from CARLA are highlighted

Parameter	Normal	Aggressive	Description
maxSpeed	31.4	36.7	Maximum speed in m/s (CARLA)
Accel	2.5	3.0	Maximum acceleration in m/s² (default)
Decel	4.5	5.0	Maximum deceleration in m/s² (default)
speedFactor	0.8	0.9	Speed multiplier relative to speed limit (CARLA)
emergencyDecel	9	10	Emergency braking acceleration in m/s² (default)

Similarly, the simulation is performed for both aggressive and normal behavior. The acceleration and position data are collected via TraCI, a Python interface from SUMO, keeping frequencies in agreement with the real dataset (data generation at 100 Hz followed by downsampling to 10 Hz). SUMO does not allow sensor positioning. The simulation resulted in 504 seconds of normal driving behavior and 454 seconds of aggressive driving behavior. The simulations are available at *URL removed for blind review*.

4. Experiments and Results

We compared SUMO vs. CARLA, SUMO vs. UAH, CARLA vs. UAH, and UAH driver 1 vs. driver 2. Figure 1 plots each trajectory. Despite following the same route, total times differ due to varied acceleration and deceleration patterns.
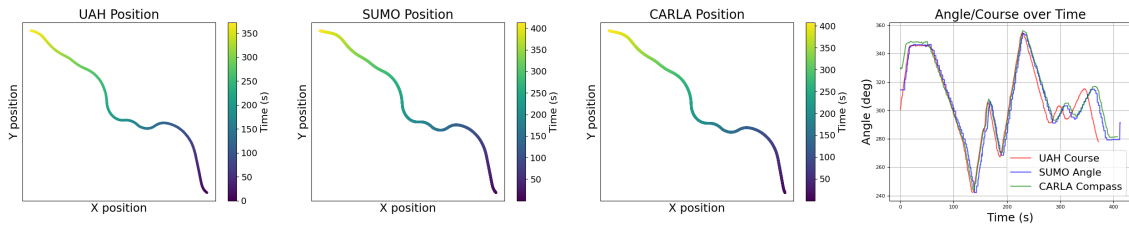

Figure 1. Position over time for UAH, SUMO and CARLA

Figure 2 shows that, although two different real drivers from the same behavior produce different sensor readings, their overall distributions are very similar. In order to keep comparisons clear, we selected a single driver to be our real-world benchmark for the comparisons that follow.
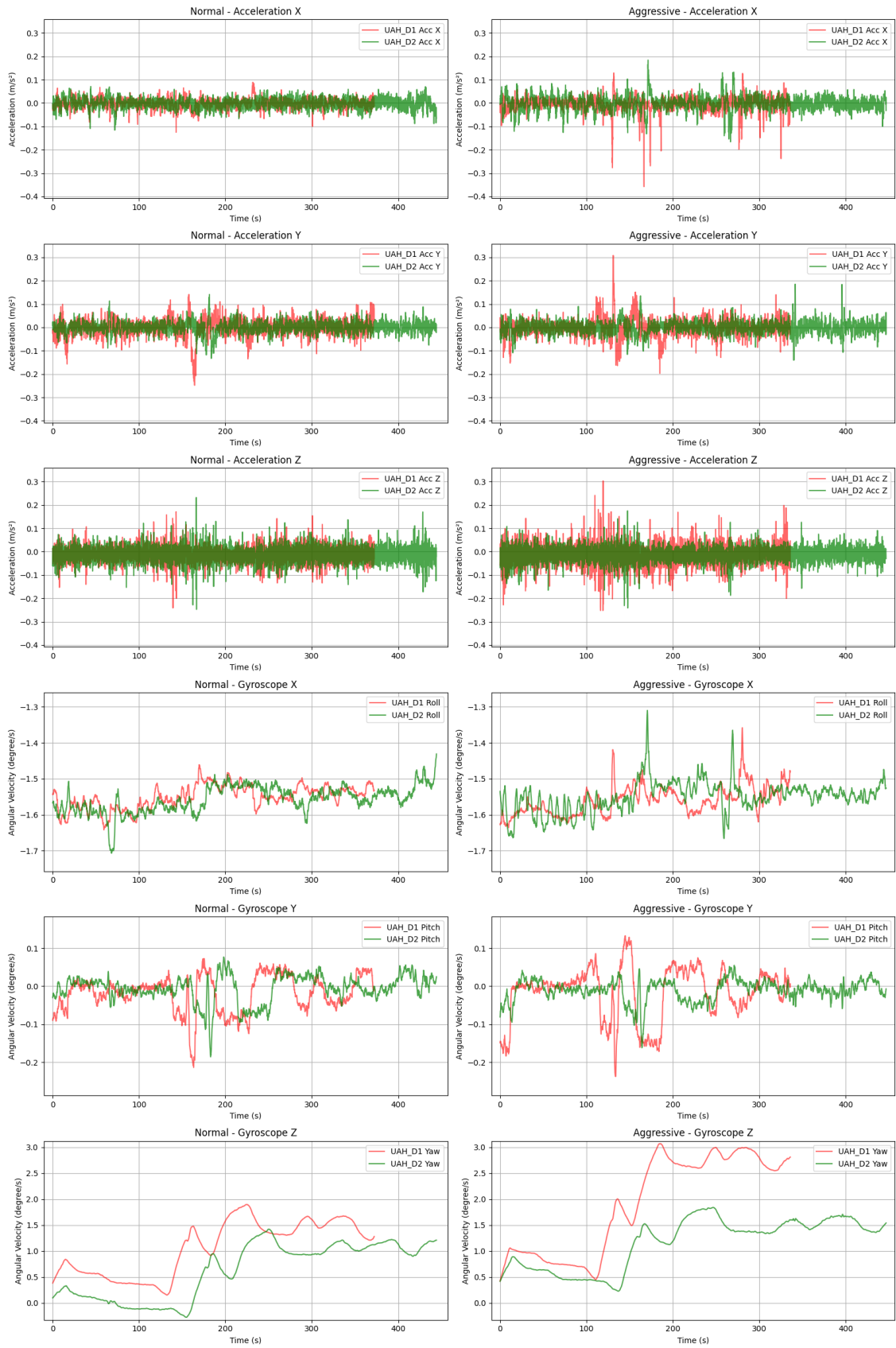
Figure 2. Real UAH data, normal driving behavior (left) versus aggressive driving behavior (right). We observe that while the route followed by both drivers is the same, their trip times are different. Also, aggressive driving behavior shows clearly higher acceleration patterns.

Figure 3 shows the comparison of the raw sensor data between CARLA and UAH for normal and aggressive behaviors, respectively. Gyroscope readings on the x and y axes, as well as acceleration on the z axis, are not shown in the graphs due to the limited vertical realism of the simulation, which results from using 2D OpenStreetMap data to generate the 3D environment in CARLA. Figure 4 contrasts SUMO with UAH for the same behaviors. From them, we can see that CARLA shows a worse alignment than SUMO for both accelerometer axes, especially because of the huge intensity variations on the y axis. Gyroscope data from the simulators have similar patterns between them, but also display very high frequency variations in intensity that are not present in the real data, making them easily distinguishable. Such behavior may be related to the control of the vehicles in the simulation and the constant discrete adjustments needed, which causes a jittery car behavior that is transmitted to the sensors collected.
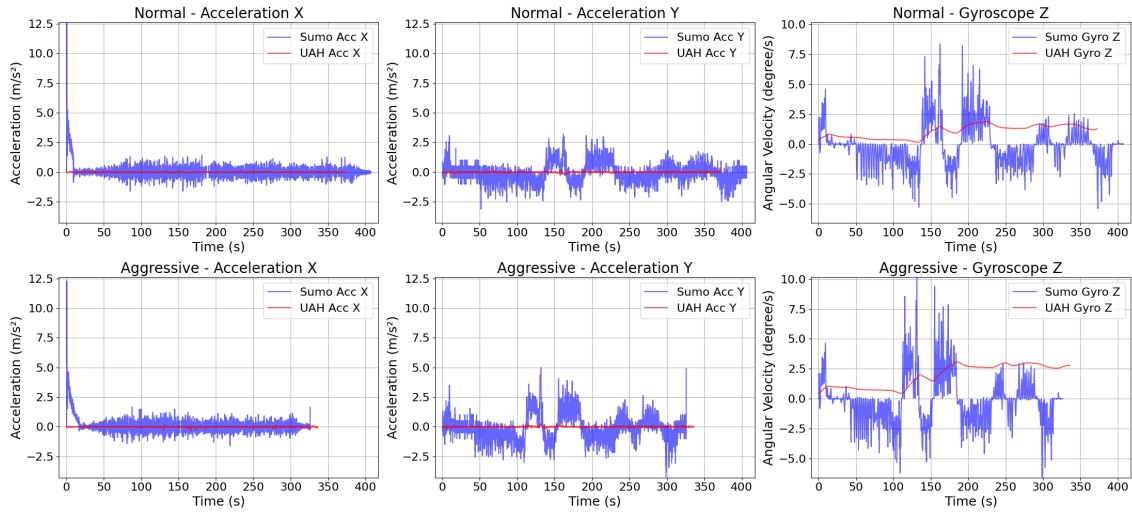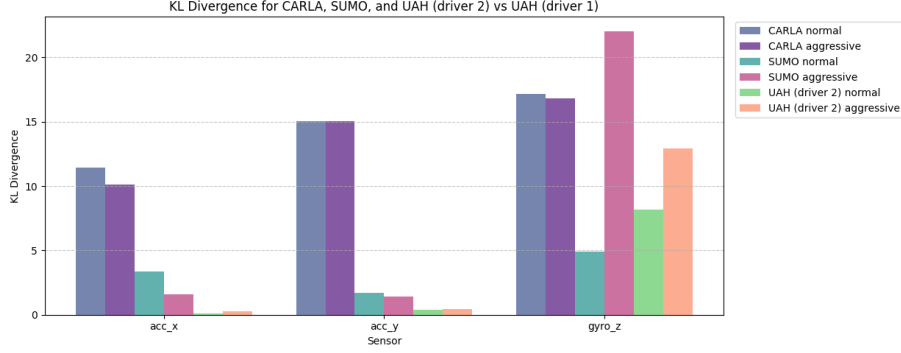

Figure 3. Comparison of CARLA (blue) and UAH (red) data for normal (top) and aggressive (bottom) drivers.
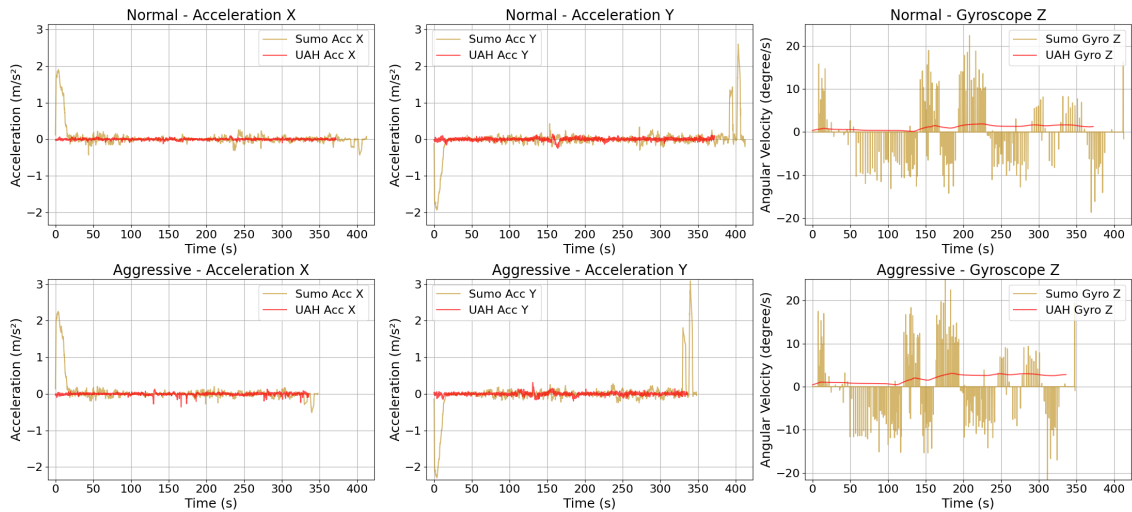
Figure 4. Comparison of SUMO (yellow) and UAH (red) data for normal (top) and aggressive (bottom) drivers.

Figure 5. KL Divergence of CARLA and SUMO sensor readings against UAH. Acceleration on the z axis is captured only by CARLA, and the gyroscope on the x and y axes is not considered for the simulators.

In order to quantitatively compare the simulated and real data distributions, we used the Kullback-Leibler divergence metric [Kullback and Leibler 1951]. This metric measures how different two probability distributions are, offering a sense of similarity between them. For instance, a low divergence score on the x axis of the accelerometer indicates that the distribution of simulated values closely resembles that of the real data, regardless of time alignment, thus reflecting how realistic the simulated magnitudes are. Each axis is treated independently, as they are fixed relative to the vehicle's frame and consistently oriented (e.g., the x axis always points forward), which helps preserve distributional comparability across experiments.

Figure 5 presents the KL divergence of SUMO and CARLA against the UAH-driveset, alongside the divergence between the two UAH drivers. Divergence between UAH drivers is minimal for both acceleration axes and the gyroscope, as expected, and will be used as a baseline. SUMO's accelerometer data consistently yields lower divergence than CARLA's, although none of them have as low divergence as the UAH drivers. Again, gyroscope x and y readings are omitted (not used in 2D simulations), and the z axis of the gyroscope exhibits the highest divergence, even between real drivers, which is likely due to slight sensor offsets (as can be seen in Figs. 2), but it is visually clear that neither simulator was able to capture realistic z-axis patterns (Figs. 3 and 4).

To further evaluate fidelity, we applied UMAP projections (Figs. 6, 7) and Fourier transforms to the accelerometer's x and y axes for both normal and aggressive behaviors, excluding the mostly unreliable gyro z data. The UMAP embeddings on the time domain can easily separate real from synthetic CARLA simulator data, while SUMO yields tighter clusters, as expected from its lower KL divergence. In the frequency domain, SUMO's clusters become more mixed, whereas CARLA's remain clearly separated. From them, we further conclude that accelerometer data from SUMO is more reliable, although further inspection is needed to determine its utility in real-world scenarios.

5. Conclusion and Future Work

The experiments demonstrate that neither SUMO nor CARLA produces fully realistic inertial sensor data, although SUMO consistently outperforms CARLA in both qualitative assessments and quantitative KL-divergence metrics. We hypothesize that SUMO's stable, model-driven vehicle control results in more consistent and realistic acceleration
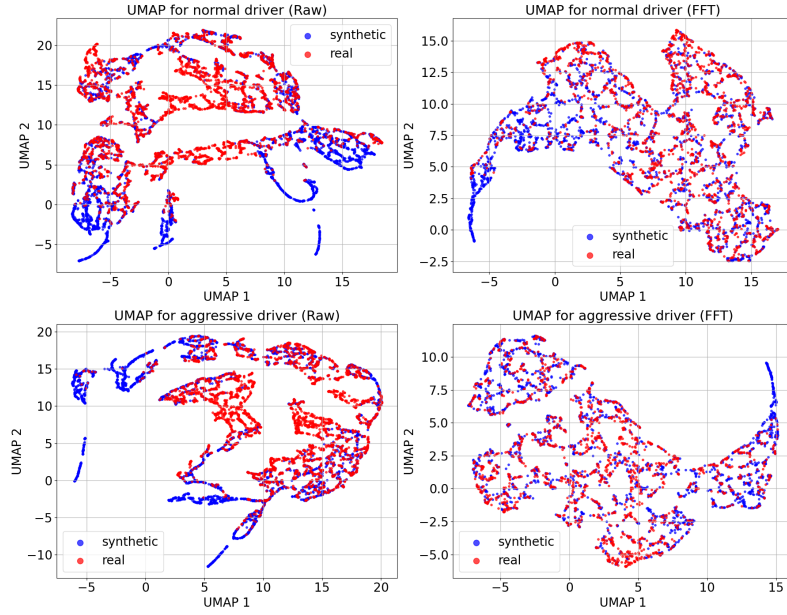
Figure 6. UMAP projection of SUMO synthetic data and UAH real data, considering only the x (longitudinal) and y (lateral) acceleration components. We observe that data simulated on SUMO covers real data distribution.
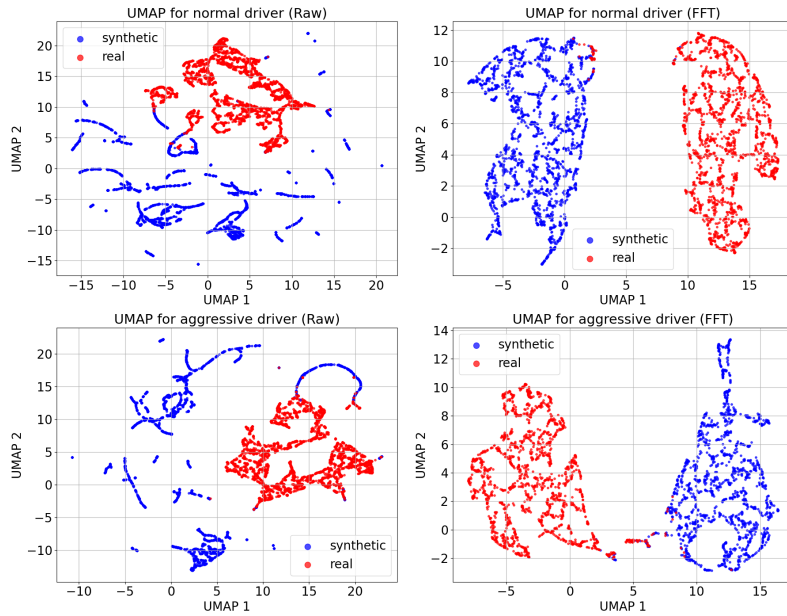
Figure 7. UMAP projection of CARLA synthetic data and UAH real data, considering only the x (longitudinal) and y (lateral) acceleration components. The well-separated clusters indicate that the real and simulated data distributions are easily distinguishable, suggesting that the synthetic data from CARLA does not closely resemble the real-world inertial patterns.

patterns. In contrast, CARLA—despite its rich 3D environment and support for a wide range of sensors—may lack the fine-tuned control stability required to generate smooth inertial data, particularly in its default agent configurations. On the other hand, SUMO's inherently 2D structure fundamentally limits its ability to replicate complex spatial scenarios or support the broader range of sensing modalities required for advanced autonomous driving research. These findings underscore the need for deeper investigation into the synthesis of realistic inertial data within driving simulators. Models trained on synthetic data from the simulators (as in [Gutiérrez-Moreno et al. 2022, Brahim et al. 2022]) may fail to generalize in practice. They also show that efforts to enhance sensor realism in SUMO and CARLA (as done in [Chah et al. 2024, Han et al. 2024]) are valuable, but if CARLA's core control logic is the actual problem, post-processing alone will hardly be able to solve it. More broadly, the results demonstrate the persistent challenge of generating useful, realistic driving data.

Future work should address the so-called *sim2real* gap by improving inertial data synthesis. Training models on synthetic IMU data and evaluating them against real-world benchmarks can help quantify the impact of simulation fidelity on generalization. Although generative approaches such as GANs and diffusion models have not yet achieved consistent success in capturing realistic inertial patterns, they remain a promising avenue for future research. Additionally, leveraging each simulator's strengths—CARLA in detailed 3D environments and SUMO in complex traffic scenarios—may further improve the quality of synthetic data and support more robust autonomy systems.

6. Acknowledgements

References

Alvarez Lopez, P., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wießner, E. (2018). Microscopic Traffic Simulation using SUMO. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 2575–2582. IEEE.

Brahim, S. B., Ghazzai, H., Besbes, H., and Massoud, Y. (2022). A Machine Learning Smartphone-based Sensing for Driver Behavior Classification. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 610–614.

Chah, B., Lombard, A., Mualla, Y., Bkakria, A., Abbas-Turki, A., and Yaich, R. (2024). Building a Database of Simulated Driver Behaviors Using the SUMO Simulator. In Arai, K., editor, *Intelligent Systems and Applications*, pages 536–555, Cham. Springer Nature Switzerland.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16.

Gómez-Huélamo, C., Del Egido, J., Bergasa, L. M., Barea, R., López-Guillén, E., Arango, F., Araluce, J., and López, J. (2021). Train Here, Drive There: Simulating Real-World

Use Cases with Fully-Autonomous Driving Architecture in CARLA Simulator. In Bergasa, L. M., Ocaña, M., Barea, R., López-Guillén, E., and Revenga, P., editors, *Advances in Physical Agents II*, pages 44–59, Cham. Springer International Publishing.

Gutiérrez-Moreno, R., Barea, R., López-Guillén, E., Araluce, J., and Bergasa, L. M. (2022). Reinforcement Learning-Based Autonomous Driving at Intersections in CARLA Simulator. *Sensors*, 22(21).

Han, Y., Liu, Z., Sun, S., Li, D., Sun, J., Yuan, C., and Jr, M. H. A. (2024). CARLA-Loc: Synthetic SLAM Dataset with Full-stack Sensor Setup in Challenging Weather and Dynamic Environments.

Krajzewicz, D., Hertkorn, G., Rössel, C., and Wagner, P. (2002). SUMO (Simulation of Urban MObility)-an open-source traffic simulation. In *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002)*, pages 183–187.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Romera, E., Bergasa, L., and Arroyo, R. (2016). Need Data for Driver Behaviour Analysis? Presenting the Public UAH-DriveSet.

Sarker, S., Maples, B., Islam, I., Fan, M., Papadopoulos, C., and Li, W. (2025). A Comprehensive Review on Traffic Datasets and Simulators for Autonomous Vehicles.

Schrader, M., Al Abdraboh, M., and Bittle, J. (2023). Comparing Measured Driver Behavior Distributions to Results from Car-Following Models using SUMO and Real-World Vehicle Trajectories from Radar: SUMO Default vs. Radar-Measured CF model Parameters. *SUMO Conference Proceedings*, 4:41–54.

Stević, S., Krunić, M., Dragojević, M., and Kaprocki, N. (2019). Development and Validation of ADAS Perception Application in ROS Environment Integrated with CARLA Simulator. In *2019 27th Telecommunications Forum (TELFOR)*, pages 1–4.