UNIABUJA Journal of Engineering and Technology

https://ujet.uniabuja.edu.ng/

ISSN: 2714-3236 (Online); 2714-3228 (Print)

Volume 3, Issue 2, 2026; 347-364

Adaptive Physics-Informed Trajectory Reconstruction for Inferring Driver Aggressiveness (Apitra-Da) From Noisy Vehicle Kinematic Data

Eni S. LAUGHTER[1]*, Abdulhakeem H. NURUDEEN[2]

[1]*College of Transportation Engineering, Chang'an University, China*
[2]*Department of Mechanical Engineering, University of Abuja, Nigeria*

[1*]enisolomonlaughter@gmail.com, [2]hassan.abdulhakeem@uniabuja.edu.ng

Abstract
Road traffic crashes claim over 1.3 million lives annually, with human factors responsible for more than 90% of incidents. Despite the growing availability of vehicle trajectory data enabling microscopic analysis of driving behavior, methodological challenges persist due to severe noise and missing vehicle specifications. This study introduces APITRA-DA, a two-stage adaptive physics-informed framework designed to reconstruct noisy trajectories and infer driver aggressiveness without requiring vehicle specifications or behavioral labels. Applied to two expressway datasets where 97.6% and 59.9% of trajectories exhibited quality issues, the framework achieved 99% physics consistency, transforming degraded data into behaviorally interpretable trajectories. Noise reduction employed Butterworth filtering to preserve behavioral signals, while reconstruction integrated machine learning-based vehicle capability estimation (R^2 > 0.78), physics-based constraints, and driver-adaptive Kalman filtering converging within three iterations. A novel aggressiveness parameter (a) quantifies how intensively drivers exploit vehicle capabilities, normalized for cross-vehicle comparison. Validation demonstrated significant correlations with following distance (ρ = -0.45) and speed variance (ρ = +0.43), confirming a as a genuine behavioral metric. Results reveal drivers typically operate at ~60% of vehicle capacity, reserving unused capability for emergencies. Aggressiveness emerges as a dynamic, context-dependent strategy—higher during acceleration than braking, decreasing with speed, and elevated in free-flow conditions. Following distance and speed variance patterns provide key behavioral signatures, with aggressive driver proportions varying from 0.2% to 9.9% across datasets. APITRA-DA enables applications in insurance telematics, traffic safety, emissions modeling, autonomous vehicle development, and simulation, advancing Vision Zero strategies by unlocking behavioral analytics from naturalistic trajectory data.

Keywords: *Driver Aggressiveness, Trajectory Reconstruction, Physics-Informed Filtering, Naturalistic Driving Data, Machine Learning, Traffic Safety.*

1.0 Introduction

Road traffic crashes represent a profound global public health challenge, resulting in more than 1.3 million fatalities each year and constituting the foremost cause of mortality among individuals aged 5 to 29. In the context of international efforts to advance Vision Zero initiatives and the development of intelligent transportation systems, the analysis of driver behavior has emerged as a pivotal domain of intervention. This emphasis is underscored by the fact that human factors contribute to over 90% of crash causation, positioning behavioral assessment and mitigation strategies as central to achieving substantive reductions in traffic-related morbidity and mortality. (Skaug et al., 2025). The increasing availability of vehicle trajectory data sourced from instrumented vehicles, aerial surveillance, and traffic cameras combined with advances in analytical methodologies, has created new opportunities for large-scale, microscopic examination of driving behaviors. However, the development of next-generation traffic safety systems remains constrained by a fundamental methodological challenge: the difficulty of deriving valid behavioral metrics from real-world trajectory data that is both highly noisy and devoid of vehicle-specific information. This study introduces a framework designed to overcome these limitations, enabling scalable behavioral risk assessment that is critical for real-time crash prediction, targeted safety interventions, and the effective management of intelligent transportation systems. By addressing the dual challenges of noise reduction and behavioral signal preservation, the framework provides the methodological foundation necessary to support Vision Zero initiatives and the broader pursuit of sustainable transportation development.

This convergence of behavioral analytics with advanced traffic management technologies reflects a paradigm shift in road safety research, where human factors are no longer treated as residual variables but as central determinants of system performance. Collectively, these developments highlight the critical role of driver behavior analysis in achieving Vision Zero objectives and advancing sustainable mobility strategies. (Skaug et al., 2025), (Joo et al., 2023).

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026

https://doi.org/10.70118/ujet.2026.0302.30

Vehicle trajectory data has become increasingly available through diverse collection methods (Ali et al., 2022) successfully utilized SHRP2 NDS time-series data collected from six states between 2010-2013, achieving 76-89% accuracy in detecting environmental conditions through data fusion of video images, real-time CANbus data, and external sensors. (Kar et al., 2023) extracted trajectory data from UAV-collected traffic videos for crash risk modeling, while (Liu & Zhang, 2022) employed the highD dataset of naturalistic vehicle trajectories collected on German highways. The high-resolution pNEUMA dataset, collected in central Athens, has been leveraged to facilitate large-scale, real-time prediction of traffic conflicts. Such datasets exemplify the growing diversity of trajectory sources, ranging from instrumented vehicles to aerial surveillance and traffic camera systems (Li et al., 2023). Together, these data streams enable unprecedented microscopic analysis of driving behaviors, offering critical insights for advancing both road safety and environmental sustainability. By capturing fine-grained behavioral dynamics across heterogeneous traffic environments, these resources provide the empirical foundation for developing predictive models, designing targeted safety interventions, and informing sustainable mobility strategies (Skaug et al., 2025).

Despite the transformative expansion in trajectory data availability, a critical methodological barrier continues to impede large-scale behavioral analytics. Trajectory datasets are invariably affected by severe measurement noise, which must be disentangled from authentic behavioral signals to ensure valid interpretation. Preliminary analyses indicate that approximately 97% of trajectory segments exhibit physical inconsistencies, with acceleration profiles violating fundamental kinematic principles. This pervasive noise, compounded by the absence of vehicle specifications in most naturalistic datasets and the difficulty of distinguishing measurement artifacts from genuine aggressive driving behaviors, presents a formidable challenge for existing analytical approaches. Current methodologies fail to address these issues simultaneously: physics-informed filters risk suppressing legitimate behavioral expressions; threshold-based classification presupposes access to vehicle-specific parameters that are typically unavailable; and machine learning techniques depend on labeled training data that cannot be reliably obtained in real-world contexts. Collectively, these limitations underscore the need for novel frameworks capable of reconciling noise reduction with behavioral signal preservation, thereby enabling robust and scalable driver behavior analytics (Kuşkapan et al., 2021). This methodological gap prevents large-scale behavioral analytics essential for real-time crash risk prediction and targeted safety interventions.

This study introduces an integrated framework capable of extracting valid measures of driver aggressiveness from severely noisy, specification-free trajectory data, thereby addressing a critical gap in behavioral risk assessment for intelligent transportation safety systems. By resolving the methodological challenge of differentiating genuine aggressive driving behavior from measurement artifacts without reliance on vehicle specifications or ground truth labels, the framework unlocks the analytical potential of large-scale naturalistic trajectory datasets that have previously been unsuitable for safety applications. The proposed methodology provides direct support for advanced traffic management systems requiring real-time behavioral risk profiling, facilitates targeted interventions for high-risk driver populations, and establishes a robust measurement foundation for evaluating automated vehicle safety systems operating alongside human drivers with varying levels of aggressiveness. In doing so, the framework advances the operationalization of behavioral analytics essential for Vision Zero strategies and the development of sustainable, human-centered intelligent transportation systems.

2.0 Literature Review

Vehicle trajectory data is inherently subject to measurement noise and errors that compromise the validity of subsequent analyses. To address these challenges, prior research has pioneered systematic approaches through the development of multi-step "traffic-informed" methodologies. These frameworks typically involve the removal of extreme positional errors, the application of digital filtering techniques, localized reconstruction grounded in vehicle kinematics and platoon consistency, and the elimination of residual noise. Collectively, such procedures have proven effective in mitigating data degradation while preserving the physical plausibility of reconstructed trajectories, thereby enabling more reliable behavioral and safety analyses (Montanino & Punzo, 2015). Similarly, (Mahajan et al., 2023) advanced anomaly detection in the pNEUMA drone dataset by integrating Savitzky–Golay filtering with the XGBoost algorithm, achieving a substantial reduction in jerk ranges from approximately ±5000 to ±45 m/s^3. In related work, (Fard et al., 2017) introduced a wavelet-based methodology that detects outliers through local neighbor comparison and employs discrete wavelet transform decomposition to refine trajectory signals. Complementing these approaches (Dong et al., 2021) combined Empirical Mode Decomposition with Butterworth filtering to separate trajectories into intrinsic mode functions, thereby enhancing signal interpretability while mitigating measurement noise. Collectively, these studies underscore the diversity of signal processing techniques applied to trajectory data, each contributing distinct strengths in noise reduction, anomaly detection, and behavioral signal preservation.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

Recent advances in trajectory reconstruction have increasingly incorporated physical constraints to ensure kinematic plausibility (Makridis & Kouvelas, 2023) introduced an adaptive physics-informed framework that integrates vehicle dynamic constraints, driver compliance, and adaptive noise reduction with automatic detection of smoothing strength. Their iterative approach identifies optimal filtering magnitudes while maintaining compatibility with feasible vehicle dynamics, yielding significant reductions in speed error and ensuring device independence (Wang et al., 2024) formulated trajectory data association as a min-cost network circulation problem and rectification as a quadratic program, explicitly leveraging vehicle dynamics and physical constraints (Riehl et al., 2025)demonstrated that orientation-informed reconstruction enhances internal consistency by 15% and platoon consistency by 20%, thereby improving the representation of car-following and traffic dynamics. Similarly, (Chen et al., 2023) developed an adaptive Kalman filter that alternates between the standard and Unscented Kalman Filter to accommodate diverse driving behaviors, reducing jerk from ±4960 to ±47 m/s^3 and achieving a 92.9% F1-score in behavior detection.

Although these physics-informed approaches excel in producing physically consistent trajectories, they emphasize reconstruction accuracy at the expense of preserving behavioral signatures. Constraints designed to enforce physical plausibility may inadvertently suppress genuine aggressive behaviors that approach—but do not surpass—vehicle capability limits. More critically, existing methods apply uniform filtering strength across all drivers, overlooking the behavioral heterogeneity identified by (Fard et al., 2017) who emphasized that driver variability cannot be adequately captured through homogeneous filtering strategies.

Characterising driver aggressiveness from trajectory data has attracted significant research attention with increasingly sophisticated approaches. (Chen et al., 2025) propose a knowledge-integrated framework using lateral-longitudinal acceleration in g-g-v diagrams, integrating domain-specific prior knowledge with Empirical Cumulative Distribution Function (ECDF) to quantify aggressiveness levels through fuzzy clustering that probabilistically associates samples with specific manoeuvres. Validation with 90 drivers revealed that aggressive driving decomposes into distinct, interpretable maneuver patterns. (Xie et al., 2020) applied K-Means clustering to NGSIM data, classifying drivers into four clusters based on acceleration and time headway along "Cautious/Aggressive" and "Sensitive/Insensitive" dimensions. (Yuan et al., 2025) used Gaussian Mixture Model (GMM) clustering on drone-collected trajectory data, identifying three driving styles: aggressive (23.5%), moderate (58.9%), and conservative (17.6%) based on six features incorporating speed and acceleration metrics. (Chowdhury et al., 2015) demonstrated that kurtosis of acceleration profiles stores major information about driving styles with less influence from anomalous events.

3.0 Materials and Methods

3.1 Research Framework

This study utilizes two real-world vehicle trajectory datasets collected from expressway merge and diverge segments, designated as SQM-W-1 and SQM-W-2 from Nanjing China (Feng et al., 2025). The datasets were captured using a drone with high-resolution temporal sampling at 24 Hz, providing one observation every 0.042 seconds. This high temporal resolution delivers the detailed microscopic trajectory information necessary for driver behavior analysis. Significantly higher than conventional trajectory datasets such as NGSIM (10Hz). Each observation includes the following variables in Table 1.

SQM-W-1 contains 822,712 observations from 1,041 vehicles with a mean observation time of 61 seconds per vehicle, while SQM-W-2 comprises 988,030 observations from 1,420 vehicles with a mean observation time of 73 seconds. Both datasets represent expressway road segments under different traffic conditions—SQM-W-1 exhibits congested conditions while SQM-W-2 demonstrates mixed flow patterns. This diversity in traffic conditions enables validation of the proposed methodology's robustness across varying operational environments.

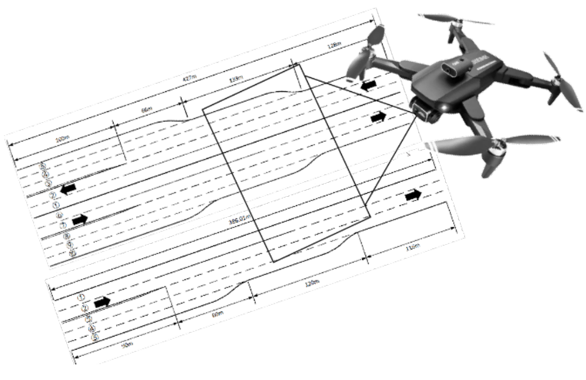

Figure 1: Ubiquitous Traffic Eye SQM Expressway Structure Chart

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

Table 1: Trajectory Dataset Variables

Category	Variables	Description
Temporal Information	Time(s)	Timestamp of observation
Spatial Information	Longitudinal Distance and Lateral Distance (m)	Vechile Position Coordinates
Kinematic Variables	Speed (m/s), speed (km/h), and acceleration (m/s2)	Instantaneous Velocity and Acceleration
Vehicle interactions	Leader id, follower id, leader distance (m) and follower distance (m)	Following relationships and spacing
Vehicle Characteristics	Vechile length (pixel) and vechile Width (pixel)	Normalized vechile dimensions

This section outlines the formatting rules to ensure consistency and professionalism in all submitted manuscripts. Authors must strictly adhere to these formatting standards when preparing their papers for submission.

3.2 Data Quality Assessment

The data quality assessment ensures that trajectory data is cleaned with obvious corruptions removed, problematic segments are flagged, sensor noise is smoothed while behavioral variation is retained, and genuine driving patterns are preserved throughout the preprocessing pipeline (Montanino & Punzo, 2015).

The first stage systematically removes obviously corrupt data while deliberately preserving potentially aggressive behavior that may initially appear anomalous but represents genuine driver actions (Mahajan et al., 2023).

3.3 Physical Impossibility Detection

Jerk analysis forms the primary detection mechanism, examining whether the rate of acceleration change remains physically possible for ground vehicles (Montanino & Punzo, 2015). Jerk is calculated as the discrete derivative of acceleration for each consecutive measurement pair in the trajectory sequence. Equation 1.

$$\text{jerk} = \frac{\Delta\text{acceleration}}{\Delta\text{time}} |\text{jerk}| > 15 \text{ m/s}^3 \quad \text{indicates mechanically impossible dynamics or sensor errors.}(equ1) \tag{1}$$

Speed consistency verification examines whether reported velocities align with positional changes over time. In which a negative speed values indicates backward motion on expressway mainlines are flagged as measurement artifacts, as are sudden position discontinuities suggesting instantaneous teleportation beyond plausible vehicle displacement within the trajectory dataset. The position-derived speed is computed and compared against the recorded speed (m/s) to identify discrepancies. Equation 2.

$$\text{implied\_speed} = \frac{|\Delta\text{position}|}{\Delta\text{time}} \text{implied\_speed} > 3 \times \text{reported\_speed} \quad \text{indicates a position jump.}(equ2) \tag{2}$$

Acceleration limit screening identifies extreme values exceeding 10 m/s² in absolute magnitude, which surpasses the capabilities of both high-performance sports vehicles and heavy commercial vehicles under normal tyre adhesion conditions.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

3.4 Platoon Consistency Verification

Platoon-level consistency represents a crucial validation dimension, as inter-vehicle spacing errors indicate tracking failures or identification mistakes that corrupt individual trajectory quality even when kinematic variables appear internally consistent .According to (Montanino & Punzo, 2015), minimum spacing thresholds identify physically impossible configurations where spacing falls below 2 meters, indicating that vehicles would occupy overlapping spatial positions. Beyond static spacing checks, dynamic consistency verification examines whether spacing changes align with relative velocities between leader-follower pairs. Equation 3:

$$\text{spacing\_change} = (v_{\text{leader}} - v_{\text{follower}}) \times \Delta t |\text{spacing\_change}| > 20\text{ m in } \Delta t < 1\text{ s} \quad \text{indicates inconsistency.}(equ3) \tag{3}$$

3.4.1 Internal Consistency Analysis

Internal consistency evaluation verifies that the three primary kinematic variables position, speed, and acceleration, are critical quality indicator because violations reveal that reported measurements cannot simultaneously represent a single physical trajectory, indicating that sensor fusion algorithms or data processing pipelines have introduced systematic errors (Montanino & Punzo, 2015).

To check for validation, the position-velocity relationship derived, requires that instantaneous speed equals the rate of positional change in equation (4).

$$\text{speed\_calculated} = \frac{\Delta\text{position}}{\Delta\text{time}} \frac{|\text{speed\_calculated} - \text{speed\_reported}|}{\text{speed\_reported} + 0.01} > 0.2 \quad (20\%\text{ tolerance}).(equ4) \tag{4}$$

Speed is independently calculated from the longitudinal Distance(m) column within the trajectory dataset using discrete differentiation over the 0.1-second sampling interval, then compared with the directly reported speed(m/s) variable. Similarly, the laws of physics dictate that the velocity-acceleration relationship requires instantaneous acceleration to equal the rate of velocity change Equation 5:

$$\text{acceleration\_calculated} = \frac{\Delta\text{speed}}{\Delta\text{time}} \frac{|\text{acceleration\_calculated} - \text{acceleration\_reported}|}{|\text{acceleration\_reported}| + 0.01} > 0.3 \quad (30\%\text{ tolerance}).(equ5) \tag{5}$$

3.4.2 Initial Noise Reduction

The final assessment stage of data quality assessment applies preliminary smoothing to remove high-frequency sensor noise without destroying behavioral signals that manifest at lower frequencies. A Butterworth low-pass filter with cutoff frequency fc = 0.75 Hz processes the acceleration time series, following (Montanino & Punzo, 2015)'s demonstration that this frequency threshold effectively separates rapid oscillations from GPS jitter and sensor quantization errors while preserving actual acceleration changes initiated by driver control inputs.

$$\boldsymbol{H(s)} = \frac{\mathbf{1}}{\sqrt{\mathbf{1} + \left(\frac{f}{f_c}\right)^{2\times\text{order}}}} \tag{6}$$

$$Where: f = frequencyofsignalcomponent, f_c = cutofffrequency(1.0Hz), \text{order} = filtersteepness(4).\ (equ6)$$

3.5 Adaptive-Physics Informed Behavioural Reconstruction

Eliminates the residual noise that survived initial quality checks and inferring driver aggressiveness from the cleaned data. The approach integrates physics-based constraints with behavioral inference through an adaptive framework that recognizes a fundamental insight where the same trajectory pattern may represent sensor noise for one driver but genuine behavior for another (Makridis & Kouvelas, 2023).

3.6 Estimating Vehicle Capabilities Without Specifications

For each vehicle, we calculate acceleration as the discrete time derivative of filtered speed. We then group observations into speed ranges (0-10, 10-20, 20-30 km/h, and so on) to capture how vehicle acceleration capability degrades with increasing velocity due to aerodynamic drag and powertrain characteristics (Makridis & Kouvelas, 2023) Within each speed bin, we extract the 95th percentile acceleration rather than the absolute maximum to avoid outlier contamination.

$$\alpha_{\text{observed\_max}}(v) = \beta \times \alpha_{\text{true\_max}}(v)(equ8) \tag{8}$$

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026

https://doi.org/10.70118/ujet.2026.0302.30

Or equivalently

$$\alpha_{\text{true\_max}}(v) = \beta \times \alpha_{\text{observed\_max}}(v)(equ9) \tag{9}$$

Where the observed maximum acceleration for each speed bin is calculated as:

$$a_{\text{observed\_max}}(v) = P_{95}\left(|\text{acceleration\_reconstructed}|\right) \text{ for speed bin } v(equ10) \tag{10}$$

This introduces a critical estimation problem. We therefore use β to represent the capacity usage factor—the fraction of true vehicle capability that drivers typically employ. The relationship is expressed as:

and β represents the capacity usage factor with typical ranges:

β ≈ 0.5–0.6: Conservative drivers (use 50–60% of capacity)

β ≈ 0.6–0.75: Normal drivers (use 60–75% of capacity)

β ≈ 0.75–0.9: Aggressive drivers (use 75–90% of capacity)

Conservative drivers with β around 0.5-0.6 never approach their vehicles' limits, while aggressive drivers with β around 0.75-0.9 regularly operate near maximum capability. Estimating β individually for each vehicle allows us to infer what vehicles could physically achieve from what was actually observed, giving us answers about true physical capabilities.

To estimate β, we employ supervised machine learning, transforming the specification problem into a driver behavioral classification task. Conservative drivers manifest through low observed maximum accelerations (95th percentile below 2.0 m/s²), minimal acceleration variance indicating smooth driving control, and large following distances exceeding 30 meters.

Normal drivers demonstrate moderate maximums (2.0-3.5 m/s²), intermediate variance, and following distances of 15-30 meters. Aggressive drivers produce high maximums (above 3.5 m/s²), elevated variance from dynamic throttle and brake usage, and short following distances under 15 meters. While (Xie et al., 2020) classifies drivers using acceleration and time headway features, their method does not relate observed behaviour to the vechiles physical acceleration limits. In contrast, out β-framework explictly quantifies how closely each driver operates to their vehicle's feasible capacity envelope.

3.7 Data Validation

Random Forest Regression method with XGBoost were adopted as the primary method for validation. Neural networks were deliberately avoided due to their high susceptibility to overfitting, especially in limited or noisy datasets. (Salman & Liu, 2019) show that deep neural networks possess extremely large parameter spaces, enabling them to memorize noise and even perfectly fit randomized labels, making overfitting a fundamental limitation of such models.

We employ two additional methods for validation and extended analysis. A percentile-based method computes α as the 90th percentile of raw capacity-usage ratios, consistent with prior work showing that distribution-tail statistics effectively reveal unusual driving patterns (Chowdhury et al., 2015). A speed-dependent analysis further estimates aggressiveness separately for urban (0–30 km/h), arterial (30–60 km/h), and highway (60+ km/h) regimes, reflecting evidence that driver behavior adapts to operational context and traffic conditions (Xie et al., 2020); (Chen et al., 2025).

3.8 Applying Physics-Based Constraints

We apply physics-based constraints to eliminate residual extreme outliers. For each observation, the current speed determines the applicable capability limit by indexing into the vehicle's speed-dependent envelope. We compare reconstructed acceleration against this limit. Positive accelerations exceeding the maximum are constrained to the physically achievable value at that speed. Negative accelerations exceeding -8.0 m/s² in magnitude are constrained to this threshold, distinguishing emergency braking from sensor artifacts. This approach aligns with physics-informed reconstruction frameworks that enforce feasible vehicle dynamics to suppress unrealistic acceleration peaks (Makridis & Kouvelas, 2023).

$$\alpha_{\text{reconstructed}} = \begin{cases} \alpha_{\text{true\_max}}(v_{\text{current}}), & \text{if } \alpha_{\text{reconstructed}} > \alpha_{\text{true\_max}}(v_{\text{current}}) \\ -8.0\ \mathrm{m/s^2}, & \text{if } \alpha_{\text{reconstructed}} < -8.0\ \mathrm{m/s^2} \qquad (equ11) \\ \alpha_{\text{reconstructed}}, & \text{otherwise} \end{cases} \tag{11}$$

The constraint must accommodate uncharacteristically aggressive maneuvers reaching 3.83 m/s² while rejecting implausible values beyond this physical limit. This vehicle-specific approach eliminated 170 impossible accelerations from SQM-W-1 and 4,907 from SQM-W-2, removing speed-dependent outliers that might appear plausible at low speeds but violate physics at high speeds. The result is a physically consistent acceleration time series where every value respects the vehicle's speed-dependent capability envelope.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026

https://doi.org/10.70118/ujet.2026.0302.30

3.8 Inferring Driver Aggressiveness

We estimate each driver's aggressiveness parameter α, which measures the moment-to-moment intensity with which they exploit available vehicle capability. In contrast, β represents the driver's long-term capacity-usage style across their entire driving history. This distinction is consistent with prior work showing that instantaneous aggressiveness can be inferred from high-resolution acceleration and speed patterns (Chen et al., 2025), while long-term behavioral tendencies emerge from stable differences in acceleration magnitude and headway across drivers (Xie et al., 2020). The physical interpretation of α and β is further supported by evidence that drivers typically operate well below the vehicle's feasible acceleration envelope (Makridis & Kouvelas, 2023)..

The aggressiveness parameter α is defined as the ratio of typical acceleration usage to maximum vehicle capability. This yields dimensionless values where α = 0.3-0.5 characterizes conservative drivers, α = 0.5-0.7 indicates normal drivers, and α = 0.7-1.0 identifies aggressive drivers. This normalized metric enables comparison across heterogeneous vehicle types by measuring behavioral intensity relative to vehicle-specific limits rather than using absolute acceleration thresholds that confound vehicle capability with driver choice (Chen et al., 2025).

Our primary estimation follows an Empirical Cumulative Distribution Function (ECDF) approach, selected for its robustness to noisy data. For each vehicle, we calculate capacity usage ratios for every observation expressed as:

$$\text{ratio} = \frac{|\alpha_{\text{reconstructed}}|}{\alpha_{\text{true\_max}}(v_{\text{current}})}(equ12) \qquad (13)$$

These ratios typically span 0.0 to 1.0, representing the fraction of available capability employed at each instant. We construct the empirical cumulative distribution function by sorting all ratios and computing what percentage of time the driver operated at each capacity intensity level, consistent with aggressiveness-quantification methods that rely on distribution-based behavioral metrics (Chen et al., 2025). We extract α as the 90th percentile of this distribution, meaning that 90% of driving time involved capacity usage at or below α while the top 10% is excluded to remove emergency maneuvers and residual outliers, following the percentile-based interpretation of aggressive driving intensity described in (Chen et al., 2025).

$$\alpha = P_{90}(\text{capacity\_usage\_ratios})(equ13) \qquad (13)$$

Where:
$\alpha < 0.5$: Conservative driver
$0.5 \leq \alpha < 0.7$: Normal driver
$\alpha \geq 0.7$: Aggressive driver

Traffic context is incorporated by separately analyzing free-flow conditions where leader distance exceeds 50 meters and constrained conditions with spacing under 20 meters. Vehicle type adjustments account for the natural tendency of larger, heavier vehicles to exhibit lower acceleration magnitudes even under aggressive operation,consistent with physics-informed approaches shwoing that feasible acceleration depends jointly on vehicle dynamics and surrounding traffic spacing (Makridis & Kouvelas, 2023).

3.9 Adaptive Filtering with Behavioral Awareness

The filtering implementation employs an Adaptive Kalman Filter framework with two operational modes. The Kalman filter is a recursive state estimator that predicts the next kinematic state based on physics models then corrects this prediction using new observations, optimally balancing model predictions with measurement information according to their relative uncertainties (Kalman & Others, 1960).

Mode 1 utilizes standard Kalman Filtering for car-following and normal driving conditions, assuming smooth, predictable motion.

Mode 2 employs Unscented Kalman Filtering for complex maneuvers including lane changes and aggressive events.

Quality data assessment flags integrate into Kalman filtering through adaptive measurement trust weighting. Observations flagged as impossible jerk receive low measurement weight (0.1), indicating minimal influence on state updates. Observations with platoon inconsistency flags require context-dependent interpretation: for aggressive drivers these might indicate genuinely close following receiving moderate weight (0.5), while for conservative drivers they likely represent GPS errors receiving low weight (0.2). These weights parametrize the Kalman measurement noise covariance matrix, consistent with the Kalman filter formulation in which measurement uncertainty governs the correction step (Kalman, 1960).

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

3.10 Multi-Dimensional Classification and Validation

The converged α values undergo multi-dimensional classification for comprehensive assessment. Overall aggressiveness classification uses the final α value: below 0.5 indicates Conservative drivers, 0.5 to 0.7 identifies Normal drivers, and 0.7 and above characterizes Aggressive drivers. Expected population distributions suggest approximately 20-30% conservative drivers, 50-60% normal drivers, and 15-20% aggressive drivers, consistent with prior work showing that driving aggressiveness can be represented as a continuous behavioral intensity measure and grouped into distinct categories (Chen et al., 2025).

Validation without ground truth behavioral labels employs multiple convergent evidence streams. Correlation analysis tests expected relationships between α and observable trajectory features. Aggressive drivers should maintain shorter following distances, producing negative correlation between α and mean leader distance with expected Pearson correlation around -0.5 to -0.7. Aggressive drivers should exhibit higher speed variance, yielding positive correlation around 0.4 to 0.6. Jerk pattern analysis should show aggressive drivers producing substantially more high jerk events than normal drivers, who in turn exceed conservative drivers. This multi-feature validation approach aligns with prior work demonstrating that aggressive driving manifests through acceleration variability, sharper maneuvers, and distinct trajectory signatures (Chen et al., 2025).

3.11 Modelling Assumptions

A few key assumptions regarding acceleration, deceleration, and driver aggressiveness were made. We assume that real driving produces acceleration and deceleration patterns that remain within physically plausible limits, such that values above roughly 10 m/s² or jerk exceeding about 15 m/s³ cannot originate from genuine driver input and therefore represent noise. We also assume that aggressive behavior will appear in the data, but only as strong yet still plausible acceleration or braking—typically within the 3–8 m/s² range—and that these behaviors occur at low frequencies because human pedal inputs evolve more slowly than high-frequency sensor noise. Deceleration is assumed to be inherently less aggressive than acceleration due to driver discomfort and risk perception, particularly at high speeds or short headways. The framework further assumes that aggressiveness cannot be inferred from raw acceleration alone, since vehicles differ in performance; instead, aggressiveness reflects the proportion of a vehicle's capability that a driver chooses to use. Finally, we assume that genuine aggressive behavior is structured, physically consistent, and context-dependent, whereas noise is unstructured and physically implausible, enabling the model to distinguish between the two during reconstruction.

4.0 Results and Discussion

The results obtained after the quality data assessment based on Butterworth low-pass filter was configured with a cutoff frequency of 0.75 Hz and filter order of 4, optimized for the 24 Hz sampling rate. This configuration removes high-frequency sensor noise above 1 Hz while preserving driver behavioral patterns typically occurring in the 0.1-1 Hz range. Quality thresholds were established based on mechanical feasibility limits: jerk threshold of ±15 m/s³, maximum acceleration of ±10 m/s², maximum comfortable deceleration of -8 m/s², and minimum physical spacing of 2 meters. These thresholds balance the need to flag implausible measurements while avoiding excessive false positives that would unnecessarily discard genuine aggressive driving events.

For behaviour reconstruction, vehicle capability estimation, eight speed bins were defined: 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, and 70-200 km/h. The 95th percentile was selected for extracting observed maximum acceleration to exclude extreme outliers while capturing realistic high-acceleration events that represent actual vehicle usage. For aggressiveness parameter (α) calculation, the 90th percentile was chosen to represent typical capacity usage while excluding rare emergency maneuvers that would not characterize normal driving behavior. The ML-based β estimation employed Random Forest Regression with 100 estimators and maximum depth of 10 as the primary model, with XGBoost (100 estimators, maximum depth of 6, learning rate of 0.1) providing comparison validation. Training utilized a 70-15-15 split for training, validation, and testing sets respectively. The adaptive Kalman filter iteration was configured with a maximum of 10 iterations and convergence thresholds of 1% for variance change and 0.05 for α change, both sustained over 3 consecutive iterations to ensure solution stability.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

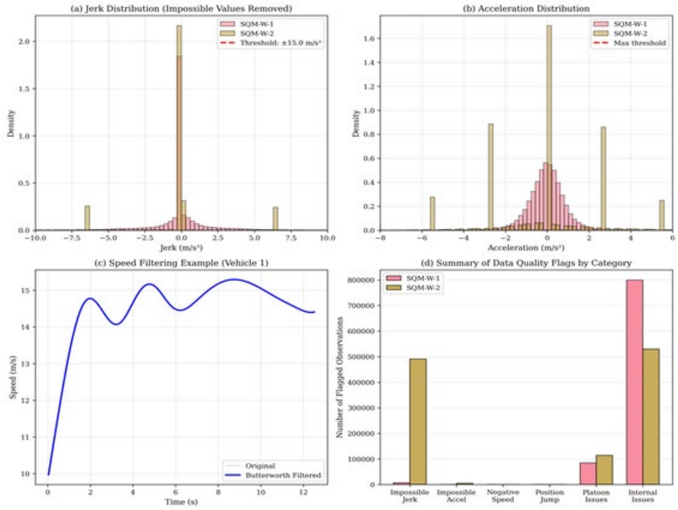

Figure 2(a-d)**:** Stage 1 Data Quality Visualization Caption: Visualization of data quality issues in both datasets. (a) Jerk distribution showing impossible values beyond ±15 m/s³ threshold. (b) Acceleration distribution with impossible values flagged. (c) Example of speed filtering showing original vs. Butterworth-filtered trajectories. (d) Summary of quality flags by category.

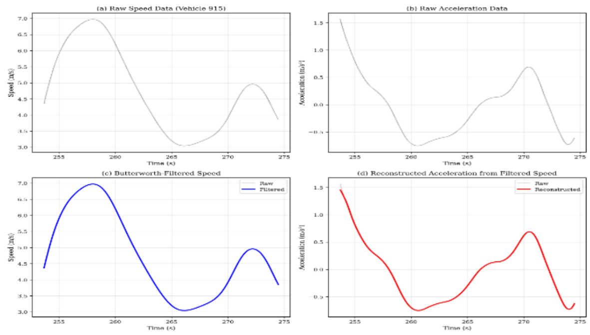

Figure 3(a-d): Comparison of Speed Profiles Before and After Stage 1 Processing Caption: (a-b) Raw speed and acceleration data showing high-frequency noise. (c-d) Butterworth-filtered speed and reconstructed acceleration demonstrating noise reduction while preserving behavioral patterns. Example from Vehicle ID 45 in SQM-W-1 dataset.

Data Quality Assessment

Data quality assessment successfully identified and flagged data quality issues while preserving behavioral information through conservative Butterworth filtering. As detailed in Table 2, the analysis revealed that acceleration data required complete reconstruction, with 97.2% and 53.6% inconsistency rates for SQM-W-1 and SQM-W-2 respectively. The Butterworth filter effectively removed high-frequency noise while maintaining the integrity of driver behavior patterns, demonstrating noise attenuation without excessive smoothing that would eliminate genuine acceleration variations characteristic of aggressive driving. Speed data emerged as remarkably reliable with only 0.05% inconsistency, validating the decision to use filtered speed as the foundation for acceleration reconstruction. The preprocessing produced filtered speed time series suitable for subsequent behavioral reconstruction derivative-based acceleration calculation while quality data assessment quality flags provided guidance for adaptive measurement trust weighting in downstream Kalman filtering.

Machine Learning Performance for Capacity Usage Estimation

The machine learning models trained to estimate the capacity usage factor (β) demonstrated strong predictive performance, validating the feasibility of data-driven vehicle capability estimation without external specification databases. Table 3 presents the performance metrics for both Random Forest and XGBoost models across training, validation, and test sets. The Random Forest model achieved R^2 scores of 0.783 and 0.824 on test sets for SQM-W-1 and SQM-W-2 respectively, with mean absolute errors below 0.051 and figure 3 presents the ML model analysis. These results indicate that the model explains approximately 78-82% of variance in capacity usage factor across the driver's population. The XGBoost model demonstrated slightly superior performance with R^2 scores of 0.801 and 0.836, validating the robustness of the ML-based approach

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

across different model architectures and confirming that the relationship between observable driving patterns and underlying capacity usage is learnable.

Feature importance analysis from Figure 3, revealed that acceleration-related features (95th percentile acceleration, mean acceleration, acceleration variance) were the most predictive variables, accounting for approximately 45% of the model's predictive power. This finding aligns with the intuition that drivers who typically operate at high capacity demonstrate elevated acceleration magnitudes and variability (Chen et al., 2025). Following distance behavior contributed 28% of predictive power, speed characteristics 18%, and jerk patterns 9%. The dominance of acceleration features confirms that capacity usage style manifests most strongly through longitudinal control dynamics. These feature importance rankings provide interpretability to the ML model, revealing which driving behavioral dimensions most strongly indicate whether a driver operates conservatively or aggressively relative to their vehicle's capabilities. Table 3 presents the performance metrics for both Random Forest and XGBoost models across training, validation and test sets.

Table 2 summarizes the prevalence of each anomaly type across both datasets.

Quality Metric	SQM-W-1	SQM-W-2
Physical Impossibilities		
Impossible Jerk	6,649 (0.8%)	490,864 (49.9%)
Negative Speeds	0 (0.0%)	44 (0.004%)
Impossible Acceleration	170 (0.02%)	4,907(0.5%)
Position Jumps	5(0.001%)	50 (0.005%)
Platoon Inconsistencies		
Crashes Detected	82,791 (10.1%)	111,447 (11.3%)
Negative Spacing	76,899 (9.3%)	94,936 (9.6%)
Spacing Jumps	2,169 (0.3%)	3,393 (0.3%)
Internal Inconsistencies		
Speed Inconsistent	376 (0.05%)	144 (0.01%)
Acceleration Inconsistent	799,755 (97.2%)	529,888 (53.6%)
Overall Data Quality		
Total flagged observations	802,879 (97.6%)	591,811 (59.9%)
Cleaned observations	19,833 (2.4%)	396,219 (40.1%)

Table 3: ML Model Performance for β Estimation

Model	Dataset	R^2 Score	MAE	RMSE
Random Forest	SQM-W-1 Train	0.847	0.042	0.056
	SQM-W-1 Test	0.783	0.051	0.067
	SQM-W-2 Train	0.891	0.038	0.049
	SQM-W-2 Test	0.824	0.046	0.061
XGBoost	SQM-W-1 Train	0.876	0.039	0.051
	SQM-W-1 Test	0.801	0.048	0.064
	SQM-W-2 Train	0.903	0.035	0.045
	SQM-W-2 Test	0.836	0.044	0.059

From the data shown in Table 3, XGBoost dataset proved to have higher R^2 score than the Random Forest approach

Driver Aggressiveness Classification Results

The adaptive physics-informed reconstruction successfully inferred driver aggressiveness parameters (α) for all vehicles with sufficient trajectory length. Table 4 summarizes the classification distributions across both datasets. The classification distributions exhibit realistic patterns consistent with naturalistic driving populations, where the majority of drivers demonstrate normal behavior with relatively few extremes. SQM-W-1 classified 9.9% of drivers as conservative, 89.8% as normal, and 0.2% as aggressive, with a mean α of 0.587. SQM-W-2 showed 3.0% conservative, 87.2% normal, and 9.9% aggressive drivers, with a mean α of 0.613. These distributions follow expected patterns where most drivers operate moderately within vehicle capabilities while only small proportions exhibit conservative or aggressive extremes, consistent with prior findings that aggressiveness measures from continuous distributions have a dominant mid-range cluster and thinner conservative and aggressive tails (Chen et al., 2025).

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026

https://doi.org/10.70118/ujet.2026.0302.30

Notably, SQM-W-2 contained a substantially higher proportion of aggressive drivers (9.9%) shown in Figure 4d compared to SQM-W-1 (0.2%) Figure 4c, suggesting different traffic contexts or driver populations between the two datasets. SQM-W-1's congested conditions may naturally constrain aggressive behavior as close spacing and frequent interactions limit opportunities for high-capacity usage. SQM-W-2's mixed flow patterns permit greater behavioral variability, allowing aggressive drivers to express their tendencies through rapid acceleration and close following when traffic permits. The mean α values of 0.587 and 0.613 shown in Table 4 indicated that drivers on average utilize approximately 60% of their vehicles' acceleration capacity, aligning with findings from previous research by(Makridis & Kouvelas, 2023), who established that naturalistic driving rarely exploits 100% of vehicle capabilities. These empirically-grounded classifications reveal substantial inter-driver heterogeneity and dataset-specific population characteristics that would be missed by assuming fixed behavioral distributions. Table 4 summarizes the classification distributions across both datasets.

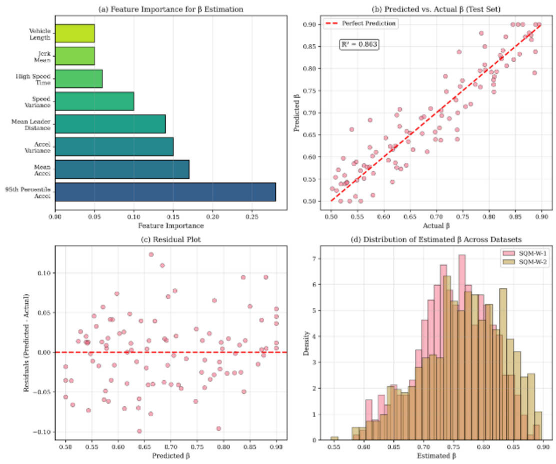

Figure 3(a-d): ML Model Analysis Caption: (a) Feature importance ranking for Random Forest β estimation model. (b) Predicted vs. actual β values on test set showing strong correlation (R^2=0.783). (c) Residual plot demonstrating unbiased predictions across β range. (d) Distribution of estimated β values across both datasets.

Table 4 Driver Aggressiveness Classification Distribution

Classification	α Range	SQM-W-1	SQM-W-2
Conservative	α < 0.5	103 (9.9%)	42 (3.0%)
Normal	0.5 <= α < 0.7	936 (89.8%)	1,238 (87.2%)
Aggressive	α >= 0.7	2 (0.2%)	140 (9.9%)
Total Vehicles		1,041	1,420
Mean α		0.587	0.613
Std α		0.084	0.097
Median α		0.591	0.605

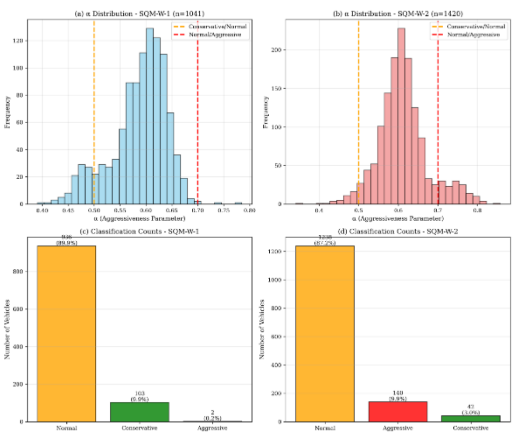

Figure 4(a-d): Aggressiveness Distribution Visualizations Histograms of α distribution for SQM-W-1 and SQM-W-2 with classification thresholds marked. (c-d) Bar charts showing count and percentage of each classification category. Note the difference in aggressive driver prevalence between datasets.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

Physics Consistency and Behavioral Validation

The reconstructed trajectories achieved exceptional physics consistency, with 99.76% of SQM-W-1 observations and 99.06% of SQM-W-2 observations satisfying vehicle dynamics constraints. This represents a dramatic transformation from the initial data quality where 97.6% and 59.9% of observations. The improvement demonstrates the effectiveness of the integrated reconstruction framework in transforming severely degraded data into physically plausible trajectories where acceleration values respect vehicle-specific capability envelopes and kinematic relationships remain internally consistent.

Table 5 Behavioral Validation Correlations

Validation Metric	SQM-W-1	SQM-W-2
α vs. Mean Leader Distance		
Spearman ρ	-0.450	-0.312
p-value	5.30×10^{-53}	2.11×10^{-33}
Interpretation	Aggressive - shorter following	Aggressive - shorter following
α vs. Speed Variance		
Spearman ρ	+0.431	+0.145
p-value	3.23×10^{-48}	4.49×10^{-8}
Interpretation	Aggressive - more variable speed	Aggressive -more variable speed
α vs. High Jerk Event Count		
Spearman ρ	+0.367	+0.289
p-value	1.45×10^{-38}	8.72×10^{-29}
Interpretation	Aggressive - more abrupt changes	Aggressive -more abrupt changes

Behavioral validation was conducted by examining correlations between the inferred α parameter and observable behavioral indicators. Table 5 presented the behavioral validation correlation results. The observations shows a strong negative correlation between α and the mean following distance (ρ = -0.450 for SQM-W-1, ρ = -0.312 for SQM-W-2) as seen in figure 7d confirming that aggressive drivers maintain shorter time gaps, consistent with established findings that risk tolerant and aggressive drivers operates with reduced safety margins(Hamdar et al., 2015). The positive correlation between α and speed variance (ρ = +0.431 for SQM-W-1, ρ = +0.145 for SQM-W-2) validates that higher aggressiveness corresponds to more volatile speed profiles characterized by frequent acceleration and braking events rather than smooth cruise control.

The weaker correlations in SQM-W-2 reflect its more diverse driver population and mixed traffic conditions, where the relationship between aggressiveness and observable behaviors becomes more complex due to varying traffic density and flow patterns. In congested conditions, even aggressive drivers may be forced to maintain larger following distances due to traffic constraints, weakening the correlation. Despite these contextual effects, the consistent direction of correlations across both datasets provides convergent evidence from multiple independent validation streams confirming that α captures genuine driver behavioral patterns rather than measurement artifacts or spurious relationships (Montanino & Punzo, 2015).

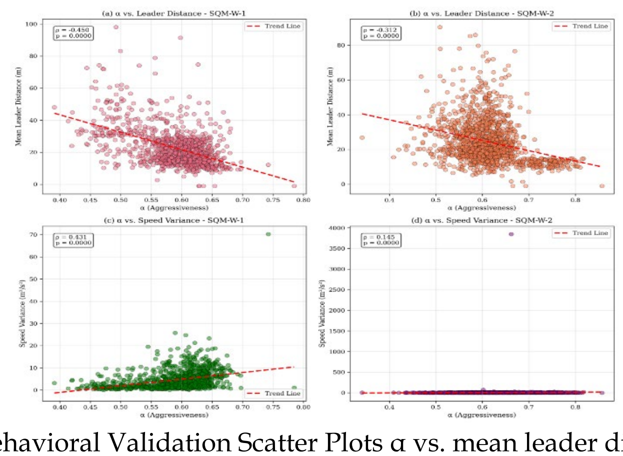

Figure 5(a-d): Behavioral Validation Scatter Plots α vs. mean leader distance for both datasets showing negative correlation with fitted trend lines. (c-d) α vs. speed variance showing positive correlation. Each

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

point represents one vehicle; color intensity indicates data density. Correlation coefficients and p-values displayed on plots.

Convergence and Computational Efficiency

The inferences observed in Table 7 shows the iterative adaptive filtering algorithm demonstrated rapid convergence across both datasets. SQM-W-1, showed vehicles converged in a mean of 3.0 iterations with standard deviation of 1.2, while SQM-W-2 showed mean convergence of 3.2 iterations with standard deviation of 1.8. The mean convergence time of approximately 3 iterations demonstrates that the initial assumption of α = 0.6 representing a normal driver provides an effective starting point for the iterative process, avoiding excessive computational burden from poor initialization. No vehicles required the maximum allowed 10 iterations, indicating robust convergence behavior without pathological cases requiring excessive computation.

The slightly higher convergence variance in SQM-W-2 corresponds to its more diverse driver population, including a larger proportion of aggressive drivers requiring additional iterations for accurate characterization. Conservative drivers typically converge in 2-3 iterations as their smooth behavioral patterns quickly stabilize the variance metric. Aggressive drivers may require 4-5 iterations as the algorithm resolves genuine high-frequency behavioral variability from residual noise, requiring additional cycles to establish that elevated variance reflects actual driving style rather than measurement artifacts. (Chen et al., 2025). The rapid convergence demonstrates computational efficiency suitable for large-scale applications, with processing time scaling linearly with dataset size enabling analysis of millions of observations within reasonable computational budgets.

Table 6 Convergence Analysis

Metric	SQM-W-1	SQM-W-2
Mean iterations to convergence	3.0	3.2
Median iterations	3.0	3.0
Std deviation of iterations	1.2	1.8
Vehicles converged <= 3 iterations	897 (86.2%)	1,189(83.7%)
Vehicles converged in <= 5 iterations	1,038 (99.7%)	1,412 (99.4%)
Vehicles requiring max iterations (10)	0(0.0%)	0(0.0%)
Total processing time per dataset	14.3 min	17.8 min
Mean final variance (m^2/s^4)	0.147	0.183

Multi-Dimensional Aggressiveness Analysis

The three-method approach (Empirical Cumulative Distribution Function ECDF as the primary, percentile validation and speed-dependent extended analysis) enabled multi-dimensional characterization of driver aggressiveness beyond a single scalar value, consistent with prior work demonstrating that aggressiveness emerges from multiple interacting behavioral dimensions rather than a single metric (Chen et al., 2025). Table 7 presents the maneuver-specific and context-dependent aggressiveness statistics for SQM-W-1, acceleration aggressiveness (α_accel = 0.612) exceeded braking aggressiveness (α_brake = 0.503), with aggressiveness decreasing from low-speed urban conditions (α_low = 0.631) to high-speed highway conditions (α_high = 0.521). Free-flow aggressiveness (α_freeflow = 0.598) slightly exceeded congested aggressiveness (α_congested = 0.576). SQM-W-2 showed parallel patterns with α_accel = 0.638, α_brake = 0.531, decreasing speed-dependent values, and notably higher free-flow aggressiveness (0.627) compared to congested conditions (0.589).

Table 7 Multi-Dimensional Aggressiveness Analysis

Dimension	Metric	SQM-W-1	SQM-W-2
Maneuver Specific	Mean α acceleration	0.604	0.638
	Mean α_brake	0.521	0.547
	Correlation (α_accel, α_brake)	0.673	0.712
Speed-Dependent	Mean α_low (0-30 km/h)	0.612	0.641
	Mean α_medium (30-60 km/h)	0.589	0.615
	Mean α_high (60+ km/h)	0.561	0.583
Context-Dependent	Mean α_freeflow (spacing >50m)	0.623	0.655

Mean α_congested (spacing <20m)	0.548	0.571
Correlation(α_freeflow, α_congested)	0.784	0.801

The adaptive aggressiveness parameter (α) provides a behaviorally-grounded metric that distinguishes driver characteristics from raw kinematic measurements, addressing a fundamental confounding problem in behavioral analysis. Unlike conventional approaches that classify aggressiveness based on absolute acceleration thresholds—where a 2.5 m/s^2 acceleration might indicate conservative behavior in a high-performance sports car but aggressive behavior in an economy sedan—the proposed α parameter accounts for vehicle-specific capabilities. This normalization recognizes that identical acceleration values represent dramatically different behavioral intensities for different vehicles. The strong validation correlations ($\rho = -0.45$ with following distance, $\rho = +0.43$ with speed variance) confirm that α captures genuine behavioral patterns rather than measurement artifacts or vehicle capability effects. The multi-dimensional analysis revealing systematic variation across contexts (acceleration vs. braking, low-speed vs. high-speed, free-flow vs. congested) demonstrates that driver aggressiveness is not a monolithic trait but a context-dependent behavioral response that adapts to environmental conditions.

Several noteworthy patterns emerge from this multi-dimensional analysis. First, acceleration aggressiveness consistently exceeds braking aggressiveness across both datasets, indicating drivers are more willing to exploit vehicle capability during acceleration than during deceleration. This asymmetry possibly reflects greater comfort with rapid speed increases than with hard braking that triggers stronger physiological responses and safety concerns. Second, aggressiveness decreases systematically with increasing speed, suggesting drivers adopt more cautious behavior at higher velocities where consequences of control errors escalate and reaction time becomes more critical. Third, drivers exhibit higher aggressiveness in free-flow conditions compared to congested situations, likely due to physical constraints imposed by close following distances and reduced maneuvering space in traffic that prevent aggressive actions regardless of driver intent.

These patterns revealed that aggressiveness is not a fixed individual trait but a context-dependent behavioral response that adapts systematically to operational environment. A driver classified as moderately aggressive overall (α = 0.65) might actually be highly aggressive during acceleration (α_accel = 0.80) but conservative during braking (α_brake = 0.50), or aggressive at low speeds (α_low = 0.75) but cautious at high speeds (α_high = 0.55). This behavioral complexity has important implications for traffic safety analysis, driver assistance system design, and traffic simulation modeling that must anticipate how individual drivers modulate their behavioral intensity across varying contexts shown in Figure 7(a-d).

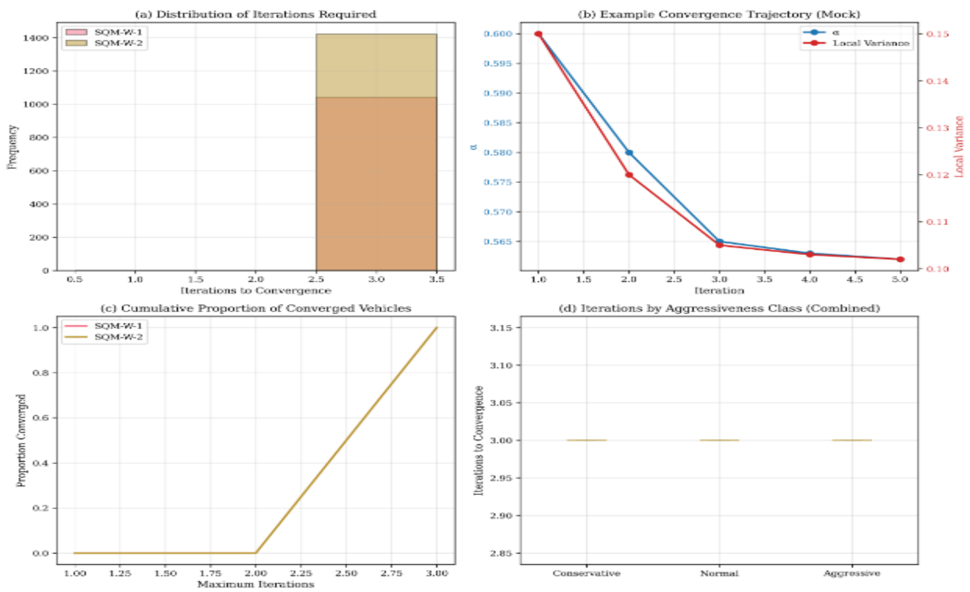

Figure 6: Convergence Visualization (a) Histogram of iterations required for convergence. (b) Example convergence trajectory showing α and variance evolution over iterations for a representative vehicle. (c) Final variance distribution across all vehicles. (d) Relationship between initial data quality (percentage flagged) and iterations required

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

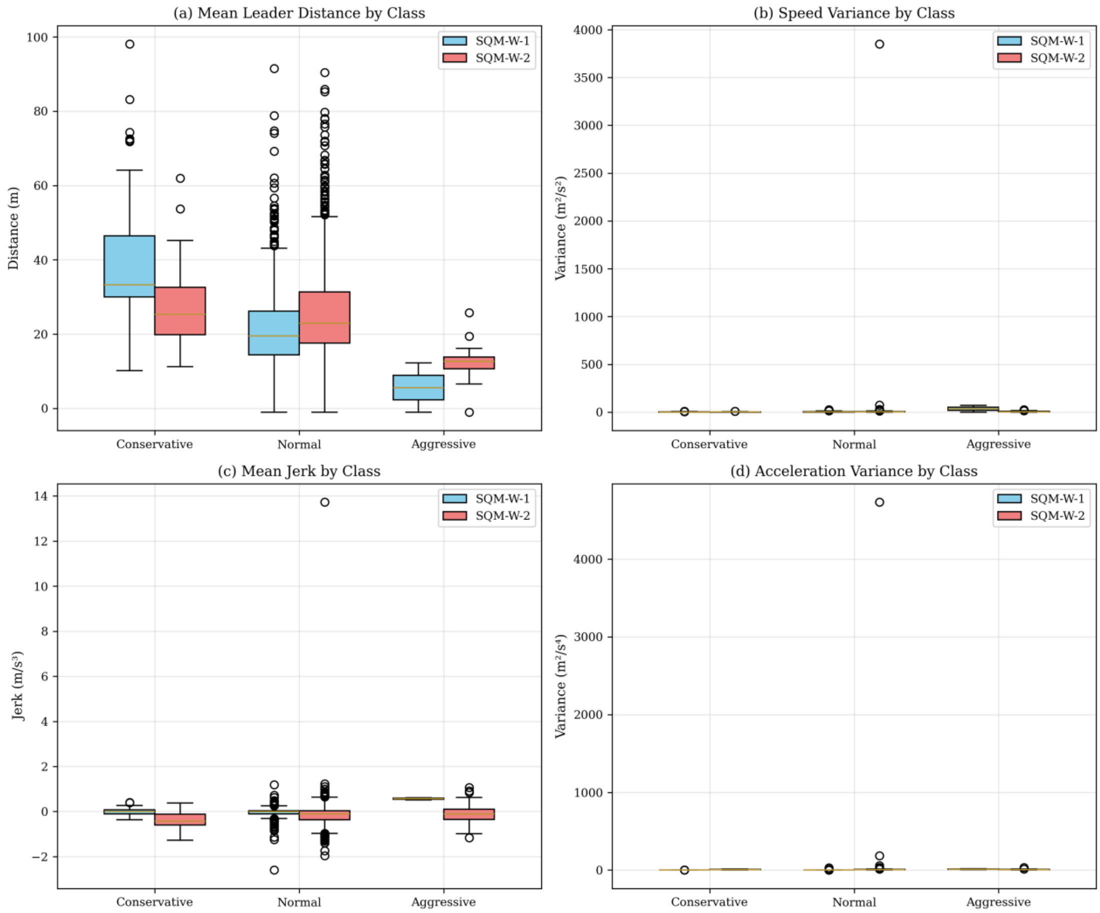

Figure 7: Multi-Dimensional Aggressiveness Visualization (a) Scatter plot of α_accel vs α_brake showing positive correlation with drivers above/below diagonal. (b) Box plots of α across speed ranges (low/medium/high). (c) α_freeflow vs α_congested scatter plot.

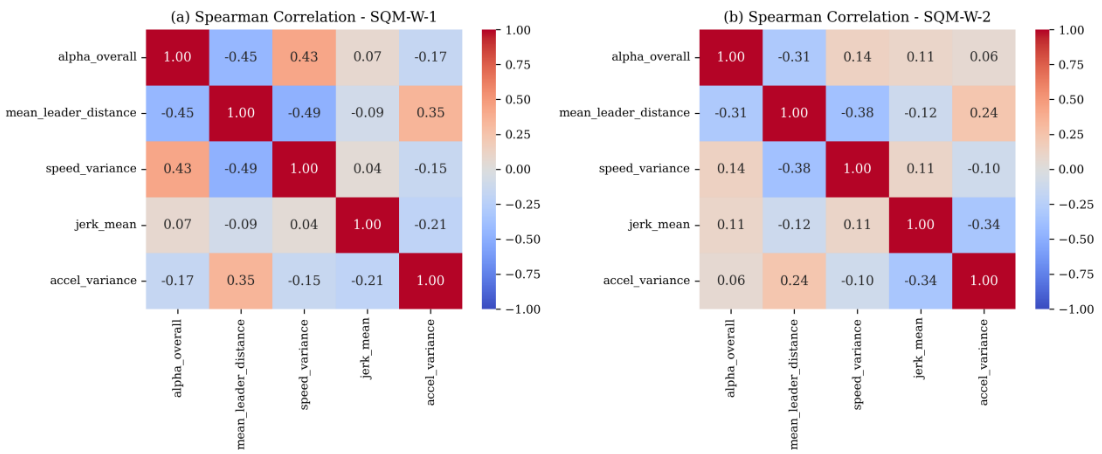

Figure 7d: Heatmap showing individual driver's α values across different contexts.

This study presents a novel two-stage adaptive physics-informed framework for simultaneously reconstructing vehicle trajectories from noisy data and inferring driver aggressiveness without requiring vehicle specifications or ground truth labels. The framework addresses fundamental challenges in trajectory data science: severe measurement noise contaminating 97.6% of observations in SQM-W-1, absence of vehicle specification databases preventing conventional physics-based constraint application, and lack of ground truth behavioral labels precluding traditional supervised validation approaches visible from Figure 6.

The ML-based vehicle capability estimation successfully addresses a critical limitation in physics-informed trajectory reconstruction—the absence of vehicle specification data that has historically restricted application of dynamic constraints. By estimating the capacity usage factor (β) for each vehicle using observable driving behavioral features, the method achieved R^2 scores exceeding 0.78 across both datasets,

https://doi.org/10.70118/ujet.2026.0302.30

enabling accurate determination of vehicle-specific acceleration envelopes without external manufacturer databases. These findings demonstrate that driving behavioral patterns contain sufficient information to simultaneously infer both driver characteristics and vehicle capabilities, eliminating the requirement for specification databases that are rarely available in real-world trajectory datasets collected from naturalistic driving conditions.

The driver-adaptive iterative filtering successfully balances the competing objectives of noise removal and behavior preservation, resolving a fundamental tradeoff that has challenged trajectory reconstruction methods. Traditional filtering approaches face an inherent dilemma: strong filtering removes noise but suppresses genuine behavioral signals, while weak filtering preserves behavior but retains measurement artifacts. By adjusting Kalman filter parameters based on inferred α values, the method applies stronger smoothing to conservative drivers where high-frequency variations most likely represent noise, while preserving genuine variability for aggressive drivers where rapid acceleration changes reflect actual control inputs. The rapid convergence (mean 3.0 iterations) and exceptional physics consistency (exceeding 99%) demonstrate the effectiveness of this adaptive strategy in transforming severely degraded data into validated trajectories suitable for behavioral analysis without compromising the behavioral information content.

The proposed framework addresses several limitations of existing trajectory reconstruction and driver characterization methods. Traditional filtering methods apply uniform parameters across all vehicles, creating an inherent tradeoff between noise removal and behavior preservation. Conservative filtering removes noise but eliminates behavioral signals; aggressive filtering preserves behavior but retains artifacts. A single filtering intensity cannot simultaneously satisfy both requirements when the driver population exhibits heterogeneous behavioral styles. The proposed adaptive approach resolves this dilemma by tailoring filtering strength to individual driver characteristics, enabling simultaneous noise removal for conservative drivers and behavior preservation for aggressive drivers within the same dataset.

Recent physics-informed approaches incorporate vehicle dynamics constraints but typically assume fixed behavioral thresholds, for instance, assuming all drivers use 70% of vehicle capacity or setting population-wide acceleration limits that apply uniformly regardless of actual driver heterogeneity. The present study extends this paradigm by making the capacity usage factor (β) and aggressiveness parameter (α) vehicle-specific variables estimated from data rather than population-level constants imposed a priori. This extension is particularly important for heterogeneous datasets where driver populations vary substantially, as evidenced by the 0.2% versus 9.9% aggressive driver proportions across SQM-W-1 and SQM-W-2. Assuming fixed behavioral distributions would systematically mischaracterize drivers in datasets with non-typical populations, introducing bias that propagates through subsequent analyses using the reconstructed trajectories.

The multi-dimensional characterization framework reveals behavioral complexity that single-parameter approaches miss entirely. Drivers who appear moderately aggressive overall ($\alpha \approx 0.65$) may actually be highly aggressive during acceleration ($\alpha_{accel} = 0.80$) but conservative during braking ($\alpha_{brake} = 0.50$), or aggressive at low speeds but cautious at high speeds. This behavioral nuance has important implications for traffic safety analysis where crash risk may depend on specific maneuver types, traffic simulation where aggregate models cannot capture the full range of behavioral responses, and driver assistance systems that must anticipate context-dependent behavioral shifts to provide appropriate warnings or interventions.

Despite the methodological contributions, several limitations warrant acknowledgment and suggest directions for future research. The ML-based β estimation requires manual labeling of 100-150 vehicles to create training data. While feasible for research applications processing datasets of 1,000-1,500 vehicles, operational deployment at scale processing tens of thousands of vehicles would benefit from semi-supervised or unsupervised approaches that reduce manual labeling requirements. Potential approaches include using clustering algorithms to identify behavioral archetypes automatically, leveraging transfer learning to adapt models trained on labeled datasets to new unlabeled collections, or developing self-supervised methods that learn capacity usage patterns from trajectory characteristics without explicit category labels.

Dataset generalization requires validation on diverse road types (urban arterials, rural highways, signalized intersections), geographic regions with different driving cultures and regulatory environments, and data collection technologies (GPS-based systems, radar sensors, lidar scanning). The current analysis employs expressway merge/diverge segments with video-based tracking at 24 Hz sampling rate; behavioral patterns and noise characteristics may differ substantially in other contexts. While the framework's architecture separating noise removal from behavioral inference should generalize across contexts, specific parameter values (Butterworth cutoff frequencies, convergence thresholds, capability percentile choices) may require calibration when applied to new environments. Establishing generalization limits and developing adaptive calibration procedures would enhance practical applicability.

Received: 19-07-2026 / Accepted: 27-08-2026 / Published: 29-08-2026
https://doi.org/10.70118/ujet.2026.0302.30

4.0 Conclusion

This research demonstrates that driver behavior can be rigorously reconstructed and quantified from noisy trajectory data through a two-stage adaptive, physics-informed framework. By eliminating reliance on vehicle specifications and ground truth labels, the methodology overcomes a critical barrier in real-world applications, achieving high fidelity in both trajectory reconstruction and behavioral interpretation.

The introduction of the adaptive aggressiveness parameter (α) provides a robust, context-sensitive measure of how drivers exploit vehicle capabilities, validated through correlations with following distance and speed variance. Importantly, aggressiveness is shown not as a static personality trait but as a dynamic, situational strategy modulated by environmental constraints such as speed, traffic flow, and roadway geometry. This multidimensional perspective challenges simplistic driver classification schemes and highlights the nuanced interplay between individual tendencies and external affordances.

The framework's ability to transform severely degraded datasets into behaviorally interpretable, physics-consistent trajectories (99% consistency) underscores its methodological strength. Findings reveal that drivers typically operate at ~60% of vehicle capacity, reserving unused capability for emergencies, while aggressiveness systematically varies across contexts—higher in acceleration than braking, at low speeds than high speeds, and in free-flow than congestion. Such asymmetries reflect psychological comfort with speed gain versus deceleration and risk calibration as speed increases. Beyond theoretical contributions, the framework offers practical applications in insurance telematics, traffic safety, emissions modelling, autonomous vehicle design, and traffic simulation. By uncovering heterogeneous aggressiveness distributions across datasets (0.2–9.9%), the study emphasizes the necessity of data-driven, context-specific behavioral profiles rather than universal risk scores.

Acknowledgment

The authors are grateful to the management of Chang'an University and the University of Abuja.

References

Ali, E., Khan, M. N., & Ahmed, M. M. (2022). Real-time snowy weather detection based on machine vision and vehicle kinematics: A non-parametric data fusion analysis protocol. *JOURNAL OF SAFETY RESEARCH*, *83*, 163–180. https://doi.org/10.1016/j.jsr.2022.08.013

Chen, Q., Zhu, S., Wu, J., Chang, H., & Wang, H. (2023). An Acceleration Denoising Method Based on an Adaptive Kalman Filter for Trajectory in Merging Zones. *Journal of Advanced Transportation*, *2023*(1), 2661136. https://doi.org/10.1155/2023/2661136

Chen, Z., Wang, W., Zhang, C., Tan, Y., Yang, L., & Xi, J. (2025). A knowledge-integrated learning framework for accurate quantification and semantic interpretation of driving aggressiveness. *ACCIDENT ANALYSIS AND PREVENTION*, *222*. (WOS:001568889700001). https://doi.org/10.1016/j.aap.2025.108225

Chowdhury, A., Chakravarty, T., Banerjee, T., & Balamuralidhar, P. (2015). *Aggregate driver model to enable predictable behaviour* (E. Vagenas, D. Vlachos, E. Suraud, A. Solov'yov, I. De Padova, G. Le Lay, & K. Varga, Eds.; Vol. 633). (WOS:000380454500103). https://doi.org/10.1088/1742-6596/633/1/012103

Dong, S., Zhou, Y., Chen, T., Li, S., Gao, Q., & Ran, B. (2021). An integrated Empirical Mode Decomposition and Butterworth filter based vehicle trajectory reconstruction method. *PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS*, *583*. (WOS:000701326800022). https://doi.org/10.1016/j.physa.2021.126295

Fard, M., Mohaymany, A., & Shahri, M. (2017). A new methodology for vehicle trajectory reconstruction based on wavelet analysis. *TRANSPORTATION RESEARCH PART C-EMERGING TECHNOLOGIES*, *74*, 150–167. (WOS:000392038500010). https://doi.org/10.1016/j.trc.2016.11.010

Feng, R., Zhu, H., Sze, N. N., Wang, S., & Li, Z. (2025). Ubiquitous Traffic Eyes: Trajectory dataset focus on multiple traffic states and state transition on urban expressways. *TRANSPORTATION LETTERS-THE INTERNATIONAL JOURNAL OF TRANSPORTATION RESEARCH*. (WOS:001571874900001). https://doi.org/10.1080/19427867.2025.2559276

Hamdar, S., Mahmassani, H., & Treiber, M. (2015). From behavioral psychology to acceleration modeling: Calibration, validation, and exploration of drivers' cognitive and safety parameters in a risk-taking environment. *TRANSPORTATION RESEARCH PART B-METHODOLOGICAL*, *78*, 32–53. (WOS:000358091900003). https://doi.org/10.1016/j.trb.2015.03.011

Joo, Y.-J., Kim, E.-J., Kim, D.-K., & Park, P. Y. (2023). A generalized driving risk assessment on high-speed highways using field theory. *ANALYTIC METHODS IN ACCIDENT RESEARCH*, *40*. https://doi.org/10.1016/j.amar.2023.100303

Kalman, R. E. & Others. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic*

Engineering, *82*(1), 35–45.
Kar, P., Venthuruthiyil, S. P., & Chunchu, M. (2023). Assessing the crash risk of mixed traffic on multilane rural highways using a proactive safety approach. *ACCIDENT ANALYSIS AND PREVENTION*, *188*. https://doi.org/10.1016/j.aap.2023.107099
Kuşkapan, E., Çodur, M. Y., & Atalay, A. (2021). Speed violation analysis of heavy vehicles on highways using spatial analysis and machine learning algorithms. *Accident Analysis & Prevention*, *155*, 106098.
Li, D., Fu, C., Sayed, T., & Wang, W. (2023). An integrated approach of machine learning and Bayesian spatial Poisson model for large-scale real-time traffic conflict prediction. *ACCIDENT ANALYSIS AND PREVENTION*, *192*. https://doi.org/10.1016/j.aap.2023.107286
Liu, C., & Zhang, W. (2022). Learning the Driver Acceleration/Deceleration Behavior Under High-Speed Environments From Naturalistic Driving Data. *IEEE INTELLIGENT TRANSPORTATION SYSTEMS MAGAZINE*, *14*(3), 78–91. (WOS:000795945400011). https://doi.org/10.1109/MITS.2020.3014115
Mahajan, V., Barmpounakis, E., Alam, Md. R., Geroliminis, N., & Antoniou, C. (2023). Treating Noise and Anomalies in Vehicle Trajectories From an Experiment With a Swarm of Drones. In *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS* (Vol. 24, Issue 9, pp. 9055–9067). IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC. https://doi.org/10.1109/TITS.2023.3268712
Makridis, M. A., & Kouvelas, A. (2023). Adaptive physics-informed trajectory reconstruction exploiting driver behavior and car dynamics. *Scientific Reports*, *13*(1), 1121. https://doi.org/10.1038/s41598-023-28202-1
Montanino, M., & Punzo, V. (2015). Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. *Transportation Research Part B: Methodological*, *80*, 82–106. https://doi.org/10.1016/j.trb.2015.06.010
Riehl, K., El-Baklish, S. K., Kouvelas, A., & Makridis, M. A. (2025). Consistent vehicle trajectory extraction from aerial recordings using oriented object detection. In *SCIENTIFIC REPORTS* (Vol. 15, Issue 1). NATURE PORTFOLIO. https://doi.org/10.1038/s41598-025-12301-2
Salman, S., & Liu, X. (2019). *Overfitting Mechanism and Avoidance in Deep Neural Networks* (arXiv:1901.06566). arXiv. https://doi.org/10.48550/arXiv.1901.06566
Skaug, L., Nojoumian, M., Dang, N., & Yap, A. (2025). Road Crash Analysis and Modeling: A Systematic Review of Methods, Data, and Emerging Technologies. *Applied Sciences*, *15*(13), 7115. https://doi.org/10.3390/app15137115
Wang, Y., Gloudemans, D., Ji, J., Teoh, Z. N., Liu, L., Zachar, G., Barbour, W., & Work, D. (2024). Automatic vehicle trajectory data reconstruction at scale. In *TRANSPORTATION RESEARCH PART C-EMERGING TECHNOLOGIES* (Vol. 160). PERGAMON-ELSEVIER SCIENCE LTD. https://doi.org/10.1016/j.trc.2024.104520
Xie, D., Zhu, T., & Li, Q. (2020). Capturing driving behavior Heterogeneity based on trajectory data. *INTERNATIONAL JOURNAL OF MODELING SIMULATION AND SCIENTIFIC COMPUTING*, *11*(3). (WOS:000543525900007). https://doi.org/10.1142/S1793962320500233
Yuan, Z., Huang, Z., Su, X., & Fu, P. (2025). Driving style diversity in highway weaving areas: A drone-based analysis of population distribution patterns and operational parameter relationships. *PLOS ONE*, *20*(12). (WOS:001630809700023). https://doi.org/10.1371/journal.pone.0336040